On continuous-time generalized AR(1) processes: models, statistical inference, and applications to non-normal time series.

by

Rong Zhu

B.Sc., USTC 1986

M.Sc., USTC 1988

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Department of Statistics)

we accept this thesis as conforming
to the required standard

The University of British Columbia

March 2002

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics

The University of British Columbia
Vancouver, Canada

Date March 28, 2002

Abstract

This thesis develops the theory of continuous-time generalized AR(1) processes and presents their use for non-normal time series models. The theory is of dual interest in probability and statistics. From the probabilistic viewpoint, this study generalizes a type of Markov process which has a similar representation structure to the Ornstein-Uhlenbeck process (or continuous-time Gaussian AR(1) process). However, the stationary distributions can now have support on non-negative integers, or positive reals, or reals; the dependence structures are no longer restricted to be linear. From the statistical viewpoint, this study is dedicated to modelling unequally-spaced or equally-spaced non-normal time series with non-negative integer, or positive, or real-valued observations. The research on both the probabilistic and statistical sides contribute to a complete modelling procedure which consists of model construction, choice and diagnosis.

The main contributions in this thesis include the following new concepts: self-generalized distributions, extended-thinning operators, generalized Ornstein-Uhlenbeck stochastic differential equations, continuous-time generalized AR(1) processes, generalized self-decomposability, generalized discrete self-decomposability, P-P plots and diagonal P-P plots. These concepts play crucial roles in the newly developed theory.

We take a dynamic view to construct the continuous-time stochastic processes. Part II is devoted to the construction of the continuous-time generalized AR(1) process, which is obtained from the generalized Ornstein-Uhlenbeck stochastic differential equation, and the proposed stochastic integral. The resulting continuous-time generalized AR(1) process consists of a dependent term and an innovation term. The dependent term involves an extended-thinning stochastic operation which

generalizes the commonly used operation of constant multiplier. Such a Markov process can have a simple interpretation in modelling non-normal time series. In addition, the family of continuous-time generalized AR(1) processes is surprisingly rich. Both stationary and non-stationary situations of the process are considered.

In Part III, we answer the question of what kind of stationary distributions are obtained from the family of continuous-time generalized AR(1) processes, as well as the converse question of whether a specific distribution can be the stationary distribution of a continuous-time generalized AR(1) process. This leads to the characterization of distributions according to the extended-thinning operations. The characterization results are meaningful in statistical modelling, because under steady state, the marginal distributions of a Markov process are the same as the stationary distribution. They will guide us to choose appropriate processes to model a non-normal time series. The probabilistic study also shows that the autocorrelation function is of exponential form in the time difference, like that of the Ornstein-Uhlenbeck or Ornstein-Uhlenbeck-type process.

Part IV deals with statistical inference and modelling. We have studied parameter estimation for various situations such as equally-spaced time, unequally-spaced time, finite marginal mean, infinite marginal mean, and so on. The graphical tools, the P-P plot and diagonal P-P plot, are proposed for use in identifying the marginal distribution and serial dependence, and diagnosing the fitted model. Three data examples are given to illustrate the new modelling procedure, and the application capacity of this theory of continuous-time generalized AR(1) processes. These time series are non-negative integer or positive-valued, with equally-spaced or unequally-spaced time observations.

Contents

Abstract **ii**

Contents **iv**

List of Tables **ix**

List of Figures **xi**

Notations, abbreviations and conventions **xv**

Acknowledgements **xviii**

I Introduction **1**

1 Overview **2**

1.1 Motivation and literature review 2

1.2 Highlights of our new research 6

II Theory for model construction **12**

2 Relevant background on characteristic tools, distribution families and stochastic processes **13**

2.1 Ornstein-Uhlenbeck processes and Ornstein-Uhlenbeck-type processes 14

2.2 Characterization tools of distributions and examples 16
2.2.1 Probability generating function 17
2.2.2 Laplace transformation, moment generating function and characteristic function 34
2.3 Particular families of distributions 43
2.3.1 Infinitely divisible, self-decomposable and stable distributions 43
2.3.2 Tweedie exponential dispersion family 50
2.3.3 Generalized convolutions 54
2.4 Independent increment processes and examples 64

3 Self-generalized distributions and extended-thinning operations 72

3.1 Self-generalized distributions 73
3.1.1 Non-negative integer case and examples 73
3.1.2 Positive case and examples 76
3.2 Properties of self-generalized distributions 80
3.3 Construction of new self-generalized distributions 88
3.4 Extended-thinning operation 92

4 Generalized Ornstein–Uhlenbeck stochastic differential equations and their possible solutions 103

4.1 Stochastic differentiation and integration 104
4.2 Generalized Ornstein-Uhlenbeck equations 109
4.3 Explanations, innovation types, non-stationary situations and examples 112
4.4 Construction of possible solutions for the generalized Ornstein-Uhlenbeck SDE . . . 116
4.5 Summary and discussion 126

5 Results for continuous-time generalized AR(1) processes 128

5.1 Main results for continuous-time GAR(1) processes 128
5.2 Non-negative integer innovation processes and examples 137
5.3 Positive-valued innovation processes and examples 152

5.4 Real-valued innovation processes and examples 157
5.5 Tweedie innovation processes 162

III Probabilistic and statistical properties **165**

6 Stationary distributions, steady states and generalized AR(1) time series **166**
6.1 Stationary distributions 167
6.2 Marginal distributions under steady state 169
6.2.1 Non-negative integer margins 170
6.2.2 Positive-valued margins 174
6.2.3 Real-valued margins 176
6.3 Customizing margins 178
6.4 Generalized AR(1) time series 196

7 Characterization of stationary distribution families **202**
7.1 Self-decomposable and discrete self-decomposable classes 203
7.2 Generalized self-decomposable, generalized discrete self-decomposable classes and their infinite divisibility property 220
7.3 Relationships among the classes of generalized self-decomposable and discrete self-decomposable distributions 236

8 Transition and sojourn time **244**
8.1 Infinitesimal transition analysis 245
8.1.1 Non-negative integer margin 245
8.1.2 Positive-valued margin 255
8.2 Characteristic feature of the PDE of the conditional pgf or LT 257
8.2.1 Non-negative integer margin: PDE of the conditional pgf 258
8.2.2 Positive-valued margin: PDE of the conditional LT 263
8.2.3 Summary: margins, self-generalized distribution and increment of innovation 266

8.3 Distributions of sojourn time 268

9 Conditional and joint distributions **271**

9.1 Consistency in process construction: the view from distribution theory 271

9.2 Conditional properties 283

9.3 Joint properties 302

9.3.1 Bivariate distributions 303

9.3.2 Multivariate distributions 314

IV Statistical inference and applications **320**

10 Parameter estimation **321**

10.1 Maximum likelihood estimation 322

10.2 Conditional least squares estimation and variations 325

10.3 Empirical characteristic function estimation approach and variations 341

10.4 Other estimation approaches 346

10.5 Numerical solution of optimization 352

11 Asymptotic study of estimators **354**

11.1 Random sampling scheme, assumptions and fundamental theorem 355

11.2 Asymptotic properties of MLE 362

11.3 Asymptotic properties of conditional least squares estimator 373

12 Autocorrelation detection, model selection, testing, diagnosis, forecasting and process simulation **385**

12.1 Assessing autocorrelation 386

12.2 Model selection 398

12.3 Model diagnostics and hypothesis testing 400

12.3.1 Graphical diagnostic method 400

12.3.2 Parameter testing 404

12.4 Forecasting 405
12.5 Simulation of the continuous-time GAR(1) processes 407

13 Applications and data analyses **415**
13.1 Introduction to modelling procedure 415
13.2 Manuscript data study 417
13.3 WCB claims data study 428
13.4 Ozone data study 444

V Discussion **456**

14 Conclusions and further research topics **457**
14.1 Discussion of continuous-time generalized AR(1) processes 458
14.2 Some thoughts on model construction 459
14.3 Future research 460

Appendix A Data sets **462**
A.1 Manuscripts data 462
A.2 WCB claims data 463
A.3 Abbotsford daily maximum ozone concentrations data 466

Bibliography **468**

List of Tables

2.1 *Summary of Tweedie exponential dispersion models (S = support set).* 54

3.1 *Some results from Theorem 3.3.1.* 89

6.1 *Partial derivative of pgf, $H(s)$, for self-generalized distributions with non-negative integer support.* 181

6.2 *Partial derivative of negative log LT, $H(s)$, for self-generalized distributions with positive support.* 183

6.3 *Conditional pgf $G_{(\alpha)_K \circledast X_{i-1}|X_{i-1}=x}(s)$ when K is a non-negative integer self-generalized random variable.* 198

6.4 *Conditional LT $\phi_{(\alpha)_K \circledast X_{i-1}|X_{i-1}=x}(s)$ when K is a positive self-generalized random variable.* 199

9.1 *Mean and variance of non-negative integer and positive self-generalized random variable $K(\alpha)$.* 287

9.2 *Auto-covariance and auto-correlation function of the stationary continuous-time GAR(1) process associated with known self-generalized random variable $K\left(e^{-\mu(t_2-t_1)}\right)$. Here the variance function is $V(t) = V$.* 305

13.1 *Summary of the frequencies of the number of manuscripts in refereeing queue.* . . . 418

13.2 *Summary of the number of pairs by lag for the manuscripts data.* 419

13.3 *Summary of different estimates of parameter μ and λ in the $GAR(1)$ model for the manuscript data.* 425

13.4 *Summary of the series C3 in WCB claims data.* 431

13.5 *Estimated conditional probabilities:* $\Pr[X(t') = y \mid X(t) = x]$. *The highest probability in each column is highlighted with an asterisk.* 443

13.6 *Estimated conditional cdf:* $\Pr[X(t') \leq y \mid X(t) = x]$. *The median in each column is highlighted with an asterisk.* 443

13.7 *One month predictions:* $\hat{y}_{mode}$, $\hat{y}_{median}$ *and* $\hat{y}_{PI}$. 443

13.8 *Summary of the series of daily maximum ozone concentration.* 444

List of Figures

3.1 *Illustration of* $\{J_K(t); -\infty < t < \infty\}$ *in Cases 1 and 2. (a) corresponds to non-negative integer* X *in Case 1, where dotted vertical lines indicate the discrete time points* $\{0, 1, 2, \ldots\}$*. (b) corresponds to positive* X *in Case 2, where* t *is continuous on* $[0, \infty)$. 95

3.2 *Illustration of* $\{J_1(t); t \geq 0\}$, $\{J_1(t); t \geq 0\}$ *and* $\{J_k(t); -\infty < t < \infty\}$ *in Case 3.* . . 96

4.1 *Illustration of increment in the deterministic and stochastic cases. (a) corresponds to deterministic function* $x(t)$*. (b), (c) and (d) correspond to three different paths of the stochastic process* $\{X(t); t \geq 0\}$. 105

4.2 *Illustration of stochastic integration via infinitesimal partition. (a) and (b) correspond to two different paths of the stochastic process* $\{X(t); t \geq 0\}$. 108

4.3 *Illustration of the geometrical explanation of the extended-thinning operation. (a) corresponds to a constant multiplier* cX*;* X *can be either real or positive-valued. (b) corresponds to a non-negative integer* X*. (c) corresponds to a positive* X. 110

4.4 *Illustration of the mechanism of the generalized Ornstein-Uhlenbeck SDE.* 113

7.1 *The relationship of ID, DSD and GDSD*(**I2**). 238

7.2 *The relationship of ID, SD and GSD*(**P2**). 240

12.1 *Sunflower plots of two time series count data. The left one is from the model in (12.1.1), while the right one is from an independent Poisson*(5) *series.* 389

12.2 *Scatterplots and diagonal P-P plots of positively correlated and negatively correlated bivariate normal data. The left side corresponds to positive correlation, while the right side corresponds to negative one.* 392

12.3 *Diagonal P-P plots of two time series count data. The left one is from the model in (12.1.1), while the right one is from an independent Poisson*(5) *series.* 394

12.4 *Randomized quantile transformation scatterplots of two time series count data. The left one is from the model in (12.1.1), while the right one is from an independent Poisson*(5) *series.* 396

12.5 *The Auto-correlation function (ACF) plot of the count time series from the model in (12.1.1).* 397

12.6 *Simulation of a time series with length* 100 *from (12.5.2) with* $\lambda = 2.15$ *and* $\mu = 0.43$. 409

12.7 *Simulation of time series with length* 100 *from (12.5.3) and (12.5.4). Both processes have Gamma*$(8.1, 0.17)$ *margins.* 411

12.8 *Simulation of a continuous-time path from (12.5.2) with* $\lambda = 2.15$ *and* $\mu = 0.43$. . . 413

13.1 *The time series plot of refereeing queue length of manuscripts.* 419

13.2 *The histogram of the manuscript data, and its P-P plot against* Poisson(2.412). . . 420

13.3 *The ACF plot of the manuscript data. The dotted horizontal line indicates the* 95% *boundary of the estimate of correlation coefficient for* 85 *pairs of independent Poisson*(2.412) *random variables; the boundary is obtained by simulation.* 421

13.4 *The sunflower, randomized quantile transformation, and diagonal P-P plots for the pairs with lag* 1, 2 *and* 3 *months from the manuscript data.* 422

13.5 *The plot of function* $R_{CLS(ME)}(\alpha_0)$ *of the manuscript data.* 424

13.6 *Model diagnosis for manuscript data: diagonal P-P plots for estimates of the CLS(ME) (top row), MLE (middle row) and CLS (bottom row).* 426

13.7 *Model diagnosis for manuscript data: diagonal P-P plots for estimates of the CWLS (top row) and diagonal PLS (bottom row).* 427

13.8 *Model diagnosis for manuscript data: diagonal P-P plots for the estimate* $(\mu, \lambda) = (0.433, 1.04)$. 429

13.9 *The histogram of the series C3 (left), and* 1000 *simulated sample variances from Poisson*(6.13)*; the dotted vertical line is the sample variance of C3 (right).* 431

13.10 *The P-P plots of C3 in WCB claims data against Poisson*(6.13) *(left), NB*(6.64, 0.48) *(middle) and GP*(4.42, 0.28) *(right).* 433

13.11 *The time series plot (top) and ACF plot (bottom) of C3 in WCB claims data.* . . . 434

13.12 *Serial dependence: sunflower plots (1st row), randomized quantile transformation plots (2nd row) and diagonal P-P plots (3rd row) of pairs in C3 with lag 1, 2 and 3 months.* 435

13.13 *Model diagnosis for WCB claims data: diagonal P-P plots of C3 against the fitted (method of moments) NB GAR*(1) *model.* 438

13.14 *Model diagnosis for WCB claims data: diagonal P-P plots of C3 against the fitted (method of moments) GP GAR*(1) *model.* 439

13.15 *Model diagnosis for WCB claims data: diagonal P-P plots of C3 against the fitted (MLE) NB GAR*(1) *model.* 440

13.16 *Model diagnosis for WCB claims data: diagonal P-P plots of C3 against the fitted (MLE) GP GAR*(1) *model.* 441

13.17 *The histogram of the daily maximum ozone concentration, and the P-P plot against* Gamma(8.03, 0.17). 445

13.18 *The time series plot and ACF plot for the daily maximum ozone concentration.* . . 446

13.19 *The scatterplots and diagonal P-P plots of lag one day to three days for the daily maximum ozone concentration.* 447

13.20 *Model diagnosis: diagonal P-P plots of lag one day for* $\gamma = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 *and* 1/1.17 *in the ozone data study.* 450

13.21 *One step ahead predictions (dotted line) by the conditional mean for the daily maximum ozone concentrations (solid line).* 451

13.22 *The analysis of differences between observations and one step ahead predictions by the conditional mean for the daily maximum ozone concentration.* 453

13.23 *Model diagnosis: diagonal P-P plot of lag one day for Model (13.4.3) at* $\widehat{\delta}_{MLE} = 8.31$, $\widehat{\beta}_{MLE} = 0.17$ *and* $\widehat{\alpha}_{MLE} = 0.51$. 455

Notations, abbreviations and conventions

We follow the widely used conventions throughout the thesis. Latin upper-case letters, often X, Y, Z, usually with subscripts or incorporated with other variables such as t, α, are used for random variables. Bold Latin upper-case letters, often $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, are used for random vectors. Greek lower-case letters, say α, β, γ, with or without subscripts, are used for parameters in models. Bold Greek lower-case letters stand for parameter vectors, e.g., $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$. For a vector or matrix, the transpose is indicated with a superscript of T. All vectors are column vectors; hence transposed vectors such as X^T, x^T are row vectors.

List of notation and abbreviations:

$\Re$	$(-\infty, \infty)$
$\Re_0$	$[0, \infty)$
$-\Re_0$	$(-\infty, 0]$
$\Re_+$	$(0, \infty)$
$\Re_-$	$(-\infty, 0)$
$\mathcal{N}$	$\{1, 2, \ldots\}$
$\mathcal{N}_0$	$\{0, 1, 2, \ldots\}$
$\sim$	"distributed as" or "has its distribution in"
$\stackrel{d}{=}$	equality in distribution
$\propto$	proportional to
$\prec^{st}$	stochastically smaller than
$\xrightarrow{L}$	convergence in distribution
$\xrightarrow{P}$	convergence in probability
$\xrightarrow{a.s.}$	convergence almost surely
$G_X(s)$	probability generating function of rv X
$M_X(s)$	moment generating function of rv X
$\phi_X(s)$	Laplace transformation (LT) of rv X
$\varphi_X(s)$	characteristic function of rv X

$h^{-1}(x)$	functional inverse function of $h(x)$
$*$	binomial thinning operator
$\circledast$	extended-thinning
$\hat{\alpha}_{\text{CLS}}$	Conditional least squares estimate of α
$\hat{\alpha}_{\text{CWLS}}$	Conditional weighted least squares estimate of α
$\hat{\alpha}_{\text{DPLS}}$	Diagonal probability least squares estimate of α
$\hat{\alpha}_{\text{ECF}}$	Empirical characteristic function estimate of α
$\hat{\alpha}_{\text{M}}$	Method of moments estimate of α
$\hat{\alpha}_{\text{MLE}}$	MLE of α
ACF	Autocorrelation function
AIC	Akaike information criterion
AM	absolute monotonicity (absolutely monotone)
AR(p)	autoregressive of order p
cdf	cumulative distribution function
cf	characteristic function
CLS	conditional least square
CM	complete monotonicity (completely monotone)
DSD	discrete self-decomposable
ECF	empirical characteristic function
iid	independent and identically distributed
IIP	independent increment process
ID	infinitely divisible (infinite divisibility)
iff	if and only if
GAR	generalized autoregressive
GC	generalized convolution
GDSD	generalized discrete self-decomposable
GLM	generalized linear model
GNBC	generalized negative binomial convolution
GP	generalized Poisson

GSD	generalized self-decomposable
LT	Laplace transform
mgf	moment generating function
ML	maximum likelihood
MLE	maximum likelihood estimate (estimation)
NB	negative binomial
PDE	partial differential equation
pdf	probability density function
pgf	probability generating function
pmf	probability mass function
P-P plot	Probability-probability plot
Pr	probability of
rv	random variable
SD	self-decomposable
SDE	stochastic differential equation
SE	standard error
SG	self-generalized (self-generalizability)
$X\|Y$	rv X conditioned on rv Y
$[X\|Y = y]$	rv X conditioned on rv $Y = y$

Acknowledgements

I'd like to warmest acknowledge my supervisor, Prof. Harry Joe, a tremendous mentor, for his constant inspiration and guidance. Without him, the theory of continuous-time generalized AR(1) processes would never have been developed. He casts a great impact on me in various aspects of statistical research. His sound intuition and experience always guide me on the right track. I have received numerous suggestions and discussions from him; these were of crucial help to me in the development of this theory. I would say that I am standing on the shoulders of a great statistician.

I would also thank Prof. John Petkau and Prof. Lang Wu for serving on my supervisory committee. Their comments are very valuable to this thesis. In addition, I would thank Dr. Jian Liu, a former supervisory committee member, for his kind help and encouragement.

Many thanks go to Prof. Paul Gustafson, Prof. Martin Puterman, Prof. Martin Barlow, Prof. Nancy Heckman, Prof. James Zidek, Prof. Reg Kulperger for their help and support in many circumstances. Particularly, Prof. Edwin Perkins had carefully corrected mistakes and given valuable suggestions in the probabilistic part of the theory of continuous-time generalized AR(1) processes.

Special thanks to Ryan Woods, Mahbub Latif, Lee Shean Er, Yinshan Zhao, Rachel MacKay, Dana Scott Aeschliman, Weiliang Qiu, Jafar Ahmed Khan, Hongbin Zhang, Dr. Steven Xiaogang Wang, Dr. Renjun Ma, Dr. Li Sun, Dr. Huiying Sun, Dr. Kathy Huiqing Li, Dr. Peter Xuekun Song and others for their help, discussion, encouragement and support in diverse ways.

No doubt, the support from my wife Yalin Chen and my son Si Chen is unforgettable. Their great patience and constant reminding are one of the key sources to impel me to complete this work.

Financial support from Prof. Harry Joe and the Department of Statistics in the past five years are highly appreciated.

Finally, I take this chance to appreciate our student services coordinator Christine Graham. She offers far more than professional service to us. One example is that she always register us for the summer term, because many of us forget it for naively assuming that is vacation. Her excellent work and kind help deeply move me, and make my study at UBC very smooth.

The Department of Statistics at UBC is an incredibly integrated and cozy study environment. I would say that it is truly a statistician's home.

Part I

Introduction

Chapter 1

Overview

This thesis is devoted to the development of a theory to construct models for non-normal equally-spaced or unequally-spaced time series. The time series models are based on continuous-time stochastic processes in a class called generalized AR(1) processes, and these are constructed based on classes of random operators or stochastic differential equations.

Section 1.1 briefly explains the motivation of this study, and reviews the relevant literature. In Section 1.2, we summarize the key ideas that led to our direction of theoretical development, and highlight the new concepts and main results in subsequent chapters; these may help readers to navigate through the details and obtain an integrated understanding of the theory of continuous-time generalized AR(1) process.

1.1 Motivation and literature review

Dynamic phenomena exist in diverse disciplines like chemistry, physics, economics, actuarial science, epidemiology, biology, management science, and so on. It means an event evolving over time. People have been developing various stochastic process models to try to describe or approximate these phenomena. A series of observations of a dynamic process lead to a time series.

Traditionally, for a real-valued time series, we use the Gaussian or normal time series model, which has a Gaussian or normal marginal distribution. However, in reality, there are many situations where the observed series are discrete or positive-valued. Such issues arise especially in the longitudinal studies or clinical trials. The marginal distributions are often skewed and have large variations. Hence, the normal marginal distribution is no longer directly suitable for such situations. This has motivated the development of non-normal time series models, where the marginal distributions could be like the Poisson or Gamma distribution, to handle discrete or positive-valued data.

Such a transition is similar to the transition from the linear model to the generalized linear model where the response variables are discrete or positive-valued. However, unlike the GLM where distributions for modelling discrete or positive-valued responses are well developed, there has been little past research for stochastic processes for discrete or positive-valued time series. For example, suppose we find the marginal distribution for a count data time series is well modelled by the generalized Poisson distribution, what kind of stochastic processes should we use? Or in other words, is there any simple stochastic process which has the generalized Poisson margin? We believe most people will face a difficulty when encountering such a problem. Therefore, it is important to construct probabilistic models which haven't been considered previously.

In addition, the sampling scheme is another serious question. Usually, we take the equally-spaced sampling scheme when we design an experiment study. However, for practical reasons, we may obtain unequally-spaced observations. Many reasons could lead to such phenomena:

- subjects can't be observed on the original schedule plan, say the patients can't visit the clinic for the scheduled appointments due to personal matters;
- there exist missing values;
- or even more extreme, the schedule can't be made equally-spaced, it is random.

For a stationary process, an equally-spaced sampling scheme can guarantee the dependence structure between two adjacent margins is always the same. However, this is not true when the sampling scheme is unequally-spaced. Unequally-spaced time series are sometimes called irregular time series.

Many methods have been developed toward this issue. One reasonable approach is to construct continuous-time stochastic process models as pointed out by Jones [1993], p. 56, because only the continuous-time underlying process can allow the observations taken at arbitrary time points.

In this study, we focus on unequally-spaced count or positive-valued time series. We try to develop the continuous-time stochastic processes for them in a systematic approach. Before we proceed, we take a literature review for both discrete-time and continuous-time stochastic processes with marginal distributions whose support is the non-negative integers or the positive reals.

For positive-valued margins, an incomplete list is Gaver and Lewis [1980] Lawrance and Lewis [1980], Wolfe [1982], Sato and Yamazato [1983], Lewis, McKenzie and Hugus [1989], Anděl [1988, 1989a, 1989b], Rao and Johnson [1988], Hutton [1990], Sim [1990, 1993, 1994], Adke and Balakrishna [1992], Jayakumar and Pillai [1993], Jørgensen and Song [1998], Barndorff-Nielsen [1998b], etc. These marginal distributions include gamma, exponential, and so on. Most of them are discrete-time processes which can not be extended to continuous-time.

For non-negative integer-valued margins, there are: Phatarfod and Mardia [1973], van Harn, Steutel and Vervaat [1981], McKenzie [1985, 1986, 1988], Al-Osh and Alzaid [1987], Al-Osh and Aly [1992], Alzaid and Al-Osh [1993], Aly and Bouzar [1994]. These marginal distributions include Poisson, negative binomial, generalized Poisson, etc. Some of the processes come from the birth-death processes, especially for the linear birth-death processes; one can even trace them to Kendall [1948, 1949].

Joe [1996] proposed a class of discrete-time stochastic processes with infinite divisible margins, which include both count and positive-valued margins.

These processes are first-order Markov processes. Some of them can be generalized to higher order Markov processes. Although the constructions of these processes differ from one another, there are three major approaches: constructing the birth-death process by the generating function method, constructing the process by specifying multivariate distributions for adjacent margins, and constructing the process by solving stochastic differential equations. Next we give a brief comments on these three approaches.

The approach of constructing the birth-death process by the generating function method was established by Kendall [1948, 1949]. It will yield a continuous-time stochastic process with state space being non-negative integers. This approach is still active in finding models for population processes in biological and cancer research. By sampling on equally-spaced time points, we can obtain the discrete-time processes. Two examples from the resulting linear birth-death processes with Poisson and negative binomial margins respectively, are often cited in the literature to model count data time series. However, the birth-death process approach can not yield the processes with state space being the real numbers.

In the area of multivariate non-normal statistics, researchers (see Joe [1997]; Kotz, Balakrishnan and Johnson [2000]) have used copulas and other approaches to construct multivariate distributions with given univariate margins and desirable dependence structures. The theory extends to construct discrete-time Markov processes with given non-normal margins by specifying appropriate multivariate distributions for adjacent margins. One famous example is the one defined by binomial thinning when the marginal distribution is discrete self-decomposable. However, some of these models, for example, random coefficient models, are quite isolated without a systematic method. We can't extend most of them from discrete-time case to continuous-time case because of the consistency requirement for stochastic processes. Moving from the discrete-time to the continuous-time situation, we will experience the change from finite or countably infinite dimensions to uncountably infinite dimensions. This makes it harder to develop theory for continuous-time stochastic processes with given margins.

The third approach is to define a type of stochastic differential equation, and find the solution which yields a continuous-time stochastic process. The obvious benefit is that it could provide a large family of Markov processes with desired margins. For example, Ornstein-Uhlenbeck and Ornstein-Uhlenbeck-type processes obtained from their corresponding SDE's lead to self-decomposable margins (see Section 7.1), known as the class L in Feller [1966b]. Since the theory of stochastic differential equations is dominated by the Itô integral which is involved in Brownian motion, the stochastic differential equations defined for processes with positive-valued margins was not developed until the early 1980s when the Ornstein-Uhlenbeck-type process evolved. Probably this is the first one appearing in that area. To our knowledge, we have not seen any stochastic differential equation defined for processes with non-negative integer-valued margins. The reason could be that we don't know how to define such kind of stochastic differential equations and how to define their solutions. However, the counterpart of self-decomposable distribution was proposed a little bit earlier than the Ornstein-Uhlenbeck-type process, and it leads to the concept of discrete self-decomposable distribution (see definition in Section 7.1). This discrete self-decomposability property leads to continuous-time Markov processes with a special stochastic representation, which involves the binomial thinning operation. The linear birth-death process with Poisson margins discovered by Kendall [1948] is fortunately a concrete example in this family.

1.2 Highlights of our new research

Our study is dedicated to developing continuous-time stochastic processes with count or positive-valued margins which can be used to model equally-spaced or unequally-spaced count or positive-valued time series. To achieve this, for reasons of simplicity, we focus on first-order Markov processes, rather than on more general classes.

We take the dynamic view of building the continuous-time stochastic process with desired margins. Based on the infinitesimal analysis for the stochastic representations of the two linear birth-death processes with Poisson and negative binomial margins, we propose the stochastic differential equation for a continuous-time process with non-negative integer-valued margins. We introduce the concepts of a self-generalized distribution and the extended-thinning operation to define the stochastic differential equation:

$$\begin{aligned} dX(t) &= [K(1-\mu dt) \circledast X(t) - X(t)] + d\epsilon(t) \\ &= [(1-\mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t), \end{aligned} \tag{1.2.1}$$

which we call the generalized Ornstein-Uhlenbeck equation. Here $K(\alpha)$ is a self-generalized rv with respect to parameter α, and "$\circledast$" denotes the extended-thinning operation. The new stochastic integral in our theory is defined by convergence in distribution, rather than in L^2 or probability.

The solution of the generalized Ornstein-Uhlenbeck equation has a simple stochastic representation,

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t), \tag{1.2.2}$$

with a dependent term $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)$ and an innovation term $\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$, quite similar to the structure of first-order auto-regressive process. Hence, we call it the continuous-time generalized AR(1) process. One special case of extended-thinning operations is binomial thinning. In Section 1.1, we mentioned that the binomial thinning operation can lead to continuous-time processes with non-negative integer-valued margins. Such a class is included in the class of continuous-time generalized AR(1) processes. In this way, we can obtain the two linear birth-death processes with Poisson and negative binomial margins again.

By the hint of a correspondence between self-decomposability and discrete self-decomposability, we obtain the positive real counterpart of the discrete self-generalized distribution. This leads to the positive real counterpart of the extended-thinning operation and stochastic differential equation, as well as the solution. Finally, we generalize the extended-thinning operation to the real case; the only known operator is the constant multiplier, and consequently the common AR(1) process obtains.

In summary, we unite the cases of non-negative integer, positive-valued and real-valued state space by the self-generalized distribution and extended-thinning. The corresponding generalized Ornstein-Uhlenbeck equation (1.2.1) leads to the continuous-time generalized AR(1) process (1.2.2). This type of Markov process has a simple stochastic representation, which provides an easy explanation when modelling, and a wide range of stationary infinitely divisible distributions such as Poisson, negative binomial, generalized Poisson, Gamma, exponential, inverse Gaussian etc, to cover diverse problems arising in various disciplines.

In addition, the generalized Ornstein-Uhlenbeck equation can allow us to obtain continuous-time process which is not only stationary, but also non-stationary with time-varying parameters. For example, replacing only the constant parameter μ by a time-varying parameter $\mu(t)$ in (1.2.1), we obtain the stochastic differential equation

$$X(t+h) = (1-\mu(t)h)_K \circledast X(t) + \Delta\epsilon(t), \tag{1.2.3}$$

which leads to non-stationary continuous-time generalized AR(1) process

$$X(t_2) \stackrel{d}{=} \left(e^{-\int_{t_1}^{t_2} \mu(t)dt}\right)_K \circledast X(t_1) + \int_{t_1}^{t_2} \left(e^{-\int_{t}^{t_2} \mu(\tau)d\tau}\right)_K \circledast d\epsilon(t). \tag{1.2.4}$$

Such flexibility is very meaningful in developing models for non-negative integer, or positive, or real-valued time series with non-stationarity like trend, or seasonality, or covariate effects.

In the context of the continuous-time generalized AR(1) process, the constant multiplier operation leads to self-decomposability and the binomial thinning operation leads to discrete self-decomposability. This is well known in the literature. Now the development of the theory of continuous-time generalized AR(1) processes certainly extends the existing concepts of self-decomposability and discrete self-decomposability to other operators: a self-generalized distribution with non-negative integer support leads to generalized discrete self-decomposability, while a self-generalized distribution with positive real support corresponds to generalized self-decomposability. These concepts of generalized self-decomposability and discrete self-decomposability help us to develop continuous-time generalized AR(1) processes with specific marginal distributions to fit practical needs.

This work presents the theory of continuous-time generalized AR(1) processes in the order of model constructions, properties and applications. Now we highlight by chapter the new concepts and key results to help readers gain an overview of the theory of continuous-time generalized AR(1) processes.

Chapter 2 defines the basic distribution families and independent increment processes for the subsequent theoretical developments. Some new distributions are discovered; these include four generalized convolution families: GC I, GC II, GC III and GC IV, which will be used in constructing independent increment processes.

We propose the concepts of self-generalized distributions and extended-thinning operations in Chapter 3. These generalize the binomial thinning and constant multiplier operators for random variables with support on non-negative integers and positive reals. Besides a general theory, four new pairs of families of self-generalized distributions are discovered; there is a one–one mapping of operators with the two types of support. This theory has its origins from a careful study of the conditional probabilities of the linear birth-death process. The self-generalized operator in Example

3.2 is the operator associated with the linear birth-death process whose stationary distribution is negative binomial. In addition, the discovery of extended-thinning operations for positive-valued rv's enlarges our vision on obtaining a positive linear conditional expectation; we need not restrict ourselves on the commonly used constant multiplier operation to achieve this property. These operators also give us more choices in modelling the correlation between two positive random variables.

Chapter 4 develops the generalized Ornstein-Uhlenbeck SDE's to include processes with support on non-negative integers, and construction of solutions of these equations (in the sense of convergence in distribution). The solution has a simple representation in terms of an extended thinning operator and an independent increment innovation process. These resulting processes are called continuous-time generalized AR(1) processes to emphase the similarity of their conditional expectation with that of the continuous-time Gaussian AR(1) process.

Applying the theory in Chapter 4, we obtain interesting results from the generalized Ornstein-Uhlenbeck equations by choosing different extended-thinning operations and independent increment processes; their state spaces cover the non-negative integers, positive reals and reals. Both stationary and non-stationary processes are considered. Many special cases are developed and studied in Chapter 5.

In Chapter 6, we study the stationary distributions of the continuous-time generalized AR(1) processes. This study is to answer the question whether a specific distribution can be the marginal distribution of a continuous-time generalized AR(1) process. It guides us to choose proper processes with certain margins when modelling. Time series with diverse marginal distributions from the stationary continuous-time generalized AR(1) processes are also obtained. Key theorems are Theorems 6.1.1 and 6.3.1. The latter theorem is a result on the pgf or LT of the independent increment innovation process based on the pgf or LT of the stationary distribution, with a given extended-thinning operator. Many special cases are developed and studied.

Chapter 7 further studies the stationary distributions under different extended-thinning operations. The generalized self-decomposable (GSD) and generalized discrete self-decomposable (GDSD) classes are defined in a similar way to the self-decomposable and discrete self-decomposable classes associated with the constant multiple and binomial thinning operators. Several ways are developed to check if a given distribution is in one of the GSD or GDSD classes. Key theorems are Theorems 7.2.3, 7.2.5 (possibly simpler ways to check if a distribution is GSD or GDSD), and Theorem 7.2.7 (infinite divisibility of the classes). Relations between different GSD and GDSD classes are studied, as well as analog results between the cases of discrete and continuous margins.

Chapter 8 investigates infinitesimal transition and duration features of the continuous-time generalized AR(1) processes. A PDE characterization is given for the conditional pgf or LT; a key result is that the pgf or LT of a self-generalized distribution is determined by its partial derivative evaluated at a boundary. For the continuous-time generalized AR(1) process with non-negative integer support, the infinitesimal generator matrix has the downwardly skip-free property. Another key result is that a steady-state continuous-time generalized AR(1) process can be determined based on two of the following three elements: marginal distribution, self-generalized distribution for the operator, increment of the innovation process.

In Chapter 9, we present some differences for stochastic process constructions for the discrete-time and continuous-time situations. We also study conditional and multivariate distributions associated with some specific cases of the continuous-time generalized AR(1) process. A by-product is a new approach to construct families of multivariate distributions with given univariate margins. Interesting stochastic representations are given for some special processes and it is shown that some new discrete-time time series with gamma margins have better properties in the innovation random variable, compared with time series based on self-decomposability.

We give a thorough study on parameter estimation methods in Chapter 10; these estimators including MLE, CLS, ECF etc, are desired in different situations and have their own advantages and disadvantages.

Chapter 11 looks into the asymptotic properties of the commonly used estimates like MLE and CLS in the unequally-spaced setting. A random sampling scheme is proposed, and results build on the techniques of proof in Billingsley [1961a] and Klimko and Nelson [1978].

Chapter 12 discusses a variety of topics like detection of serial dependence, model diagnosis and selection, hypothesis testing, forecasting and process simulation. The graphical methods, called the P-P plot and diagonal P-P plot, are proposed for assessing autocorrelation and model diagnosis.

In Chapter 13, we illustrate the capability of the theory of continuous-time generalized AR(1) processes for real problems with three applications. These time series include non-negative integer and positive-valued observations.

Finally in Chapter 14, we summarize the strengths and weakness of the continuous-time generalized AR(1) processes in modelling. Also we briefly discuss some thoughts on construction of stochastic processes. Areas for future research are also mentioned.

Part II

Theory for model construction

Chapter 2

Relevant background on characteristic tools, distribution families and stochastic processes

In this chapter, we cover background concepts needed in the development of the new theory of continuous-time generalized AR(1) processes. We try to select a minimum of necessary materials for the subsequent chapters.

This chapter is organized in the following way: Section 2.1 briefly introduces the Ornstein-Uhlenbeck processes and Ornstein-Uhlenbeck-type processes; we will generalize these processes to a wider range, leading to the continuous-time generalized AR(1) processes. In Section 2.2, we discuss some characteristic tools for probability distributions and prove some new results. We present some particular distribution families in Section 2.3, and independent increment processes in Section 2.4. These results are used in subsequent chapters for special examples of generalized AR(1) processes.

2.1 Ornstein-Uhlenbeck processes and Ornstein-Uhlenbeck-type processes

The Ornstein-Uhlenbeck process comes from the Ornstein-Uhlenbeck stochastic differential equation (SDE), which has another name, the Langevin equation (see Ornstein and Uhlenbeck [1930], also Nelson [1967], Øksendal [1995], Hsu and Park [1988], Schuss [1988]). Let $\{X(t);\ t \geq 0\}$ be a continuous-time process. The Ornstein-Uhlenbeck equation is defined as

$$dX(t) = -\mu X(t)dt + \sigma dW(t), \qquad \mu > 0,\ \sigma > 0,$$

where $\{W(t); t \geq 0\}$ is a standard Brownian motion independent of $X(0)$. The solution of this SDE is well known as

$$X(t) \overset{d}{=} e^{-\mu t} X(0) + \sigma \int_0^t e^{-\mu(t-\tau)} dW(\tau),$$

where $\int_0^t e^{-\mu\tau} dW(\tau)$ is the Itô integral, which is the limit of a sequence of rv's in the sense of convergence in L^2, and is normally distributed. Hence, the support of the margin $X(t)$ is $\Re$. Furthermore, $X(t)$ can be represented as

$$X(t) \overset{d}{=} e^{-\mu(t-s)} X(s) + \sigma \int_0^{t-s} e^{-\mu(t-s-\tau)} dW(\tau), \qquad s < t.$$

Note that $\int_0^{t-s} e^{-\mu(t-s-\tau)} dW(\tau)$ can be written as $\int_s^t e^{-\mu(t-\tau)} dW(\tau)$, and is independent of $X(s)$ because $X(s)$ is independent of $\{W(\tau); \tau \geq s\}$. This feature shows that the process is a Markov process, and a discrete-time AR(1) process can be readily obtained from it. If $X(0)$ is normally distributed, then $X(t)$ is normally distributed for all $t > 0$. This model serves continuous-time time series very well, and is named as continuous-time AR(1) (CAR(1)). See Brockwell and Davis [1996] and references therein. Because of normal margins under steady state, it is sometimes called a continuous-time AR(1) Gaussian process. The Ornstein-Uhlenbeck process has applications in mathematical finance (see Neftci [1996]).

Wolfe [1982] initiated the study of Ornstein-Uhlenbeck-type processes. Almost at the same time, Sato and Yamazato [1982], Jurek and Vervaat [1983] studied this process too. The Ornstein-Uhlenbeck SDE is extended to

$$dX(t) = -\mu X(t)dt + dW(t), \qquad \mu > 0,$$

where $\{W(t); t \geq 0\}$ is a homogeneous Lévy process independent of $X(t)$. The solution has the same form as the Ornstein-Uhlenbeck SDE:

$$X(t) \stackrel{d}{=} e^{-\mu(t-s)} X(s) + \int_0^{t-s} e^{-\mu(t-s-\tau)} dW(\tau), \qquad s < t.$$

but $\int_0^t e^{-\mu(t-s-\tau)} dW(\tau)$ is not the Itô integral. This stochastic integral is the limit of a sequence of rv's in the sense of convergence in probability. The existence of such stochastic integral can be found in Lukacs [1968], where the characteristic function of the integral has the form

$$\exp\left\{\int_0^{t-s} \log \varphi_{W(1)}\left(se^{-\mu(t-s-\tau)}\right) d\tau\right\}.$$

Similarly, $\int_0^{t-s} e^{-\mu(t-s-\tau)} dW(\tau)$ $\left(= \int_s^t e^{-\mu(t-\tau)} dW(\tau)\right)$ is independent of $X(s)$. The support of $W(t)$ can be positive real-valued. Hence, $\{X(t); t \geq 0\}$ can be a positive-valued process. Similar to the Ornstein-Uhlenbeck process, a generalized time series (other than classical Gaussian distributed time series) can be easily obtained if sampling on equally-spaced time points. This feature allows the Ornstein-Uhlenbeck-type processes to model positive-valued observed data. Wolfe [1982] showed two possible applications: the study of radioactive material in stockpile, and bank currency. Later Barndorff-Nielsen *et al.* [1993, 1998a] applied this kind of process with specific marginal distributions to turbulence and finance.

Now we discuss some common features of Ornstein-Uhlenbeck and Ornstein-Uhlenbeck-type processes:

- Nice stochastic representation form: the sum of two independent terms. One governs the dependence relation with the previous state, one explains the input (noise or innovation). Note that $|e^{-\mu(t-s)} X(s)| \leq |X(s)|$, hence, the term $e^{-\mu(t-s)} X(s)$ looks "thinner" than $X(s)$.

- First order Markov: this Markov property is very helpful in the study of conditional properties, stationary distribution or margin under steady state, transition properties, and joint finite-dimensional distributions.

- Simple auto-correlation: the auto-correlation function, if it exists, under steady state has the exponential form $\mathbf{Cor}[X(s), X(t)] = e^{-\mu|t-s|}$. This implies that for a bigger time difference, there is less influence on the future state.

Although the Lévy process $\{W(t);\ t \geq 0\}$ can have increments which are non-negative integer valued, the term $e^{-\mu(t-s)}X(s)$ excludes the possibility of non-negative integer valued margins. Such a disadvantage precludes the application to count data time series.

In the study of continuous-time generalized AR(1) process, we extend the stochastic operation of a constant multiplier to an extended-thinning operation, and define generalized stochastic integrals. Such modifications allow us to obtain a similar representation for processes with non-negative integer state space. The continuous-time generalized AR(1) processes includes the Ornstein-Uhlenbeck process and Ornstein-Uhlenbeck-type process as special cases.

2.2 Characterization tools of distributions and examples

In this section, we review the common characterization tools which are heavily used in the theory of continuous-time generalized AR(1) process. Note that this is a simplification of terminology; the processes are AR(1)-like with AR(1) autocorrelation, but not always autoregressive. These tools include the probability generating function (pgf), Laplace transformation (LT), moment generating function (mgf) and characteristic function (cf).

This section consists of results that are used in subsequent chapters. It can be skimmed in the first reading. Proposition 2.2.2 is especially important.

Any kind of generating function has the property that there is one-to-one mapping between the generating functions and the distributions. Hence, by investigating the generating function, we can know the corresponding distribution. In principle, the cf can be used in all types of random variables because it always exists. However, for specific types of random variables or distribution families, other generating functions may be more convenient. For example, the pgf is often used in non-negative integer-valued rv's, while the LT is adopted for positive real-valued rv's. This is for convenience of theorems for pgf's and LT's that can be applied. In exponential dispersion models, the mgf is used because the definition of that kind of model is related to the cumulant generating function.

We also prove some new results concerning pgf's and LT's. These will play certain roles in the theory of continuous-time generalized AR(1) processes.

2.2.1 Probability generating function

The probability generating function is used for discrete distributions with non-negative integer support $\mathcal{N}_0$. Assume X is a non-negative integer-valued rv with probability mass function

$$\Pr[X = i] = p_i \geq 0, \qquad i = 0, 1, 2, \ldots.$$

The pgf of X is defined as

$$G_X(s) = \mathbf{E}\,(s^X) = \sum_{i=0}^{\infty} p_i s^i, \qquad 0 \leq s \leq 1.$$

Usually the pgf is defined on $[0, 1]$ because the power series on the right hand side always exists when $0 \leq s \leq 1$. This domain is sufficient for our need. Of course, it can be extended to $|s| \leq 1$. As for $|s| > 1$, the finiteness of $G_X(s)$ depends on the individual probability mass function. The function $G_X(s)$ is increasing from p_0 to 1 as s increases from 0 to 1. Once we have the pgf, we can obtain the probability masses:

$$p_i = G_X^{(i)}(0)/i!, \qquad i = 0, 1, 2, \ldots.$$

Also the mean and variance are derived as

$$\mathbf{E}\,(X) = G'_X(1), \qquad \mathbf{Var}\,(X) = \mathbf{E}\,(X^2) - (\mathbf{E}\,(X))^2 = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2.$$

The **index of dispersion**, D, is defined as $D(X) = \mathbf{Var}\,(X)/\mathbf{E}\,(X)$, and is referred as an index of dispersion for distributions for count data. If $D(X) > 1$, there is overdispersion relative to Poisson. If $D(X) < 1$, there is underdispersion relative to Poisson.

The following theorem characterizes the pgf; it is useful to verify if a function $G(s)$ is a pgf. Refer to Bondesson [1992], p. 9.

Theorem 2.2.1 *Suppose $G(s)$ is a Taylor series in s. Then, $G(s)$ is a pgf iff*

1. *$G(s)$ is absolutely monotone (AM), i.e.,* $G^{(i)}(s) \geq 0, \; i \in \mathcal{N}_0, \; s \in [0,1)$.

2. $G(s) \longrightarrow 1$, *as* $s \to 1$.

This is equivalent to checking $G^{(i)}(0) \geq 0$ for all i and $G(1) = 1$. It is viewed as the discrete counterpart of Bernstein's theorem (Theorem 2.2.5). See Bondesson [1992], p. 9.

Perhaps the simplest distribution is the Bernoulli distribution. It is often used to build other distributions, say Binomial, Poisson, etc. Let $X \sim \text{Bernoulli}(p)$. X takes only two values, 0 and 1, with

$$\Pr[X = 1] = p, \qquad \Pr[X = 0] = 1 - p;$$

and the pgf is $G_X(s) = (1-p) + ps$. Consequently, the mean and variance are

$$\mathbf{E}(X) = p \quad \text{and} \quad \mathbf{Var}(X) = p(1-p).$$

Some distributions with non-negative integer support are listed below; these can be used in modelling count data. All of them are discussed to some extent in Bondesson [1992]. Also refer to Johnson and Kotz [1969].

(a) Poisson: $X \sim \text{Poisson}(\lambda)$. Then

$$p_i = \Pr[X = i] = \frac{\lambda^i}{i!}e^{-\lambda}; \quad i = 0, 1, 2, \ldots; \quad \lambda > 0.$$

The pgf is $G_X(s) = \exp\{\lambda(s-1)\}$, and $\mathbf{E}(X) = \mathbf{Var}(X) = \lambda$. Thus, $D(X) = 1$, which is referred to as equidispersion.

(b) generalized Poisson: Let $X \sim \text{GP}(\theta, \eta)$. Then

$$p_i = \Pr[X = i] = \theta(\theta + \eta i)^{i-1}e^{-\theta-\eta i}/i!, \quad i = 0, 1, 2, \ldots; \quad \theta > 0, \;\; 0 \leq \eta \leq 1.$$

The pgf is

$$G_X(s) = \exp\left\{\theta\left(\sum_{k=1}^{\infty}\eta(k\eta)^{k-1}e^{-k\eta}s^k/k! - 1\right)\right\},$$

The mean, variance and index of dispersion are

$$\mathbf{E}(X) = \theta(1-\eta)^{-1}, \quad \mathbf{Var}(X) = \theta(1-\eta)^{-3}, \quad D(X) = (1-\eta)^{-2} \geq 1.$$

Note that if $\eta = 0$, it becomes Poisson(θ). A good reference is Consul [1989]; there η can be negative to obtain an underdispersed distribution. However, this case has nothing to do with the study of continuous-time generalized AR(1) processes, because the marginal distribution of continuous-time generalized AR(1) process should have probability mass on the whole non-negative integer set, not on a bounded subset. When $\eta < 0$, the generalized Poisson rv has an upper bound of support.

(c) negative binomial: Let $X \sim \mathrm{NB}(\gamma, q)$. [Note that this is a non-standard parametrization.] Then the pmf is

$$p_i = \Pr[X = i] = \binom{\gamma + i - 1}{i}(1-q)^\gamma q^i; \quad i = 0, 1, 2, \ldots; \quad \gamma > 0 \ \ , 0 \le q \le 1.$$

The pgf is

$$G_X(s) = \left(\frac{1-q}{1-qs}\right)^\gamma = \exp\left\{\gamma \log(1-q)^{-1}\left[\frac{\log(1-qs)}{\log(1-q)} - 1\right]\right\},$$

and

$$\mathbf{E}(X) = \gamma q/(1-q), \quad \mathbf{Var}(X) = \gamma q/(1-q)^2, \quad D(X) = 1/(1-q) \ge 1.$$

When γ is an integer, the negative binomial distribution can be explained by Bernoulli trials with success probability $1-q$ or failure probability q, in which the experiment stops until the γ-th successes, and X is the total number of trials in the experiment.

The geometric distribution is the special case in the negative binomial family. It is obtained when $\gamma = 1$ with pmf:

$$p_i = \Pr[X = i] = (1-q)^\gamma q^i; \quad i = 0, 1, 2, \ldots; \quad \gamma > 0 \ \ , 0 \le q \le 1,$$

and pgf $G_X(s) = \frac{1-q}{1-qs}$. Note that X can take value 0. But sometimes people treat $X' = X+1$ as the geometric distribution which is positive integer-valued and has pgf $G_{X'}(s) = \frac{(1-q)s}{1-qs}$. Unless stated otherwise, we will take the former as geometric distribution throughout the thesis.

(d) discrete stable: Steutel and van Harn [1979] proposed this discrete stable distribution. Let X be a rv from discrete stable distribution. Then the pgf is defined as

$$G_X(s) = \exp\{-\lambda(1-s)^\theta\} = \exp\{\lambda[1-(1-s)^\theta - 1]\}, \qquad \lambda > 0,\ 0 < \theta \le 1.$$

The pmf can be obtained by expanding the pgf in a power series:

$$p_0 = e^{-\lambda},\ p_1 = \lambda\theta e^{-\lambda},\ p_i = (-1)^i \sum_{j=1}^{\infty} \frac{(-1)^j \lambda^j}{j!} \frac{\Gamma(j\theta+1)}{\Gamma(j\theta+1-i)i!}, \qquad i = 2, 3, \ldots.$$

However, since $G'_X(s) = G_X(s) \cdot \lambda\theta/(1-s)^{1-\theta}$, the expectation will be infinite if $0 < \theta < 1$. When $\theta = 1$, it becomes Poisson(λ).

(e) logarithmic series distribution: Let X be a rv from logarithm series distribution. Then the pmf is defined as

$$p_i = \Pr[X = i] = \frac{c^{i+1}}{(i+1)\theta}; \quad i = 0, 1, 2, \ldots; \quad c = 1 - e^{-\theta}, \ \ \theta > 0.$$

The pgf is

$$G_X(s) = -s^{-1}\log(1-cs)/\theta = s^{-1}\log(1-cs)/\log(1-c),$$

and

$$\mathbf{E}(X) = c\theta^{-1}/(1-c) - 1, \quad \mathbf{Var}(X) = c\theta^{-1}(1-c\theta^{-1})/(1-c)^2, \quad D(X) = \frac{c\theta^{-1}(1-c\theta^{-1})}{(1-c)(c\theta^{-1}+c-1)}.$$

Note that this logarithm series distribution is left shifted to 0 compared to the usual definition in Johnson and Kotz [1969], p. 166. Therefore, this logarithm series distribution is sometimes overdispersed, and sometimes underdispersed depending on the parameter θ. Let θ_0 be the solution of $\frac{c\theta^{-1}(1-c\theta^{-1})}{(1-c)(c\theta^{-1}+c-1)} = 1$. Then if $\theta > \theta_0$, it has overdispersion; otherwise, underdispersion.

(f) power series distribution: Let X be distributed in power series distribution. Then the pmf is

$$p_0 = \theta,\ p_i = \Pr[X = i] = \theta \prod_{k=1}^{i} (k-\theta) \Big/ (i+1)!, \quad i = 1, 2, \ldots; \quad 0 < \theta \le 1.$$

When $\theta = 1$, X degenerates to 0. The pgf is

$$G_X(s) = s^{-1}[1 - (1-s)^\theta].$$

Note that when $0 < \theta < 1$, X has no moments. See Bondesson [1992], p. 128 and p. 132. This is also related to LTC (Laplace transform family C) in Joe [1997], p. 375; there is a left shift. Also the discrete stable distribution is compound Poisson with the distribution of $X + 1$. See the pgf expression in (d).

(g) Zeta (discrete Pareto) distribution: Let X be a rv from Zeta(ρ). Then the pmf is defined as

$$p_i = c \cdot (i+1)^{-(\rho+1)}, \quad i = 0, 1, 2, \ldots; \quad \rho > 0,$$

where $c = \sum_{i=1}^{\infty} i^{-(\rho+1)}$. Note that this distribution comes from left shifting the Zeta distribution in Johnson and Kotz [1969], p. 240; it is commonly called the Zipf-Estoup law in linguistic studies.

Unfortunately, the pgf, expectation and variance of Zeta distribution have no explicit expressions.

Stochastic operations can lead to new pgf's. Here we summarize some of the facts regarding operations on one rv.

Proposition 2.2.2

(1) $G(s)$ *pgf* $\Rightarrow (1-\alpha) + \alpha G(s)$, $0 \le \alpha \le 1$, *is a pgf* [random zero-truncation operation].

(2) $G(s)$ *pgf* $\Rightarrow G(\alpha s + 1 - \alpha)$, $0 \le \alpha \le 1$, *is a pgf* [binomial-thinning operation].

(3) $G(s;\beta)$ *pgf for* $\beta \in B$ *and* F *a distribution on* $B \Rightarrow \int_B G(s;\beta)dF(\beta)$ *is a pgf* [mixture operation].

(4) $G(s)$ *pgf* $\Rightarrow e^{\lambda[G(s)-1]}$ $(\lambda > 0)$ *is a pgf* [compound Poisson operation].

(5) $G(s)$ *pgf* $\Rightarrow (1-\alpha) + \alpha s G(s)$, $0 \le \alpha \le 1$, *is a pgf* [zero-modification operation].

Proof: Suppose X has pgf $G(s)$ or $G(s;\beta)$.

(1) Consider $Y = I \cdot X$, where $I \sim \text{Bernoulli}(\alpha)$. Then

$$G_Y(s) = \mathbf{E}(s^Y) = \mathbf{E}(s^{I\cdot X}) = \Pr[I=0] + \Pr[I=1]\mathbf{E}(s^X) = (1-\alpha) + \alpha G(s).$$

(2) Let $Y = \sum_{i=0}^{X} I_i$, where $I_0 = 0,\ I_i \overset{i.i.d.}{\sim} \text{Bernoulli}(\alpha), i = 1, 2, \ldots$. Then

$$\begin{aligned} G_Y(s) &= \mathbf{E}(s^Y) = \mathbf{E}\left(s^{\sum_{i=0}^{X} I_i}\right) = \mathbf{E}\left(\mathbf{E}\left(s^{\sum_{i=0}^{X} I_i}\middle| X\right)\right) \\ &= \mathbf{E}\left(\mathbf{E}\left(s^{I_1}\right)^X\right) = \mathbf{E}\left([\alpha s + 1 - \alpha]^X\right) = G(\alpha s + 1 - \alpha). \end{aligned}$$

(3) Suppose Y conditioned on $\beta^* = \beta$ has the same pgf $G(s;\beta)$, and β^* is distributed in F on B. Then

$$G_Y(s) = \mathbf{E}(s^Y) = \mathbf{E}\left(\mathbf{E}(s^Y|\beta^*)\right) = \mathbf{E}(G(s;\beta^*)) = \int_B G(s;\beta)dF(\beta).$$

(4) Let $Y = \sum_{i=0}^{Z} X_i$, where $X_0 = 0,\ X_i \overset{i.i.d.}{\sim}$ with pgf $G(s)$ $(i = 1, 2, \ldots)$, and $Z \sim \text{Poisson}(\lambda)$. Then

$$\begin{aligned} G_Y(s) &= \mathbf{E}(s^Y) = \mathbf{E}\left(s^{\sum_{i=0}^{Z} X_i}\right) = \mathbf{E}\left(\mathbf{E}\left(s^{\sum_{i=0}^{Z} X_i}\middle| Z\right)\right) \\ &= \mathbf{E}\left(\mathbf{E}\left(s^{X_1}\right)^Z\right) = \mathbf{E}\left(G^Z(s)\right) = e^{\lambda[G(s)-1]}. \end{aligned}$$

(5) Consider $Y = I \cdot (X+1)$, where $I \sim \text{Bernoulli}(\alpha)$. Then

$$G_Y(s) = \mathbf{E}(s^Y) = \mathbf{E}(s^{I\cdot(X+1)}) = \Pr[I=0] + \Pr[I=1]s\mathbf{E}(s^X) = (1-\alpha) + \alpha s G(s).$$

Note that random zero-truncation operation is very similar to zero-modification operation. Both involve truncation. The random zero-truncation operation directly applies truncation to a rv X, while the zero-modification operation first shifts X to right as $X+1$, then applies truncation. Hence, both primarily keep the shape of the pmf of X with slight differences. However, there does exist a difference between two operations. The random zero-truncation operation increases the probability mass at zero:

$$\Pr[I \cdot X = 0] = (1-\alpha) + \alpha \Pr[X=0] = (1-\alpha) - (1-\alpha)\Pr[X=0] + \Pr[X=0] \geq \Pr[X=0].$$

But the zero-modification operation relocates the probability mass at zero:

$$\Pr[I \cdot (X+1) = 0] = \Pr[I = 0] = \alpha,$$

which could be bigger or smaller than $\Pr[X = 0]$.

The fraction of zeros in count data is one concern when modelling. Both random zero-truncated distribution and zero-modified distribution of X are alternative choices for data with zero fraction if the original distribution of X does fit the data very well. However, the zero-modified distribution of X is more flexible then the random zero-truncated distribution, because it can be used to either lower or higher zero fraction situation, while the random zero-truncated distribution can only applied in higher zero fraction situation (sometimes called zero-inflated).

Example 2.1 *Poisson(λ) compounded with NB$(1, q)$ will have pgf*

$$\exp\left\{\lambda\left(\frac{1-q}{1-qs} - 1\right)\right\} = \exp\left\{\lambda\frac{q(s-1)}{1-qs}\right\},$$

where $\lambda > 0$ and $0 \le q < 1$. This is the basis of GC I introduced in Section 2.3.3.

Another example of a compound Poisson distribution leads to the GC II in Section 2.3.3. We claim that

$$\exp\left\{\frac{q(s-1)(1-\gamma s)}{(1-qs)}\right\} = \exp\left\{(1-\gamma s)\left(\frac{1-q}{1-qs} - 1\right)\right\}, \qquad 0 < \gamma \le q < 1, \tag{2.2.1}$$

is a pgf. This is because

$$\begin{aligned}
\frac{1-\gamma s}{1-\gamma} \cdot \frac{1-q}{1-qs} &= \frac{1-q}{1-\gamma}(1-\gamma s)(1+qs+q^2s^2+q^3s^3+\cdots) \\
&= \frac{1-q}{1-\gamma}[1+qs+q^2s^2+q^3s^3+\cdots \\
&\qquad -\gamma s - \gamma q s^2 - \gamma q^2 s^3 - \cdots] \\
&= \frac{1-q}{1-\gamma} + \frac{(1-q)(q-\gamma)}{1-\gamma}s + \frac{(1-q)q(q-\gamma)}{1-\gamma}s^2 + \cdots.
\end{aligned}$$

Let $G(s) = 1 + \frac{1}{d}\frac{1-\gamma s}{1-\gamma}\left(\frac{1-q}{1-qs} - 1\right)$, where $d \ge \frac{q}{1-\gamma} > 0$. Then

$$\begin{aligned}
G(s) &= 1 + \frac{1}{d}\left[\frac{1-q}{1-\gamma} + \frac{(1-q)(q-\gamma)}{1-\gamma}s + \frac{(1-q)q(q-\gamma)}{1-\gamma}s^2 + \cdots - \frac{1-\gamma s}{1-\gamma}\right] \\
&= 1 - \frac{1}{d}\frac{q}{1-\gamma} + \frac{\gamma + (1-q)(q-\gamma)}{d(1-\gamma)}s + \frac{(1-q)q(q-\gamma)}{d(1-\gamma)}s^2 + \cdots.
\end{aligned}$$

Since all coefficients of series expansion of $G(s)$ are non-negative and $G(1) = 1$, $G(s)$ is a pgf. Hence we can represent 2.2.1 as

$$\exp\left\{\frac{q(s-1)(1-\gamma s)}{1-qs}\right\} = \exp\left\{d(1-\gamma)[G(s)-1]\right\}.$$

This implies that $\exp\left\{\frac{q(s-1)(1-\gamma s)}{1-qs}\right\}$ *is the pgf of a compound Poisson distribution.*

Example 2.2 *Following the zero-modification operation, we can show that*

$$L(s) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}, \quad 0 \le \alpha \le 1,\ 0 < \gamma < 1.$$

is a pgf. This is because the following decomposition:

$$L(s) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s} = \left[1 - \frac{(1-\gamma)\alpha}{1-\alpha\gamma}\right] + \frac{(1-\gamma)\alpha}{1-\alpha\gamma} \cdot s \cdot \frac{(1-\gamma)/(1-\alpha\gamma)}{1-(1-\alpha)\gamma s/(1-\alpha\gamma)}.$$

Here we know that $0 \le \frac{(1-\gamma)\alpha}{1-\alpha\gamma} \le 1$ *and* $0 \le 1 - \frac{1-\gamma}{1-\alpha\gamma} = \frac{(1-\alpha)\gamma}{1-\alpha\gamma} \le 1$. *Let* $I \sim \text{Bernoulli}\left(\frac{(1-\gamma)\alpha}{1-\alpha\gamma}\right)$, $Z \sim \text{NB}\left(1, \frac{(1-\alpha)\gamma}{1-\alpha\gamma}\right)$. *Then* $I(Z+1)$ *has the* $L(s)$ *as its pgf. When* $\alpha = 0$, $X = 0$, *while* $\alpha = 1$, $X = Z + 1$.

We are not clear if such a distribution has been previously studied. Since Z is Geometric, we call this distribution the **zero-modified geometric distribution**.

Sometimes operations are carried out on more than one random variable. The well known **convolution**, which is the sum of independent rv's, is an example. For two independent non-negative integer valued rv's X_1 and X_2, the pgf of convolution $X_1 + X_2$ is

$$G_{X_1+X_2}(s) = \mathbf{E}(s^{X_1+X_2}) = \mathbf{E}(s^{X_1})\mathbf{E}(s^{X_2}) = G_{X_1}(s)G_{X_2}(s).$$

For more than two, say n independent rv's, we have

$$G_{\sum_{i=1}^n X_i}(s) = \mathbf{E}(s^{\sum_{i=1}^n X_i}) = \prod_{i=1}^n \mathbf{E}(s^{X_i}) = \prod_{i=1}^n G_{X_i}(s).$$

Furthermore, the convolution concept can be naturally extended to the situation of uncountably many rv's, leading to the **generalized convolution**. See Section 2.3.3.

Next we prove that some functions are probability generating functions; these will be used in the study of continuous-time generalized AR(1) processes in later chapters.

Theorem 2.2.3 *The following functions $L(s)$ are pgf's.*

(1) $L(s) = \frac{1}{s} - \frac{\theta(1-s)^\theta}{1-(1-s)^\theta}, \quad 0 < \theta \le 1.$

(2) $L(s) = c^{-1}[1 - e^{-\theta(1-\alpha)}(1-cs)^\alpha], \quad 0 \le \alpha \le 1,\ c = 1 - e^{-\theta},\ \theta \ge 0,$

(3) $L(s) = 1 - \alpha^\theta(1-\gamma)^\theta \left[(1-\alpha)\gamma + (1-\gamma)(1-s)^{-1/\theta}\right]^{-\theta}, \quad 0 \le \alpha \le 1,\ 0 < \gamma < 1,\ \theta \ge 1.$

(4) $L(s) = (s-1)\frac{G'(s)}{G(s)} + 1,$ *where* $G(s)$ *is the pgf of Zeta*(ρ), $\rho > 0$.

(5) $L(s) = 1 + \frac{s-1}{(1-\beta s)(1-\frac{\beta-\alpha}{1-\alpha}s)}, \quad 0 < \alpha < 1,\ 0 \le \beta \le 1,\ \frac{\alpha}{1+(1-\alpha)^{3/2}} \le \beta \le \alpha,$ *or* $\alpha < \beta \le \frac{1}{2-\alpha}.$

(6) $L(s) = 1 + \frac{(1-\gamma s)(s-1)}{(1-\beta s)(1-\frac{\beta-\alpha}{1-\alpha}s)}, \quad 0 < \alpha < 1,\ 0 \le \beta \le 1,\ 0 \le \gamma < 1,$ *and either* $\alpha < \beta \le \frac{1+\gamma-\alpha\gamma}{2-\alpha},$ *or* $\max\left(\gamma, \frac{\alpha}{2-\alpha}\right) \le \beta \le \alpha$ *and* $\beta(\beta-\gamma)(1-\alpha)^3 \ge (\alpha-\beta+\gamma-\alpha\gamma)(\alpha-\beta).$

Proof: It is obvious that for all cases, $L(1) = 1$. Suppose $L(s)$ has series expansion of form

$$L(s) = r_0 + r_1 s + r_2 s^2 + \cdots = \sum_{i=0}^{\infty} r_i s^i.$$

It suffices to show that all coefficients $r_i \ge 0$ $(i = 0, 1, 2, \ldots)$.

(1) Rewrite $L(s)$ as

$$L(s) = \frac{1}{s} - \frac{\theta(1-s)^\theta}{1-(1-s)^\theta} = \frac{1}{s} - \left[\frac{(1-s)^{-\theta} - 1}{\theta}\right]^{-1}.$$

Since

$$(1-s)^{-\theta} = 1 + \theta s + \frac{\theta(\theta+1)}{2!}s^2 + \frac{\theta(\theta+1)(\theta+2)}{3!}s^3 + \cdots = 1 + \theta s + \sum_{j=2}^{\infty} \frac{\theta \prod_{k=1}^{j-1}(k+\theta)}{j!} s^j,$$

we have

$$L(s) = \frac{1}{s} - \left[s + \sum_{j=2}^{\infty} \frac{\prod_{k=1}^{j-1}(k+\theta)}{j!} s^j\right]^{-1} = \frac{1}{s}\left(1 - \left[1 + \sum_{j=1}^{\infty} \frac{\prod_{k=1}^{j}(k+\theta)}{(j+1)!} s^j\right]^{-1}\right).$$

Assume the Taylor expansion:

$$\left[1 + \sum_{j=1}^{\infty} \frac{\prod_{k=1}^{j}(k+\theta)}{(j+1)!} s^j\right]^{-1} = 1 - q_1 s - q_2 s^2 - q_3 s^3 - q_4 s^4 - \cdots.$$

Then

$$L(s) = s^{-1}\left(1 - \left[1 - q_1 s - q_2 s^2 - q_3 s^3 - q_4 s^4 - \cdots\right]\right) = q_1 + q_2 s + q_3 s^2 + q_4 s^3 + \cdots.$$

We now need to prove that $q_j \geq 0$ for $j = 1, 2, \ldots$. Because

$$\left[1 + \sum_{j=1}^{\infty} \frac{\prod_{k=1}^{j}(k+\theta)}{(j+1)!} s^j\right] \cdot \left[1 - q_1 s - q_2 s^2 - q_3 s^3 - q_4 s^4 - \cdots\right] = 1,$$

it follows that

$$\begin{aligned}
q_1 &= \frac{1+\theta}{2}, \\
q_2 &= \frac{(1+\theta)(2+\theta)}{3!} - \frac{1+\theta}{2} q_1, \\
&\vdots \\
q_j &= \frac{\prod_{k=1}^{j}(k+\theta)}{(j+1)!} - \sum_{l=1}^{j-1} \frac{\prod_{k=1}^{j-l}(k+\theta)}{(j-l+1)!} q_l, \\
&\vdots
\end{aligned}$$

Note that $\frac{a}{b} \leq \frac{a+1}{b+1}$ if $0 < a \leq b$. For $j \geq 2$,

$$\begin{aligned}
q_j &= \frac{\prod_{k=1}^{j}(k+\theta)}{(j+1)!} - \sum_{l=1}^{j-2} \frac{\prod_{k=1}^{j-l}(k+\theta)}{(j-l+1)!} q_l - \frac{1+\theta}{2} q_{j-1} \\
&= \frac{j+\theta}{j+1} \cdot \frac{\prod_{k=1}^{j-1}(k+\theta)}{j!} - \sum_{l=1}^{j-2} \frac{j-l+\theta}{j-l+1} \cdot \frac{\prod_{k=1}^{j-l-1}(k+\theta)}{(j-l)!} q_l - \frac{1+\theta}{2} q_{j-1} \\
&> \frac{j-1+\theta}{j} \cdot \frac{\prod_{k=1}^{j-1}(k+\theta)}{j!} - \frac{j-1+\theta}{j} \sum_{l=1}^{j-2} \frac{\prod_{k=1}^{j-l-1}(k+\theta)}{(j-l)!} q_l - \frac{1+\theta}{2} q_{j-1} \\
&= \frac{j-1+\theta}{j} q_{j-1} - \frac{1+\theta}{2} q_{j-1} = \left(\frac{j-1+\theta}{j} - \frac{1+\theta}{2}\right) q_{j-1}.
\end{aligned}$$

Due to the fact that $q_1 = \frac{1+\theta}{2} > 0$, by induction, we obtain $q_j > 0$ $(j \geq 2)$, which means that $L(s)$ is a pgf.

(2) Note that for $0 \leq \alpha \leq 1$, $(1 - cs)^{\alpha} = 1 - \alpha cs - \frac{\alpha(1-\alpha)}{2!} c^2 s^2 - \frac{\alpha(1-\alpha)(2-\alpha)}{3!} c^3 s^3 - \cdots$. It is straightforward to show that all coefficients of series expansion of $L(s)$ are non-negative, which shows that $L(s)$ is a pgf.

(3) Let $r_j = L^{(j)}(0)/j!$. It suffices to show that $L^{(j)}(0) \geq 0$ for $j = 0, 1, 2, \ldots$. Check it when $j = 0$.

$$L(0) = 1 - \alpha^\theta (1-\gamma)^\theta \left[(1-\alpha)\gamma + (1-\gamma)\right]^{-\theta} = 1 - \left(\frac{\alpha(1-\gamma)}{1-\alpha\gamma}\right)^\theta \geq 0.$$

Now consider the derivatives.

$$\begin{aligned}
L'(s) &= -\alpha^\theta (1-\gamma)^\theta (-\theta) \left[(1-\alpha)\gamma + (1-\gamma)(1-s)^{-1/\theta}\right]^{-\theta-1} (1-\gamma)(1/\theta)(1-s)^{-1/\theta-1} \\
&= \alpha^\theta (1-\gamma)^{\theta+1} \left[(1-\alpha)\gamma + (1-\gamma)(1-s)^{-1/\theta}\right]^{-\theta-1} (1-s)^{-1/\theta-1} \\
&= \alpha^\theta \left[1 + \frac{(1-\alpha)\gamma}{1-\gamma}(1-s)^{1/\theta}\right]^{-(\theta+1)}, \\
L''(s) &= \frac{\theta+1}{\theta} \cdot \frac{(1-\alpha)\gamma}{1-\gamma} \alpha^\theta (1-s)^{1/\theta-1} \left[1 + \frac{(1-\alpha)\gamma}{1-\gamma}(1-s)^{1/\theta}\right]^{-(\theta+2)}.
\end{aligned}$$

When $0 \leq s < 1$, it follows that $L'(s) > 0$ and $L''(s) > 0$. Starting from $j = 3$, higher order derivatives are non-negative linear combination of products of form

$$(1-s)^{1/\theta - l} \left[1 + \frac{(1-\alpha)\gamma}{1-\gamma}(1-s)^{1/\theta}\right]^{-(\theta+k)}, \quad l \geq 1,\ k \geq 2. \tag{2.2.2}$$

Since for $0 \leq s < 1$

$$\begin{aligned}
&\frac{d}{ds}(1-s)^{1/\theta - l} = \frac{\theta l - 1}{\theta}(1-s)^{1/\theta-(l+1)} > 0, \\
&\frac{d}{ds}\left[1 + \frac{(1-\alpha)\gamma}{1-\gamma}(1-s)^{1/\theta}\right]^{-(\theta+k)} \\
&\quad = \frac{\theta+k}{\theta} \cdot \frac{(1-\alpha)\gamma}{1-\gamma}(1-s)^{1/\theta-1}\left[1 + \frac{(1-\alpha)\gamma}{1-\gamma}(1-s)^{1/\theta}\right]^{-(\theta+k+1)} \\
&\quad > 0,
\end{aligned}$$

(2.2.2) follows by induction, and we can conclude that all higher order derivatives are positive when $0 \leq s < 1$. This leads to that $L^{(j)}(0) > 0\ (j \geq 1)$. Hence, $L(s)$ is a pgf.

(4) Since the pmf of Zeta(ρ) is

$$p_i = c \cdot (i+1)^{-(\rho+1)}, \quad i = 0, 1, 2, \ldots; \quad \rho > 0,\ c = \sum_{i=1}^{\infty} i^{-(\rho+1)},$$

the pgf of Zeta(ρ) is

$$G(s) = \sum_{i=0}^{\infty} p_i s^i = \sum_{i=0}^{\infty} c \cdot (i+1)^{-(\rho+1)} s^i.$$

Because $L(s)G(s) = (s-1)G'(s) + G(s)$, and

$$L(s)G(s) = p_0r_0 + (p_0r_1 + p_1r_0)s + (p_0r_2 + p_1r_1 + p_2r_0)s^2 + \cdots + \left(\sum_{k=0}^{i} p_k r_{i-k}\right) s^i + \cdots,$$

$$\begin{aligned}
&(s-1)G'(s) + G(s) \\
&\quad = (s-1)\cdot[p_1 + 2p_2s + 3p_3s^2 + \cdots] + [p_0 + p_1s + p_2s^2 + p_3s^3 + \cdots] \\
&\quad = (p_0 - p_1) + 2(p_1 - p_2)s + 3(p_2 - p_3)s^2 + \cdots + (i+1)(p_i - p_{i+1})s^i + \cdots,
\end{aligned}$$

we obtain

$$\left\{\begin{array}{l}
p_0r_0 = p_0 - p_1, \\
p_0r_1 + p_1r_0 = 2(p_1 - p_2), \\
p_0r_2 + p_1r_1 + p_2r_0 = 3(p_2 - p_3), \\
\quad\vdots \\
\sum_{k=0}^{i} p_k r_{i-k} = (i+1)(p_i - p_{i+1}), \\
\quad\vdots
\end{array}\right.$$

or more specifically,

$$\left\{\begin{array}{l}
r_0 = 1 - 2^{-(\rho+1)}, \\
r_1 + 2^{-(\rho+1)}r_0 = 2[2^{-(\rho+1)} - 3^{-(\rho+1)}], \\
\quad\vdots \\
r_i + 2^{-(\rho+1)}r_{i-1} + \cdots + (i+1)^{-(\rho+1)}r_0 = (i+1)[(i+1)^{-(\rho+1)} - (i+2)^{-(\rho+1)}], \\
\quad\vdots
\end{array}\right.$$

Thus

$$\begin{aligned}
r_0 &= 1 - 2^{-(\rho+1)} > 0, \\
r_1 &= 2[2^{-(\rho+1)} - 3^{-(\rho+1)}] - 2^{-(\rho+1)}[1 - 2^{-(\rho+1)}] = 2^{-(\rho+1)} + 4^{-(\rho+1)} - 2\cdot 3^{-(\rho+1)} \\
&> 2\sqrt{2^{-(\rho+1)}4^{-(\rho+1)}} - 2\cdot 3^{-(\rho+1)} = 2[\sqrt{8^{-(\rho+1)}} - \sqrt{9^{-(\rho+1)}}] > 0.
\end{aligned}$$

Assume $r_i \geq 0$ $(i \geq 2)$. We show that $r_{i+1} \geq 0$. To prove it, we apply the contradiction method, which supposes $r_{i+1} < 0$. Note that $\frac{1}{2} < \frac{2}{3} < \cdots < \frac{i}{i+1}$. This leads to

$$(k+1)^{-(\rho+1)} < k^{-(\rho+1)} \left(\frac{i+1}{i+2}\right)^{(\rho+1)}, \quad k = 1, 2, \ldots, i,$$

Consider the new equation

$$\begin{aligned}
&\left(\frac{i+1}{i+2}\right)^{(\rho+1)}\left[r_i + 2^{-(\rho+1)}r_{i-1} + \cdots + (i+1)^{-(\rho+1)}r_0\right] \\
&\qquad - \left[r_{i+1} + 2^{-(\rho+1)}r_i + \cdots + (i+2)^{-(\rho+1)}r_0\right] \qquad (2.2.3)\\
&= \left(\frac{i+1}{i+2}\right)^{(\rho+1)}(i+1)\left[\frac{1}{(i+1)^{\rho+1}} - \frac{1}{(i+2)^{\rho+1}}\right] - (i+2)\left[\frac{1}{(i+2)^{\rho+1}} - \frac{1}{(i+3)^{\rho+1}}\right].
\end{aligned}$$

The left hand side is

$$\begin{aligned}
LHS \;=\; & -r_{i+1} + \left[\left(\frac{i+1}{i+2}\right)^{(\rho+1)} - 2^{-(\rho+1)}\right]r_i + \left[2^{-(\rho+1)}\left(\frac{i+1}{i+2}\right)^{(\rho+1)} - 3^{-(\rho+1)}\right]r_{i-1} \\
& + \cdots + \left[i^{-(\rho+1)}\left(\frac{i+1}{i+2}\right)^{(\rho+1)} - (i+1)^{-(\rho+1)}\right]r_1 \\
& + \left[(i+1)^{-(\rho+1)}\left(\frac{i+1}{i+2}\right)^{(\rho+1)} - (i+2)^{-(\rho+1)}\right]r_0 \\
\geq\; & -r_{i+1} \;>\; 0.
\end{aligned}$$

Denote the right hand side of (2.2.3) as a function of ρ:

$$RHS \;=\; h(\rho) \;=\; -\frac{1}{(i+2)^{\rho+1}} - \frac{(i+1)^{\rho+2}}{(i+2)^{2\rho+2}} + \frac{i+2}{(i+3)^{\rho+1}}.$$

Then

$$h'(\rho) = \frac{\log(i+2)}{i+2}\frac{1}{(i+2)^{\rho}} - \frac{i+2}{i+3}\log(i+3)\frac{1}{(i+3)^{\rho}} + \log\left(\frac{(i+2)^2}{i+1}\right)\frac{(i+1)^{\rho+2}}{(i+2)^{2\rho+2}}.$$

As ρ increases, $\left(\frac{i+2}{i+3}\right)^{\rho}$ decreases to 0, thus, $\frac{\log(i+2)}{i+2}\frac{1}{(i+2)^{\rho}} - \frac{i+2}{i+3}\log(i+3)\frac{1}{(i+3)^{\rho}}$ will eventually be positive. This means $h'(\rho)$ will eventually be positive when ρ increases, although it could be negative at the beginning. Hence, $h(\rho)$ could be either always increasing, or decreasing first and then increasing. Thus

$$h(\rho) \leq \max(h(0), h(\infty)).$$

By calculation, we have

$$h(0) = -\frac{1}{(i+2)^2(i+3)} < 0, \quad h(\infty) = \lim_{\rho\to\infty} h(\rho) = 0.$$

Thus $h(\rho) \leq \max(h(0), h(\infty)) = h(\infty) = 0$. This contradicts to that $LHS = RHS$, which implies that $r_{i+1} \geq 0$. Therefore, $L(s)$ is a pgf.

(5) Rewrite and expand

$$
\begin{aligned}
L(s) &= 1+\frac{s-1}{(1-\beta s)\left(1-\frac{\beta-\alpha}{1-\alpha}s\right)} = 1+\frac{1-\alpha}{\alpha}\times\frac{1}{1-\beta s}-\frac{1}{\alpha}\times\frac{1}{\left(1-\frac{\beta-\alpha}{1-\alpha}s\right)} \\
&= 1+\frac{1-\alpha}{\alpha}\left[1+\beta s+\beta^2 s^2+\beta^3 s^3+\cdots\right] \\
&\quad -\frac{1}{\alpha}\left[1+\left(\frac{\beta-\alpha}{1-\alpha}\right)s+\left(\frac{\beta-\alpha}{1-\alpha}\right)^2 s^2+\left(\frac{\beta-\alpha}{1-\alpha}\right)^3 s^3+\cdots\right] \\
&= \frac{1}{\alpha}\left[(1-\alpha)\beta-\left(\frac{\beta-\alpha}{1-\alpha}\right)\right]s+\frac{1}{\alpha}\left[(1-\alpha)\beta^2-\left(\frac{\beta-\alpha}{1-\alpha}\right)^2\right]s^2 \\
&\quad +\frac{1}{\alpha}\left[(1-\alpha)\beta^3-\left(\frac{\beta-\alpha}{1-\alpha}\right)^3\right]s^3+\cdots \\
&= \frac{1}{\alpha}\sum_{i=1}^{\infty}\left[(1-\alpha)\beta^i-\left(\frac{\beta-\alpha}{1-\alpha}\right)^i\right]s^i.
\end{aligned}
$$

Now we verify if all coefficients are non-negative, i.e.,

$$(1-\alpha)\beta^i-\left(\frac{\beta-\alpha}{1-\alpha}\right)^i \ge 0, \quad i=1,2,3,\ldots.$$

This is equivalent to

$$(1-\alpha)^{i+1} \ge \left(1-\frac{\alpha}{\beta}\right)^i, \quad i=1,2,3,\ldots. \tag{2.2.4}$$

If $\alpha<\beta$, then $1-\frac{\alpha}{\beta}>0$. Thus, (2.2.4) holds iff

$$1-\frac{\alpha}{\beta} \le (1-\alpha)^{(i+1)/i}, \quad i=1,2,3,\ldots,$$

iff

$$1-\frac{\alpha}{\beta} \le \min_{i\in\mathcal{N}}\left\{(1-\alpha)^{(i+1)/i}\right\} = (1-\alpha)^2, \quad \beta \le \frac{\alpha}{1-(1-\alpha)^2} = \frac{1}{2-\alpha}.$$

In this situation, we obtain the range of β:

$$\alpha<\beta\le\frac{1}{2-\alpha}.$$

If $\alpha\ge\beta$, $1-\frac{\alpha}{\beta}\le 0$. We only need to consider i being even integers. (2.2.4) can be rewritten as

$$(1-\alpha)^{2j+1} \ge \left(\frac{\alpha}{\beta}-1\right)^{2j}, \quad j=1,2,3,\ldots.$$

This holds if

$$\frac{\alpha}{\beta} - 1 \le \min_{j \in \mathcal{N}} \left\{ (1-\alpha)^{(2j+1)/(2j)} \right\} = (1-\alpha)^{3/2}.$$

Hence, the range of β in this situation is

$$\frac{\alpha}{1 + (1-\alpha)^{3/2}} \le \beta \le \alpha.$$

These imply that for β in above ranges, the function $L(s)$ is a pgf.

(6) We rewrite and expand

$$\begin{aligned}
L(s) &= 1 + \frac{(1-\gamma s)(s-1)}{(1-\beta s)\left(1 - \frac{\beta-\alpha}{1-\alpha} s\right)} \\
&= 1 + (1-\gamma s) \left\{ \frac{1-\alpha}{\alpha} \times \frac{1}{1-\beta s} - \frac{1}{\alpha} \times \frac{1}{\left(1 - \frac{\beta-\alpha}{1-\alpha} s\right)} \right\} \\
&= 1 + \frac{1-\alpha}{\alpha} \times \frac{1-\gamma s}{1-\beta s} - \frac{1}{\alpha} \times \frac{1-\gamma s}{1 - \frac{\beta-\alpha}{1-\alpha} s} \\
&= 1 + \frac{1-\alpha}{\alpha} \left(\frac{\gamma}{\beta} + \frac{\beta-\gamma}{\beta} \times \frac{1}{1-\beta s} \right) \\
&\quad - \frac{1}{\alpha} \left(\frac{(1-\alpha)\gamma}{\beta-\alpha} + \frac{(\beta-\alpha)-(1-\alpha)\gamma}{\beta-\alpha} \times \frac{1}{1 - \frac{\beta-\alpha}{1-\alpha} s} \right) \\
&= 1 + \frac{(1-\alpha)\gamma}{\alpha\beta} - \frac{(1-\alpha)\gamma}{\alpha(\beta-\alpha)} + \frac{(1-\alpha)(\beta-\gamma)}{\alpha\beta} \times \frac{1}{1-\beta s} \\
&\quad - \frac{\beta-\gamma-\alpha+\alpha\gamma}{\alpha(\beta-\alpha)} \times \frac{1}{\left(1 - \frac{\beta-\alpha}{1-\alpha} s\right)} \\
&= 1 + \frac{(1-\alpha)\gamma}{\alpha\beta} - \frac{(1-\alpha)\gamma}{\alpha(\beta-\alpha)} + \frac{(1-\alpha)(\beta-\gamma)}{\alpha\beta} \times \left[1 + \beta s + \beta^2 s^2 + \beta^3 s^3 + \cdots\right] \\
&\quad - \frac{\beta-\gamma-\alpha+\alpha\gamma}{\alpha(\beta-\alpha)} \times \left[1 + \left(\frac{\beta-\alpha}{1-\alpha}\right) s + \left(\frac{\beta-\alpha}{1-\alpha}\right)^2 s^2 + \left(\frac{\beta-\alpha}{1-\alpha}\right)^3 s^3 + \cdots\right] \\
&= \frac{1}{\alpha} \sum_{i=1}^{\infty} \left[(1-\alpha)(\beta-\gamma)\beta^{i-1} - (\beta-\gamma-\alpha+\alpha\gamma)(\beta-\alpha)^{i-1}(1-\alpha)^{-i}\right] s^i.
\end{aligned}$$

We want

$$(1-\alpha)(\beta-\alpha)\beta^{i-1} - (\beta-\gamma-\alpha+\alpha\gamma)(\beta-\alpha)^{i-1}(1-\alpha)^{-i} \ge 0, \qquad \text{for } i \ge 1.$$

The inequality can be written as

$$(\beta-\gamma)(1-\alpha)^2 \geq (\beta-\gamma-\alpha+\alpha\gamma)\left(\frac{1-\alpha/\beta}{1-\alpha}\right)^{i-1}, \qquad \text{for } i \geq 1.$$

When $\alpha < \beta$, $0 \leq \frac{1-\alpha/\beta}{1-\alpha} < 1$. In this case, it simply results in $(\beta-\gamma)(1-\alpha)^2 \geq (\beta-\gamma-\alpha+\alpha\gamma)$. Further calculations lead to

$$\alpha < \beta \quad \text{and} \quad 1+\gamma-\alpha\gamma-\beta(2-\alpha) \geq 0,$$

i.e., $\alpha < \beta \leq \frac{1+\gamma-\alpha\gamma}{2-\alpha}$.

When $\alpha \geq \beta$, $\frac{1-\alpha/\beta}{1-\alpha} \leq 0$. We obtain

$$\begin{aligned}(\beta-\gamma)(1-\alpha)^2 &\geq (\beta-\gamma-\alpha+\alpha\gamma)(-1)^{i-1}\left(\frac{\alpha/\beta-1}{1-\alpha}\right)^{i-1} \\ &= (\alpha-\beta+\gamma-\alpha\gamma)(-1)^{i}\left(\frac{\alpha/\beta-1}{1-\alpha}\right)^{i-1}, \qquad \text{for } i \geq 1.\end{aligned}$$

Note that $\alpha-\beta+\gamma-\alpha\gamma \geq 0$. First, β can't be smaller than γ. Otherwise, the left hand side is always negative while the right hand side alternates in sign, which is a contradiction. Thus, $\beta \geq \gamma$, and we only need to consider those situations where the right hand side is non-negative. Secondly, if $\frac{\alpha/\beta-1}{1-\alpha} > 1$, i.e., $\beta < \frac{\alpha}{2-\alpha}$, the right hand side can go to infinity. This is impossible because the left hand side is finite. Hence, it must be $\frac{\alpha}{2-\alpha} \leq \beta$, and consequently, $0 \leq \frac{\alpha/\beta-1}{1-\alpha} \leq 1$. Under such a situation, we can simplify the inequality to be

$$(\beta-\gamma)(1-\alpha)^2 \geq (\alpha-\beta+\gamma-\alpha\gamma) \times \frac{\alpha/\beta-1}{1-\alpha},$$

or equivalently, $\beta(\beta-\gamma)(1-\alpha)^3 \geq (\alpha-\beta+\gamma-\alpha\gamma)(\alpha-\beta)$.

In summary, we obtain two groups of conditions: (1) $\alpha < \beta \leq (1+\gamma-\alpha\gamma)/(2-\alpha)$; (2) $\max(\gamma, \alpha/(2-\alpha)) \leq \beta \leq \alpha$ and $\beta(\beta-\gamma)(1-\alpha)^3 \geq (\alpha-\beta+\gamma-\alpha\gamma)(\alpha-\beta)$. Under these conditions, the function $L(s)$ is a pgf.

Remark: More on the compounding operation. The compound Poisson rv can be represented as random summation:

$$Z = \sum_{i=0}^{Y} X_i, \quad X_0 = 0,\ X_i\ iid,\ Y \text{ is distributed in Poisson.} \tag{2.2.5}$$

Here X_i can be not only a non-negative integer-valued rv, but also positive-valued or real-valued rv. The random variable Y can be extended to any other non-negative integer-valued rv. For example, if Y is a Bernoulli rv, the compound operation is the random zero-truncation operation, it is often used in **Zero-inflated** models for economic applications. See Winkelmann [1997], p. 107-108, and references therein.

Now we briefly study a few properties of this extended compounding operation such as the mean, variance and probability mass at zero.

$$
\begin{aligned}
\mathbf{E}(Z) &= \mathbf{E}\left(\sum_{i=0}^{Y} X_i\right) = \mathbf{E}\left(\mathbf{E}\left(\sum_{i=0}^{Y} X_i \Big| Y\right)\right) = \mathbf{E}(X_1)\mathbf{E}(Y), \\
\mathbf{Var}(Z) &= \mathbf{Var}\left(\sum_{i=0}^{Y} X_i\right) = \mathbf{E}\left(\mathbf{Var}\left(\sum_{i=0}^{Y} X_i \Big| Y\right)\right) + \mathbf{Var}\left(\mathbf{E}\left(\sum_{i=0}^{Y} X_i \Big| Y\right)\right) \\
&= \mathbf{E}(\mathbf{Var}(X_1)Y) + \mathbf{Var}(\mathbf{E}(X_1)Y) \\
&= \mathbf{Var}(X_1)\mathbf{E}(Y) + (\mathbf{E}(X_1))^2\,\mathbf{Var}(Y), \\
D(Z) &= \frac{\mathbf{Var}(Z)}{\mathbf{E}(Z)} = \frac{\mathbf{Var}(X_1)\mathbf{E}(Y) + (\mathbf{E}(X_1))^2\,\mathbf{Var}(Y)}{\mathbf{E}(X_1)\mathbf{E}(Y)} \\
&= \frac{\mathbf{Var}(X_1)}{\mathbf{E}(X_1)} + \mathbf{E}(X_1)\frac{\mathbf{Var}(Y)}{\mathbf{E}(Y)}, \\
\Pr[Z=0] &= \Pr\left[\sum_{i=0}^{Y} X_i = 0\right] = \Pr[Y=0] + \Pr[X_1=0]\Pr[Y=1] + (\Pr[X_1=0])^2\Pr[Y=2] \\
&\qquad + \cdots + (\Pr[X_1=0])^i\Pr[Y=i] + \cdots \\
&= \sum_{i=0}^{\infty}(\Pr[X_1=0])^i\Pr[Y=i] \\
&= G_Y(\Pr[X_1=0]), \qquad\qquad \text{where } G_Y(s) \text{ is the pgf of } Y.
\end{aligned}
$$

If X_1 is also a non-negative integer-valued rv, we have

$$
G_Z(s) = \mathbf{E}\left(s^Z\right) = \mathbf{E}\left(s^{\sum_{i=0}^{Y} X_i}\right) = \mathbf{E}\left(\mathbf{E}\left(s^{\sum_{i=0}^{Y} X_i}\Big| Y\right)\right) = \mathbf{E}\left([G_{X_1}(s)]^Y\right) = G_Y(G_{X_1}(s)),
$$

where $G_{X_1}(s)$, $G_Y(s)$ are the pgf of X_1 and Y.

With these results, we have the following proposition.

Proposition 2.2.4 *Suppose X_1 is a non-negative integer-valued rv, and (2.2.5) holds.*

(1) If $\mathbf{E}(X_1)$ is positive and $\mathbf{Var}(X_1)$ exists, then $D(Z) > D(X_1)$,

(2) If $\Pr[Z = 0] \geq 1 - \Pr[Y = 1]$, *then* $\Pr[Z = 0] > \Pr[X_1 = 0]$.

Proof: *Apply the above results.*

(1) Since $\mathbf{E}(X_1) > 0$ *and* $\mathbf{E}(Y) > 0$, *we have*

$$D(Z) = \frac{\mathbf{Var}(X_1)}{\mathbf{E}(X_1)} + \mathbf{E}(X_1)\frac{\mathbf{Var}(Y)}{\mathbf{E}(Y)} > \frac{\mathbf{Var}(X_1)}{\mathbf{E}(X_1)} = D(X_1).$$

This means that after compounding, the index of dispersion becomes larger.

(2) Because $G_Y(\Pr[X_1 = 0]) = \sum_{i=0}^{\infty}(\Pr[X_1 = 0])^i \Pr[Y = i] \geq \Pr[Y = 0] + \Pr[X_1 = 0]\Pr[Y = 1]$, *it follows that*

$$\begin{aligned} \Pr[Z = 0] &\geq \Pr[Y = 0] + \Pr[X_1 = 0]\Pr[Y = 1] \\ &= \Pr[Y = 0] + \Pr[X_1 = 0](1 - \Pr[Y = 0]) \\ &= \Pr[Y = 0](1 - \Pr[X_1 = 0]) + \Pr[X_1 = 0] \\ &\geq \Pr[X_1 = 0], \end{aligned}$$

where the equality holds only when $\Pr[Y = 1] = 1$ *or* $\Pr[X_1 = 0] = 1$, *which are extreme cases. Hence, the compound operation results in a larger mass at zero in general.*

The compound Poisson and random zero-truncation operations result in a larger probability mass at zero. Such a property of a larger probability mass at zero makes the compound distribution an alternative candidate in modelling count data with a higher fraction of zeros.

2.2.2 Laplace transformation, moment generating function and characteristic function

The **Laplace transformation (LT), moment generating function (mgf)** and **characteristic function (cf)** are widely used in probability and statistics. They are defined as below.

Definition 2.1 *Let* X *be a rv. Then*

(1) the LT of X is

$$\phi_X(s) = \mathbf{E}(e^{-sX}), \quad s \in S_1, \quad \textit{where } S_1 = \{s : \mathbf{E}(e^{-sX}) < \infty\}.$$

(2) the mgf of X is

$$M_X(s) = \mathbf{E}(e^{sX}), \quad s \in S_2, \quad \textit{where } S_2 = \{s : \mathbf{E}(e^{sX}) < \infty\}.$$

(3) the cf of X is

$$\varphi_X(s) = \mathbf{E}(e^{isX}), \quad s \in (-\infty, \infty).$$

The relationships among the pgf, LT, mgf and cf, over approriate domains, are listed below:

$$\begin{aligned}
G_X(s) &= \phi_X(-\log s) = M_X(\log s) = \varphi_X(-i\log s),\\
\phi_X(s) &= G_X(e^{-s}) = M_X(-s) = \varphi_X(is),\\
M_X(s) &= G_X(e^{s}) = \phi_X(-s) = \varphi_X(-is),\\
\varphi_X(s) &= G_X(e^{is}) = M_X(is) = \phi_X(-is).
\end{aligned}$$

For positive-valued rv X, the LT is more convenient than mgf, because its convergence domain of s includes $\Re_0 = [0, \infty)$, a fixed set, while the domain of s for the mgf depends on the individual distribution. The LT is decreasing while mgf is increasing.

The mean and variance can be derived from both LT and mgf.

$$\mathbf{E}(X) = \begin{cases} \phi'_X(0), \\ M'_X(0), \\ -i\varphi'_X(0). \end{cases} \quad \text{and} \quad \mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = \begin{cases} \phi''_X(0) - (\phi'_X(0))^2, \\ M''_X(0) - (M'_X(0))^2, \\ -\varphi''_X(0) + (\varphi'_X(0))^2. \end{cases}$$

Note that the probability mass at zero can be obtained as

$$\Pr[X = 0] = \lim_{s\to\infty} \phi_X(s) = \phi_X(\infty) = \lim_{s\to-\infty} M_X(s) = M(-\infty).$$

Example 2.3 *Gamma distribution: Let X be the rv of Gamma(α, β), where α is the shape parameter and β is the rate parameter. Then the pdf is*

$$f_X(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad x > 0; \quad \alpha,\ \beta > 0.$$

The LT is

$$\phi_X(s) = \left(\frac{\beta}{\beta + s}\right)^{\alpha}, \quad s \in [0, \infty);$$

the mgf is

$$M_X(s) = \left(\frac{\beta}{\beta - s}\right)^{\alpha}, \quad s \in (-\infty, \beta);$$

and the cf is

$$\varphi_X(s) = \left(\frac{\beta}{\beta - is}\right)^{\alpha}, \quad s \in (-\infty, \infty).$$

The expectation and variance are

$$\mathbf{E}(X) = \alpha\beta^{-1} \quad \textit{and} \quad \mathbf{Var}(X) = \alpha\beta^{-2}.$$

The Gamma family contains a couple of special distribution cases:

- when $\alpha = 1$, it is Exponential(β),
- when $\alpha = k/2,\ \beta = 1/2$ (k is an integer), it is Chi squared, χ^2_k.

More examples of LT and mgf can be seen in the successive subsections.

Because the theory of continuous-time generalized AR(1) process heavily involves non-negative rv's, we focus on the LT in the rest of this subsection. The following theorem is important to characterize the LT of a non-negative rv. Refer to Bondesson [1992], p. 8-9.

Theorem 2.2.5 (Bernstein's theorem)
$\phi(s)$ is a LT iff

1. *$\phi(s)$ is* **completely monotone (CM)**, *i.e.,* $(-1)^i\phi^{(i)}(s) \geq 0,\ i \in \mathcal{N}_0,\ s \in (0, \infty)$.
2. $\phi(s) \longrightarrow 1, \quad$ *as* $s \to 0$.

Following two theorems are very useful to identify new LT. Very nice proofs for the first two can be found in Joe [1997], p. 374.

Theorem 2.2.6 *Let $\phi(s)$ be a LT. Then $\phi^{\alpha}(s)$ is a LT for all $\alpha > 0$ if and only if $-\log\phi(s)$ is an infinitely differentiable increasing function of $[0, \infty)$ onto $[0, \infty)$, with alternating signs for the derivatives.*

Theorem 2.2.7 *Suppose $\phi_1(s)$ and $\phi_2(s)$ are LT's. If $-\log\phi_1(s)$ is an infinitely differentiable increasing function of $[0,\infty)$ onto $[0,\infty)$, with alternating signs for the derivatives, then $\phi_2(-\log\phi_1(s))$ is a LT.*

An explanation of $\phi_2(-\log\phi_1(s))$ is discussed later in Section 3.4; refer to Bondesson [1992], p. 17. Note that for a LT $\phi(s)$, $\phi^\alpha(s)$ being a LT for all $\alpha > 0$ means that $\phi(s)$ is the LT of an infinitely divisible distribution. See Section 2.3.1.

By Theorem 2.2.6, we can prove that the exponential function $\phi(s) = \exp\{\phi_0(s)\}$ is a LT if $\phi_0(s)$ satisfies that $\phi_0(s) \le 0$, $(-1)^i\phi_0^{(i)}(s) \ge 0$, $i = 1,2,3,\ldots$, $s \in (0,\infty)$, and $\phi_0(s) \to 0$ as $s \to 0$. In some exponential form situations, the conditions of Theorem 2.2.7 are easier to be verified than Theorem 2.2.5.

Applying these theorems, we can obtain the following results, which are used in the theory of continuous-time generalized AR(1) processes.

Theorem 2.2.8 *The following functions $\phi(s)$ are LT of distribution with support on $[0,\infty)$.*

(1) $\phi(s) = \exp\left\{-\lambda\cdot\frac{s(1-\gamma+\gamma s)}{\beta+s}\right\}$, *where* $\lambda,\beta > 0$, $0 \le \gamma < 1$ *and* $\gamma \le \frac{1}{1+\beta}$.

(2) $\phi(s) = \exp\left\{\frac{1-\gamma}{1-\gamma-\gamma u}\log\left(\frac{(1-\gamma+\gamma s)u}{(1-\gamma)(u+s)}\right)\right\} = \left(\frac{(1-\gamma+\gamma s)u}{(1-\gamma)(u+s)}\right)^{\frac{1-\gamma}{1-\gamma-\gamma u}}$, *where* $u \ge 0$, $0 \le \gamma < 1$.

(3) $\phi(s) = \exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right\}$, *where* $0 \le \alpha \le 1$, $\theta \ge 0$.

(4) $\phi(s) = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\}$, *where* $0 \le \alpha \le 1$, $0 \le \gamma < 1$.

(5) $\phi(s) = \exp\left\{-\left[\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma+(1-\gamma)s^{-1/\theta}}\right]^\theta\right\}$, *where* $0 \le \alpha \le 1$, $\theta \ge 1$ *and* $0 < \gamma < 1$.

(6) $\phi(s) = \exp\left\{\lambda\cdot\frac{-s}{(1+\beta s)^{1/2}}\right\}$, *where* λ, $\beta > 0$.

Proof:

(1) When $s = 0$, $\phi(s) = e^0 = 1$. Hence, it suffices to show the complete monotonicity of function $\phi(s)$. Since $\gamma \le \frac{1}{1+\beta}$, we know that $1-\gamma(1+\beta) \ge 0$. Taking the first order derivative, we

obtain

$$\begin{aligned}
\phi'(s) &= -\lambda \cdot \frac{(1-\gamma+2\gamma s)(\beta+s)-s(1-\gamma+\gamma s)}{(\beta+s)^2} \exp\left\{-\lambda \cdot \frac{s(1-\gamma+\gamma s)}{\beta+s}\right\} \\
&= -\lambda \cdot \frac{\beta(1-\gamma)+2\beta\gamma s+\gamma s^2}{(\beta+s)^2} \exp\left\{-\lambda \cdot \frac{s(1-\gamma+\gamma s)}{\beta+s}\right\} \\
&= -\lambda\gamma \exp\left\{-\lambda \cdot \frac{s(1-\gamma+\gamma s)}{\beta+s}\right\} \\
&\qquad -\lambda\beta[1-\gamma(1+\beta)](\beta+s)^{-2} \exp\left\{-\lambda \cdot \frac{s(1-\gamma+\gamma s)}{\beta+s}\right\} \\
&\leq 0.
\end{aligned}$$

By induction, the higher derivatives $\phi^{(i)}(s)$ $(i \geq 2)$ are the sum of terms of form (omitting the coefficients)

$$(-1)^i(\beta+s)^{-k} \exp\left\{-\lambda \cdot \frac{s(1-\gamma+\gamma s)}{\beta+s}\right\}, \quad k \geq 0. \tag{2.2.6}$$

This follows because derivatives of (2.2.6) lead to two terms having the same form (ignoring coefficients that don't depend on s). With this property, we can conclude that the derivatives alternate the signs. By Theorem 2.2.5, $\phi(s)$ is a LT.

(2) First we prove that $\frac{1+as}{1+bs}$ $(0 \leq a \leq b)$ is a LT. This is because

$$\frac{1+as}{1+bs} = \frac{a}{b} + \left(1-\frac{a}{b}\right)\frac{1}{1+bs} = \frac{a}{b} + \left(1-\frac{a}{b}\right)\frac{1/b}{1/b+s},$$

the LT of the zero truncation of the Exponential$(1/b)$ distribution. Secondly, we show that $-\log\frac{1+as}{1+bs}$ is an infinitely differentiable increasing function of $[0,\infty)$ onto $[0,\infty)$, with alternating signs for the derivatives. Since $\frac{1+as}{1+bs}$ is decreasing, $-\log\frac{1+as}{1+bs}$ is increasing, and

$$\begin{aligned}
-\log\frac{1+as}{1+bs} &= \log(1+bs)-\log(1+as), \\
\left(-\log\frac{1+as}{1+bs}\right)' &= \frac{b}{1+bs}-\frac{a}{1+as} = \frac{b-a}{(1+as)(1+bs)}, \\
&\vdots \\
\left(-\log\frac{1+as}{1+bs}\right)^{(i)} &= \frac{(-1)^{i-1}b^i}{(1+bs)^i}-\frac{(-1)^{i-1}a^i}{(1+as)^i} = (-1)^{i-1}\frac{(b+abs)^i-(a+abs)^i}{(1+as)^i(1+bs)^i}, \\
&\vdots
\end{aligned}$$

Because $\frac{(b+abs)^i-(a+abs)^i}{(1+as)^i(1+bs)^i} \geq 0$ for all $i \geq 1$, we conclude that the derivatives of $-\log\frac{1+as}{1+bs}$ take alternating signs. Hence, by Theorem 2.2.6, $\left(\frac{1+as}{1+bs}\right)^c$ $(c > 0)$ is a LT. Rewrite

$$\phi(s) = \left(\frac{(1-\gamma+\gamma s)u}{(1-\gamma)(u+s)}\right)^{\frac{1-\gamma}{1-\gamma-\gamma u}} = \begin{cases} \left(\frac{1+\frac{u\gamma}{(1-\gamma)u}s}{1+\frac{1-\gamma}{(1-\gamma)u}s}\right)^{\frac{1-\gamma}{1-\gamma-\gamma u}}, & \text{if } 1-\gamma-\gamma u \geq 0, \\ \left(\frac{1+\frac{1-\gamma}{(1-\gamma)u}s}{1+\frac{u\gamma}{(1-\gamma)u}s}\right)^{-\frac{1-\gamma}{1-\gamma-\gamma u}}, & \text{if } 1-\gamma-\gamma u \leq 0. \end{cases}$$

When either $1-\gamma-\gamma u \geq 0$ or $1-\gamma-\gamma u \leq 0$, $\phi(s)$ has the form of $\left(\frac{1+as}{1+bs}\right)^c$ $(0 \leq a \leq b,\ c > 0)$. Therefore, $\phi(s)$ is a LT.

(3) It is clear that $\phi(s) \longrightarrow 1$ when $s \to 0$. Hence, it suffices to show the complete monotonicity of function $\phi(s)$. Taking the first and second order derivatives, we obtain

$$\begin{aligned} \phi'(s) &= \left(\exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right\}\right)' \\ &= -\alpha\left[1+(e^\theta-1)s\right]^{\alpha-1}\exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right\}, \\ \phi''(s) &= \left(-\alpha\left[1+(e^\theta-1)s\right]^{\alpha-1}\exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right\}\right)' \\ &= (\alpha)(1-\alpha)\left[1+(e^\theta-1)s\right]^{\alpha-2}\exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right\} \\ &\quad +\alpha^2\left[1+(e^\theta-1)s\right]^{2\alpha-2}\exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right\}. \end{aligned}$$

Obviously, $\phi'(s)$ has negative sign while $\phi''(s)$ has positive sign. By induction, the higher order derivatives are the sum of terms of form (omitting the coefficients)

$$\left[1+(e^\theta-1)s\right]^{m\cdot\alpha-n}\exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right\}, \quad 1 \leq m \leq n.$$

Such a term has a derivative with negative sign, just like $\phi''(s)$ changes the sign of $\phi'(s)$. Hence, the derivatives $\phi^{(i)}(s)$ alternate in sign. This shows the CM property of $\phi(s)$. By Theorem 2.2.5, we conclude that $\phi(s)$ is a LT.

(4) Rewrite

$$\phi(s) = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\} = \exp\left\{\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma}\left[\frac{(1-\gamma)/[(1-\alpha)\gamma]}{(1-\gamma)/[(1-\alpha)\gamma]+s}-1\right]\right\}.$$

Note that $\frac{(1-\gamma)/[(1-\alpha)\gamma]}{(1-\gamma)/[(1-\alpha)\gamma]+s}$ is the LT of Exponential$((1-\gamma)/[(1-\alpha)\gamma])$. Hence, $\phi(s)$ is the LT of compound Poisson with the Exponential$((1-\gamma)/[(1-\alpha)\gamma])$ distribution.

(5) Rewrite

$$\begin{aligned}
\phi(s) &= \exp\left\{-\left[\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma+(1-\gamma)s^{-1/\theta}}\right]^{\theta}\right\} \\
&= \exp\left\{-\left[\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma}\right]^{\theta}\left[1+\frac{(1-\gamma)}{(1-\alpha)\gamma}s^{-\frac{1}{\theta}}\right]^{-\theta}\right\} \\
&= \exp\left\{\left[\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma}\right]^{\theta}\left(1-\left(\frac{1-\gamma}{(1-\alpha)\gamma}\right)^{-\theta}\left[\frac{(1-\alpha)\gamma}{(1-\gamma)}+s^{-\frac{1}{\theta}}\right]^{-\theta}-1\right)\right\}.
\end{aligned}$$

Let $\phi_0(s) = 1-\left(\frac{1-\gamma}{(1-\alpha)\gamma}\right)^{-\theta}\left[\frac{(1-\alpha)\gamma}{(1-\gamma)}+s^{-\frac{1}{\theta}}\right]^{-\theta}$. We prove that $\phi_0(s)$ is a LT. First, when $s\to 0$,

$$\phi_0(s) = 1-\left(\frac{1-\gamma}{(1-\alpha)\gamma}\right)^{-\theta}\frac{s}{\left[1+\frac{(1-\alpha)\gamma}{(1-\gamma)}s^{\frac{1}{\theta}}\right]^{\theta}} \longrightarrow 1.$$

Hence, it suffices to show the CM property of $\phi_0(s)$. Now check the first and second order derivative of $\phi_0(s)$. Denote $C = \left(\frac{1-\gamma}{(1-\alpha)\gamma}\right)^{-\theta}$. We obtain

$$\begin{aligned}
\phi_0'(s) &= -C\cdot(-\theta)\left[\frac{(1-\alpha)\gamma}{(1-\gamma)}+s^{-\frac{1}{\theta}}\right]^{-\theta-1}\left(-\frac{1}{\theta}\right)s^{-\frac{1}{\theta}-1} \\
&= -C\left[\frac{(1-\alpha)\gamma}{(1-\gamma)}s^{\frac{1}{\theta}}+1\right]^{-(\theta+1)}, \\
\phi_0''(s) &= -C\cdot[-(\theta+1)]\left[\frac{(1-\alpha)\gamma}{(1-\gamma)}s^{\frac{1}{\theta}}+1\right]^{-(\theta+1)-1}\left(\frac{(1-\alpha)\gamma}{(1-\gamma)}\cdot\frac{1}{\theta}\right)s^{\frac{1}{\theta}-1} \\
&= C\frac{\theta+1}{\theta}\frac{(1-\alpha)\gamma}{(1-\gamma)}\left[\frac{(1-\alpha)\gamma}{(1-\gamma)}s^{\frac{1}{\theta}}+1\right]^{-(\theta+2)}s^{\frac{1}{\theta}-1}.
\end{aligned}$$

Note that $\phi_0''(s)$ alternates the sign of $\phi_0'(s)$. By induction, the higher order derivatives are the sum of terms of form (omitting the coefficients)

$$\left[\frac{(1-\alpha)\gamma}{(1-\gamma)}s^{\frac{1}{\theta}}+1\right]^{-(\theta+k)}s^{m/\theta-n}, \qquad k\geq 2,\ 1\leq m\leq n.$$

Since $\left[\frac{(1-\alpha)\gamma}{(1-\gamma)} s^{1/\theta} + 1\right]^{-(\theta+k)}$ and $s^{m/\theta-n}$ have derivatives with negative sign, it is straightforward to show that higher order derivatives $\phi_0^{(i)}(s)$ $(i \geq 3)$ alternate in sign. Lastly, we check that if $\phi_0(s) \geq 0$ for $s \in (0, \infty)$. Since $\phi_0(s)$ is decreasing, it follows that

$$\phi_0(s) \geq \phi_0(\infty) = 1 - C(C^{-1} + 0) = 1 - 1 = 0.$$

Thus, $\phi_0(s)$ is non-negative. This completes the proof that $\phi_0(s)$ is a LT. Therefore, $\phi(s)$ is the LT of the compound Poisson with the distribution characterized by the LT $\phi_0(s)$.

(6) The proof is similar to that of (3). Rewrite

$$\phi(s) = \exp\left\{\lambda \cdot \frac{-s}{(1+\beta s)^{1/2}}\right\} = \exp\left\{\frac{\lambda}{\beta}\left[(1+\beta s)^{-1/2} - (1+\beta s)^{1/2}\right]\right\}.$$

Then, the first order derivative is

$$\begin{aligned} \phi'(s) &= \exp\left\{\lambda \cdot \frac{-s}{(1+\beta s)^{1/2}}\right\} \times \frac{\lambda}{\beta}\left[(1+\beta s)^{-1/2} - (1+\beta s)^{1/2}\right]' \\ &= -\frac{\lambda}{2}\left[(1+\beta s)^{-3/2} + (1+\beta s)^{-1/2}\right] \exp\left\{\lambda \cdot \frac{-s}{(1+\beta s)^{1/2}}\right\} \\ &< 0. \end{aligned}$$

By induction, the higher order derivatives $\phi^{(i)}(s)$ $(i \geq 2)$ are the sum of terms of form (omitting the coefficients)

$$(1+\beta s)^{-k/2} \exp\left\{\lambda \cdot \frac{-s}{(1+\beta s)^{1/2}}\right\}, \quad k \geq 1,$$

By the same reasoning in (3), we know that $\phi(s)$ is a LT.

The LT's in (1) and (2) of Theorem 2.2.8 lead to GC IV and GC III in Section 2.3.3. This distribution corresponding to (2) has non-zero probability mass at zero, and the mass is:

$$\phi(\infty) = \left(\frac{u\gamma}{1-\gamma}\right)^{\frac{1-\gamma}{1-\gamma-\gamma u}}.$$

The LT's in (3), (4) and (5) of Theorem 2.2.8 will be adopted as positive self-generalized distributions denoted as **P4**, **P2** and **P5** in Section 3.1.2. The LT in (3) belongs to Tweedie exponential dispersion family. See Section 2.3.2. Comparing with the LT of $\mathrm{Tw}_d(\mu, \sigma^2)$ there, we find it is the LT with specific parameter

$$d = \frac{\alpha - 2}{\alpha - 1} \geq 2, \quad \mu = 1/\alpha > 0, \quad \sigma^2 = (1 - \alpha)(e^\theta - 1)/\alpha^{\frac{1}{1-\alpha}} > 0.$$

The pdf is given in Section 2.3.2. This distribution does not have probability mass at zero, because $\phi(\infty) = 0$. However, the LT's in (4) and (5) are not in the Tweedie exponential dispersion family. This can be verified by comparing their LT forms. But they have non-zero (positive) probability masses at zero, which are

$$\exp\left\{-\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma}\right\} \quad \text{and} \quad \exp\left\{-\left[\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma}\right]^\theta\right\}$$

respectively. The LT in (6) will serve as the innovation of a stationary continuous-time generalized AR(1) process with inverse Gaussian margins (see Section 6.3). As to their explicit pdf forms, we are not clear at this moment.

Perhaps the most enjoyable and popular distribution is the Normal distribution. It has a lot of good properties, such as bell-shaped density and limiting distribution of an average of rv's. To enlarge and modify this family of distributions, the variance mixture of normal distributions was introduced:

$$X = \sqrt{Y} Z, \quad Z \sim \mathrm{N}(0, 1) \text{ and } Y > 0 \text{ is a rv independent of } Z.$$

Since X is a real rv, we prefer to calculate its cf. Let $\phi_Y(s)$ be the LT of Y. Because $\varphi_Z(s) = e^{-s^2/2}$, we have

$$\varphi_X(s) = \mathbf{E}\left(e^{isX}\right) = \mathbf{E}\left(e^{is\sqrt{Y}Z}\right) = \mathbf{E}\left(\mathbf{E}\left(e^{is\sqrt{Y}Z}\middle|Y\right)\right) = \mathbf{E}\left(e^{-Ys^2/2}\right) = \phi_Y(s^2/2).$$

See Bondesson [1992], p. 115. An equivalent version is the scale mixture of normal distributions, which is defined as

$$X = Z/Y, \quad Z \sim \mathrm{N}(0, 1) \text{ and } Y > 0 \text{ is a rv independent of } Z.$$

However, its cf can not be explicitly expressed via $\phi_Y(s)$:

$$\varphi_X(s) = \mathbf{E}\left(e^{isX}\right) = \mathbf{E}\left(e^{isZ/Y}\right) = \mathbf{E}\left(\mathbf{E}\left(e^{isZ/Y}\middle|Y\right)\right) = \mathbf{E}\left(e^{-s^2/(2Y^2)}\right).$$

Refer to Joe [1997], p. 132-134 and references therein. Any symmetric Stable distribution is the variance mixture of the normal distribution (see Bondesson [1992], p. 116). Other examples of the variance mixture of the normal distribution are shown in the EGGC family in the Section 2.3.3.

2.3 Particular families of distributions

We review and investigate some distribution families which are used either as distribution of innovation or as marginal distributions in the theory of continuous-time generalized AR(1) process.

2.3.1 Infinitely divisible, self-decomposable and stable distributions

Infinitely divisible (ID), self-decomposable and stable distributions appear quite often in the study of the continuous-time generalized AR(1) process. For ease of reference, we briefly review them here. Good references are Bondesson [1992] and Feller [1966a, 1966b].

Definition 2.2 (Infinite Divisibility) *Suppose $X \sim F$. If for each $n \geq 1$, X can be decomposed into the sum of n independent and identically distributed rv's, namely*

$$X \stackrel{d}{=} X_{n1} + X_{n2} + \cdots + X_{nn}, \qquad \textit{where } X_{n1}, X_{n2}, \ldots, X_{nn} \textit{ iid},$$

then the probability distribution F is said to be infinitely divisible (ID).

By the definition, it follows that $\varphi_X(s) = (\varphi_{X_{n1}}(s))^n$. This leads to that $\varphi_X^{1/n}(s)$ is a cf for any non-negative integer n. Thus, the ID is equivalent to that $\varphi_X^{\alpha}(s)$ is a cf for all $\alpha > 0$. This class of ID distributions is closed under convolution and weak limits. Some canonical representations of the mgf of the ID distributions are summarized below.

Proposition 2.3.1 *Suppose $X \sim F$, an ID distribution.*

- *When the support of F is $\Re$, the Lévy(-Khintchine) representation of mgf is*

$$M_X(s) = \exp\left\{as + \frac{\sigma^2}{2}\cdot s^2 + \int_{y\neq 0}\left(e^{sy} - 1 - \frac{sy}{1+y^2}\right)L(dy)\right\}, \quad \Re e\ s = 0,\ a \in \Re,$$

where the measure L satisfies $\int_{y\neq 0}\min(1, y^2)L(dy) < \infty$.

- *When the support of F is $\Re_+$, the representation of LT is*

$$\phi_X(s) = \exp\left\{-as + \int_{(0,\infty)}\left(e^{-sy} - 1\right)L(dy)\right\}, \quad a \geq 0,$$

where the Lévy measure L is non-negative and satisfies $\int_{(0,\infty)}\min(1, y)L(dy) < \infty$. The parameter a is called the left-extremity.

- *When the support of F is $\mathcal{N}_0$, the representation of pgf is*

$$G_X(s) = \exp\left\{\int_{(0,\infty)}\left(s^y - 1\right)L(dy)\right\} = \exp\left\{\lambda[Q(s) - 1]\right\},$$

where the Lévy measure L is non-negative and satisfies $\int_{(0,\infty)}\min(1, y)L(dy) < \infty$. Here $\lambda = \int_{(0,\infty)} L(dy)$, the total Lévy measure, and the pgf

$$Q(s) = \lambda^{-1}\sum_{k=1}^{\infty} s^k L(\{k\}).$$

Note that the term e^{as} corresponds to the mgf of the constant a. Hence, for the case that the support is $\Re_+$, the lower bound is $a \geq 0$. Also for this case, there is a nice stochastic explanation. Ignoring a and considering $\lambda = \int_{(0,\infty)} L(dy) < \infty$, we know that $L(dy)/\lambda$ is a probability measure on $(0,\infty)$. Assume $Y_0 = 0, Y_j, j = 1, 2, \ldots,$ be iid rv's with probability measure $L(dy)/\lambda$, and $Z \sim \text{Poisson}(\lambda)$. Define the compound Poisson $X = \sum_{j=0}^{Z} Y_j$, namely Poisson compound with a distribution with support on $\Re_+$. Then the mgf of X is to be

$$\begin{aligned} M_X(s) &= \mathbf{E}\left(e^{sX}\right) = \mathbf{E}\left(e^{s\sum_{j=0}^{Z} Y_j}\right) = \mathbf{E}\left(\mathbf{E}\left(e^{s\sum_{j=0}^{Z} Y_j} \mid Z\right)\right) \\ &= \mathbf{E}\left(\left[\int_{(0,\infty)}\lambda^{-1}e^{sy}L(dy)\right]^Z\right) = \exp\left\{\lambda\left[\int_{(0,\infty)}\lambda^{-1}e^{sy}L(dy) - 1\right]\right\} \end{aligned}$$

$$
\begin{aligned}
&= \exp\left\{\lambda\left[\int_{(0,\infty)} \lambda^{-1} e^{sy} L(dy) - \int_{(0,\infty)} \lambda^{-1} L(dy)\right]\right\} \\
&= \exp\left\{\int_{(0,\infty)} (e^{sy}-1)\, L(dy)\right\}.
\end{aligned}
$$

This shows that X is distributed in the ID distribution represented in the second case of this proposition. When $\lambda = \infty$, the explanation is a little bit complicated. Interested readers can refer to Bondesson [1992], p. 16.

For support being $\mathcal{N}_0$, the ID distribution is compound Poisson too, i.e., Poisson compound with another discrete distribution with support on $\mathcal{N}$ (excluding 0). A nice proof for this case can be found in Feller [1966a], p. 271-272.

Non-negative ID rv's are of particular interest in our research. In practice, we may not have the Lévy representation of its pgf or LT. However, there is an simple verification approach: check the absolute monotonicity of $M'_X(s)/M_X(s)$. See Bondesson [1992], p. 16. This absolute monotonicity is equivalent to the infinite divisibility of a non-negative rv. Note that this verification approach is equivalent to Theorem 2.2.6 given by Joe [1997].

Common examples of infinitely divisible distributions are: Gamma, Negative Binomial, Stable distributions, and so on.

Next we turn to self-decomposable distributions.

Definition 2.3 (Self-decomposability) *Suppose $X \sim F$. If for each c, $0 < c \le 1$, there exists a rv ϵ_c such that*

$$X \overset{d}{=} cX + \epsilon_c,$$

where ϵ_c is independent of X, then the probability distribution F is said to be self-decomposable (SD).

An equivalent definition is that $X(s)$ is SD iff for each c, $0 < c \le 1$, $\phi_X(s)/\phi_X(cs)$ [or $\varphi_X(s)/\varphi_X(cs)$] is the LT [or cf] of a probability distribution. In probability, this class is sometimes called the $L-$class. For example, Gamma and Stable distributions are SD. The property of self-decomposability can be applied to construct the stationary discrete-time or continuous-time first order autoregressive process by setting

$$X(n+1) = cX(n) + \epsilon_n, \quad n \in \mathcal{N}_0,\ X(n),\ \epsilon_n \text{ are independent},$$

(refer to Vervaat [1979]) and

$$X(t+h) \stackrel{d}{=} e^{-\rho h} X(t) + \epsilon(h), \qquad h > 0,\ t \in \Re_+,\ X(t),\ \epsilon(h) \text{ are independent.}$$

The latter corresponds to Ornstein-Uhlenbeck-type processes; see Section 1.2. However, the support of marginal distributions of these processes can not be $\mathcal{N}_0$. For this reason, the concept of SD is generalized to discrete distributions with support $\mathcal{N}_0$ by replacing the constant multiplier with binomial thinning.

Definition 2.4 (Discrete Self-decomposability) *Suppose $X \sim F$. If for each c, $0 \le c \le 1$, there exists a rv ϵ_c such that*

$$X \stackrel{d}{=} c * X + \epsilon_c = \sum_{i=0}^{X} I_i + \epsilon_c, \quad I_0 = 0,\ I_1, I_2, \ldots \stackrel{i.i.d.}{\sim} \text{Bernoulli}(c),$$

where ϵ_c is independent of X, then the probability distribution F is said to be discrete self-decomposable (DSD).

This is credited to Steutel and van Harn [1979]. In the sense of pgf, this definition is equivalent to that $G_X(s)/G_X(cs+1-c)$ is a pgf for each $0 < c < 1$. Similar to SD, the property of DSD leads to applications in construction of stationary discrete-time or continuous-time first order autoregressive processes with non-negative integer-valued margins in the literature.

We will show examples of DSD distributions; these are analogues of continuous SD distributions and are given in the end of this subsection.

Now we consider the Stable distributions.

Definition 2.5 (Stability) *Suppose $X \sim F$. If for each $n \ge 1$, there exists constants b_n and c_n such that X can be decomposed as*

$$X \stackrel{d}{=} b_n + c_n(X_{n1} + X_{n2} + \ldots + X_{nn}) = \sum_{i=1}^{n} \left(\frac{b_n}{n} + c_n X_{ni} \right), \qquad \text{where } X_{n1}, X_{n2}, \ldots, X_{nn} \stackrel{i.i.d.}{\sim} F,$$

then the probability distribution F is said to be Stable. If $b_n = 0$ for all n, then F is said to be strictly Stable.

This class has the mgf of form

$$M_X(s) = \exp\left\{-\int_{(0,1]}(-s)^\alpha K(d\alpha)\right\}, \quad K \text{ is a non-negative measure.}$$

See Urbanik [1972].

Finally, we discuss the relationship among ID, SD and Stable. Obviously, Stable is a subset of ID by their definitions. As to others, however, it's not clear by their definitions. Further research has shown that

$$\text{Stable} \subset \text{SD} \subset \text{ID}.$$

The converses are not true. For example, the Gamma distribution is ID and SD, but not Stable. The proof of SD $\subset$ ID can be found in Feller [1966b], p. 553-555, and a brief explanation of Stable $\subset$ SD can be seen in Bondesson [1992], p. 19. As for the discrete self-decomposability, Steutel and van Harn [1979] (Theorem 2.2) proved that a DSD distribution is ID.

For a continuous distribution with positive support, it is of interest to find and study its discrete analogue, because they may share some common features in analysis. Essentially, the discrete analogue is defined in such a way:

Definition 2.6 (Discrete analogue) *Assume the LT of a continuous distribution with positive support is $\phi(s)$. Then, its discrete analogue with non-negative integer support is defined to have pgf of form $G(s) = \phi(1-s)$.*

This definition sometimes can be modified to be $G(s) = \phi(d(1-s))$ $(d > 0)$ to enlarge the family of discrete analogue (see Example 2.5).

Common examples are: Poisson is the discrete analogue of a degenerate rv on a positive point; Negative binomial is the discrete analogue of Gamma; in particular, Geometric is the discrete analogue of Exponential.

If $\phi(s)$ is a LT, then $G(s) = \phi(1-s)$ is always a pgf. This follows from Theorem 2.2.1 by checking the $[0,1]$ domain of s, $G(1) = 1$ and AM feature [follows from CM of ϕ]. Thus, for any positive continuous distribution, we can always obtain its analogue by defining its pgf in terms of the LT. One may wonder what's the explanation of the discrete analogue. Suppose Λ is a positive rv with the LT $\phi(s)$. Given $\Lambda = \lambda$, $Y \sim$ Poisson(λ). Hence Y is a Poisson mixture, and is a non-negative integer rv. The pgf of Y is then

$$G(s) = \mathbf{E}\left(s^Y\right) = \mathbf{E}\left(\mathbf{E}\left(s^Y|\Lambda = \lambda\right)\right) = \mathbf{E}\left(e^{\Lambda(s-1)}\right) = \phi(1-s), \qquad 0 \leq s \leq 1.$$

This means that the discrete analogue is the Poisson mixture and the positive continuous distribution is just the mixing distribution. Therefore, by Poisson mixing, there is one-to-one mapping between the class of positive continuous distributions and the class of discrete Poisson mixtures. For a discrete Poisson mixture distribution, we call the corresponding positive mixing distribution as the **continuous analogue** of that discrete distribution, and by algebra, it has the LT in terms of the pgf: $\phi(s) = G(1-s)$.

Note that in general, we can't define a LT by an arbitrary pgf in this way. The big problem is whether $G(s)$ can be extended from domain $[0, 1]$ to $(-\infty, 1]$. It is sure to work for a Poisson mixture, but not certain for a non-Poisson mixture.

We end this subsection with two examples of continuous SD distributions and their discrete analogues, DSD distributions.

Example 2.4 (Positive stable distribution and discrete stable distribution) *The positive stable distribution has LT*

$$\phi(s) = \exp\left\{-\lambda s^{\gamma}\right\}, \qquad \lambda > 0,\ 0 < \gamma < 1.$$

The discrete stable distribution was introduced by Steutel and van Harn [1979] to have pgf

$$G(s) = \phi(1-s) = \exp\left\{-\lambda(1-s)^{\gamma}\right\}, \qquad \lambda > 0,\ 0 < \gamma < 1.$$

The first one is SD, while the latter is DSD (refer to Steutel and van Harn [1979]).

Example 2.5 (Mittag-Leffler distribution and discrete Mittag-Leffler distribution) *Refer to Bondesson [1992], p. 15. Assume $X = Y^{1/\gamma} \cdot Z$, where $Y \sim \mathrm{Gamma}(\beta, 1)$ and Z is distributed* *in positive stable with LT* e^{-s^γ}, $0<\gamma<1$. *Extending to* $\gamma=1$ *so that* $0<\gamma\le 1$, *we will have* $Z=1$ *as a special case at the upper bound of* γ. *Then the LT of* X *is*

$$\phi_X(s) = \mathbf{E}\left(e^{-sY^{1/\gamma}\cdot Z}\right) = \mathbf{E}\left(\mathbf{E}\left(e^{-sY^{1/\gamma}\cdot Z}\middle|Z\right)\right) = \mathbf{E}\left(e^{-s^\gamma Y}\right) = \frac{1}{(1+s^\gamma)^\beta}.$$

This LT family is labeled as LTE in Joe [1997], p. 376 where it is a special case of Theorem 2.2.7. Taking $\beta=1$, *we obtain*

$$\phi(s) = \frac{1}{1+s^\gamma}, \qquad 0<\gamma\le 1,$$

which is the LT of Mittag-Leffler distribution named by Pillai [1990], because the corresponding cdf is linked to the Mittag-Leffler function. When $\gamma=1$, *it is exponential. Hence, the Mittag-Leffler distribution can be viewed as a generalization of the exponential distribution.*

It seems that the Mittag-Leffler distribution is unlikely to be a stable distribution since one special case is the exponential distribution, which is in the Gamma family, and the Gamma family is not stable.

The discrete Mittag-Leffler distribution was introduced by Pillai and Jayakumar [1995], and has pgf of form

$$G(s) = \frac{1}{1+d(1-s)^\gamma}, \qquad d>0,\ 0<\gamma\le 1.$$

Pillai and Jayakumar [1995] also gave an explanation for this distribution. Consider an infinite sequence of Bernoulli trials where the $k-th$ *trial has success probability* γ/k, $0<\gamma\le 1$, $k=1,2,3,\ldots$. *Denote* Y *as the trial number in which the first success happens. Then the pmf and pgf of* Y *are*

$$\begin{aligned} p_k &= \Pr[Y=k] = (1-\gamma)\left(1-\frac{\gamma}{2}\right)\cdots\left(1-\frac{\gamma}{k-1}\right)\frac{\gamma}{k} \\ &= (-1)^{k-1}\frac{\gamma(\gamma-1)\cdots(\gamma-k+1)}{k!}, \qquad\qquad k=1,2,3,\ldots, \\ G_Y(s) &= 1-(1-s)^\gamma. \end{aligned}$$

Hence Y *has a power series distribution with lower support point* 1. *Let* Z *be from Geometric with pgf* $G_Z(s) = \frac{1-1/(1+d^{-1})}{1-s/(1+d^{-1})} = \frac{1}{1-ds}$, *and* $X=\sum_{i=0}^{Z} Y_i$, *where* $Y_0=0$, Y_i $(i\ge 1)$ *iid from the power series distribution. Then* X *has pgf of form*

$$G_X(s) = \mathbf{E}\left(s^{\sum_{i=0}^Z Y_i}\right) = \mathbf{E}\left(\mathbf{E}\left(s^{\sum_{i=0}^Z Y_i}\middle|Z\right)\right) = \mathbf{E}\left([1-(1-s)^\gamma]^Z\right) = \frac{1}{1+d(1-s)^\gamma}.$$

Similarly, the discrete Mittag-Leffler distribution can be seen as a generalization of Geometric distribution, because it becomes Geometric distribution when $\gamma = 1$.

The Mittag-Leffler distribution is SD, and the discrete Mittag-Leffler distribution is DSD. See Section 7.1.

2.3.2 Tweedie exponential dispersion family

The Tweedie exponential dispersion family is a major member in the class of exponential dispersion models, which has been systematically studied by Prof. Bent Jørgensen. Important references are Jørgensen [1986, 1987, 1992, 1997]. The following is extracted from Jørgensen [1997] and Song [1996].

This section is referred to in a few places in subsequent chapters. It can be skimmed in the first reading.

Suppose $X \sim ED^*(\theta, \lambda)$, the exponential dispersion distribution with probability density (mass) function proportional to

$$c(x; \lambda) \exp\{\theta x - \lambda \kappa(\theta)\}, \qquad x \in \Re,$$

where $c(x; \lambda)$ is a density with respect to a suitable measure (typically Lebesgue measure or counting measure), and

$$\kappa(\theta) = \log\left(\int e^{\theta x} c(x; \lambda) dx\right),$$

the cumulant generator. Hence, a suitable measure $v(dx)$ is required so that

$$\int c(x; \lambda) e^{\theta x} v(dx) = e^{\lambda \kappa(\theta)}.$$

This kind of distribution, $ED^*(\theta, \lambda)$, is called the **additive exponential dispersion model** with the **canonical parameter** θ and the **index parameter** λ. Let $\Theta = \{\theta \in \Re : \kappa(\theta) < \infty\}$ be the **canonical parameter domain**, intΘ be the interior of Θ. Denote the **mean value mapping** $\tau : \text{int}\Theta \to \Re$, and the **mean domain** defined by

$$\tau(\theta) = \kappa'(\theta) \quad \text{and} \quad \Omega = \tau(\text{int}\Theta)$$

respectively. Define the **unit variance function** $V : \Omega \to \Re_+$ as $V(\mu) = \tau'\left(\tau^{-1}(\mu)\right)$. By the property of exponential family, the cumulant generating function is

$$\begin{aligned} K^*(s;\theta,\lambda) &= \log \mathbf{E}[e^{sX}] = \log\left(e^{-\lambda\kappa(\theta)}\int e^{sx}c(x;\lambda)e^{\theta x}v(dx)\right) \\ &= \log\left(e^{-\lambda\kappa(\theta)}e^{\lambda\kappa(\theta+s)}\right) = \lambda[\kappa(\theta+s)-\kappa(\theta)], \quad s\in\Theta-\theta. \end{aligned}$$

Note that s takes value $\theta^*-\theta$ to guarantee that $\kappa(\theta+s)=\kappa(\theta^*)<\infty$, where $\theta^*\in\Theta$. Differentiating $K^*(s;\theta,\lambda)$ twice with respect to s and setting $s=0$, we find the mean and variance to be

$$\mathbf{E}(X) = \lambda\tau(\theta) \quad \text{and} \quad \mathbf{Var}(X) = \lambda V(\tau(\theta)) = \lambda\tau'(\theta).$$

Let $Y = X/\lambda,\ \mu=\tau(\theta), \sigma^2 = 1/\lambda$. Then by definition, $Y\sim ED(\mu,\sigma^2)$, with probability density (mass) function

$$\bar{c}(y;\lambda)\exp\{\lambda[\theta y-\kappa(\theta)]\}, \qquad y\in\Re,$$

where $\bar{c}(y;\lambda)$ is a density with respect to a fixed measure, and the cumulant generator is

$$\kappa(\theta) = \log\left(\int e^{\theta y}\bar{c}(y;\lambda)dy\right),$$

the cumulant generating function is

$$K(s;\theta,\lambda) = \log\mathbf{E}[e^{sY}] = \lambda[\kappa(\theta+s/\lambda)-\kappa(\theta)], \quad s\in\Theta-\theta,$$

and the mean and variance are

$$\mathbf{E}(Y) = \mathbf{E}(X)/\lambda = \mu(=\tau(\theta)) \quad \text{and} \quad \mathbf{Var}(Y) = \mathbf{Var}(X)/\lambda^2 = \sigma^2V(\mu)(=\sigma^2\tau'(\theta)).$$

This kind of distribution, $ED(\mu,\sigma^2)$, is called the **reproductive exponential dispersion model.** Here $\sigma^2 = 1/\lambda$ is called the **dispersion parameter**.

Concrete examples in exponential dispersion models include the Binomial, Negative binomial, Poisson, Gamma, Normal, hyperbolic secant and generalized hyperbolic secant distributions.

The Tweedie family is a special member in the class of exponential dispersion models; it was first studied by Tweedie [1947]. Following the reproductive form, it has special form of unit variance function:

$$V(\mu) = \mu^d, \qquad \mu\in\Omega, \quad d\in\Re.$$

Hence, the ratio of variance and mean is

$$\frac{\mathbf{Var}(Y)}{\mathbf{E}(Y)} = \mu^{d-1}.$$

A model with this reproductive form is denoted as $\mathrm{Tw}_d(\mu, \sigma^2)$. Tweedie models are closed with respect to scale transformation, i.e., if $Y \sim \mathrm{Tw}_d(\mu, \sigma^2)$, then $cY \sim \mathrm{Tw}_d(c\mu, c^{2-d}\sigma^2)$.

Since $V(\mu) = \mu^d = \tau^d(\theta) = \tau'(\theta)$, $\tau(\theta)$ must be

$$\tau(\theta) = \begin{cases} [(1-d)\theta]^{1/(1-d)}, & d \neq 1, \\ e^{\theta}, & d = 1. \end{cases}$$

For the sake of convenience, let $\beta = \frac{d-2}{d-1}$. Then $1/(1-d) = \beta - 1,\ 1 - d = 1/(\beta - 1)$. This leads to

$$\tau(\theta) = \begin{cases} \left(\frac{\theta}{\beta-1}\right)^{\beta-1}, & d \neq 1, \text{ or } \beta \neq \infty \\ e^{\theta}, & d = 1, \text{ or } \beta = \infty. \end{cases}$$

Since $\tau(\theta) = \kappa'(\theta)$, for $\mathrm{Tw}_d(\mu, \sigma^2)$, the cumulant generator $\kappa(\theta)$ has explicit form

$$\kappa(\theta) = \begin{cases} \frac{\beta-1}{\beta}\left(\frac{\theta}{\beta-1}\right)^{\beta}, & d \neq 1,\ 2, \text{ or } \beta \neq 0,\ \infty, \\ -\log(-\theta), & d = 2, \text{ or } \beta = 0, \\ e^{\theta}, & d = 1, \text{ or } \beta = \infty. \end{cases}$$

(ignoring the arbitrary constants in the integrations will not affect the final results of cumulant generating function.) Thus, one of the other advantages of Tweedie model we are appreciating is that it has explicit expression for the cumulant generating function and mgf. The cumulant generating function of $\mathrm{Tw}_d(\mu, \sigma^2)$ is

$$K(s;\theta,\lambda) = \begin{cases} \lambda\frac{\beta-1}{\beta}\left(\frac{\theta}{\beta-1}\right)^{\beta}\left[\left(1+\frac{s}{\theta\lambda}\right)^{\beta} - 1\right], & d \neq 1, 2; \\ -\lambda\log\left(1+\frac{s}{\theta\lambda}\right), & d = 2; \\ \lambda e^{\theta}\left[e^{s/\lambda} - 1\right], & d = 1, \end{cases}$$

and the mgf of this family has special exponential form:

$$M(s;\theta,\lambda) = \mathbf{E}\left[e^{sX}\right] = \begin{cases} \exp\left\{\lambda\frac{\beta-1}{\beta}\left(\frac{\theta}{\beta-1}\right)^{\beta}\left[\left(1+\frac{s}{\theta\lambda}\right)^{\beta} - 1\right]\right\}, & d \neq 1, 2; \\ \left(1+\frac{s}{\theta\lambda}\right)^{-\lambda}, & d = 2; \\ \exp\left\{\lambda e^{\theta}\left[e^{s/\lambda} - 1\right]\right\}, & d = 1. \end{cases}$$

Further research shows that $d \in (-\infty, 0] \cup [1, \infty]$. Thus, if $d \in (-\infty, 0]$, then $\beta \in (1, 2]$; if $d \in [1, \infty]$, then $\beta \in [-\infty, 1]$. This leads to $\beta \in (1, 2] \cup [-\infty, 1] = [-\infty, 2]$. Now for the future use, we impose the subscript β on $\tau(\theta), \kappa(\theta), K(s; \theta, \lambda), M(s; \theta, \lambda), c(x; \lambda)$ and $\bar{c}(y; \lambda)$ to indicate that they are linked to the specific parameter β.

Also, the probability density (mass) function of Tweedie model can be obtained, though it is complicated. Recall $X = \lambda Y$ has the additive model if Y has the reproductive model. The probability density (mass) function is

$$f_X(x; \theta, \lambda, \beta) = c_\beta(x; \lambda) \exp\{\theta x - \lambda \kappa_\beta(\theta)\},$$

where

$$c_\beta(x; \lambda) = \begin{cases} 1, & \beta < 0, x = 0; \\ \sum_{k=1}^{\infty} \frac{\lambda^k \kappa_\beta^k(-1/x)}{x\Gamma(-k\beta)k!}, & \beta < 0, x > 0; \\ \sum_{k=1}^{\infty} \frac{\Gamma(1+k\beta)}{\pi x k!} \lambda^k \kappa_\beta^k(-1/x) \sin(-k\pi\beta), & 0 < \beta < 1, x > 0; \\ \sum_{k=1}^{\infty} \frac{\Gamma(1+k/\beta)}{\pi x k!} \left(\frac{-x}{\lambda \kappa_\beta^{1/\beta}(1)} \right)^k \sin(-k\pi\beta), & 1 < \beta < 2, x \in \Re. \end{cases}$$

Here $\kappa_\beta(\theta)$ is just the previously calculated $\kappa(\theta)$ for $\mathrm{Tw}_d(\mu, \sigma^2)$. The corresponding reproductive model then has has probability density (mass) function

$$f_Y(y; \theta, \lambda, \beta) = \lambda c_\beta(\lambda y; \lambda) \exp\{\lambda[\theta y - \kappa_\beta(\theta)]\}.$$

Tweedie family includes distributions with support on $\Re, \Re_+, \mathcal{N}_0$, corresponding to real-valued, positive-valued and non-negative integer-valued random variables. These are related to different ranges of d, in which we view the endpoints of ranges as boundaries. Table 2.1 summarizes the different types of Tweedie models. From the table, we know that the real support $\Re$ appears when $d \leq 0$ and $d = \infty$; this corresponds to $1 \leq \beta \leq 2$, while the non-negative support $\mathcal{N}_0$, or $\Re_0$, or $\Re_+$ appears when $1 \leq d \leq \infty$; it corresponds to $\beta \leq 1$. Some boundary cases are well known distributions: normal ($d = 0$), Poisson ($d = 1$), Gamma ($d = 2$), inverse Gaussian ($d = 3$) and extreme stable ($d = \infty$). When $1 < d < 2$ (corresponding to $\beta < 0$), the compound Poisson with Gamma obtains, that is,

$$Y = \sum_{i=0}^{N} Z_i,$$

Table 2.1: *Summary of Tweedie exponential dispersion models (S = support set).*

Distributions	d	S	Ω	Θ
Extreme stable	$d < 0$	$\Re$	$\Re_+$	$\Re_0$
Normal	$d = 0$	$\Re$	$\Re$	$\Re$
(Do not exist)	$0 < d < 1$	—	—	—
Poisson	$d = 1$	$\mathcal{N}_0$	$\Re_+$	$\Re$
Compound Poisson	$1 < d < 2$	$\Re_0$	$\Re_+$	$\Re_-$
Gamma	$d = 2$	$\Re_+$	$\Re_+$	$\Re_-$
Positive stable	$2 < d < 3$	$\Re_+$	$\Re_+$	$-\Re_0$
Inverse Gaussian	$d = 3$	$\Re_+$	$\Re_+$	$-\Re_0$
Positive stable	$d > 3$	$\Re_+$	$\Re_+$	$-\Re_0$
Extreme stable	$d = \infty$	$\Re$	$\Re$	$\Re_-$

where $Z_0 = 0, Z_i \overset{iid}{\sim} \text{Gamma}(\theta, -\beta)$ and $N \sim \text{Poisson}(\lambda\kappa_d(\theta))$. This distribution has a positive probability on zero,

$$\Pr[Y = 0] = \Pr[N = 0] = \exp\{-\lambda\kappa_d(\theta)\},$$

and density function

$$f_Y(y; \theta, \lambda, \beta) = \frac{1}{y}\sum_{i=1}^{\infty} \frac{\lambda^i \kappa_d^i(-1/y)}{i!\Gamma(-i\beta)} \exp\{\theta y - \lambda\kappa_d(\theta)\}.$$

2.3.3 Generalized convolutions

The generalized convolution is a natural extension of a finite convolution. It helps to connect those individual distributions which seem to have quite different forms in the pdf or cdf. Fortunately, we find that the generalized convolution provides a huge ammunition for the theory of continuous-time generalized AR(1) process. A good reference on the generalized convolution is Bondesson [1992]. The following materials regarding GGC, EGGC, GCMED and GNBC are extracted from that book.

First, we review the generalized Gamma convolution, which was introduced by Thorin [1977a, 1977b], to understand the mechanism of construction of generalized convolution.

Since the LT of Gamma(u_i, β_i) is $\left(\frac{u_i}{u_i+s}\right)^{\beta_i}$ $(i = 1, \ldots, n)$, the LT of the sum of n such Gamma rv's, i.e., finite convolution, is

$$\phi_n(s) = \prod_{i=1}^{n} \left(\frac{u_i}{u_i + s}\right)^{\beta_i} = \exp\left\{\sum_{i=1}^{n} \beta_i \log\left(\frac{u_i}{u_i + s}\right)\right\}.$$

Consider pointwise limits of $\phi_n(s)$ and permit a Gamma distribution to be degenerate at a point $a \geq 0$ with LT e^{-as}. This leads to the following definition.

Definition 2.7 *A generalized Gamma convolution (GGC) is defined as a distribution with support on $[0, \infty)$ and LT of the form*

$$\phi(s) = \exp\left\{-as + \int_{(0,\infty)} \log\left(\frac{u}{u+s}\right) U(du)\right\},$$

where $a \geq 0$ and $U(du)$ is a non-negative measure on $(0, \infty)$ satisfying

$$\int_{(0,1]} |\log u| U(du) < \infty \quad \textit{and} \quad \int_{(1,\infty)} u^{-1} U(du) < \infty.$$

Therefore, the GGC is the limiting distribution for a sequence of sums of independent Gamma variables with possibly different rate parameters. Extending this idea to random variables from other families, we obtain the concept of the **generalized convolution**, which is defined as the limit distribution for a sequence of sums of independent variables from a parametric family. In the sense of limit, we know that there are usually numerous rv's involved in the convolution. In our study, we tentatively call the distribution involved in the sum as the **base distribution** of the generalized convolution.

For a rv X distributed in the GGC class, we have

$$\mathbf{E}[X] = -\phi'(0) = a + \int_{(0,\infty)} u^{-1} U(du), \quad \mathbf{Var}[X] = \phi''(0) - (\phi'(0))^2 = \int_{(0,\infty)} u^{-2} U(du).$$

This class is surprisingly rich. Some of the examples include

- (strictly) positive Stable distribution on $(0, \infty)$: the LT $\phi(s) = \exp\{-s^\gamma\}$ $(0 < \gamma < 1)$, and the measure $U(du) = \frac{\gamma \sin(\gamma\pi)}{\pi} u^{\gamma-1} du$ which leads to $\int_{(0,\infty)} U(du) = \infty$.

- Pareto distribution: the pdf $f(x) = \gamma\lambda^\gamma(x+\lambda)^{-\gamma-1}$ $(x,\gamma,\lambda > 0)$, and the density of the measure U is $U'(u) = \frac{\gamma}{\Gamma(\gamma)}\lambda^\gamma u^{\gamma-1}e^{-\lambda u}$.

- Generalized inverse Gaussian distribution: the pdf

$$f(x) = Cx^{\beta-1}\exp\{-c_1x - c_2x^{-1}\}, \quad x > 0 \ (c_1, c_2 > 0, \beta \in \Re).$$

The U-measure has density

$$U'(u) = \begin{cases} 0, & u < c_1, \\ \Big[(u-c_1)^{\beta+1}\int_0^\infty\int_0^\infty(\lambda-c_1)^{-\beta-1}\rho^{-1}e^{-1/\rho} & \\ \qquad \times\exp\left\{-c_2(s-1)^2\rho(\lambda-c_1)^{-1}\right\}d\rho d\lambda\Big]^{-1}, & u > c_1. \end{cases}$$

Letting $c_1 \to 0$ leads to inverse Gamma distribution. A good reference on the inverse Gaussian distribution is Seshadri [1999], where applications can be found in reliability, survival analysis and actuarial science.

- Generalized Gamma distribution (power of Gamma random variable): the pdf is

$$f(x) = Cx^{\beta-1}\exp\left\{-x^\alpha\right\}, \quad x > 0 \quad (0 < \alpha < 1, \beta > 0).$$

The density of U-measure is

$$U(u) = \pi^{-1}\arg\left[Cu^{-\beta}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\Gamma(\beta+k\alpha)\exp\{i\pi(\beta+k\alpha)\}u^{-k\alpha}\right].$$

- Beta distribution of the second kind (Ratio of Gamma variables): the pdf is

$$f(x) = Cx^{\beta-1}(1+cx)^{-\gamma}, \quad x > 0 \quad (\gamma > \beta > 0, c > 0).$$

However, the density of the U-measure does not have a simple expression.

- Lognormal distributions: the pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}}\sigma^{-1}x^{-1}\exp\left\{-\frac{(\log x)^2}{2\sigma^2}\right\}, \quad x > 0.$$

Unfortunately, no simple expression for the U-measure exists.

Next we visit the extended generalized Gamma convolutions (EGGC), which have support on the whole real line $\Re$. This class is needed to cover the limit distributions for sums of independent positive and negative Gamma variables. It was also introduced by O. Thorin. See Thorin (1978).

Definition 2.8 *An extended generalized Gamma convolution (EGGC) is defined as a distribution with support on $\Re$ and cf of the form*

$$\varphi(s) = \exp\left\{ ibs - \frac{cs^2}{2} + \int_{(-\infty,\infty)} \left(\log\left(\frac{u}{u-is}\right) - \frac{isu}{1+u^2} \right) U(du) \right\},$$

where $b \in \Re$, $c \geq 0$ and $U(du)$ is a non-negative measure on $\Re \setminus \{0\}$ satisfying

$$\int_{\Re\setminus\{0\}} \frac{1}{1+u^2} U(du) < \infty \quad \textit{and} \quad \int_{|u|\leq 1} |\log u^2| U(du) < \infty.$$

Remark: The term $su/(1+u^2)$ is added to guarantee the convergence of the integral. When $\int_{|u|>1} |u|^{-1} U(du) < \infty$, it can be omitted. Hence, GGC is a subclass of EGGC. Further research shows that the symmetric EGGC is the variance mixture of the Normal distribution, with cf of form

$$\varphi(s) = \exp\left\{ -\frac{cs^2}{2} + \int_{(0,\infty)} \log\left(\frac{u^2}{u^2+s^2}\right) U(du) \right\},$$

where U is symmetric on $\Re \setminus \{0\}$. Thorin [1978] proved the EGGC is SD.

Some examples of the EGGC class are listed below. They are verified by characterizations other than specifying U-measure.

- Stable distribution: the cf of the general Stable distribution of index α $(0 < \alpha \leq 2)$ is

  $$\varphi(s) = \exp\{i\mu s - C|s|^\alpha (1 - i\beta \text{sign}(s)\omega(s,\alpha))\}, \quad \mu \in \Re,\ C \geq 0,\ |\beta| \leq 1,$$

  where

  $$\omega(s,\alpha) = \begin{cases} \tan(\alpha\pi/2), & \alpha \neq 1, \\ -2\pi^{-1}\log|s|, & \alpha = 1. \end{cases}$$

  The Cauchy distribution is within the Stable distribution family with cf

  $$\varphi(s) = \exp\{-C|s|\}, \quad C \geq 0.$$

- Generalized Logistic distribution: This rv is derived from two Exponential rv's as $X = \log(Y_1/Y_2)$, where $Y_i \sim \text{Gamma}(\beta_i, 1)$ $(i = 1, 2)$. The pdf of X is

$$f(x) = \frac{1}{\mathrm{B}(\beta_1, \beta_2)} e^{-\beta_1 x}/(1 + e^{-x})^{\beta_1+\beta_2}, \quad x \in \Re.$$

When $\beta_1 = \beta_2 = 1/2$, it is

$$f(x) = \frac{1}{\pi}(e^{x/2} + e^{-x/2})^{-1}, \quad x \in \Re.$$

While $\beta_1 = \beta_2 = 1$, it is the logistic distribution with pdf

$$f(x) = e^{-x}/(1 + e^{-x})^2, \quad x \in \Re;$$

and mgf

$$M(s) = \Gamma(1 + s)\Gamma(1 - s) = \frac{\pi s}{\sin(\pi s)} = \prod_{k=1}^{\infty} \frac{1}{(1 - s^2/k^2)} = \prod_{k=1}^{\infty} \frac{1}{(1 - s/k)(1 + s/k)}.$$

A stochastic representation of scale mixture of the Normal distribution for this logistic rv X can be found in Joe [1997], p. 133-134. See also Andrews and Mallows [1974] and Stefanski [1991]. It is

$$X \stackrel{d}{=} Z/V,$$

where $Z \sim \mathrm{N}(0, 1)$, and V is a positive rv with pdf

$$f_V(x) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} k^2 x^{-3} \exp\{-k^2/(2x^2)\}.$$

Hence, X is the rv of a variance mixture of the Normal distribution.

- Logarithm of Gamma variable: $X = \log Y$, $Y \sim \text{Gamma}(\beta, 1)$.

- Other symmetric EGGC distributions with pdf:

  (i) $f(x) = C(1 + cx^2)^{-\gamma}$, t-distribution, essentially

  (ii) $f(x) = C(1 + c|x|)^{-\gamma}$, two-sided Pareto distribution

  (iii) $f(x) = C \exp\{-c\sqrt{x^2 + \delta}\}$; Hyperbolic distribution

(iv) $f(x) = C\exp\{-c|x|^{2/k}\} \quad (k = 1, 2, \ldots)$

The generalized convolutions of mixtures of Exponential distributions (GCMED) is another extension of GGC (see Bondesson [1992], p139-140). This class has support on $[0, \infty)$.

Definition 2.9 *A generalized convolution of mixtures of Exponential distribution (GCMED) is defined as a distribution with support on* $[0, \infty)$ *and LT of the form*

$$\phi(s) = \exp\left\{-as + \int_{(0,\infty)} \left(\frac{1}{u+s} - \frac{1}{u}\right) U(du)\right\} = \exp\left\{-as + \int_{(0,\infty)} \frac{-s}{u(u+s)} U(du)\right\},$$

where $a \geq 0$ *and* $U(du)$ *is a non-negative measure on* $(0, \infty)$ *satisfying*

$$\int_{(0,\infty)} \frac{1}{u(1+u)} U(du) < \infty.$$

Remark: A mixture of Exponential distributions (MED) is defined as a probability distribution on $[0, \infty)$ with pdf

$$f(x) = \int_{(0,\infty)} ue^{-xu} U(du),$$

or cdf

$$F(x) = \int_{(0,\infty]} (1 - e^{-xu}) U(du).$$

Here $U(du)$ is the mixing measure (for the inverse of the scale parameter), which is non-negative and satisfies $\int_{(0,\infty]} U(du) = 1$. $U(\{\infty\}) > 0$ implies the distribution has an atom at 0. The LT of an MED is

$$\phi(s) = U(\{\infty\}) + \int_{(0,\infty)} \frac{u}{u+s} U(du).$$

The LT of compound Poisson with Exponential has the form of

$$\phi(s) = \exp\left\{\lambda\left(\frac{u}{u+s} - 1\right)\right\} = \exp\left\{\lambda u\left(\frac{1}{u+s} - \frac{1}{u}\right)\right\}.$$

Hence, the GCMED is a generalized compound Poisson-Exponential convolution following the convention of GGC.

Some examples of GCMED are

- Compound Poisson with Exponential distribution: Let $X = \sum_{i=0}^{N} Y_i$, where $N \sim \text{Poisson}(\lambda)$ and $Y_0 = 0, Y_i \sim \text{Exponential}(u)$ $(i = 1, 2, \ldots)$. Then

$$\phi_X(s) = \exp\{\lambda(\phi_{Y_1}(s) - 1)\} = \exp\left\{\lambda u \left(\frac{1}{u+s} - \frac{1}{u}\right)\right\}.$$

- Non-central χ^2-distribution: Let $Z_i \sim \text{N}(\mu_i, 1)$ $(i = 1, 2, \ldots, n)$ be independent. Then

$$X = \sum_{i=1}^{n} Z_i^2 = \sum_{j=0}^{N} Y_j,$$

where $\lambda = \sum_{i=1}^{n} \mu_i^2$, $N \sim \text{Poisson}(\lambda/2)$ and $Y_0 = 0$, $Y_j \sim \text{Exponential}(1/2)$.

- Logarithm of Beta variable: $X = -\log Y$, where $Y \sim \text{Beta}(\alpha, \beta)$. The LT of X is

$$\phi_X(s) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+s)}{\Gamma(\alpha)\Gamma(\alpha+\beta+s)}.$$

- Inverse Gaussian mixture distribution (introduced by Jorgensen, Seshadri & Whitmore [1991]): the pdf is

$$f(x) = C'(p + q\sqrt{c_1/c_2}x)f_1(x), \quad q = 1-p,\ 0 \le p \le 1,$$

where $f_1(x)$ is the pdf of inverse Gaussian distribution

$$f_1(x) = Cx^{-3/2}\exp\{-c_1x - c_2x^{-1}\},$$

which has LT

$$\phi_1(s) = \exp\{c_3(1 - \sqrt{1 + c_4 s})\}, \quad c_3 = 2\sqrt{c_1c_2},\ c_4 = 1/c_1.$$

The LT of the inverse Gaussian mixture distribution is

$$\phi(s) = (p + q/\sqrt{1+c_4s})\phi_1(s).$$

This family includes the well-known life distribution of Birnbaum & Saunders (1969) when $p = 1/2$.

The last generalized convolution discussed in Bondesson [1992] is the generalized Negative Binomial convolution (GNBC), which has support on non-negative integer $\{0, 1, 2, \ldots\}$. The discrete analogue of the Gamma distribution is the Negative Binomial distribution; hence, the GNBC is the discrete analogue of the GGC.

Definition 2.10 *A generalized Negative Binomial convolution (GNBC) is defined as a distribution with support on non-negative integer and pgf of the form*

$$G(s) = \exp\left\{a(s-1) + \int_{(0,1)} \log\left(\frac{p}{1-qs}\right) V(dq)\right\},$$

where $a \geq 0$, $p = 1 - q$ *and* $U(du)$ *is a non-negative measure on* $(0,1)$ *satisfying*

$$\int_{(0,1/2]} qV(dq) < \infty \quad \textit{and} \quad \int_{(1/2,1)} \log(p)V(dq) < \infty.$$

Some examples of the GNBC class are

- Discrete Stable distribution: the pgf is

$$G(s) = \exp\{-c(1-s)^{\alpha}\}, \quad c > 0, \; 0 < \alpha \leq 1.$$

(refer to Steutel and van Harn [1979]).

- Generalized Waring distribution: this family is defined with probability mass

$$p_j = C\frac{\beta^{[j]}\alpha^{[j]}}{j!(\alpha+\beta+\gamma)^{[j]}}, \quad j = 0, 1, 2, \ldots, \quad \alpha, \beta, \gamma > 0, \quad C = \frac{\Gamma(\beta+\gamma)\Gamma(\alpha+\gamma)}{\Gamma(\alpha+\beta+\gamma)\Gamma(\gamma)},$$

where $\beta^{[0]} = 1$, $\beta^{[j]} = \beta \cdot (\beta+1) \cdots (\beta+j-1)$. The pgf is the sum of a Hypergeometric series. Furthermore, the generalized Waring distribution is the Poisson(Λ)-mixture, where $\Lambda = Y \cdot X_1/X_2$ and Y, X_1, X_2 are independent with

$$Y \sim \text{Gamma}(\beta, 1), \quad X_1 \sim \text{Gamma}(\alpha, 1), \quad X_2 \sim \text{Gamma}(\gamma, 1).$$

This family leads to several distributions:

(i) Waring distribution (see Johnson & Kotz [1969], p. 250): $\beta = 1$.

A special case is the power series distribution when $\alpha = 1 - \eta,\ \gamma = \eta\ (0 < \eta < 1)$, which has pmf and pgf:

$$p_j = \eta \frac{(1-\eta)^{[j]}}{(j+1)!}, \quad j = 0, 1, \ldots, \qquad \text{and} \quad G(s) = \frac{1-(1-s)^\eta}{s}.$$

(ii) Yule distribution: $\alpha = \beta = 1$ and $\gamma \to 0$.

(iii) NB(β, q)-distribution: $q = \alpha/(\alpha + \gamma)$ and let $\alpha \to \infty,\ \gamma \to \infty$.

Distributions (i) and (ii) have applications in modelling word size in prose.

- Logarithmic series distribution (shifted): the pmf and pgf are

$$p_i = \frac{c^{i+1}}{\theta(i+1)}\ (i = 0, 1, 2, \ldots), \quad c = 1 - e^\theta,\ \theta > 0, \quad \text{and} \quad G(s) = \frac{\log(1-cs)}{s\log(1-c)}.$$

In the study of continuous-time generalized AR(1) process, we have discovered four new generalized convolutions, which we tentatively name as GC I, GC II, GC III and GC IV. These generalized convolutions play an important role in customizing marginal distributions of a steady state Markov process (see Chapter 6).

Definition 2.11 *A generalized convolution I (GC I) is defined as a distribution with support on non-negative integer and pgf of the form*

$$G(s) = \exp\left\{-as + \int_{(0,1)} \frac{q(s-1)}{1-qs} V(dq)\right\},$$

where $a \geq 0$ and $V(dq)$ is a non-negative measure on $(0,1)$ satisfying

$$\int_{(0,1/2]} qV(dq) < \infty \quad \textit{and} \quad \int_{(1/2,1)} (1-q)^{-1} V(dq) < \infty.$$

The base $\exp\left\{\frac{q(s-1)}{1-qs}\right\}$ was proved to be a pgf in Example 2.1 in Section 2.2.1.

Definition 2.12 *A generalized convolution II (GC II) is defined as a distribution with support on non-negative integer and pgf of the form*

$$G(s) = \exp\left\{-as + \int_{[\gamma,1)} \frac{q(s-1)(1-\gamma s)}{1-qs} V(dq)\right\},$$

where $a \geq 0$, $\gamma > 0$ and $V(dq)$ is a non-negative measure on $[\gamma, 1)$ satisfying

$$\int_{[\gamma,1)} (1-q)^{-1} V(dq) < \infty.$$

The base $\exp\left\{\frac{q(s-1)(1-\gamma s)}{1-qs}\right\}$, $0 < \gamma \leq q < 1$, was also proved to be a pgf in Example 2.1 in Section 2.2.1.

Definition 2.13 *A generalized convolution III (GC III) is defined as a distribution with support on $[0, \infty)$ and LT of the form*

$$\phi(s) = \exp\left\{-as + \int_{(0,\infty)} \frac{1-\gamma}{1-\gamma-\gamma u} \log\left(\frac{(1-\gamma+\gamma s)u}{(1-\gamma)(u+s)}\right) U(du)\right\},$$

where $a \geq 0$, $0 \leq \gamma < 1$ (γ is fixed) and $U(du)$ is a non-negative measure on $(0, \infty)$ satisfying

$$\int_{(0,1])} |\log u| U(du) < \infty \quad \textit{and} \quad \int_{(1,\infty)} u^{-2} U(du) < \infty.$$

Note that when $\gamma = 0$, GC III will become GGC. Hence, GGC is a special case of GC III.

The base $\exp\left\{\log\left(\frac{(1-\gamma+\gamma s)u}{(1-\gamma)(u+s)}\right)\right\}$ was proved to be a LT in (2) of Theorem 2.2.8.

Definition 2.14 *A generalized convolution IV (GC IV) is defined as a distribution with support on $[0, \infty)$ and LT of the form*

$$\phi(s) = \exp\left\{-as + \lambda \int_{(0,\gamma^{-1}-1]} -\frac{s(1-\gamma+\gamma s)}{u+s} U(du)\right\},$$

where $a \geq 0$, $\lambda > 0$, $0 \leq \gamma < 1$ (γ is fixed) and $U(du)$ is a non-negative measure on $(0, \gamma^{-1} - 1]$ satisfying

$$\int_{(0,\gamma^{-1}-1]} u^{-1} U(du) < \infty.$$

Note that when $\gamma = 0$, the LT will be

$$\phi(s) = \exp\left\{-as + \lambda \int_{(0,\infty)} \frac{-s}{u+s} U(du)\right\} = \exp\left\{-as + \int_{(0,\infty)} \frac{-s}{u(u+s)} U'(du)\right\},$$

where $U'(du) = \lambda u U(du)$. Thus, GC IV will become GCMED. This shows that GCMED is a special case of GC IV.

The base $\exp\left\{-\lambda \cdot \frac{s(1-\gamma+\gamma s)}{u+s}\right\}$ was proved to be a LT in (1) of Theorem 2.2.8.

Specific distributions which are in the new generalized convolution families are not known at this moment; thus, further investigations are under study.

2.4 Independent increment processes and examples

The independent increment process is well studied, and is intimately connected with infinitely divisible distributions. The latter links to the study of Lévy process. Refer to Prabhu [1980], p. 69, Feller [1966b], p. 177-179, Bhattacharya & Waymire [1990], p. 349-356, Protter [1990], Section 5 in Chapter 1. We review this process family and will choose some of them to be the noise process or innovation process in the theory of continuous-time generalized AR(1) processes defined in subsequent chapters.

Definition 2.15 Stationary independent increment process (IIP):
A process $\{X(t); t \geq 0\}$ is said to have stationary independent increments if it satisfies the following properties:

(i) *For $0 \leq t_1 < t_2 < \cdots < t_n (n \geq 2)$, the random variables*

$$X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$$

are independent.

(ii) *The distribution of the increment $X(t_k) - X(t_{k-1})$ depends on (t_{k-1}, t_k) only through the difference $t_k - t_{k-1}$.*

Without loss of generality, $X(0)$ is usually taken as 0. This is because that if $X(0) \neq 0$, we can subtract it from the process which results: $Y(t) = X(t) - X(0)$. The new process $\{Y(t); t \geq 0\}$ has stationary independent increments and starts from $Y(0) = 0$. This means that the starting point is independent of any increment for a stationary independent increment process. Since

$$X(t) = \sum_{i=1}^{n} \left[X\left(\frac{i}{n}t\right) - X\left(\frac{i-1}{n}t\right)\right],$$

$X(t)$ can be viewed to be the sum of n independent random variables, all of which have the same distribution as $X(t/n)$. This is true for all $n \geq 1$. Hence, it follows that $X(t)$ has an infinitely divisible distribution.

Note that it's easy to extend the stationary IIP to non-stationary IIP by loosening condition (ii), in which case the distribution of $X(t)$ may not be infinitely divisible if the small increments are not identically distributed.

With the following additional conditions, the stationary independent increment process becomes the **Lévy process**:

(iii) *$X(t)$ is continuous in probability, namely, for any $\epsilon > 0$,*

$$\lim_{t \to 0+} \Pr[|X(t)| > \epsilon] \longrightarrow 0.$$

This is equivalent to stochastic continuity: $\lim_{t_1 \to t_2} \Pr[|X(t_2) - X(t_1)| > \epsilon] = 0.$

(iv) *There exist left and right limits $X(t-)$ and $X(t+)$. Assume that $X(t)$ is right continuous: $X(t+) = X(t)$. Here the difference $X(t+) - X(t-) = X(t) - X(t-)$ is called the jump of the process at time t.*

Note that the number of conditions required for a Lévy process may appear as three to five in the literature, depending on the author's view. For example, some impose $X(0) = 0$, some don't.

Because the increment $X(t_2) - X(t_1)$ is infinitely divisible for the Lévy process, we can characterize this process by the mgf, LT or pgf of the infinitely divisible distribution discussed in Proposition 2.3.1. Some scholars even define the Lévy process in this way, e.g., Bondesson [1992], p. 16. Since we are particularly interested in three kinds of supports of increment: $(-\infty, +\infty)$, $(0, +\infty)$ and $\{0, 1, 2, \ldots\}$, we summarize the results in these three cases in the following proposition.

Proposition 2.4.1 *Suppose $\{X(t); t \geq 0\}$ is a Lévy process with $X(0) = 0$. Consider $X(t)$, the margin at time t.*

- *When the support of $X(t)$ is $\Re$, the mgf of $X(t)$ is*

$$\phi_{X(t)}(s) = \exp\left\{t\left[as + \frac{\sigma^2}{2} \cdot s^2 + \int_{y \neq 0} \left(e^{sy} - 1 - \frac{sy}{1+y^2}\right) L(dy)\right]\right\}, \quad \Re e\ s = 0,\ a \in \Re,$$

where the measure L satisfies $\int_{y\neq 0} \min(1, y^2)L(dy) < \infty$.

- *When the support of $X(t)$ is $\Re_+$, the LT of $X(t)$ is*

$$\phi_{X(t)}(s) = \exp\left\{t\left[-as + \int_{(0,\infty)} \left(e^{-sy} - 1\right) L(dy)\right]\right\}, \quad a \geq 0,$$

where the Lévy measure L is non-negative and satisfies $\int_{(0,\infty)} \min(1, y)L(dy) < \infty$. *Now the parameter a is called the left-extremity.*

- *When the support of $X(t)$ is $\mathcal{N}_0$, the pgf of $X(t)$ is*

$$G_{X(t)}(s) = \exp\left\{t\left[\int_{(0,\infty)} \left(s^y - 1\right) L(dy)\right]\right\} = \exp\left\{t\lambda[Q(s) - 1]\right\},$$

where the Lévy measure L is non-negative and satisfies $\int_{(0,\infty)} \min(1, y)L(dy) < \infty$. *Here* $\lambda = \int_{(0,\infty)} L(dy)$, *the total Lévy measure, and the pgf is*

$$Q(s) = \lambda^{-1} \sum_{k=1}^{\infty} s^k L(\{k\}).$$

When the support of $X(t)$ is $\mathcal{N}_0$, obviously it is compound Poisson based on another discrete distribution which also has support $\mathcal{N}_0$. When the support of $X(t)$ is $\Re_+$, it is also compound Poisson. See the explanation in Bondesson [1992], p. 16. When the support of $X(t)$ is $\Re$, further research shows that the Lévy process can be decomposed as a Brownian motion plus drift and a jump process. And the only one in Lévy process family, which have a.s. continuous sample paths, is the Brownian motion. See Bhattacharya & Waymire [1990], p. 349-356, and Protter [1990], Section 5 in Chapter 1.

The **compound Poisson process** is a concrete example in Lévy process family, which is defined to have compound Poisson increments. Assume $X(0) = 0$. Then

$$X(t) = \sum_{i=0}^{N(t)} Y_i, \quad \text{where } Y_0 = 0,\ Y_i\ (i \geq 1) \text{ iid and } N(t) \sim \text{Poisson}(\lambda t).$$

The margin can be real, positive or non-negative integer valued depending on the support of Y_i. The cf, LT or pgf of $X(t)$ is then

$$\begin{cases} \varphi_{X(t)}(s) = \exp\left\{\lambda t(\varphi_{Y_1}(s) - 1)\right\}, & \text{if } Y_1 \text{ is a real rv,} \\ \phi_{X(t)}(s) = \exp\left\{\lambda t(\phi_{Y_1}(s) - 1)\right\}, & \text{if } Y_1 \text{ is a positive rv,} \\ G_{X(t)}(s) = \exp\left\{\lambda t(G_{Y_1}(s) - 1)\right\}, & \text{if } Y_1 \text{ is a non-negative integer rv,} \end{cases}$$

where $\varphi_{Y_1}(s)$, $\phi_{Y_1}(s)$ or $G_{Y_1}(s)$ is the cf, LT or pgf of Y_1 respectively. This family contains many processes such as Poisson process, Negative Binomial process, Gamma process, etc.

The increment with three kinds of domains: $(-\infty, +\infty)$, $(0, +\infty)$ and $\{0, 1, 2, \ldots\}$ are of our special interests in the theory of continuous-time generalized AR(1) process. In the rest of this section, we list some specific stationary IIP $\{X(t); t \geq 0\}$ with non-negative integer rv, positive rv and real rv margins respectively for the future use. They are used to construct specific models in the theory of continuous-time generalized AR(1) processes. The non-stationary case can be easily generalized by allowing the the time difference $t_2 - t_1$ to be a function of t_1 and t_2, say a function of $t_2 - t_1$. All starting points are assumed as 0, namely $X(0) = 0$.

Case 1: Non-negative integer rv margins

Example 2.6 *Poisson IIP. The increment* $X(t_2) - X(t_1) \sim$ *Poisson($\lambda(t_2 - t_1)$), with pgf*

$$G_{X(t_2)-X(t_1)}(s) = \mathbf{E}\left(s^{X(t_2)-X(t_1)}\right) = \exp\{\lambda(t_2 - t_1)(s - 1)\},$$

where $\lambda > 0$. *Thus, the margin* $X(t) \sim$ *Poisson(λt) with pgf* $G_{X(t)}(s) = \exp\{\lambda t(s - 1)\}$.

Example 2.7 *Compound Poisson IIP.* $\{Y(t); t \geq 0\}$ *is a Poisson IIP defined as in Example 2.6.* Z *is a non-negative integer rv with pgf* $G_Z(s) = \mathbf{E}(s^Z)$. *The increment of* $\{X(t); t \geq 0\}$ *is defined as*

$$X(t_2) - X(t_1) = \sum_{i=0}^{Y(t_2)-Y(t_1)} Z_i,$$

where $Z_0 = 0$ *and* $Z_1, Z_2, \ldots \overset{i.i.d.}{\sim} Z$. *Thus, the pgf of* $X(t_2) - X(t_1)$ *is*

$$\begin{aligned} G_{X(t_2)-X(t_1)}(s) &= \mathbf{E}\left(s^{X(t_2)-X(t_1)}\right) = \mathbf{E}\left\{\mathbf{E}\left[s^{\sum_{i=0}^{Y(t_2)-Y(t_1)} Z_i} \Big| Y(t_2) - Y(t_1)\right]\right\} \\ &= \mathbf{E}\{G_Z^{Y(t_2)-Y(t_1)(s)}\} = \exp\left\{\lambda(t_2 - t_1)[G_Z(s) - 1]\right\}, \end{aligned}$$

and the pgf of $X(t)$ *is* $G_{X(t)}(s) = \exp\{\lambda t[G_Z(s) - 1]\}$.

$G_Z(s)$ *has a variety of choices. For instance, we can take* $G_Z(s) = p_0 + p_1 s + \cdots + p_n s^n$, *where the* p_i*'s are non-negative and sum to 1. If* $G_Z(s) = s$, *the Poisson IIP in Example 2.6 obtains.*

Z *can be generalized to a continuous-time process* $\{Z(t); t \geq 0\}$ $(Z(0) = 0)$ *with the property that for* $t_1 < t_2 < t_3$,

$$(t_2 - t_1)[G_{Z(t_2-t_1)}(s) - 1] + (t_3 - t_2)[G_{Z(t_3-t_2)}(s) - 1] = (t_3 - t_1)[G_{Z(t_3-t_1)}(s) - 1].$$

Define the increment of $\{X(t); t \geq 0\}$ *as*

$$X(t_2) - X(t_1) = \sum_{i=0}^{Y(t_2)-Y(t_1)} Z_i(t_2 - t_1),$$

where $Z_0(t_2 - t_1) = 0$ *and* $Z_1(t_2 - t_1), Z_2(t_2 - t_1), \ldots \overset{i.i.d.}{\sim} Z(t_2 - t_1)$. *Hence, the pgf of* $X(t_2) - X(t_1)$ *is*

$$G_{X(t_2)-X(t_1)}(s) = \exp\left\{\lambda(t_2 - t_1)\Big[G_{Z(t_2-t_1)}(s) - 1\Big]\right\},$$

Checking the pgf of $X(t_3) - X(t_1)$, *we obtain*

$$\begin{aligned}
G_{X(t_3)-X(t_1)}(s) &= \mathbf{E}\left(s^{X(t_3)-X(t_1)}\right) = \mathbf{E}\left(s^{X(t_3)-X(t_2)+X(t_2-X(t_1))}\right) \\
&= G_{X(t_2)-X(t_1)}(s) G_{X(t_3)-X(t_2)}(s) \\
&= \exp\left\{\lambda(t_2 - t_1)\Big[G_{Z(t_2-t_1)}(s) - 1\Big] + \lambda(t_3 - t_2)\Big[G_{Z(t_3-t_2)}(s) - 1\Big]\right\} \\
&= \exp\left\{\lambda(t_3 - t_1)\Big[G_{Z(t_3-t_1)}(s) - 1\Big]\right\}.
\end{aligned}$$

Therefore, the pgf of $X(t)$ *is* $\exp\left\{\lambda t\Big[G_{Z(t)}(s) - 1\Big]\right\}$.

Example 2.8 *Negative Binomial IIP. Let the increment* $X(t_2) - X(t_1) \sim NB(\theta(t_2 - t_1), \gamma)$, *with pgf*

$$G_{X(t_2)-X(t_1)}(s) = \left(\frac{1-\gamma}{1-\gamma s}\right)^{\theta(t_2-t_1)},$$

where $\theta > 0$ *and* $0 < \gamma < 1$. *So* $X(t)$ *has pgf* $\left(\frac{1-\gamma}{1-\gamma s}\right)^{\theta t}$.

Example 2.9 *Discrete stable IIP. Let the increment $X(t_2) - X(t_1)$ be distributed with discrete stable, i.e., the pgf is*

$$G_{X(t_2)-X(t_1)}(s) = \exp\{-\lambda(t_2 - t_1)(1 - s)^\alpha\},$$

where $\lambda > 0$ and $0 < \alpha < 1$. Then $X(t)$ has pgf $G_{X(t)}(s) = \exp\{-\lambda t(1 - s)^\alpha\}$.

Example 2.10 *Generalized Negative Binomial convolution (GNBC) IIP. Let the increment $X(t_2)-X(t_1)$ be distributed in GNBC with such kind of pgf*

$$G_{X(t_2)-X(t_1)}(s) = \exp\left\{(t_2 - t_1)\int_{(0,1)} \log(\frac{p}{1-qs})V(dq)\right\}.$$

Then $X(t)$ has pgf $G_{X(t)}(s) = \exp\left\{t\int_{(0,1)} \log(\frac{p}{1-qs})V(dq)\right\}$.

Example 2.11 *GC I IIP. Let the increment $X(t_2) - X(t_1)$ be distributed in GC I with pgf of the form*

$$G_{X(t_2)-X(t_1)}(s) = \exp\left\{(t_2 - t_1)\int_{(0,1)} \frac{q(s-1)}{1-qs}V(dq)\right\},$$

Then $X(t)$ has pgf

$$G_{X(t)}(s) = \exp\left\{t\int_{(0,1)} \frac{q(s-1)}{1-qs}V(dq)\right\}.$$

Example 2.12 *GC II IIP. Let the increment $X(t_2) - X(t_1)$ be distributed with GC II with pgf of the form*

$$G_{X(t_2)-X(t_1)}(s) = \exp\left\{(t_2 - t_1)\int_{(0,1)} \frac{q(s-1)(1-\gamma s)}{1-qs}V(dq)\right\}, \quad \gamma > 0.$$

Then $X(t)$ has pgf

$$G_{X(t)}(s) = \exp\left\{t\int_{(0,1)} \frac{q(s-1)(1-\gamma s)}{1-qs}V(dq)\right\}.$$

Case 2: Positive rv margins

Example 2.13 *Gamma IIP. Let the increment $X(t_2) - X(t_1) \sim$ Gamma$(\alpha(t_2 - t_1), \beta)$, with LT*

$$\phi_{X(t_2)-X(t_1)}(s) = \mathbf{E}\left[e^{s(X(t_2)-X(t_1))}\right] = \left(\frac{\beta}{\beta+s}\right)^{\alpha(t_2-t_1)},$$

where $\alpha, \beta > 0$. The LT of $X(t)$ is $\left(\frac{\beta}{\beta+s}\right)^{\alpha t}$, i.e., the LT of Gamma$(\alpha t, \beta)$.

Example 2.14 *Inverse Gaussian IIP. Inverse Gaussian rv X has pdf*

$$f_X(x;\mu,\lambda) = \sqrt{\lambda/(2\pi x^3)}\exp\{-\lambda(x-\mu)^2/(2\mu^2 x)\}, \quad x>0,$$

where $\mu, \lambda > 0$, *and the LT is*

$$\phi_X(s) = \mathbf{E}[e^{sX}] = \exp\left\{\frac{\lambda}{\mu}\left[1-\left(1+\frac{2\mu^2 s}{\lambda}\right)^{1/2}\right]\right\}.$$

Now let $\lambda = k\mu^2$, *where* k *is a constant. Then*

$$\phi_X(s) = \exp\left\{k\mu\left[1-\left(1+\frac{2}{k}s\right)^{1/2}\right]\right\}.$$

For this special form, we can construct Inverse Gaussian IIP $\{X(t); t\geq 0\}$, *such that the increment* $X(t_2)-X(t_1)$ *has LT*

$$\phi_{X(t_2)-X(t_1)}(s) = \exp\left\{k(t_2-t_1)\left[1-\left(1+\frac{2}{k}s\right)^{1/2}\right]\right\}.$$

Hence, the LT of $X(t)$ *is*

$$\phi_{X(t)}(s) = \exp\left\{kt\left[1-\left(1+\frac{2}{k}s\right)^{1/2}\right]\right\}.$$

Example 2.15 *GGC IIP. Let the increment* $X(t_2)-X(t_1)$ *be distributed in generalized Gamma convolution distributed with LT*

$$\phi_{X(t_2)-X(t_1)}(s) = \exp\left\{(t_2-t_1)\int \log\left(\frac{u}{u+s}\right)U(du)\right\}.$$

Hence, $X(t)$ *has LT*

$$\phi_{X(t)}(s) = \exp\left\{t\int \log\left(\frac{u}{u+s}\right)U(du)\right\}.$$

This family is a big class, consisting of many known distributions.

Example 2.16 *GCMED IIP. Let the increment* $X(t_2)-X(t_1)$ *be distributed in GCMED with LT*

$$\phi_{X(t_2)-X(t_1)}(s) = \exp\left\{(t_2-t_1)\int_{(0,+\infty)}\frac{-s}{u+s}U(du)\right\},$$

Then $X(t)$ *has LT*

$$\phi_{X(t)}(s) = \exp\left\{t\int_{(0,+\infty)}\frac{-s}{u+s}U(du)\right\}.$$

Example 2.17 *GC III IIP. Let the increment $X(t_2) - X(t_1)$ be distributed in GC III with LT*

$$\phi_{X(t_2)-X(t_1)}(s) = \exp\left\{(t_2 - t_1)\int_{(0,\infty)} \frac{1-\gamma}{1-\gamma-\gamma u} \log\left(\frac{(1-\gamma+\gamma s)u}{(1-\gamma)(u+s)}\right) U(du)\right\}, \quad 0 \le \gamma < 1.$$

Then $X(t)$ has LT

$$\phi_{X(t)}(s) = \exp\left\{t\int_{(0,\infty)} \frac{1-\gamma}{1-\gamma-\gamma u} \log\left(\frac{(1-\gamma+\gamma s)u}{(1-\gamma)(u+s)}\right) U(du)\right\}.$$

Case 3: Real rv margins

Example 2.18 *Gaussian IIP (Brownian Motion). This is well known. The increment $X(t_2) - X(t_1) \sim N(0, t_2 - t_1)$.*

Example 2.19 *Cauchy IIP. A Cauchy$(0, \lambda)$ rv X has pdf*

$$f_X(x|\lambda) = \frac{1}{\pi}\frac{\lambda}{\lambda^2 + x^2}, \quad -\infty < x < +\infty, \quad \lambda > 0,$$

and cf

$$\varphi_X(s) = \mathbf{E}[e^{isX}] = e^{-\lambda|s|}.$$

To obtain a Cauchy IIP, just set the increment $X(t_2) - X(t_1) \sim$ Cauchy$(0, \lambda(t_2 - t_1))$.

Example 2.20 *Stable Paretian (or stable non-Gaussian) IIP. Consider a special case in stable Paretian family, which has cf of form*

$$\varphi(s) = \exp\{-\lambda|s|^\alpha\}, \quad 0 < \alpha \le 2.$$

(When $\alpha = 2$, it's normal distribution.) To obtain a stable Paretian IIP, just let the cf of the increment $X(t_2) - X(t_1)$ be

$$\varphi_{X(t_2)-X(t_1)}(s) = \exp\{-\lambda(t_2 - t_1)|s|^\alpha\}.$$

Thus, the cf of $X(t)$ is $\exp\{-\lambda t|s|^\alpha\}$.

In summary, all pgf's, LT's or cf's in these examples are of exponential form with $(t_2 - t_1)$ as linear parameter in the exponent. We pick up such a form because we want to change the product form to summation form in obtaining the pgf, LT or cf form of the continuous-time generalized AR(1) process. In all these cases, $X(t)$ is infinitely divisible.

Chapter 3

Self-generalized distributions and extended-thinning operations

In this chapter, we shall propose a new concept of closure for probability distributions. Families with this closure property are called **self-generalized distributions**. The support of these families can be non-negative integer or positive real. They induce a class of stochastic operators, which we call **extended-thinning operators**. These stochastic operators will be applied in generalized Ornstein-Uhlenbeck stochastic differential equations, and the property of self-generalizability plays a crucial role in model construction of continuous-time generalized AR(1) processes (see Chapter 4).

In Section 3.1, we shall define the self-generalized distribution in the non-negative integer-valued case and the positive-valued case respectively, and give some examples as well. We discuss the properties of self-generalized distributions in Section 3.2, as well as construction of self-generalized distributions in Section 3.3. Finally, we propose the extended-thinning operations in Section 3.4.

3.1 Self-generalized distributions

A family of self-generalized distributions has a pgf or LT which is closed under some compound one-parameter operation. The support is non-negative integer or positive real. We give the thorough discussion on both cases in the following subsections.

3.1.1 Non-negative integer case and examples

Suppose K is a non-negative integer random variable, taking value on $\{0, 1, 2, \dots\}$. Now we define the self-generalized distribution in non-negative integer case.

Definition 3.1 *Let Λ be a subset of reals that is closed under multiplication. Suppose K has cdf $F(x;\alpha)$ depending on a parameter α, $\alpha \in \Lambda$. The probability generating function is*

$$G_K(s;\alpha) = \mathbf{E}[s^K] = \int_0^\infty s^x dF(x;\alpha) = \sum_{i=0}^\infty s^i \Pr[K = i].$$

If

$$G_K(G_K(s;\alpha);\alpha') = G_K(s;\alpha\alpha'),$$

then the distribution family $\{F(x;\alpha); \alpha \in \Lambda\}$ is said to be self-generalized with respect to parameter α. For brevity and convenience, we say that K is self-generalized with respect to parameter α to refer to the self-generalizablity of the distribution family $\{F(x;\alpha); \alpha \in \Lambda\}$.

In non-negative integer case, the self-generalizability is closed under the compound operation for the probability generating function. This closure operation corresponds to an interesting stochastic representation (refer to Property 3.6, which leads us to call it self-generalizability).

To illustrate this new family in non-negative integer case, we give five examples in the remainder of this subsection. For the sake of saving space and reducing redundancy in the later study, we label them from **I1** to **I5**.

Example 3.1 (I1): *Let $K \sim$ Bernoulli(α) ($0 \le \alpha \le 1$). The pgf of K is $G_K(s;\alpha) = (1-\alpha) + \alpha s$. Thus*

$$\begin{aligned} G_K(G_K(s;\alpha);\alpha') &= (1-\alpha') + \alpha' G_K(s;\alpha) = (1-\alpha') + \alpha'[(1-\alpha)+\alpha s] \\ &= (1-\alpha\alpha') + \alpha\alpha' s = G_K(s;\alpha\alpha'). \end{aligned}$$

Therefore K is self-generalized with respect to parameter α.

Example 3.2 (I2): *Consider $K = ZI$, where $I \sim$ Bernoulli(a), $Z = Z' + 1$, $Z' \sim NB(1, \frac{1}{1+b})$, and $a = \frac{(1-\gamma)\alpha}{1-\gamma\alpha}$, $b = \frac{1-\gamma}{(1-\alpha)\gamma}$, $0 \le \alpha \le 1$, $0 \le \gamma < 1$. Here the parameter γ is fixed. Note that Z, Z' have Geometric distributions with positive integer support and non-negative integer support respectively. The pgf of Z is $(1-q)s/(1-qs)$ where $q = (1+b)^{-1}$. A straightforward calculation leads to*

$$G_K(s;\alpha) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}.$$

It follows that

$$\begin{aligned} G_K(G_K(s;\alpha);\alpha') &= \frac{(1-\alpha')+(\alpha'-\gamma)G_K(s;\alpha)}{(1-\alpha'\gamma)-(1-\alpha')\gamma G_K(s;\alpha)} \\ &= \frac{(1-\alpha')+(\alpha'-\gamma)\frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}}{(1-\alpha'\gamma)-(1-\alpha')\gamma\frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}} \\ &= \frac{(1-\alpha')[(1-\alpha\gamma)-(1-\alpha)\gamma s]+(\alpha'-\gamma)[(1-\alpha)+(\alpha-\gamma)s]}{(1-\alpha'\gamma)[(1-\alpha\gamma)-(1-\alpha)\gamma s]-(1-\alpha')\gamma[(1-\alpha)+(\alpha-\gamma)s]} \\ &= \frac{[(1-\alpha')(1-\alpha\gamma)+(\alpha'-\gamma)(1-\alpha)]+[-(1-\alpha')(1-\alpha)\gamma+(\alpha'-\gamma)(\alpha-\gamma)]s}{[(1-\alpha'\gamma)(1-\alpha\gamma)-(1-\alpha')(1-\alpha)\gamma]-[(1-\alpha'\gamma)(1-\alpha)\gamma+(1-\alpha')(\alpha-\gamma)\gamma]s} \\ &= \frac{(1-\gamma)(1-\alpha\alpha')+(1-\gamma)(\alpha\alpha'-\gamma)s}{(1-\gamma)(1-\alpha\alpha'\gamma)-(1-\gamma)(1-\alpha\alpha')\gamma s} \\ &= \frac{(1-\alpha\alpha')+(\alpha\alpha'-\gamma)s}{(1-\alpha\alpha'\gamma)-(1-\alpha\alpha')\gamma s} \\ &= G_K(s;\alpha\alpha'). \end{aligned}$$

Hence, K is self-generalized with respect to α.

When $\gamma = 0$, this becomes Example 3.1.

Example 3.3 (I3): *Let K be a right-shift power series random variable, taking values in $\{1, 2, 3, \ldots\}$. The pgf is*

$$G_K(s;\alpha) = 1 - (1-s)^\alpha, \qquad 0 < \alpha \le 1.$$

It follows that

$$\begin{aligned} G_K(G_K(s;\alpha);\alpha') &= 1 - (1 - G_K(s;\alpha))^{\alpha'} = 1 - ((1-s)^\alpha)^{\alpha'} = 1 - (1-s)^{\alpha\alpha'} \\ &= G_K(s;\alpha\alpha'). \end{aligned}$$

This shows that K is self-generalized.

Example 3.4 (I4): *Suppose the non-negative integer random variable K has pgf*

$$G_K(s;\alpha) = c^{-1}[1 - e^{-\theta(1-\alpha)}(1-cs)^\alpha],$$

where $0 \le \alpha \le 1$, $c = 1 - e^{-\theta}$, $\theta \ge 0$. The parameter θ is fixed. Then

$$\begin{aligned} G_K(G_K(s;\alpha);\alpha') &= c^{-1}[1 - e^{-\theta(1-\alpha')}(1 - cG_K(s;\alpha))^{\alpha'}] \\ &= c^{-1}[1 - e^{-\theta(1-\alpha')}\left(e^{-\theta(1-\alpha)}(1-cs)^\alpha\right)^{\alpha'}] \\ &= c^{-1}[1 - e^{-\theta(1-\alpha'+\alpha'-\alpha\alpha')}(1-cs)^{\alpha\alpha'}] \\ &= c^{-1}[1 - e^{-\theta(1-\alpha\alpha')}(1-cs)^{\alpha\alpha'}] \\ &= G_K(s;\alpha\alpha'). \end{aligned}$$

Thus, K is self-generalized with respect to α.

Since $\lim_{\theta\to 0} c^{-1}[1 - e^{-\theta(1-\alpha)}(1-cs)^\alpha] = 1 - \alpha + \alpha s$, the lower boundary leads to Example 3.1.

Example 3.5 (I5): *Consider the non-negative integer random variable K which has pgf*

$$G_K(s;\alpha) = 1 - \alpha^\theta(1-\gamma)^\theta\left[(1-\alpha)\gamma + (1-\gamma)(1-s)^{-1/\theta}\right]^{-\theta},$$

where $0 \le \alpha \le 1$, $0 < \gamma < 1$ and $\theta \ge 1$. Here the parameter γ and θ are fixed. Then, it follows that

$$\begin{aligned} G_K(G_K(s;\alpha);\alpha') &= 1 - (\alpha')^\theta(1-\gamma)^\theta\left[(1-\alpha')\gamma + (1-\gamma)(1 - G_K(s;\alpha))^{-1/\theta}\right]^{-\theta} \\ &= 1 - (\alpha')^\theta(1-\gamma)^\theta\left[(1-\alpha')\gamma + \alpha^{-1}\left((1-\alpha)\gamma + (1-\gamma)(1-s)^{-1/\theta}\right)^{-\theta}\right]^{-\theta} \end{aligned}$$

$$
\begin{aligned}
&= 1-(\alpha')^{\theta}(1-\gamma)^{\theta}\left[\frac{(1-\alpha\alpha')\gamma+(1-\gamma)(1-s)^{-1/\theta}}{\alpha}\right]^{-\theta} \\
&= 1-(\alpha\alpha')^{\theta}(1-\gamma)^{\theta}\left[(1-\alpha\alpha')\gamma+(1-\gamma)(1-s)^{-1/\theta}\right]^{-\theta} \\
&= G_K(s;\alpha\alpha').
\end{aligned}
$$

Hence, K is self-generalized with respect to α.

When $\theta = 1$, the pgf becomes

$$
G_K(s;\alpha) = 1-\alpha(1-\gamma)\left[(1-\alpha)\gamma+(1-\gamma)(1-s)^{-1}\right]^{-1} = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s},
$$

which is the pgf of Example 3.2. Therefore, Example 3.2 is a special case in this family.

We summarize the existing relationship among these classes: **I1** $\subset$ **I2** $\subset$ **I5** and **I1** $\subset$ **I4**.

3.1.2 Positive case and examples

In this section, we define self-generalizability for positive rv's.

Definition 3.2 *Let Λ be a subset of reals that is closed under multiplication. Suppose K has cdf $F(x;\alpha)$ depending on a parameter α, $\alpha \in \Lambda$. The Laplace transformation of K is*

$$
\phi_K(s;\alpha) = \mathbf{E}\left[e^{-sK}\right].
$$

If

$$
\phi_K(-\log\phi_K(s;\alpha);\alpha') = \phi_K(s;\alpha\alpha'),
$$

then the distribution family $\{F(x;\alpha);\alpha \in \Lambda\}$ is said to be self-generalized with respect to the parameter α. For convenience, we say that K is self-generalized with respect to the parameter α to refer to the self-generalizablity of the distribution family $\{F(x;\alpha);\alpha \in \Lambda\}$.

In positive rv case, self-generalizability is closed under the negative logarithm-compound operation for the LT. This seems quite different from non-negative integer case, where self-generalizability is closed under compounding for the pgf. However, recalling that $G_K(s;\alpha) = \mathbf{E}(s^K) = \mathbf{E}\left(e^{-(-\log s)K}\right) = \phi_K(-\log s;\alpha)$, one can induce from

$$G_K(G_K(s;\alpha);\alpha') = G_K(s;\alpha\alpha')$$

to

$$\phi_K(-\log\phi_K(-\log s;\alpha);\alpha') = \phi(-\log s;\alpha\alpha').$$

Replacing $-\log s$ with s, we see that the non-negative integer self-generalized distribution still satisfy the definition for positive rv case. This implies that both definitions are the same in principle. Of course, we can use the definition regarding LT to unite both cases; however, the pgf is more convenient than the LT for the non-negative integer case.

Similarly, this kind of closure of LT with respect to a parameter corresponds to another interesting stochastic representation (see property 3.7) leading to the terminology of self-generalizability.

The following are five positive rv self-generalizability examples. Similarly, we label them from **P1** to **P5**; they form pairs with **I1** to **I5**.

Example 3.6 (P1): *Suppose K is a degenerate rv on point α ($\alpha > 0$). Then the LT of K is $\phi_K(s;\alpha) = e^{-\alpha s}$. It is easy to check self-generalizability, because*

$$\phi_K(-\log\phi_K(s;\alpha);\alpha') = e^{-\alpha'[-\log\phi_K(s;\alpha)]} = e^{-\alpha'[\alpha s]} = e^{-\alpha\alpha' s}.$$

Example 3.7 (P2): *Suppose K is a positive random variable with LT*

$$\phi_K(s;\alpha) = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\},$$

where $0 \le \alpha \le 1$, $0 \le \gamma < 1$ and γ is fixed. This is the LT of a compound Poisson distribution with exponential rv's. It follows that

$$\begin{aligned}
\phi_K(-\log\phi_K(s;\alpha);\alpha') &= \exp\left\{-\frac{\alpha'(1-\gamma)[-\log\phi_K(s;\alpha)]}{(1-\gamma)+(1-\alpha')\gamma[-\log\phi_K(s;\alpha)]}\right\} \\
&= \exp\left\{-\frac{\alpha'(1-\gamma)\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}}{(1-\gamma)+(1-\alpha')\gamma\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}}\right\} \\
&= \exp\left\{-\frac{\alpha\alpha'(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s+\alpha(1-\alpha')\gamma s}\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \exp\left\{-\frac{\alpha\alpha'(1-\gamma)s}{(1-\gamma)+(1-\alpha\alpha')\gamma s}\right\} \\
&= \phi_K(s;\alpha\alpha').
\end{aligned}
$$

Hence, K is self-generalized with respect to α.

When $\gamma = 0$, this becomes Example 3.6, namely **P1**.

Example 3.8 (P3): *Let K be positive stable with LT $\phi_K(s;\alpha) = \exp\{-s^\alpha\}$, where $0 < \alpha \leq 1$. Then*

$$
\phi_K(-\log\phi_K(s;\alpha);\alpha') = \exp\left\{-[-\log\phi_K(s;\alpha)]^{\alpha'}\right\} = \exp\left\{-[s^\alpha]^{\alpha'}\right\} = \exp\left\{-s^{\alpha\alpha'}\right\}
$$
$$
= \phi_K(s;\alpha\alpha').
$$

Thus K is self-generalized.

Example 3.9 (P4): *Consider positive random variable K with LT*

$$
\phi_K(s;\alpha) = \exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha - 1}{e^\theta - 1}\right\},
$$

where $0 \leq \alpha \leq 1$, $\theta \geq 0$ and θ is fixed. It follows that

$$
\begin{aligned}
\phi_K(-\log\phi_K(s;\alpha);\alpha') &= \exp\left\{-\frac{\left[1+(e^\theta-1)\{-\log\phi_K(s;\alpha)\}\right]^{\alpha'}-1}{e^\theta-1}\right\} \\
&= \exp\left\{-\frac{\left[1+(e^\theta-1)\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right]^{\alpha'}-1}{e^\theta-1}\right\} \\
&= \exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^{\alpha\alpha'}-1}{e^\theta-1}\right\} \\
&= \phi_K(s;\alpha\alpha').
\end{aligned}
$$

Thus, K is self-generalized with respect to α.

Since $\lim_{\theta\to 0}\exp\left\{-\frac{\left[1+(e^\theta-1)s\right]^\alpha-1}{e^\theta-1}\right\} = e^{-\alpha s}$, the lower boundary leads to Example 3.6.

Example 3.10 (P5): *Consider positive random variable K with LT*

$$\phi_K(s;\alpha) = \exp\left\{-\left[\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma + (1-\gamma)s^{-\frac{1}{\theta}}}\right]^\theta\right\},$$

where $0 \le \alpha \le 1$, $\theta \ge 1$ and $0 < \gamma < 1$. The parameters θ and γ are fixed. Now we check the self-generalizability.

$$\begin{aligned}
\phi_K(-\log\phi_K(s;\alpha);\alpha') &= \exp\left\{-\left[\frac{\alpha'(1-\gamma)}{(1-\alpha')\gamma + (1-\gamma)[-\log\phi_K(s;\alpha)]^{-\frac{1}{\theta}}}\right]^\theta\right\} \\
&= \exp\left\{-\left[\frac{\alpha'(1-\gamma)}{(1-\alpha')\gamma + (1-\gamma)\left(\left[\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma+(1-\gamma)s^{-\frac{1}{\theta}}}\right]^\theta\right)^{-\frac{1}{\theta}}}\right]^\theta\right\} \\
&= \exp\left\{-\left[\frac{\alpha'(1-\gamma)}{(1-\alpha')\gamma + (1-\gamma)\frac{(1-\alpha)\gamma+(1-\gamma)s^{-\frac{1}{\theta}}}{\alpha(1-\gamma)}}\right]^\theta\right\} \\
&= \exp\left\{-\left[\frac{\alpha\alpha'(1-\gamma)}{\alpha(1-\alpha')\gamma + (1-\alpha)\gamma + (1-\gamma)s^{-\frac{1}{\theta}}}\right]^\theta\right\} \\
&= \exp\left\{-\left[\frac{\alpha\alpha'(1-\gamma)}{(1-\alpha\alpha')\gamma + (1-\gamma)s^{-\frac{1}{\theta}}}\right]^\theta\right\} \\
&= \phi_K(s;\alpha\alpha').
\end{aligned}$$

This implies that K is self-generalized.

When $\theta = 1$, the LT will be

$$\phi_K(s;\alpha) = \exp\left\{-\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma + (1-\gamma)s^{-1}}\right\} = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma) + (1-\alpha)\gamma s}\right\},$$

which is the LT of the Example 3.7. Hence, Example 3.7 is a special case in this bigger family.

The relationships among these classes are: **P1** $\subset$ **P2** $\subset$ **P5** and **P1** $\subset$ **P4**.

3.2 Properties of self-generalized distributions

In this section, we shall discuss some properties of the proposed self-generalized distributions in proceeding section. These involve the properties of their means, boundaries, as well as possible stochastic representations for the compounded pgf and LT.

For non-triviality we assume that the distributions of a self-generalized family $\{F_K(\cdot;\alpha);\alpha\in\Lambda\}$ are distinct for different $\alpha\in\Lambda$. Thus, trivial cases like K being a constant 0 for the entire family are excluded.

Theorem 3.2.1 *Suppose K is a self-generalized random variable. The expectation of K is:*

$$h(\alpha)=\mathbf{E}(K)=\begin{cases} \left.\frac{\partial G_K(s,\alpha)}{\partial s}\right|_{s=1}, & \text{if } K \text{ is non-negative integer rv,}\\ -\left.\frac{\partial \phi_K(s,\alpha)}{\partial s}\right|_{s=0}, & \text{if } K \text{ is positive rv.}\end{cases}$$

Then

$$h(\alpha)h(\alpha')=h(\alpha\alpha').$$

(This is the Cauchy functional equation.)

Proof: Taking partial derivative with respect to s for both sides in the self-generalizability definitions, by the chain rule, we obtain

$$\frac{\partial G_K(G_k(s;\alpha);\alpha')}{\partial G_k(s;\alpha)}\times\frac{\partial G_k(s;\alpha)}{\partial s}=\frac{\partial G_k(s;\alpha\alpha')}{\partial s},$$

and

$$\frac{\partial \phi_K(-\log\phi_K(s;\alpha);\alpha')}{\partial(-\log\phi_K(s;\alpha))}\times\frac{\partial(-\log\phi_K(s;\alpha))}{\partial s}=\frac{\partial\phi_K(s;\alpha\alpha')}{\partial s}.$$

The latter can be further written as

$$\frac{\partial \phi_K(-\log\phi_K(s;\alpha);\alpha')}{\partial(-\log\phi_K(s;\alpha))}\times\left(-\frac{1}{\phi_K(s;\alpha)}\frac{\partial\phi_K(s;\alpha)}{\partial s}\right)=\frac{\partial\phi_K(s;\alpha\alpha')}{\partial s}.$$

By setting $s=1$ or $s=0$, we can obtain the related equations regarding to the expectations associated with parameter values α and α' for non-negative integer and positive self-generalized distributions respectively.

Since

$$G_K(1,\alpha) = \mathbf{E}(1^K) = \mathbf{E}(1) = 1 \qquad \text{and} \qquad \phi_K(0,\alpha) = \mathbf{E}(e^{-0\times K}) = \mathbf{E}(1) = 1,$$

we have

$$\frac{\partial G_K(G_k(s;\alpha);\alpha')}{\partial G_k(s;\alpha)}\bigg|_{s=1} = \frac{\partial G_K(s_1;\alpha')}{\partial s_1}\bigg|_{s_1=1} = h(\alpha')$$

and

$$\frac{\partial \phi_K(-\log\phi_K(s;\alpha);\alpha')}{\partial(-\log\phi_K(s;\alpha))}\bigg|_{s=0} = \frac{\partial \phi_K(s_2;\alpha')}{\partial s_2}\bigg|_{s_2=0} = -h(\alpha'),$$

where $s_1 = G_k(s;\alpha)$ and $s_2 = -\log\phi_K(s;\alpha)$ respectively. Thus, for the non-negative integer self-generalized distribution, it is straightforward to obtain

$$h(\alpha)h(\alpha') = h(\alpha\alpha').$$

For the positive self-generalized distribution, we first have the following equation

$$-h(\alpha') \times h(\alpha) = -h(\alpha\alpha'),$$

which simply leads to $h(\alpha)h(\alpha') = h(\alpha\alpha')$.

To distinguish the self-generalized random variables with different values of parameter α, we adopt $X(\alpha)$ to denote the one corresponding to α. Hence, $X(\alpha)$ and $X(\alpha')$ will be from the same self-generalized distribution family, but with parameter values α and α' respectively.

Since the closure property of self-generalized distribution is with respect to the parameter α, i.e., $\alpha\alpha' \in \Lambda$, the possible domain Λ for α are the real set, or the intervals $[-1,1]$ and $(-\infty,-1] \cup [1,\infty)$ (including or excluding the boundaries), or the positive real set, or the intervals $[0,1]$, $[1,\infty)$ (boundaries could be excluded).

Note that reparametrizing by taking inverse, there is one-to-one mapping between $(0,1]$ and $[1,\infty)$, and such a reparameterization keeps the self-generalizability. This feature can be seen in the following reseasoning. Suppose $G_K^*(s;\alpha) = G_K(s;1/\alpha)$, where $\alpha \in (0,1]$. Then $1/\alpha \in [1,\infty)$, and $G_K^*(G_K^*(s;\alpha);\alpha') = G_K(G_K(s;1/\alpha);1/\alpha') = G_K(s;1/(\alpha\alpha')) = G_K^*(s,\alpha\alpha')$. Hence, $(0,1]$ and $[1,\infty)$ are equivalent. We only need to consider domain $(0,1]$. However, we can't find a one-to-one mapping between $(0,1)$ and $(0,\infty)$ such that the self-generalizability is kept.

In the remainder of this section, we are only interested in non-negative set Λ: $[0, 1]$ and $[0, +\infty)$. In fact, the theory of continuous-time generalized AR(1) processes only needs $\Lambda = [0, 1]$. The boundary 0 may be excluded, but the boundary 1 is always included in Λ through the remainder of this thesis.

The inclusion of 1 has been justified from the definition of self-generalizability. This property plays an important roles in the theory of continuous-time generalized AR(1) processes.

Property 3.1 *Let K be a self-generalized rv. Then, $G_K(s; 1) = s$ or $\phi_K(s; 1) = e^{-s}$, that is, $K(\alpha) \equiv 1$ for $\alpha = 1$.*

Proof: We consider the discrete and positive case respectively.

(1) Discrete case. $G_K(s; \alpha)$ is increasing in s for any $\alpha \in \Lambda$. Hence, from

$$G_K(G_K(s; 1); \alpha) = G_K(s; \alpha), \qquad \text{for all } s,$$

we conclude that either $G_K(s; 1) = s$ or $G_K(s; \alpha) = 1$. However, $G_K(s; \alpha) = 1$ means that $K(\alpha)$ takes value 0 with probability 1 for all α. This contradicts the non-triviality assumption. Therefore, the only choice is $G_K(s; 1) = s$.

(2) Positive case. $\phi_K(s; \alpha)$ is decreasing in s for any $\alpha \in \Lambda$. Since

$$\phi_K(-\log \phi_K(s; \alpha); 1) = \phi_K(s; \alpha), \qquad \text{for all } s,$$

either $-\log \phi_K(s; 1) = s$ or $\phi_K(s; \alpha) = 1$ holds. However, the latter implies that $\phi_K(s; \alpha) = 1$, contradicting to the non-triviality assumption. Thus, $\phi_K(s; 1) = e^{-s}$.

Both cases imply that $K(1) \equiv 1$.

With extra conditions, we can obtain the functional form of the expectation of a self-generalized rv by Theorem 3.2.1.

Property 3.2 *Suppose $h(\alpha)$, the expectation $h(\alpha)$ of a self-generalized rv K, is continuous with respect to α. Then*

$$h(\alpha) = \alpha^r.$$

Here r can be positive or negative ($r = 0$ is eliminated to avoid triviality). If $h(\alpha)$ is bounded in $(0, 1)$, then $r > 0$. If $h(\alpha)$ is finite but unbounded in $(0, 1)$, then $r < 0$.

Proof: Under the continuity assumption, it is straightforward to deduce

$$h(\alpha) = \alpha^r$$

by Theorem 3.2.1. Excluding the trivial case, we know that $r > 0$ or $r < 0$. $h(\alpha) = \alpha^r$ will goes to 0 or ∞ according to $r > 0$ or $r < 0$ respectively. This completes the proof.

Note that $h(\alpha)$ may not be finite when $\alpha \neq 1$. See the cases of **(I3)** and **(P3)** in the following example.

Example 3.11 *Checking the non-negative integer and positive self-generalized distributions in last two sections, we find $h(\alpha) = \alpha$ for* **I1, I2, I4, P1, P2** *and* **P4**. *For* **I3**, *the power series distribution, and* **P3**, *the positive stable distribution, the expectations are infinity, i.e., $h(\alpha) = \infty$ when $0 < \alpha < 1$. For* **I5** *and* **P5**, *$h(\alpha) = \alpha^\theta$. Also see the summary for the mean and variance of self-generalized distribution in Table 9.1.*

Note that if $K(\alpha)$ has finite expectation for all $\alpha > 0$, namely $\mathbf{E}[K(\alpha)] = \alpha^r$, $r > 0$, then $K(\alpha)$ can be reparameterized by $\alpha^{1/r}$ so that $\mathbf{E}[K(\alpha)] = \left(\alpha^{1/r}\right)^r = \alpha$. This is because that for the reparameterization transformation, $\alpha^{1/r}(\alpha')^{1/r} = (\alpha\alpha')^{1/r}$, is closed under multiplication.

Property 3.3 *Let K be a self-generalized rv.*

(1) If the boundary 0 is included in the domain Λ of parameter α, then $G_K(s; 0) = 1$ or $\phi_K(s; 0) = 1$, that is, $K(\alpha) \equiv 0$ for $\alpha = 0$.

(2) If the boundary 0 is not in domain Λ, but the expectation of K is bounded and continuous with respect to α, then $K(\alpha) \xrightarrow{\mathrm{P}} 0$ as $\alpha \to 0$.

Proof:

(1) The boundary 0 is included in the domain Λ. Then by self-generalizability, it follows that for $\alpha \in \Lambda$,

$$G_K(G_K(s; \alpha); 0) = G_K(s; 0), \qquad \text{for all } s,$$

and

$$\phi_K(-\log \phi_K(s;\alpha);0) = \phi_K(s;0), \qquad \text{for all } s.$$

Because of the monotonicity of $G_K(s;\alpha)$ and $\phi_K(s;\alpha)$ with respect to s, the above equations yield that

$$G_K(s;\alpha) = s \qquad \text{or} \qquad G_K(s;0) = 1,$$

and

$$-\log \phi_K(s;\alpha) = s \qquad \text{or} \qquad \phi_K(s;0) = 1.$$

But $G_K(s;\alpha) = s$ and $\phi_K(s;\alpha) = s$ will lead to the triviality that $K(\alpha) \equiv 1$, thus, it must hold that $G_K(s;0) = 1$ and $\phi_K(s;0) = 1$, namely $K(0) \equiv 0$.

(2) The boundary 0 is not included in the domain Λ. Then by Property 3.2, $\lim_{\alpha \to 0^+} \mathbf{E}[K(\alpha)] = 0$. By non-negativity, we obtain that $K(\alpha) \xrightarrow{\mathrm{P}} 0$ as $\alpha \to 0$.

The support of self-generalized rv is of interests. Below is the feature of support of a non-negative integer self-generalized rv.

Property 3.4 *Suppose $G_K(s;\alpha) = p_0(\alpha) + p_1(\alpha)s + \cdots + p_n(\alpha)s^n$, with $n \geq 1$, and $p_n(\alpha) > 0$ for all $\alpha \neq 1$ if n is finite. Then the order n is either 1 or ∞.*

This is because that the polynomial degree of $G_K(G_K(s;\alpha);\alpha')$ will be n^2. Only 1 or $+\infty$ are possible choices. Therefore, any distribution with domain in a finite non-negative integer set other than $\{0,1\}$, such as Binomial distribution, can not be self-generalized.

The pgf $G_K(s;\alpha)$ and the LT $\phi_K(s;\alpha)$ are uniformly continuous in s on their range $[0,1]$ and $[0,\infty]$ respectively. How about their continuity in α? This leads to the following conclusion.

Property 3.5 *Let K be a self-generalized rv.*

(1) *For K being a non-negative integer-valued rv with pgf $G_K(s;\alpha)$, if $G_K(s;\alpha)$ is left continuous at $\alpha = 1$, then $G_K(s;\alpha)$ is continuous in α in $(0,1]$.*

Furthermore, if $G_K(s;\alpha)$ is right continuous at $\alpha = 0$ and $\lim_{\alpha \to 0^+} G_K(s;\alpha) = 1$, then $G_K(s;\alpha)$ is uniformly continuous in α in $[0,1]$.

(2) *For K being a positive-valued rv with LT $\phi_K(s;\alpha)$, if $\phi_K(s;\alpha)$ is left continuous at $\alpha = 1$, then $\phi_K(s;\alpha)$ is continuous in α in $(0,1)$.*

Furthermore, if $\phi_K(s;\alpha)$ is right continuous at $\alpha = 0$ and $\lim_{\alpha\to 0^+} \phi_K(s;\alpha) = 1$, then $\phi_K(s;\alpha)$ is uniformly continuous in α in $[0,1]$.

Proof:

(1) Suppose $\alpha' < \alpha$. It follows that

$$G_K(s;\alpha) - G_K(s;\alpha') = G_K(s;\alpha) - G_K(s;\beta\alpha) = G_K(s;\alpha) - G_K(G_K(s;\beta);\alpha),$$

where $\beta = \alpha'/\alpha$. When $\alpha' \to \alpha$, $\beta \to 1$. Since $\lim_{\alpha\to 1^-} G_K(s;\alpha) = s$, thus, $G_K(s;\beta) \to s$. By the continuity of a pgf in s, we know that $G_K(s;\alpha) - G_K(G_K(s;\beta);\alpha) \to 0$. This implies that $G_K(s;\alpha)$ is continuous in α in $(0,1)$.

If $G_K(s;\alpha)$ is left and right continuous at its two boundaries of α, then $G_K(s;\alpha)$ is continuous in α in the closed interval $[0,1]$, which shows that $G_K(s;\alpha)$ is uniformly continuous in α in $[0,1]$.

(2) Applying the same reasoning, we can obtain the similar conclusion for $\phi_K(s;\alpha)$.

Remark:

For **I3**, K does not have finite mean, and in fact, the right limit $\lim_{\alpha\to 0^+} G_K(s;\alpha) = 0$; this is not a pgf. Similarly, for **P3**, K does not have finite mean too, and the right limit $\lim_{\alpha\to 0^+} \phi_K(s;\alpha) = e^{-1}$, which is not a LT. In both cases, the pgf or LT is left continuous at $\alpha = 1$. As to **I1**, **I2**, **I4**, **I5** and **P1**, **P2**, **P4**, **P5**, K has finite mean, and its pgf or LT is continuous at boundaries $\alpha = 0$ and 1.

Stochastic representations of $G_K(G_K(s;\alpha);\alpha')$ and $\phi_K(-\log\phi_K(s;\alpha);\alpha')$ are of interest. Here we discuss their possible representations.

Property 3.6 *Suppose $K(\alpha)$ and $K(\alpha')$ are distributed from the same non-negative integer self-generalized distribution family with respective parameter values α and α'. Then $\sum_{i=0}^{K(\alpha')} K_i(\alpha)$ has pgf $G_K(G_K(s;\alpha);\alpha')$, where $K_0(\alpha) = 0$; $K_i(\alpha) \overset{i.i.d.}{\sim} F_K(\cdot;\alpha)$ and are independent of $K(\alpha')$.*

Proof:

$$\begin{aligned}
\mathbf{E}\left\{s^{\sum_{i=0}^{K(\alpha')} K_i(\alpha)}\right\} &= \mathbf{E}\left[\mathbf{E}\left\{s^{\sum_{i=0}^{K(\alpha')} K_i(\alpha)}\middle|K(\alpha')\right\}\right] = \mathbf{E}\left[\left(\mathbf{E}\left\{s^{K_1(\alpha)}\right\}\right)^{K(\alpha')}\right] \\
&= \mathbf{E}\left[G_K^{K(\alpha')}(s;\alpha)\right] = G_K(G_K(s;\alpha);\alpha').
\end{aligned}$$

Property 3.7 *Suppose the positive self-generalized rv $K(\alpha)$ has the LT $\phi_K(s;\alpha)$. $K(\alpha')$ is from the same family but with parameter value α'. Let $\{J_K(t); t \geq 0\}$ be a process with stationary and independent increments, and assume that*

$$\phi_{J_K(t)}(s) = \mathbf{E}[e^{-sJ_K(t)}] = \phi_K^t(s;\alpha), \qquad t \geq 0.$$

Also suppose that $K(\alpha')$ is independent of the process $\{J_K(t); t \geq 0\}$. Then $J(K(\alpha'))$ has the LT $\phi_K(-\log\phi_K(s;\alpha);\alpha')$.

Proof:

$$\begin{aligned}
\mathbf{E}\left\{e^{-sJ(K(\alpha'))}\right\} &= \mathbf{E}\left[\mathbf{E}\left\{e^{-sJ(K(\alpha'))}\middle|K(\alpha')\right\}\right] = \mathbf{E}\left[\phi_K^{K(\alpha')}(s;\alpha)\right] \\
&= \mathbf{E}\left[e^{-(\log\phi_K(s;\alpha))K(\alpha')}\right] = \phi_K(-\log\phi_K(s;\alpha);\alpha').
\end{aligned}$$

Since the rv of self-generalized distribution can be decomposed as sum of any number of iid rv's from the same distribution (see Property 3.6 and 3.7), it arises an interesting question: is the self-generalized distribution ID? We give a brief conclusion here.

Suppose K has the self-generalized distribution. If K is positive-valued, according to Property 3.7, $\phi_K^t(s;\alpha)$ is a LT for any $t \geq 0$. Thus, K is ID. If K is non-negative integer-valued, by Property 3.6, $G_K^n(s;\alpha)$ is a pgf for $n = 1, 2, \ldots$. However, it is not clear whether this is true for $0 < n < 1$ or $n > 0$. Hence, it may or may not be ID. For example, K from **I1** is obviously not ID. It is also a boundary case in **I2**, **I4** and **I5**. Thus, we know that at least some members in **I2**, **I4** and **I5** are not ID. There could exist ID members in these classes; their ID features can be verified by Theorem 2.2.6, or the absolute monotonicity of $M_K'(s;\alpha)/M_K(s;\alpha)$.

A few more properties of the self-generalized distribution are given below.

Property 3.8 *Let K be a self-generalized rv. Suppose Λ is $(0,1]$ or $(0,\infty)$.*

(1) Discrete case. Suppose G_K is left differentiable in α at 1. Let $H(s) = \frac{\partial G_K}{\partial \alpha}(s;\alpha)\Big|_{\alpha=1}$. If $H(s) < 0$ for $0 < s < 1$, then $G_K(s;\alpha)$ is decreasing in α for all $0 < s < 1$. Similarly if $H(s) > 0$ for $0 < s < 1$, then $G_K(s;\alpha)$ is increasing in α for all $0 < s < 1$.

(2) Positive case. Suppose ϕ_K is left differentiable in α at 1. Let $H(s) = \frac{-\partial \log \phi_K}{\partial \alpha}(s;\alpha)\Big|_{\alpha=1}$. If $H(s) > 0$ for $s > 0$, then $\phi_K(s;\alpha)$ is decreasing in α for all $s > 0$.

Proof:

(1) Fix $0 < s < 1$. $H(s) < 0$ implies that $G_K(s;\beta) > s$ for all $\beta_s < \beta < 1$ for some $\beta_s > 0$. Let $\alpha' < \alpha$. There exists a positive integer m and $\beta_s < \beta < 1$ such that $\alpha' = \alpha\beta^m$. Note that $G_K(s;\delta)$ is increasing function of s,

$$G_K(s;\delta\beta) = G_K(G_K(s;\beta);\delta) > G_K(s;\delta), \quad 0 < \delta < 1.$$

Hence by induction $G_K(s;\alpha') > G_K(s;\alpha)$ or $G_K(s;\alpha)$ is decreasing in α.

(2) Fix $s > 0$. $H(s) > 0$ implies that $-\log \phi_K(s;\beta) < s$ for all $\beta_s < \beta < 1$ for some $\beta_s > 0$. Note that

$$\phi_K(s;\delta\beta) = \phi_K(-\log \phi_K(s;\beta);\delta) > \phi_K(s;\delta), \quad 0 < \delta < 1.$$

The completion of the proof is like case (1).

Property 3.9 *Let K be a self-generalized rv.*

(1) Discrete case. Suppose $G_K(s;\alpha)$ is decreasing in $\alpha \in (0,1]$ for $0 < s < 1$. Then $\mathbf{E}[K(\alpha)] \leq 1$.

(2) Positive case. Suppose $\phi_K(s;\alpha)$ is decreasing in $\alpha \in (0,1]$ for $s > 0$. Then $\mathbf{E}[K(\alpha)] \leq 1$.

Proof:

(1) Since $G_K(s;1) = s$, the supposition implies $G_K(s;\alpha) \geq s$. Hence

$$\frac{G_K(1;\alpha) - G_K(s;\alpha)}{1-s} \leq \frac{1-s}{1-s} = 1.$$

Take a limit as $s \uparrow 1$ to get the conclusion.

(2) The supposition implies $\phi_K(s;\alpha) \geq e^{-s}$. Hence

$$\frac{\phi_K(0;\alpha) - \phi_K(s;\alpha)}{s} \leq \frac{1-e^{-s}}{s}.$$

Take a limit as $s \downarrow 0$ to get

$$E[K(\alpha)] = -\phi_K'(0;\alpha) \leq 1.$$

3.3 Construction of new self-generalized distributions

Exploring new self-generalized distributions is quite meaningful and challenging. In this section, we summarize some approaches leading to new self-generalized distributions, and conclude with results/conjectures regarding the relationship between non-negative integer self-generalized and positive self-generalized distributions.

A function $g: \Re \times \Re \to \Re$, $g(x;y)$, satisfying

$$g(g(x;y);y') = g(x;yy'),$$

is called a self-generalized function. We can search for non-negative integer self-generalized distributions in the family of self-generalized functions. If a self-generalized function $G(s;\alpha)$ is a pgf in s, then it is the pgf of a self-generalized distribution. With this idea, we have the following results.

Theorem 3.3.1 *Suppose $g_1(x)$ is a monotone real-valued function, and its inverse g_1^{-1} exists. Let $g_2(x;y)$ be a self-generalized function. Then*

$$g(x;y) = g_1^{-1}\left(g_2\left(g_1(x);y\right)\right)$$

is another self-generalized function.

Proof: A direct calculation shows

$$\begin{aligned} g(g(x;y_1);y_2) &= g_1^{-1}\left(g_2\left(g_1(g(x;y_1);y_2\right)\right) \\ &= g_1^{-1}\left(g_2\left(g_1(g_1^{-1}\left(g_2\left(g_1(x);y_1\right)\right);y_2\right)\right) \end{aligned}$$

Table 3.1: *Some results from Theorem 3.3.1.*

$g_1(s)$	$g_1^{-1}(t)$	$g_2(s;\alpha)$	$g_1^{-1}\left(g_2\left(g_1(x);y\right)\right)$
$\frac{1-\gamma}{1-\gamma s}$	$\frac{1}{\gamma}-\frac{1-\gamma}{\gamma}t^{-1}$	$(1-\alpha)+\alpha s$	$\frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}$
$-\theta^{-1}\log[1-(1-e^{-\theta})s]$	$\frac{1-e^{-\theta t}}{1-e^{-\theta}}$	$(1-\alpha)+\alpha s$	$\frac{1-e^{-\theta(1-\alpha)}[1-(1-e^{-\theta})s]^{\alpha}}{1-e^{-\theta}}$
$1-(1-s)^{1/\theta}$	$1-(1-t)^{\theta}$	$\frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}$	$1-\frac{\alpha^{\theta}(1-\gamma)^{\theta}}{[(1-\alpha)\gamma+(1-\gamma)(1-s)^{-1/\theta}]^{\theta}}$

$$
\begin{aligned}
&= g_1^{-1}\left(g_2\left(g_2\left(g_1(x);y_1\right);y_2\right)\right) \\
&= g_1^{-1}\left(g_2\left(g_1(x);y_1y_2\right)\right) \\
&= g(x;y_1y_2).
\end{aligned}
$$

Hence, $g(x;y)$ is a self-generalized function.

Certainly, g_1 can be chosen as a pgf, and g_2 a self-generalized pgf. Examples given by this approach can be found in Examples 3.2, 3.4, 3.5 illustrated in Table 3.1.

Theorem 3.3.2 *Suppose $g_1(x;y)$ is a self-generalized function. Then*

(1) $g(x;y)=\left(g_1(x^{-1};y)\right)^{-1}$ is a self-generalized function.

(2) $g(x;y)=1-g_1(1-x;y)$ is a self-generalized function.

Proof: We verify their self-generalizability by direct calculation.

(1)

$$
\begin{aligned}
g(g(x;y_1);y_2) &= \left(g_1\left(\frac{1}{g(x;y_1)};y_2\right)\right)^{-1} = \left(g_1\left(g_1(x^{-1};y_1);y_2\right)\right)^{-1} \\
&= \left(g_1(x^{-1};y_1y_2\right)^{-1} = g(x;y_1y_2).
\end{aligned}
$$

(2)

$$
\begin{aligned}
g(g(x;y_1);y_2) &= 1-g_1\left(1-g(x;y_1);y_2\right) = 1-g_1\left(g_1(x;y_1);y_2\right) \\
&= 1-g_1(x;y_1y_2) = g(x;y_1y_2).
\end{aligned}
$$

Next we study the analogues between non-negative integer self-generalized and positive self-generalized distributions. This extends an idea of McKenzie [1986]. The following result describes analogous features between these two kinds of self-generalized distributions.

Theorem 3.3.3 *Suppose $G_K(s;\alpha)$ is the pgf of a non-negative integer self-generalized rv K. Suppose $G_K(\cdot;\alpha)$ can be extended to domain $(-\infty,1]$ with self-generalizability $G_K(G_K(s;\alpha);\alpha') = G_K(s;\alpha\alpha')$ for all $0<\alpha,\alpha'<1$. Let*

$$\phi(s;\alpha) = \exp\{G_K(1-s;\alpha)-1\}, \qquad s>0.$$

If $\phi(s;\alpha)$ is a LT, then it is the LT of a positive self-generalized distribution.

Proof: We need to check the self-generalizability of $\phi(s;\alpha)$. By definition, it follows that

$$\log\phi(s;\alpha) = G_K(1-s;\alpha)-1.$$

Thus,

$$\begin{aligned}
\phi\left(-\log\phi(s;\alpha);\alpha'\right) &= \phi\left(1-G_K(1-s;\alpha);\alpha'\right)\\
&= \exp\left\{G_K\left(1-[1-G_K(1-s;\alpha)];\alpha'\right)-1\right\}\\
&= \exp\left\{G_K\left(G_K(1-s;\alpha);\alpha'\right)-1\right\}\\
&= \exp\left\{G_K(1-s;\alpha\alpha')-1\right\}\\
&= \phi(s;\alpha\alpha').
\end{aligned}$$

Examples 3.1 to 3.5 are just the analogues of Examples 3.6 to 3.10 respectively. The resulting positive self-generalized rv denoted as K' has expectation and variance:

$$\begin{aligned}
\mathbf{E}(K') &= -\phi'(0;\alpha) \;=\; G_K'(1;\alpha) \;=\; \mathbf{E}(K),\\
\mathbf{Var}(K') &= \phi''(0;\alpha)-\left(\phi'(0;\alpha)\right)^2 \;=\; G_K''(1;\alpha)+\left(G_K'(1;\alpha)\right)^2-\left(-G_K'(1;\alpha)\right)^2\\
&= G_K''(1;\alpha) \;=\; \mathbf{Var}(K)+\mathbf{E}^2(K)-\mathbf{E}(K).
\end{aligned}$$

Furthermore, we have the following open questions.

Conjecture 1: If $G(s;\alpha)$ is the pgf of a non-negative integer self-generalized distribution, then

$$\phi(s;\alpha) = \exp\{G(1-s;\alpha) - 1\}$$

is the LT of a positive self-generalized distribution.

To show $\phi(s;\alpha)$ is a LT, we need to:

(1) extend the range of s in $G(1-s;\alpha)$ from $0 \le s \le 1$ to $s \ge 0$,

(2) prove the completely monotone property of $\phi(s;\alpha)$.

For (1), it is equivalent to extend the domain of s in pgf $G(s;\alpha)$ from $0 \le s \le 1$ to $-\infty < s \le 1$. This is fine for the interval $-1 \le s < 0$. As to (2), the completely monotone property holds for $0 \le s \le 1$, but this is not clear for $s > 1$.

However, there is no need for domain extension if we define the pgf for the discrete analogue by the LT of a positive self-generalized distribution. Thus, under minor conditions, the counterpart of Conjecture 1 holds. This leads to the following theorem.

Theorem 3.3.4 *Let $\phi_K(s;\alpha)$ be the LT of a positive self-generalized distribution. Define*

$$G(s;\alpha) = \log \phi_K(1-s;\alpha) + 1, \qquad 0 \le s \le 1,\ 0 < \alpha \le 1.$$

Suppose (1) $\phi_K(1;\alpha) \ge e^{-1}$, $0 < \alpha \le 1$, and (2) $G(s;\alpha)$ has a Taylor series expansion in s. Then $G(s;\alpha)$ is the pgf of a non-negative integer self-generalized distribution.

Proof: First we check the self-generalized condition:

$$\begin{aligned} G(G(s;\alpha);\alpha') &= \log \phi_K(1-G(s;\alpha);\alpha') + 1 = \log \phi_K(-\log \phi_K(1-s;\alpha);\alpha') + 1 \\ &= \log \phi_K(1-s;\alpha\alpha') + 1 = G(s;\alpha\alpha'). \end{aligned}$$

Next note that $G(s;\alpha) = \log \phi_K(1-s;\alpha) + 1$ is increasing in s, $G(0;\alpha) = \log \phi_K(1;\alpha) + 1 \ge -1 + 1 = 0$, $G(1;\alpha) = 1$.

On the other hand, $G(s;\alpha) = G_{K'}(s;\alpha) = \mathbf{E}\left(s^{K'(\alpha)}\right)$ is a proper pgf iff

$$\phi_{K'}(s;\alpha) = G_{K'}(e^{-s};\alpha), \qquad s \ge 0,$$

is a proper LT. We will show that $\phi_{K'}(s;\alpha)$ is completely monotone. Note that

$$\phi_{K'}(s;\alpha) = \log \phi_K(1 - e^{-s};\alpha) + 1. \tag{3.3.1}$$

K infinitely divisible implies that (by Theorem 2.2.6) the derivatives of $\chi(s) = -\log \phi_K(s;\alpha)$ alternate in sign, that is, $(-1)^{j-1}\chi^j(s) \geq 0$. Then $\omega(s) = \omega(s;\alpha) = \chi(1 - e^{-s};\alpha)$ has the same property:

$$\omega'(s) = \chi'(1 - e^{-s})\, e^{-s} \geq 0, \qquad \omega''(s) = \chi''(1 - e^{-s})\, e^{-2s} - \chi'(1 - e^{-s})\, e^{-s} \leq 0,$$

and the derivatives of each term of the form $\chi^{(j)}(1 - e^{-s})\, e^{-ms}$ will continue to be opposite in sign to the original term. Hence $\phi_{K'}(s;\alpha)$, given in (3.3.1), is completely monotone.

Finally, $\phi_{K'}(s;\alpha)$ is the LT of an nonnegative integer-valued rv, if $G_{K'}(s;\alpha)$ has a Taylor series expansion. Because of condition (2), we know that $\phi_{K'}(s;\alpha)$ is a LT, and consequently $G(s;\alpha)$ is a pgf. This completes the proof.

Theorem 3.3.3 and Theorem 3.3.4 disclose the relationship between a self-generalized operator for positive reals and one for non-negative integers.

3.4 Extended-thinning operation

In this section, we propose an extended-thinning operation which is one of the essentials to the model construction of continuous-time stochastic processes with given univariate margins. This extends binomial thinning (see (2) in Proposition 2.2.2). In fact, we hinted at this topic in Section 2.3, where we studied the stochastic representations of the compound self-generalized pgf and logarithm-compound self-generalized LT.

Now we study the stochastic operation between two independent rv's X and K, which have LT's $\phi_X(s)$ and $\phi_K(s)$ respectively. We wish to define the operation: $K \circledast X$, such that its LT has the form

$$\phi_{K \circledast X}(s) = \mathbf{E}\left[e^{-s(K \circledast X)}\right] = \phi_X\left(-\log \phi_K(s)\right).$$

We shall give the stochastic representation of this definition in three cases where X is non-negative integer-valued, positive-valued, and real-valued respectively.

Case 1: X is a non-negative integer rv. Define a discrete-time process $\{J_K(t); t = 0, 1, 2, \ldots\}$ independent of X as

$$J_K(t) = \sum_{i=0}^{t} K_i,$$

where $K_0 = 0,\ K_1, \ldots, K_i, \ldots$ are iid with LT: $\phi_{K_i}(s) = \phi_K(s)$. Let

$$K \circledast X = J_K(X) = \sum_{i=0}^{X} K_i,$$

the random summation over the process $\{J_K(t); t = 0, 1, 2, \ldots\}$.

Direct calculations show that

$$\begin{aligned} \phi_{K\circledast X}(s) &\doteq \mathbf{E}\left[e^{-s(K\circledast X)}\right] = \mathbf{E}\left[\mathbf{E}\left(e^{-s\sum_{i=0}^{X} K_i}\middle| X\right)\right] = \mathbf{E}\left[\phi_K^X(s)\right] \\ &= \mathbf{E}\left[e^{-(-\log\phi_K(s))X}\right] = \phi_X\left(-\log\phi_K(s)\right). \end{aligned}$$

The illustration can be seen in (a) of Figure 3.1.

Case 2: X is a positive rv. Consider a continuous-time process $\{J_K(t); t \geq 0\}$ independent of X which has stationary and independent increments, such that the LT of $J_K(t)$ is:

$$\phi_{J_K(t)}(s) = \phi_K^t(s), \quad t \geq 0.$$

Define

$$K \circledast X = J_K(X).$$

Then

$$\begin{aligned} \phi_{K\circledast X}(s) &= \mathbf{E}\left[e^{-s(K\circledast X)}\right] = \mathbf{E}\left[\mathbf{E}\left(e^{-s(K\circledast X)}\middle| X\right)\right] = \mathbf{E}\left[\mathbf{E}\left(e^{-sJ_K(X)}\middle| X\right)\right] \\ &= \mathbf{E}\left[\phi_K^X(s)\right] = \mathbf{E}\left[e^{-(-\log\phi_K(s))X}\right] = \phi_X\left(-\log\phi_K(s)\right). \end{aligned}$$

See (b) in Figure 3.1 for the illustration.

One example of the defined process $\{J_K(t); t \geq 0\}$ is that in the family of Lévy processes with LT

$$\phi_{J_K(t)}(s) = \exp\left\{t\left[-as + \int_0^\infty (e^{-sy} - 1)L(dy)\right]\right\},$$

where $L(\cdot)$ is the Lévy measure. Certainly, in this case, the LT of the K is

$$\exp\left\{\left[-as + \int_0^\infty (e^{-sy} - 1)L(dy)\right]\right\}.$$

Case 3: X is a real rv. Consider two stationary independent increment processes $\{J_1(t); t \geq 0\}$ and $\{J_2(t); t \geq 0\}$ independent of X with LT

$$\phi_{J_1(t)}(s) = \phi_K^t(s) \quad \text{and} \quad \phi_{J_2(t)}(s) = [\phi_K(s)]^{-t}$$

respectively. Note that $\phi_{J_2(t)}(s)$ is the reciprocal of $\phi_{J_1(t)}(s)$. Hence, under the requirements of a LT, $\phi_K(s)$ can not be arbitrary. Construct a new process over the whole real axis $\{J_K(t); t \in (-\infty, +\infty)\}$ such that

$$J_K(t) = \begin{cases} J_1(t), & \text{if } t \geq 0; \\ J_2(|t|), & \text{if } t < 0. \end{cases}$$

For this new process, the LT of $J_K(t)$ is

$$\begin{aligned} \phi_{J_K(t)}(s) &= \mathbf{E}\left[e^{-sJ_K(t)}\right] = \begin{cases} \phi_K^t(s), & t \geq 0; \\ \phi_K^{-|t|}(s), & t < 0; \end{cases} \\ &= \phi_K^t(s). \end{aligned}$$

Define

$$K \circledast X = J_K(X).$$

Then

$$\begin{aligned} \phi_{K \circledast X}(s) &= \mathbf{E}\left[e^{-s(K \circledast X)}\right] = \mathbf{E}\left[\mathbf{E}\left(e^{-sJ_K(X)} \middle| X\right)\right] \\ &= \mathbf{E}\left[\phi_K^X(s)\right] = \phi_X\left(-\log \phi_K(s)\right). \end{aligned}$$

See the illustration in Figure 3.2.

In Case 3, process $\{J_2(t); t \geq 0\}$ is in fact artificially developed by process $\{J_1(t); t \geq 0\}$ for $\phi_{J_2(t)}(s) = [\phi_{J_1(t)}(s)]^{-1}$. Hence, $J_1(t)$ and $J_2(t)$ can not both be positive, because $\phi_{J_1(t)}(s)$ and $\phi_{J_2(t)}(s)$ both be bounded above by 1 and can't satisfy the completely monotone property at the same time. Suppose we restrict the process $\{J_1(t); t \geq 0\}$ to the Lévy process family, with

$$\phi_{J_1(t)}(s) = \exp\left\{t\left[-as + \int_0^\infty (e^{-sy} - 1)L(dy)\right]\right\},$$

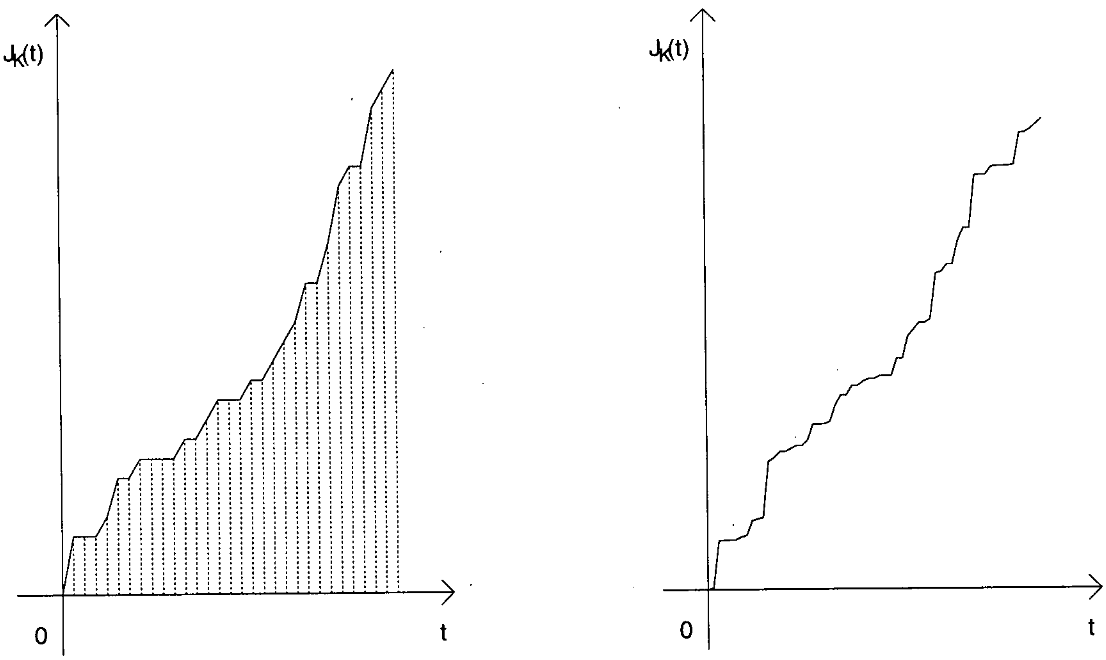


Figure 3.1: *Illustration of* $\{J_K(t); -\infty < t < \infty\}$ *in Cases 1 and 2. (a) corresponds to non-negative integer* X *in Case 1, where dotted vertical lines indicate the discrete time points* $\{0, 1, 2, \ldots\}$. *(b) corresponds to positive* X *in Case 2, where* t *is continuous on* $[0, \infty)$.

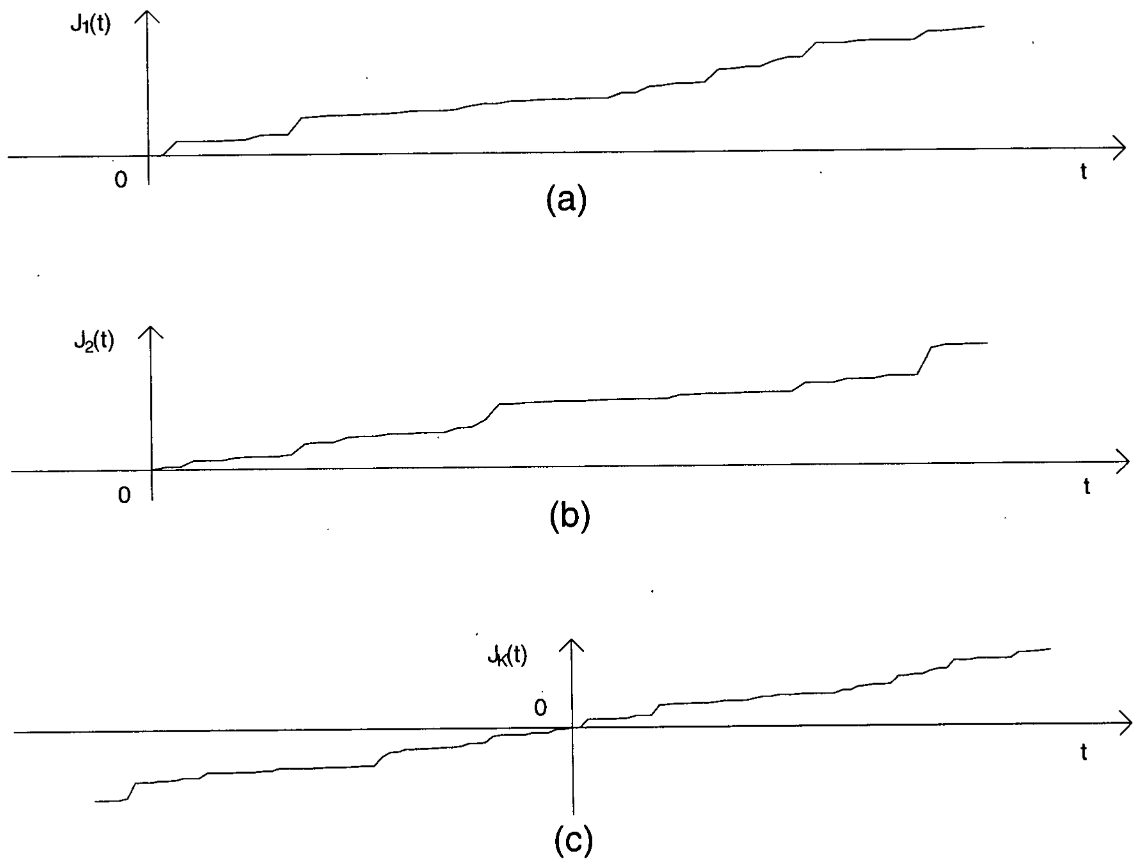


Figure 3.2: *Illustration of* $\{J_1(t); t \geq 0\}$, $\{J_1(t); t \geq 0\}$ *and* $\{J_k(t); -\infty < t < \infty\}$ *in Case 3.*

where $L(\cdot)$ is the Lévy measure, non-negative and satisfies $\int_0^\infty \min(1,y)L(dy) < \infty$. Also assume that the process $\{J_2(t); t \geq 0\}$ is a Lévy process. Then

$$\phi_{J_2(t)}(s) = \left[\phi_{J_1(t)}(s)\right]^{-1} = \exp\left\{t\left[as + \int_0^\infty (e^{-sy}-1)(-L)(dy)\right]\right\}.$$

Since $-L(dy) \leq 0$, the only possible choice is $L(dy) = 0$. This implies

$$\phi_{J_1(t)}(s) = e^{-ats}, \quad \text{and} \quad \phi_{J_2(t)}(s) = e^{ats},$$

for some constant a, i.e., $J_1(t) = at$, and $J_2(t) = -at$, degenerate at points at and $-at$ respectively.

In summary, we propose the extended-thinning operation as below.

Definition 3.3 *Suppose $\{J_K(t); t \in T_0\}$ is an appropriate stationary independent increment process constructed via rv K such that*

$$\phi_{J_K(t)}(s) = \phi_K^t(s),$$

where T_0 could be $\{0,1,2,\ldots\}$, or $[0,\infty)$ or $(-\infty,+\infty)$ (refer to cases 1, 2 and 3). The extended-thinning operation is defined as a stochastic operation between J_K and X with X independent of $\{J_K(t)\}$,

$$K \circledast X = J_K(X).$$

Such an operation results in a rv with LT

$$\phi_{K\circledast X}(s) = \phi_X\left(-\log \phi_K(s)\right).$$

The notation $K\circledast$ means an independent copy of rv K, which has the same distribution as that of K. Hence, notations $K \circledast X$ and $K \circledast Y$ do not means the rv K in both is the same, but their distributions are the same.

Note that K and X may not be arbitrary random variables, restrictions on them in the different domain cases should be imposed. In other words, $\phi_X\left(-\log \phi_K(s)\right)$ being a LT requires conditions on ϕ_K and ϕ_X.

We can calculate the cf of $K \circledast X$:

$$\begin{aligned}\varphi_{K\circledast X}(s) &\doteq \mathbf{E}\left[e^{isK\circledast X}\right] = \phi_{K\circledast X}(is) = \phi_X\left(-\log \phi_K(is)\right) \\ &= \begin{cases} \phi_X\left(-\log \varphi_K(s)\right), & \text{if } X \text{ is non-negative;} \\ \varphi_X\left(i \log \varphi_K(s)\right), & \text{if } X \text{ is real.} \end{cases}\end{aligned}$$

A natural property regarding expectation is given next.

Property 3.10 *If $\mathbf{E}[K(\alpha)]$ is finite and continuous with respect to α, i.e., $\mathbf{E}[K(\alpha)] = \alpha^r$, then*

$$\mathbf{E}[K(\alpha) \circledast X] = \mathbf{E}[K(\alpha)] \cdot \mathbf{E}[X] = \alpha^r \mathbf{E}[X].$$

Proof: This can be readily derived by taking the derivative of the LT.

Hence, $\mathbf{E}[K(\alpha) \circledast X] \leq \mathbf{E}[X]$ if α is within $[0,1]$ and $r > 0$. In general, the extended-thinning operation rescales the expectation.

Now we look into the examples of extended-thinning operation in statistical practice.

Example 3.12 *The well-known binomial-thinning is one special case of extended-thinning operation, for*

$$\alpha * X = \sum_{i=0}^{X} K_i, \quad K_0 = 0, \; K_1, \; \ldots, \; K_i, \; \ldots \overset{i.i.d.}{\sim} \mathit{Bernoulli}(\alpha).$$

One special feature of binomial-thinning is that $\alpha * X \leq X$, which means $\alpha * X$ does become "thinner" than X almost surely. However, in general, the extended-thinning may not retain this feature, but the expectation is "thinner" than $\mathbf{E}[X]$ if we restrict the domain of parameter α to $[0,1]$.

Example 3.13 *A branching process has an operation similar to the extended-thinning operation. This kind of processes $\{X(n) : n = 0, 1, 2, \ldots\}$ is defined as*

$$X(n+1) = \sum_{i=0}^{X(n)} Z_i, \quad Z_0 = 0, \; Z_1, \; \ldots, \; Z_i, \; \ldots \; \mathit{iid}.$$

$X(n)$ is the size of the n^{th} generation.

Example 3.14 *The product of a constant α with a rv X, αX, is another example. This may not be straightforward at the first glance. However, we can check its LT. In this case, we can view α as a rv degenerate at point α. Hence, it has LT: $\phi_\alpha(s) = e^{-\alpha s}$. The LT of αX, then, is*

$$\phi_{\alpha X}(s) = \mathbf{E}\left(e^{-s\alpha X}\right) = \phi_X(\alpha s) = \phi_X\left(-\log\phi_\alpha(s)\right).$$

Therefore, αX is an extended-thinning operation.

We use the notation '$\circledast$' for the extended-thinning operation based on the consideration to unite the constant multiplier '$\bullet$' and the binomial-thinning operation '$*$', in a simple expression, namely,

$$(\bullet) \quad \cup \quad (*) \quad \Longrightarrow \quad (\circledast).$$

Following are two properties of extended-thinning operations; these are very important to the construction of the continuous-time generalized AR(1) processes in Chapter 4.

Property 3.11 (Distributive law) *Suppose K is a self-generalized rv, and let X and Y be independent rv's. Then*

$$K \circledast (X+Y) \stackrel{d}{=} K \circledast X + K \circledast Y.$$

Proof: Since X and Y are independent, we have

$$\phi_{K\circledast(X+Y)}(s) = \phi_{X+Y}\left(-\log\phi_K(s)\right) = \phi_X\left(-\log\phi_K(s)\right)\phi_Y\left(-\log\phi_K(s)\right).$$

This implies that the distributive law holds.

Property 3.12 (Associative law) *Let K_1, K_2 be two different self-generalized rv's acting as operators. Then*

$$K_1 \circledast (K_2 \circledast X) \stackrel{d}{=} (K_1 \circledast K_2) \circledast X.$$

Proof: Direct calculation shows

$$\begin{aligned} \phi_{K_1\circledast(K_2\circledast X)}(s) &= \phi_{K_2\circledast X}\left(-\log\phi_{K_1}(s)\right) = \phi_X\left(-\log\phi_{K_2}\left(-\log\phi_{K_1}(s)\right)\right) \\ &= \phi_X\left(-\log\phi_{K_1\circledast K_2}(s)\right) = \phi_{(K_1\circledast K_2)\circledast X}(s). \end{aligned}$$

Hence, the associative law holds.

Recalling self-generalizability, we find that it is closed under the extended-thinning operation, i.e.,

$$K(\alpha) \circledast K(\alpha') \stackrel{d}{=} K(\alpha\alpha').$$

To keep symbolic consistency with constant multiplier $\bullet$ and binomial-thinning operator $*$, we rewrite the notation for the self-generalized rv $K(\alpha)$ in extended-thinning operation as α_K such that

$$(\alpha)_K \circledast X \stackrel{\text{def}}{=} K(\alpha) \circledast X.$$

This change makes the extended-thinning operation with a self-generalized rv looks like constant multiplier or binomial-thinning operator. For instance, we can rewrite the closure property of self-generalizability via new notation:

$$(\alpha)_K \circledast (\alpha')_K \stackrel{d}{=} (\alpha\alpha')_K.$$

But remember $(\alpha)_K$ is a rv, not a parameter, and it is valid only with the extended-thinning operator $\circledast$. The reason we impose the subscript K on α is to try to avoid the misunderstanding of $(\alpha)_K$, a rv, to the parameter α, and the convention of binomial-thinning operation as well. This new notation will benefit us immediately with the following commutative law.

Property 3.13 (Commutative law) $\quad (\alpha)_K \circledast (\alpha')_K = (\alpha')_K \circledast (\alpha)_K.$

Proof: This is simply because

$$(\alpha)_K \circledast (\alpha')_K \stackrel{d}{=} (\alpha\alpha')_K = (\alpha'\alpha)_K(\alpha')_K \circledast (\alpha)_K.$$

Note that the commutative law only holds for two self-generalized rv's from the same family.

Property 3.14 (Weak convergence) *Let $\alpha_n \to \alpha$, where $\alpha_n \in A$ for all n, and $\alpha \in A$. If $G_K(s;\alpha)$ or $\phi_K(s;\alpha)$ is continuous with respect to α, then*

$$(\alpha_n)_K \circledast X \xrightarrow{\text{L}} (\alpha)_K \circledast X.$$

Proof: This follows from the continuity of G_K or ϕ_k in α.

Suppose Λ, the domain of the parameter α, is the interval $[0,1]$. Then on boundaries, extended-operation behaviors like the constant multiplier. Here we assume boundary 0 is included in Λ.

Property 3.15 $\quad (0)_K \circledast X \overset{d}{=} 0 \quad$ *and* $\quad (1)_K \circledast X \overset{d}{=} X.$

Proof: By Properties 3.1 and 3.3, we have

$$(0)_K = K(0) \overset{d}{=} 0 \qquad \text{and} \qquad (1)_K = K(1) \overset{d}{=} 1.$$

Hence, the resulting process $\{J_K(t); t \geq 0\}$ has margins

$$J_{K(0)}(t) = 0, \qquad \text{and} \qquad J_{K(1)}(t) = t, \qquad \text{for } t \geq 0.$$

Thus,

$$J_{K(0)}(X) \overset{d}{=} 0, \qquad J_{K(1)}(X) \overset{d}{=} X.$$

This completes the proof.

Remark: If the boundary 0 is excluded from Λ, but the expectation of the self-generalized rv K is bounded and continuous in α, then $\lim_{\alpha \to 0^+} (\alpha)_K \circledast X = (0)_K \circledast X = 0$. This is because of (2) of Property 3.3.

Lastly, we discuss the variable type of $K \circledast X$: non-negative integer, positive, or real. This is basically determined by the variable type of K, not the variable type of X. Correspondingly, we study its pgf, or LT, or cf. Recall $K \circledast X = J_K(X)$. We have:

(1) if K is non-negative integer, then $\{J_K(t); t = 0, 1, 2, \ldots\}$ or $\{J_K(t); t \geq 0\}$ is a process with non-negative integer increments, so $J_K(X)$ is non-negative integer no matter if X is non-negative integer or positive real. And it follows that

$$G_{K \circledast X}(s) = \mathbf{E}\left\{s^{J_K(X)}\right\} = \begin{cases} G_X(G_K(s)) & \text{if } X \text{ is non-negative integer;} \\ \phi_X(-\log G_K(s)) & \text{if } X \text{ is positive,} \end{cases}$$

(2) if K is positive, then $\{J_K(t); t = 0, 1, 2, \ldots.\}$ or $\{J_K(t); t \geq 0\}$ is a process with positive increments, so $J_K(X)$ is non-negative no matter if X is non-negative integer or positive real. And it follows that

$$\phi_{K \circledast X}(s) = \mathbf{E}\left\{e^{sJ_K(X)}\right\} = \begin{cases} G_X(\phi_K(s)) & \text{if } X \text{ is non-negative integer;} \\ \phi_X(-\log \phi_K(s)) & \text{if } X \text{ is positive,} \end{cases}$$

(3) if K is degenerate, namely being a real number α, then

$$K \circledast X = \alpha X,$$

and $\{J_K(t); t \geq 0\}$ is a process with real-valued increments, so $J_K(X)$ is real no matter if X is positive or real. And it follows that

$$\varphi_{K \circledast X}(s) = \mathbf{E}\left\{e^{is(\alpha X)}\right\} = \begin{cases} \phi_X(i\alpha s) & \text{if } X \text{ is positive;} \\ \varphi_X(\alpha s) & \text{if } X \text{ is real,} \end{cases}$$

The extended-thinning operation will be discussed again in Section 4.1, where a geometrical explanation is given.

Chapter 4

Generalized Ornstein–Uhlenbeck stochastic differential equations and their possible solutions

Ornstein-Uhlenbeck stochastic differential equations is a classical topic well discussed in the literature. Its applications can be found in mathematical finance, physics, and so on. Refer to Hsu and Park [1988], Neftci [1996]. In this chapter, we shall propose generalized Ornstein-Uhlenbeck stochastic differential equations, and define the corresponding generalized stochastic integration. These are fundamental techniques and key ideas in the model construction of a class of continuous-time Markov processes given in this chapter and the next chapter.

We start with the introduction to stochastic differentiation and integration in Section 4.1. The generalization of Ornstein-Uhlenbeck equations will be given in Section 4.2, and the explanation and examples are shown in Section 4.3. We construct the solutions for the generalized Ornstein-Uhlenbeck equation in Section 4.4, and summarize the resulting processes in Section 4.5.

4.1 Stochastic differentiation and integration

The dynamic feature of a continuous-time process $\{X(t); t \geq 0\}$ is of interest, as it describes the instantaneous behavior of the process. To address this feature, we need the concept of stochastic differentiation and integration.

In the literature, some scholars explain the concept of stochastic differentiation via stochastic integration, while others illustrate it in terms of the infinitesimal increment of the process. For the beginner, the former is not a direct approach, and furthermore, the definition of stochastic integration requires the concepts of infinitesimal increment. Hence, we shall take the latter approach. However, it may not be very strict in the mathematical sense. We just focus on main ideas. There are many references on this area, such as: Chung and Williams [1990], Lukacs [1968], Øksendal [1995], Protter [1990], etc. A good introductory book, which clearly explains the concepts of SDE and stochastic integration without measure theory is Neftci [1996].

To clearly state the idea, let's recall the concept of differentiation in calculus, where it is defined as the infinitesimal increment. For instance, if $x(t)$ is a function of t, denote by

$$\Delta x(t) = x(t+h) - x(t),$$

the increment of $x(t)$ when the argument changes from t to $t+h$. When h, the argument increment, is infinitesimal, we denote it as dt. Correspondingly, the infinitesimal increment of $x(t)$ is written as $dx(t)$, which can be, expressed as

$$dx(t) = x(t+dt) - x(t).$$

Thus, the differential is the infinitesimal increment.

For a stochastic process $\{X(t); t \geq 0\}$, this well-known idea can be borrowed to define the differentiation of the process.

Definition 4.1 *The differential $dX(t)$ of a continuous-time stochastic process $\{X(t); t \geq 0\}$ is defined as the infinitesimal increment $X(t+h) - X(t)$, where the increment of time h is infinitesimal.*

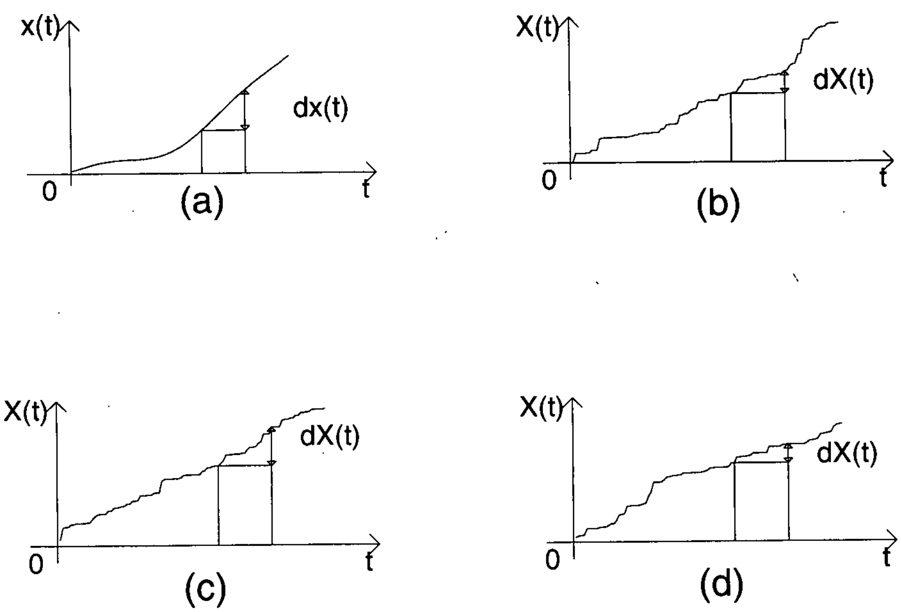


Figure 4.1: *Illustration of increment in the deterministic and stochastic cases. (a) corresponds to deterministic function $x(t)$. (b), (c) and (d) correspond to three different paths of the stochastic process $\{X(t); t \geq 0\}$.*

Note that the increment of a function $x(t)$ is a number, however, the increment of a process is a rv. Hence, the differential $dX(t)$ is understood as a rv, not a number unless in the degenerate case.

Figure 4.1 illustrates the increments in both cases. Note that in (b), (c) and (d) of Figure 4.1, the infinitesimal increment $dX(t)$ are different because they correspond to three different paths of the process $\{X(t); t \geq 0\}$. This clearly shows that $dX(t)$ is a random variable.

The Riemann integration of a function $x(t)$ over $[t_1, t_2]$, the area with sign, is constructed via the infinitesimal partition approach. Here we roughly review its idea. Consider the argument range $[t_1, t_2]$. Divide this interval in n equal pieces, i.e.,

$$[t_1, t_1 + h),\ [t_1 + h, t_1 + 2h),\ \ldots,\ [t_1 + (n-1)h, t_1 + nh] = [t_1 + (n-1)h, t_2],$$

where $h = (t_2 - t_1)/n$. When n goes to infinity, each piece will become an infinitesimal interval. Use a finite Riemann sum over these small intervals

$$\sum_{i=0}^{n-1} x(t_1 + ih)h$$

to approximate the integrated "area". When n goes to ∞, the limit is defined as the integration.

This method was introduced to stochastic integration over a half century ago. However, the difference between common integration and stochastic integration is how to define the limit. In probability theory, the common modes of convergence include in distribution, in probability, in L^1, in L^2, *a.s.*, etc. Hence, different stochastic integrations arise. For example, the Itô integration is the limit in L^2. In our study, the convergence mode that we adopt is "in distribution".

Suppose $\{X(t); t \geq 0\}$ is a continuous-time process. We now define the stochastic integral $\int_{t_1}^{t_2} g(X(t))dX(t)$. Consider $n+1$ equally spaced points

$$t_1, t_1 + h, t_1 + 2h, \ldots, t_1 + (n-1)h, t_1 + nh = t_2$$

over $[t_1, t_2]$, where $h = (t_2 - t_1)/n$. Let

$$S_n = \sum_{i=0}^{n-1} g(X(t + ih))[X(t + (i+1)h) - X(t + ih)].$$

If there exists a rv Y such that

$$S_n \xrightarrow{L} Y, \qquad \text{as } n \to \infty,$$

then Y is defined as $\int_{t_1}^{t_2} g(X(t))dX(t)$. This leads to the following definition.

Definition 4.2 *Let $\{X(t); t \geq 0\}$ be a continuous-time process. Divide $[t_1, t_2]$ into n equally small intervals. Then*

$$\int_{t_1}^{t_2} g(X(t))dX(t) \stackrel{d}{=} \lim_{n\to\infty} \sum_{i=0}^{n-1} g(X(t + ih))[X(t + (i+1)h) - X(t + ih)],$$

where $h = (t_2 - t_1)/n$. The summation on the right hand side converges in distribution.

Note that stochastic integral is still a rv, not a number. Figure 4.2 shows the idea of stochastic integration via the infinitesimal partition approach. Note that the "area" in (a) and (b) may not be the same, because they correspond to two different path of the process $\{X(t); t \geq 0\}$. This clearly indicates that the stochastic integration $\int_{t_1}^{t_2} g(X(t))dX(t)$ is a random variable.

Since the increment, $\Delta X(t) = X(t+h) - X(t)$, of a continuous-time process $\{X(t);\ t \geq 0\}$ can be written as

$$\Delta X(t) = X(t+h) - X(t) = \sum_{i=0}^{n-1} \left[X\left(t + \frac{i+1}{n}h\right) - X\left(t + \frac{i}{n}h\right) \right]$$

for any positive integer n, we can rewrite the increment via a stochastic differential and integral as

$$\Delta X(t) = X(t+h) - X(t) = \int_t^{t+h} dX(s).$$

This kind of expression is used to formally define the stochastic differentiation by many authors.

Look back at the extended-thinning where $K \circledast X = J_K(X)$. When X is a non-negative integer rv, the extended-thinning $K \circledast X = J_K(X) = \sum_{i=0}^{X} K_i$ is a random summation. What will it be if X is a positive or real-valued rv? Note that in these two cases, $\{J_K(t); t \geq 0\}$ is a continuous-time process. Thus,

$$J_K(X) = J_K(X) - J_K(0) = \int_0^X dJ_K(t),$$

a random stochastic integral. Therefore, in principle, the extended-thinning operation is a random summation or a stochastic integration.

Now we make up a geometric explanation for the extended-thinning operation. Let's consider αX, a special case of an extended-thinning operation, as the area of random rectangle with length X (in the horizontal direction) and width α (in the vertical direction). Imagine a random rectangle in this way: the length is a rv. However, the width is not a fixed constant or rv. On every slice orthogonal to the length, the cutting width is a rv. All these cutting widths are iid rv's. One can use the sliced bread to mimic this random rectangle. Because $\int_0^X dJ_K(t)$ is the limit of a Riemann sum, and in small time intervals the increments of process $\{J_K(t); t \geq 0\}$ are iid, a natural explanation of $J_K(X)$ is that it is the limit of sums of areas of rectangles with widths rv X/n and iid heights $K_{n1}, \ldots, K_{nn}$ (with distribution of $J_K(X/n)$). This area is random. Thus, it is a new random rv.

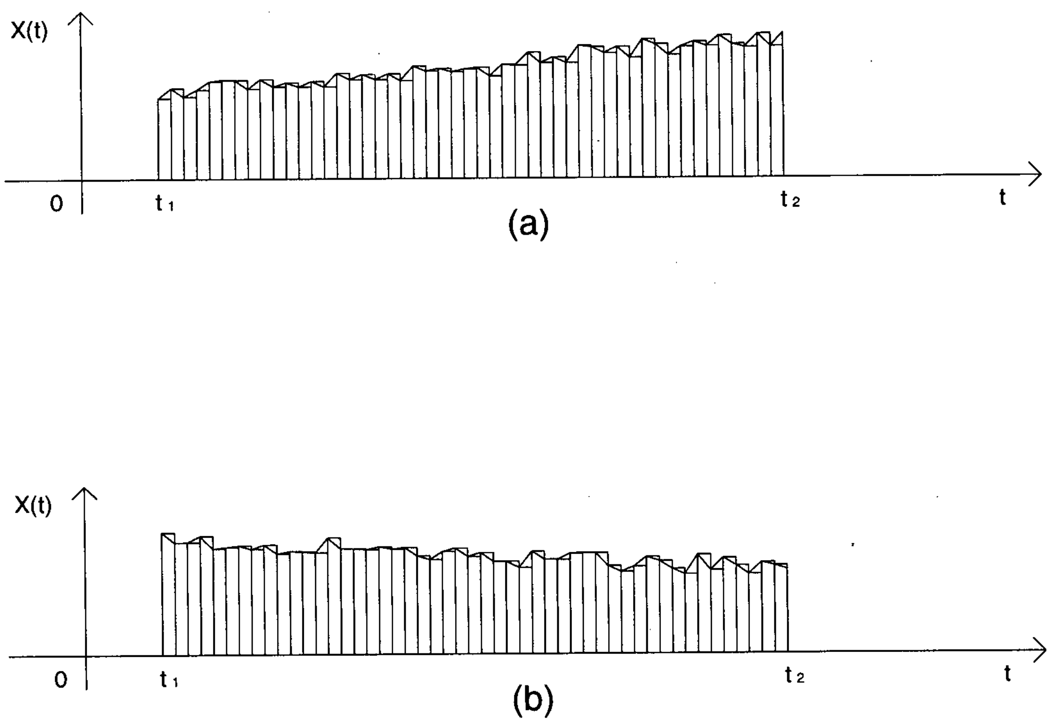


Figure 4.2: *Illustration of stochastic integration via infinitesimal partition. (a) and (b) correspond to two different paths of the stochastic process* $\{X(t); t \geq 0\}$.

Figure 4.3 illustrates the idea of this geometrical explanation of the extended-thinning operation. We give the "imagined" random rectangles in three cases corresponding to X being (a) real, (b) non-negative integer, and (c) positive valued.

Without loss of generality, we assume the mode of convergence for stochastic integration is convergence in distribution throughout the remainder of this thesis.

4.2 Generalized Ornstein-Uhlenbeck equations

Like a differential equation which expresses the dynamic characteristic of a function, a stochastic differential equation describes the dynamic feature of a continuous-time process. However, because the derivative of a process commonly doesn't exist, we can not include the derivative of a process in the equation. Instead, we include the differential of a process into the equation.

Recall the Ornstein-Uhlenbeck process (see Section 2.1), which is defined by the following stochastic differential equation (SDE) for real-valued process rv $X(t)$,

$$dX(t) = -\mu X(t)dt + \sigma dW(t),$$

where $\{W(t); t \geq 0\}$ is a Brownian motion independent of $X(t)$. To keep consistency with the literature, we absorb σ into the innovation process so that it becomes

$$dX(t) = -\mu X(t)dt + dW(t).$$

This SDE shows that the infinitesimal increment of $X(t)$ in the near future depends on the present circumstance and the innovation term. Note that $X(t)$ has support on the range of $(-\infty, +\infty)$.

Replacing the innovation term from a Wiener process (Brownian motion) with a more general Lévy process (Brownian motion is a special process in Lévy process family) leads to the Ornstein-Uhlenbeck-type process (see Barndorff-Nielsen et al. (1998) and references therein), namely,

$$dX(t) = -\mu X(t)dt + dL(t),$$

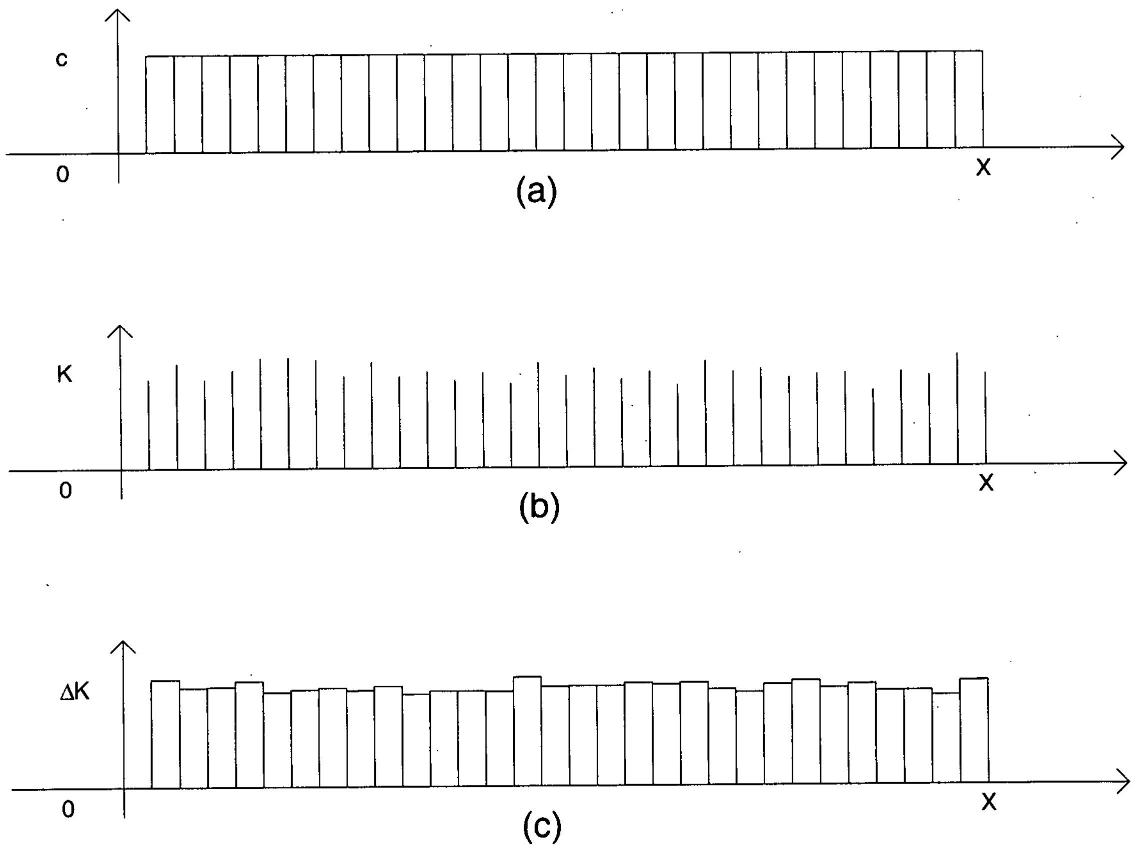


Figure 4.3: *Illustration of the geometrical explanation of the extended-thinning operation. (a) corresponds to a constant multiplier cX; X can be either real or positive-valued. (b) corresponds to a non-negative integer X. (c) corresponds to a positive X.*

where $\{L(t); t \geq 0\}$ is a Lévy process. This offers possible marginal distributions for $X(t)$ with support $[0, \infty)$. However, it doesn't provide any marginal distributions with support on the non-negative integers, because $\mu X(t)dt$ is unlikely to be an integer.

The extension of innovation processes seems to be ideal. It covers distributions with domain on the non-negative integers. But only extending the innovation processes won't help us to construct models with marginal distributions having the non-negative integer support. Therefore, extending the dependence term from a product to a generalized stochastic operation may lead to a successful approach. This inspiration comes from Joe [1996].

Note that the dependence structure of such kind of processes $\{X(t); t \geq 0\}$ is determined by $-\mu X(t)dt$, and an independent innovation process is introduced to explain the fluctuation. Hence, the process is simply governed by the dependence mechanism part, $-\mu X(t)dt$, and the independent input part $\{\epsilon(t); t \geq 0\}$.

Recall that αX is a special operation in the class of extended-thinning operations (see Section 3.4). We can rewrite the dependent mechanism part $-\mu X(t)dt$ as

$$-\mu X(t)dt = -\mu dt X(t) = (1 - \mu dt)X(t) - X(t).$$

Hence, a natural generalization for this dependent mechanism term is

$$K(1 - \mu dt) \circledast X(t) - X(t) = (1 - \mu dt)_K \circledast X(t) - X(t).$$

However, we will restrict K to be within a self-generalized family. Since this new term could be a non-negative integer, or positive, or real rv, we may hopefully obtain marginal distributions with support on non-negative integer, or positive, or real values respectively.

We now formally define the generalized Ornstein-Uhlenbeck SDE below.

Definition 4.3 *Suppose $\{X(t); t \geq 0\}$ is a continuous-time process, and $\{\epsilon(t); t \geq 0\}$ is an innovation IIP. The generalized Ornstein-Uhlenbeck SDE is defined as*

$$dX(t) = [K(1 - \mu dt) \circledast X(t) - X(t)] + d\epsilon(t) = [(1 - \mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t),$$

where $K(\alpha)$ is a self-generalized rv with respect to parameter α.

Because the generalized Ornstein-Uhlenbeck SDE involves an extended-thinning operation, we should name the corresponding stochastic integration as *generalized stochastic integration.*

Definition 4.4 *Let $\{X(t); t \geq 0\}$ be a continuous-time process. Divide $[t_1, t_2]$ into n equally small intervals. Then*

$$\begin{aligned}\int_{t_1}^{t_2} (g(t))_K \circledast dX(t) &= \int_{t_1}^{t_2} K(g(t)) \circledast dX(t) \\ &\stackrel{d}{=} \lim_{n\to\infty} \sum_{i=0}^{n-1} K(g(t+ih)) \circledast [X(t+(i+1)h) - X(t+ih)] \\ &= \lim_{n\to\infty} \sum_{i=0}^{n-1} (g(t+ih))_K \circledast [X(t+(i+1)h) - X(t+ih)],\end{aligned}$$

where $h = (t_2 - t_1)/n$, $g(\cdot)$ is a function with range $[0,1]$, and $K(\alpha)$ is a self-generalized rv with respect to parameter α. This is well defined if the summation on right hand side converges in distribution.

This generalized stochastic integration will be applied to solving generalized Ornstein-Uhlenbeck SDE in Section 4.4.

4.3 Explanations, innovation types, non-stationary situations and examples

We may give a further explanation of the generalized Ornstein-Uhlenbeck SDE in this section.

In the generalized Ornstein-Uhlenbeck SDE

$$dX(t) = [(1 - \mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t),$$

$X(t)$ means present state, while dt, $d\epsilon(t)$ and $dX(t)$ means the infinitesimal increment in the near future infinitesimal time interval. Hence, we can comprehend the generalized Ornstein-Uhlenbeck SDE as a forward expression, not a backward expression.

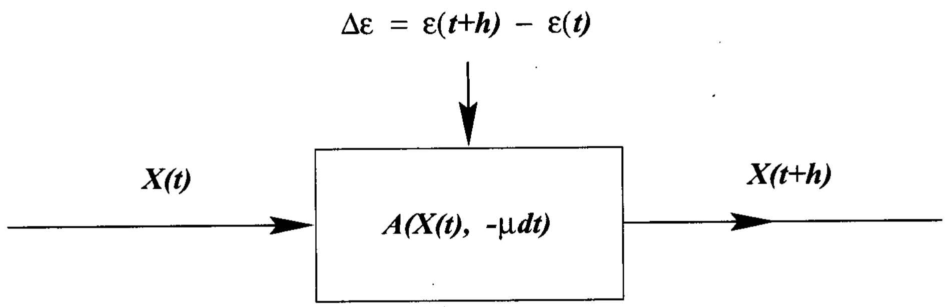


Figure 4.4: *Illustration of the mechanism of the generalized Ornstein-Uhlenbeck SDE.*

With such an understanding, we can write down the difference equation from the generalized Ornstein-Uhlenbeck SDE. It is

$$X(t+h) - X(t) = [(1-\mu h)_K \circledast X(t) - X(t)] + \Delta\epsilon, \quad \Delta\epsilon = \epsilon(t+h) - \epsilon(t),$$

which can be simplified as

$$X(t+h) = (1-\mu h)_K \circledast X(t) + \Delta\epsilon. \tag{4.3.1}$$

Denote $A(X(t), -\mu h) = (1-\mu h)_K \circledast X(t)$, the dependence mechanism. We can further write it as

$$X(t+h) = A(X(t), -\mu h) + \Delta\epsilon.$$

This uncovers the stochastic representation of the process in an infinitesimal time interval. Figure 4.4 roughly shows the mechanism idea of the process generated from the generalized Ornstein-Uhlenbeck SDE when h is infinitesimal.

From the discussion, we know that these type of continuous-time processes are completely governed by the dependence mechanism term and the innovation term.

Next we investigate the innovation types: non-negative integer, or positive, or real-valued increment. From the stochastic difference equation (4.3.1), we see that the dependent term and the innovation term are independent. Besides, the dependent term can take value 0 if K or $X(t)$ can be 0. Thus, we deduce the following.

(1) When K is a non-negative integer self-generalized rv, $X(t+h)$ and $(1-\mu h)_K \circledast X(t)$ are non-negative integer, thus, $\Delta\epsilon$ is non-negative integer. This implies that the innovation process $\{\epsilon(t); t \geq 0\}$ has non-negative integer-valued increment.

(2) When K is a positive self-generalized rv, $X(t+h)$ and $(1-\mu h)_K \circledast X(t)$ are non-negative real, thus, $\Delta\epsilon$ is positive. This implies that the innovation process $\{\epsilon(t); t \geq 0\}$ has positive-valued increment.

(3) When K is a positive constant c, $X(t+h)$ and $c \cdot X(t)$ are real or positive; thus, $\Delta\epsilon$ is real or positive respectively. Note that positive case has been included in (2). We only consider real case. Therefore, the innovation process $\{\epsilon(t); t \geq 0\}$ has real-valued increment.

In summary, the type of the increment of the innovation process is the same as the margins of the process $\{X(t); t \geq 0\}$.

In reality, we often encounter dynamic phenomena modelled by a process $\{X(t); t \geq 0\}$ which could be stationary or non-stationary over time. Stationarity is a simple and natural requirement for a process model. Non-stationarity usually arises from seasonality, increasing or declining trend, heteroscedasticity, etc. Thus, appropriate model settings should be considered. A good process model theory should be able to accommodate both stationary and non-stationary situations.

For the stationary case, we may just simply assume that $\{\epsilon(t); t \geq 0\}$ is stationary, and that μ, in the dependence mechanism term, is a constant.

For the non-stationary case, we can modify either the dependence mechanism term or the innovation term to be time-dependent. Hence, the SDE becomes

$$dX(t) = [(1-\mu(t)dt)_K \circledast X(t) - X(t)] + d\epsilon(t),$$

where $\{\epsilon(t); t \geq 0\}$ may be a non-stationary independent increment process. However, the modification should correspond to what the non-stationary situation is. Sometimes it is a time-varying marginal mean or variance, sometimes it is a time-varying autocorrelation.

Finally, we look at some examples, where the innovation processes have non-negative integer, or positive, or real increments. Also the stochastic operations include binomial-thinning, and other extended-thinning operators. We just mention their SDE's to illustrate the existence of generalized Ornstein-Uhlenbeck SDE. Their solutions will be given in Section 4.4, as well as Chapter 5.

Example 4.1 *Let $\{X(t); t \geq 0\}$ be a process with non-negative integer margins. Consider the binomial-thinning operation. Then the following is the corresponding generalized Ornstein-Uhlenbeck SDE:*

$$dX(t) = [(1 - \mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t) = (1 - \mu dt) * X(t) - X(t) + d\epsilon(t),$$

where $\{\epsilon(t); t \geq 0\}$ is a stationary Poisson process, and the increment $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ has pgf $G_{\Delta\epsilon} = \exp\{\mu\lambda h(s-1)\}$, $\mu > 0$, $\lambda > 0$.

Example 4.2 *Let $\{X(t); t \geq 0\}$ be a process with non-negative integer margins. Still consider binomial-thinning operation. But change the innovation process to be an IIP with an increment whose pgf is*

$$G_{\Delta\epsilon} = \exp\left\{\mu\theta\gamma h \frac{s-1}{1-\gamma s}\right\}, \quad \mu > 0,\ \theta > 0,\ 0 < \gamma < 1.$$

Then the following is another generalized Ornstein-Uhlenbeck SDE:

$$dX(t) = (1 - \mu dt) * X(t) - X(t) + d\epsilon(t).$$

Example 4.3 *Let $\{X(t); t \geq 0\}$ be a process with non-negative integer margins. Consider generalized Ornstein-Uhlenbeck SDE with operator* **I2** *(Example 3.2):*

$$dX(t) = [(1 - \mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t) = \left[\sum_{i=0}^{X(t)} K_i - X(t)\right] + d\epsilon(t),$$

where $K_0 = 0$, K_1, ..., K_i, ... are iid, with pgf

$$G_{K(1-\mu h)} = \frac{\mu h + (1 - \mu h - \gamma)s}{(1 - \gamma - \gamma\mu h) - \mu\gamma h s}, \quad \mu > 0,\ 0 < \gamma < 1.$$

$\{\epsilon(t); t \geq 0\}$ is a stationary Poisson process with such increment $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ that the pgf is

$$G_{\Delta\epsilon} = \exp\left\{\mu\frac{\theta\gamma}{1-\gamma}h(s-1)\right\}, \quad \theta > 0.$$

Example 4.4 *Let $\{X(t); t \geq 0\}$ be a process with positive margins. Suppose $\{J_K(t); t \geq 0\}$ is a stationary IIP such that $\phi_{J_K(t)}(s;\alpha) = \phi_K^t(s;\alpha)$ $(\alpha > 0)$, where $\phi_K(s;\alpha) = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\}$, $0 < \gamma < 1$. Choose $\{\epsilon(t); t \geq 0\}$ to be a stationary IIP with positive increment $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$, whose LT is*

$$\phi_{\Delta\epsilon} = \frac{1}{1+s} \times \frac{1 + [1-\gamma-\mu h + 2\mu\gamma h](1-\gamma)^{-1}s}{1+\mu\gamma(1-\gamma)^{-1}hs}, \quad \mu > 0.$$

Then

$$dX(t) = [(1-\mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t) = \left[\int_0^{X(t)} dJ_K(s) - X(t)\right] + d\epsilon(t)$$

is a generalized Ornstein-Uhlenbeck SDE with operator **P2** *(see Example 3.7).*

Example 4.5 *We now return to constant multiplier operation, but choose the innovation process $\{\epsilon(t); t \geq 0\}$ to be a stationary IIP with real increment $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ such that its cf $\varphi_{\Delta\epsilon} = \exp\{-\lambda h|s|^\alpha\}$, $\lambda > 0$, $0 < \alpha \leq 2$. Then*

$$dX(t) = [(1-\mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t) = -\mu X(t)dt + d\epsilon(t)$$

is a generalized Ornstein-Uhlenbeck SDE.

4.4 Construction of possible solutions for the generalized Ornstein-Uhlenbeck SDE

We define the generalized Ornstein-Uhlenbeck SDE as

$$dX(t) = [(1-\mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t) \qquad \text{for the stationary case,}$$

or

$$dX(t) = [(1-\mu(t)dt)_K \circledast X(t) - X(t)] + d\epsilon(t) \qquad \text{for the non-stationary case,}$$

where $\{K(\alpha)\}$ is a family of self-generalized rv with respect to parameter α. This shows the differential, or in other words, the infinitesimal increment of the process $\{X(t); t \geq 0\}$ can be split into two terms: a dependence term associated with the extended-thinning operation on the current observation, and an innovation term introduced to explain the remaining fluctuation. Assuming that a solution exists, our tasks are

(1) what does the solution mean?

(2) how to find it?

The unknown in the generalized Ornstein-Uhlenbeck SDE is the entire process $\{X(t); t \geq 0\}$, not just $X(t)$ at a single time point. Thus, we need to find such a continuous-time process that satisfies the generalized Ornstein-Uhlenbeck SDE. Such a process is called the solution of the generalized Ornstein-Uhlenbeck SDE.

Next we have to figure out a way to obtain the solution. For this purpose, we resort to infinitesimal partition method well known in calculus. The following is the rough idea of how this method works in stochastic calculus.

Suppose the continuous-time process is $\{X(t); t \geq 0\}$. We study some kind of feature or behavior of this process between time t_1 and t_2, namely the time interval $[t_1, t_2]$. Divide this interval into n equal pieces, i.e.,

$$[t_1, t_1 + h),\ [t_1 + h, t_1 + 2h),\ \ldots,\ [t_1 + (n-1)h, t_1 + nh] = [t_1 + (n-1)h, t_2],$$

where $h = (t_2 - t_1)/n$. When n goes to infinity, each piece will become an infinitesimal interval. We consider the feature or behavior of the process in each small interval $[t_1 + (k-1)h, t_1 + kh)$ ($k = 1, 2, \ldots, n$), and apply an approximation in each small interval. Then, we sum these approximations, and finally let n increase to infinity to obtain the limit. This resulting limit is the desired process on the interval $[t_1, t_2]$.

In summary, the infinitesimal partition method applied to a continuous-time phenomena on a certain time interval involves the following steps:

(1) Discretize the continuous-time phenomena by dividing the time interval into n equal small pieces;

(2) Carry out relevant measures such as approximation and summation over these n small intervals;

(3) Make it continuous over time again for those n discretized pieces by letting $n \to +\infty$.

For the generalized Ornstein-Uhlenbeck SDE, the finite difference approximation in a small interval is

$$X(t+h) - X(t) = [(1-\mu h)_K \circledast X(t) - X(t)] + \Delta\epsilon, \quad \Delta\epsilon = \epsilon(t+h) - \epsilon(t),$$

or simply

$$X(t+h) \;=\; (1-\mu h)_K \circledast X(t) + \Delta\epsilon, \tag{4.4.1}$$

for the stationary case; and

$$X(t+h) - X(t) = [(1-\mu(t)h)_K \circledast X(t) - X(t)] + \Delta\epsilon(t), \quad \Delta\epsilon(t) = \epsilon(t+h) - \epsilon(t),$$

or simply

$$X(t+h) \;=\; (1-\mu(t)h)_K \circledast X(t) + \Delta\epsilon(t), \tag{4.4.2}$$

for the non-stationary case. These will be applied to construct the solution of the generalized Ornstein-Uhlenbeck SDE next.

Before proceeding to the solution, we list a useful lemma below.

Lemma 4.4.1 *If $\{a_k\}$ is a bounded sequence such that $\lim\limits_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} a_k = a$, then*

$$\prod_{k=0}^{n-1}\left(1 - \frac{a_k}{n}\right) \longrightarrow e^{-a} \qquad \textit{as } n \to \infty.$$

Proof: Expand the product and take a limit term by term.

In the rest of this section, we construct the possible solution $\{X(t); t \geq 0\}$ for the generalized Ornstein-Uhlenbeck SDE. We are interested in the conditional stochastic representation form of $\{X(t); t \geq 0\}$.

First consider the generalized Ornstein-Uhlenbeck SDE in the stationary case. Concretely, the setting of generalized Ornstein-Uhlenbeck SDE consists of constant parameter μ, and the innovation being a stationary independent increment process, namely

$$dX(t) = [(1 - \mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t).$$

We apply the infinitesimal partition method to obtain an explicit expression for $X(t_2)$ given $X(t_1)$, where $t_1 < t_2$.

Let $h = (t_2 - t_1)/n$, and $\Delta\epsilon_k = \epsilon(t_1 + kh) - \epsilon(t_1 + (k-1)h)$, $k = 1, 2, \ldots, n$. Then from (4.4.1), we have

$$X(t_1 + h) = (1 - \mu h)_K \circledast X(t_1) + \Delta\epsilon_1;$$

$$X(t_1 + 2h) = (1 - \mu h)_K \circledast X(t_1 + h) + \Delta\epsilon_2;$$

$$\vdots$$

$$X(t_2) = X(t_1 + nh) = (1 - \mu h)_K \circledast X(t_1 + (n-1)h) + \Delta\epsilon_n.$$

By induction and employing the properties of the extended-thinning operation,

$$\begin{aligned}
X(t_1 + 2h) &= (1 - \mu h)_K \circledast [(1 - \mu h)_K \circledast X(t_1) + \Delta\epsilon_1] + \Delta\epsilon_2 \\
&= (1 - \mu h)_K \circledast (1 - \mu h)_K \circledast X(t_1) + (1 - \mu h)_K \circledast \Delta\epsilon_1 + \Delta\epsilon_2 \\
&= (1 - \mu h)_K^2 \circledast X(t_1) + (1 - \mu h)_K \circledast \Delta\epsilon_1 + \Delta\epsilon_2, \\
X(t_1 + 3h) &= (1 - \mu h)_K^3 \circledast X(t_1) + (1 - \mu h)_K^2 \circledast \Delta\epsilon_1 + (1 - \mu h)_K \circledast \Delta\epsilon_2 + \Delta\epsilon_3, \\
&\vdots \\
X(t_2) &= (1 - \mu h)_K^n \circledast X(t_1) + \sum_{k=0}^{n-1} (1 - \mu h)_K^k \circledast \Delta\epsilon_{n-k}.
\end{aligned}$$

Let $Y_n = (1 - \mu h)_K^n \circledast X(t_1)$, and

$$Z_n = \sum_{k=0}^{n-1} (1 - \mu h)_K^k \circledast \Delta\epsilon_{n-k}. \tag{4.4.3}$$

Note that Y_n and Z_n are independent. When n goes to $+\infty$,

$$(1 - \mu h)^n = \left(1 - \mu \frac{t_2 - t_1}{n}\right)^n \longrightarrow e^{-\mu(t_2 - t_1)}.$$

Hence, by Property 3.14,

$$Y_n \xrightarrow{\mathrm{d}} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1), \quad \text{as } n \longrightarrow +\infty.$$

Assume that $\{Z_n\}$ converges in distribution. Then this limit will be

$$\begin{aligned}
\lim_{n\to+\infty} Z_n &\stackrel{d}{=} \lim_{n\to+\infty}\sum_{k=0}^{n-1}(1-\mu h)_K^k \circledast \Delta\epsilon_{n-k} = \lim_{n\to+\infty}\sum_{k=0}^{n-1}\left(e^{-\mu(kh)}\right)_K \circledast \Delta\epsilon_{n-k} \\
&= \lim_{n\to+\infty}\sum_{j=1}^{n}\left(e^{-\mu(t_2-t_1-jh)}\right)_K \circledast \Delta\epsilon_j = \int_0^{t_2-t_1}\left(e^{-\mu(t_2-t_1-t)}\right)_K \circledast d\epsilon(t).
\end{aligned}$$

However, in the stationary situation, since $\Delta\epsilon_1$, $\Delta\epsilon_2$, $\ldots$, $\Delta\epsilon_n$ are iid, we can derive a simpler expression,

$$\lim_{n\to+\infty}\sum_{k=0}^{n-1}\left(e^{-\mu(kh)}\right)_K \circledast \Delta\epsilon_{n-k} \stackrel{d}{=} \lim_{n\to+\infty}\sum_{k=0}^{n-1}\left(e^{-\mu(kh)}\right)_K \circledast \Delta\epsilon_k = \int_0^{t_2-t_1}\left(e^{-\mu t}\right)_K \circledast d\epsilon(t).$$

Finally, by the independence of Y_n and Z_n, we obtain

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1}\left(e^{-\mu t}\right)_K \circledast d\epsilon(t). \tag{4.4.4}$$

Next we turn to the non-stationary case, where we allow μ to be a function of t, i.e., $\mu(t)$, and the innovation process could be a stationary independent increment process or a non-stationary independent increment process. Then the generalized Ornstein-Uhlenbeck SDE becomes

$$dX(t) = [(1-\mu(t)dt)_K \circledast X(t) - X(t)] + d\epsilon(t).$$

We follow the convention that a null product $\prod_{j=1}^{0} a_j$ is 1. We have a slightly different version of the approximation of differences in a small interval based on (4.4.2):

$$\begin{aligned}
X(t_1+2h) &= \left(1-\mu(t_1+h)h\right)_K \circledast \left[\left(1-\mu(t_1)h\right)_K \circledast X(t_1) + \Delta\epsilon_1\right] + \Delta\epsilon_2 \\
&= \left(1-\mu(t_1+h)h\right)_K \circledast \left(1-\mu(t_1)h\right)_K \circledast X(t_1) \\
&\quad + \left(1-\mu(t_1+h)h\right)_K \circledast \Delta\epsilon_1 + \Delta\epsilon_2 \\
&= \left([1-\mu(t_1)h][1-\mu(t_1+h)h]\right)_K \circledast X(t_1) \\
&\quad + \left(1-\mu(t_1+h)h\right)_K \circledast \Delta\epsilon_1 + \Delta\epsilon_2,
\end{aligned}$$

$$\begin{aligned} X(t_1+3h) &= \Big([1-\mu(t_1)h][1-\mu(t_1+h)h][1-\mu(t_1+2h)h]\Big)_K \circledast X(t_1) \\ &\quad +\Big([1-\mu(t_1+h)h][1-\mu(t_1+2h)h]\Big)_K \circledast \Delta\epsilon_1 \\ &\quad +\Big([1-\mu(t_1+2h)h]\Big)_K \circledast \Delta\epsilon_2 + \Delta\epsilon_3, \\ &\;\;\vdots \\ X(t_2) &= \Big(\prod_{k=0}^{n-1}[1-\mu(t_1+kh)h]\Big)_K \circledast X(t_1) \\ &\quad +\sum_{k=0}^{n-1}\Big(\prod_{j=1}^{k}[1-\mu(t_1+(n-j)h)h]\Big)_K \circledast \Delta\epsilon_{n-k}. \end{aligned}$$

Similar to before, let

$$Y_n = \Big(\prod_{k=0}^{n-1}[1-\mu(t_1+kh)h]\Big)_K \circledast X(t_1)$$

and

$$\begin{aligned} Z_n &= \sum_{k=0}^{n-1}\Big(\prod_{j=1}^{k}[1-\mu(t_1+(n-j)h)h]\Big)_K \circledast \Delta\epsilon_{n-k} \\ &= \sum_{k=0}^{n-1}\Big(\prod_{j=1}^{k}[1-\mu(t_2-jh)h]\Big)_K \circledast \Delta\epsilon_{n-k}. \end{aligned} \tag{4.4.5}$$

Then Y_n and Z_n are independent. Note that by Lemma 4.4.1, as $n \to +\infty$,

$$\prod_{k=0}^{n-1}[1-\mu(t_1+kh)h] \doteq \prod_{k=0}^{n-1} e^{-\mu(t_1+kh)h} = \exp\left\{-\sum_{k=0}^{n-1}\mu(t_1+kh)h\right\} \longrightarrow e^{-\int_{t_1}^{t_2}\mu(t)dt}.$$

Hence,

$$Y_n \xrightarrow{\mathrm{d}} \left(e^{-\int_{t_1}^{t_2}\mu(t)dt}\right)_K \circledast X(t_1), \quad \text{as } n \longrightarrow +\infty.$$

Assume that $\{Z_n\}$ converges in distribution. Then

$$\begin{aligned} \lim_{n\to+\infty} Z_n &= \lim_{n\to+\infty}\sum_{k=0}^{n-1}\Big(\prod_{j=1}^{k}[1-\mu(t_2-jh)h]\Big)_K \circledast \Delta\epsilon_{n-k} \\ &= \lim_{n\to+\infty}\sum_{k=0}^{n-1}\left(\prod_{j=1}^{k} e^{-\mu(t_2-jh)h}\right)_K \circledast \Delta\epsilon_{n-k} \\ &= \lim_{n\to+\infty}\sum_{k=0}^{n-1}\left(e^{-\sum_{j=1}^{k}\mu(t_2-jh)h}\right)_K \circledast \Delta\epsilon_{n-k} \end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to+\infty}\sum_{k=1}^{n}\left(e^{-\sum_{j=1}^{n-k}\mu(t_2-jh)h}\right)_K \circledast \Delta\epsilon_k \\
&\stackrel{d}{=} \int_{t_1}^{t_2}\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t).
\end{aligned}
$$

Finally, by the independence of Y_n and Z_n, we obtain

$$
X(t_2) \stackrel{d}{=} \left(e^{-\int_{t_1}^{t_2}\mu(t)dt}\right)_K \circledast X(t_1) + \int_{t_1}^{t_2}\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t). \tag{4.4.6}
$$

The stochastic integral $\int_{t_1}^{t_2}\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$ can be viewed as a cumulative innovation. The more recent innovation has more influence on the cumulative innovation, because $e^{-\int_t^{t_2}\mu(\tau)d\tau}$ is increasing as t approaches t_2. For a stationary generalized Ornstein-Uhlenbeck SDE, although the stochastic integrals $\int_0^{t_2-t_1}\left(e^{-\mu(t_2-t_1-t)}\right)_K \circledast d\epsilon(t)$ and $\int_0^{t_2-t_1}\left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$ have slightly different interpretations, it doesn't matter because they are equal in distribution.

The stochastic representations of the solution of the generalized Ornstein-Uhlenbeck SDE are shown in (4.4.4) and (4.4.6), where the current state can be split into two independent terms: the first is the dependent part related to previous state; the second is the cumulative innovation, a generalized stochastic integration. Now the natural questions arise:

- When does the generalized stochastic integral exist?
- How to find this stochastic integral?

Recall that the generalized stochastic integral is defined as the limit in distribution. Hence the convenient tools to investigate such a generalized stochastic integral are the pgf, or LT, or cf, depending on if the innovation process has non-negative integer, or positive, or real-valued margins.

Assume K has pgf $G_K(s,\alpha)$, or LT $\phi_K(s,\alpha)$, or cf $\varphi_K(s,\alpha)$; and $\Delta\epsilon_1$ has pgf $G_{\Delta\epsilon}(s)$, or $\phi_{\Delta\epsilon}(s)$, or $\varphi_{\Delta\epsilon}(s)$. Now we study the pgf, or LT, or cf for Z_n.

For the stationary case, since $\Delta\epsilon_1, \Delta\epsilon_2, \ldots, \Delta\epsilon_n$ are iid, it follows that

$$
\begin{aligned}
G_{Z_n}(s) &= \mathbf{E}\left[s^{\sum_{k=0}^{n-1}((1-\mu h)^k)_K \circledast \Delta\epsilon_{n-k}}\right] = \prod_{k=0}^{n-1}\mathbf{E}\left[s^{((1-\mu h)^k)_K \circledast \Delta\epsilon_{n-i}}\right] \\
&= \prod_{k=0}^{n-1} G_{\Delta\epsilon}\left(G_K\left(s;(1-\mu h)^k\right)\right),
\end{aligned}
$$

or

$$
\begin{aligned}
\phi_{Z_n}(s) &= \mathbf{E}\left[e^{-s\sum_{k=0}^{n-1}((1-\mu h)^k)_K \circledast \Delta\epsilon_{n-k}}\right] = \prod_{k=0}^{n-1} \mathbf{E}\left[e^{-s((1-\mu h)^k)_K \circledast \Delta\epsilon_{n-i}}\right] \\
&= \prod_{k=0}^{n-1} \phi_{\Delta\epsilon}\left(-\log \phi_K\left(s;(1-\mu h)^k\right)\right),
\end{aligned}
$$

or

$$
\begin{aligned}
\varphi_{Z_n}(s) &= \mathbf{E}\left[e^{is\sum\limits_{k=0}^{n-1}\left((1-\mu h)^k\right)_K \circledast \Delta\epsilon_{n-k}}\right] = \prod_{k=0}^{n-1} \mathbf{E}\left[e^{is\left((1-\mu h)^k\right)_K \circledast \Delta\epsilon_{n-k}}\right] \\
&= \prod_{k=0}^{n-1} \phi_{\Delta\epsilon}\left(-\log \varphi_K\left(s;(1-\mu h)^k\right)\right) \\
&= \prod_{k=0}^{n-1} \varphi_{\Delta\epsilon}\left(i\log \varphi_K\left(s;(1-\mu h)^k\right)\right).
\end{aligned}
$$

If as $n \longrightarrow +\infty$,

$$
G_{Z_n}(s) \longrightarrow G(s), \quad \text{or} \quad \phi_{Z_n}(s) \longrightarrow \phi(s), \quad \text{or} \quad \varphi_{Z_n}(s) \longrightarrow \varphi(s),
$$

and

$$
\begin{gathered}
G(s) \text{ is continuous at } s=1 \text{ with } G(1)=1, \quad \textit{or} \\
\phi(s) \text{ is continuous at } s=0 \text{ with } \phi(0)=1, \quad \textit{or} \\
\varphi(s) \text{ is continuous at } s=0 \text{ with } \varphi(0)=1,
\end{gathered}
$$

then $G(s)$, or $\phi(s)$ or $\varphi(s)$ is the pgf, or LT, or cf of $\int_0^{t_2-t_1}\left(e^{-\mu t)}\right)_K \circledast d\epsilon(t)$.

For the non-stationary case, $\Delta\epsilon_1, \Delta\epsilon_2, \ldots, \Delta\epsilon_n$ are independent, they may or may not be identically distributed. The corresponding pgf, or LT, or cf of Z_n are

$$
\begin{aligned}
G_{Z_n}(s) &= \mathbf{E}\left[s^{\sum\limits_{k=0}^{n-1}\left(\prod\limits_{j=1}^{k}[1-\mu(t_2-jh)h]\right)_K \circledast \Delta\epsilon_{n-k}}\right] \\
&= \prod_{k=0}^{n-1} \mathbf{E}\left[s^{\left(\prod\limits_{j=1}^{k}[1-\mu(t_2-jh)h]\right)_K \circledast \Delta\epsilon_{n-k}}\right] \\
&= \prod_{k=0}^{n-1} G_{\Delta\epsilon_{n-k}}\left(G_K\left(s;\prod_{j=1}^{k}[1-\mu(t_2-jh)h]\right)\right),
\end{aligned}
$$

or

$$\begin{aligned}
\phi_{Z_n}(s) &= \mathbf{E}\left[e^{-s\sum\limits_{k=0}^{n-1}\left(\prod\limits_{j=1}^{k}[1-\mu(t_2-jh)h]\right)_K \circledast \Delta\epsilon_{n-k}}\right] \\
&= \prod_{k=0}^{n-1}\mathbf{E}\left[e^{-s\left(\prod\limits_{j=1}^{k}[1-\mu(t_2-jh)h]\right)_K \circledast \Delta\epsilon_{n-k}}\right] \\
&= \prod_{k=0}^{n-1}\phi_{\Delta\epsilon_{n-k}}\left(-\log\phi_K\left(s;\prod_{j=1}^{k}[1-\mu(t_2-jh)h]\right)\right),
\end{aligned}$$

or

$$\begin{aligned}
\varphi_{Z_n}(s) &= \mathbf{E}\left[e^{is\sum\limits_{k=0}^{n-1}\left(\prod\limits_{j=1}^{k}[1-\mu(t_2-jh)h]\right)_K \circledast \Delta\epsilon_{n-k}}\right] \\
&= \prod_{k=0}^{n-1}\mathbf{E}\left[e^{is\left(\prod\limits_{j=1}^{k}[1-\mu(t_2-jh)h]\right)_K \circledast \Delta\epsilon_{n-k}}\right] \\
&= \prod_{k=0}^{n-1}\phi_{\Delta\epsilon_{n-k}}\left(-\log\varphi_K\left(s;\prod_{j=1}^{k}[1-\mu(t_2-jh)h]\right)\right) \\
&= \prod_{k=0}^{n-1}\varphi_{\Delta\epsilon_{n-k}}\left(i\log\varphi_K\left(s;\prod_{j=1}^{k}[1-\mu(t_2-jh)h]\right)\right),
\end{aligned}$$

Similarly, if

$$G_{Z_n}(s) \longrightarrow G(s), \quad \text{or} \quad \phi_{Z_n}(s) \longrightarrow \phi(s), \quad \text{or} \quad \varphi_{Z_n}(s) \longrightarrow \varphi(s), \quad \text{as } n \longrightarrow +\infty,$$

and

$$\begin{aligned}
&G(s) \text{ is continuous at } s=1 \text{ with } G(1)=1, \quad \textit{or} \\
&\phi(s) \text{ is continuous at } s=0 \text{ with } \phi(0)=1, \quad \textit{or} \\
&\varphi(s) \text{ is continuous at } s=0 \text{ with } \varphi(0)=1,
\end{aligned}$$

then we can conclude that $G(s)$, or $\phi(s)$, or $\varphi(s)$ is the pgf, or LT, or cf of $\int_{t_1}^{t_2}\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$.

These analyses show that finding the generalized stochastic integral is equivalent to finding the limit of pgf, or LT, or cf of Z_n, where the limit should be a pgf, or LT, or cf. Due to self-generalizability, the dependence part Y_n in the approximation always converges. Hence, we obtain the following theorem.

Theorem 4.4.2 (Solutions of the generalized Ornstein-Uhlenbeck SDE)

(1) For the stationary generalized Ornstein-Uhlenbeck SDE

$$dX(t) = [(1 - \mu dt)_K \circledast X(t) - X(t)] + d\epsilon(t),$$

if $\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$ *exists, then the solution is*

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t).$$

(2) For the non-stationary generalized Ornstein-Uhlenbeck SDE, where $\mu(t)$ is bounded,

$$dX(t) = [(1 - \mu(t)dt)_K \circledast X(t) - X(t)] + d\epsilon(t),$$

if $\int_{t_1}^{t_2} \left(e^{-\int_t^{t_2} \mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$ *exists, then the solution is*

$$X(t_2) \stackrel{d}{=} \left(e^{-\int_{t_1}^{t_2} \mu(t)dt}\right)_K \circledast X(t_1) + \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2} \mu(\tau)d\tau}\right)_K \circledast d\epsilon(t).$$

Note that in the sense of convergence in distribution, the classical Ornstein-Uhlenbeck SDE has the same solution as in the sense of convergence in L^2.

In summary, the stochastic representations of $X(t_2)$ conditioned on $X(t_1)$ in both the stationary and non-stationary cases show that $\{X(t); t \geq 0\}$ (if existing) is a first-order Markov process. Since the classical Ornstein-Uhlenbeck SDE leads to the continuous-time AR(1) Gaussian process, we name the new processes, constructed by the generalized Ornstein-Uhlenbeck SDE, the continuous-time generalized AR(1) processes, or in short, the continuous-time GAR(1) processes. Specifically, they are of the forms given in Theorem 4.4.2. Some comments about the comparison with traditional AR(1) processes are given in the next section.

Our next main task is to search for appropriate IIP innovations, which guarantee that the generalized stochastic integral exists.

4.5 Summary and discussion

In this section, we compare among the Ornstein-Uhlenbeck processes, the Ornstein-Uhlenbeck-type processes, and the new continuous-time generalized AR(1) processes. We summarize their features only for the stationary situation.

The Ornstein-Uhlenbeck process and the Ornstein-Uhlenbeck-type process have the same stochastic representation:

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} \bullet X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t).$$

If $\{\epsilon(t); t \geq 0\}$ is a Wiener process, it's the ordinary Ornstein-Uhlenbeck process. If $\{\epsilon(t); t \geq 0\}$ is a Lévy process, it is then the Ornstein-Uhlenbeck-type processes. Extending the constant multiplier operation to extended-thinning operation, we obtain the continuous-time generalized AR(1) process:

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t).$$

Now we investigate their representation structure. The expressions of the stochastic representation show that $X(t_2)$ consists of two independent parts: one related to $X(t_1)$ only and one related to the innovation process only. When $t_2 - t_1$ goes to infinity, $e^{-\mu(_2-t_1)}$ goes to zero. Thus, the first part will diminish to zero, which means $X(t_1)$ will gradually have less and less influence on $X(t_2)$ until the influence reduces to null. Then the influence will exclusively come from the innovation process. This shows us a dynamic process picture: after continuously repeated treatment by the dependence mechanism device (i.e., $A(X(t), -\mu dt)$), the original input $X(t_1)$ will be diminishing to nothing. On the other hand, the innovation during this period will be treated by the same mechanism; however, it has accumulated as a stochastic integral and finally accounts for $X(t_2)$ solely.

What are the restrictions to applying the Ornstein-Uhlenbeck process and the Ornstein-Uhlenbeck-type process in modelling dynamic phenomena? Since the Ornstein-Uhlenbeck process has real-valued margins, it only works for those with real observations over time, otherwise, some transformation for the data, say log-transformation, should be take to fit the model. The Ornstein-Uhlenbeck-type process extends to positive real-valued margins, but can not have non-negative integer-valued margins, because $e^{-\mu(t_2-t_1)} \bullet X(t_1)$ is not likely to be an integer. However, the continuous-time generalized AR(1) process offers non-negative integer-valued, positive real-valued and real-valued margins. Therefore, it is quite flexible in modelling different types of margins. In Chapter 5, we give abundant examples.

For the Ornstein-Uhlenbeck process and the Ornstein-Uhlenbeck-type process, the dependence part is a linear function of $X(t_1)$. Hence, conditioned on $X(t_1)$, this dependence part is fixed, not random. The conditional variation of $X(t_2)$ only comes from the part related to innovation. For the continuous-time generalized AR(1) process, however, this part may not be a linear function of $X(t_1)$. Furthermore, conditioned on $X(t_1)$, it's no longer fixed, but random, so it looks like a random effect, and also contributes to the conditional variation of $X(t_2)$.

The reader may wonder why we name this type process as continuous-time generalized AR(1) process, instead of following the conventional way to name it like the generalized Ornstein-Uhlenbeck-type process. This is because we focus on the statistical point of view. We wish to emphasize its advantage, the auto-regression like property, in statistical modelling. One big concern in modelling dynamic phenomena is to capture the dependence structure over time. Since the processes we study possess the same auto-correlation as the continuous-time AR(1) Gaussian process, we propose the name continuous-time generalized AR(1) process to clearly show that one can apply this kind of process to model, or in another word, to approximate the real dynamic problems which have obvious dependence structure over time. However, the concept "auto- regression" in the new processes is not strictly autoregression, because $X(t_2)$ is not equal in distribution to a linear function of $X(t_1)$. This jump in concept resembles the relationship of the generalized linear model to the linear model. If one like, one may call such an autoregression a generalized autoregression to distinguish it with the classical autoregression, which is linear in the previous observation.

Chapter 5

Results for continuous-time generalized AR(1) processes

In this chapter, we shall deduce some concrete results of continuous-time GAR(1) processes discussed in Chapter 4. We consider special IIP's as the innovation processes. These innovation processes are classified as having non-negative integer, positive and real increment.

The general conclusion for these special innovation processes is given in Section 5.1. From Sections 5.2 to 5.3, we discuss the non-negative integer, positive and real increment cases and examples respectively. Finally, in Section 5.4 we explore the Tweedie IIP as the innovation process to study or revisit the models from the view of dispersion.

These innovation processes lead to the continuous-time GAR(1) processes with non-negative integer, positive and real margins. The abundant resulting processes could be potential models for real phenomena that statisticians seek to explore.

5.1 Main results for continuous-time GAR(1) processes

In this section, we apply the theory in Chapter 4 to construct concrete examples of the continuous-time GAR(1) processes.

First, we choose relevant IIP's as innovation processes. They could have non-negative integer, or positive, or real margins. Secondly, we specify the extended-thinning operations. These two steps determine the generalized Ornstein-Uhlenbeck SDE, and thus, the corresponding continuous-time GAR(1) processes.

The key point in such procedure is to calculate $G(s)$, or $\phi(s)$, or $\varphi(s)$ discussed in Section 4.4. If this is a pgf, or LT, or cf, then we obtain the corresponding generalized stochastic integral, as well as the corresponding continuous-time GAR(1) process.

Both stationary and non-stationary situations are investigated. However, for the sake of simplicity, we restrict the innovation processes as stationary IIP's so that we can easily give explicit results. This idea can be readily extended to the case of the non-stationary IIP's being the innovation processes.

Now we probe the issue of the support type of margins of continuous-time GAR(1) processes. Recall the stochastic representations of this kind of processes in stationary case:

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t).$$

When $t_2 - t_1 \to \infty$, the term $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)$ will converge to zero. Hence, the margins and their support are essentially governed by the rv K and the innovation process $\{\epsilon(t); t \geq 0\}$. We may want the marginal support to be the non-negative integer, or positive, or real set. This can be realized by choosing the appropriate self-generalized rv K and IIP $\{\epsilon(t);\ t \geq 0\}$.

Recall that the generalized stochastic integral involved in the continuous-time GAR(1) process, $\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$ or $\int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$ have pgf, or LT, or cf as the following

$$G_{Z_n}(s) = \begin{cases} \prod_{i=0}^{n-1} G_{\Delta\epsilon}\left(G_K\left(s;(1-\mu h)^i\right)\right), & \text{for constant } \mu; \\ \prod_{i=0}^{n-1} G_{\Delta\epsilon}\left(G_K\left(s;\prod_{j=1}^{i}\left[(1-\mu(t_2-jh)h\right]\right)\right), & \text{for } \mu(t); \end{cases}$$

or

$$\phi_{Z_n}(s) = \begin{cases} \prod_{i=0}^{n-1} \phi_{\Delta\epsilon}\left(-\log\phi_K\left(s;(1-\mu h)^i\right)\right), & \text{for constant } \mu; \\ \prod_{i=0}^{n-1} \phi_{\Delta\epsilon}\left(-\log\phi_K\left(s;\prod_{j=1}^{i}\left[(1-\mu(t_2-jh)h\right]\right)\right), & \text{for } \mu(t); \end{cases}$$

or or

$$\varphi_{Z_n}(s) = \begin{cases} \prod_{i=0}^{n-1} \varphi_{\Delta\epsilon}\left(-i\log\varphi_K\left(s;(1-\mu h)^i\right)\right). & \text{for constant } \mu; \\ \prod_{i=0}^{n-1} \varphi_{\Delta\epsilon}\left(-i\log\varphi_K\left(s;\prod_{j=1}^{i}[(1-\mu(t_2-jh)h]\right)\right), & \text{for } \mu(t); \end{cases}$$

where Z_n is defined in (4.4.3) and (4.4.5). Since they are all products, we may choose innovation processes in which the pgf, or LT, or cf of increment is of the exponential form. Such a form has the advantage that it can change the product to summation of their exponents so that the limit will be an integral.

The following proposition from calculus is essential to our study of results of continuous-time GAR(1) processes.

Proposition 5.1.1 *Suppose $R(x)$ is a differentiable real-valued function with bounded first order derivative. Let $h = (t_2 - t_1)/n$, where $t_2 - t_1 > 0$. Then*

(1) for constant μ,

$$h\sum_{i=0}^{n-1} R\left((1-\mu h)^i\right) \longrightarrow \int_0^{t_2-t_1} R\left(e^{-\mu t}\right)dt, \qquad \text{as } n\to\infty,$$

(2) for function $\mu(t)$,

$$h\sum_{i=0}^{n-1} R\left(\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right) \longrightarrow \int_{t_1}^{t_2} R\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)dt, \qquad \text{as } n\to\infty.$$

Proof: The key step is to show that

$$\begin{aligned} h\sum_{i=0}^{n-1} R\left((1-\mu h)^i\right) &= h\sum_{i=0}^{n-1} R\left(e^{-\mu i h}\right) + o(h), \\ h\sum_{i=0}^{n-1} R\left(\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right) &= h\sum_{i=0}^{n-1} R\left(e^{-\sum_{j=1}^{i}\mu(t_2-jh)h}\right) + o(h). \end{aligned}$$

Then by the definition of Riemann integration, the conclusions hold.

According to Inequality 3.6.2 in Mitrinović and Vasić [1970] (Section 3.6, p. 266), we have

$$\left|\left(1-\frac{x}{n}\right)^n - e^{-x}\right| = O\left(\frac{1}{n}\right).$$

Thus,

$$(1-\mu h)^i = \left(1-\frac{\mu(t_2-t_1)}{n}\right)^{n\times\frac{i}{n}} = \left(e^{-\mu(t_2-t_1)}+\frac{A}{n}\right)^{i/n} = e^{-\mu i h}+\frac{A_i}{n}+o\left(\frac{1}{n}\right),$$

and consequently,

$$\begin{aligned} h\sum_{i=0}^{n-1} R\left((1-\mu h)^i\right) &= h\sum_{i=0}^{n-1} R\left(e^{-\mu i h}+\frac{A_i}{n}+o\left(\frac{1}{n}\right)\right) \\ &= h\sum_{i=0}^{n-1}\left\{R\left(e^{-\mu i h}\right)+R'\left(e^{-\mu i h}\right)\cdot\frac{A_i}{n}+o\left(\frac{1}{n}\right)\right\} \\ &= h\sum_{i=0}^{n-1} R\left(e^{-\mu i h}\right)+o\left(\frac{1}{n}\right). \end{aligned}$$

Similarly,

$$\begin{aligned} \prod_{j=1}^{i}\left[1-\mu(t_2-jh)h\right] &= \prod_{j=1}^{i}\left[1-\frac{\mu(t_2-jh)(t_2-t_1)}{n}\right]^{n\times\frac{1}{n}} = \prod_{j=1}^{i}\left[e^{-\mu(t_2-jh)(t_2-t_1)}+\frac{A_j}{n}\right]^{\frac{1}{n}} \\ &= \prod_{j=1}^{i}\left[e^{-\mu(t_2-jh)h}+\frac{A_{ij}}{n}+o\left(\frac{1}{n}\right)\right] = e^{-\sum_{j=1}^{i}\mu(t_2-jh)h}+\frac{B_i}{n}+o\left(\frac{1}{n}\right), \end{aligned}$$

which leads to

$$\begin{aligned} h\sum_{i=0}^{n-1} R\left(\prod_{j=1}^{i}\left[1-\mu(t_2-jh)h\right]\right) &= h\sum_{i=0}^{n-1} R\left(e^{-\sum_{j=1}^{i}\mu(t_2-jh)h}+\frac{B_i}{n}+o\left(\frac{1}{n}\right)\right) \\ = h\sum_{i=0}^{n-1}&\left\{R\left(e^{-\sum_{j=1}^{i}\mu(t_2-jh)h}\right)+R'\left(e^{-\sum_{j=1}^{i}\mu(t_2-jh)h}\right)\cdot\frac{B_i}{n}+o\left(\frac{1}{n}\right)\right\} \\ = h\sum_{i=0}^{n-1}& R\left(e^{-\sum_{j=1}^{i}\mu(t_2-jh)h}\right)+o\left(\frac{1}{n}\right). \end{aligned}$$

Remark: The technique of proof uses a bounded first order derivative, but it may be possible to prove the result without this condition by a different method.

Now we choose a special innovation process, in which the increment, $\epsilon(t+h)-\epsilon(t)$, has the exponential form of pgf, or LT, or cf. Applying Proposition 5.1.1, we can obtain the following theorem.

Theorem 5.1.2 *Assume the innovation process $\{\epsilon(t);\ t \geq 0\}$ has increment $\epsilon(t+h)-\epsilon(t)$ such that its pgf, or LT, or cf is of form $e^{hC(s)}$, depending on the increment being non-negative integer-valued, or positive-valued, or real-valued. $C(s)$ is assumed to be differentiable with bounded first order derivative.*

(1) For the stationary situation, let Z_n be defined in (4.4.3). If $G_K(s;\alpha)$, or $\log\phi_K(s;\alpha)$, or $\log\varphi_K(s;\alpha)$ have bounded first order derivative with respect to α in $[0,1]$ (boundaries could be excluded), then it follows that

$$\begin{aligned} G_{Z_n}(s) &= \prod_{i=0}^{n-1} G_{\Delta\epsilon}\left(G_K\left(s;(1-\mu h)^i\right)\right) = \prod_{i=0}^{n-1} \exp\left\{hC\left(G_K\left(s;(1-\mu h)^i\right)\right)\right\} \\ &\overset{n\to\infty}{\longrightarrow} \exp\left\{\int_0^{t_2-t_1} C\left(G_K\left(s;e^{-\mu t}\right)\right)dt\right\} = G(s), \end{aligned}$$

or

$$\begin{aligned} \phi_{Z_n}(s) &= \prod_{i=0}^{n-1} G_{\Delta\epsilon}\left(\phi_K\left(s;(1-\mu h)^i\right)\right) \\ &\overset{n\to\infty}{\longrightarrow} \exp\left\{\int_0^{t_2-t_1} C\left(\phi_K\left(s;e^{-\mu t}\right)\right)dt\right\} = \phi(s), \end{aligned}$$

or

$$\begin{aligned} \varphi_{Z_n}(s) &= \prod_{i=0}^{n-1} \varphi_{\Delta\epsilon}\left(-i\log\varphi_K\left(s;(1-\mu h)^i\right)\right) \\ &\overset{n\to\infty}{\longrightarrow} \exp\left\{\int_0^{t_2-t_1} C\left(-i\log\varphi_K\left(s;e^{-\mu t}\right)\right)dt\right\} = \varphi(s). \end{aligned}$$

And the $G(s)$, $\phi(s)$ and $\varphi(s)$ are the pgf, LT and cf of the generalized stochastic integral

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$$

respectively.

(2) For the non-stationary situation, let Z_n be defined in (4.4.5). If $G_K(s;\alpha)$, or $\log\phi_K(s;\alpha)$, or $\log\varphi_K(s;\alpha)$ have bounded first order derivative in $[0,1]$ (boundaries could be excluded), then it follows that

$$\begin{aligned} G_{Z_n}(s) &= \prod_{i=0}^{n-1} G_{\Delta\epsilon}\left(G_K\left(s;\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right)\right) \\ &= \prod_{i=0}^{n-1}\exp\left\{hC\left(G_K\left(s;\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right)\right)\right\} \\ &\xrightarrow{n\to\infty} \exp\left\{\int_{t_1}^{t_2} C\left(G_K\left(s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)\right)dt\right\} = G(s), \end{aligned}$$

or

$$\begin{aligned} \phi_{Z_n}(s) &= \prod_{i=0}^{n-1} \phi_{\Delta\epsilon}\left(-\log\phi_K\left(s;\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right)\right) \\ &\xrightarrow{n\to\infty} \exp\left\{\int_{t_1}^{t_2} C\left(-\log\phi_K\left(s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)\right)dt\right\} = \phi(s), \end{aligned}$$

or

$$\begin{aligned} \varphi_{Z_n}(s) &= \prod_{i=0}^{n-1} \varphi_{\Delta\epsilon}\left(-i\log\varphi_K\left(s;\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right)\right) \\ &\xrightarrow{n\to\infty} \exp\left\{\int_{t_1}^{t_2} C\left(-i\log\varphi_K\left(s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)\right)dt\right\} = \varphi(s). \end{aligned}$$

The $G(s)$, $\phi(s)$ and $\varphi(s)$ are the pgf, LT and cf of the generalized stochastic integral

$$\int_{t_1}^{t_2}\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

respectively.

Proof: With suppression of the dependence on s, let

$$R(\alpha) = \begin{cases} C(G_K(s;\alpha)), & \text{if } K \text{ is a non-negative integer rv,} \\ C(-\log\phi_K(s;\alpha)), & \text{if } K \text{ a is positive rv,} \\ C(-i\log\varphi_K(s;\alpha)), & \text{if } K \text{ a is real rv.} \end{cases}$$

Then the derivative of $R(\alpha)$ is

$$R'(\alpha) = \begin{cases} C'(G_K(s;\alpha))\frac{\partial}{\partial\alpha}G_K(s;\alpha), & \text{if } K \text{ is a non-negative integer rv,} \\ -C'(-\log\phi_K(s;\alpha))\frac{\partial}{\partial\alpha}\log\phi_K(s;\alpha), & \text{if } K \text{ a is positive rv,} \\ -iC'(-i\log\varphi_K(s;\alpha))\frac{\partial}{\partial\alpha}\log\varphi_K(s;\alpha), & \text{if } K \text{ a is real rv.} \end{cases}$$

Hence, $R'(\alpha)$ is bounded in $[0,1]$ (boundaries could be excluded) under the conditions of this theorem. To save space, we only verify them for K being a non-negative integer rv and the increment of innovation process being non-negative integer-valued.

(1) For the stationary situation,

$$\begin{aligned} G_{Z_n}(s) &= \prod_{i=0}^{n-1} G_{\Delta\epsilon}\left(G_K\left(s;(1-\mu h)^i\right)\right) = \prod_{i=0}^{n-1}\exp\left\{hC\left(G_K\left(s;(1-\mu h)^i\right)\right)\right\} \\ &= \exp\left\{h\sum_{i=0}^{n-1} R\left((1-\mu h)^i\right)\right\} \end{aligned}$$

under the conditions in this theorem, Proposition 5.1.1 holds. Thus,

$$G_{Z_n}(s) \overset{n\to\infty}{\longrightarrow} \exp\left\{\int_0^{t_2-t_1} C\left(G_K\left(s;e^{-\mu t}\right)\right)dt\right\} = G(s).$$

When $s=1$, since $C(1) = \log(G_{\Delta\epsilon}(1))/h = \log(1)/h = 0$,

$$G(1) = \exp\left\{\int_0^{t_2-t_1} C\left(G_K\left(1;e^{-\mu t}\right)\right)dt\right\} = \exp\left\{\int_0^{t_2-t_1} C(1)dt\right\} = e^0 = 1.$$

Therefore, $G(s)$ is a pgf. We can conclude that $\int_0^{t_2-t_1}\left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$ exists with pgf $G(s)$.

(2) For the non-stationary situation,

$$\begin{aligned} G_{Z_n}(s) &= \prod_{i=0}^{n-1} G_{\Delta\epsilon}\left(G_K\left(s;\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right)\right) \\ &= \prod_{i=0}^{n-1}\exp\left\{hC\left(G_K\left(s;\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right)\right)\right\} \\ &= \exp\left\{h\sum_{i=0}^{n-1}R\left(\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\right)\right\} \end{aligned}$$

By Proposition 5.1.1,

$$G_{Z_n}(s) \overset{n\to\infty}{\longrightarrow} \exp\left\{\int_{t_1}^{t_2} C\left(G_K\left(s; e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)\right) dt\right\} = G(s),$$

and $G(1) = 1$. Hence $\displaystyle\int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$ exists with pgf $G(s)$.

Recall Examples 3.1 to 3.5 (labeled from **I1** to **I5**) where all of the self-generalized rv K are non-negative integer-valued. We check the partial derivatives

$$\frac{\partial}{\partial\alpha}G_K(s;\alpha) = \begin{cases} (s-1), & \text{for } \mathbf{I1}, \\ (1-\gamma)(s-1)(1-\gamma s)[\gamma(s-1)\alpha + (1-\gamma s)]^{-2}, & \text{for } \mathbf{I2}, \\ -(1-s)^{\alpha}\log(1-s), & \text{for } \mathbf{I3}, \\ -\frac{1}{e^{\theta}-1}\left[e^{\theta}-(e^{\theta}-1)s\right]^{\alpha}\log\left[e^{\theta}-(e^{\theta}-1)s\right], & \text{for } \mathbf{I4}, \\ -\theta(1-\gamma)^{\theta}\left[1+(1-\gamma)(1-s)^{-1/\theta}\right] & \\ \qquad \times\alpha^{\theta-1}\left[(1-\alpha)\gamma+(1-\gamma)(1-s)^{-1/\theta}\right]^{-(\theta+1)}, & \text{for } \mathbf{I5}. \end{cases}$$

Thus, they are all bounded if $0 < s < 1$.

For Examples 3.6 to 3.10 (labeled from **P1** to **P5**) where K is a positive self-generalized rv, because the following relationship corresponding to the **I1** to **I5** holds

$$\phi_K(s;\alpha) = \exp\{G_K(1-s;\alpha) - 1\},$$

we obtain

$$\frac{\partial}{\partial\alpha}\log\phi_K(s;\alpha) = \frac{\partial}{\partial\alpha}G_K(1-s;\alpha).$$

Therefore, the $\frac{\partial}{\partial\alpha}\log\phi_K(s;\alpha)$ is bounded in $[0,1]$ too.

Essentially, to apply Theorem 5.1.2, we only need to check the boundedness of $C'(s)$ if we consider K being from **I1** to **I5** and **P1** to **P5**.

The type of the generalized stochastic integral

$$\int_0^{t_2-t_1}\left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2}\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

is determined by the self-generalized rv K and the innovation process $\{\epsilon(t);\ t \geq 0\}$. It could be non-negative integer-valued, or positive-valued, or real-valued. Under the cases that K is from **I1** to **I5** and **P1** to **P5**, we classify them in the following theorem.

Theorem 5.1.3 *Assume the innovation process $\{\epsilon(t);\ t \geq 0\}$ has increment $\Delta\epsilon(t) = \epsilon(t+h) - \epsilon(t)$ whose pgf, or LT, or cf has form $e^{hC(s)}$ depending on the increment being non-negative integer-valued, or positive-valued, or real-valued. $C(s)$ is differentiable with bounded first order derivative. We classify the possible type of the generalized stochastic integral.*

(1) non-negative integer-valued: K could be from **I1** *to* **I5**, *the increment of innovation process should be non-negative integer-valued.*

In this circumstance, if the increment has pgf $G_{\Delta\epsilon(t)}(s) = e^{hC(s)}$, then

$$G(s) = \begin{cases} \exp\left\{\int_0^{t_2-t_1} C\left(G_K\left(s; e^{-\mu t}\right)\right) dt\right\}, & \text{for constant } \mu, \\ \exp\left\{\int_{t_1}^{t_2} C\left(G_K\left(s; e^{-\int_t^{t_2} \mu(\tau) d\tau}\right)\right) dt\right\}, & \text{for } \mu(t). \end{cases}$$

(2) positive-valued: K could be from **P1** *to* **P5**, *the increment of innovation process should be positive-valued.*

In this circumstance, if the increment has LT $\phi_{\Delta\epsilon(t)}(s) = e^{hC(s)}$, then

$$\phi(s) = \begin{cases} \exp\left\{\int_0^{t_2-t_1} C\left(-\log\phi_K\left(s; e^{-\mu t}\right)\right) dt\right\}, & \text{for constant } \mu, \\ \exp\left\{\int_{t_1}^{t_2} C\left(-\log\phi_K\left(s; e^{-\int_t^{t_2} \mu(\tau) d\tau}\right)\right) dt\right\}, & \text{for } \mu(t). \end{cases}$$

(3) real-valued: K is only from **P1**, *the increment of innovation process is only real-valued.*

In this circumstance, $\varphi_{\Delta\epsilon(t)}(s) = e^{hC(s)}$, and

$$\varphi(s) = \begin{cases} \exp\left\{\int_0^{t_2-t_1} C\left(s e^{-\mu t}\right) dt\right\}, & \text{for constant } \mu, \\ \exp\left\{\int_{t_1}^{t_2} C\left(s e^{-\int_t^{t_2} \mu(\tau) d\tau}\right) dt\right\}, & \text{for } \mu(t). \end{cases}$$

Proof: The proof is straightforward. We omit it to save space.

Fortunately, there are several distribution families which have exponential form of pgf, or LT, or cf. For instance, the compound Poisson, GNBC, GGC, stable, Tweedie families are well-known examples. We will discuss them in the case of innovation processes with non-negative integer, positive and real increment respectively.

5.2 Non-negative integer innovation processes and examples

In the following we mainly consider four process families as the innovations: compound Poisson IIP, generalized Negative Binomial convolution IIP, GC I IIP and GC II IIP. These four families lead to a lot of well-known distributions as margins.

As to the self-generalized distributions, because the increment of the innovation process $\{\epsilon(t); t \geq 0\}$ is non-negative integer-valued, K should be a non-negative integer rv, which further leads to the non-negative integer generalized stochastic integral. We pick up those five non-negative integer rv's in Examples 3.1 to 3.5 (labeled from **I1** to **I5**) for the extended-thinning operations.

First, we consider the Compound Poisson IIP as innovation process. By Theorem 5.1.2, we have

Theorem 5.2.1 *Let $\{\epsilon(t); t \geq 0\}$ be a Compound Poisson IIP with pgf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$*

$$G_{\Delta\epsilon}(s) = \exp\left\{\lambda h[g(s) - 1]\right\},$$

where $g(s)(= \sum_{i=0}^{\infty} p_i s^i)$ is a pgf, and differentiable with bounded first order derivative. Suppose $K(\alpha)$ is a non-negative integer self-generalized rv with pgf $G_K(s;\alpha)$, which is differentiable with bounded first order derivative with respect to α. Then, it follows that

$$G(s) = \begin{cases} \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\Big(G_K\big(s; e^{-\mu t}\big)\Big) - 1\right] dt\right\}, & \textit{for constant } \mu; \\ \\ \exp\left\{\lambda \int_{t_1}^{t_2} \left[g\Big(G_K\big(s; e^{-\int_t^{t_2} \mu(\tau) d\tau}\big)\Big) - 1\right] dt\right\}, & \textit{for } \mu(t); \end{cases}$$

and $G(s)$ is a pgf. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \textit{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2} \mu(\tau) d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are non-negative integer rv's.

For a specific self-generalized rv K, we know the form of its pgf $G_K(s)$. Thus, we can obtain the full expression of $G(s)$ by Theorem 5.2.1. This leads to the following corollary, where the K is chosen from **I1** to **I5** (non-negative integer case).

For the sake of saving space, we only list the results for the stationary case. The non-stationary case can be straightforward to deduce without any difficulty.

Corollary 5.2.2 *Consider the innovation being the Compound Poisson IIP. In Theorem 5.2.1, by specifying $G_K(s;\alpha)$ or $\phi_K(s;\alpha)$ for the self-generalized rv K, we can get the further form of $g(G_K(s;\alpha))$ or $g(\phi_K(s;\alpha))$. The following are the results for K being from* **I1** *to* **I5** *under the stationary case.*

I1: $$G(s) = \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\left([1-e^{-\mu t}] + e^{-\mu t}s\right) - 1\right] dt\right\}.$$

Furthermore, if $g(s) = \sum\limits_{i=0}^{\infty} p_i s^i$, *then*

$$G(s) = \exp\left\{\frac{\lambda}{\mu}\sum_{j=1}^{\infty}\left(\frac{1}{j}\left[\sum_{i=j}^{\infty}\binom{i}{j}p_i\right][1-e^{-j\mu(t_2-t_1)}](s-1)^j\right)\right\}.$$

I2: $$G(s) = \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\left(\frac{(1-e^{-\mu t}) + (e^{-\mu t}-\gamma)s}{(1-e^{-\mu t}\gamma) - (1-e^{-\mu t})\gamma s}\right) - 1\right] dt\right\}.$$

I3: $$G(s) = \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\left(1-(1-s)^{e^{-\mu t}}\right) - 1\right] dt\right\}.$$

I4: $$G(s) = \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\left(c^{-1}[1-e^{-\theta(1-e^{-\mu t})}(1-cs)^{e^{-\mu t}}]\right) - 1\right] dt\right\}.$$

I5: $$G(s) = \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\left(1-\frac{e^{-\theta\mu t}(1-\gamma)^{\theta}}{\left[(1-e^{-\mu t})\gamma + (1-\gamma)(1-s)^{-\frac{1}{\theta}}\right]^{-\theta}}\right) - 1\right] dt\right\}.$$

For the non-stationary case, just replace the $e^{-\mu t}$ *with* $e^{-\int_t^{t_2}\mu(\tau)d\tau}$.

Proof: The second half part of **I1** needs some details.

$$\begin{aligned}
G(s) &= \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\Big([1-e^{-\mu t}] + e^{-\mu t}s\Big) - 1\right] dt\right\} \\
&= \exp\left\{\lambda \int_0^{t_2-t_1} \left[\sum_{i=0}^{\infty} p_i\Big([1-e^{-\mu t}] + e^{-\mu t}s\Big)^i - 1\right] dt\right\} \\
&= \exp\left\{\lambda \int_0^{t_2-t_1} \sum_{i=1}^{\infty} p_i\left[\Big([1-e^{-\mu t}] + e^{-\mu t}s\Big)^i - 1\right] dt\right\} \\
&= \exp\left\{\lambda \sum_{i=1}^{\infty} p_i\left[\int_0^{t_2-t_1} \Big([1+(s-1)e^{-\mu t}]^i - 1\Big) dt\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \exp\Big\{\lambda\sum_{i=1}^{\infty} p_i\Big[\int_0^{t_2-t_1}\Big(\sum_{j=1}^{i}\binom{i}{j}(s-1)^j e^{-j\mu t} - 1\Big)dt\Big]\Big\} \\
&= \exp\Big\{\lambda\sum_{i=1}^{\infty} p_i\Big[\int_0^{t_2-t_1}\sum_{j=1}^{i}\binom{i}{j}(s-1)^j e^{-j\mu t}dt\Big]\Big\} \\
&= \exp\Big\{\lambda\sum_{i=1}^{\infty} p_i\Big[\sum_{j=1}^{i}\binom{i}{j}(s-1)^j\int_0^{t_2-t_1} e^{-j\mu t}dt\Big]\Big\} \\
&= \exp\Big\{\lambda\sum_{i=1}^{\infty} p_i\Big[\sum_{j=1}^{i}\binom{i}{j}(s-1)^j\frac{[1-e^{-j\mu(t_2-t_1)}]}{j\mu}\Big]\Big\} \\
&= \exp\Big\{\frac{\lambda}{\mu}\sum_{j=1}^{\infty}\Big(\frac{1}{j}\Big[\sum_{i=j}^{\infty}\binom{i}{j}p_i\Big][1-e^{-j\mu(t_2-t_1)}](s-1)^j\Big)\Big\}.
\end{aligned}
$$

By choosing an appropriate pgf $g(s)$, we can find the exact form of the pgf $G(s)$ of interest so that we can obtain the corresponding generalized stochastic integral. The following are a few examples.

Example 5.1 *Consider the compound Poisson IIP innovation. Let $g(s) = s$ and let K be from* **I1**, *so that the extended-thinning operator is binomial-thinning. Then*

$$
\begin{aligned}
G(s) &= \exp\Big\{\lambda\int_0^{t_2-t_1}\Big[g\Big([1-e^{-\mu t}]+e^{-\mu t}s\Big)-1\Big]dt\Big\} \\
&= \exp\Big\{\lambda\int_0^{t_2-t_1}\Big[\Big([1-e^{-\mu t}]+e^{-\mu t}s\Big)-1\Big]dt\Big\} \\
&= \exp\Big\{\lambda\int_0^{t_2-t_1}(s-1)e^{-\mu t}dt\Big\} \\
&= \exp\Big\{\lambda(s-1)\int_0^{t_2-t_1}e^{-\mu t}dt\Big\} \\
&= \exp\Big\{\frac{\lambda}{\mu}[1-e^{-\mu(t_2-t_1)}](s-1)\Big\},
\end{aligned}
$$

or

$$
\begin{aligned}
G(s) &= \exp\Big\{\lambda\int_{t_1}^{t_2}\Big[g\Big([1-e^{-\int_t^{t_2}\mu(\tau)d\tau}]+e^{-\int_t^{t_2}\mu(\tau)d\tau}s\Big)-1\Big]dt\Big\} \\
&= \exp\Big\{\lambda\int_{t_1}^{t_2}\Big[\Big([1-e^{-\int_t^{t_2}\mu(\tau)d\tau}]+e^{-\int_t^{t_2}\mu(\tau)d\tau}s\Big)-1\Big]dt\Big\}
\end{aligned}
$$

$$= \exp\left\{\lambda \int_{t_1}^{t_2} (s-1) e^{-\int_t^{t_2} \mu(\tau)d\tau} dt\right\}$$
$$= \exp\left\{\lambda\Big(\int_{t_1}^{t_2} e^{-\int_t^{t_2} \mu(\tau)d\tau} dt\Big)(s-1)\right\}.$$

These results correspond to the following models:

$$X(t_2) = e^{-\mu(t_2-t_1)} * X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t),$$

where $\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t) \sim Poisson\Big(\frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right]\Big)$, *and*

$$X(t_2) = e^{-\int_{t_1}^{t_2} \mu(\tau)d\tau} * X(t_1) + \int_{t_1}^{t_2} e^{-\int_t^{t_2} \mu(\tau)d\tau} * d\epsilon(t),$$

where $\int_{t_1}^{t_2} e^{-\int_t^{t_2} \mu(\tau)d\tau} * d\epsilon(t) \sim Poisson\left(\lambda \int_{t_1}^{t_2} e^{-\int_t^{t_2} \mu(\tau)d\tau} dt\right)$.

Example 5.2 *Still consider the compound Poisson IIP innovation, and suppose K remains in* **I1**. *Now we choose $g(s) = \frac{1-\gamma}{1-\gamma s}$, i.e., the pgf of NB$(1,\gamma)$. Then for the stationary case,*

$$\begin{aligned}
G(s) &= \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\Big([1-e^{-\mu t}] + e^{-\mu t}s\Big) - 1\right] dt\right\} \\
&= \exp\left\{\lambda \int_0^{t_2-t_1} \left[\frac{1-\gamma}{1-\gamma\Big([1-e^{-\mu t}]+e^{-\mu t}s\Big)} - 1\right] dt\right\} \\
&= \exp\left\{\lambda \int_0^{t_2-t_1} \frac{\gamma(s-1)e^{-\mu t}}{(1-\gamma)-\gamma(s-1)e^{-\mu t}} dt\right\} \\
&= \exp\left\{\frac{\lambda}{\mu} \log\left[(1-\gamma)-\gamma(s-1)e^{-\mu t}\right]\Big|_0^{t_2-t_1}\right\} \\
&= \exp\left\{\frac{\lambda}{\mu} \log\left[\frac{(1-\gamma)-\gamma(s-1)e^{-\mu(t_2-t_1)}}{(1-\gamma)-\gamma(s-1)}\right]\right\} \\
&= \left(\frac{(1-\gamma)-\gamma(s-1)e^{-\mu(t_2-t_1)}}{1-\gamma s}\right)^{\lambda/\mu} \\
&= \left(e^{-\mu(t_2-t_1)} + [1-e^{-\mu(t_2-t_1)}]\frac{1-\gamma}{1-\gamma s}\right)^{\lambda/\mu}.
\end{aligned}$$

This leads to the model

$$X(t_2) = e^{-\mu(t_2-t_1)} * X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t),$$

where $\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t)$ *has pgf* $\left(e^{-\mu(t_2-t_1)} + [1-e^{-\mu(t_2-t_1)}]\frac{1-\gamma}{1-\gamma s}\right)^{\lambda/\mu}$.

For the non-stationary case, it follows that

$$\begin{aligned}
G(s) &= \exp\left\{\lambda \int_{t_1}^{t_2} \left[g\left([1-e^{-\int_t^{t_2}\mu(\tau)d\tau}] + e^{-\int_t^{t_2}\mu(\tau)d\tau}s\right) - 1\right]dt\right\} \\
&= \exp\left\{\lambda \int_{t_1}^{t_2} \left[\frac{1-\gamma}{1-\gamma\left([1-e^{-\int_t^{t_2}\mu(\tau)d\tau}] + e^{-\int_t^{t_2}\mu(\tau)d\tau}s\right)} - 1\right]dt\right\} \\
&= \exp\left\{\lambda \int_{t_1}^{t_2} \frac{\gamma(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}{1-\gamma-\gamma(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}dt\right\}.
\end{aligned}$$

Hence, we have model

$$X(t_2) = e^{-\int_{t_1}^{t_2}\mu(\tau)d\tau} * X(t_1) + \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} * d\epsilon(t),$$

where $\int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} * d\epsilon(t)$ *has pgf* $\exp\left\{\lambda \int_{t_1}^{t_2} \frac{\gamma(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}{1-\gamma-\gamma(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}dt\right\}$.

Example 5.3 *Let* $g(s) = \frac{1}{s} - \frac{\theta^{-1}(1-s)^{1/\theta}}{1-(1-s)^{1/\theta}}$ $(\theta \geq 1)$. *Consider the compound Poisson IIP innovation, and suppose* K *remains in* **I1**. *Then for the stationary case,*

$$\begin{aligned}
G(s) &= \exp\left\{\lambda \int_0^{t_2-t_1} \left[g\left([1-e^{-\mu t}] + e^{-\mu t}s\right) - 1\right]dt\right\} \\
&= \exp\left\{\lambda \int_0^{t_2-t_1} \left[\frac{1}{1+(s-1)e^{-\mu t}} - \frac{\theta^{-1}(1-s)^{1/\theta}e^{-\mu t/\theta}}{1-(1-s)^{1/\theta}e^{-\mu t/\theta}} - 1\right]dt\right\} \\
&= \exp\left\{\lambda \int_0^{t_2-t_1} \left[\frac{-(s-1)e^{-\mu t}}{1+(s-1)e^{-\mu t}} - \frac{\theta^{-1}(1-s)^{1/\theta}e^{-\mu t/\theta}}{1-(1-s)^{1/\theta}e^{-\mu t/\theta}}\right]dt\right\} \\
&= \exp\left\{\frac{\lambda}{\mu}\left[\log\left[1+(s-1)e^{-\mu t}\right]\Big|_0^{t_2-t_1} - \log\left[1-(1-s)^{1/\theta}e^{-\mu t/\theta}\right]\Big|_0^{t_2-t_1}\right]\right\} \\
&= \left(\frac{1-(1-s)^{1/\theta}}{s} \times \frac{1+(s-1)e^{-\mu(t_2-t_1)}}{1-(1-s)^{1/\theta}e^{-\mu(t_2-t_1)/\theta}}\right)^{\lambda/\mu},
\end{aligned}$$

which is the pgf of $\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t)$. *Thus, the model is*

$$X(t_2) = e^{-\mu(t_2-t_1)} * X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t).$$

λ *may take the same value as* μ. *In this case,* $G(s) = s^{-1}\left[1-(1-s)^{1/\theta}\right]$ *(as* $t_2 - t_1 \to \infty$*), which is just the pgf of power series distribution (this is an alternative way to show its DSD property).*

For the non-stationary case, the pgf becomes

$$\begin{aligned} G(s) &= \exp\left\{\lambda \int_{t_1}^{t_2}\left[g\left([1-e^{-\int_t^{t_2}\mu(\tau)d\tau}] + e^{-\int_t^{t_2}\mu(\tau)d\tau}s\right) - 1\right]\right\} \\ &= \exp\left\{\lambda \int_{t_1}^{t_2}\left[\frac{1}{1+(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}} - \frac{\theta^{-1}(1-s)^{1/\theta}e^{-\theta^{-1}\int_t^{t_2}\mu(\tau)d\tau}}{1-(1-s)^{1/\theta}e^{-\theta^{-1}\int_t^{t_2}\mu(\tau)d\tau}} - 1\right]dt\right\} \\ &= \exp\left\{\lambda \int_{t_1}^{t_2}\left[\frac{-(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}{1+(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}} - \frac{\theta^{-1}(1-s)^{1/\theta}e^{-\theta^{-1}\int_t^{t_2}\mu(\tau)d\tau}}{1-(1-s)^{1/\theta}e^{-\theta^{-1}\int_t^{t_2}\mu(\tau)d\tau}}\right]dt\right\}, \end{aligned}$$

the pgf of $\int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} * d\epsilon(t)$*, and the model is*

$$X(t_2) = e^{-\int_{t_1}^{t_2}\mu(\tau)d\tau} * X(t_1) + \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} * d\epsilon(t).$$

Example 5.4 *Suppose innovation remains as compound Poisson IIP, and* $g(s) = s$. *Now consider* K *from* **I2** *with pgf* $G_K(s;\alpha) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}$, *then*

$$G_K(s;\alpha) - 1 = \frac{\alpha(1-\gamma)(s-1)}{(1-\alpha\gamma)-(1-\alpha)\gamma s},$$

and for the stationary case,

$$\begin{aligned} G(s) &= \exp\left\{\lambda \int_0^{t_2-t_1}\left[g\Big(\frac{(1-e^{-\mu t})+(e^{-\mu t}-\gamma)s}{(1-e^{-\mu t}\gamma)-(1-e^{-\mu t})\gamma s}\Big) - 1\right]dt\right\} \\ &= \exp\left\{\lambda \int_0^{t_2-t_1}\frac{(1-\gamma)(s-1)e^{-\mu t}}{(1-\gamma e^{-\mu t})-(1-e^{-\mu t})\gamma s}dt\right\} \\ &= \exp\left\{\lambda \int_0^{t_2-t_1}\frac{(1-\gamma)(s-1)e^{-\mu t}}{(1-\gamma s)+\gamma(s-1)e^{-\mu t}}dt\right\} \\ &= \exp\left\{-\frac{\lambda}{\mu}\frac{1-\gamma}{\gamma}\log\left[(1-\gamma s)+\gamma(s-1)e^{-\mu t}\right]\Big|_0^{t_2-t_1}\right\} \\ &= \exp\left\{\frac{\lambda}{\mu}\frac{1-\gamma}{\gamma}\log\left[\frac{1-\gamma}{(1-\gamma s)+\gamma(s-1)e^{-\mu(t_2-t_1)}}\right]\right\} \\ &= \left(\frac{1-\gamma}{(1-\gamma s)+\gamma(s-1)e^{-\mu(t_2-t_1)}}\right)^{\frac{\lambda}{\mu}\frac{1-\gamma}{\gamma}} \\ &= \left(\frac{\frac{1-\gamma}{1-\gamma e^{-\mu(t_2-t_1)}}}{1-\frac{\gamma(1-e^{-\mu(t_2-t_1)})}{1-\gamma e^{-\mu(t_2-t_1)}}s}\right)^{\lambda(1-\gamma)/(\mu\gamma)} = \left(\frac{1-\frac{\gamma(1-e^{-\mu(t_2-t_1)})}{1-\gamma e^{-\mu(t_2-t_1)}}}{1-\frac{\gamma(1-e^{-\mu(t_2-t_1)})}{1-\gamma e^{-\mu(t_2-t_1)}}s}\right)^{\lambda(1-\gamma)/(\mu\gamma)}. \end{aligned}$$

Note that this last form is a NB pgf. Hence the model is

$$X(t_2) = \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t),$$

where $\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \sim NB\Big(\frac{\lambda(1-\gamma)}{\mu\gamma}, \frac{\gamma(1-e^{-\mu(t_2-t_1)})}{1-\gamma e^{-\mu(t_2-t_1)}}\Big)$.

Similarly, the model for the non-stationary case is

$$X(t_2) = \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast X(t_1) + \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t),$$

where $\int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$ *has pgf*

$$\begin{aligned} G(s) &= \exp\Big\{\lambda\int_{t_1}^{t_2}\Big[g\Big(\frac{(1-e^{-\int_t^{t_2}\mu(\tau)d\tau}) + (e^{-\int_t^{t_2}\mu(\tau)d\tau}-\gamma)s}{(1-e^{-\int_t^{t_2}\mu(\tau)d\tau}\gamma) - (1-e^{-\int_t^{t_2}\mu(\tau)d\tau})\gamma s}\Big) - 1\Big]dt\Big\} \\ &= \exp\Big\{\lambda\int_0^{t_2-t_1}\frac{(1-\gamma)(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}{(1-\gamma s)+\gamma(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}dt\Big\}. \end{aligned}$$

Example 5.5 *Keep the previous stochastic setting, but choose* $g(s) = \frac{1-\beta}{1-\beta s}$ $(0 < \beta < 1)$. *Then,*

$$\begin{aligned} g(G_K(s,\alpha)) - 1 &= \frac{1-\beta}{1-\beta\frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}} - 1 \\ &= \frac{(1-\beta)[(1-\gamma s)-\gamma(1-s)\alpha]}{(1-\beta)(1-\gamma s)+(\beta-\gamma)(1-s)\alpha} - 1 \\ &= \frac{-\beta(1-\gamma)(1-s)\alpha}{(1-\beta)(1-\gamma s)+(\beta-\gamma)(1-s)\alpha}. \end{aligned}$$

For the stationary case,

$$\begin{aligned} G(s) &= \exp\Big\{\lambda\int_0^{t_2-t_1}\Big[g\Big(\frac{(1-e^{-\mu t})+(e^{-\mu t}-\gamma)s}{(1-e^{-\mu t}\gamma)-(1-e^{-\mu t})\gamma s}\Big)-1\Big]dt\Big\} \\ &= \exp\Big\{\lambda\int_0^{t_2-t_1}\frac{-\beta(1-\gamma)(1-s)e^{-\mu t}}{(1-\beta)(1-\gamma s)+(\beta-\gamma)(1-s)e^{-\mu t}}dt\Big\} \\ &= \exp\Big\{\frac{\lambda\beta(1-\gamma)}{\mu(\beta-\gamma)}\log\left[(1-\beta)(1-\gamma s)+(\beta-\gamma)(1-s)e^{-\mu t}\right]\Big|_0^{t_2-t_1}\Big\} \\ &= \exp\Big\{\frac{\lambda\beta(1-\gamma)}{\mu(\beta-\gamma)}\log\frac{(1-\beta)(1-\gamma s)+(\beta-\gamma)(1-s)e^{-\mu(t_2-t_1)}}{(1-\beta)(1-\gamma s)+(\beta-\gamma)(1-s)}\Big\} \\ &= \Big(\frac{(1-\beta)(1-\gamma s)+(\beta-\gamma)(1-s)e^{-\mu(t_2-t_1)}}{(1-\beta)(1-\gamma s)+(\beta-\gamma)(1-s)}\Big)^{\frac{\lambda\beta(1-\gamma)}{\mu(\beta-\gamma)}}, \end{aligned}$$

and the model is

$$X(t_2) = \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t),$$

where $\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$ *has the above* $G(s)$ *as its pgf.*

For the non-stationary case,

$$\begin{aligned} G(s) &= \exp\Big\{\lambda \int_{t_1}^{t_2} \Big[g\Big(\frac{(1-e^{-\int_t^{t_2}\mu(\tau)d\tau}) + (e^{-\int_t^{t_2}\mu(\tau)d\tau} - \gamma)s}{(1-e^{-\int_t^{t_2}\mu(\tau)d\tau}\gamma) - (1-e^{-\int_t^{t_2}\mu(\tau)d\tau})\gamma s}\Big) - 1\Big]dt\Big\} \\ &= \exp\Big\{\lambda \int_0^{t_2-t_1} \frac{-\beta(1-\gamma)(1-s)e^{-\int_t^{t_2}\mu(\tau)d\tau}}{(1-\beta)(1-\gamma s) + (\beta-\gamma)(1-s)e^{-\int_t^{t_2}\mu(\tau)d\tau}} dt\Big\}, \end{aligned}$$

and corresponding model becomes

$$X(t_2) = \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast X(t_1) + \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t),$$

where the second $G(s)$ *is the pgf of* $\int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$.

Secondly, we consider GNBC IIP as innovation processes. In this case, we have following theorem, from Theorem 5.1.2.

Theorem 5.2.3 *Let* $\{\epsilon(t); t \geq 0\}$ *be a GNBC IIP with pgf of* $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$

$$G_{\Delta\epsilon}(s) = \exp\left\{h \int_{(0,1)} \log\left(\frac{p}{1-qs}\right) V(dq)\right\}.$$

Suppose $K(\alpha)$ *is a non-negative integer self-generalized rv with pgf* $G_K(s;\alpha)$, *which is differentiable with bounded first order derivative. Then, it follows that*

$$G(s) = \begin{cases} \exp\left\{\int_{(0,1)} \left(\int_0^{t_2-t_1} \log\left[\frac{p}{1-qG_K(s;e^{-\mu t})}\right] dt\right) V(dq)\right\}, & \text{for constant } \mu; \\ \exp\left\{\int_{(0,1)} \left(\int_{t_1}^{t_2} \log\left[\frac{p}{1-qG_K\left(s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)}\right] dt\right) V(dq)\right\}, & \text{for } \mu(t), \end{cases}$$

and $G(s)$ *is a pgf. Hence, the generalized stochastic integrals*

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are non-negative integer rv's.

Proof: Omitted.

Consequently, we have derived the following corollary by direct calculation.

Corollary 5.2.4 *Here we consider the specific self-generalized rv K given in* **I1** *to* **I5**. *The innovation process is the GNBC IIP.*

I1:

$$G(s)=\begin{cases}\exp\left\{(t_2-t_1)\int_{(0,1)}\log\left(\frac{p}{1-qs}\right)V(dq)\right.\\ \qquad\left.+\frac{\mu(t_2-t_1)^2}{2}\int_{(0,1)}\frac{q(1-s)}{1-qs}V(dq)\right\} & \textit{for constant } \mu;\\ \exp\left\{(t_2-t_1)\int_{(0,1)}\log\left(\frac{p}{1-qs}\right)V(dq)\right.\\ \qquad\left.+\left(\int_{t_1}^{t_2}\left[\int_t^{t_2}\mu(\tau)d\tau\right]dt\right)\int_{(0,1)}\frac{q(1-s)}{1-qs}V(dq)\right\} & \textit{for } \mu(t).\end{cases}$$

I2:

$$G(s)=\begin{cases}\exp\left\{(t_2-t_1)\int_{(0,1)}\log\left(\frac{p}{1-qs}\right)V(dq)\right.\\ \qquad\left.+\frac{\mu(t_2-t_1)^2}{2}\int_{(0,1)}\left[\frac{q}{1-qs}\frac{(1-s)(1-\gamma s)}{1-\gamma}\right]V(dq)\right\} & \textit{for constant } \mu;\\ \exp\left\{(t_2-t_1)\int_{(0,1)}\log\left(\frac{p}{1-qs}\right)V(dq)\right.\\ \qquad\left.+\left(\int_{t_1}^{t_2}\left[\int_t^{t_2}\mu(\tau)d\tau\right]dt\right)\int_{(0,1)}\left[\frac{q}{1-qs}\frac{(1-s)(1-\gamma s)}{1-\gamma}\right]V(dq)\right\} & \textit{for } \mu(t).\end{cases}$$

I3:

$$G(s)=\begin{cases}\exp\left\{(t_2-t_1)\int_{(0,1)}\log\left(\frac{p}{1-qs}\right)V(dq)\right.\\ \qquad\left.+\frac{\mu(t_2-t_1)^2}{2}\int_{(0,1)}\frac{q(1-s)\log(1-s)}{1-qs}V(dq)\right\} & \textit{for constant } \mu;\\ \exp\left\{(t_2-t_1)\int_{(0,1)}\log\left(\frac{p}{1-qs}\right)V(dq)\right.\\ \qquad\left.+\left(\int_{t_1}^{t_2}\left[\int_t^{t_2}\mu(\tau)d\tau\right]dt\right)\int_{(0,1)}\frac{q(1-s)\log(1-s)}{1-qs}V(dq)\right\} & \textit{for } \mu(t).\end{cases}$$

I4:

$$G(s)=\begin{cases}\exp\left\{(t_2-t_1)\int_{(0,1)}\log\left(\frac{p}{1-qs}\right)V(dq)\right.\\ \qquad\left.+\frac{\mu(t_2-t_1)^2}{2}\int_{(0,1)}\frac{qc^{-1}(1-cs)[\theta+\log(1-cs)]}{1-qs}V(dq)\right\} & \textit{for constant } \mu;\\ \exp\left\{(t_2-t_1)\int_{(0,1)}\log\left(\frac{p}{1-qs}\right)V(dq)\right.\\ \qquad\left.+\left(\int_{t_1}^{t_2}\left[\int_t^{t_2}\mu(\tau)d\tau\right]dt\right)\int_{(0,1)}\frac{qc^{-1}(1-cs)[\theta+\log(1-cs)]}{1-qs}V(dq)\right\} & \textit{for } \mu(t).\end{cases}$$

I5:

$$G(s) = \begin{cases} \exp\Big\{(t_2-t_1)\int_{(0,1)} \log\left(\frac{p}{1-qs}\right) V(dq) \\ \qquad + \frac{\mu(t_2-t_1)^2}{2} \int_{(0,1)} \frac{q(1-\gamma)^{\beta-\theta}(1-s)[\beta+\theta\frac{\gamma}{1-\gamma}(1-s)^{\frac{1}{\theta}}]}{1-qs} V(dq)\Big\} & \textit{for constant } \mu; \\ \exp\Big\{(t_2-t_1)\int_{(0,1)} \log\left(\frac{p}{1-qs}\right) V(dq) + \left(\int_{t_1}^{t_2}\left[\int_t^{t_2}\mu(\tau)d\tau\right]dt\right) \\ \qquad \times \int_{(0,1)} \frac{q(1-\gamma)^{\beta-\theta}(1-s)[\beta+\theta\frac{\gamma}{1-\gamma}(1-s)^{\frac{1}{\theta}}]}{1-qs} V(dq)\Big\} & \textit{for } \mu(t). \end{cases}$$

However, usually the measure $V(\cdot)$ on $(0,1)$ is not clear. Hence, Corollary 5.2.4 is not helpful in obtaining the pgf or LT of the generalized stochastic integral. For each specific member in the GNBC family, we have to calculate the pgf or LT individually. Following are some examples resulting from the GNBC IIP innovation family.

Example 5.6 *Consider the NB IIP innovation, in which the increment* $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ *has pgf*

$$G_{\Delta\epsilon}(s) = \left(\frac{p}{1-qs}\right)^{h\theta},$$

where $p, q > 0, p+q = 1$. *This is a special case of the GNBC when* $V(\cdot)$ *has a mass of* θ *at a single value* q.

When K *is from* **I1**, *we obtain*

$$G(s) = \begin{cases} \exp\left\{\theta \int_0^{t_2-t_1} \log\left[\frac{p}{p+q(s-1)e^{-\mu t}}\right] dt\right\}, & \textit{for constant } \mu; \\ \exp\left\{\theta \int_{t_1}^{t_2} \log\left[\frac{p}{p+q(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}\right] dt\right\}, & \textit{for } \mu(t). \end{cases}$$

which leads to the models

$$X(t_2) = e^{-\mu(t_2-t_1)} * X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t),$$

and

$$X(t_2) = e^{-\int_{t_1}^{t_2}\mu(\tau)d\tau} * X(t_1) + \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} * d\epsilon(t).$$

for the stationary and non-stationary case respectively.

When K is from **I2**, *then*

$$G(s)=\begin{cases}\exp\left\{\theta\int_0^{t_2-t_1}\log\left[\dfrac{p(1-\gamma s)+p\gamma(s-1)e^{-\mu t}}{p(1-\gamma s)+(\gamma-q)(s-1)e^{-\mu t}}\right]dt\right\}, & \text{for constant } \mu;\\ \\ \exp\left\{\theta\int_{t_1}^{t_2}\log\left[\dfrac{p(1-\gamma s)+p\gamma(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}{p(1-\gamma s)+(\gamma-q)(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}\right]dt\right\}, & \text{for } \mu(t).\end{cases}$$

The resulting models are

$$X(t_2)=\left(e^{-\mu(t_2-t_1)}\right)_K\circledast X(t_1)+\int_0^{t_2-t_1}\left(e^{-\mu t}\right)_K\circledast d\epsilon(t),$$

and

$$X(t_2)=\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K\circledast X(t_1)+\int_{t_1}^{t_2}\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K\circledast d\epsilon(t).$$

Example 5.7 *Consider discrete stable IIP innovation, in which the increment $\epsilon(t+h)-\epsilon(t)$ has pgf*

$$G_{\Delta\epsilon}(s)=\exp\{-\lambda h(1-s)^{\beta}\},\quad \lambda>0,\quad 0<\beta<1.$$

Case 1: *K is from* **I1**. *For the stationary situation,*

$$\begin{aligned}G(s)&=\lim_{n\to\infty}\prod_{i=0}^{n-1}\exp\{-\lambda h[1-(1-(1-s)(1-\mu h)^i]^{\beta}\}\\&=\lim_{n\to\infty}\exp\{-\lambda\sum_{i=0}^{n-1}h[(1-s)(1-\mu h)^i]^{\beta}\}\\&=\lim_{n\to\infty}\exp\Big\{-\lambda(1-s)^{\beta}\sum_{i=0}^{n-1}he^{-\beta\mu ih}\Big\}\\&=\exp\Big\{-\lambda(1-s)^{\beta}\int_0^{t_2-t_1}e^{-\beta\mu t}dt\Big\}\\&=\exp\Big\{-\frac{\lambda[1-e^{-\beta\mu(t_2-t_1)}]}{\beta\mu}(1-s)^{\beta}\Big\},\end{aligned}$$

which leads to the model

$$X(t_2)=e^{-\mu(t_2-t_1)}*X(t_1)+\int_0^{t_2-t_1}e^{-\mu t}*d\epsilon(t),$$

where $\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t)$ *has pgf* $G(s)$, *and is distributed as discrete stable also.*

For the non-stationary situation,

$$\begin{aligned} G(s) &= \lim_{n\to\infty}\prod_{i=0}^{n-1}\exp\Big\{-\lambda h[1-\Big(1-(1-s)\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\Big)]^{\beta}\Big\} \\ &= \exp\Big\{-\lambda(1-s)^{\beta}\int_{t_1}^{t_2} e^{-\beta\int_t^{t_2}\mu(\tau)d\tau}dt\Big\} \end{aligned}$$

which leads to model

$$X(t_2) = e^{-\int_{t_1}^{t_2}\mu(\tau)d\tau} * X(t_1) + \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} * d\epsilon(t).$$

where $\int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} * d\epsilon(t)$ *remains in discrete stable family.*

Case 2: *K is from* **I2**. *Then, by straightforward calculation,*

$$G(s) = \begin{cases} \exp\Big\{-\lambda(1-\gamma)^{\beta}(1-s)^{\beta}\int_0^{t_2-t_1}\dfrac{\exp(-\beta\mu t)}{[(1-\gamma s)-\gamma(1-s)\exp(-\mu t)]^{\beta}}dt\Big\} & \textit{for constant } \mu; \\ \\ \exp\Bigg\{-\lambda(1-\gamma)^{\beta}(1-s)^{\beta}\int_{t_1}^{t_2}\dfrac{\exp\Big(-\beta\int_t^{t_2}\mu(\tau)d\tau\Big)}{\Big[(1-\gamma s)-\gamma(1-s)\exp\Big(-\int_t^{t_2}\mu(\tau)d\tau\Big)\Big]^{\beta}}dt\Bigg\} & \textit{for } \mu(t). \end{cases}$$

Hence, resulting models are

$$X(t_2) = e^{-\mu(t_2-t_1)} \circledast X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} \circledast d\epsilon(t),$$

and

$$X(t_2) = e^{-\int_{t_1}^{t_2}\mu(\tau)d\tau} \circledast X(t_1) + \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} \circledast d\epsilon(t).$$

corresponding to the stationary and the non-stationary case respectively. In Case 2, the generalized stochastic integrals are not in discrete stable family.

In the remainder of this section, we study two generalized convolution families: GC I and GC II.

Theorem 5.2.5 *Let $\{\epsilon(t); t \geq 0\}$ be a GC I IIP with pgf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ of form*

$$G_{\Delta\epsilon}(s) = \exp\left\{h \int_{(0,1)} \frac{q(s-1)}{1-qs} V(dq)\right\}.$$

Suppose $K(\alpha)$ is a non-negative integer self-generalized rv with pgf $G_K(s;\alpha)$, which is differentiable with bounded first order partial derivative with respect to α. Then, it follows that

$$G(s) = \begin{cases} \exp\left\{\displaystyle\int_{(0,1)} \left(\int_0^{t_2-t_1} \frac{q\left[G_K\left(s;e^{-\mu t}\right)-1\right]}{1-qG_K\left(s;e^{-\mu t}\right)} dt\right) V(dq)\right\}, & \textit{for constant } \mu; \\ \\ \exp\left\{\displaystyle\int_{(0,1)} \left(\int_{t_1}^{t_2} \frac{q\left[G_K\left(s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)-1\right]}{1-qG_K\left(s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)} dt\right) V(dq)\right\}, & \textit{for } \mu(t), \end{cases}$$

and $G(s)$ is a pgf. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \textit{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are non-negative integer rv's.

Proof: Omitted.

As in the case of GNBC innovation processes case, we can obtain the further expression for a specific K. However, it may not be useful since the measure $V(\cdot)$ is not clear. Nevertheless, we use two specific K in the following.

Corollary 5.2.6 *Consider the specific self-generalized rv K from* **I1** *and* **I2**. *The innovation is the GC I IIP. Then, it follows that*

I1:

$$G(s) = \begin{cases} \exp\left\{\frac{1}{\mu}\displaystyle\int_{(0,1)} \log\left(\frac{1-q-q(s-1)e^{-\mu(t_2-t_1)}}{1-qs}\right) V(dq)\right\}, & \textit{for constant } \mu; \\ \\ \exp\left\{\displaystyle\int_{(0,1)} \left[\int_{t_1}^{t_2} \frac{q(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}{1-q-q(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}} dt\right] V(dq)\right\}, & \textit{for } \mu(t). \end{cases}$$

I2:

$$G(s) = \begin{cases} \exp\Big\{ \displaystyle\int_{(0,1)} \Big[\frac{q(1-\gamma)}{\mu(q-\gamma)} \log\Big(\frac{(1-q)(1-\gamma s) + (q-\gamma)(1-s)e^{-\mu(t_2-t_1)}}{(1-\gamma)(1-qs)}\Big)\Big] V(dq)\Big\}, \\ \hfill \textit{for constant } \mu; \\ \\ \exp\Big\{ \displaystyle\int_{(0,1)} \Big[\int_{t_1}^{t_2} \frac{q(1-\gamma)(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}}{(1-q)(1-\gamma s) + (q-\gamma)(1-s)e^{-\int_t^{t_2}\mu(\tau)d\tau}} dt\Big] V(dq)\Big\}, \\ \hfill \textit{for } \mu(t). \end{cases}$$

Proof: Omitted.

Example 5.8 *Example 5.2 can be revisited as one member of the GC I family.*

Theorem 5.2.7 *Let $\{\epsilon(t); t \geq 0\}$ be a GC II IIP with pgf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ of form*

$$G_{\Delta\epsilon}(s) = \exp\Big\{ h \int_{(0,1)} \frac{-q(1-s)(1-\gamma s)}{1-qs} V(dq)\Big\}.$$

Suppose $K(\alpha)$ is a non-negative integer self-generalized rv with pgf $G_K(s;\alpha)$, which is differentiable with bounded first order partial derivatives with respect to α. Then, it follows that

$$G(s) = \begin{cases} \exp\left\{ \displaystyle\int_{[\gamma,1)} \left(\int_0^{t_2-t_1} \frac{-q\left[1 - G_K\left(s; e^{-\mu t}\right)\right]\left[1 - \gamma G_K\left(s; e^{-\mu t}\right)\right]}{1 - qG_K\left(s; e^{-\mu t}\right)} dt \right) V(dq) \right\}, \\ \hfill \textit{for constant } \mu; \\ \\ \exp\left\{ \displaystyle\int_{[\gamma,1)} \left(\int_{t_1}^{t_2} \frac{-q\left[1 - G_K\left(s; e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)\right]\left[1 - \gamma G_K\left(s; e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)\right]}{1 - qG_K\left(s; e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)} dt \right) V(dq) \right\}, \\ \hfill \textit{for } \mu(t), \end{cases}$$

and $G(s)$ is a pgf. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \textit{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are non-negative integer rv's.

Proof: Omitted.

Like previous innovation cases, we can specify the self-generalized rv K to obtain the complete expressions of pgf or LT of generalized stochastic integral. However, most of expressions involve in integration over time t are not simply expressed. Hence, we only give two self-generalized rv K for GC II innovation in the following corollary.

Corollary 5.2.8 *Consider the two special self-generalized rv's discussed in* **I1** *and* **I2**, *and the innovation is the GC II IIP.*

I1:

$$G(s) = \begin{cases} \exp\left\{ \frac{\gamma}{\mu}\left(1 - e^{-\mu(t_2 - t_1)}\right)(s-1)\int_{[\gamma,1)} V(dq) + \right. \\ \qquad \left. \frac{1}{\mu}\int_{[\gamma,1)} \left(\frac{1-\gamma}{q} \log \frac{1 - q - q(s-1)e^{-\mu(t_2-t_1)}}{1-qs} \right) V(dq) \right\}, \\ \hfill \textit{for constant } \mu; \\ \exp\left\{ \int_{[\gamma,1)} \left(\int_{t_1}^{t_2} \frac{q(s-1)[(1-\gamma) - \gamma(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}]e^{-\int_t^{t_2}\mu(\tau)d\tau}}{1 - q - q(s-1)e^{-\int_t^{t_2}\mu(\tau)d\tau}} dt \right) V(dq) \right\}, \\ \hfill \textit{for } \mu(t). \end{cases}$$

I2:

$$G(s) = \begin{cases} \exp\left\{ \frac{1}{\mu}\int_{[\gamma,1)} \left[\log\left(\frac{(1-q)(1-\gamma s) + (q-\gamma)(1-s)e^{-\mu(t_2-t_1)}}{(1-q)(1-\gamma s) - \gamma(1-q)(1-s)e^{-\mu(t_2-t_1)}} \right) + \log\frac{1-q}{1-qs} \right] V(dq) \right\}, \\ \hfill \textit{for constant } \mu; \\ \exp\left\{ (1-\gamma)(1-\gamma s)\int_{[\gamma,1)} \left[(1-q)\int_{t_1}^{t_2} \left(\frac{1}{(1-q)(1-\gamma s) + (q-\gamma)(1-s)e^{-\int_t^{t_2}\mu(\tau)d\tau}} \right.\right.\right. \\ \qquad \left.\left.\left. - \frac{1}{(1-q)(1-\gamma s) - \gamma(1-q)(1-s)e^{-\int_t^{t_2}\mu(\tau)d\tau}} \right) dt \right] V(dq) \right\}, \\ \hfill \textit{for } \mu(t). \end{cases}$$

Proof: Omitted.

5.3 Positive-valued innovation processes and examples

For positive innovation processes, we choose the compound Poisson (with a distribution with positive support) IIP, Generalized Gamma Convolution IIP and GCMED IIP. The LT's of the increment in these three kind processes are of exponential form. The families of Generalized Gamma Convolution and GCMED include many distributions having domain on $(0,\infty)$. Hence, these lead to many continuous-time GAR(1) processes with positive-valued margins.

The results are summarized by Theorem 5.3.1, 5.3.2 and 5.3.3 in the following.

Theorem 5.3.1 *Let $\{\epsilon(t); t \geq 0\}$ be a compound Poisson IIP with pgf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$*

$$\phi_{\Delta\epsilon}(s) = \exp\left\{\lambda h[\phi_0(s) - 1]\right\},$$

where $\phi_0(s)$ is a LT, and differentiable with bounded first order derivative. Suppose $K(\alpha)$ is a positive self-generalized rv with LT $\phi_K(s;\alpha)$, which is differentiable with bounded first order derivative with respect to α. Then it follows that

$$\phi(s) = \begin{cases} \exp\left\{\lambda \int_0^{t_2-t_1} \left[\phi_0\Big(-\log\phi_K\big(s; e^{-\mu t}\big)\Big) - 1\right] dt\right\}, & \text{for constant } \mu; \\ \\ \exp\left\{\lambda \int_{t_1}^{t_2} \left[\phi_0\Big(-\log\phi_K\big(s; e^{-\int_t^{t_2}\mu(\tau)d\tau}\big)\Big) - 1\right] dt\right\}, & \text{for } \mu(t), \end{cases}$$

and $\phi(s)$ is a LT. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are positive rv's.

Proof: It is straightforwardly derived by Theorem 5.1.2.

Example 5.9 *Consider the innovation being a compound Poisson with Gamma IIP, in which the increment $\epsilon(t+h) - \epsilon(t)$ has LT*

$$\phi_{\Delta\epsilon}(s) = \exp\left\{\lambda h\left[\left(\frac{\theta}{\theta+s}\right)^{\gamma} - 1\right]\right\}, \quad \lambda, \theta, \gamma > 0.$$

Choose K a positive self-generalized rv. By (2) of Theorem 5.3.1, we have

$$\phi(s) = \begin{cases} \exp\left\{\lambda \int_0^{t_2-t_1} \left[\left(\frac{\theta}{\theta - \log\phi_K\left(s; e^{-\mu t}\right)}\right)^{\gamma} - 1\right] dt\right\}, & \text{for constant } \mu; \\ \exp\left\{\lambda \int_{t_1}^{t_2} \left[\left(\frac{\theta}{\theta - \log\phi_K\left(s; e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)}\right)^{\gamma} - 1\right] dt\right\}, & \text{for } \mu(t). \end{cases}$$

Let K be from **P1**, *then*

$$\phi(s) = \begin{cases} \exp\left\{\lambda \int_0^{t_2-t_1} \left[\left(\frac{\theta}{\theta + se^{-\mu t}}\right)^{\gamma} - 1\right] dt\right\}, & \text{for constant } \mu; \\ \exp\left\{\lambda \int_{t_1}^{t_2} \left[\left(\frac{\theta}{\theta + se^{-\int_t^{t_2}\mu(\tau)d\tau}}\right)^{\gamma} - 1\right] dt\right\}, & \text{for } \mu(t). \end{cases}$$

When $\gamma = 1$ and for constant μ, this becomes

$$\phi(s) = \left(\frac{\theta + e^{-\mu(t_2-t_1)}s}{\theta + s}\right)^{\lambda/\mu}.$$

Let K be from **P2**, *then*

$$\phi(s) = \begin{cases} \exp\left\{\lambda \int_0^{t_2-t_1} \left[\left(1 - \frac{(1-\gamma)se^{-\mu t}}{\theta(1-\gamma) + \theta\gamma s(1-e^{-\mu t})}\right)^{-\gamma} - 1\right] dt\right\}, & \text{for constant } \mu; \\ \exp\left\{\lambda \int_{t_1}^{t_2} \left[\left(1 - \frac{(1-\gamma)se^{-\int_t^{t_2}\mu(\tau)d\tau}}{\theta(1-\gamma) + \theta\gamma s\left(1 - e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)}\right)^{-\gamma} - 1\right] dt\right\}, & \text{for } \mu(t). \end{cases}$$

When $\gamma = 1$ and for constant μ,

$$\phi(s) = \left(\frac{1 + \left[\frac{\gamma}{1-\gamma} + e^{-\mu(t_2-t_1)}\frac{1-\gamma-\theta\gamma}{\theta(1-\gamma)}\right]s}{1 + \frac{s}{\theta}}\right)^{\frac{\lambda}{\mu}\frac{1-\gamma}{1-\gamma-\theta\gamma}}.$$

Theorem 5.3.2 *Let $\{\epsilon(t); t \geq 0\}$ be a GGC IIP with LT of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$*

$$\phi_{\Delta\epsilon}(s) = \exp\left\{h \int_{(0,\infty)} \log\left(\frac{u}{u+s}\right) U(du)\right\},$$

where $U(du)$ is a non-negative measure on $(0,\infty)$ satisfying

$$\int_{(0,1]} |\log u| U(du) < \infty, \quad \text{and} \quad \int_{(1,\infty)} u^{-1} U(du) < \infty.$$

Suppose $K(\alpha)$ is a positive self-generalized rv with LT $\phi_K(s;\alpha)$, which is differentiable with bounded first order partial derivative with respect to α. Then it follows that

$$\phi(s) = \begin{cases} \exp\left\{\int_{(0,\infty)} \left[\int_0^{t_2-t_1} \log\left(\frac{u}{u - \log\phi_K\left(s; e^{-\mu t}\right)}\right) dt\right] U(du)\right\}, & \text{for constant } \mu; \\ \exp\left\{\int_{(0,\infty)} \left[\int_{t_1}^{t_2} \log\left(\frac{u}{u - \log\phi_K\left(s; e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)}\right) dt\right] U(du)\right\}, & \text{for } \mu(t), \end{cases}$$

and $\phi(s)$ is a LT. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are positive rv's.

Proof: Omitted.

Similar to GNBC and other generalized convolutions innovation situations, we have to calculate the LT for specific member in GGC family by Theorem 5.1.2 or 5.1.3.

Example 5.10 *Consider the Gamma IIP innovation, in which the increment $\epsilon(t+h) - \epsilon(t)$ has LT*

$$\phi_{\Delta\epsilon}(s) = \left(\frac{1}{1+\beta s}\right)^{\gamma h} = \exp\{-\gamma h \log(1+\beta s)\},$$

where $\alpha, \beta > 0$. Let K be from **P1**. *By Theorem 5.1.2, we have*

$$\phi(s) = \begin{cases} \exp\left\{-\gamma \int_0^{t_2-t_1} \log\left(1 + \beta s e^{-\mu t}\right) dt\right\}, & \text{for constant } \mu; \\ \exp\left\{-\gamma \int_{t_1}^{t_2} \log\left(1 + \beta s e^{-\int_t^{t_2}\mu(\tau)d\tau}\right) dt\right\}, & \text{for } \mu(t), \end{cases}$$

which seems to be in the GGC family.

Example 5.11 *Consider the inverse Gaussian IIP innovation, in which the increment $\epsilon(t+h)-\epsilon(t)$ has LT*

$$\phi_{\Delta\epsilon}(s) = \exp\left\{\gamma h\left[1-\left(1+2\gamma^{-1}s\right)^{1/2}\right]\right\},$$

where $\gamma > 0$. Let K be from **P1**. *Then, we have*

$$\phi(s) = \begin{cases} \exp\left\{\int_0^{t_2-t_1} \gamma\left[1-\left(1+2\gamma^{-1}se^{-\mu t}\right)^{1/2}\right]dt\right\}, & \text{for constant } \mu; \\ \exp\left\{\int_{t_1}^{t_2} \gamma\left[1-\left(1+2\gamma^{-1}se^{-\int_t^{t_2}\mu(\tau)d\tau}\right)^{1/2}\right]dt\right\}, & \text{for } \mu(t). \end{cases}$$

Theorem 5.3.3 *Let $\{\epsilon(t); t \geq 0\}$ be a GCMED IIP with LT of $\Delta\epsilon = \epsilon(t+h)-\epsilon(t)$*

$$\phi_{\Delta\epsilon}(s) = \exp\left\{h\int_{(0,\infty)} \frac{-s}{u+s}U(du)\right\},$$

where $U(du)$ is a non-negative measure on $(0,\infty)$ satisfying $\int_{(0,\infty)} u^{-1}U(du) < \infty$. Suppose $K(\alpha)$ is a positive self-generalized rv with LT $\phi_K(s;\alpha)$, which is differentiable with bounded first order partial derivative with respect to α. Then it follows that

$$\phi(s) = \begin{cases} \exp\left\{\int_{(0,\infty)}\left[\int_0^{t_2-t_1} \frac{\log\phi_K\left(s;e^{-\mu t}\right)}{u-\log\phi_K\left(s;e^{-\mu t}\right)}dt\right]U(du)\right\}, & \text{for constant } \mu; \\ \exp\left\{\int_{(0,\infty)}\left[\int_{t_1}^{t_2} \frac{\log\phi_K\left(s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)}{u-\log\phi_K\left(s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)}dt\right]U(du)\right\}, & \text{for } \mu(t), \end{cases}$$

and $\phi(s)$ is a LT. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are positive rv's.

Proof: Omitted.

For K from **P1**, we can calculate the LT under the stationary case:

$$\begin{aligned}\phi(s) &= \exp\Big\{\int_{(0,\infty)}\Big[\int_0^{t_2-t_1}\frac{-se^{-\mu t}}{u+se^{-\mu t}}dt\Big]U(du)\Big\}\\ &= \exp\Big\{\int_{(0,\infty)}\Big[\frac{1}{\mu}\log(u+e^{-\mu t})\Big|_0^{t_2-t_1}\Big]U(du)\Big\}\\ &= \exp\Big\{\frac{1}{\mu}\int_{(0,\infty)}\log\Big(\frac{u+se^{-\mu(t_2-t_1)}}{u+s}\Big)U(du)\Big\}.\end{aligned}$$

When $t_2-t_1\to\infty$,

$$\exp\Big\{\frac{1}{\mu}\int_{(0,\infty)}\log\Big(\frac{u+se^{-\mu(t_2-t_1)}}{u+s}\Big)U(du)\Big\}\longrightarrow\exp\Big\{\frac{1}{\mu}\int_{(0,\infty)}\log\Big(\frac{u}{u+s}\Big)U(du)\Big\},$$

which is the GGC class. For K from **P2**, we can calculate the LT under the stationary case:

$$\begin{aligned}\phi(s) &= \exp\Big\{\int_{(0,\infty)}\Big[\int_0^{t_2-t_1}\frac{-(1-\gamma)se^{-\mu t}}{u(1-\gamma+\gamma s)+(1-\gamma-u\gamma)se^{-\mu t}}dt\Big]U(du)\Big\}\\ &= \exp\left\{\int_{(0,\infty)}\left[\frac{1-\gamma}{\mu(1-\gamma-u\gamma)}\int_0^{t_2-t_1}d\log\left(\frac{u(1-\gamma+\gamma s)}{(1-\gamma-u\gamma)s}+e^{-\mu t}\right)\right]U(du)\right\}\\ &= \exp\left\{\int_{(0,\infty)}\left[\frac{1-\gamma}{\mu(1-\gamma-u\gamma)}\log\frac{u(1-\gamma+\gamma s)+(1-\gamma-u\gamma)se^{-\mu(t_2-t_1)}}{(1-\gamma)(u+s)}\right]U(du)\right\}\end{aligned}$$

When $t_2-t_1\to\infty$,

$$\begin{aligned}&\exp\left\{\int_{(0,\infty)}\left[\frac{1-\gamma}{\mu(1-\gamma-u\gamma)}\log\frac{u(1-\gamma+\gamma s)+(1-\gamma-u\gamma)se^{-\mu(t_2-t_1)}}{(1-\gamma)(u+s)}\right]U(du)\right\}\\ &\longrightarrow\exp\left\{\int_{(0,\infty)}\left[\frac{1-\gamma}{\mu(1-\gamma-u\gamma)}\log\frac{u(1-\gamma+\gamma s)}{(1-\gamma)(u+s)}\right]U(du)\right\},\end{aligned}$$

which is the GC III class.

Furthermore, we can construct the following example.

Example 5.12 *Consider the measure $U(du)$ is θ on point β, and 0 elsewhere. If K is from* **P1**, *it follows that*

$$\phi(s)=\begin{cases}\exp\Big\{\frac{\theta}{\mu}\log\Big(\frac{\beta+se^{-\mu(t_2-t_1)}}{\beta+s}\Big)\Big\}=\Big(e^{-\mu(t_2-t_1)}+\big[1-e^{-\mu(t_2-t_1)}\big]\frac{\beta}{\beta+s}\Big)^{\theta/\mu}, & \textit{for constant } \mu;\\[2ex] \exp\left\{\theta\int_{t_1}^{t_2}\left(\frac{-se^{-\int_t^{t_2}\mu(\tau)d\tau}}{\beta+se^{-\int_t^{t_2}\mu(\tau)d\tau}}\right)dt\right\}, & \textit{for } \mu(t).\end{cases}$$

Thus, resulting models are

$$X(t_2) = e^{-\mu(t_2-t_1)} \bullet X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t),$$

and

$$X(t_2) = e^{-\int_{t_1}^{t_2} \mu(\tau)d\tau} \bullet X(t_1) + \int_{t_1}^{t_2} e^{-\int_t^{t_2} \mu(\tau)d\tau} \bullet d\epsilon(t),$$

corresponding to the stationary and the non-stationary case respectively.

5.4 Real-valued innovation processes and examples

Finally, we consider the real-valued innovation processes to include all possible cases for the theory of continuous-time GAR(1) processes. Since for real rv's, the only choice among extended-thinning operations is the constant multiplier, our task is simply to choose proper innovation processes. Also the cf of K in this case is of form $e^{-i\alpha s}$; hence, its first partial derivative with respect to α is bounded.

First, we choose the compound Poisson (with a variance mixture of the normal distribution) IIP as the innovation process. Then, we shall choose the EGGC IIP as the innovation process. In particularly, we will calculate for a special case, the stable non-Gaussian distribution family, which includes Gaussian (when $\gamma = 2$) and Cauchy (when $\gamma = 1$). Hence, the classical continuous-time GAR(1) Gaussian process is included in our theory, but the process is defined in the sense of convergence in distribution, not in L^2. Note that Cauchy distribution has no expectation. Therefore, it's impossible to construct a continuous-time GAR(1) Cauchy process in the sense of the Itô integral, but it works in the theory of continuous-time GAR(1) processes where convergence in distribution is used. Note that these processes with stable stationary distributions are already known in the literature, however, the convergence for stochastic integration there is in probability, not in distribution; thus, they can induce the processes with stable stationary distributions in this section. Interested readers can see Samorodnitsky and Taqqu [1994]. Here we just show that they can be unified by the theory of continuous-time generalized AR(1) processes.

Theorem 5.4.1 *Let $\{\epsilon(t); t \geq 0\}$ be a IIP of the compound Poisson with the variance mixture of the normal distribution, and the pgf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$*

$$\varphi_{\Delta\epsilon}(s) = \exp\left\{\lambda h[\phi_0(s^2/2) - 1]\right\},$$

where $\phi_0(s)$ is a LT of a positive rv, and differentiable with bounded first order derivative. Then it follows that

$$\varphi(s) = \begin{cases} \exp\left\{\lambda \int_0^{t_2-t_1} \left[\phi_0\left(\frac{s^2 e^{-2\mu t}}{2}\right) - 1\right] dt\right\}, & \textit{for constant } \mu; \\ \\ \exp\left\{\lambda \int_{t_1}^{t_2} \left[\phi_0\left(\frac{s^2 e^{-2\int_t^{t_2} \mu(\tau)d\tau}}{2}\right) - 1\right] dt\right\}, & \textit{for } \mu(t), \end{cases}$$

and $\varphi(s)$ is a cf. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t) \quad \textit{and} \quad \int_{t_1}^{t_2} e^{-\int_t^{t_2} \mu(\tau)d\tau} \bullet d\epsilon(t)$$

exist and are positive rv's.

Proof: This is a direct conclusion from Theorem 5.1.3 since the extended-thinning operation is very simple, just the constant multiplier operation. We can show the rough calculation for the stationary case, i.e., constant μ case in the following.

For K from **P1**, the cf is

$$\varphi_K(s;\alpha) = e^{i\alpha s}, \quad \text{and} \quad -i\log\varphi_K(s;\alpha) = -i(i\alpha s) = \alpha s.$$

Thus

$$\begin{aligned} \varphi_{Z_n}(s) &= \prod_{i=0}^{n-1} \varphi_{\Delta\epsilon}\left(-i\log\varphi_K(s;(1-\mu h)^i)\right) = \prod_{i=0}^{n-1} \varphi_{\Delta\epsilon}\left((1-\mu h)^i s\right) \\ &= \prod_{i=0}^{n-1} \exp\left\{\lambda h\left[\phi_0\left(\frac{(1-\mu h)^{2i}s^2}{2}\right) - 1\right]\right\} \\ &= \exp\left\{\lambda h \sum_{i=0}^{n-1}\left[\phi_0\left(\frac{(1-\mu h)^{2i}s^2}{2}\right) - 1\right]\right\}. \end{aligned}$$

By Theorem 5.1.2, as $n \to \infty$, this goes to

$$\varphi(s) = \exp\left\{\lambda \int_0^{t_2-t_1} \Big[\phi_0\Big(\frac{s^2 e^{-2\mu t}}{2}\Big) - 1\Big] dt\right\},$$

and $\varphi(0) = \exp\left\{\lambda \int_0^{t_2-t_1} \Big[\phi_0(0) - 1\Big] dt\right\} = \exp\{0\} = 1$. Therefore, $\varphi(s)$ is a LT.

Note that the compound Poisson distribution has a positive mass at zero. These kind of distributions are useful in modelling zero-inflated data.

Theorem 5.4.2 *Let $\{\epsilon(t); t \geq 0\}$ be a EGGC IIP with cf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$*

$$\varphi_{\Delta\epsilon}(s) = \exp\left\{-h\frac{c}{2}s^2 + h\int_{(-\infty,\infty)} \left[\log\left(\frac{u}{u-is}\right) - \frac{isu}{1+u^2}\right] U(du)\right\},$$

where $U(du)$ is a non-negative measure on $(0,\infty)$ satisfying

$$\int_{\Re\setminus\{0\}} \frac{1}{1+u^{-2}} U(du) < \infty \quad \text{and} \quad \int_{|u|\leq 1} |\log u^2| U(du) < \infty.$$

Then

$$\varphi(s) = \begin{cases} \exp\Big\{ -\frac{c}{2}s^2\left[1 - e^{-2\mu(t_2-t_1)}\right] + \int_{(-\infty,\infty)} \Big[\int_0^{t_2-t_1} \log\left(\frac{u}{u - ise^{-\mu t}}\right) dt \\ \qquad -\frac{ius}{1+u^2}\left[1 - e^{-2\mu(t_2-t_1)}\right]\Big] U(du)\Big\}, & \text{for constant } \mu; \\ \\ \exp\Big\{ -\frac{c}{2}s^2 \int_{t_1}^{t_2} e^{-\int_t^{t_2} 2\mu(\tau)d\tau} dt + \int_{(-\infty,\infty)} \Big[\int_{t_1}^{t_2} \log\left(\frac{u}{u - ise^{-\int_t^{t_2}\mu(\tau)d\tau}}\right) dt \\ \qquad -\frac{ius}{1+u^2} \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} dt\Big] U(du)\Big\}, & \text{for } \mu(t), \end{cases}$$

and $\varphi(s)$ is a cf. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} \bullet d\epsilon(t)$$

exist and are real rv's.

Proof: This is the straightforward result from Theorem 5.1.3.

However, we usually are not clear on the form of measure $U(\cdot)$. Thus, for a specific distribution in the EGGC to be the innovation process, we have to calculate the $\varphi(s)$ based on the specific form of the cf of that distribution.

Theorem 5.4.3 *Let $\{\epsilon(t); t \geq 0\}$ be a stable non-Gaussian IIP with cf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$*

$$\varphi_{\Delta\epsilon}(s) = \exp\{-\lambda h|s|^\gamma\},$$

where $\lambda > 0$ and $1 \leq \gamma \leq 2$. Then, it follows that

$$\varphi(s) = \begin{cases} \exp\left\{-\frac{\lambda[1-e^{-\gamma\mu(t_2-t_1)}]}{\gamma\mu}|s|^\gamma\right\}, & \text{for constant } \mu; \\ \exp\left\{-\lambda\left(\int_{t_1}^{t_2}\left[e^{-\gamma\int_t^{t_2}\mu(\tau)d\tau}\right]dt\right)|s|^\gamma\right\}, & \text{for } \mu(t). \end{cases}$$

This leads to the model

$$X(t_2) = e^{-\mu(t_2-t_1)} \bullet X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t),$$

and

$$X(t_2) = e^{-\int_{t_1}^{t_2}\mu(\tau)d\tau} \bullet X(t_1) + \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} \bullet d\epsilon(t),$$

corresponding to stationary and non-stationary case respectively. The stochastic integrals

$$\int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} e^{-\int_t^{t_2}\mu(\tau)d\tau} \bullet d\epsilon(t)$$

remain in the same distribution family as the innovation.

Proof: Directly applying (3) of Theorem 5.1.3 can lead to this theorem. However, we can check some calculations by using Theorem 5.1.2. For the stationary case, we have

$$\begin{aligned} \varphi(s) &= \lim_{n\to\infty}\prod_{i=0}^{n-1}\varphi_{\Delta\epsilon}\Big((1-\mu h)^i s\Big) = \lim_{n\to\infty}\prod_{i=0}^{n-1}\exp\Big\{-\lambda h|(1-\mu h)^i s|^\gamma\Big\} \\ &= \lim_{n\to\infty}\exp\Big\{-\lambda\sum_{i=0}^{n-1}h(1-\mu h)^{\gamma i}|s|^\gamma\Big\} = \exp\Big\{-\lambda|s|^\gamma\int_0^{t_2-t_1}e^{-\gamma\mu t}dt\Big\} \\ &= \exp\Big\{-\frac{\lambda[1-e^{-\gamma\mu(t_2-t_1)}]}{\gamma\mu}|s|^\gamma\Big\}. \end{aligned}$$

Similarly, for the non-stationary case,

$$\varphi(s) = \lim_{n\to\infty}\prod_{i=0}^{n-1}\varphi_{\Delta\epsilon}\Big(s\prod_{j=1}^{i}[1-\mu(t_2-jh)h]\Big) = \lim_{n\to\infty}\prod_{i=0}^{n-1}\exp\Big\{-\lambda h|s\prod_{j=1}^{i}[1-\mu(t_2-jh)h]|^\gamma\Big\}$$

$$= \lim_{n\to\infty} \exp\Big\{ -\lambda|s|^\gamma \sum_{i=0}^{n-1} h\Big(\prod_{j=1}^{i} [1-\mu(t_2-jh)h] \Big)^\gamma \Big\}$$
$$= \exp\Big\{ -\lambda\Big(\int_{t_1}^{t_2} [e^{-\gamma \int_t^{t_2} \mu(\tau)d\tau}]dt \Big)|s|^\gamma \Big\}.$$

Since $\varphi(s)$ is still of the form $\exp\{-\beta|s|^\gamma\}$, we conclude that

$$\int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} e^{-\int_t^{t_2} \mu(\tau)d\tau} \bullet d\epsilon(t)$$

with cf $\varphi(s)$ in the same family as the innovation. Applying the generalized Ornstein-Uhlenbeck SDE theory, we obtain the models in this theorem corresponding to stationary and non-stationary case respectively.

Example 5.13 *Consider Cauchy IIP. Then the increment $\epsilon(t+h)-\epsilon(t)$ has cf:*

$$\varphi_{\Delta\epsilon}(s) = \exp\{-h|s|\}.$$

By the above theorem, we have

$$\varphi(s) = \begin{cases} \exp\left\{-\mu^{-1}[1-e^{-\mu(t_2-t_1)}]\cdot|s|\right\}, & \text{for constant } \mu; \\ \\ \exp\left\{-\left(\int_{t_1}^{t_2} \left[e^{-\int_t^{t_2}\mu(\tau)d\tau}\right] dt\right)|s|\right\}, & \text{for } \mu(t). \end{cases}$$

Example 5.14 *Consider Brownian motion. Then the increment $\epsilon(t+h)-\epsilon(t)$ has cf:*

$$\varphi_{\Delta\epsilon}(s) = \exp\{-hs^2/2\}.$$

By the above theorem, we have

$$\varphi(s) = \begin{cases} \exp\left\{-(4\mu)^{-1}\left[1-e^{-2\mu(t_2-t_1)}\right]s^2\right\}, & \text{for constant } \mu; \\ \\ \exp\left\{-\frac{1}{2}\left(\int_{t_1}^{t_2} \left[e^{-\int_t^{t_2}\mu(\tau)d\tau}\right] dt\right)s^2\right\}, & \text{for } \mu(t). \end{cases}$$

This means for the stationary case,

$$\int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t) \sim N\left(0, \frac{1-e^{-2\mu(t_2-t_1)}}{2\mu}\right),$$

and for the non-stationary case,

$$\int_{t_1}^{t_2} e^{-\int_t^{t_2} \mu(\tau)d\tau} \bullet d\epsilon(t) \sim N\left(0, \int_{t_1}^{t_2} [e^{-\int_t^{t_2} \mu(\tau)d\tau}]dt\right).$$

5.5 Tweedie innovation processes

This is another viewpoint to choose the innovation processes. The Tweedie family includes many of the distributions discussed in the previous sections in this chapter such as compound Poisson, Gamma, inverse Gaussian, stable distribution, and so on. Although this overlaps with the previous discussion, we would like to revisit or summarize this case from the perspective of dispersion.

The Tweedie family consists of three types of distributions: non-negative integer support, positive support and real support. All the distributions in this family have the mgf of special exponential form:

$$M_X(s;\theta,\lambda,\beta) = \mathbf{E}\left[e^{sX}\right] = \begin{cases} \exp\left\{\lambda\frac{\beta-1}{\beta}\left(\frac{\theta}{\beta-1}\right)^{\beta}\left[\left(1+\frac{s}{\theta\lambda}\right)^{\beta}-1\right]\right\}, & d \neq 1,2; \\ \left(1+\frac{s}{\theta\lambda}\right)^{-\lambda}, & d = 2; \\ \exp\left\{\lambda e^{\theta}\left[e^{s/\lambda}-1\right]\right\}, & d = 1, \end{cases}$$

where $d = \frac{\beta-2}{\beta-1}$ or $\beta = \frac{d-2}{d-1}$. In specific, the non-negative integer case includes only one distribution, that is Poisson distribution when $\theta = 1$. The positive case includes the compound Poisson with Gamma distribution ($1 < d < 2$), Gamma ($d = 2$), positive stable ($2 < d < 3$ or $d > 3$) and inverse Gaussian ($d = 3$). The real case includes normal ($d = 0$) and extreme stable ($d < 0$ or $d = \infty$). Refer to Section 2.3.2.

The innovation and the self-generalized rv should be of the same type. That is, if the increment of innovation is to be Poisson, then the self-generalized rv K should be non-negative integer-valued, which leads to the choice like **I1**, etc. If the increment of innovation is to be a positive rv, we can choose a positive self-generalized rv K like **P1**. For the real case like normal and extreme stable, K can only be **P1**. We give the following theorem without proof.

Theorem 5.5.1 *Let $\{\epsilon(t); t \geq 0\}$ be a Tweedie IIP with mgf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$*

$$M_{\Delta\epsilon}(s;\theta,\lambda,\beta) = \mathbf{E}\left[e^{sX}\right] = \begin{cases} \exp\left\{h\lambda\frac{\beta-1}{\beta}\left(\frac{\theta}{\beta-1}\right)^{\beta}\left[\left(1+\frac{s}{\theta\lambda}\right)^{\beta}-1\right]\right\}, & d \neq 1,2; \\ \left(1+\frac{s}{\theta\lambda}\right)^{-h\lambda}, & d = 2; \\ \exp\left\{h\lambda e^{\theta}\left[e^{s/\lambda}-1\right]\right\}, & d = 1, \end{cases}$$

1. *Suppose $K(\alpha)$ is a non-negative integer self-generalized rv with pgf $G_K(s;\alpha)$. In this situation, the only non-negative integer distribution is Poisson and $d = 1$. Then, it follows that*

$$G(s) = \begin{cases} \exp\left\{\lambda e^{\theta}\int_0^{t_2-t_1}\left[G_K^{1/\lambda}\left(s; e^{-\mu t}\right)-1\right]dt\right\}, & \textit{for constant } \mu, \\ \exp\left\{\lambda e^{\theta}\int_{t_1}^{t_2}\left[G_K^{1/\lambda}\left(s; e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)-1\right]dt\right\}, & \textit{for } \mu(t), \end{cases}$$

   and $G(s)$ is a pgf. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1}\left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \textit{and} \quad \int_{t_1}^{t_2}\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

   exist and are non-negative integer rv's.

2. *Suppose $K(\alpha)$ is a positive self-generalized rv with LT $\phi_K(s;\alpha)$. In this situation, d corresponds to $(1,2), \{2\}, (2,3), \{3\}$ and $(3,\infty)$. Then it follows that for constant μ*

$$M(s) = \begin{cases} \exp\left\{\lambda\frac{\beta-1}{\beta}\left(\frac{\theta}{\beta-1}\right)^{\beta}\int_0^{t_2-t_1}\left[\left(1+\frac{\log\phi_K\left(-s;e^{-\mu t}\right)}{\theta\lambda}\right)^{\beta}-1\right]dt\right\}, & d \neq 2; \\ \exp\left\{-\lambda\int_0^{t_2-t_1}\left(1+\frac{\log\phi_K\left(-s;e^{-\mu t}\right)}{\theta\lambda}\right)dt\right\}, & d = 2; \end{cases}$$

   and for $\mu(t)$,

$$M(s) = \begin{cases} \exp\left\{\lambda\frac{\beta-1}{\beta}\left(\frac{\theta}{\beta-1}\right)^{\beta}\int_{t_1}^{t_2}\left[\left(1+\frac{\log\phi_K\left(-s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)}{\theta\lambda}\right)^{\beta}-1\right]dt\right\}, & d \neq 2; \\ \exp\left\{-\lambda\int_{t_1}^{t_2}\left(1+\frac{\log\phi_K\left(-s;e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)}{\theta\lambda}\right)dt\right\}, & d = 2; \end{cases}$$

and $M(s)$ is a mgf. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2} \mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are positive rv's.

3. *Suppose $K(\alpha)$ is from* **P1** *with LT $\phi_K(s;\alpha) = e^{-\alpha s}$. In this situation, d corresponds to $(-\infty, 0), \{0\}$ and $\{\infty\}$. Then it follows that*

$$M(s) = \begin{cases} \exp\left\{\lambda^{\frac{\beta-1}{\beta}} \left(\frac{\theta}{\beta-1}\right)^{\beta} \int_0^{t_2-t_1} \left[\left(1+\frac{se^{-\mu t}}{\theta\lambda}\right)^{\beta} - 1\right] dt\right\}, & \text{for constant } \mu, \\ \exp\left\{\lambda^{\frac{\beta-1}{\beta}} \left(\frac{\theta}{\beta-1}\right)^{\beta} \int_{t_1}^{t_2} \left[\left(1+\frac{se^{-\int_t^{t_2}\mu(\tau)d\tau}}{\theta\lambda}\right)^{\beta} - 1\right] dt\right\}, & \text{for } \mu(t), \end{cases}$$

and $M(s)$ is a mgf. Hence, the generalized stochastic integrals

$$\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \quad \text{and} \quad \int_{t_1}^{t_2} \left(e^{-\int_t^{t_2} \mu(\tau)d\tau}\right)_K \circledast d\epsilon(t)$$

exist and are real rv's.

Part III

Probabilistic and statistical properties

Chapter 6

Stationary distributions, steady states and generalized AR(1) time series

The continuous-time GAR(1) processes constructed in Chapter 4 are first order Markov processes. Hence, it's possible that the stationary distribution, namely the limiting distribution of the process, could exist. These are discussed in Section 6.1. Also if the stationary distribution exists, then the process will evolve under steady state when starting from the stationary distribution. This means that $X(t)$ is distributed as the stationary distribution for all t. We study three cases of margins in Section 6.2. Such a steady state process offers a reasonable good model for a stationary time series. This motivates us to study the possible margins under steady states. For margins with specific distributions of interest, we propose a general approach to fit such a need in Section 6.3. In other words, we are trying to investigate the continuous-time GAR(1) processes from the perspective of state space.

In Section 6.4, we discuss the generalized AR(1) time series obtained from the continuous-time GAR(1) processes via equally-spaced time observations. They cover many of the first order autoregressive non-Gaussian time series existing in the literature.

6.1 Stationary distributions

Assume in this section that $\{K(\alpha)\}$ has bounded expectation for $\alpha \in (0,1]$. The stationary distribution, if it exists, is the long run result of a stationary or homogeneous process. Hence, it is independent of "time". This means that as the process evolves, the distribution of the margin $X(t)$ will finally reach a fixed or invariant equilibrium. From the view of state space, i.e., the support of $X(t)$, we are interested in that if there is a stationary distribution for the continuous-time GAR(1) process.

Now we look into the structure of the continuous-time GAR(1) process. By part (1) of Theorem 4.4.2,

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t), \quad t_1 < t_2,$$

where $\int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t) \stackrel{\text{def}}{=} E(0, t_2-t_1)$, the integrated innovation or cumulative innovation, has the following pgf $G_{E(0,t_2-t_1)}(s)$, or LT $\phi_{E(0,t_2-t_1)}(s)$, or cf $\varphi_{E(0,t_2-t_1)}(s)$:

$$\begin{cases} G_{E(0,t_2-t_1)}(s) = \lim\limits_{n\to\infty} \prod\limits_{i=0}^{n-1} G_{\Delta\epsilon}\left(G_K\left(s;(1-\mu h)^i\right)\right), & \text{if the support of } \epsilon(t) \text{ is } \mathcal{N}, \\ \phi_{E(0,t_2-t_1)}(s) = \lim\limits_{n\to\infty} \prod\limits_{i=0}^{n-1} \phi_{\Delta\epsilon}\left(-\log\phi_K\left(s;(1-\mu h)^i\right)\right), & \text{if the support of } \epsilon(t) \text{ is } \Re_+, \\ \varphi_{E(0,t_2-t_1)}(s) = \lim\limits_{n\to\infty} \prod\limits_{i=0}^{n-1} \varphi_{\Delta\epsilon}\left(i\log\varphi_K\left(s;(1-\mu h)^i\right)\right), & \text{if the support of } \epsilon(t) \text{ is } \Re. \end{cases}$$

Here $h = (t_2-t_1)/n$, and $\{\epsilon(t); t \geq 0\}$ is a stationary independent increment process of innovation. $\Delta\epsilon$ is the increment with time lag h in the innovation process.

First we study the dependence term $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)$ to see its limiting behavior. Fix time t_1, and let $t_2 \to \infty$. Then the time difference $t_2 - t_1 \to \infty$, which leads $e^{-\mu(t_2-t_1)} \to 0$. For the margin being a real rv, the extended-thinning becomes the constant multiplier. Hence,

$$\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) = e^{-\mu(t_2-t_1)} \cdot X(t_1) \to 0, \quad t_2 - t_1 \to \infty.$$

This means that the dependence term will finally diminish to zero. It leaves us a hint that this could be true for margins being positive or non-negative integer-valued. Revisiting Property 3.14

and 3.15, we obtain

$$\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) \longrightarrow (0)_K \circledast X(t_1) = 0, \quad \text{as } t_2 - t_1 \to \infty.$$

Therefore, as time goes to infinity, the margin of a continuous-time GAR(1) process will be

$$X(\infty) \stackrel{d}{=} \lim_{t_2-t_1\to\infty} \int_0^{t_2-t_1} \left(e^{-\mu t)}\right)_K \circledast d\epsilon(t) = \int_0^{\infty} \left(e^{-\mu t)}\right)_K \circledast d\epsilon(t).$$

Consequently, this integral will have the pgf

$$G_\infty(s) = \lim_{t_2-t_1\to\infty} G_{E(0,t_2-t_1)}(s),$$

or LT

$$\phi_\infty(s) = \lim_{t_2-t_1\to\infty} \phi_{E(0,t_2-t_1)}(s),$$

or cf

$$\varphi_\infty(s) = \lim_{t_2-t_1\to\infty} \varphi_{E(0,t_2-t_1)}(s),$$

where $G_\infty(s)$ is continuous at $s = 1$ with $G_\infty(1) = 1$, or $\phi_\infty(s)$ is continuous at $s = 0$ with $\phi_\infty(0) = 1$, or $\varphi_\infty(s)$ is continuous at $s = 0$ with $\varphi_\infty(0) = 1$.

It is possible to calculate the explicit form of pgf, or LT, or cf of the stationary distribution. Note that the $G_\infty(s)$, or $\phi_\infty(s)$ or $\varphi_\infty(s)$ is of product form of the pgf, or LT, or cf of the increment of the innovation process. Hence, we can choose special innovation process, in which the increment, $\epsilon(t+h) - \epsilon(t)$, has the exponential form of pgf, or LT, or cf. Of course, the choice of extended-thinning should be appropriate too. This leads to the following theorem.

Theorem 6.1.1 *Assume the innovation process $\{\epsilon(t);\ t \geq 0\}$ has increment $\epsilon(t+h)-\epsilon(t)$ such that its pgf, or LT, or cf is of form $e^{hC(s)}$, depending on the increment being non-negative integer-valued, or positive-valued, or real-valued. $C(s)$ is assumed to be differentiable with bounded first order derivative. Also assume $G_K(s;\alpha)$ and $\log G_K(s;\alpha)$, or $\phi_K(s;\alpha)$ and $\log \phi_K(s;\alpha)$, or $\log \varphi_K(s;\alpha)$ have bounded first order derivative with respect to α in $[0,1]$ (boundaries could be excluded). It follows that for the stationary continuous-time GAR(1) process*

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t), \quad t_1 < t_2,$$

if the following integrals exist:

$$\int_0^\infty C\left(G_K\left(s; e^{-\mu t}\right)\right) dt, \quad \int_0^\infty C\left(-\log \phi_K\left(s; e^{-\mu t}\right)\right) dt, \quad \int_0^\infty C\left(-i \log \varphi_K\left(s; e^{-\mu t}\right)\right) dt,$$

then the stationary distribution exists, and has the following pgf, or LT, or cf according to non-negative integer, or positive, or real margins:

$$\begin{cases} G_\infty(s) = \exp\left\{\int_0^\infty C\left(G_K\left(s; e^{-\mu t}\right)\right) dt\right\}, & \text{if the support of } \epsilon(t) \text{ is } \mathcal{N}, \\ \phi_\infty(s) = \exp\left\{\int_0^\infty C\left(-\log \phi_K\left(s; e^{-\mu t}\right)\right) dt\right\}, & \text{if the support of } \epsilon(t) \text{ is } \Re_+, \\ \varphi_\infty(s) = \exp\left\{\int_0^\infty C\left(-i \log \varphi_K\left(s; e^{-\mu t}\right)\right) dt\right\}, & \text{if the support of } \epsilon(t) \text{ is } \Re. \end{cases}$$

The corresponding rv is the generalized stochastic integral $\int_0^\infty \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$.

Proof: It is straightforward to derive them by Theorem 5.1.2.

By Theorem 5.1.3, we know if K is from **I1**, **I2**, **I4**, **I5**, **P1**, **P2**, **P4** and **P5**, then Theorem 6.1.1 holds. This theorem is not valid for **I3**, **P3**, because $G_K(s; 0) \neq 1$ for **I3**, and **I3**, **P3** do not have finite expectations.

In the next section, we will discuss the situations of steady state where the relevant stationary distributions have support on $\mathcal{N}$, or $\Re_+$ and $\Re$.

6.2 Marginal distributions under steady state

The stationary distribution is a particular feature of a homogeneous Markov process. If the stationary distribution exists, then this process has steady state when starting just from this stationary distribution. Typically, the margins of a Markov process, do not have the same distributions when the process evolves. However, under steady state, all marginal distributions are the same as the stationary distribution.

In this section, we shall investigate the particular types of marginal distributions under steady state; they may have non-negative integer, or positive, or real support. These results mainly correspond to those continuous-time GAR(1) processes constructed in Chapter 5.

In statistical practice, often we encounter time series data (observed on equally or unequally spaced time points) which are near stationary. Hence, an assumption of stationarity is reasonable. Of course, obvious non-stationary situations like trend, seasonality, and so on can happen too. This leads to a general principle of modelling for observations over time: define a stationary process for the time series first, then make parameters depend on covariates to define a process with time-varying marginal distribution.

For this purpose, constructing a steady state process which has the same distribution on every margin is the first consideration of modelling. Thus, it gains more attention from statisticians.

6.2.1 Non-negative integer margins

First, we turn to the continuous-time GAR(1) processes constructed in Chapter 5, which have non-negative integer-valued margins. We shall study the limiting behavior of these processes as $t_2 - t_1$ goes to infinity. We only consider the stationary SDE case, namely, constant μ and stationary IIP innovation. These limiting behaviors lead to the stationary distributions, and the resulting processes have the same distribution as the marginal distributions under steady state.

To guarantee the margins being non-negative integer-valued, the self-generalized rv K involved in extended-thinning operation should be non-negative integer-valued, and the increment of innovation process $\{\epsilon(t); t \geq 0\}$ should be non-negative integer-valued too. Hence, K could be from **I1** to **I5**, while the innovation process can be the compound Poisson (with a non-negative integer distribution) IIP, GNBC IIP, GC I IIP and GC II IIP. The results for the general non-negative integer-valued self-generalized rv K are given in the following theorem.

Theorem 6.2.1 *Suppose $K(\alpha)$ is a non-negative integer self-generalized rv with pgf $G_K(s;\alpha)$, which is differentiable with bounded first order derivative with respect to α. Consider the stationary generalized Ornstein-Uhlenbeck SDE where μ is a constant.*

(1) *Let* $\{\epsilon(t); t \geq 0\}$ *be a non-negative integer-valued compound Poisson IIP with pgf of* $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$

$$G_{\Delta\epsilon}(s) = \exp\left\{\lambda h[g(s) - 1]\right\},$$

where $g(s)(= \sum_{i=0}^{\infty} p_i s^i)$ *is a pgf, and differentiable with bounded first order derivative. Then, the limiting pgf is*

$$G_\infty(s) = \exp\left\{\lambda \int_0^\infty \left[g\Big(G_K\big(s; e^{-\mu t}\big)\Big) - 1\right] dt\right\}.$$

(2) *Let* $\{\epsilon(t); t \geq 0\}$ *be a GNBC IIP with pgf of* $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$

$$G_{\Delta\epsilon}(s) = \exp\left\{h \int_{(0,1)} \log\left(\frac{p}{1-qs}\right) V(dq)\right\}.$$

If $\int_0^\infty \log\left[\frac{p}{1 - qG_K\left(s; e^{-\mu t}\right)}\right] dt < \infty$ *for any* $q \in (0,1)$, *then the limiting pgf is*

$$G_\infty(s) = \exp\left\{\int_{(0,1)} \left(\int_0^\infty \log\left[\frac{p}{1 - qG_K\left(s; e^{-\mu t}\right)}\right] dt\right) V(dq)\right\}.$$

(3) *Let* $\{\epsilon(t); t \geq 0\}$ *be a GC I IIP with pgf of* $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ *of form*

$$G_{\Delta\epsilon}(s) = \exp\left\{h \int_{(0,1)} \frac{q(s-1)}{1-qs} V(dq)\right\}.$$

If $\int_0^\infty \frac{q\left[G_K\left(s; e^{-\mu t}\right) - 1\right]}{1 - qG_K\left(s; e^{-\mu t}\right)} dt < \infty$ *for any* $q \in (0,1)$, *then the limiting pgf is*

$$G_\infty(s) = \exp\left\{\int_{(0,1)} \left(\int_0^\infty \frac{q\left[G_K\left(s; e^{-\mu t}\right) - 1\right]}{1 - qG_K\left(s; e^{-\mu t}\right)} dt\right) V(dq)\right\}.$$

(4) *Let* $\{\epsilon(t); t \geq 0\}$ *be a GC II IIP with pgf of* $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ *of form*

$$G_{\Delta\epsilon}(s) = \exp\left\{h \int_{(0,1)} \frac{-q(1-s)(1-\gamma s)}{1-qs} V(dq)\right\}.$$

If $\int_0^\infty \frac{-q\left[1 - G_K\left(s; e^{-\mu t}\right)\right]\left[1 - \gamma G_K\left(s; e^{-\mu t}\right)\right]}{1 - qG_K\left(s; e^{-\mu t}\right)} dt < \infty$ *for any* $q \in (0,1)$, *then the limiting pgf is*

$$G_\infty(s) = \exp\left\{\int_{[\gamma,1)} \left(\int_0^\infty \frac{-q\left[1 - G_K\left(s; e^{-\mu t}\right)\right]\left[1 - \gamma G_K\left(s; e^{-\mu t}\right)\right]}{1 - qG_K\left(s; e^{-\mu t}\right)} dt\right) V(dq)\right\}.$$

All these $G_\infty(s)$ are the pgf's of the generalized stochastic integral $\int_0^\infty \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$.

Proof: Straightforward to derive from Theorem 6.1.1.

We now verify the existence of stationary distributions for specific self-generalized rv's K and innovation processes. These stationary distributions are the Poisson, negative binomial, geometric, power series, discrete stable distributions, and the GNBC class.

Example 6.1 (Poisson) *Consider Example 5.1, where*

$$G_{E(0,t_2-t_1)}(s) = \exp\left\{\frac{\lambda}{\mu}[1 - e^{-\mu(t_2-t_1)}](s-1)\right\}.$$

Thus, as $t_2 - t_1 \to \infty$, the limiting pgf is

$$G_\infty(s) = \exp\left\{\frac{\lambda}{\mu}(s-1)\right\},$$

which implies that the stationary distribution is Poisson(λ/μ).

Example 6.2 (Negative binomial and geometric) *First consider Example 5.2. Then*

$$G_{E(0,t_2-t_1)}(s) = \left(e^{-\mu(t_2-t_1)} + [1 - e^{-\mu(t_2-t_1)}]\frac{1-\gamma}{1-\gamma s}\right)^{\lambda/\mu}.$$

This leads to the limit

$$G_\infty(s) = \left(\frac{1-\gamma}{1-\gamma s}\right)^{\lambda/\mu},$$

indicating that the stationary distribution is NB$(\lambda/\mu, \gamma)$. When $\lambda = \mu$, it's the geometric distribution with parameter γ.

Secondly consider Example 5.4, where

$$G_{E(0,t_2-t_1)}(s) = \left(\frac{1 - \frac{\gamma(1-e^{-\mu(t_2-t_1)})}{1-\gamma e^{-\mu(t_2-t_1)}}}{1 - \frac{\gamma(1-e^{-\mu(t_2-t_1)})}{1-\gamma e^{-\mu(t_2-t_1)}}s}\right)^{\lambda(1-\gamma)/(\mu\gamma)}.$$

As $t_2 - t_1 \to \infty$, we have

$$G_\infty(s) = \left(\frac{1-\gamma}{1-\gamma s}\right)^{\lambda(1-\gamma)/(\mu\gamma)},$$

which shows that the stationary distribution is NB$\left(\frac{\lambda(1-\gamma)}{\mu\gamma}, \gamma\right)$. When $\lambda(1-\gamma) = \mu\gamma$, it's the geometric distribution with parameter γ.

Example 6.3 (Power series) *Consider Example 5.3. We choose $\lambda = \mu$, then*

$$G_{E(0,t_2-t_1)}(s) = \frac{1-(1-s)^{1/\theta}}{s} \times \frac{1+(s-1)e^{-\mu(t_2-t_1)}}{1-(1-s)^{1/\theta}e^{-\mu(t_2-t_1)/\theta}}.$$

It goes to $G_\infty(s) = s^{-1}\left[1-(1-s)^{1/\theta}\right]$ as $t_2 - t_1 \to \infty$, namely the stationary distribution is the power series distribution.

Example 6.4 (Discrete stable) *Consider case 1 in Example 5.7. Then*

$$G_{E(0,t_2-t_1)}(s) = \exp\Big\{-\frac{\lambda[1-e^{-\beta\mu(t_2-t_1)}]}{\beta\mu}(1-s)^\beta\Big\},$$

When $t_2 - t_1 \to \infty$, we obtain $G_\infty(s) = \exp\Big\{-\frac{\lambda}{\beta\mu}(1-s)^\beta\Big\}$, thus, the discrete stable distribution.

Example 6.5 (GNBC) *First, we consider Case* **I1** *in Corollary 5.2.6, where*

$$G_{E(0,t_2-t_1)}(s) = \exp\Big\{\frac{1}{\mu}\int_{(0,1)}\log\Big(\frac{1-q-q(s-1)e^{-\mu(t_2-t_1)}}{1-qs}\Big)V(dq)\Big\}$$

when $t_2 - t_1 \to \infty$, the limit is

$$G_\infty(s) = \exp\Big\{\frac{1}{\mu}\int_{(0,1)}\log\Big(\frac{1-q}{1-qs}\Big)V(dq)\Big\},$$

which implies that the stationary distribution is a GNBC.

Secondly, we consider Case **I2** *in Corollary 5.2.8. Then $G_{E(0,t_2-t_1)}(s)$ is*

$$\exp\Big\{\frac{1}{\mu}\int_{[\gamma,1)}\Big[\log\Big(\frac{(1-q)(1-\gamma s)+(q-\gamma)(1-s)e^{-\mu(t_2-t_1)}}{(1-q)(1-\gamma s)-\gamma(1-q)(1-s)e^{-\mu(t_2-t_1)}}\Big)+\log\frac{1-q}{1-qs}\Big]V(dq)\Big\},$$

which goes to $\exp\left\{\frac{1}{\mu}\int_{[\gamma,1)}\log\frac{1-q}{1-qs}V(dq)\right\}$ as $t_2 - t_1 \to \infty$. Since $[\gamma,1) \subset (0,1)$, it is the pgf of a GNBC where $V(dq)$ has zero measure on $(0,\gamma)$.

Since the GNBC class covers many distributions like the logarithmic series distribution, the two kind processes offer many specific continuous-time GAR(1) processes with the same GNBC stationary distributions.

6.2.2 Positive-valued margins

Secondly, as $t_2 - t_1$ goes to infinity, we will study the limiting behavior of those continuous-time GAR(1) processes having positive-valued margins. Like non-negative integer margin situation, we only consider the stationary SDE case, and the limiting behaviors lead to the stationary distributions, as well as the processes being steady state.

For the sake of positive-valued margins, we require the self-generalized rv K involved in extended-thinning operation to be positive-valued, and the increment of innovation process $\{\epsilon(t); t \geq 0\}$ to be positive-valued too. Therefore, K could be from **P1** to **P5**, while the innovation process can be compound Poisson (with a positive distribution) IIP, GGC IIP, and GCMED IIP. Similarly, we have the following results for the general positive self-generalized rv K.

Theorem 6.2.2 *Suppose $K(\alpha)$ is a positive self-generalized rv with LT $\phi_K(s;\alpha)$, which is differentiable with bounded first order derivative with respect to α. Consider the stationary generalized Ornstein-Uhlenbeck SDE where μ is a constant.*

(1) Let $\{\epsilon(t); t \geq 0\}$ be a positive-valued compound Poisson IIP with LT of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$

$$\phi_{\Delta\epsilon}(s) = \exp\{\lambda h[\phi_0(s) - 1]\},$$

where $\phi_0(s)$ is a LT, and differentiable with bounded first order derivative. Then, the limiting LT is

$$\phi_\infty(s) = \exp\Big\{\lambda \int_0^\infty \Big[\phi_0\Big(-\log \phi_K\Big(s; e^{-\mu t}\Big)\Big) - 1\Big] dt\Big\}.$$

(2) Let $\{\epsilon(t); t \geq 0\}$ be a GGC IIP with LT of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$

$$\phi_{\Delta\epsilon}(s) = \exp\Big\{h \int_{(0,\infty)} \log\big(\frac{u}{u+s}\big) U(du)\Big\},$$

where $U(du)$ is a non-negative measure on $(0,\infty)$ satisfying

$$\int_{(0,1]} |\log u| U(du) < \infty, \quad \text{and} \quad \int_{(1,\infty)} u^{-1} U(du) < \infty.$$

If $\left|\int_0^\infty \log\left(\frac{u}{u-\log\phi_K(s;e^{-\mu t})}\right)dt\right| < \infty$ *for any* $u \in (0,\infty)$, *then the limiting LT is*

$$\phi_\infty(s) = \exp\left\{\int_{(0,\infty)}\left[\int_0^\infty \log\left(\frac{u}{u-\log\phi_K(s;e^{-\mu t})}\right)dt\right]U(du)\right\}.$$

(3) Let $\{\epsilon(t); t \geq 0\}$ *be a GCMED IIP with LT of* $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$

$$\phi_{\Delta\epsilon}(s) = \exp\left\{h\int_{(0,\infty)}\frac{-s}{u+s}U(du)\right\},$$

where $U(du)$ *is a non-negative measure on* $(0,\infty)$ *satisfying* $\int_{(0,\infty)} u^{-1}U(du) < \infty$.

If $\left|\int_0^\infty \frac{\log\phi_K(s;e^{-\mu t})}{u-\log\phi_K(s;e^{-\mu t})}dt\right| < \infty$ *for any* $u \in (0,\infty)$, *then the limiting LT is*

$$\phi_\infty(s) = \exp\left\{\int_{(0,\infty)}\left[\int_0^\infty \frac{\log\phi_K(s;e^{-\mu t})}{u-\log\phi_K(s;e^{-\mu t})}dt\right]U(du)\right\}.$$

All these $\phi_\infty(s)$ *are the LT's of the generalized stochastic integral* $\int_0^\infty (e^{-\mu t})_K \circledast d\epsilon(t)$.

Proof: It is straightforward to derive them from Theorem 6.1.1.

For specific self-generalized rv K and innovation processes, we can check if the stationary distributions exist. Following are some examples which appeared in Chapter 5. These stationary distributions include the exponential, Gamma, inverse Gaussian, etc. Thus, we can obtain the steady state processes with marginal distributions being the exponential, Gamma, GGC and GC III.

Example 6.6 (Gamma and exponential) *Consider Example 5.12. Then*

$$\phi_{E(0,t_2-t_1)}(s) = \left(e^{-\mu(t_2-t_1)} + [1-e^{-\mu(t_2-t_1)}]\frac{\beta}{\beta+s}\right)^{\theta/\mu}.$$

Thus, as $t_2 - t_1 \to \infty$, *the limiting LT is*

$$\phi_{E(0,t_2-t_1)}(s;\infty) = \left(\frac{\beta}{\beta+s}\right)^{\theta/\mu},$$

which implies that the stationary distribution is Gamma$(\theta/\mu,\beta)$. *By setting* $\theta = \mu$, *we obtain the Exponential*(β) *stationary distribution.*

Example 6.7 (GGC) *Consider Theorem 5.3.3. We choose K being from* **P1**. *Then*

$$\phi_{E(0,t_2-t_1)}(s) = \exp\left\{\frac{1}{\mu}\int_{(0,\infty)} \log\left(\frac{u+se^{-\mu(t_2-t_1)}}{u+s}\right)U(du)\right\}.$$

When $t_2 - t_1 \to \infty$, we obtain

$$\phi_\infty(s) = \exp\left\{\frac{1}{\mu}\int_{(0,\infty)} \log\left(\frac{u}{u+s}\right)U(du)\right\},$$

which shows that the stationary distribution is a GGC. Note that the GGC class covers a lot of distributions such as the positive stable distribution, inverse Gaussian distribution, etc.

Example 6.8 (GC III) *Still consider Theorem 5.3.3, but choose K being from* **P2**. *Then*

$$\phi_{E(0,t_2-t_1)}(s) = \exp\left\{\int_{(0,\infty)} \left[\frac{1-\gamma}{\mu(1-\gamma-u\gamma)} \log\frac{u(1-\gamma+\gamma s)+(1-\gamma-u\gamma)se^{-\mu(t_2-t_1)}}{(1-\gamma)(u+s)}\right] U(du)\right\}.$$

When $t_2 - t_1 \to \infty$, the limiting LT is

$$\phi_\infty(s) = \exp\left\{\int_{(0,\infty)} \left[\frac{1-\gamma}{\mu(1-\gamma-u\gamma)} \log\frac{u(1-\gamma+\gamma s)}{(1-\gamma)(u+s)}\right] U(du)\right\}.$$

This leads to the stationary distribution being GC III.

6.2.3 Real-valued margins

Lastly, we study those continuous-time GAR(1) processes with real margins. In this case, the job seems easier because the only known extended-thinning operation is the constant multiplier, or in our terminology, **P1**. We still consider the stationary situation of the processes, i.e., constant μ and stationary innovation processes. The innovation processes we choose here are the compound Poisson with the variance mixture of the normal distribution, EGGC, stable non-Gaussian IIP. It gives us the following result.

Theorem 6.2.3 *Consider the stationary generalized Ornstein-Uhlenbeck SDE where μ is a constant.*

(1) Let $\{\epsilon(t); t \geq 0\}$ be the IIP of a compound Poisson with the variance mixture of the normal distribution, and the cf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ is

$$\varphi_{\Delta\epsilon}(s) = \exp\left\{\lambda h[\phi_0(s^2/2) - 1]\right\},$$

where $\phi_0(s)$ is a LT of a positive rv, and differentiable with bounded first order derivative. Then, the limiting cf is

$$\varphi_\infty(s) = \exp\Big\{\lambda \int_0^\infty \Big[\phi_0\Big(\frac{s^2 e^{-2\mu t}}{2}\Big) - 1\Big] dt\Big\}.$$

(2) Let $\{\epsilon(t); t \geq 0\}$ be a EGGC IIP with cf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ is

$$\varphi_{\Delta\epsilon}(s) = \exp\left\{-h\frac{c}{2}s^2 + h\int_{(-\infty,\infty)} \left[\log\left(\frac{u}{u - is}\right) - \frac{isu}{1+u^2}\right] U(du)\right\},$$

where $U(du)$ is a non-negative measure on $(0, \infty)$ satisfying

$$\int_{\Re\setminus\{0\}} \frac{1}{1+u^{-2}} U(du) < \infty \quad \textit{and} \quad \int_{|u|\leq 1} |\log u^2| U(du) < \infty.$$

If $\left|\int_0^\infty \log\left(\frac{u}{u - ise^{-\mu t}}\right) dt\right| < \infty$ for any $u \in (-\infty, \infty)$, then the limiting cf is

$$\varphi_\infty(s) = \exp\left\{-\frac{c}{2}s^2 + \int_{(-\infty,\infty)} \left[\int_0^\infty \log\left(\frac{u}{u - ise^{-\mu t}}\right) dt - \frac{ius}{1+u^2}\right] U(du)\right\}.$$

(3) Let $\{\epsilon(t); t \geq 0\}$ be a stable non-Gaussian IIP with cf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ is

$$\varphi_{\Delta\epsilon}(s) = \exp\{-\lambda h |s|^\gamma\},$$

where $\lambda > 0$ and $1 \leq \gamma \leq 2$. then the limiting cf is

$$\varphi_\infty(s) = \exp\Big\{-\frac{\lambda}{\gamma\mu}|s|^\gamma\Big\}.$$

All these $\varphi_\infty(s)$ are the cf's of the generalized stochastic integral $\int_0^\infty e^{-\mu t} \bullet d\epsilon(t)$.

Proof: Case (1) and (2) are derived from Theorem 6.1.1. Case (3) is simply setting $t_2 - t_1 \to \infty$ in Theorem 5.4.3 so that the component $e^{-\gamma\mu(t_2 - t_1)}$ vanishes.

Case (3) in Theorem 6.2.3 shows that stationary distribution is still a stable non-Gaussian distribution. Here we list two boundary cases as examples: $\gamma = 1$ (Cauchy) and $\gamma = 2$ (Gaussian).

Example 6.9 (Cauchy) *Consider Example 5.13, where the innovation process is a Cauchy IIP. When $t_2 - t_1 \to \infty$, the limiting cf is*

$$\varphi_\infty(s) = \exp\left\{-\frac{|s|}{\mu}\right\}.$$

This leads to the stationary distribution being Cauchy again.

Example 6.10 (Gaussian) *Consider Example 5.14, where the innovation process is a Brownian motion. When $t_2 - t_1 \to \infty$, the limiting cf is*

$$\varphi_\infty(s) = \exp\left\{-\frac{s^2}{4\mu}\right\}.$$

This leads to the stationary distribution being the Gaussian distribution.

6.3 Customizing margins

Modelling is one of the biggest concerns of statisticians. In the context of time series, we encounter observations record over time. These observations may be count data (non-negative integer-valued), positive data, or real-valued data. Each type of data may be modelled by several potential distributions. Hence, one typical approach is to propose appropriate marginal distributions for each time point. These marginal distributions could be Poisson, negative binomial, generalized Poisson, Gamma, exponential, inverse Gaussian, and so on. With these in mind, for the obtained data, we may fix a couple of distributions to be the possible choices of the marginal distributions. In other words, we wish to model the marginal distributions with certain known distributions which are widely used in statistical practice for a univariate response. This leads to the question of how to customize the margins, which is part of the model considerations. The method in Chapter 5 is a passive way to construct models, because we don't know in advance the possible stationary distributions or the marginal distributions under steady state.

Recall the idea in Chapter 5. We first fix the extended-thinning operation and the innovation process, namely K and $\{\epsilon(t); t \geq 0\}$, in the generalized Ornstein-Uhlenbeck SDE. Then we obtain the continuous-time GAR(1) process. Certain appropriate innovation processes will lead to stationary continuous-time GAR(1) processes as we discussed in the last section. Under steady state, the marginal distributions are the same as the stationary distribution.

Now we assume to have a steady state continuous-time GAR(1) process with known marginal distributions for a certain self-generalized rv K. Our task is to determine if there is an appropriate innovation process. If such an innovation process exists, we know the assumed continuous-time GAR(1) process exists which possesses the marginal distribution we prescribe. We shall describe this idea more accurately in notation.

The possible stochastic representation for a stationary generalized Ornstein-Uhlenbeck SDE has form

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t).$$

Suppose it is under steady state. Now consider a small increment on $[t, t+h]$. This leads to

$$X(t+h) \stackrel{d}{=} \left(e^{-\mu h}\right)_K \circledast X(t) + E(t; t+h).$$

Here we use $E(t;t+h)$ to replace the cumulative innovation on $[t, t+h]$, namely $\int_0^h \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$.

First, we consider a non-negative integer self-generalized rv K with pgf $G_K(s;\alpha)$. In this case, the margins are non-negative integer-valued and the innovation process has non-negative integer-valued increment. Hence, we prescribe the pgf of stationary distribution as $G_X(s)$. We deduce the following:

$$\begin{aligned}
\mathbf{E}\left(s^{X(t+h)}\right) &= \mathbf{E}\left(s^{\left(e^{-\mu h}\right)_K \circledast X(t) + E(t;t+h)}\right) \\
&= \mathbf{E}\left(s^{\left(e^{-\mu h}\right)_K \circledast X(t)}\right) \mathbf{E}\left(s^{E(t;t+h)}\right), \\
G_{X(t+h)}(s) &= G_{X(t)}\left(G_K\left(s; e^{-\mu h}\right)\right) \mathbf{E}\left(s^{E(t;t+h)}\right).
\end{aligned}$$

Since under steady state, $G_{X(t+h)}(s) = G_{X(t)}(s) = G_X(s)$, it follows that

$$G_X(s) = G_X\left(G_K\left(s; e^{-\mu h}\right)\right) \mathbf{E}\left(s^{E(t;t+h)}\right).$$

This leads to

$$\mathbf{E}\left(s^{E(t;t+h)}\right) = \frac{G_X(s)}{G_X\left(G_K\left(s; e^{-\mu h}\right)\right)} = \exp\left\{\log \frac{G_X(s)}{G_X\left(G_K\left(s; e^{-\mu h}\right)\right)}\right\},$$

namely

$$G_{E(t;t+h)}(s) = \exp\left\{\log \frac{G_X(s)}{G_X\left(G_K\left(s; e^{-\mu h}\right)\right)}\right\}.$$

We wish to find out this stationary innovation process $\{\epsilon(t); t \geq 0\}$, or equivalently, find out the independent increment $\Delta\epsilon(h) = \epsilon(t+h) - \epsilon(t)$. Let h be infinitesimal, expand $G_{E(t;t+h)}(s)$ in h, keep the order one term of h and omit the terms with higher order of h. This procedure will lead to the pgf or LT of $\Delta\epsilon(h)$ so that the increment is determined in the sense of distribution. One can recall the idea that we obtain the infinitesimal increment $\Delta f(t)$ of a function $f(t)$ by expanding

$$f(t+h) = f(t) + f'(t)h + o(h),$$

and retaining $f'(t)h$. Here we follow the same procedure. In this point of view of infinitesimal analysis, one can imagine that $\Delta\epsilon(h)$ is the first order differential of $E(t; t+h)$, the cumulative integration.

Denote $H(s) = \left.\frac{\partial G_K(s;\alpha)}{\partial \alpha}\right|_{\alpha=1}$. We expand the following at $h = 0$, or in the form of $e^{-\mu h}$, around $e^0 = 1$. By Property 3.1, we know that $G_K(s; 1) = s$. Expansions are:

$$\begin{aligned}
e^{-\mu h} &= 1 - \mu h + o(h), \\
G_K\left(s; e^{-\mu h}\right) &= G_K(s;1) + H(s)(e^{-\mu h} - 1) + o(e^{-\mu h} - 1) \\
&= s - H(s)\mu h + o(h), \\
G_X\left(G_K\left(s; e^{-\mu h}\right)\right) &= G_X(s - H(s)\mu h + o(h)) \\
&= G_X(s) + G'_X(s)(-H(s)\mu h) + o(h), \\
\log \frac{G_X(s)}{G_X\left(G_K\left(s; e^{-\mu h}\right)\right)} &= -\log\left(1 - \frac{G'_X(s)}{G_X(s)} H(s)\mu h + o(h)\right) \\
&= \frac{G'_X(s)}{G_X(s)} H(s)\mu h + o(h).
\end{aligned}$$

[Note that by the infinite divisibility property proved in Theorem 7.2.7, $G_X(0) > 0$ and we don't have to worry about the denominator as $s \to 0$.] Leaving out the term of $o(h)$ in the last expression, we derive

$$\mathbf{E}\left(s^{\Delta\epsilon(h)}\right) = \exp\left\{\frac{G'_X(s)}{G_X(s)} H(s)\mu h\right\}.$$

Table 6.1: *Partial derivative of pgf, $H(s)$, for self-generalized distributions with non-negative integer support.*

K	$H(s)$
I1	$s-1$
I2	$(1-\gamma s)(s-1)/(1-\gamma)$
I3	$(s-1)\log(1-s)$
I4	$(s-c^{-1})[\theta+\log(1-cs)]$
I5	$\theta(s-1)\left[\frac{\gamma}{1-\gamma}(1-s)^{1/\theta}+1\right]$

Therefore, we obtain the form of potential pgf of $\Delta\epsilon(h)$ as:

$$G_{\Delta\epsilon(h)}(s) = \exp\left\{\frac{G'_X(s)}{G_X(s)}H(s)\mu h\right\}.$$

If we can verify that it is indeed a pgf, then we can conclude that $\{\epsilon(t); t \geq 0\}$ stipulated by independent increment with such pgf is appropriate for the assumed continuous-time AR(1) process. Under steady state, this process has the prescribed margins. Table 6.1 list the $H(s)$ of all five self-generalized distributions with non-negative integer support.

Secondly, we consider the positive self-generalized rv K with LT $\phi_K(s;\alpha)$. In this situation, the margins of the process will be positive-valued and the increment of innovation process is positive-valued too. Assume the LT of stationary distribution is $\phi_X(s)$.

$$\begin{aligned}
\mathbf{E}\left(e^{-sX(t+h)}\right) &= \mathbf{E}\left(e^{-s\left[\left(e^{-\mu h}\right)_K \circledast X(t)+E(t;t+h)\right]}\right) \\
&= \mathbf{E}\left(e^{-s\cdot\left(e^{-\mu h}\right)_K \circledast X(t)}\right)\mathbf{E}\left(e^{-sE(t;t+h)}\right), \\
\phi_{X(t+h)}(s) &= \phi_{X(t)}\left(-\log\phi_K\left(s;e^{-\mu h}\right)\right)\mathbf{E}\left(e^{-sE(t;t+h)}\right).
\end{aligned}$$

Similarly, under steady state, we have $\phi_{X(t+h)}(s) = \phi_{X(t)}(s) = \phi_X(s)$, thus,

$$\phi_X(s) = \phi_X\left(-\log\phi_K\left(s;e^{-\mu h}\right)\right)\mathbf{E}\left(e^{-sE(t;t+h)}\right),$$

which leads to

$$\mathbf{E}\left(e^{-sE(t;t+h)}\right) = \frac{\phi_X(s)}{\phi_X\left(-\log\phi_K\left(s;e^{-\mu h}\right)\right)} = \exp\left\{\log\frac{\phi_X(s)}{\phi_X\left(-\log\phi_K\left(s;e^{-\mu h}\right)\right)}\right\},$$

namely

$$\phi_{E(t;t+h)}(s) = \exp\left\{\log \frac{\phi_X(s)}{\phi_X\left(-\log \phi_K\left(s; e^{-\mu h}\right)\right)}\right\}.$$

We shall apply the same reasoning as before to deduce the LT of $\Delta\epsilon(h)$, the increment of stationary innovation process.

Denote $H(s) = \frac{\partial}{\partial \alpha}[-\log \phi_K(s;\alpha)]\Big|_{\alpha=1}$. We expand the following at $h = 0$. Note that $\phi_K(s;1) = e^{-s}$ (see Property 3.1) and $e^{-\mu h} - 1 = -\mu h + o(h)$.

$$\begin{aligned}
-\log \phi_K\left(s; e^{-\mu h}\right) &= -\log \phi_K(s;1) + H(s)(e^{-\mu h} - 1) + o(e^{-\mu h} - 1) \\
&= s - H(s)\mu h + o(h), \\
\phi_X\left(-\log \phi_K\left(s; e^{-\mu h}\right)\right) &= \phi_X(s - H(s)\mu h + o(h)) \\
&= \phi_X(s) + \phi_X'(s)(-H(s)\mu h) + o(h), \\
\log \frac{\phi_X(s)}{\phi_X\left(-\log \phi_K\left(s; e^{-\mu h}\right)\right)} &= -\log\left(1 - \frac{\phi_X'(s)}{\phi_X(s)} H(s)\mu h + o(h)\right) \\
&= \frac{\phi_X'(s)}{\phi_X(s)} H(s)\mu h + o(h).
\end{aligned}$$

By leaving out the $o(h)$ term, we obtain

$$\mathbf{E}\left(e^{s\Delta\epsilon(h)}\right) = \exp\left\{\frac{\phi_X'(s)}{\phi_X(s)} H(s)\mu h\right\}.$$

Thus, if the appropriate innovation process $\{\epsilon(t); t \geq 0\}$ exists, the LT of the increment $\Delta\epsilon(h)$ must be of the form:

$$\phi_{\Delta\epsilon(h)}(s) = \exp\left\{\frac{\phi_X'(s)}{\phi_X(s)} H(s)\mu h\right\}.$$

Now the issue comes down to prove that it is a LT. One can resort to relevant techniques of proof used in Section 2.2.2. If proved, the assumed continuous-time GAR(1) process exists, and has the prescribed margins. Table 6.2 list the $H(s)$ of all five self-generalized distributions with positive support.

Lastly, we discuss this method for real margins. Now the only self-generalized rv K involved in extended-thinning operation is from **P1**. Hence, in this case, we have

$$X(t+h) = e^{-\mu h} \bullet X(t) + E(t; t+h).$$

Table 6.2: *Partial derivative of negative log LT, $H(s)$, for self-generalized distributions with positive support.*

K	$H(s)$
P1	s
P2	$(1-\gamma+\gamma s)s/(1-\gamma)$
P3	$s \log s$
P4	$\left(s+\frac{1}{e^{\theta}-1}\right)\log\left[1+(e^{\theta}-1)s\right]$
P5	$\theta s\left(1+\frac{\gamma}{1-\gamma}s^{1/\theta}\right)$

This leads to

$$\varphi_X(s) = \varphi_X\left(e^{-\mu h}s\right)\varphi_{E(t;t+h)}(s),$$

where $\varphi_X(s)$ is the cf of the margins of an assumed steady state continuous-time GAR(1) process. Furthermore,

$$\varphi_{E(t;t+h)}(s) = \frac{\varphi_X(s)}{\varphi_X\left(e^{-\mu h}s\right)} = \exp\left\{\log\frac{\varphi_X(s)}{\varphi_X\left(e^{-\mu h}s\right)}\right\}.$$

Expanding it at $h=0$, we obtain

$$\begin{aligned}
\frac{\partial\varphi_X(\alpha s)}{\partial\alpha} &= \varphi'_X(\alpha s)s, \\
\log\frac{\varphi_X(s)}{\varphi_X\left(e^{-\mu h}s\right)} &= -\log\frac{\varphi_X\left(e^{-\mu h}s\right)}{\varphi_X(s)} \\
&= -\log\frac{\varphi_X(s) - s\varphi'_X(s)\mu h + o(h)}{\varphi_X(s)} \\
&= -\log\left(1-\frac{s\varphi'_X(s)}{\varphi_X(s)}\mu h + o(h)\right) \\
&= \frac{s\varphi'_X(s)}{\varphi_X(s)}\mu h + o(h),
\end{aligned}$$

which leads to

$$\varphi_{\Delta\epsilon(h)}(s) = \exp\left\{\frac{s\varphi'_X(s)}{\varphi_X(s)}\mu h\right\}.$$

If it is a cf, we can claim that our assumed continuous-time GAR(1) process is appropriate.

Note that h can be arbitrary once we obtain the form of the pgf or LT or cf of $\Delta\epsilon(h)$. These exponential forms show that they are infinitely divisible. Hence, the characterization form of ID pgf, or LT or cf will be very helpful to prove that if these obtained expressions are pgf's, LT and cf or not. Refer back to Section 2.3.1 for those characterization forms.

One byproduct is that once the assumed continuous-time GAR(1) process exists, we can easily obtain the pgf, or LT or cf for the cumulative innovation

$$E(t_1;t_2) = \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$$

in the stationary situation. It is simply the ratio:

$$\begin{aligned}
\mathbf{E}\left(s^{E(t_1;t_2)}\right) &= \frac{G_X(s)}{G_X\left(G_K\left(s;e^{-\mu(t_2-t_1)}\right)\right)}, \\
\mathbf{E}\left(e^{-sE(t_1;t_2)}\right) &= \frac{\phi_X(s)}{\phi_X\left(-\log\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)\right)}, \\
\varphi_{E(t_1;t_2)}(s) &= \frac{\varphi_X(s)}{\varphi_X\left(e^{-\mu(t_2-t_1)}s\right)}.
\end{aligned}$$

This will avoid the tedious calculation in Chapter 5.

We summarize these results in the following theorem.

Theorem 6.3.1 *Assume $\{X(t);t \geq 0\}$ is a continuous-time GAR(1) process with stationary distribution. Under steady state, the margins have the same pgf $G_X(s)$, or LT $\phi_X(s)$, or cf $\varphi_X(s)$.*

(1) If K is a non-negative integer self-generalized rv with pgf $G_K(s;\alpha)$, and $H(s) = \left.\frac{\partial G_K(s;\alpha)}{\partial\alpha}\right|_{\alpha=1}$, then the innovation process must have such stationary independent increment that its pgf has form:

$$G_{\Delta\epsilon(h)}(s) = \exp\left\{\frac{G_X'(s)}{G_X(s)}H(s)\mu h\right\}.$$

(2) If K is a positive self-generalized rv with LT $\phi_K(s;\alpha)$, and $H(s) = \left.\frac{\partial}{\partial\alpha}\left[-\log\phi_K(s;\alpha)\right]\right|_{\alpha=1}$, then the innovation process must have such stationary independent increment that its LT has form:

$$\phi_{\Delta\epsilon(h)}(s) = \exp\left\{\frac{\phi_X'(s)}{\phi_X(s)}H(s)\mu h\right\}.$$

(3) If K is from **P1** *and the margins of the continuous-time AR(1) process being real-valued, then the innovation process must have such stationary independent increment that its cf has form:*

$$\varphi_{\Delta\epsilon(h)}(s) = \exp\left\{\frac{s\varphi_X'(s)}{\varphi_X(s)}\mu h\right\}.$$

Proof: The stochastic representation of a stationary continuous-time GAR(1) process is

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t).$$

Consider a small time increment h from time t. Then it will be

$$X(t+h) \stackrel{d}{=} \left(e^{-\mu h}\right)_K \circledast X(t) + E(t;t+h).$$

Here we use $E(t;t+h)$ to replace the cumulative innovation on $[t, t+h]$, namely $\int_0^h \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$. We want to find the corresponding stochastic difference equation (see Section 4.3)

$$X(t+h) - X(t) \stackrel{d}{=} [(1-\mu h)_K \circledast X(t) - X(t)] + \Delta\epsilon, \quad \Delta\epsilon = \epsilon(t+h) - \epsilon(t),$$

or

$$X(t+h) \stackrel{d}{=} (1-\mu h)_K \circledast X(t) + \Delta\epsilon.$$

The pgf, or LT, or cf of $E(t;t+h)$ can be determined by the pgf, or LT, or cf of the stationary distribution and the self-generalized distribution. They are all the ratios

$$\frac{G_X(s)}{G_X\left(G_K\left(s;e^{-\mu h}\right)\right)}, \quad \text{or} \quad \frac{\phi_X(s)}{\phi_X\left(-\log\phi_K\left(s;e^{-\mu h}\right)\right)}, \quad \text{or} \quad \frac{\varphi_X(s)}{\varphi_X\left(e^{-\mu h}s\right)}.$$

Now we need to find the pgf, or LT, or cf of $\Delta\epsilon$, the increment of innovation. Recall that the innovation process $\{\epsilon(t);\ t \geq 0\}$ is additive. Hence, the pgf, or LT, or cf of $\Delta\epsilon = \epsilon(t+h) - \epsilon(t)$ is expressed in the exponential form with exponent being linear in h. i.e., the form like $\exp\{hg(s)\}$. Our task will simply become the expansion of the logarithm of these ratios in terms of h, and omitting those terms with higher order.

The key step is that if we can expand the $G_K\left(s;e^{-\mu h}\right)$, or $\phi_K\left(s;e^{-\mu h}\right)$ in terms of h. This leads to the requirement of conditions of existence of the partial derivative with respect to α at boundary $\alpha = 1$ for the non-negative integer and positive stationary distribution situation. $G_X(s)$ and $\phi_X(s)$ are positive and continuous in their domains. Following the previous discussion, we can complete the proof without further difficulty.

Remarks:

(1) Given the marginal distribution, the form of pgf, or LT, or cf of the increment of innovation is

$$\exp\left\{\frac{G'_X(s)}{G_X(s)}H(s)\mu h\right\}, \quad \text{or} \quad \exp\left\{\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h\right\}, \quad \text{or} \quad \exp\left\{\frac{s\varphi'_X(s)}{\varphi_X(s)}\mu h\right\}.$$

However, for any prescribed marginal distribution, it may not be a pgf, or LT, or cf. We need to check whether it is a pgf, or LT, or cf. If yes, we obtain the increment of innovation, and hence the innovation process.

On the other hand, if the form is indeed a pgf, or LT, or cf, we can claim that

$$\exp\left\{\frac{G'_X(s)}{G_X(s)}H(s)\right\}, \quad \text{or} \quad \exp\left\{\frac{\phi'_X(s)}{\phi_X(s)}H(s)\right\}, \quad \text{or} \quad \exp\left\{\frac{s\varphi'_X(s)}{\varphi_X(s)}\right\}$$

is the pgf, or LT, or cf of an ID distribution, because h is arbitrary and these exponents are linear in h. This ID feature may further help us to simplify the proof of pgf by taking advantage of the fact that non-negative integer ID distribution is compound Poisson distribution.

(2) This customizing approach can be naturally extended to the non-stationary continuous-time GAR(1) process situation. In this situation, we assume that μ is still a constant to simplify the case, but the innovation process is additive with time-varying increment. Then the stochastic difference equation becomes

$$X(t+h) = (1-\mu h)_K \circledast X(t) + \Delta\epsilon(h), \quad \Delta\epsilon(h) = \epsilon(t+h) - \epsilon(t).$$

Now we have to specify the pgf, or LT, or cf for every margin $X(t)$, instead of the only one pgf, or LT, or cf for all margins. Besides, we should assume that the partial derivative $\frac{\partial G_{X(t)}(s)}{\partial t}$, or $\frac{\partial \phi_{X(t)}(s)}{\partial t}$, or $\frac{\partial \varphi_{X(t)}(s)}{\partial t}$ exist for all times $t \geq 0$. Combining the assumption on self-generalized distribution, we can obtain the pgf, or LT, or cf of the time-varying increment of innovation $\Delta\epsilon(h) = \epsilon(t+h) - \epsilon(t)$ by the same reasoning, but in slightly different modifications. This is essentially because that the innovation process is additive. The related expansions are

$$\begin{aligned}
G_K\left(s; e^{-\mu h}\right) &= s - H(s)\mu h + o(h), \\
G_{X(t+h)}\left(G_K\left(s; e^{-\mu h}\right)\right) &= G_{X(t+h)}(s - H(s)\mu h + o(h)) \\
&= G_{X(t)}(s) - \mu H(s)\frac{\partial G_{X(t)}(s)}{\partial s}h + \frac{\partial G_{X(t)}(s)}{\partial t}h + o(h),
\end{aligned}$$

$$
\begin{aligned}
\log \frac{G_{X(t)}(s)}{G_{X(t+h)}\left(G_K\left(s;e^{-\mu h}\right)\right)} &= -\log\Bigg(1-h\left[\mu H(s)\frac{\partial G_{X(t)}(s)}{\partial s}-\frac{\partial G_{X(t)}(s)}{\partial t}\right]\Big/G_{X(t)}(s) \\
&\qquad +o(h)\Bigg) \\
&= h\left[\mu H(s)\frac{\partial G_{X(t)}(s)}{\partial s}-\frac{\partial G_{X(t)}(s)}{\partial t}\right]\Big/G_{X(t)}(s)+o(h), \\
-\log\phi_K\left(s;e^{-\mu h}\right) &= s-H(s)\mu h+o(h), \\
\phi_{X(t+h)}\left(-\log\phi_K\left(s;e^{-\mu h}\right)\right) &= \phi_{X(t+h)}(s-H(s)\mu h+o(h)) \\
&= \phi_{X(t)}(s)-\mu H(s)\frac{\partial \phi_{X(t)}(s)}{\partial s}h+\frac{\partial \phi_{X(t)}(s)}{\partial t}h+o(h), \\
\log \frac{\phi_{X(t)}(s)}{\phi_{X(t+h)}\left(-\log\phi_K\left(s;e^{-\mu h}\right)\right)} &= -\log\Bigg(1-h\left[\mu H(s)\frac{\partial \phi_{X(t)}(s)}{\partial s}-\frac{\partial \phi_{X(t)}(s)}{\partial t}\right]\Big/\phi_{X(t)}(s) \\
&\qquad +o(h)\Bigg) \\
&= h\left[\mu H(s)\frac{\partial \phi_{X(t)}(s)}{\partial s}-\frac{\partial \phi_{X(t)}(s)}{\partial t}\right]\Big/\phi_{X(t)}(s)+o(h), \\
\log \frac{\varphi_{X(t)}(s)}{\varphi_{X(t+h)}\left(e^{-\mu h}s\right)} &= -\log\frac{\varphi_{X(t+h)}\left(e^{-\mu h}s\right)}{\varphi_{X(t)}(s)} \\
&= -\log\frac{\varphi_{X(t)}(s)-\mu s\frac{\partial \varphi_{X(t)}(s)}{\partial s}h+\frac{\partial \varphi_{X(t)}(s)}{\partial t}h+o(h)}{\varphi_{X(t)}(s)} \\
&= -\log\Bigg(1-h\left[\mu s\frac{\partial \varphi_{X(t)}(s)}{\partial s}-\frac{\partial \varphi_{X(t)}(s)}{\partial t}\right]\Big/\varphi_{X(t)}(s) \\
&\qquad +o(h)\Bigg) \\
&= h\left[\mu s\frac{\partial \varphi_{X(t)}(s)}{\partial s}-\frac{\partial \varphi_{X(t)}(s)}{\partial t}\right]\Big/\varphi_{X(t)}(s)+o(h).
\end{aligned}
$$

Thus, the pgf, or LT, or cf of $\Delta\epsilon(h)=\epsilon(t+h)-\epsilon(t)$ is

$$\exp\left\{h\left[\mu H(s)\frac{\partial G_{X(t)}(s)}{\partial s}-\frac{\partial G_{X(t)}(s)}{\partial t}\right]\Big/G_{X(t)}(s)\right\},$$

or

$$\exp\left\{h\left[\mu H(s)\frac{\partial \phi_{X(t)}(s)}{\partial s}-\frac{\partial \phi_{X(t)}(s)}{\partial t}\right]\Big/\phi_{X(t)}(s)\right\},$$

or

$$\exp\left\{h\left[\mu s\frac{\partial \varphi_{X(t)}(s)}{\partial s}-\frac{\partial \varphi_{X(t)}(s)}{\partial t}\right]\Big/\varphi_{X(t)}(s)\right\}.$$

respectively.

When in steady state for the stationary situation, it follows that

$$\frac{\partial G_{X(t)}(s)}{\partial t} = 0, \quad \text{or} \quad \frac{\partial \phi_{X(t)}(s)}{\partial t} = 0, \quad \text{or} \quad \frac{\partial \varphi_{X(t)}(s)}{\partial t} = 0,$$

which lead to the statements in Theorem 6.3.1.

We will illustrate Theorem 6.3.1 by some examples. They were basically discussed before. Now we revisit them from another perspective. In these examples, we prescribe the marginal distributions as Poisson, negative binomial, modified Geometric, exponential or Gamma, inverse Gaussian, and so on.

Example 6.11 (Poisson) *Consider a continuous-time GAR(1) process under steady state. Prescribe that the margins are distributed in Poisson(λ). Thus*

$$G_X(s) = \exp\{\lambda(s-1)\} \quad \textit{and} \quad \frac{G'_X(s)}{G_X(s)} = \frac{\lambda \exp\{\lambda(s-1)\}}{\exp\{\lambda(s-1)\}} = \lambda.$$

By (1) in Theorem 6.3.1, the pgf or LT of increment of innovation process must be of form:

$$G_{\Delta\epsilon(h)}(s) = \exp\left\{\lambda H(s)\mu h\right\},$$

where $H(s) = \left.\frac{\partial G_K(s;\alpha)}{\partial \alpha}\right|_{\alpha=1}$. *Since the known types of K are from* **I1** *to* **I5**, *we look into Table 6.1 to find the appropriate self-generalized rv's so that the expression of $G_{\Delta\epsilon(h)}(s)$ or $\phi_{\Delta\epsilon(h)}(s)$ is a pgf.*

By the feature of ID, $\Delta\epsilon(h)$ should be a compound Poisson rv. This fact leads to that $H(s)$ must satisfy:

$C \cdot H(s) = g(s) - 1$, *C is some positive constant which may be bounded above, and $g(s)$ is a pgf.*

In another word, $C \cdot H(s) + 1$ must be a pgf. Checking with the form of $H(s)$ in Table 6.1, we find that only K being **I1** *works; the others lead to negative coefficients in the power series of $H(s)$ which implies that* **I2** *to* **I5** *are excluded.*

With K is from **I1**,

$$G_{\Delta\epsilon(h)}(s) = \exp\left\{\lambda\mu h(s-1)\right\},$$

the pgf of Poisson($\lambda\mu h$). It basically corresponds to the Poisson IIP innovation process appearing in Example 5.1. The assumed continuous-time GAR(1) process matches the process obtained in Example 5.1 too.

Example 6.12 (Negative Binomial/Geometric) *Consider a continuous-time GAR(1) process under steady state. Prescribe that the margins are distributed in NB(β,γ). When $\beta = 1$, it is geometric distribution. Thus*

$$G_X(s) = \left(\frac{1-\gamma}{1-\gamma s}\right)^\beta \quad \text{and} \quad \frac{G'_X(s)}{G_X(s)} = \frac{\beta\gamma}{1-\gamma s}.$$

The form of the pgf or LT of increment of the potential innovation process must be:

$$G_{\Delta\epsilon(h)}(s) = \exp\left\{\beta\gamma\mu h \frac{H(s)}{1-\gamma s}\right\},$$

where $H(s) = \left.\frac{\partial G_K(s;\alpha)}{\partial\alpha}\right|_{\alpha=1}$. Now we investigate $H(s)$ in Table 6.1 to see which will lead to a proper innovation process. Recall the fact that $C \cdot \frac{H(s)}{1-\gamma s} + 1$ should be a pgf for a positive constant C.

For K being from **I1**, *take $C = \gamma$. Then*

$$\gamma \cdot \frac{H(s)}{1-\gamma s} + 1 = \gamma \cdot \frac{s-1}{1-\gamma s} + 1 = \frac{1-\gamma}{1-\gamma s},$$

which is the pgf of NB($1,\gamma$). This corresponds to the innovation process appearing in Example 5.2.

For K being from **I2**, *in which the fixed parameter γ is exactly the same as prescribed for the marginal distribution here, we take $C = 1-\gamma$. It gives*

$$(1-\gamma) \cdot \frac{H(s)}{1-\gamma s} + 1 = (1-\gamma) \cdot \frac{(1-\gamma s)(s-1)}{(1-\gamma)(1-\gamma s)} + 1 = s,$$

which is, of course, a pgf. The corresponding innovation process is Poisson IIP appearing in Example 5.4.

For K being from **I3**, *it seems that $C \cdot \frac{(s-1)\log(1-s)}{1-\gamma s} + 1$ is not a pgf.*

Example 6.13 (Modified Geometric) *Since a zero-modification operation can adjust the mass at zero, we may be interested in considering a steady state continuous-time GAR(1) process with modified Geometric margins, in which the pgf is*

$$G_X(s) = (1-p) + ps\frac{1-\beta}{1-\beta s}, \quad 0 < p,\ \beta < 1.$$

Then,

$$\begin{aligned} G'_X(s) &= p(1-\beta)(1-\beta s)^{-1} + ps\beta(1-\beta)(1-\beta s)^{-2}, \\ \frac{G'_X(s)}{G_X(s)} &= \frac{p(1-\beta)}{(1-\beta s)[(1-p)+(p-\beta)s]}. \end{aligned}$$

The form of the pgf of increment of the potential innovation process should be:

$$G_{\Delta\epsilon(h)}(s) = \exp\left\{p(1-\beta)\mu h \frac{H(s)}{(1-\beta s)[(1-p)+(p-\beta)s]}\right\},$$

where $H(s) = \left.\frac{\partial G_K(s;\alpha)}{\partial \alpha}\right|_{\alpha=1}$. *We search for suitable* $H(s)$ *in Table 6.1. Bear in mind that* $1 + C \cdot \frac{H(s)p(1-\beta)}{(1-\beta s)[(1-p)+(p-\beta)s]}$ *should be a pgf for a positive constant* C.

For K *being from* **I1**, *take* $C = \frac{1-p}{p(1-\beta)}$. *Then we have*

$$1 + C \cdot \frac{H(s)p(1-\beta)}{(1-\beta s)[(1-p)+(p-\beta)s]} = 1 + \frac{s-1}{(1-\beta s)\left(1 - \frac{\beta-p}{1-p}s\right)}.$$

If $\frac{p}{1+(1-p)^{3/2}} \leq \beta \leq p$ *or* $p < \beta \leq \frac{1}{2-p}$, *this is a pgf; see (5) in Theorem 2.2.3. Thus, the assumed continuous-time GAR(1) process exists, and is new (to our knowledge).*

For K *being from* **I2**, *taking* $C = \frac{(1-\gamma)(1-p)}{p(1-\beta)}$, *leads to*

$$1 + C \cdot \frac{H(s)p(1-\beta)}{(1-\beta s)[(1-p)+(p-\beta)s]} = 1 + \frac{(1-\gamma s)(s-1)}{(1-\beta s)\left(1 - \frac{\beta-p}{1-p}s\right)},$$

which is a pgf when $p < \beta \leq \frac{1+\gamma-p\gamma}{2-p}$, *or* $\max\left(\gamma, \frac{p}{2-p}\right) \leq \beta \leq p$ *and* $\beta(\beta-\gamma)(1-p)^3 \geq (p-\beta+\gamma-p\gamma)(p-\beta)$; *see (6) in Theorem 2.2.3. This implies that the assumed continuous-time GAR(1) process exists, which is also new.*

For K *being from* **I3**, *it seems there is no such pgf. As to* **I4** *and* **I5**, *further study is needed.*

Example 6.14 (GNBC) *Since the GNBC family includes many distributions, we now study this kind of margin. For a continuous-time GAR(1) process being under steady state, prescribe the marginal distribution as GNBC with pgf*

$$G_X(s) = \exp\left\{\frac{1}{\mu}\int_{(0,1)} \log\left(\frac{1-q}{1-qs}\right)V(dq)\right\}.$$

Then

$$\frac{G'_X(s)}{G_X(s)} = \frac{1}{\mu}\int_{(0,1)} \left[\log\left(\frac{1-q}{1-qs}\right)\right]' V(dq) = \frac{1}{\mu}\int_{(0,1)} \frac{q}{1-qs}V(dq).$$

The form of the pgf of the increment of the potential innovation process must be:

$$G_{\Delta\epsilon(h)}(s) = \exp\left\{h\int_{(0,1)} \frac{qH(s)}{1-qs}V(dq)\right\},$$

where $H(s) = \left.\frac{\partial G_K(s;\alpha)}{\partial \alpha}\right|_{\alpha=1}$. *Now we resort to Table 6.1 to see which form of* $H(s)$ *will lead to a proper innovation process. We realize that this is the pgf of a generalized convolution and the base distribution has pgf of form* $\exp\left\{h\frac{qH(s)}{1-qs}\right\}$. *Following previous examples,* $C \cdot \frac{qH(s)}{1-qs} + 1$ *should be a pgf for a positive constant.*

For K *being from* **I1**, *we obtain*

$$\exp\left\{h\int_{(0,1)} \frac{qH(s)}{1-qs}V(dq)\right\} = \exp\left\{h\int_{(0,1)} \frac{q(s-1)}{1-qs}V(dq)\right\},$$

which is the pgf of GC I. Therefore, the assumed process exists and the innovation process corresponds to the one appearing in **(I1)** *of Corollary 5.2.6.*

For K *being from* **I2**, *it is*

$$\exp\left\{h\int_{(0,1)} \frac{qH(s)}{1-qs}V(dq)\right\} = \exp\left\{h\int_{(0,1)} \frac{q(1-\gamma s)(s-1)}{(1-\gamma)(1-qs)}V(dq)\right\}.$$

When $V(\cdot)$ *has* 0 *measure on* $(0,\gamma)$, *namely* $q \geq \gamma$, *it is the pgf of GC II:*

$$\exp\left\{h\int_{(0,1)} \frac{q(1-\gamma s)(s-1)}{(1-\gamma)(1-qs)}V(dq)\right\} = \exp\left\{h\int_{[\gamma,1)} \frac{q(1-\gamma s)(s-1)}{(1-\gamma)(1-qs)}V(dq)\right\}.$$

This means the innovation process is GC II IIP, which corresponds to the one in **(I2)** *of Corollary 5.2.8. Thus, the assumed process exists.*

It is possible that a steady state continuous-time AR(1) process with GNBC margins exists for K *being from other non-negative integer self-generalized distributions. They are under further study.*

Example 6.15 (Gamma/Exponential) *Consider a continuous-time GAR(1) process under steady state. Prescribe that the margins are distributed in Gamma*(δ,β). *When* $\delta = 1$, *it is exponential distribution. Thus*

$$\phi_X(s) = \left(\frac{\beta}{\beta+s}\right)^{\delta} \quad \text{and} \quad \frac{\phi'_X(s)}{\phi_X(s)} = -\frac{\delta}{\beta+s}.$$

By (2) in Theorem 6.3.1, the form of the LT of the increment of the potential innovation process must be:

$$\phi_{\Delta\epsilon(h)}(s) = \exp\left\{-\frac{\delta}{\beta+s}H(s)\mu h\right\},$$

where $H(s) = \frac{\partial}{\partial\alpha}\left[-\log\phi_K(s;\alpha)\right]\Big|_{\alpha=1}$. *Now we investigate* $H(s)$ *in Table 6.2 to search for proper self-generalized distributions which lead to an innovation process.*

For K *being from* **P1**, *then*

$$\exp\left\{-\frac{\delta}{\beta+s}H(s)\mu h\right\} = \exp\left\{-\frac{\delta}{\beta+s}s\mu h\right\} = \exp\left\{\delta\mu h\left(\frac{\beta}{\beta+s}-1\right)\right\},$$

which is the LT of the compound Poisson($\delta\mu h$) with Gamma($1,\beta$). This corresponds to the innovation process appearing in Example 5.9 when $\gamma = 1$.

For K *being from* **P2**, *we have*

$$\exp\left\{-\frac{\delta}{\beta+s}H(s)\mu h\right\} = \exp\left\{-\frac{\delta}{\beta+s}\frac{(1-\gamma+\gamma s)s}{1-\gamma}\mu h\right\} = \exp\left\{-\frac{\delta\mu h}{1-\gamma}\frac{s(1-\gamma+\gamma s)}{\beta+s}\right\}.$$

If $\gamma \leq \frac{1}{1+\beta}$, *this function is a LT (See (1) in Theorem 2.2.8). Hence, it can be the LT of the increment of an innovation process, and the assumed continuous-time GAR(1) process exists. This process is new, with stochastic representation*

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1}\left(e^{-\mu t}\right)_K \circledast d\epsilon(t),$$

where K *is from* **P2**, *and the stochastic integral has LT*

$$\phi(s) = \frac{\phi_X(s)}{\phi_X\left(-\log\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)\right)} = \left(\frac{\beta + \frac{e^{-\mu(t_2-t_1)}(1-\gamma)s}{(1-\gamma)+(1-e^{-\mu(t_2-t_1)})\gamma s}}{\beta+s}\right)^{\delta}, \qquad \gamma \leq \frac{1}{1+\beta}.$$

For K *being from* **P3**, *it is not a LT, because* $\log s < 0$ *as* $0 < s < 1$, *hence,*

$$\exp\left\{-\frac{\delta}{\beta+s}H(s)\mu h\right\} > 1, \qquad \text{when } 0 < s < 1,$$

contradicting to that a LT is less than or equal to 1.

Other cases of K *being from* **P4** *to* **P5** *are under further study.*

Example 6.16 (Inverse Gaussian) *The inverse Gaussian distribution has many applications; see Seshadri [1999]. Hence, it is of interest to see if there are any continuous-time GAR(1) processes with inverse Gaussian margins in steady state.*

Assume the LT of the inverse Gaussian margin is

$$\phi_X(s) = \exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 + \frac{2\mu^2}{\lambda}s\right)^{1/2}\right]\right\}, \quad \lambda,\ \mu > 0.$$

This form is from Johnson and Kotz [1970a], p. 139, and credited to M. C. K. Tweedie. Then

$$\frac{\phi_X'(s)}{\phi_X(s)} = \frac{\lambda}{\mu}\left[1 - \left(1 + \frac{2\mu^2}{\lambda}s\right)^{1/2}\right]' = -\mu\left(1 + \frac{2\mu^2}{\lambda}s\right)^{-1/2}.$$

Hence, by (2) in Theorem 6.3.1, the form of the LT of increment of the potential innovation process must be:

$$\phi_{\Delta\epsilon(h)}(s) = \exp\left\{-\mu^2 h\left(1 + \frac{2\mu^2}{\lambda}s\right)^{-1/2} H(s)\right\},$$

where $H(s) = \frac{\partial}{\partial\alpha}\left[-\log\phi_K(s;\alpha)\right]\Big|_{\alpha=1}$. *We search* $H(s)$ *in Table 6.2.*

For K *being from* **P1**, *it follows that*

$$\exp\left\{-\mu^2 h\left(1 + \frac{2\mu^2}{\lambda}s\right)^{-1/2} H(s)\right\} = \exp\left\{-\mu^2 hs\left(1 + \frac{2\mu^2}{\lambda}s\right)^{-1/2}\right\}.$$

By (6) in Theorem 2.2.8, it is a LT. Hence, the innovation process is well defined so that the assumed continuous-time GAR(1) process exists. This continuous-time GAR(1) process has stochastic representation

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} \bullet X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t),$$

where the stochastic integral has LT

$$\begin{aligned}
\phi_{E(t_1;t_2)}(s) &= \frac{\phi_X(s)}{\phi_X\left(e^{-\mu(t_2-t_1)}s\right)} \\
&= \exp\left\{\frac{\lambda}{\mu}\left[\left(1 + \frac{2\mu^2}{\lambda}e^{-\mu(t_2-t_1)}s\right)^{1/2} - \left(1 + \frac{2\mu^2}{\lambda}s\right)^{1/2}\right]\right\} \\
&= \exp\left\{\frac{-2\mu(1 - e^{-\mu(t_2-t_1)})s}{\sqrt{1 + \frac{2\mu^2}{\lambda}e^{-\mu(t_2-t_1)}s} + \sqrt{1 + \frac{2\mu^2}{\lambda}s}}\right\}
\end{aligned}$$

Other cases of K *being other self-generalized rv's are under further study.*

Example 6.17 (GGC) *The GGC class includes many distributions. It's very meaningful to investigate the GGC margins. Consider a steady state continuous-time GAR(1) process with prescribed GGC margin. Thus*

$$\phi_X(s) = \exp\left\{\frac{1}{\mu}\int_{(0,\infty)} \log\left(\frac{u}{u+s}\right) U(du)\right\},$$

and

$$\frac{\phi'_X(s)}{\phi_X(s)} = \frac{1}{\mu}\int_{(0,\infty)} \left[\log\left(\frac{u}{u+s}\right)\right]' U(du) = \frac{1}{\mu}\int_{(0,\infty)} \frac{-1}{u+s} U(du).$$

By (2) in Theorem 6.3.1, the form of the LT of increment of the potential innovation process must be:

$$\phi_{\Delta\epsilon(h)}(s) = \exp\left\{h\int_{(0,\infty)} \frac{-H(s)}{u+s} U(du)\right\},$$

where $H(s) = \frac{\partial}{\partial\alpha}[-\log\phi_K(s;\alpha)]\Big|_{\alpha=1}$*. Following the last example (with Gamma/Exponential margins), we turn to Table 6.2 to find proper self-generalized distributions which lead to an innovation process.*

For K being from **P1**, *then*

$$\begin{aligned}\exp\left\{h\int_{(0,\infty)} \frac{-H(s)}{u+s} U(du)\right\} &= \exp\left\{h\int_{(0,\infty)} \frac{-s}{u+s} U(du)\right\} \\ &= \exp\left\{h\int_{(0,\infty)} \frac{-s}{u(u+s)} U^*(du)\right\}, \qquad U^*(du) = uU(du),\end{aligned}$$

which is the LT of the GCMED. This essentially corresponds to the innovation process appearing in Theorem 5.3.3 when K is from **P1**.

For K being from **P2**, *we have*

$$\exp\left\{h\int_{(0,\infty)} \frac{-H(s)}{u+s} U(du)\right\} = \exp\left\{\frac{h}{1-\gamma}\int_{(0,\infty)} \frac{-s(1-\gamma+\gamma s)}{u+s} U(du)\right\}$$

If $U(\cdot)$ has zero measure on $(\gamma^{-1}-1,\infty)$, then it becomes

$$\exp\left\{\frac{h}{1-\gamma}\int_{(0,\infty)} \frac{-s(1-\gamma+\gamma s)}{u+s} U(du)\right\} = \exp\left\{\frac{h}{1-\gamma}\int_{(0,\gamma^{-1}-1]} \frac{-s(1-\gamma+\gamma s)}{u+s} U(du)\right\},$$

which is the LT of GC IV (refer to Section 2.3.3). Hence, we get an appropriate innovation process, and can conclude the assumed continuous-time GAR(1) process exists. This process is new. Its stochastic representation is

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t),$$

where K is from **P2**, *and the stochastic integral has LT*

$$\begin{aligned} \phi(s) &= \frac{\phi_X(s)}{\phi_X\left(-\log\phi_K\left(s; e^{-\mu(t_2-t_1)}\right)\right)} \\ &= \exp\left\{\frac{1}{\mu}\int_{(0,\gamma^{-1}-1]} \log\left(\frac{u + \frac{e^{-\mu(t_2-t_1)}(1-\gamma)s}{(1-\gamma)+(1-e^{-\mu(t_2-t_1)})\gamma s}}{u+s}\right) U(du)\right\}. \end{aligned}$$

For other cases in which K is from **P3** *to* **P5**, *further investigations are under study.*

From these examples, we see that there could be several steady state continuous-time GAR(1) processes corresponding to the same marginal distribution. Such flexibility allows us to try different models when we deal with data. However, this is also somehow vexing for us. Which one should we use? This leads to the question that what are the features of these models with common margins.

For the case of a positive-valued margin, one feature of the cumulative innovation, $E(t_1; t_2) = \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t)$, is whether it has mass at zero, namely $\Pr[E(t_1;t_2) = 0] > 0$. This can be checked with $\lim_{s\to\infty} \phi_{E(t_1;t_2)}(s)$. We can look the following example.

Example 6.18 *Consider two models obtained in Example 6.15. One is*

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} \bullet X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t),$$

where the stochastic integral has LT

$$\phi(s) = \frac{\phi_X(s)}{\phi_X\left(e^{-\mu(t_2-t_1)}s\right)} = \left(\frac{\beta + e^{-\mu(t_2-t_1)}s}{\beta+s}\right)^\delta.$$

Hence,

$$\Pr[E(t_1;t_2) = 0] = \phi(\infty) = e^{-\delta\mu(t_2-t_1)}.$$

The other model is

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t),$$

where K is from **P2**, *and the stochastic integral has LT*

$$\phi(s) = \frac{\phi_X(s)}{\phi_X\left(-\log \phi_K\left(s; e^{-\mu(t_2-t_1)}\right)\right)} = \left(\frac{\beta + \frac{e^{-\mu(t_2-t_1)}(1-\gamma)s}{(1-\gamma)+(1-e^{-\mu(t_2-t_1)})\gamma s}}{\beta + s}\right)^\delta, \quad 0 < \gamma \leq \frac{1}{1+\beta}.$$

In this case, it is

$$\Pr[E(t_1;t_2) = 0] = \phi(\infty) = 0,$$

and there is no mass at zero.

Even the model in Example 6.16 can be an alternate choice for the first model in Example 6.15 whose cumulative innovation has no mass at zero, because as $s \to \infty$,

$$\phi_{E(t_1;t_2)}(s) = \exp\left\{\frac{-2\mu(1-e^{-\mu(t_2-t_1)})s}{\sqrt{1+\frac{2\mu^2}{\lambda}e^{-\mu(t_2-t_1)}s} + \sqrt{1+\frac{2\mu^2}{\lambda}s}}\right\} \longrightarrow e^{-\infty} = 0.$$

The cumulative innovation with mass at zero is the primary cause for the phenomena of sharp jump down or sharp drop pattern. Of course, the dependent term with mass at zero can lead to the same phenomena too. Such a pattern can be imagined as the behavior of a small kid who is climbing a high slide; he struggles to climb a little bit by a little bit. When he reaches a certain height, he suddenly drops down some distance back. He repeats this game without ever tiring.

6.4 Generalized AR(1) time series

Time series data usually means that the observations are obtained at equally spaced time points. It corresponds to the discrete-time process in the probabilistic point of view. Looking back to the history of processes development, it is very common that often the discrete-time processes have been proposed earlier than the continuous-time processes. For example, this phenomena has being happened in the development of Ornstein-Uhlenbeck-type processes and the continuous-time

GAR(1) process proposed in this thesis. It may lead to a false appearance that the continuous-time processes are derived from the corresponding discrete-time processes. However, there is no such simple conclusion that the continuous-time case comes out from the discrete, or vice versa, although we may apply approaches like differentiating the continuous-time processes or integrating the discrete-time processes. This point of view can be seen from the perspective of distribution theory discussed later in Chapter 9, where we will see that construction of the discrete-time processes is relatively easier than the continuous-time processes in some situations.

This clarification won't prevent us from investigating the time series sampled from a continuous-time process. In principle, sampling the observations on equally spaced time points from a continuous-time GAR(1) process $\{X(t); t \geq 0\}$, we will obtain a time series, denoted as

$$\{X_0, X_1, X_2, \ldots, X_n, \ldots\}$$

which we call **the generalized AR(1) time series**. In the literature, there are some generalized AR(1) time series models, but not rich classes. We will look into the generalized AR(1) time series models from the continuous-time GAR(1) processes, and compare with those models in the literature.

There is always an issue about stationary and non-stationary time series. If sampling from a stationary continuous-time GAR(1) process, we will have a stationary generalized AR(1) time series; while sampling from a non-stationary continuous-time GAR(1) process, we will obtain a non-stationary generalized AR(1) time series. For the sake of simplicity, we only consider the stationary case.

Now we introduce some notations. In the equally spaced case, the time difference between any two consecutive points is the same. Denote

$$\Delta = t_i - t_{i-1}, \quad \alpha = e^{-\mu\Delta}, \quad E_i = \int_0^{\Delta} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t), \qquad i = 1, 2, \ldots.$$

Then the stochastic representation for a generalized AR(1) time series model is

$$X_i \stackrel{d}{=} (\alpha)_K \circledast X_{i-1} + E_i, \quad i \in \mathcal{N}. \tag{6.4.1}$$

The classical time series data are real-valued, and the distribution of each observation is stipulated as Gaussian. But for a generalized AR(1) time series, the data may be non-negative

Table 6.3: *Conditional pgf* $G_{(\alpha)_K \circledast X_{i-1}|X_{i-1}=x}(s)$ *when* K *is a non-negative integer self-generalized random variable.*

K	$G_{(\alpha)_K \circledast X_{i-1}\|X_{i-1}=x}(s)$
I1	$((1-\alpha)+\alpha s)^x$
I2	$\left(\frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}\right)^x$
I3	$(1-(1-s)^\alpha)^x$
I4	$\left(c^{-1}[1-e^{-\theta(1-\alpha)}(1-cs)^\alpha]\right)^x$
I5	$\left(1-\alpha^\theta(1-\gamma)^\theta\left[(1-\alpha)\gamma+(1-\gamma)(1-s)^{-1/\theta}\right]^{-\theta}\right)^x$

integer, positive, or real-valued. Hence, the corresponding distributions to model such kind of data are non-Gaussian. Perhaps Phatarfod and Mardia [1973] was the first study for the count data time series, where they defined an AR(1) time series model, in which the stationary (or in another view, the marginal distribution under steady state) include Binomial, Negative Binomial, Geometric and Poisson. Since the 1980's, more non-normal time series models have appeared in the literature. Note that for the generalized AR(1) time series with non-negative integer margins, it corresponds to the type of Galton-Watson process with immigration which has the form of the sum of a branching part and an immigration part; see Nanthi [1983], p. 180-181 for the definition.

It is meaningful to give the conditional pgf or LT of $(\alpha)_K \circledast X_{i-1}$ in equation 6.4.1 when $X_{i-1}=x$. They are

$$G_{(\alpha)_K \circledast X_{i-1}|X_{i-1}=x}(s) = (G_K(s;\alpha))^x, \quad \text{and} \quad \phi_{(\alpha)_K \circledast X_{i-1}|X_{i-1}=x}(s) = (\phi_K(s;\alpha))^x.$$

For specific self-generalized random variables K, see Table 6.3 and 6.4. They are useful when comparing with the models in the literature.

In the following, we will classify those generalized AR(1) time series models by their marginal distributions, similar to what we have done in Section 6.3.

Example 6.19 (Poisson) *Consider the model in Example 6.11. Its margins are distributed in Poisson(λ). Hence, we can derive the generalized AR(1) time series model, which is*

$$X_i \overset{d}{=} \alpha * X_{i-1} + E_i, \quad i \in \mathcal{N},$$

Table 6.4: *Conditional LT $\phi_{(\alpha)_K \circledast X_{i-1}|X_{i-1}=x}(s)$ when K is a positive self-generalized random variable.*

K	$\phi_{(\alpha)_K \circledast X_{i-1}\|X_{i-1}=x}(s)$
P1	$\exp\{-\alpha x s\}$
P2	$\exp\left\{-\frac{\alpha(1-\gamma)xs}{(1-\gamma)+(1-\alpha)\gamma s}\right\}$
P3	$\exp\left\{-xs^{\alpha}\right\}$
P4	$\exp\left\{-x \cdot \frac{\left[1+(e^{\theta}-1)s\right]^{\alpha}-1}{e^{\theta}-1}\right\}$
P5	$\exp\left\{-x\left[\frac{\alpha(1-\gamma)}{(1-\alpha)\gamma+(1-\gamma)s^{-\frac{1}{\theta}}}\right]^{\theta}\right\}$

where $X_i \sim$ Poisson(λ), and $G_{E_i}(s) = \exp\{\lambda(1-\alpha)(s-1)\}$.

This model has been discussed by many researchers, such as Phatarfod and Mardia [1973], McKenzie [1985, 1988], Joe [1996], Jørgensen and Song [1998].

Example 6.20 (Negative binomial/Geometric) *Consider two models in Example 6.12. Their margins are distributed in NB(β,γ) (when $\beta = 1$, it is Geometric distribution). Hence, we can derive two generalized AR(1) time series models. One is*

$$X_i \stackrel{d}{=} \alpha * X_{i-1} + E_i, \quad i \in \mathcal{N},$$

where $X_i \sim$ NB(β,γ), and $G_{E_i}(s) = \left(\frac{(1-\gamma+\alpha\gamma)-\alpha\gamma s}{1-\gamma s}\right)^{\beta}$. This model has been discussed by many researchers, such as McKenzie [1985, 1986], Aly and Bouzar [1994].

The other generalized AR(1) time series model is

$$X_i \stackrel{d}{=} (\alpha)_K \circledast X_{i-1} + E_i, \quad i \in \mathcal{N}.$$

where K is from **I2**, *$X_i \sim$ NB(β,γ), and $G_{E_i}(s) = \left(\frac{1-\gamma}{(1-\alpha\gamma)-(1-\alpha)\gamma s}\right)^{\beta}$, namely $E_i \sim$ NB$\left(\beta, \frac{(1-\alpha)\gamma}{1-\alpha\gamma}\right)$. This model or a restriction of it has been discussed by many researchers, such as Phatarfod and Mardia [1973], McKenzie [1985], Al-Osh and Alzaid [1992], Aly and Bouzar [1994]. Here the parameter γ of* **I2** *and of the NB margin are the same. In Chapter 7, we show that these can be different.*

The stochastic representation of the cumulative innovation can be found in McKenzie [1987].

Example 6.21 (Modified Geometric) *Consider two models in Example 6.13. Their margins are distributed as Modified Geometric. From them, we can obtain two generalized AR(1) time series models. One is*

$$X_i \stackrel{d}{=} \alpha * X_{i-1} + E_i, \quad i \in \mathcal{N},$$

where X_i has pgf $G_X(s) = (1-p) + ps\frac{1-\beta}{1-\beta s}$, and $G_{E_i}(s) = \frac{[(1-p)-(\beta-p)s]\cdot[(1-\beta+\alpha\beta)-\alpha\beta s]}{(1-\beta s)\cdot[(1-\beta+\alpha\beta-\alpha p)-\alpha(\beta-p)s]}$.

The other generalized AR(1) time series model is

$$X_i \stackrel{d}{=} (\alpha)_K \circledast X_{i-1} + E_i, \quad i \in \mathcal{N}.$$

where K is from **I2**, *X_i has pgf $G_X(s) = (1-p) + ps\frac{1-\beta}{1-\beta s}$, and*

$$G_{E_i}(s) = \frac{[(1-p)-(\beta-p)s]\cdot[(1-\beta-\alpha\gamma+\alpha\beta)+(-\gamma-\alpha\beta+\alpha\gamma+\beta\gamma)s]}{(1-\beta s)\cdot[(1-\beta-\alpha p+\alpha\beta-\alpha\gamma+\alpha p\gamma)+(-\gamma+\alpha p-\alpha\beta+\alpha\gamma+\beta\gamma-\alpha p\gamma)s]}.$$

These two models are new.

Example 6.22 (Gamma/Exponential/Chi-square) *Consider two models in Example 6.15. Their margins are distributed in Gamma(δ, β). Special cases include*

- *When $\delta = 1$, it is Exponential(β);*
- *When $\delta = k/2$, $\beta = 1/2$ (k is an integer), it is χ^2_K.*

Hence, we can derive two generalized AR(1) time series models. One is

$$X_i \stackrel{d}{=} \alpha \bullet X_{i-1} + E_i, \quad i \in \mathcal{N},$$

where $X_i \sim \text{Gamma}(\delta, \beta)$, and $\phi_{E_i}(s) = \left(\frac{\beta+\alpha s}{\beta+s}\right)^\delta$. This model has been discussed by Gaver and Lewis [1980], Hutton [1990]. A stochastic representation of cumulative innovation can be seen in Walker [2000].

The other generalized AR(1) time series model is

$$X_i \stackrel{d}{=} (\alpha)_K \circledast X_{i-1} + E_i, \quad i \in \mathcal{N}.$$

where K is from **P2**, *$X_i \sim \text{Gamma}(\delta, \beta)$, and*

$$\phi_{E_i}(s) = \left(\frac{\beta(1-\gamma)+(\alpha-\alpha\gamma+\beta\gamma-\alpha\beta\gamma)s}{(\beta+s)[(1-\gamma)+(1-\alpha)\gamma s]}\right)^\delta.$$

This model generalizes Sim [1990], Adke and Balakrishna [1992], where the cumulative innovation is an exponential rv, a special case when γ takes the upper boundary $\frac{1}{1+\beta}$, and has LT $\phi_{E_i}(s) = \left(\frac{\beta}{\beta+(1-\alpha)s}\right)^{\delta}$.

Example 6.23 (Inverse Gaussian) *Consider the model in Example 6.16. Its margins are distributed in Inverse Gaussian. Hence, we can derive the following generalized AR(1) time series model:*

$$X_i \stackrel{d}{=} \alpha \bullet X_{i-1} + E_i, \quad i \in \mathcal{N},$$

where X_i has LT $\phi_X(s) = \exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 + \frac{2\mu^2}{\lambda}s\right)^{1/2}\right]\right\}$, and $\phi_{E_i}(s) = \exp\left\{\frac{-2\mu(1-\alpha)s}{\sqrt{1+\frac{2\mu^2}{\lambda}\alpha s}+\sqrt{1+\frac{2\mu^2}{\lambda}s}}\right\}$. We are not clear if this model has been previously studied.

In the literature, there are some generalized AR(1) time series models with other marginal distributions. We will visit them in Chapter 7 when we study the generalized self-decomposability. The auto-correlation feature of the generalized time series model will be discussed in Chapter 9. A discussion of simulation of the processes can be found in Chapter 12.

Since the family of continuous-time GAR(1) processes is very rich, we can produce many many generalized AR(1) time series models from them. These models differ in either the marginal distributions or the extended-thinning operations. We have realized that the names of the generalized AR(1) time series models appearing in the literature are quite diverse. Hence, it is a serious problem to name them in a clear way to avoid the potential confusion for the readers. We don't want to open a "drug store" to sell the models in the fancy names like medicines.

Chapter 7

Characterization of stationary distribution families

In this chapter, we further investigate the stationary distributions resulting from the continuous-time GAR(1) processes. This will be from the perspective of extended-thinning operations.

Since the continuous-time GAR(1) processes are essentially characterized by the defining extended-thinning operation, it seems that the resulting stationary distributions may be characterized by it too. This leads to the study of self-decomposability (SD) and discrete self-decomposability (DSD), and furthermore, the proposal of generalized self-decomposability (GSD) and generalized discrete self-decomposability (GDSD). Similar to the SD and DSD case, we find that generalized self-decomposable and generalized discrete self-decomposable distributions are infinitely divisible distributions.

This study is not only for probabilistic interest, but also for statistical interest, because it points out the equivalence between the continuous-time GAR(1) process and the corresponding generalized self-decomposable or generalized discrete self-decomposable distribution. In other words, a generalized (discrete) self-decomposable distribution can lead to the construction of a continuous-time GAR(1) process. This feature attracts particular attention and interests of statisticians in modelling data. For the SD and DSD situation, it has been studied by the statistical pioneers for a long time (see references in Section 7.1). Here, we extend this idea to generalized self-decomposability and generalized discrete self-decomposability in Section 7.2. In Section 7.3, we discuss the relationship among the classes of the generalized self-decomposable distributions and the generalized discrete self-decomposable distributions respectively; that is, if they overlap with one another, or if they are covered by one another.

7.1 Self-decomposable and discrete self-decomposable classes

In Section 2.3.1, we've reviewed the concept of SD and DSD, as well as their advantages to construct a generalized discrete-time or continuous-time GAR(1) process. Now we do the reverse. We shall investigate the self-decomposability and discrete self-decomposability of the stationary distribution from a continuous-time GAR(1) process where K is from **P1** and **I1**.

Now we consider the continuous-time GAR(1) process with K being from **P1** and **I1**, namely

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} \bullet X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t),$$

and

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} * X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t).$$

The first process involves the constant multiplier operation which allows positive and real-valued margins. The second involves the binomial-thinning operation which leads to non-negative integer-valued margins. Suppose they have stationary distributions. Let X have a stationary distribution. Under steady state, we have the following stochastic representation for X

$$X \stackrel{d}{=} e^{-\mu(t_2-t_1)} \bullet X + \int_0^{t_2-t_1} e^{-\mu t} \bullet d\epsilon(t) \quad \text{and} \quad X \stackrel{d}{=} e^{-\mu(t_2-t_1)} * X + \int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t).$$

Denote the stochastic integral as $E(t_1; t_2) = E\left(e^{-\mu(t_2-t_1)}\right)$ in the two cases. Since $e^{-\mu(t_2-t_1)}$ can be any value in $(0, 1)$, hence, the previous representations are equivalent to that for any $c \in (0, 1)$, X can be decomposed as

$$X \stackrel{d}{=} c \bullet X + E(c) \qquad \text{and} \qquad X \stackrel{d}{=} c * X + E(c).$$

By the definition of SD and DSD (see Section 2.3.1), we conclude that X is self-decomposable and discrete self-decomposable in the two cases respectively. This is summarized in the following theorem.

Theorem 7.1.1 *The stationary distribution from the continuous-time GAR(1) process with constant multiplier and binomial-thinning operation is SD and DSD respectively.*

This discloses a new way to prove SD and DSD by the corresponding continuous-time GAR(1) process. Traditionally, to prove SD or DSD, we need to show that $\varphi(s)/\varphi(cs)$ is a cf for a distribution with real support, or $\phi(s)/\phi(cs)$ is a LT for a distribution with positive support, or $G(s)/G(1-c+cs)$ is a pgf for a distribution with non-negative integer support. They correspond to the cumulative innovation in the stochastic representation of a continuous-time GAR(1) process with K being from **P1** or **I1**. In the new approach, we first specify a particular innovation process, and then use the stochastic integration approach discussed in Chapter 4 and 5 to obtain the cumulative innovation. These particular innovation processes are based on the specific distributions.

Using this new approach, the key point is how to find the appropriate innovation process. This has been answered in Section 6.3. Now we formalize it in a reverse way in the following theorem.

Theorem 7.1.2 *Suppose X is distributed in a specific distribution.*

(1) Assume the distribution has real support with cf $\varphi_X(s)$. If $\exp\left\{C \cdot \frac{s\varphi'_X(s)}{\varphi_X(s)}\right\}$ is a cf for all $C > 0$, then X is SD.

(2) Assume the distribution has positive support with LT $\phi_X(s)$. If $\exp\left\{C \cdot \frac{s\phi'_X(s)}{\phi_X(s)}\right\}$ is a LT for all $C > 0$, then X is SD.

(3) Assume the distribution has non-negative integer support with pgf $G_X(s)$. If $1 + C \cdot \frac{(s-1)G'_X(s)}{G_X(s)}$ is a pgf for some $C > 0$, then X is DSD.

Proof: The key step in the proof is that for any $0 < c < 1$, we express $\varphi(s)/\varphi(cs)$, or $\phi(s)/\phi(cs)$, or $G(s)/G(1-c+cs)$ in terms of $\exp\left\{C \cdot \frac{s\varphi'_X(s)}{\varphi_X(s)}\right\}$, or $\exp\left\{C \cdot \frac{s\phi'_X(s)}{\phi_X(s)}\right\}$, or $1 + C \cdot \frac{(s-1)G'_X(s)}{G_X(s)}$.

(1) Since $\exp\left\{C\cdot\frac{s\varphi'_X(s)}{\varphi_X(s)}\right\}$ is a cf for any $C>0$, it follows that $\exp\left\{\int_c^1 \frac{s\varphi'_X(\beta s)}{\varphi_X(\beta s)}d\beta\right\}$ is a cf. This comes from the idea of generalized convolution. By integration, we obtain

$$\exp\left\{\int_c^1 \frac{s\varphi'_X(\beta s)}{\varphi_X(\beta s)}d\beta\right\} = \exp\left\{\int_c^1 \frac{\varphi'_X(\beta s)}{\varphi_X(\beta s)}d(\beta s)\right\} = \exp\left\{\int_c^1 \frac{d[\varphi_X(\beta s)]}{\varphi_X(\beta s)}\right\}$$
$$= \exp\left\{\log\varphi_X(\beta s)\Big|_c^1\right\} = \exp\left\{\log\varphi_X(s) - \log\varphi_X(cs)\right\}$$
$$= \frac{\varphi_X(s)}{\varphi_X(cs)}.$$

This shows that $\frac{\varphi_X(s)}{\varphi_X(cs)}$ is a cf. Therefore, X is SD.

(2) Similar to (1), $\exp\left\{\int_c^1 \frac{s\phi'_X(\beta s)}{\phi_X(\beta s)}d\beta\right\}$ is also a LT. We compute the integration in the exponent,

$$\exp\left\{\int_c^1 \frac{s\phi'_X(\beta s)}{\phi_X(\beta s)}d\beta\right\} = \exp\left\{\int_c^1 \frac{\phi'_X(\beta s)}{\phi_X(\beta s)}d(\beta s)\right\} = \exp\left\{\int_c^1 \frac{d[\phi_X(\beta s)]}{\phi_X(\beta s)}\right\}$$
$$= \exp\left\{\log\phi_X(\beta s)\Big|_c^1\right\} = \exp\left\{\log\phi_X(s) - \log\phi_X(cs)\right\}$$
$$= \frac{\phi_X(s)}{\phi_X(cs)},$$

which means that $\frac{\phi_X(s)}{\phi_X(cs)}$ is a LT. Hence, X is SD.

(3) Since $1+C\cdot\frac{(s-1)G'_X(s)}{G_X(s)}$ is a pgf for a $C>0$, $1+C\cdot\frac{\beta(s-1)G'_X(\beta s+1-\beta)}{G_X(\beta s+1-\beta)}$ is a pgf too (operation (2) in Proposition 2.2.2). By the mixture operation in (3) of Proposition 2.2.2, it follows that

$$g(s) = [-\log c]^{-1}\int_c^1\left(1+C\cdot\frac{\beta(s-1)G'_X(\beta s+1-\beta)}{G_X(\beta s+1-\beta)}\right)\beta^{-1}d\beta$$

is a pgf. Direct calculation shows

$$\begin{aligned} g(s) &= [-\log c]^{-1}\int_c^1\left[C\cdot\beta(s-1)\frac{G'_X(\beta s+1-\beta)}{G_X(\beta s+1-\beta)}+1\right]\beta^{-1}d\beta \\ &= [-\log c]^{-1}\left\{C\int_c^1\frac{G'_X(\beta[s-1]+1)}{G_X(\beta[s-1]+1)}d(\beta[s-1]+1)+\int_c^1 d(\log\beta)\right\} \\ &= [-\log c]^{-1}C\log\frac{G_X(s)}{G_X(cs+1-c)}+1. \end{aligned}$$

This leads to

$$\frac{G_X(s)}{G_X(cs+1-c)} = \exp\left\{\log(c^{-1})C^{-1}[g(s)-1]\right\},$$

the pgf of a compound Poisson. Hence, X is DSD.

Remark: Taking K specifically from **I1** and **P1** in Theorem 6.3.1, we can prove the necessity. Hence, the conditions in Theorem 7.1.2 are actually sufficient and necessary.

The key idea in the proof for SD is to use the concept of generalized convolution established by O. Thorin in 1977. He used this concept to prove the ID of Pareto and lognormal distribution. Steutel and van Harn [1979] gave the result for DSD where there was the shade of generalized convolution. Here we prove it again from the view of continuous-time GAR(1) process theory. Note that in (1) and (2), we require $\exp\left\{C \cdot \frac{s\varphi'_X(s)}{\varphi_X(s)}\right\}$ and $\exp\left\{C \cdot \frac{s\phi'_X(s)}{\phi_X(s)}\right\}$ to be a cf and LT for any $C > 0$ respectively. This is equivalent to say that they are the cf or LT of an ID distribution. Recall that the stationary innovation process, an IIP, is just a Lèvy process. It is reasonable to impose such requirements when we prepare to use them as the increment to construct the innovation process. In (3), we only require the condition holds for a positive constant C, not all constants. In fact, this is because that the discrete ID distribution with non-negative integer support is a compound Poisson and the constant C is just to guarantee that $1 + C \cdot \frac{-G'_X(0)}{G_X(0)} \geq 0$. Therefore, we can pick $C = \frac{G_X(0)}{G'_X(0)}$ which equates the preceding inequality to 0.

A SD distribution is ID (See Feller [1966b], p. 550-555). Accordingly, Steutel and van Harn [1979] proved that a DSD distribution is ID too. Since a discrete ID distribution is compound Poisson, then a DSD distribution must be compound Poisson. This leads us to consider the compound Poisson as the marginal distribution, and consequently leads to the following useful theorem.

Theorem 7.1.3 *For a compound Poisson rv X with pgf*

$$G_X(s) = \exp\{\lambda[g(s) - 1]\}, \quad \lambda > 0,$$

where $g(s) = q_0 + q_1 s + q_2 s^2 + \ldots + q_k s^k + \ldots$, a pgf on the non-negative integers. If

$$kq_k - (k+1)q_{k+1} \geq 0, \quad k = 1, 2, 3, \ldots,$$

then X is DSD.

Proof: Consider $1 + C \cdot \frac{(s-1)G'_X(s)}{G_X(s)}$. By calculation, we obtain

$$\begin{aligned} 1 + C \cdot \frac{(s-1)G'_X(s)}{G_X(s)} &= 1 + \lambda C(s-1)g'(s) \\ &= 1 + \lambda C(s-1)[q_1 + 2q_2 s + 3q_3 s^2 + \ldots + kq_k s^{k-1} + \ldots] \\ &= (1 - \lambda C q_1) + \lambda C \sum_{k=1}^{\infty} [kq_k - (k+1)q_{k+1}] s^k. \end{aligned}$$

Thus, $1 + C \cdot \frac{(s-1)G'_X(s)}{G_X(s)}$ being a pgf is equivalent to

$$1 - \lambda C q_1 \geq 0, \quad kq_k - (k+1)q_{k+1} \geq 0, \qquad k = 1, 2, 3, \ldots.$$

We can always take $C = \lambda^{-1}$ to guarantee $1 - \lambda C q_1 = 1 - q_1 \geq 0$. This completes the proof.

Note that the conditions in Theorem 7.1.3 are also necessary because of Theorem 6.3.1 where we can take K from **I1**. Now we give further results about SD and DSD which are not directly related to Theorems 7.1.2 and 7.1.3.

We encounter many generalized convolutions in our study. We now consider whether these are SD or DSD. This leads to the following theorem.

Theorem 7.1.4 *Consider the real, or positive, or non-negative integer-valued generalized convolution.*

(1) If the base distribution of a real or positive-valued generalized convolution is SD, then the generalized convolution is SD.

(2) If the base distribution of a non-negative integer-valued generalized convolution is DSD, then the generalized convolution is DSD.

Proof: Apply the fact that the generalized convolution is the limit distribution for the sum of independent rv's from the base distribution. If the base distribution is SD or DSD, then the distribution of the sum is SD or DSD (using the distributive law in Property 3.11). This leads to the SD or DSD of the generalized convolution.

The same idea can be applied to the stochastic integral $\int_0^\infty e^{-\mu t} d\epsilon(t)$ and $\int_0^\infty e^{-\mu t} * d\epsilon(t)$, where the increment of the innovation process is SD or DSD. Theorem 7.1.2 stipulates the form of cf, or LT, or pgf of the increment, which further indicates that they should be ID, but it does not require that the increment is SD or DSD. Here we consider these two particular innovation processes.

Theorem 7.1.5 *If the distribution of the increment of the innovation process in a continuous-time GAR(1) process is SD or DSD, then the stationary distribution of the continuous-time GAR(1) process with operator* **P1** *or* **I1** *is SD or DSD respectively.*

Proof: The stationary distribution is the distribution of the stochastic integral $\int_0^\infty e^{-\mu t} d\epsilon(t)$ or $\int_0^\infty e^{-\mu t} * d\epsilon(t)$, which is the limit of $\int_0^{t_2-t_1} e^{-\mu t} d\epsilon(t)$ or $\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t)$ as $t_2 - t_1 \to \infty$.

Let $h = (t_2 - t_1)/n$, and $\Delta\epsilon_i = \epsilon(t_1 + ih) - \epsilon(t_1 + (i-1)h)$, $i = 1, 2, \ldots, n$. Then by the commutative law and distributive law,

$$\begin{aligned}
\int_0^{t_2-t_1} e^{-\mu t} d\epsilon(t) &= \lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i \Delta\epsilon_{n-i} = \lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i [c\Delta\epsilon_{n-i} + Y_i] \\
&= c\left(\lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i \Delta\epsilon_{n-i}\right) + \lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i Y_i \\
&= c\left(\int_0^{t_2-t_1} e^{-\mu t} d\epsilon(t)\right) + \lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i Y_i, \\
\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t) &= \lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i * \Delta\epsilon_{n-i} = \lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i * [c * \Delta\epsilon_{n-i} + Y_i] \\
&= c * \left(\lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i * \Delta\epsilon_{n-i}\right) + \lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i * Y_i \\
&= c * \left(\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t)\right) + \lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i * Y_i.
\end{aligned}$$

Because of the existence of $\int_0^{t_2-t_1} e^{-\mu t} d\epsilon(t)$ and $\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t)$, $\lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i Y_i$ and $\lim_{n\to\infty} \sum_{i=0}^{n-1} (1-\mu h)^i * Y_i$ exist. This means that $\int_0^{t_2-t_1} e^{-\mu t} d\epsilon(t)$ and $\int_0^{t_2-t_1} e^{-\mu t} * d\epsilon(t)$ are SD and DSD respectively. Accordingly, their limits are SD and DSD too.

Remark: This approach works for the sum of a finite or infinite number of random variables. The key point in this approach is that the two decomposed parts should be independent; which may be naively neglected. For a finite sum, the two decomposed parts are of course independent.

For an infinite sum, since the limits of two independent sequences are still independent, the two decomposed parts are then independent. However, for a random sum, this approach will lead to two dependent decomposed parts like:

$$X \stackrel{d}{=} \sum_{i=0}^{N} Y_i \stackrel{d}{=} c \cdot \sum_{i=0}^{N} Y_i + \sum_{i=0}^{N} Z_i \stackrel{d}{=} c \cdot X + \sum_{i=0}^{N} Z_i,$$

or

$$X \stackrel{d}{=} \sum_{i=0}^{N} Y_i \stackrel{d}{=} c * \left(\sum_{i=0}^{N} Y_i \right) + \sum_{i=0}^{N} Z_i \stackrel{d}{=} c * X + \sum_{i=0}^{N} Z_i.$$

Here N is a non-negative integer rv. $c \cdot \sum_{i=0}^{N} Y_i$ and $c * \left(\sum_{i=0}^{N} Y_i \right)$ are dependent to the other part $\sum_{i=0}^{N} Z_i$. Such phenomena could happen in other situations like the variance mixture of the normal distribution.

In other words, SD and DSD distributions have the closure property under the finite or infinite sum of independent rv's, but not under random sums.

Next, we investigate the SD property for variance mixtures of the normal distribution.

Theorem 7.1.6 *For a variance mixture of the normal distribution with representation* $X \stackrel{d}{=} \sqrt{Y} Z$, *where* Y *is a positive rv with LT* $\phi_Y(s)$ *and independent of* $Z \sim \mathrm{N}(0,1)$, *if* $\exp\left\{ C \cdot \frac{s\phi_Y'(s)}{\phi_Y(s)} \right\}$ *is a LT for any* $C > 0$, *then* X *is SD.*

Proof: Recall that the variance mixture of the normal distribution has cf $\varphi_X(s) = \phi_Y(s^2/2)$. Hence,

$$\frac{\varphi_X'(s)}{\varphi_X(s)} = \frac{s\phi_Y'(s^2/2)}{\phi_Y(s^2/2)}.$$

This leads to

$$\exp\left\{ C \cdot \frac{s\varphi_X'(s)}{\varphi_X(s)} \right\} = \exp\left\{ C \cdot \frac{s^2\phi_Y'(s^2/2)}{\phi_Y(s^2/2)} \right\} = \exp\left\{ C' \cdot \frac{\frac{s^2}{2}\phi_Y'(s^2/2)}{\phi_Y(s^2/2)} \right\}, \qquad C' = 2C.$$

If for any $C > 0$, $\exp\left\{ C \cdot \frac{s\phi_Y'(s)}{\phi_Y(s)} \right\}$ is a LT, then $\exp\left\{ C \cdot \frac{\frac{s^2}{2}\phi_Y'(s^2/2)}{\phi_Y(s^2/2)} \right\}$ is the cf of $X' = \sqrt{Y'}Z$. where Y' has LT $\exp\left\{ C \cdot \frac{s\phi_Y'(s)}{\phi_Y(s)} \right\}$. By (1) in Theorem 7.1.2, X is SD.

Note that an equivalent statement is that if Y is SD, then X is SD. One conclusion deduced by this theorem can be seen in Example 7.2.

For the SD distribution with positive support, it is of interest to investigate its discrete analogue; and vice versa. McKenzie [1987] implicitly mentioned this kind of relationship for a specific pair of distributions: the Gamma and the negative binomial; there his original purpose was to claim that the alternative pgf $A(s) = G(1-s)$ is much helpful. The following theorem gives a complete explanation to his description.

Theorem 7.1.7 *If the positive rv X is SD, then its discrete analogue Y is DSD (see Definition 2.6 for the discrete analogue).*

Proof: Assume X has LT $\phi_X(s)$ $(s > 0)$ and Y has pgf $G_Y(s)$ $(0 < s < 1)$, where $G_Y(s) = \phi_X(1-s)$ $(0 < s < 1)$ or $G_Y(s) = \phi_X(d(1-s))$ $(d > 0)$. Without loss of generality, we consider the modified version, namely the latter. Thus, we obtain $\phi_X(s) = G_Y(1-s/d)$. This leads to that for $0 < c < 1$,

$$g(s) \stackrel{\text{def}}{=} \frac{G_Y(s)}{G_Y(1-c+cs)} = \frac{\phi_X(d(1-s))}{\phi_X(d(1-[1-c+cs]))} = \frac{\phi_X(d(1-s))}{\phi_X(cd(1-s))},$$

and

$$h(s) \stackrel{\text{def}}{=} \frac{\phi_X(s)}{\phi_X(cs)} = \frac{G_Y(1-s/d)}{G_Y(1-cs/d)} = \frac{G_Y(1-s/d)}{G_Y(1-c+c[1-cs/d])}.$$

Thus, $g(s) = h(d(1-s))$ and $h(s) = g(1-s/d)$.

If X is SD, then $h(s) = \frac{\phi_X(s)}{\phi_X(cs)}$ is a LT. Now we need to prove that $g(s) = h(d(1-s))$ is a pgf. First, $g(0) = h(d) \geq 0$ and $g(1) = h(0) = 1$. Secondly, we take derivatives for $g(s)$: $g^{(n)}(s) = (-1)^n d^n h^{(n)}(d(1-s)) \geq 0$ for $n \geq 1$. These indicate that $g(s)$ is a pgf, so Y is DSD.

Open question: Conversely, for X, if the discrete analogue Y is DSD, is X SD?

We can show that $h(s)$ is CM when $s \in [0,1]$ and $h(0) = g(1) = 1$. But the difficulty is how to extend the domain of $h(s)$ from $[0,1]$ to $[0,\infty)$. We leave it as a conjecture.

This link between SD and DSD may help us to prove or disprove SD by looking at its discrete analogue or DSD by looking at its continuous analogue.

In the following, we shall illustrate the applications of Theorem 7.1.2, 7.1.3 and 7.1.4 by some examples. Many of them have been shown in the literature, but some are new. First we look at the distribution with real support.

Example 7.1 (Logistic) *The SD of the logistic distribution is proved in Sim [1993] where he proposed a discrete-time logistic GAR(1) process. Here we give another proof to illustrate the power of generalized convolution. The logistic distribution has cf*

$$\varphi_X(s) = M_X(is) = \prod_{k=1}^{\infty} \frac{1}{(1-(is)^2/k^2)} = \prod_{k=1}^{\infty} \frac{1}{(1+s^2/k^2)}.$$

Hence, we obtain

$$\begin{aligned}
\frac{s\varphi_X'(s)}{\varphi_X(s)} &= s\sum_{k=1}^{\infty} \frac{-2s}{k^2}(1+s^2/k^2)^{-1}\varphi_X(s)\Big/\varphi_X(s) \\
&= 2\sum_{k=1}^{\infty} \frac{-s^2}{k^2+s^2} = 2\sum_{k=1}^{\infty}\left(\frac{k^2}{k^2+s^2}-1\right) \\
&= 2\sum_{k=1}^{\infty}\left(\frac{k^2/2}{k^2/2+s^2/2}-1\right).
\end{aligned}$$

Now we need to show that for any $C>0$, $\exp\left\{2C\sum_{k=1}^{\infty}\left(\frac{k^2/2}{k^2/2+s^2/2}-1\right)\right\}$ *is a cf. Denote*

$$\phi(s) = \exp\left\{2C\sum_{k=1}^{\infty}\left(\frac{k^2/2}{k^2/2+s}-1\right)\right\}.$$

Then

$$\exp\left\{2C\sum_{k=1}^{\infty}\left(\frac{k^2/2}{k^2/2+s^2/2}-1\right)\right\} = \phi(s^2/2).$$

This suggests that $\exp\left\{2C\sum_{k=1}^{\infty}\left(\frac{k^2/2}{k^2/2+s^2/2}-1\right)\right\}$ *is the cf of a variance mixture of the normal distribution if* $\phi(s)$ *is the LT of a positive rv.*

Recall GCMED (see Section 2.3.3), where the LT is defined as

$$\exp\left\{-as+\int_{(0,\infty)}\left(\frac{1}{u+s}-\frac{1}{u}\right)U(du)\right\} = \exp\left\{-as+\int_{(0,\infty)}\frac{-s}{u(u+s)}U(du)\right\},$$

where $a\geq 0$, $\int_{(0,\infty)}\frac{1}{u(1+u)}U(du)<\infty$. *Taking* $a=0$ *and rewriting its LT, we have*

$$\exp\left\{\int_{(0,\infty)}\left(\frac{u}{u+s}-1\right)\frac{1}{u}U(du)\right\} = \exp\left\{\int_{(0,\infty)}\left(\frac{u}{u+s}-1\right)U^*(du)\right\},$$

where $U^*(du) = u^{-1}U(du)$ *and* $\int_{(0,\infty)} \frac{1}{1+u} U^*(du) < \infty$. *It becomes* $\phi(s)$ *if we choose the positive measure* $U^*(\cdot)$ *such that it only has mass* $2C$ *on discrete points* $\{k^2/2;\ k = 1, 2, 3, \dots.\}$. *Obviously,* $\sum_{k=1}^{\infty} \frac{1}{1+k^2/2} < \infty$. *Hence,* $\phi(s)$ *is a LT in GCMED family. This indicates that*

$$\exp\left\{2C \sum_{k=1}^{\infty} \left(\frac{k^2/2}{k^2/2 + s^2/2} - 1\right)\right\}$$

is a cf for any $C > 0$. *By Theorem 7.1.2, the logistic distribution is SD.*

Since logistic distribution is in symmetric EGGC family, a natural question is how about other distributions in this family. This leads to the following conclusion for the symmetric EGGC family.

Example 7.2 (symmetric EGGC) *A symmetric EGGC distribution has cf of form*

$$\varphi(s) = \exp\left\{-\frac{ds^2}{2} + \int_{(0,\infty)} \log\left(\frac{u^2}{u^2+s^2}\right) U(du)\right\}, \qquad d \geq 0,$$

where $U(du)$ *is a symmetric non-negative measure on* $\Re \setminus \{0\}$ *satisfying*

$$\int_{\Re\setminus\{0\}} \frac{1}{1+u^2} U(du) < \infty \quad \text{and} \quad \int_{|u|\leq 1} |\log u^2| U(du) < \infty.$$

Hence, it is the convolution of a $N(0, d)$ *and another distribution which has the cf*

$$\exp\left\{\int_{(0,\infty)} \log\left(\frac{u^2}{u^2+s^2}\right) U(du)\right\}.$$

We know that the normal distribution is SD. If we can prove the latter generalized convolution is SD, then the symmetric EGGC is SD. To do so, we consider the base distribution of the latter. It has cf

$$\exp\left\{\log\left(\frac{u^2}{u^2+s^2}\right)\right\} = \frac{u^2}{u^2+s^2} = \frac{u^2/2}{u^2/2+s^2/2}.$$

This can be viewed as the cf of a variance mixture of the normal distribution with representation $X \stackrel{d}{=} \sqrt{Y} Z$, *where* Y *has the LT* $\phi_Y(s) = \frac{u^2/2}{u^2/2+s}$, *the LT of an exponential. The exponential distribution is SD (see Example 7.3). By Theorem 7.1.6, the base distribution is SD. Therefore, by Theorem 7.1.4, the symmetric EGGC is SD.*

Besides the logistic distribution, other common members in this symmetric EGGC family include the t distribution, stable non-Gaussian distribution, and so on.

In fact, in a more general case, EGGC was proved to be SD by Thorin [1978]. Also one can refer to Bondesson [1992], p. 107.

Secondly, we turn to examples of distributions with positive support.

Example 7.3 (Gamma) *Consider the Gamma distribution which has LT*

$$\phi_X(s) = \left(\frac{\beta}{\beta+s}\right)^{\gamma}, \qquad \beta,\ \gamma > 0.$$

Then for any $C > 0$,

$$\exp\left\{C \cdot \frac{s\phi_X'(s)}{\phi_X(s)}\right\} = \exp\left\{C \cdot \frac{-\gamma s}{\beta+s}\right\} = \exp\left\{C\gamma\left[\frac{\beta}{\beta+s} - 1\right]\right\},$$

which is the LT of a compound Poisson with exponential distribution. By Theorem 7.1.2, the Gamma distribution is SD.

Special cases are the exponential and χ^2 distributions.

Example 7.4 (GGC) *It was shown in Bondesson [1992], p. 30 that a GGC distribution is SD. This is simply because that Gamma distribution is SD. We now revisit it from the view of the theory of continuous-time GAR(1) process.*

Recall GGC in Section 2.3.3. It has LT

$$\phi_X(s) = \exp\left\{-as + \int_{(0,\infty)} \log\left(\frac{u}{u+s}\right) U(du)\right\}, \qquad a \geq 0,$$

where the non-negative measure $U(du)$ on $(0,\infty)$ satisfies

$$\int_{(0,1]} |\log u| U(du) < \infty \quad \textit{and} \quad \int_{(1,\infty)} u^{-1} U(du) < \infty.$$

Without loss of generality, we take $a = 0$, because any degenerate rv on a point a is always SD. By calculation, for any $C > 0$, we have

$$\exp\left\{C \cdot \frac{s\phi_X'(s)}{\phi_X(s)}\right\} = \exp\left\{C \int_{(0,\infty)} \frac{-s}{u+s} U(du)\right\} = \exp\left\{C \int_{(0,\infty)} \left[\frac{u}{u+s} - 1\right] U(du)\right\},$$

The conditions imposed on $U(\cdot)$ can lead to $\int_{(0,\infty)} \frac{1}{1+u} U(du) < \infty$. It is indeed the LT of a GCMED (see Example 7.1). By Theorem 7.1.2, GGC is SD.

This big family includes a lot of well known distributions. Common members are Gamma, Pareto, strictly positive stable, lognormal, etc.

Example 7.5 (inverse Gaussian) *Consider the inverse Gaussian which has LT*

$$\phi_X(s) = \exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 + \frac{2\mu^2}{\lambda}s\right)^{1/2}\right]\right\}, \quad \lambda,\ \mu > 0.$$

Then, for any $C > 0$,

$$\exp\left\{C \cdot \frac{s\phi'_X(s)}{\phi_X(s)}\right\} = \exp\left\{-\mu C \cdot s\left(1 + \frac{2\mu^2}{\lambda}s\right)^{-1/2}\right\}.$$

By (6) in Theorem 2.2.8, it is a LT. Therefore, the inverse Gaussian distribution is SD.

Example 7.6 (Mittag-Leffler distribution) *The SD property of Mittag-Leffler distribution can be obtained from Jayakumar and Pillai [1993] where they construct a discrete-time Mittag-Leffler GAR(1) process. Here we will use Theorem 7.1.2 to give another proof.*

The Mittag-Leffler distribution has LT:

$$\phi_X(s) = \frac{1}{1+s^\gamma}, \quad 0 < \gamma < 1.$$

Hence,

$$\frac{s\phi'_X(s)}{\phi_X(s)} = s\frac{-\gamma s^{\gamma-1}(1+s^\gamma)^{-2}}{(1+s^\gamma)^{-1}} = \frac{-\gamma s^\gamma}{1+s^\gamma} = \gamma\left[\frac{1}{1+s^\gamma} - 1\right].$$

This means that for any $C > 0$,

$$\exp\left\{C \cdot \frac{s\phi'_X(s)}{\phi_X(s)}\right\} = \exp\left\{C\gamma\left[\frac{1}{1+s^\gamma} - 1\right]\right\}$$

is the cf of a compound Poisson with the Mittag-Leffler distribution. By Theorem 7.1.2, the Mittag-Leffler distribution is SD.

Lastly, we study examples of distributions with non-negative integer support. We shall use the power of Theorem 7.1.3, which is very convenient when we know the pmf stipulated by $g(s)$ in the exponent. We only apply the arithmetic operation and do comparison, instead of the pgf verification.

Example 7.7 (Poisson/some special compound Poisson) *For Poisson, the pgf is* $G_X(s) = \exp\{\lambda(s-1)\}$ $(\lambda > 0)$*, thus,* $q_1 = 1$*,* $q_k = 0$ $(k > 1)$*. Obviously, the conditions in Theorem 7.1.3 are satisfied. Therefore, Poisson is DSD.*

Following are some examples of compound Poisson distribution.

(1) with Geometric:

The geometric distribution has the pgf

$$g(s) = \frac{1-q}{1-qs}, \qquad 0 \le q < 1.$$

Thus, $q_k = (1-q)q^k$ $(k \ge 1)$*. By algebra, we have*

$$\begin{aligned} kq_k - (k+1)q_{k+1} &= k(1-q)q^k - (k+1)(1-q)q^{k+1} = (1-q)q^k[k-(k+1)q] \\ &= (1-q)q^k(k+1)\left[\frac{k}{k+1} - q\right]. \end{aligned}$$

If $q \le 1/2$*, then* $kq_k - (k+1)q_{k+1} \ge 0$ *for all* $k \ge 1$*. Otherwise, some may be negative. Therefore, when* $q \le 1/2$*, the compound Poisson with geometric is DSD, and when* $q > 1/2$*, the compound Poisson with geometric is not DSD.*

This leads to the result that GC I is DSD if $V(\cdot)$ *has zero measure on* $(1/2, 1)$ *by Theorem 7.1.4. Refer to Section 2.3.3 for the pgf form of GC I.*

The compound Poisson with geometric distribution is the discrete analogue of the compound Poisson with exponential distribution. By Theorem 7.1.7, this result discloses that the continuous analogue in the compound Poisson with exponential distribution (corresponding to $q > 1/2$ *in discrete case) is not SD. However, to directly prove or disprove the SD of the compound Poisson with exponential distribution is not an easy job.*

(2) with another Poisson:

The Poisson distribution has pgf

$$g(s) = e^{\gamma(s-1)}, \qquad \gamma > 0.$$

Thus, $q_k = \frac{\gamma^k}{k!}e^{-\gamma}$ $(k \geq 0)$. *This compound Poisson with another Poisson is called Neyman's Type A distribution. See Johnson and Kotz [1969], p. 186. By algebra, we have*

$$\frac{q_{k+1}}{q_k} = \frac{\frac{\gamma^{k+1}}{(k+1)!}e^{-\gamma}}{\frac{\gamma^k}{k!}e^{-\gamma}} = \frac{\gamma}{k+1}, \qquad k \geq 1.$$

If $\gamma \leq 1$, *then it holds that* $\frac{q_{k+1}}{q_k} \leq \frac{k}{k+1}$ *for all* $k \geq 1$, *which leads to* $kq_k - (k+1)q_{k+1} \geq 0$ *for all* $k \geq 1$. *Otherwise, some inequalities may not hold. Therefore, when* $\gamma \leq 1$, *the Neyman's Type A distribution is DSD; and when* $\gamma > 1$, *the Neyman's Type A distribution is not DSD.*

(3) with Katz family:

It is defined by recursive probability system, i.e., the relationship between two successive probability masses is

$$\frac{q_{k+1}}{q_k} = \frac{\alpha + \beta k}{1+k}, \qquad k = 0, 1, 2, \ldots;\ \alpha > 0;\ \beta < 1.$$

To guarantee that the ratio is non-negative, $k \leq \frac{\alpha}{-\beta}$ *when* $\beta < 0$. *See Winkelmann [1997], p. 36, Johnson and Kotz [1969], p. 37. Some well known members in this family include Poisson, Negative Binomial, geometric, binomial, etc.*

Since $\beta < 1$, $1 - \beta > 0$. *Thus, if* $\alpha + \beta < 1$, *then* $\alpha < 1 - \beta \leq k(1-\beta)$, *for* $k = 1, 2, 3, \ldots$. *This leads to* $\alpha + \beta k \leq k$ *for* $k \geq 1$, *and* $\frac{q_{k+1}}{q_k} \leq \frac{k}{k+1}$, *which indicates that the compound Poisson with Katz family is DSD.*

(4) with Yousry and Srivastava family:

The recursive probability system has been expressed by three parameters as

$$\frac{q_{k+1}}{q_k} = \frac{\alpha + \beta k}{k + \gamma}, \qquad k = 0, 1, 2, \ldots;\ \alpha,\ \gamma > 0;\ \ \beta < 1.$$

This results in the hyper-negative binomial model; see Winkelmann [1997], p. 37. When $\gamma = 1$, *it becomes the Katz family. When* $\beta = 0$, *it leads to the hyper-Poisson distribution (see Johnson and Kotz [1969], p. 43).*

We want $\frac{\alpha+\beta k}{k+\gamma} \leq \frac{k}{1+k}$ *for all* $k \geq 1$. *By algebra, we have*

$$(1-\beta)k^2 - (\alpha + \beta - \gamma)k - \alpha \geq 0, \qquad k \geq 1;$$

$$\left(k - \frac{\alpha+\beta-\gamma}{2(1-\beta)}\right)^2 - \frac{\alpha}{2(1-\beta)} - \left(\frac{\alpha+\beta-\gamma}{2(1-\beta)}\right)^2 \geq 0, \qquad k \geq 1.$$

Let $k_0 = \left[\frac{\alpha+\beta-\gamma}{2(1-\beta)}\right]$, *the integer part. Then the minimum of the left hand side will be reached at* $k = k_0$ *or* $k = k_0 + 1$. *Therefore, if* $\left(k_0 - \frac{\alpha+\beta-\gamma}{2(1-\beta)}\right)^2 - \frac{\alpha}{2(1-\beta)} - \left(\frac{\alpha+\beta-\gamma}{2(1-\beta)}\right)^2 \geq 0$ *and* $\left(k_0 + 1 - \frac{\alpha+\beta-\gamma}{2(1-\beta)}\right)^2 - \frac{\alpha}{2(1-\beta)} - \left(\frac{\alpha+\beta-\gamma}{2(1-\beta)}\right)^2 \geq 0$, *then the compound Poisson with the Yousry and Srivastava family is DSD.*

(5) with the Kulasekera and Tonkyn family:

The recursive probability system for this family has been formulated as

$$\frac{q_{k+1}}{q_k} = \beta\left(\frac{1+k}{k}\right)^\alpha, \qquad k = 1, 2, \ldots;\ \alpha \in \Re,\ 0 < \beta < 1.$$

See Winkelmann [1997], p. 37. It includes the shifted negative binomial, the logarithmic series and the discrete Pareto distribution.

We want $\beta\left(\frac{1+k}{k}\right)^\alpha \leq \frac{k}{1+k}$ *for all* $k \geq 1$. *This is equivalent to* $\beta \leq \left(\frac{k}{1+k}\right)^{\alpha+1}$ *for all* $k \geq 1$. *Thus, it follows that for* $\alpha + 1 \geq 0$, $\beta \leq \left(\frac{1}{2}\right)^{\alpha+1}$, *or for* $\alpha + 1 < 0$, $\beta \leq 1$. *But in the second case, it always holds because* $\beta < 1$. *Therefore, when* $\alpha + 1 \geq 0$ *and* $\beta \leq \left(\frac{1}{2}\right)^{\alpha+1}$, *or when* $\alpha + 1 < 0$, *the compound Poisson with the Kulasekera and Tonkyn family is DSD.*

Example 7.8 (generalized Poisson) *The pgf is*

$$G_X(s) = \exp\left\{\theta\left(\sum_{k=1}^{\infty}\eta(k\eta)^{k-1}e^{-k\eta}s^k/k! - 1\right)\right\}, \qquad \text{where } \theta > 0,\ 0 \leq \eta \leq 1.$$

Thus, $q_k = \eta(k\eta)^{k-1}e^{-k\eta}/k!$ $(k \geq 1)$. *By algebra, we know that*

$$kq_k - (k+1)q_{k+1} = \frac{k^k\eta^k e^{-k\eta}}{k!} - \frac{(k+1)^{k+1}\eta^{k+1}e^{-(k+1)\eta}}{(k+1)!} = \frac{k^k\eta^k e^{-k\eta}}{k!}\left[1 - \eta e^{-\eta}\left(1 + \frac{1}{k}\right)^k\right].$$

It is well known that $\left(1 + \frac{1}{k}\right)^k$ *is increasing to* e *as* $k \to \infty$. *Thus,*

$$1 - \eta e^{-\eta}\left(1 + \frac{1}{k}\right)^k \geq 1 - \eta e^{1-\eta} \stackrel{\text{def}}{=} l(\eta).$$

Since $l'(\eta) = -e^{1-\eta} + \eta e^{1-\eta} = -e^{1-\eta}(1-\eta) \leq 0$, $l(\eta) \geq l(1) = 1 - 1 \cdot e^{1-1} = 1 - 1 = 0$. *This leads to* $kq_k - (k+1)q_{k+1} \geq 0$ *for* $k \geq 1$. *Therefore, the generalized Poisson distribution is DSD.*

Example 7.9 (negative binomial) *The pgf is*

$$G_X(s) = \left(\frac{1-p}{1-ps}\right)^{\gamma} = \exp\left\{\gamma\left[\log(1-p)^{-1}\right]\left[\frac{\log(1-ps)}{\log(1-p)} - 1\right]\right\}, \qquad \gamma > 0,\ 0 < p < 1.$$

Thus $q_k = \frac{p^k}{-k\log(1-p)}$ $(k \geq 1)$. *This leads to that for* $k \geq 1$,

$$kq_k - (k+1)q_{k+1} = k\frac{p^k}{-k\log[(1-p)]} - (k+1)\frac{p^{k+1}}{-(k+1)\log[(1-p)]} = \frac{p^k}{-\log(1-p)}(1-p) \geq 0.$$

Hence, the negative binomial distribution is DSD.

Example 7.10 (discrete stable) *The pgf is*

$$G_X(s) = \exp\{-\lambda(1-s)^{\theta}\} = \exp\{\lambda[1-(1-s)^{\theta} - 1]\}, \qquad \lambda > 0,\ 0 < \theta \leq 1.$$

Hence, $q_k = -\prod_{i=1}^{k}(i-1-\theta)/k!$ $(k \geq 1)$. *It follows that for* $k \geq 1$,

$$\begin{aligned}
kq_k - (k+1)q_{k+1} &= k\frac{-\prod_{i=1}^{k}(i-1-\theta)}{k!} - (k+1)\frac{-\prod_{i=1}^{k+1}(i-1-\theta)}{(k+1)!} \\
&= \frac{-\prod_{i=1}^{k}(i-1-\theta)}{(k-1)!}\left[1 - \frac{k-\theta}{k}\right] = \frac{-\theta\prod_{i=1}^{k}(i-1-\theta)}{k!} \\
&\geq 0.
\end{aligned}$$

This shows that discrete stable distribution is DSD.

Note that the negative binomial and discrete stable distributions are the discrete analogues of the Gamma and positive stable distribution respectively. Since the Gamma and positive stable distributions are SD, we can also conclude that the negative binomial and discrete stable distributions are DSD by Theorem 7.1.7. Sometimes, like in the situation of logarithmic series and power series distribution, because it is very difficult to directly prove that $G_X(s)/G_X(1-c+cs)$ is a pgf, we have to resort to Theorem 7.1.2 to prove the DSD feature.

Example 7.11 (discrete Mittag-Leffler distribution) *The DSD property of the discrete Mittag-Leffler distribution can be derived from Pillai and Jayakumar [1995]. It also can be obtained by*

Theorem 7.1.7 for the Mittag-Leffler distribution is SD. Now we try to prove it using Theorem 7.1.2.

The pgf of the discrete Mittag-Leffler distribution is

$$G_X(s) = \frac{1}{1 + d(1-s)^\gamma}, \qquad d > 0,\ 0 < \gamma \leq 1.$$

Let $g(s) = 1 + \gamma^{-1}(s-1)G'_X(s)/G_X(s)$. *Then, it follows that*

$$g(s) = 1 + \gamma^{-1}(s-1)\frac{\gamma d(1-s)^{\gamma-1}}{1+d(1-s)^\gamma} = 1 - \frac{d(1-s)^\gamma}{1+d(1-s)^\gamma} = \frac{1}{1+d(1-s)^\gamma}.$$

This completes the proof.

Example 7.12 (power series distribution) *The pgf has form*

$$G_X(s) = s^{-1}[1 - (1-s)^\theta], \qquad 0 < \theta \leq 1.$$

Consider the function $1 + (s-1)\frac{G'_X(s)}{G_X(s)} = \frac{1}{s} - \frac{\theta(1-s)^\theta}{1-(1-s)^\theta}$. *By (1) in Theorem 2.2.3, it is a pgf. Therefore, the power series distribution is DSD.*

Example 7.13 (Zeta distribution) *The pgf* $G_X(s)$ *does not have closed form. However,* $1+(s-1)\frac{G'_X(s)}{G_X(s)}$ *is a pgf. See (4) in Theorem 2.2.3. This implies that Zeta distribution is DSD.*

Example 7.14 (GNBC) *Because the GNBC is the limit distribution of sums of independent negative binomial rv's, and NB distribution is DSD, the GNBC class is DSD.*

Taking advantage of SD or DSD feature of a distribution, we sometimes can prove new LT or pgf. For example, suppose Zeta distribution is known to be DSD, then we can conclude that $L(s) = 1 + C(s-1)\frac{G'_X(s)}{G_X(s)}$ (C chosen so that $L(0) \geq 0$) is a pgf, which we had spent a lot energy and time to prove in Section 2.2.1.

Now we finish this section with the discussion about the Tweedie exponential dispersion family. Tweedie model can be categorized as extreme stable, positive stable, Gamma, compound Poisson with Gamma, inverse Gaussian, Normal and Poisson. From previous examples, we know that the Poisson distribution is DSD, and the extreme stable, positive stable, Gamma, inverse

Gaussian and Normal distributions are SD. The only category left is the compound Poisson with Gamma, which can not always be SD but a subfamily of it is SD. One can refer to Example 7.7 to see the DSD feature of the compound Poisson with the geometric distribution, which is the discrete analogue of the compound Poisson with the exponential distribution. Only part of the family is DSD. Hence, we can not expect the whole family of the compound Poisson with Gamma to be SD. This fact tells us that constructing a steady state continuous-time GAR(1) process with the margins of the compound Poisson with the Gamma distribution may be futile if we just consider the constant multiplier operation. A possible approach to solve this problem may be to resort to the GSD classes proposed in the next section.

7.2 Generalized self-decomposable, generalized discrete self-decomposable classes and their infinite divisibility property

In the previous section, we have considered the pair of binomial-thinning and constant multiplier stochastic operation **(I1, P1)**, which induce the DSD and SD. Now we turn to other pairs of extended-thinning operations: **(I2, P2)**, **(I3, P3)**, **(I4, P4)**, **(I5, P5)**. These will lead to new concepts.

In general, for a continuous-time GAR(1) process under steady state, the following kind of decomposition holds:

$$X \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X + \int_0^{t_2-t_1} \left(e^{-\mu t}\right)_K \circledast d\epsilon(t), \qquad \mu > 0, \quad t_1 < t_2,$$

or in another form

$$X \stackrel{d}{=} (c)_K \circledast X + E(t_1; t_2), \qquad c \in (0,1).$$

With K is from **(I1, P1)**, they are called DSD and SD. Naturally, such a question arises: how about other extended-thinning operations? It seems that this is new to researchers.

Therefore, we introduce the generalized self-decomposability and generalized discrete self-de-composability in the following.

Definition 7.1 (Generalized discrete self-decomposability (GDSD)) *Suppose $X \sim F$. Let $\{K(\alpha)\}$ be a family of self-generalized distributions with non-negative integer support. If for each c, $0 \le c \le 1$, there exists a non-negative integer-valued rv ϵ_c such that*

$$X \stackrel{d}{=} (c)_K \circledast X + \epsilon_c,$$

where ϵ_c is independent of X, then the probability distribution F is called generalized discrete self-decomposable **(GDSD)** *with respect to $\{K(\alpha)\}$.*

This definition is equivalent to that $G_X(s)/G_X(G_K(s;c))$ is a pgf for each $0 < c < 1$.

Definition 7.2 (Generalized self-decomposability) *Suppose $X \sim F$. Let $\{K(\alpha)\}$ be a family of self-generalized distributions with positive support. If for each c; $0 \le c \le 1$, there exists a non-negative rv ϵ_c such that*

$$X \stackrel{d}{=} (c)_K \circledast X + \epsilon_c,$$

where ϵ_c is independent of X, then the probability distribution F is said to be generalized self-decomposable **(GSD)** *with respect to $\{K(\alpha)\}$.*

An equivalent definition is that X is GSD with respect to $\{K(\alpha)\}$ iff for each c, $0 \le c \le 1$, $\phi_X(s)/\phi_X(-\log \phi_K(s;c))$ is the LT of a probability distribution.

An obvious fact is that for a non-negative integer GDSD(DSD) rv X or positive-valued GDS(SD) rv X, it is always stochastically larger than $(c)_K \circledast X$, i.e.,

$$(c)_K \circledast X \prec^{st} X.$$

This is because $\epsilon_c \ge 0$ and

$$\Pr[X \le x] \;=\; \Pr\left[(c)_K \circledast X + \epsilon_c \le x\right] \;\le\; \Pr[(c)_K \circledast X \le x].$$

The self-generalized distribution family may consist of many members although currently we only know of the subclasses from **I1** to **P5**. Each subclass could be associated with a fixed parameter or parameter vector. For example, the distributions from **I2** have an associated fixed parameter γ where $0 \leq \gamma < 1$, and the distributions from **I5** have an associated fixed parameter vector (γ, θ) where $0 \leq \gamma < 1$ and $\theta \geq 1$. To distinguish these GDSD and GSD families associated with different self-generalized distributions as well as their fixed parameter or parameter vector, we shall adopt the notations like: GDSD(**I2**(γ)), GDSD(**I3**), GDSD(**I4**(θ)), GDSD(**I5**(γ, θ)), and GSD(**P2**(γ)), GSD(**P3**), GSD(**P4**(θ)), GSD(**P5**(γ, θ)) to clearly indicate the attribute of associated self-generalized distribution and its corresponding fixed parameter or parameter vector. The label number can be extended for classes of self-generalized distributions discovered in the future. Hence, we generally denote a specific GDSD or GSD class as GDSD($\mathbf{Ii}(\boldsymbol{\theta})$) and GSD($\mathbf{Pi}(\boldsymbol{\theta})$) respectively. Here $\boldsymbol{\theta}$ is the corresponding fixed parameter or parameter vector. In addition, the union of all members from a self-generalized distribution family over the space of the fixed parameter or parameter vector is denoted as GDSD(**Ii**) or GSD(**Pi**), namely

$$\mathrm{GDSD}(\mathbf{Ii}) = \bigcup_{\boldsymbol{\theta}} \mathrm{GDSD}(\mathbf{Ii}(\boldsymbol{\theta})) \qquad \text{and} \qquad \mathrm{GDSD}(\mathbf{Pi}) = \bigcup_{\boldsymbol{\theta}} \mathrm{GDSD}(\mathbf{Pi}(\boldsymbol{\theta})).$$

In previous sections, we have seen many examples where stationary distributions of the continuous-time generalized AR(1) processes exist; these mean that the resulting GDSD or GSD classes corresponding to their extended-thinning operations are not empty. However, we are not sure if the extended-thinning operation associated with a family of self-generalized distributions leads to a GDSD or GSD class. Perhaps some of these classes are empty. The following theorem presents a necessary condition for non-empty GDSD or GSD classes.

Theorem 7.2.1 *Suppose a family of self-generalized distributions have pgf* $G_K(s;\alpha)$ *or LT* $\phi_K(s;\alpha)$.

(1) A necessary condition for the existence of a GDSD class is that $G_K(s;\alpha) \geq s$ *for all* $0 < s < 1$ *and* $0 < \alpha < 1$.

(2) A necessary condition for the existence of a GSD class is that $\phi_K(s;\alpha) \geq e^{-s}$ *for all* $s > 0$ *and* $0 < \alpha < 1$.

Proof: The proofs of the two cases have the same reasoning. To save space, we only show the proof of (1).

If X is GDSD with respect to the self-generalized distribution family, then $X \stackrel{d}{=} (\alpha)_K * X + \epsilon_\alpha$ for all $0 < \alpha < 1$, where ϵ_α is a non-negative integer random variable. Hence $(\alpha)_K * X \prec^{st} X$ ($(\alpha)_K * X$ is stochastically smaller than X), and

$$G_{(\alpha)_K * X}(s) = \mathbf{E}\left(s^{(\alpha)_K * X}\right) \geq \mathbf{E}(s^X) = G_X(s), \qquad \text{for all } 0 < s < 1,$$

because $h(x) = s^x$ is decreasing in x for all $0 < s < 1$. Since $G_{(\alpha)_K * X}(s) = G_X(G_K(s;\alpha))$ and $G_X(s)$ is increasing in s, $G_X(G_K(s;\alpha)) \geq G_X(s)$ iff $G_K(s;\alpha) \geq s$.

In the study of SD and DSD, we found that the discrete analogue of positive SD is DSD. This is also true for each pair {GDSD($\mathbf{Ii}(\boldsymbol{\theta})$), GSD($\mathbf{Pi}(\boldsymbol{\theta})$)}.

Theorem 7.2.2 *If the positive rv X is GSD($\mathbf{Pi}(\boldsymbol{\theta})$), then its discrete analogue Y is GDSD($\mathbf{Ii}(\boldsymbol{\theta})$). (see Definition 2.6 for the discrete analogue).*

Proof: Assume X has LT $\phi_X(s)$ $(s > 0)$ and Y has pgf $G_Y(s)$ $(0 < s < 1)$, and $G_Y(s) = \phi_X(1 - ds)$ $(0 < s < 1)$. Thus, $\phi_X(s) = G_Y(1 - s/d)$. For each pair of the self-generalized distribution with non-negative integer and positive support **(Ii, Pi)**, the relationship between their pgf and LT are

$$-\log \phi_{K_1}(s;\alpha) = 1 - G_{K_2}(1 - s;\alpha).$$

where K_1 is from $(\mathbf{Pi}(\boldsymbol{\theta}))$ and K_2 is from GDSD($\mathbf{Ii}(\boldsymbol{\theta})$); see Section 3.3. It follows that for $0 < c < 1$, with $d = 1$,

$$g(s) \stackrel{\text{def}}{=} \frac{G_Y(s)}{G_Y(G_{K_2}(s;c))} = \frac{\phi_X(1-s)}{\phi_X(1 - G_{K_2}(s;c))} = \frac{\phi_X(1-s)}{\phi_X(-\log \phi_{K_1}(1-s;c))},$$

and

$$h(s) \stackrel{\text{def}}{=} \frac{\phi_X(s)}{\phi_X(-\log \phi_{K_1}(s;c))} = \frac{G_Y(1-s)}{G_Y(1 + \log \phi_{K_1}(s;c))} = \frac{G_Y(1-s)}{G_Y(G_{K_2}(1-s;c))}.$$

Thus, $g(s) = h(1-s)$ and $h(s) = g(1-s)$, $0 < s < 1$.

If X is GSD($\mathbf{Pi}(\boldsymbol{\theta})$), then $h(s) = \frac{\phi_X(s)}{\phi_X(-\log \phi_{K_1}(s;c))}$ is a LT. Now we need to prove that $g(s) = h(1-s)$ is a pgf. First, $g(0) = h(1) \geq 0$ and $g(1) = h(0) = 1$. Secondly, we take derivatives for $g(s)$: $g^{(n)}(s) = (-1)^n d^n h^{(n)}(1-s) \geq 0$ for $n \geq 1$. These indicates that $g(s)$ is indeed a pgf, so Y is GDSD($\mathbf{Ii}(\boldsymbol{\theta})$).

However, the converse is not true. A counterexample is shown in Example 7.18.

The GDSD and GSD classes are always associated with a self-generalized distribution and its fixed parameter or parameter vector. Bearing this fact in mind, we shall suppress the associated self-generalized distribution and its fixed parameter or parameter vector in the rest of this section, i.e., leaving $\mathbf{Ii}(\boldsymbol{\theta})$ and $\mathbf{Pi}(\boldsymbol{\theta})$ out from the previous notations, unless we have special reason to address them. This may lead to simpler writing.

Similar to Theorem 7.1.2, we have the following result, which is useful in proving the GSD and GDSD.

Theorem 7.2.3 *Let X be a non-negative rv.*

(1) Assume X has positive support with LT $\phi_X(s)$. Let K be a positive self-generalized rv, and $H(s) = \frac{\partial}{\partial\alpha}[-\log\phi_K(s,\alpha)]\Big|_{\alpha=1}$. If $\exp\left\{C\cdot\frac{H(s)\phi_X'(s)}{\phi_X(s)}\right\}$ is a LT for all $C>0$, then X is GSD.

(2) Assume X has non-negative integer support with pgf $G_X(s)$. Let K be a non-negative integer self-generalized rv, and $H(s) = \frac{\partial G_K(s,\alpha)}{\partial\alpha}\Big|_{\alpha=1}$. If $1 + C\cdot\frac{H(s)G_X'(s)}{G_X(s)}$ is a pgf for some $C>0$, then X is GDSD.

Proof: Like Theorem 7.1.2, for any $0<c<1$, we express $\phi(s)/\phi(-\log\phi_K(s;))$, or $G(s)/G(G_K(s;c))$ in terms of $\exp\left\{C\cdot\frac{s\phi_X'(s)}{\phi_X(s)}\right\}$, or $1 + C\cdot\frac{H(s)G_X'(s)}{G_X(s)}$.

(1) First, it follows that

$$\begin{aligned}
H(-\log\phi_K(s;\beta)) &= \frac{\partial}{\partial\alpha}[-\log\phi_K(-\log\phi_K(s;\beta),\alpha)]\Big|_{\alpha=1} \\
&= \frac{\partial}{\partial\alpha}[-\log\phi_K(s;\alpha\beta)]\Big|_{\alpha=1} = \left(\frac{\partial}{\partial\gamma}[-\log\phi_K(s;\gamma)]\Big|_{\gamma=\beta}\right)\beta \\
&= \left(\frac{\partial}{\partial\beta}[-\log\phi_K(s;\beta)]\right)\beta.
\end{aligned}$$

Since $\exp\left\{C\cdot\frac{H(s)\phi_X'(s)}{\phi_X(s)}\right\}$ is a LT for any $C>0$,

$$\exp\left\{\int_c^1 \frac{H(-\log\phi_K(s;\beta))\phi_X'(-\log\phi_K(s;\beta))}{\phi_X(-\log\phi_K(s;\beta))}\frac{1}{\beta}d\beta\right\}$$

is a LT too. This comes from the idea of generalized convolution. We compute the integration in the exponent,

$$\begin{aligned}
&\exp\left\{\int_c^1 \frac{H(-\log\phi_K(s;\beta))\phi'_X(-\log\phi_K(s;\beta))}{\phi_X(-\log\phi_K(s;\beta))}\frac{1}{\beta}d\beta\right\} \\
&= \exp\left\{\int_c^1 \frac{\beta\left(\frac{\partial}{\partial\beta}[-\log\phi_K(s;\beta)]\right)\phi'_X(-\log\phi_K(s;\beta))}{\phi_X(-\log\phi_K(s;\beta))}\frac{1}{\beta}d\beta\right\} \\
&= \exp\left\{\int_c^1 \frac{\phi'_X(-\log\phi_K(s;\beta))}{\phi_X(-\log\phi_K(s;\beta))}d(-\log\phi_K(s;\beta))\right\} \\
&= \exp\left\{\int_c^1 d\log\phi_X(-\log\phi_K(s;\beta))\right\} \\
&= \exp\left\{\log\phi_X(-\log\phi_K(s;\beta))\Big|_c^1\right\} \\
&= \exp\{\log\phi_X(s) - \log\phi_X(-\log\phi_K(s;c))\} \\
&= \frac{\phi_X(s)}{\phi_X(-\log\phi_K(s;c))},
\end{aligned}$$

which means that $\frac{\phi_X(s)}{\phi_X(-\log\phi_K(s;c))}$ is a LT. Hence, X is GSD.

(2) Similarly, it follows that

$$\begin{aligned}
H(G_K(s;\beta)) &= \frac{\partial}{\partial\alpha}[G_K(G_K(s;\beta),\alpha)]\Big|_{\alpha=1} \\
&= \frac{\partial}{\partial\alpha}[G_K(s;\alpha\beta)]\Big|_{\alpha=1} = \left(\frac{\partial}{\partial\gamma}[G_K(s;\gamma)]\Big|_{\gamma=\beta}\right)\beta \\
&= \beta\left(\frac{\partial}{\partial\beta}[G_K(s;\beta)]\right).
\end{aligned}$$

Since $1 + C\cdot\frac{H(s)G'_X(s)}{G_X(s)}$ is a pgf for some $C > 0$,

$$1 + C\cdot\frac{H(G_K(s;\beta))G'_X(G_K(s;\beta))}{G_X(G_K(s;\beta))}$$

from the extended-thinning operation, is a pgf too. By the mixture operation in (3) of Proposition 2.2.2, it follows that

$$g(s) = [-\log c]^{-1}\int_c^1\left(1 + C\cdot\frac{H(G_K(s;\beta))G'_X(G_K(s;\beta))}{G_X(G_K(s;\beta))}\right)\frac{1}{\beta}d\beta$$

is a pgf. By algebra, we obtain

$$\begin{aligned} g(s) &= [-\log c]^{-1} \int_c^1 \left[C \cdot \frac{H(G_K(s;\beta))G'_X(G_K(s;\beta))}{G_X(G_K(s;\beta))} + 1 \right] \beta^{-1} d\beta \\ &= [-\log c]^{-1} \left\{ C \int_c^1 \frac{\frac{\partial G_K(s;\beta)}{\partial \beta} G'_X(G_K(s;\beta))}{G_X(G_K(s;\beta))} d\beta + \int_c^1 d(\log \beta) \right\} \\ &= [-\log c]^{-1} C \log \frac{G_X(s)}{G_X(G_K(s;c))} + 1. \end{aligned}$$

Thus

$$\frac{G_X(s)}{G_X(G_K(s;c))} = \exp\left\{\log(c^{-1})C^{-1}[g(s)-1]\right\},$$

the pgf of a compound Poisson distribution. Hence, X is GDSD.

Remark: Theorem 6.3.1 is the necessary part of the results, while Theorem 7.2.3 is the sufficient part. Also the conditions in the following Theorem 7.2.5 are necessary too when we pick up K from **I2** in Theorem 6.3.1.

Corollary 7.2.4 *Let $\{K(\alpha) : 0 < \alpha \leq 1\}$ be a family of self-generalized rv's.*

(1) If it leads to extended thinning operators and a GSD class, then

$$H(s) = \frac{\partial}{\partial \alpha} [-\log \phi_K(s, \alpha)] \Big|_{\alpha=1} \geq 0, \qquad 0 \leq s < \infty.$$

(2) If it leads to extended thinning operators and a GDSD class, then

$$H(s) = \frac{\partial G_K(s;\alpha)}{\partial \alpha} \Big|_{\alpha=1} \leq 0, \qquad 0 \leq s \leq 1.$$

Proof:

(1) Since $\phi'_X(s) < 0$, $\frac{\phi'_X(s)}{\phi_X(s)} < 0$. To guarantee that $\exp\left\{C \cdot \frac{H(s)\phi'_X(s)}{\phi_X(s)}\right\}$ is a LT for all $C > 0$, it must hold that $H(s) \geq 0$ for all $0 \leq s < \infty$.

(2) Because $L(s) = 1 + C \cdot H(s)G'_X(s)/G_X(s)$ is a pgf for some $C > 0$, $0 \leq L(s) \leq 1$. Since $G'_X(s) \geq 0$, $G_X(s) > 0$ for all $0 \leq s \leq 1$, it must hold that $H(s) \leq 0$.

In summary, both cases show that $\exp\left\{\frac{H(s)\phi'_X(s)}{\phi_X(s)}\right\}$ and $\exp\left\{\frac{H(s)G'_X(s)}{G_X(s)}\right\}$ are the LT and pgf of ID distributions. Thus, for K being a positive-valued self-generalized rv, according to Theorem 2.2.6, a necessary and sufficient condition for X being GSD is that $H(s)\frac{\phi'_X(s)}{\phi_X(s)}$ is completely monotone. For K being a non-negative integer self-generalized rv, we can further derive the results similar to Theorem 7.1.3. Here we give the corresponding result for K being from **I2**.

Theorem 7.2.5 *For a compound Poisson rv X with pgf*

$$G_X(s) = \exp\{\lambda[g(s)-1]\}, \quad \lambda > 0,$$

where $g(s) = q_0 + q_1 s + q_2 s^2 + \ldots + q_k s^k + \ldots$, a pgf on the non-negative integers. If

$$(1+\gamma)q_1 - 2q_2 \geq 0, \quad k(1+\gamma)q_k - (k-1)\gamma q_{k-1} - (k+1)q_{k+1} \geq 0, \qquad k = 2, 3, \ldots,$$

then X is GDSD$(\mathbf{I2}(\gamma))$.

Proof: Consider $1 + C \cdot \frac{(1-\gamma s)(s-1)G'_X(s)}{(1-\gamma)G_X(s)}$. By algebra, we obtain

$$\begin{aligned}
1 + \frac{C}{1-\gamma} \cdot \frac{(1-\gamma s)(s-1)G'_X(s)}{G_X(s)} &= 1 + \frac{\lambda C}{1-\gamma}(1-\gamma s)(s-1)g'(s) \\
&= 1 + \frac{\lambda C}{1-\gamma}(1-\gamma s)(s-1)[q_1 + 2q_2 s + 3q_3 s^2 + \ldots + kq_k s^{k-1} + \ldots] \\
&= 1 + \frac{\lambda C}{1-\gamma}(1-\gamma s)\left(-q_1 + \sum_{k=1}^{\infty}[kq_k - (k+1)q_{k+1}]s^k\right) \\
&= 1 + \frac{\lambda C}{1-\gamma}\Big(- q_1 + [(1+\gamma)q_1 - 2q_2]s + [2(1+\gamma)q_2 - \gamma q_1 - 3q_3]s^2 \\
&\quad + [3(1+\gamma)q_3 - 2\gamma q_2 - 4q_4]s^3 + \cdots + [k(1+\gamma)q_k - (k-1)\gamma q_{k-1} - (k+1)q_{k+1}]s^k + \cdots\Big) \\
&= \left(1 - \frac{C\lambda}{1-\gamma}q_1\right) + \frac{\lambda C}{1-\gamma}\Big([(1+\gamma)q_1 - 2q_2]s + [2(1+\gamma)q_2 - \gamma q_1 - 3q_3]s^2 \\
&\quad + [3(1+\gamma)q_3 - 2\gamma q_2 - 4q_4]s^3 + \cdots + [k(1+\gamma)q_k - (k-1)\gamma q_{k-1} - (k+1)q_{k+1}]s^k + \cdots\Big) \\
&= \left(1 - \frac{C\lambda}{1-\gamma}q_1\right) + \frac{\lambda C}{1-\gamma}\sum_{k=1}^{\infty} s^k\left[k(1+\gamma)q_k - (k-1)\gamma q_{k-1} - (k+1)q_{k+1}\right].
\end{aligned}$$

Thus, that it is a pgf is equivalent to

$$1 - \frac{C\lambda}{1-\gamma}q_1 \geq 0, \quad k(1+\gamma)q_k - (k-1)\gamma q_{k-1} - (k+1)q_{k+1} \geq 0, \qquad k = 1, 2, 3, \ldots.$$

We can always take $C = \frac{1-\gamma}{\lambda}$ to guarantee $1 - \frac{C\lambda}{1-\gamma} q_1 = 1 - q_1 \geq 0$. This completes the proof.

Recall that GDSD($\mathbf{I2}(\gamma)$) is induced by a non-negative integer self-generalized rv K from $\mathbf{I2}$, which has pgf of form

$$G_K(s;\alpha) = \frac{(1-\alpha) + (\alpha-\gamma)s}{(1-\alpha\gamma) - (1-\alpha)\gamma s}, \qquad 0 \leq \alpha \leq 1,$$

where γ is fixed and $0 \leq \gamma < 1$. Hence, the GDSD($\mathbf{I2}$) is a big class consisting of subclasses associated with the fixed parameter γ. Thus,

$$\text{GDSD}(\mathbf{I2}) = \bigcup_{\gamma \in [0,1)} \text{GDSD}(\mathbf{I2}(\gamma)).$$

Since when $\gamma = 0$, K becomes from $\mathbf{I1}$, we have DSD=GDSD($\mathbf{I2}(0)$) as the special boundary case.

Corollary 7.2.6 *For a compound Poisson rv X with pgf*

$$G_X(s) = \exp\{\lambda[g(s) - 1]\}, \quad \lambda > 0,$$

where $g(s) = q_0 + q_1 s + q_2 s^2 + \ldots + q_k s^k + \ldots$, a pgf on the non-negative integers. If $q_1 - 2q_2 \geq 0$ and X is GDSD($\mathbf{I2}(\gamma)$), then X is DSD.

Proof: We can rewrite the conditions in Corollary 7.2.5 as

$$q_1 - 2q_2 \geq -\gamma q_1, \quad kq_k - (k+1)q_{k+1} \geq \gamma[(k-1)q_{k-1} - kq_k], \qquad k = 2, 3, \ldots.$$

Since X is GDSD($\mathbf{I2}(\gamma)$) and $q_1 - 2q_2 \geq 0$, we can conclude by induction that

$$kq_k - (k+1)q_{k+1} \geq 0, \qquad k = 1, 2, 3, \ldots.$$

By Theorem 7.1.3, X is DSD.

As for the ID property, Steutel and van Harn [1979] (Theorem 2.2) proved that a DSD distribution is ID. Their idea was to express the pgf, $G(s)$, of a DSD distribution in an exponential form which is the pgf of a compound Poisson. Specifically, it takes advantage of Theorem 6.3.1 so that $\exp\left\{\frac{H(s)G'_X(s)}{G_X(s)}\right\}$ is the pgf of an ID distribution. Thus, it follows that

$$\exp\left\{\frac{H(s)G'_X(s)}{G_X(s)}\right\} = \exp\{\lambda[g(s) - 1]\},$$

where $g(s) = q_0 + q_1 s + q_2 s^2 + \ldots + q_k s^k + \ldots$, a pgf. This indicates that

$$\frac{H(s)G'_X(s)}{G_X(s)} = \lambda[g(s) - 1], \qquad \text{or} \qquad \frac{G'_X(s)}{G_X(s)} = \frac{\lambda[g(s) - 1]}{H(s)}.$$

By integration,

$$\int_s^1 \frac{G'_X(u)}{G_X(u)} du = \log(G_X(u))\Big|_s^1 = -\log G_X(s) = \int_s^1 \frac{\lambda[g(u) - 1]}{H(u)} du.$$

This leads to $G_X(s) = \exp\left\{-\int_s^1 \frac{\lambda[g(u) - 1]}{H(u)} du\right\}$. For DSD, $H(u) = u - 1$. We can expand $\frac{\lambda[g(u)-1]}{H(u)}$ in term of u and integrate it to obtain a new power series in s. All coefficients in the new power series are easily proved to be non-negative and the sum of them are finite. However, for other self-generalized distributions with non-negative integer support, it is not easy to generalize this idea, because we are not clear the form of $H(u)$. Hence, their method is restricted by the form of $H(u)$. To prove the ID property for a GDSD distribution, we have to resort to other approaches.

Feller [1966b], p. 550-555, proved that a SD distribution is ID by taking advantage of a "null array", which is a special triangular array:

$$X_{1,n}, \quad X_{2,n}, \quad X_{3,n}, \quad \ldots, \quad X_{r_n,n}; \qquad n = 1, 2, 3, \ldots, \quad r_n \text{ is finite.}$$

All of them are independent rv's. This "null array" is defined as that for given $\epsilon > 0$ and $s_0 > 0$, it has

$$|1 - \varphi_{X_{i,n}}(s)| < \epsilon, \qquad |s| < s_0, \quad i = 1, 2, \ldots, r_n;$$

for all n sufficiently large. Denote the row sum $S_n = X_{1,n} + \cdots + X_{r_n,n}$. Then, Theorem 1, on page 550 in Feller [1966b], shows that if $S_n + \beta_n$ tends in distribution to a rv U, then U is ID. Here $\{\beta_n; n \in \mathcal{N}\}$ is a sequence of real constants. Without any difficulty, we can modify the definition of "null array" for a non-negative integer rv triangular array and a positive rv triangular array by replacing their cf's with their pgf's or LT's respectively, namely for given $\epsilon > 0$ and $s_0 > 0$,

$$0 < 1 - G_{X_{i,n}}(s) < \epsilon, \qquad s_0 < s \leq 1, \quad i = 1, 2, \ldots, r_n;$$

and

$$0 < 1 - \phi_{X_{i,n}}(s) < \epsilon, \qquad 0 \leq s < s_0, \quad i = 1, 2, \ldots, r_n;$$

for all n sufficiently large respectively. In principle, these new definitions are equivalent to their original definitions. Therefore, for these two particular types of "null arrays", Theorem 1 on page 550 in Feller [1966b] is still valid.

In the following theorem, we shall follow Feller's idea to show the ID property for both of the GDSD and GSD classes.

Theorem 7.2.7 *Let K be from a self-generalized distribution. It follows that*

(1) if K is a non-negative integer rv with pgf $G_K(s;\alpha)$ continuous in α in $[0,1]$, then a GDSD distribution is ID;

(2) if K is a positive rv with LT $\phi_K(s;\alpha)$ continuous in α in $[0,1]$, then a GSD distribution is ID.

Proof:

(1) Assume $G(s) = G_X(s)$ is the pgf of a GDSD distribution with non-negative integer support. X is a rv from this distribution. Note that $G_K(s;0) = 1$ and $G_K(s;1) = s$. Since $G_K(s;\alpha)$ is continuous in α in $[0,1]$, The following identity always holds:

$$\begin{aligned} G(s) &= G\left(G_K\left(s;\frac{1}{n}\right)\right) \times \frac{G\left(G_K\left(s;\frac{2}{n}\right)\right)}{G\left(G_K\left(s;\frac{1}{n}\right)\right)} \times \cdots \times \frac{G\left(G_K\left(s;\frac{n-1}{n}\right)\right)}{G\left(G_K\left(s;\frac{n-2}{n}\right)\right)} \times \frac{G\left(s\right)}{G\left(G_K\left(s;\frac{n-1}{n}\right)\right)} \\ &= \frac{G\left(G_K\left(s;\frac{1}{n}\right)\right)}{G\left(G_K\left(s;\frac{0}{n}\right)\right)} \times \frac{G\left(G_K\left(s;\frac{2}{n}\right)\right)}{G\left(G_K\left(s;\frac{1}{n}\right)\right)} \times \cdots \times \frac{G\left(G_K\left(s;\frac{n-1}{n}\right)\right)}{G\left(G_K\left(s;\frac{n-2}{n}\right)\right)} \times \frac{G\left(G_K\left(s;\frac{n}{n}\right)\right)}{G\left(G_K\left(s;\frac{n-1}{n}\right)\right)}, \end{aligned} \tag{7.2.1}$$

where $n = 1, 2, 3, \ldots$. When X is GDSD, this identity has the following explanation:

$$\begin{aligned} X &\stackrel{d}{=} \left(\frac{n-1}{n}\right)_K \circledast X + E_1, \\ \left(\frac{n-1}{n}\right)_K \circledast X &\stackrel{d}{=} \left(\frac{n-2}{n-1}\right)_K \circledast \left(\frac{n-1}{n}\right)_K \circledast X + E_2 = \left(\frac{n-2}{n}\right)_K \circledast X + E_2, \\ &\vdots \\ \left(\frac{2}{n}\right)_K \circledast X &\stackrel{d}{=} \left(\frac{1}{2}\right)_K \circledast \left(\frac{2}{n}\right)_K \circledast X + E_{n-1} = \left(\frac{1}{n}\right)_K \circledast X + E_{n-1}, \end{aligned} \tag{7.2.2}$$

where $E_1,\ E_2,\ \ldots,\ E_{n-1}$ have respective pgf's

$$G_{E_1}(s) = \frac{G(s)}{G\left(G_K\left(s;\frac{n-1}{n}\right)\right)},\quad G_{E_2}(s) = \frac{G\left(G_K\left(s;\frac{n-1}{n}\right)\right)}{G\left(G_K\left(s;\frac{n-2}{n}\right)\right)},\quad \cdots,\quad G_{E_{n-1}}(s) = \frac{G\left(G_K\left(s;\frac{2}{n}\right)\right)}{G\left(G_K\left(s;\frac{1}{n}\right)\right)}.$$

Taking the sum for these equalities in (7.2.2) and computing the pgf for the left hand right hand sides, assuming independence, yield

$$\begin{aligned}&G_X(s) \times G_{\left(\frac{n-1}{n}\right)_K \circledast X}(s) \times \cdots \times G_{\left(\frac{2}{n}\right)_K \circledast X}(s)\\ &= G_{\left(\frac{n-1}{n}\right)_K \circledast X}(s) \times \cdots \times G_{\left(\frac{2}{n}\right)_K \circledast X}(s) \times G_{\left(\frac{1}{n}\right)_K \circledast X}(s) \times G_{E_1}(s) \times \cdots \times G_{E_{n-1}}(s).\end{aligned}$$

By substitution, this leads to (7.2.1). This means that each factor (a ratio) in (7.2.1) is a pgf. Note that it also shows a decomposition for X:

$$X \stackrel{d}{=} \left(\frac{1}{n}\right)_K \circledast X + E_{n-1} + \cdots + E_2 + E_1.$$

Hence, for each $n \in \mathcal{N}$, X can be seen as the sum of n independent rv's

$$X_{1,n},\quad X_{2,n},\quad X_{3,n},\quad \ldots,\quad X_{n,n}.$$

Therefore, it forms a triangular array, and the sum of each row has the same GDSD distribution. If we can prove that this triangular array is a "null array", namely

$$0 < 1 - G_{X_{i,n}}(s) < \epsilon, \qquad s_0 < s \leq 1, \quad i = 1, 2, \ldots, n;$$

for all n sufficiently large, then, by Theorem 1 on page 550 in Feller [1966b], X is ID.

By Property 3.5, $G_K(s;\alpha)$ is uniformly continuous in α in $[0,1]$. Hence, $G\left(G_K\left(s;\alpha\right)\right)$ is uniformly continuous in α in $[0,1]$. Meanwhile, the pgf $G\left(G_K\left(s;\alpha\right)\right)$ is also uniformly continuous in s in its range $[0,1]$. Since $G_K(s;0) = 1$ and $G(s;1) = s$, $G_K(s;\alpha)$ is bounded away from 0 for $s_0 < s < 1$ and $0 \leq \alpha \leq 1$. Hence $G(G_K(s;\alpha))$ is bounded away from 0 for $s_0 < s < 1$ and uniformly continuous in α. Therefore, given $\epsilon > 0$, for n sufficiently large, $\forall s_0 < s < 1,\ i = 1, 2, \ldots, n$,

$$\left[G\left(G_K\left(s;\frac{i}{n}\right)\right) - G\left(G_K\left(s;\frac{i+1}{n}\right)\right)\right] \Big/ G\left(G_K\left(s;\frac{i}{n}\right)\right) < \epsilon,$$

and the triangular array is a "null array". This completes the proof.

(2) Assume $\phi(s)$ is the LT of a GSD distribution with positive support. X is a rv from this distribution. Consider the identity:

$$\begin{aligned}
\phi(s) &= \phi\left(-\log\phi_K\left(s;\frac{1}{n}\right)\right) \times \frac{\phi\left(-\log\phi_K\left(s;\frac{2}{n}\right)\right)}{\phi\left(-\log\phi_K\left(s;\frac{1}{n}\right)\right)} \times \cdots \times \frac{\phi\left(-\log\phi_K\left(s;\frac{n-1}{n}\right)\right)}{\phi\left(-\log\phi_K\left(s;\frac{n-2}{n}\right)\right)} \\
&\quad \times \frac{\phi(s)}{\phi\left(-\log\phi_K\left(s;\frac{n-1}{n}\right)\right)} \qquad (7.2.3) \\
&= \frac{\phi\left(-\log\phi_K\left(s;\frac{1}{n}\right)\right)}{\phi\left(-\log\phi_K\left(s;\frac{0}{n}\right)\right)} \times \frac{\phi\left(-\log\phi_K\left(s;\frac{2}{n}\right)\right)}{\phi\left(-\log\phi_K\left(s;\frac{1}{n}\right)\right)} \times \cdots \times \frac{\phi\left(-\log\phi_K\left(s;\frac{n-1}{n}\right)\right)}{\phi\left(-\log\phi_K\left(s;\frac{n-2}{n}\right)\right)} \\
&\quad \times \frac{\phi\left(-\log\phi_K\left(s;\frac{n}{n}\right)\right)}{\phi\left(-\log\phi_K\left(s;\frac{n-1}{n}\right)\right)}, \qquad n = 1,2,3,\ldots.
\end{aligned}$$

Because

$$\frac{\phi\left(-\log\phi_K(s;\beta)\right)}{\phi\left(-\log\phi_K(s;\alpha)\right)} = \frac{\phi\left(-\log\phi_K(s;\beta)\right)}{\phi\left(-\log\phi_K\left(-\log\phi_K(s;\beta);\frac{\alpha}{\beta}\right)\right)}, \qquad \frac{\alpha}{\beta} < 1,$$

and X is GSD, the right hand side is a LT. This indicates that each ratio in (7.2.3) is a LT. Hence, for each $n \in \mathcal{N}$, X can be seen as the sum of n independent rv's

$$X_{1,n}, \quad X_{2,n}, \quad X_{3,n}, \quad \ldots, \quad X_{n,n}.$$

Therefore, it forms a triangular array, and the sum of each row is distributed in the same GSD distribution.

The remainder of the proof is exactly the same as in (1).

Accordingly, we can obtain the following results for GSD and GDSD which correspond to those for SD and DSD.

Theorem 7.2.8 *Consider the positive or non-negative integer-valued generalized convolution.*

(1) If the base distribution of a positive-valued generalized convolution is GSD, then the generalized convolution is GSD.

(2) If the base distribution of a non-negative integer-valued generalized convolution is GDSD, then the generalized convolution is GDSD.

Proof: Applying the fact that the generalized convolution is the limit distribution for the sum of independent rv's from the base distribution. If the base distribution is GSD or GDSD, then the distribution of the sum is GSD or GDSD. This leads to GSD or GDSD for the generalized convolution.

Theorem 7.2.9 *If the distribution of the increment of the innovation process in a continuous-time GAR(1) process is SD or DSD, then the stationary distribution of the continuous-time GAR(1) process is SD or DSD respectively.*

Proof: Apply the same reasoning as the proof of Theorem 7.1.5.

Next, we turn to some concrete examples. Basically, we are focusing on GDSD($\mathbf{I2}(\gamma)$) using Theorem 7.2.5 and GSD($\mathbf{P2}(\gamma)$) with Theorem 7.2.2. This is enough to illustrate the concept of GDSD and GSD for the general cases. More examples in other situations are under study.

Example 7.15 (negative binomial) *The NB distribution can be viewed as a compound Poisson with logarithmic series distribution. The NB(β, q) has pgf*

$$G_X(s) = \exp\left\{\beta\left[\log(1-q)^{-1}\right]\left[\frac{\log(1-qs)}{\log(1-q)} - 1\right]\right\}, \qquad \beta > 0,\ 0 < q < 1.$$

Thus $g(s) = \frac{\log(1-qs)}{\log(1-q)}$ *and* $q_k = \frac{q^k}{-k\log(1-q)}$ $(k \geq 1)$. *Now we want to check if it belongs to GDSD($\mathbf{I2}(\gamma)$). By verifying the conditions in Theorem 7.2.5, we obtain two inequalities:*

$$(1+\gamma) - q \geq 0, \quad (1+\gamma)q - \gamma - q^2 \geq 0.$$

The first one always holds. The second one leads to $\gamma \leq q \leq 1$. *Hence, for any* $q \in (0,1)$, *we can always find a* γ *such that* $\gamma \leq q$. *This leads to that negative binomial belongs to*

$$\mathrm{NB}(\beta, q) \in \bigcap_{\gamma\in[0,q]} \mathrm{GDSD}(\mathbf{I2}(\gamma)) \subset \mathrm{GDSD}(\mathbf{I2}).$$

Hence, any negative binomial is GDSD($\mathbf{I2}$).

Example 7.16 (compound Poisson with geometric) *The geometric distribution has the pgf*

$$g(s) = \frac{1-q}{1-qs}, \qquad 0 \leq q < 1.$$

Thus, $q_k = (1-q)q^k$ $(k \geq 1)$, *and*

$$kq_k - (k+1)q_{k+1} = k(1-q)q^k - (k+1)(1-q)q^{k+1} = (1-q)q^k[k-(k+1)q], \qquad k \geq 1.$$

We want

$$q_1 - 2q_2 \geq -\gamma q_1, \quad kq_k - (k+1)q_{k+1} \geq \gamma[(k-1)q_{k-1} - kq_k], \qquad k = 2, 3, \ldots$$

so that this compound Poisson distribution can be GDSD$(\mathbf{I2}(\gamma))$. *These inequalities lead to*

$$1 + \gamma \geq 2q, \quad (q-\gamma)(1-q)k \geq q^2 - \gamma, \qquad k = 2, 3, \ldots,$$

or in a uniform expression

$$(q-\gamma)(1-q)k \geq q^2 - \gamma, \qquad k = 1, 2, 3, \ldots.$$

If $q < \gamma$, *then* $k \leq \frac{\gamma - q^2}{(\gamma-q)(1-q)}$, *so these inequalities won't hold when* $k > \frac{\gamma-q^2}{(\gamma-q)(1-q)}$. *Hence, it is necessary that* $q > \gamma$. *To guarantee all inequalities hold, it suffices that* $1+\gamma \geq 2q$ *when* $k = 1$, *i.e.,* $\gamma \geq 2q - 1$. *Therefore, in summary,* γ *should satisfy that*

$$\max(0, 2q-1) \leq \gamma \leq q, \qquad 0 \leq q < 1.$$

If $q \leq 1/2$, *then* $\max(0, 2q-1) = 0$, *taking* $\gamma = 0$ *is enough. This will lead to the conclusion that when* $q \leq 1/2$, *the compound Poisson with geometric distribution is GDSD*$(\mathbf{I2}(0))$=*DSD, consistent with our claim in the previous section. Besides, it also belongs to any GDSD*$(\mathbf{I2}(\gamma))$, *where* $0 < \gamma \leq q$.

If $q > 1/2$, $\max(0, 2q-1) = 2q-1$, *then we can always take at least* $\gamma = 2q-1 > 0$. *Under this situation, the compound Poisson with geometric distribution won't belong to DSD, but is GDSD*$(\mathbf{I2}(\gamma))$ *with* $2q-1 < \gamma \leq q$.

In summary, the compound Poisson with geometric distribution always belong to GDSD$(\mathbf{I2})$.

Example 7.17 (discrete stable) *The pgf is*

$$G_X(s) = \exp\{-\lambda(1-s)^\theta\}, \qquad \lambda > 0, \quad 0 < \theta \leq 1.$$

Thus, $G'_X(s)/G_X(s) = \lambda\theta(1-s)^{\theta-1}$. *Now consider the class GDSD*($\mathbf{I2}(\gamma)$). *According to Table 6.1,* $H(s) = (1-\gamma)^{-1}(s-1)(1-\gamma s)$. *Applying (2) of Theorem 7.2.3, we consider*

$$L(s) = 1 + C\cdot(1-\gamma)^{-1}(s-1)(1-\gamma s)\lambda\theta(1-s)^{\theta-1} = 1 - C\cdot\lambda\theta(1-\gamma)^{-1}(1-\gamma s)(1-s)^{\theta}.$$

Taking $C = (\lambda\theta)^{-1}(1-\gamma)$, *we have*

$$\begin{aligned} L(s) &= 1-(1-\gamma s)(1-s)^{\theta} = 1-(1-\gamma s)\Big(1-\sum_{j=1}^{\infty} b_j s^j\Big) \\ &= \gamma s + \sum_{j=1}^{\infty} b_j s^j - \gamma\sum_{j=2}^{\infty} b_{j-1}s^j = (\gamma+b_1)s + \sum_{j=2}^{\infty}(b_j - \gamma b_{j-1})s^j, \end{aligned}$$

where $b_1 = \theta$ *and* $b_j = \theta\prod_{l=1}^{j-1}(l-\theta)/j!$ *for* $j \geq 2$. *Since* $b_j/b_{j-1} = (j-1-\theta)/j = 1-(1+\theta)/j$ *is increasing in* j, $L(s)$ *is a pgf if* $\gamma \leq b_2/b_1 = 1-(1+\theta)/2 = (1-\theta)/2$. *Hence, when* $\gamma < 1/2$ *and* $\theta \leq 1-2\gamma$, X *is in GDSD*($\mathbf{I2}(\gamma)$).

Example 7.18 (Counterexample of the converse of Theorem 7.2.2) *The continuous analogue of discrete stable (see Example 7.17) is positive stable with LT*

$$\phi_Y(s) = \exp\{-\lambda s^{\delta}\}, \qquad \lambda > 0 \quad 0 < \delta < 1.$$

the positive stable LT. $\phi'_Y(s)/\phi_Y(s) = -\delta s^{\delta-1}$. *For* **P2**, *according to Table 6.2, we have*

$$H(s)\phi'_Y(s)/\phi_Y(s) = -\delta(1-\gamma+\gamma s)s^{\delta}/(1-\gamma).$$

Let $\chi(s) = s^{\delta}(1-\gamma+\gamma s) = (1-\gamma)s^{\delta} + \gamma s^{\delta+1}$. *Then,* χ *is not completely monotone if* $0 < \gamma < 1$ *since*

$$\chi'(s) = (1-\gamma)\delta s^{\delta-1} + \gamma(\delta+1)s^{\delta} > 0,$$

$$\chi''(s) = (1-\gamma)\delta(\delta-1)s^{\delta-2} + \gamma(\delta+1)\delta s^{\delta-1},$$

and $\chi''(s) > 0$ *for* s *sufficiently large. By Theorem 2.2.6,* $\exp\{-C\cdot\chi(s)\}$ *can not be a LT. Thus, according to (1) of Theorem 7.2.3 and (1) of Theorem 6.3.1, the positive stable is not GSD(P2(*γ*)) for any* $0 < \gamma < 1$.

This is an example where the continuous distribution is not GSD but the discrete analogue is GDSD [when $\gamma \leq (1-\delta)/2$*].*

7.3 Relationships among the classes of generalized self-decomposable and discrete self-decomposable distributions

Relationships among the GDSD and among the GSD classes are of interest. One may wonder if one covers another, or if they are overlapping or disjoint. The difference between two classes will certainly appear in modelling, namely leading to different model classes. In general, it's hard to answer such question. In this section, we shall particularly investigate the relationship among the class of GDSD(**I2**), as well as among the class of GSD(**P2**). From this special study, we may have some impression on and partially answer this issue.

First, we look into GDSD(**I2**). As a special case, it includes DSD=GDSD(**I2**(0)). But in general, it consists of all GDSD(**I2**(γ)), where $0 \leq \gamma < 1$. A natural question is what's the relationship between DSD and GDSD(**I2**(γ)), for a fixed γ. Although Theorem 7.1.3 and Theorem 7.2.5 offer conditions for DSD and GDSD(**I2**(γ)), it won't help us because we can't deduce one from the other. Hence, we have to study some special members to investigate the relationship between DSD=GDSD(**I2**(0)) and GDSD(**I2**(γ)), where $0 < \gamma < 1$.

(1) Do all distributions in DSD belong to GDSD(**I2**(γ))?

Let's consider the Poisson and Neyman's Type A distribution. The latter is the compound Poisson with another Poisson.

The Poisson has pgf $G_X(s) = e^{\lambda(s-1)}$ $(\lambda > 0)$. In terms of the pgf form of the compound Poisson, $q_1 = 1,\ q_k = 0,\ k \geq 2$. Thus, $q_1 - 2q_2 = q_1 > 0$ and $2q_2 - 3q_3 = 0$. For any $\gamma > 0$, it never holds that $2q_2 - 3q_3 \geq \gamma(q_1 - 2q_2)$. By Theorem 7.2.5, the Poisson distribution doesn't belong to GDSD(**I2**(γ)) for $0 < \gamma < 1$.

The Neyman's Type A distribution has pgf $G_X(s) = \exp\left\{\lambda\left(e^{\eta(s-1)}\right)\right\}$ $(\lambda, \eta > 0)$, that is $q_k = \frac{\eta^k}{k!}e^{-\eta}$ $(k \geq 0)$. Hence, $q_{k+1} = \frac{\eta}{k+1}q_k$ for $k \geq 1$. We check if for any $0 < \gamma < 1$,

$$q_1 - 2q_2 \geq -\gamma q_1, \quad kq_k - (k+1)q_{k+1} \geq \gamma[(k-1)q_{k-1} - kq_k], \qquad k = 2, 3, \ldots.$$

These leads to

$$1 \geq \eta - \gamma, \quad \gamma k^2 - (\gamma + \gamma\eta + \eta)k + \eta^2 \leq 0, \qquad k = 2, 3, \ldots.$$

For the latter inequalities, when $k \to \infty$, it is impossible to hold because the left hand side will go to infinity. This means that the Neyman's Type A distribution is not GDSD($\mathbf{I2}(\gamma)$) for $0 < \gamma < 1$. However, it is DSD when $\eta \leq 1$. See (2) in Example 7.7.

There are other examples like these two. This implies that DSD is not covered by GDSD($\mathbf{I2}(\gamma)$).

(2) Do all distributions in GDSD($\mathbf{I2}(\gamma)$) belong to DSD?

Consider the compound Poisson with geometric distribution, where the geometric distribution has the pgf $g(s) = \frac{1-q}{1-qs}$ $(0 \leq q < 1)$. Now choose q such that

$$\begin{cases} \frac{1}{2} < q \leq \frac{1+\gamma}{2}, & \text{if } 0 < \gamma \leq \frac{1}{2}, \\ \gamma \leq q \leq \frac{1+\gamma}{2}, & \text{if } \frac{1}{2} < \gamma < 1. \end{cases}$$

By Example 7.16, it is GDSD($\mathbf{I2}(\gamma)$). However, it is not DSD. See (1) in Example 7.7.

This implies that GDSD($\mathbf{I2}(\gamma)$) is not covered by DSD.

(3) Are there distributions common to both of DSD and GDSD($\mathbf{I2}(\gamma)$)?

We consider NB(β, q) with the pgf $G_X(s) = \left(\frac{1-q}{1-qs}\right)^{\beta}$ $(\beta > 0,\ 0 < q < 1)$. It is known that the negative binomial is DSD. Suppose $q \geq \gamma$. Then, by Example 7.15, it is GDSD($\mathbf{I2}(\gamma)$) too.

This shows that DSD and GDSD($\mathbf{I2}(\gamma)$) have common members for each $0 < \gamma < 1$.

Therefore, based on these analyses, the relationship between DSD and GDSD($\mathbf{I2}(\gamma)$) is overlapping. Recall that these distribution members will be used as the marginal distributions for a steady state continuous-time GAR(1) process. Thus, the obtained fact means that for the common distribution members, we can construct more than one kind of continuous-time GAR(1) processes; however, for the non-common distribution members, we only can construct one continuous-time GAR(1) process.

For $\gamma_1 < \gamma_2$, by the example of the compound Poisson with geometric distribution, we know that GDSD($\mathbf{I2}(\gamma_1)$) and GDSD($\mathbf{I2}(\gamma_2)$) have common distribution members. In addition, the conditions in Theorem 7.2.5 do not support that one is a subset of the other. Hence, their relationship is overlapping.

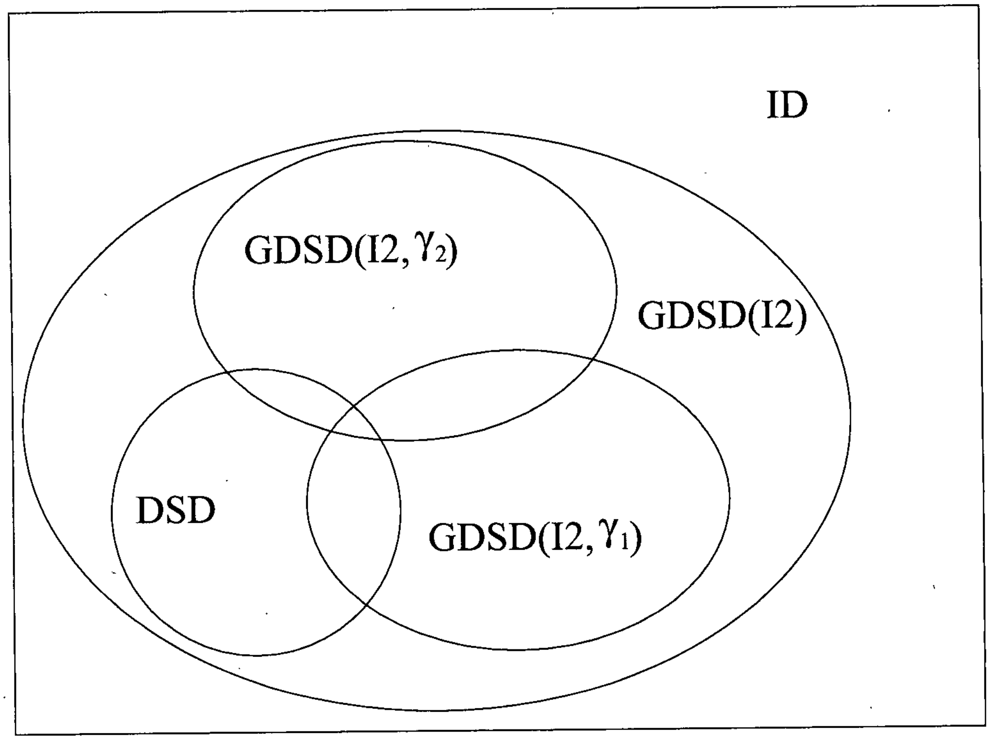


Figure 7.1: *The relationship of ID, DSD and GDSD*(**I2**).

Figure 7.1 shows the relationship of ID, DSD and GDSD(**I2**).

For a given $0 < \gamma_0 < 1$, it is interesting that we can always construct a distribution which only belongs to GDSD($\mathbf{I2}(\gamma_0)$). This fact is established by rearranging the first two probability masses in the compound Poisson with geometric distribution in Example 7.16. We describe it in the proof of the theorem below.

Theorem 7.3.1 *For any* $0 < \gamma_0 < 1$, *there exists a distribution which belongs to GDSD*($\mathbf{I2}(\gamma_0)$) *but not GDSD*($\mathbf{I2}(\gamma)$), $\gamma \neq \gamma_0$.

Proof: We shall apply a construction method. Example 7.16 shows that for any $\gamma_0 \in [0, 1)$, the compound Poisson with Geometric(γ_0) is GDSD($\mathbf{I2}(\gamma)$) for $\gamma \in [\max(0, 2\gamma_0 - 1), \gamma_0]$. The lower bound $\max(0, 2\gamma_0 - 1)$ comes from $(1+\gamma)q_1 - 2q_2 \geq 0$. Now decrease q_1 to $q_1' = \frac{2q_2}{1+\gamma_0}$ and increase q_0 to $q_0' = q_0 + (q_1 - q_1')$. Then, $q_1' = \frac{2q_2}{1+\gamma_0} \leq q_1$ and $q_0' = q_0 + q_1 - \frac{2q_2}{1+\gamma_0} \geq q_0$. The remaining q_i $(i \geq 2)$

are the same. We shall see when this new compound Poisson distribution is GDSD$(\mathbf{I2}(\gamma))$ for some γ. This is equivalent to checking

$$(1+\gamma)q_1' - 2q_2 \geq 0, \quad 2q_2 - 3q_3 \geq \gamma[q_1' - 2q_2], \quad kq_k - (k+1)q_{k+1} \geq \gamma[(k-1)q_{k-1} - kq_k], \quad k = 3, 4, \ldots$$

The first inequality will hold iff $\gamma \geq \gamma_0$, because $(1+\gamma)q_1' - 2q_2 = 2q_2\left(\frac{1+\gamma}{1+\gamma_0} - 1\right) = 2q_2 \cdot \frac{\gamma-\gamma_0}{1+\gamma}$. The second one still holds when $\max(0, 2\gamma_0 - 1) \leq \gamma \leq \gamma_0$ because

$$2q_2 - 3q_3 \geq \gamma[q_1 - 2q_2] \geq \gamma[q_1' - 2q_2].$$

The remaining inequalities hold if $\max(0, 2\gamma_0 - 1) \leq \gamma \leq \gamma_0$. Thus, the only choice of γ is γ_0. This means that the new compound Poisson is GDSD$(\mathbf{I2}(\gamma_0))$ for a fixed γ_0 only.

This fact also partially emphasizes that any two classes GDSD$(\mathbf{I2}(\gamma_1))$ and GDSD$(\mathbf{I2}(\gamma_2))$; $\gamma_1 \neq \gamma_2$, are not subsets of each other.

As to the relationship between the GDSD$(\mathbf{I2})$ class and other classes, it could be overlapping, subset inclusion or disjoint.

By Theorem 7.2.2, we can obtain the relationship among all GSD$(\mathbf{P2})$ classes (including the special case SD=GSD$(\mathbf{P2}(0))$). Figure 7.2 shows the relationship of ID, SD and GSD$(\mathbf{P2})$. Like the discrete situation, its relationship with other classes could be overlapping, subset inclusion or disjoint too.

Now we turn to other GDSD classes: GDSD$(\mathbf{I3})$, GDSD$(\mathbf{I4})$ and GDSD$(\mathbf{I5})$.

First, by (1) of Theorem 7.2.1, we find that GDSD$(\mathbf{I3})$ is empty. This is simply because that the self-generalized rv family from **I3** does not satisfy the necessary condition:

$$G_K(s;\alpha) = 1 - (1-s)^\alpha \geq s \quad \text{iff} \quad 1 - s \geq (1-s)^\alpha \quad \text{iff} \quad \alpha \geq 1.$$

So $\{G_K(s;\alpha) : 0 < \alpha \leq 1\}$ does not lead to any distribution in GDSD$(\mathbf{I3})$.

For K from **I4** with pgf

$$G_K(s;\alpha) = c^{-1}[1 - e^{-\theta(1-\alpha)}(1-cs)^\alpha], \qquad c = 1 - e^{-\theta},\ \theta > 0,$$

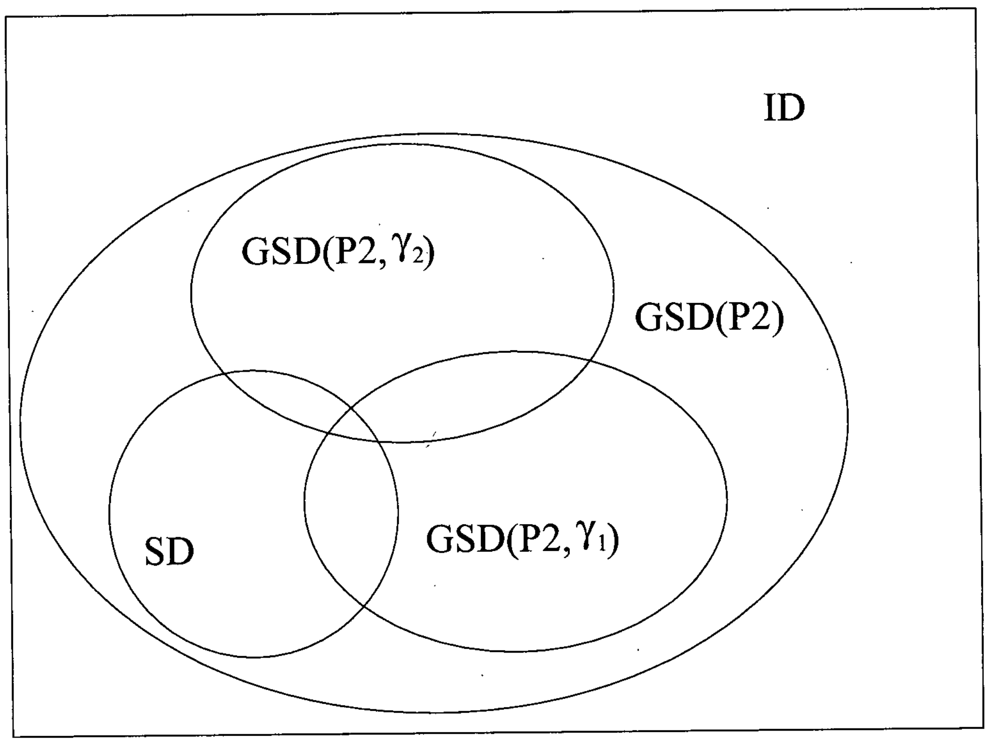


Figure 7.2: *The relationship of ID, SD and GSD*(**P2**).

according to Table 6.1, we have $H(s) = -c^{-1}(1 - cs)[\theta + \log(1 - cs)]$. By (2) of Theorem 7.2.3, $G_X \in \text{GDSD}(\mathbf{I4}(c))$ if

$$L(s) = 1 - C \cdot c^{-1}(1 - cs)[\theta + \log(1 - cs)] \cdot \frac{G'_X(s)}{G_X(s)}$$

is a pgf for some $C > 0$. The following example shows that $\text{GDSD}(\mathbf{I4}(c))$ is not empty.

Example 7.19 (negative binomial) *NB*(η, q) *has pgf* $G_X(s) = [p/(1 - qs)]^\eta$, *where* $p = 1 - q$. *Thus,* $G'_X(s)/G_X(s) = \eta p/(1 - qs)$. *Check for conditions on* c *so that*

$$L(s) = 1 - C \cdot c^{-1}(1 - cs)[\theta + \log(1 - cs)]\frac{\eta p}{1 - qs}$$

is a pgf. Choosing $C = c(\eta p)^{-1}$ *leads to*

$$L(s) = 1 - (1 - cs)[1 + \theta^{-1}\log(1 - cs)]/(1 - qs).$$

When $q = c$, $L(s) = -\theta^{-1}\log(1 - cs)$ *which is the pgf of the logarithmic series distribution. Hence* *NB*(η, q) *is GDSD*$(\mathbf{I4}(c))$ *if* $q = c$ *or* $p = 1 - c$.

For K from **I5** with pgf

$$G_K(s; \alpha) = 1 - \alpha^\theta(1 - \gamma)^\theta[(1 - \alpha)\gamma + (1 - \gamma)(1 - s)^{-1/\theta}]^{-\theta}, \qquad \theta \geq 1,\ 0 < \gamma < 1,$$

by Table 6.1, we have $H(s) = -\theta(1 - s)(1 - \gamma)^{-1}[1 - \gamma + \gamma(1 - s)^{1/\theta}]$. From (2) of Theorem 7.2.3, $G_X \in \text{GDSD}(\mathbf{I5}(\gamma, \theta))$ if

$$L(s) = 1 - C \cdot \theta(1 - \gamma)^{-1}(1 - s)[1 - \gamma + \gamma(1 - s)^{1/\theta}] \cdot \frac{G'_X(s)}{G_X(s)}$$

is a pgf for some $C > 0$. The following example shows that for every parameter vector (γ, θ), $\text{GDSD}(\mathbf{I5}(\gamma, \theta))$ is not empty.

Example 7.20 (discrete stable) *It has pgf* $G_X(s) = \exp\{-\lambda(1-s)^\delta\}$, *where* $\lambda > 0$ *and* $0 < \delta \leq 1$. *Hence,* $G'_X(s)/G_X(s) = \lambda\delta(1 - s)^{\delta - 1}$. *We check for conditions on* (γ, θ) *so that*

$$L(s) = 1 + C \cdot \theta(1 - \gamma)^{-1}(s - 1)[1 - \gamma + \gamma(1 - s)^{1/\theta}]\lambda\delta(1 - s)^{\delta - 1}$$

is a pgf. Take $C = (1-\gamma)(\theta\lambda\delta)^{-1}$ *for simplicity. Then*

$$L(s) = 1-(1-s)^{\delta}[1-\gamma+\gamma(1-s)^{1/\theta}] = (1-\gamma)[1-(1-s)^{\delta}] + \gamma[1-(1-s)^{\delta+1/\theta}].$$

Recalling the pgf form of a power series distribution, for any $0<\gamma<1$, *we know that* $L(s)$ *is a pgf if* $\delta+\theta^{-1}<1$ *or if* $\delta < 1-\theta^{-1}$ *for* $\theta>1$.

Hence, this shows that GDSD($\mathbf{I5}(\gamma,\theta)$) *is non-empty for every* $\theta>1$ *and* $0<\gamma<1$.

Similarly, for K from **P3**, the necessary condition in (2) of Theorem 7.2.1 is not satisfied, because

$$\phi_K(s;\alpha) = \exp\{-s^{\alpha}\} < e^{-s}, \qquad \text{if } 0<s<1.$$

Hence, GSD(**P3**) is empty.

As to K from **P4** and **P5**, we give a brief discussion. X $\in$GSD(**P4**) or GSD(**P5**) requires that

$$\exp\left\{C\cdot\left(s+\frac{1}{e^{\theta}-1}\right)\log\left[1+(e^{\theta}-1)s\right]\frac{\phi_X'(s)}{\phi_X(s)}\right\} \quad \text{or} \quad \exp\left\{C\cdot\theta s\left(1+\frac{\gamma}{1-\gamma}s^{1/\theta}\right)\frac{\phi_X'(s)}{\phi_X(s)}\right\}$$

must be LT for all $C>0$. The following two examples show that GSD(**P4**) and GSD(**P5**) are not empty.

Example 7.21 (Gamma) *The LT of Gamma*(δ,β) *is* $\phi_X(s) = \left(\frac{\beta}{\beta+s}\right)^{\delta}$. *Thus*

$$\frac{\phi_X'(s)}{\phi_X(s)} = \frac{-\delta\beta^{\delta}(\beta+s)^{-d-1}}{\beta^{d}(\beta+s)^{-d}} = -\frac{\delta}{\beta+s},$$

and

$$\exp\left\{C\cdot\left(s+\frac{1}{e^{\theta}-1}\right)\log\left[1+(e^{\theta}-1)s\right]\frac{\phi_X'(s)}{\phi_X(s)}\right\} = \exp\left\{-\frac{C\delta}{\beta+s}\cdot\left(s+\frac{1}{e^{\theta}-1}\right)\log\left[1+(e^{\theta}-1)s\right]\right\}.$$

For simplicity, we take $\beta = 1/(e^{\theta}-1)$ *so that*

$$\exp\left\{C\cdot\left(s+\frac{1}{e^{\theta}-1}\right)\log\left[1+(e^{\theta}-1)s\right]\right\} = \exp\left\{-C\delta\log\left[1+(e^{\theta}-1)s\right]\right\} = \left[\frac{(e^{\theta}-1)^{-1}}{(e^{\theta}-1)^{-1}+s}\right]^{C\delta},$$

which the LT of Gamma($C\delta,(e^{\theta}-1)^{-1}$). *This means that Gamma*($\delta,(e^{\theta}-1)^{-1}$) $\in$ *GSD*($\mathbf{P4}(\theta)$) *and GSD*($\mathbf{P4}(\theta)$) *is not empty for any* $\theta>0$.

Example 7.22 (Positive stable) *The positive stable distribution has LT of form* $\phi_X(s) = \exp\left\{-\lambda s^{\delta}\right\}$. *Thus,* $\frac{\phi'_X(s)}{\phi_X(s)} = -\lambda\delta s^{\delta-1}$, *and*

$$\begin{aligned}\exp\left\{C \cdot \theta s \left(1 + \frac{\gamma}{1-\gamma} s^{1/\theta}\right) \frac{\phi'_X(s)}{\phi_X(s)}\right\} &= \exp\left\{-C\lambda\delta\theta \cdot s^{\delta}\left(1 + \frac{\gamma}{1-\gamma} s^{1/\theta}\right)\right\} \\ &= \exp\left\{-C\lambda\delta\theta \cdot s^{\delta}\right\} \exp\left\{-\frac{C\lambda\delta\theta\gamma}{1-\gamma} s^{\delta+1/\theta}\right\}.\end{aligned}$$

If $\delta + 1/\theta < 1$, *then both terms on the right hand side of the above are LT's of positive stable distributions. For any* $0 < \delta < 1$, *we can always find a* $\theta > 1$ *such that* $\delta + 1/\theta < 1$, *and vice versa. Therefore, any positive stable distribution belongs to GSD*(**P5**)*; and GSD*(**P5**(θ, γ)) *is not empty for any* $\theta > 1$ *and* $0 < \gamma < 1$.

The empty property of GDSD(**I3**) and GSD(**P3**) explains why we failed to find stationary continuous-time generalized AR(1) processes resulting from extended-thinning operations by **I3** and **P3** in Chapter 6.

Chapter 8

Transition and sojourn time

The margins of continuous-time GAR(1) processes are basically divided into two types: continuous and discrete. The support of the process is the state space of a process. Given the previous observation, the conditional distribution of the current state is continuous or discrete depending on if its margins are continuous or discrete. For the discrete margins, the conditional probability for a non-negative integer is usually positive. This means the process can stay in this state for some time, then jumps to another state. However, for the continuous margins, the conditional probability at one point is usually zero although its conditional density is not zero. Hence, we are unlikely to observe a continuous-time GAR(1) process which can stay on a point over a time period.

In this chapter, we take a close look from the viewpoint of path or trajectory of the continuous-time GAR(1) processes. It motivates the transition study which describes the instantaneous change of a continuous-time Markov process. We shall investigate the self-generalized distribution involved in a continuous-time GAR(1) process via its transition property. Because for real margins, the only known self-generalized distribution is the degenerate distribution on a point, we consider the non-negative integer and positive margins only. Section 8.1 shows the feature of the change of the conditional pgf and LT. Specifically, in Section 8.1.1, we study the infinitesimal transition matrix, and compare it with relevant corresponding processes in other probabilistic areas like queuing theory, while in Section 8.1.2 we give the instantaneous change rate and relative change rate of the conditional LT. In Section 8.2, we shall apply the generating function method to discuss the corresponding partial differential equations, and make comparisons with some results in the literature. In Section 8.3, we study the distribution of the sojourn time for the continuous-time GAR(1) process with non-negative integer margins. This will be useful for situations of continuous observations like queuing.

8.1 Infinitesimal transition analysis

The infinitesimal transition approximates the probability change over a very small time period for a continuous-time process. It is commonly used in birth-death processes and in survival analysis. We shall probe the feature of transitions for the continuous-time GAR(1) processes. They are discussed separately by the type of their margins: non-negative integer or positive.

8.1.1 Non-negative integer margin

The state space of a continuous-time GAR(1) process with non-negative integer margins is $S = \{0, 1, 2, \ldots\} = \mathcal{N}_0$. Suppose the time difference h is very small. Given $X(t) = i$, the conditional probabilities of $X(t+h)$, $\Pr[X(t+h) = j \mid X(t) = i]$ $(j \in \mathcal{N}_0)$, are called the **infinitesimal transition probabilities**. This further leads to the **infinitesimal generator** $Q = (q_{i,j})$ $(i, j \in \mathcal{N}_0)$ with

$$q_{i,j} = \begin{cases} \lim\limits_{h \to 0} \frac{\Pr[X(t+h)=j|X(t)=i]}{h}, & j \neq i, \\ \lim\limits_{h \to 0} \frac{\Pr[X(t+h)=i|X(t)=i]-1}{h}, & j = i. \end{cases} \tag{8.1.1}$$

Note that $q_{i,j} \geq 0$ $(i \neq j)$ and $q_{i,i} = -\sum_{j \neq i} q_{i,j} \leq 0$. When $q_{i,j} = 0$ for $|i - j| > 1$, the process is a birth-death process with birth rates $\{q_{i,i+1};\ i = 0, 1, 2, \ldots\}$ and death rates $\{q_{i,i-1};\ i = 1, 2, \ldots\}$. This infinitesimal generator matrix remains the same for a Markov process whether it is under steady state or not. Hence, we can assume the process is under steady state when we study the infinitesimal transition. Next we will investigate the infinitesimal generator matrix of the continuous-time GAR(1) process with non-negative integer margins.

To warm up, let's first study the continuous-time GAR(1) process stipulated by the binomial-thinning operation, i.e., K from **I1**:

$$X(t+h) \stackrel{d}{=} e^{-\mu h} * X(t) + \int_0^h e^{-\mu t} * d\epsilon(t) \stackrel{\text{def}}{=} e^{-\mu h} * X(t) + E(t; t+h).$$

When h is small enough, it will become a stochastic difference equation:

$$X(t+h) \stackrel{d}{=} (1-\mu h) * X(t) + \Delta\epsilon(h), \qquad \Delta\epsilon(h) = \epsilon(t+h) - \epsilon(t).$$

Our aim is to find the infinitesimal transition probabilities with an expansion in terms of h. Assume m is a non-negative integer. $I_0 = 0$, $I_k \stackrel{i.i.d.}{\sim} \text{Bernoulli}(1-\mu h)$ $(k \geq 1)$. By algebra, we have that for $m \geq 0$

$$\begin{aligned}
\Pr[X(t+h) = m \mid X(t) = 0] &= \Pr\left[(1-\mu h) * X(t) + \Delta\epsilon(h) = m \mid X(t) = 0\right] \\
= \Pr\left[I_0 + \Delta\epsilon(h) = m\right] &= \Pr\left[\Delta\epsilon(h) = m\right], \\
\Pr[X(t+h) = 1+m \mid X(t) = 1] &= \Pr\left[(1-\mu h) * X(t) + \Delta\epsilon(h) = 1+m \mid X(t) = 1\right] \\
= \Pr\left[I_1 + \Delta\epsilon(h) = m\right] &= \Pr\left[I_1 = 1, \Delta\epsilon(h) = m\right] + \Pr\left[I_1 = 0, \Delta\epsilon(h) = m+1\right] \\
&= (1-\mu h)\Pr[\Delta\epsilon(h) = m] + (\mu h)\Pr[\Delta\epsilon(h) = m+1] \\
&= \Pr[\Delta\epsilon(h) = m] + \mu h\left(\Pr[\Delta\epsilon(h) = m+1] - \Pr[\Delta\epsilon(h) = m]\right), \\
\Pr[X(t+h) = i+m \mid X(t) = i] &= \Pr\left[(1-\mu h) * X(t) + \Delta\epsilon(h) = i+m \mid X(t) = i\right] \\
= \Pr\left[\sum_{k=0}^{i} I_k + \Delta\epsilon(h) = i+m\right] &= \sum_{l=0}^{i} \Pr\left[\sum_{k=0}^{i} I_k = l\right] \cdot \Pr[\Delta\epsilon(h) = (i+m) - l] \\
&= \sum_{l=0}^{i} \binom{i}{l} [1-\mu h]^l (\mu h)^{i-l} \Pr[\Delta\epsilon(h) = (i+m) - l] \\
&= (1-\mu h)^i \Pr[\Delta\epsilon(h) = m] + i(1-\mu h)^{i-1}(\mu h)\Pr[\Delta\epsilon(h) = m+1] \\
&\quad + \sum_{l=0}^{i-2} \binom{i}{l} (1-\mu h)^l (\mu h)^{i-l} \Pr[\Delta\epsilon(h) = (i+m) - l] \\
&= [1 - i\mu h + o(h)]\Pr[\Delta\epsilon(h) = m] + i[1 - (i-1)\mu h + o(h)](\mu h)\Pr[\Delta\epsilon(h) = m+1] \\
&\quad + (\mu h)^2 \sum_{l=0}^{i-2} \binom{i}{l} [1 - l\mu h + o(h)](\mu h)^{i-l-2} \Pr[\Delta\epsilon(h) = (i+m) - l]
\end{aligned}$$

$$= \Pr[\Delta\epsilon(h) = m] + i\mu h \left(\Pr[\Delta\epsilon(h) = m+1] - \Pr[\Delta\epsilon(h) = m]\right) + o(h), \qquad i \geq 2,$$

and for $0 < m \leq i$ $(i \geq 1)$

$$\begin{aligned}
\Pr[X(t+h) = i - i \mid X(t) = i] &= \Pr\left[(1-\mu h) * X(t) + \Delta\epsilon(h) = 0 \mid X(t) = i\right] \\
= \Pr\left[\sum_{k=0}^{i} I_k + \Delta\epsilon(h) = 0\right] &= \Pr\left[\sum_{k=0}^{i} I_k = 0, \Delta\epsilon(h) = 0\right] \\
= (\mu h)^i \Pr[\Delta\epsilon(h) = 0], & \\
\Pr[X(t+h) = i - (i-1) \mid X(t) = i] &= \Pr\left[(1-\mu h) * X(t) + \Delta\epsilon(h) = 1 \mid X(t) = i\right] \\
= \Pr\left[\sum_{k=0}^{i} I_k + \Delta\epsilon(h) = 1\right] &= \Pr\left[\sum_{k=0}^{i} I_k = 1, \Delta\epsilon(h) = 0\right] + \Pr\left[\sum_{k=0}^{i} I_k = 0, \Delta\epsilon(h) = 1\right] \\
= i(1-\mu h)(\mu h)^{i-1} \Pr[\Delta\epsilon(h) = 0] + (\mu h)^i \Pr[\Delta\epsilon(h) = 1], & \\
\Pr[X(t+h) = i - m \mid X(t) = i] &= \Pr\left[(1-\mu h) * X(t) + \Delta\epsilon(h) = i - m \mid X(t) = i\right] \\
= \Pr\left[\sum_{k=0}^{i} I_k + \Delta\epsilon(h) = i - m\right] &= \sum_{l=0}^{i-m} \Pr\left[\sum_{k=0}^{i} I_k = l\right] \cdot \Pr[\Delta\epsilon(h) = (i-m) - l] \\
= \sum_{l=0}^{i-m} \binom{i}{l} [1-\mu h]^l (\mu h)^{i-l} \Pr[\Delta\epsilon(h) = (i-m) - l] & \\
= \binom{i}{i-m} (1-\mu h)^{i-m} (\mu h)^m \Pr[\Delta\epsilon(h) = 0] & \\
\quad + \binom{i}{i-m-1} (1-\mu h)^{i-m-1} (\mu h)^{m+1} \Pr[\Delta\epsilon(h) = 1] & \\
\quad + (\mu h)^2 \sum_{l=0}^{i-m-2} \binom{i}{l} (1-\mu h)^l (\mu h)^{i-l-2} \Pr[\Delta\epsilon(h) = (i-m) - l], & \qquad i - m \geq 2.
\end{aligned}$$

In summary, the infinitesimal transition probabilities are

$$\Pr[X(t+h) = i + k \mid X(t) = i] \tag{8.1.2}$$

$$= \begin{cases} \Pr[\Delta\epsilon(h) = k] + i\mu h \left(\Pr[\Delta\epsilon(h) = k+1] - \Pr[\Delta\epsilon(h) = k]\right) + o(h), & k \geq 1, \\ \Pr[\Delta\epsilon(h) = 0] + i\mu h \left(\Pr[\Delta\epsilon(h) = 1] - \Pr[\Delta\epsilon(h) = 0]\right) + o(h), & k = 0, \\ i\mu h \Pr[\Delta\epsilon(h) = 0] + o(h), & k = -1, \\ o(h), & k < -1. \end{cases}$$

From the infinitesimal transition probabilities, we can deduce that $q_{i,j} = 0$ for all $j \leq i - 2$, i.e., the lower diagonal entries are zeros in the infinitesimal generator matrix. The other entries

$q_{i,j}$ $(j \geq i-1)$ will be further determined by $\Pr[\Delta\epsilon(h) = m]$, where $m = 0, 1, 2, \ldots$. Under steady state, by Theorem 6.3.1, the pgf of $\Delta\epsilon(h)$ is of form $\exp\left\{\frac{G'_X(s)}{G_X(s)}(s-1)\mu h\right\}$. This form will help us to obtain the expansions of the pmf of $\Delta\epsilon(h)$ in terms of h. The following consists of two examples.

Example 8.1 *Suppose the marginal distribution is Poisson(γ) with pgf $G_X(s) = e^{\gamma(s-1)}$, where $\gamma = \lambda/\mu$. Then*

$$\exp\left\{\frac{G'_X(s)}{G_X(s)}(s-1)\mu h\right\} = \exp\left\{\gamma\mu h(s-1)\right\} = e^{\lambda h(s-1)}.$$

This leads to

$$\Pr[\Delta\epsilon(h) = m] = \frac{(\lambda h)^m}{m!}e^{-\lambda h} = \begin{cases} 1 - \lambda h + o(h), & m = 0, \\ \lambda h + o(h), & m = 1, \\ o(h), & m > 1. \end{cases}$$

Therefore, by algebra, we obtain

$$\Pr[X(t+h) = i + k \mid X(t) = i] = \begin{cases} o(h), & k > 1, \\ \lambda h + o(h), & k = 1, \\ 1 - (\lambda + i\mu)h + o(h), & k = 0, \\ i\mu h + o(h), & k = -1, \\ o(h), & k < -1, \end{cases}$$

and the infinitesimal generator is

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & 0 & \cdots \\ 0 & 0 & 3\mu & -(\lambda+3\mu) & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

From the infinitesimal generator, we know it is a birth-death process with death-rates $q_{i,i-1} = i\mu$ $(i = 1, 2, 3, \ldots)$ and birth-rates $q_{i,i+1} = \lambda$ $(i = 0, 1, 2, \ldots)$. This is one special case among the linear birth-death processes, in which the birth-rates and death-rates are

$$q_{i,i+1} = a + ib, \quad q_{i,i-1} = ic, \qquad a,\ b,\ c \geq 0.$$

See Anderson [1991], Section 3.2, p. 103.

Example 8.2 *Consider the marginal distribution is NB(γ, q) with pgf* $G_X(s) = \left(\frac{1-q}{1-qs}\right)^\gamma$, *where* $0 < q < 1$ *and* $\gamma > 0$. *Then*

$$\begin{aligned}
\exp\left\{\frac{G'_X(s)}{G_X(s)}(s-1)\mu h\right\} &= \exp\left\{\gamma\mu\frac{q(s-1)}{1-qs}h\right\} = 1 + \gamma\mu\frac{q(s-1)}{1-qs}h + o(h) \\
&= 1 + \gamma\mu q(s-1)\left[1 + qs + q^2s^2 + q^3s^3 + \cdots\right] h + o(h) \\
&= 1 - \mu\gamma q h + \left[\mu\gamma(1-q)\sum_{m=1}^{\infty} q^m s^m\right] h + o(h),
\end{aligned}$$

which indicates that

$$\Pr[\Delta\epsilon(h) = 0] = 1 - \mu\gamma q h + o(h), \quad \Pr[\Delta\epsilon(h) = m] = \mu\gamma(1-q)q^m h + o(h), \quad m = 1, 2, 3, \ldots.$$

These lead to the infinitesimal transition probabilities

$$\Pr[X(t+h) = i + k \mid X(t) = i] = \begin{cases} \mu\gamma(1-q)q^k h + o(h), & k \geq 1, \\ 1 - \mu\gamma q h - i\mu h + o(h), & k = 0, \\ i\mu h + o(h), & k = -1, \\ o(h), & k < -1, \end{cases}$$

and the infinitesimal generator

$$Q = \begin{pmatrix} -\mu\gamma q & \mu\gamma(1-q)q & \mu\gamma(1-q)q^2 & \mu\gamma(1-q)q^3 & \mu\gamma(1-q)q^4 & \cdots \\ \mu & -\mu(\gamma q + 1) & \mu\gamma(1-q)q & \mu\gamma(1-q)q^2 & \mu\gamma(1-q)q^3 & \cdots \\ 0 & 2\mu & -\mu(\gamma q + 2) & \mu\gamma(1-q)q & \mu\gamma(1-q)q^2 & \cdots \\ 0 & 0 & 3\mu & -\mu(\gamma q + 3) & \mu\gamma(1-q)q & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

This is not a birth-death process.

Next we consider the more general situation where K is from any non-negative integer self-generalized distribution. Correspondingly, when the time difference h is small enough, the stochastic difference equation is

$$X(t+h) \stackrel{d}{=} (1-\mu h)_K \circledast X(t) + \Delta\epsilon(h), \qquad \Delta\epsilon(h) = \epsilon(t+h) - \epsilon(t).$$

Assume under steady state, the margins have the pgf $G_X(s)$. Then $\Delta\epsilon(h)$ has pgf of form

$$G_{\Delta\epsilon(h)}(s) = \exp\left\{\frac{G'_X(s)}{G_X(s)} H(s)\mu h\right\}, \qquad \text{where } H(s) = \left.\frac{\partial G_K(s,\alpha)}{\partial\alpha}\right|_{\alpha=1}.$$

Given $X(t) = i$, the conditional pgf of $X(t+h)$ is

$$\begin{aligned}
G_{X(t+h)|X(t)=i}(s) &= (G_K(s;1-\mu h))^i\, G_{\Delta\epsilon(h)}(s) = (G_K(s;1-\mu h))^i \exp\left\{\frac{G'_X(s)}{G_X(s)}H(s)\mu h\right\} \\
&= (s - H(s)\mu h + o(h))^i \cdot \left(1 + \frac{G'_X(s)}{G_X(s)}H(s)\mu h + o(h)\right) \\
&= s^i\left(1 - \frac{H(s)}{s}\mu h + o(h)\right)^i \cdot \left(1 + \frac{G'_X(s)}{G_X(s)}H(s)\mu h + o(h)\right) \\
&= \left(s^i - is^{i-1}H(s)\mu h + o(h)\right) \cdot \left(1 + \frac{G'_X(s)}{G_X(s)}H(s)\mu h + o(h)\right) \\
&= s^i + s^i\frac{G'_X(s)}{G_X(s)}H(s)\mu h - is^{i-1}H(s)\mu h + o(h).
\end{aligned}$$

Expanding $\frac{G'_X(s)}{G_X(s)}H(s)$ and $-H(s)$ as power series of s,

$$\frac{G'_X(s)}{G_X(s)}H(s) = a_0 + a_1 s + a_2 s^2 + \cdots + a_k s^k + \cdots, \tag{8.1.3}$$

$$-H(s) = b_0 + b_1 s + b_2 s^2 + \cdots + b_k s^k + \cdots, \tag{8.1.4}$$

we finally obtain

$$\begin{aligned}
G_{X(t+h)|X(t)=i}(s) &= s^i + s^i\frac{G'_X(s)}{G_X(s)}H(s)\mu h - is^{i-1}H(s)\mu h + o(h) \\
&= i\mu b_0 h s^{i-1} + [1 + (\mu a_0 + i\mu b_1)h]s^i + (\mu a_1 + i\mu b_2)hs^{i+1} \\
&\qquad + \cdots + (\mu a_m + i\mu b_{m+1})hs^m + \cdots + o(h) \\
&= i\mu b_0 h s^{i-1} + [1 + (\mu a_0 + i\mu b_1)h]s^i + \sum_{m=1}^{\infty}(\mu a_m + i\mu b_{m+1})hs^m + o(h).
\end{aligned}$$

This shows that the infinitesimal transition probabilities are

$$\Pr[X(t+h) = i+k \mid X(t) = i] = \begin{cases} \mu(a_k + ib_{k+1})h + o(h), & k \geq 1, \\ 1 + \mu(a_0 + ib_1)h + o(h), & k = 0, \\ i\mu b_0 h + o(h), & k = -1, \\ o(h), & k < -1. \end{cases} \tag{8.1.5}$$

and the infinitesimal generator is

$$Q = \begin{pmatrix} \mu a_0 & \mu a_1 & \mu a_2 & \mu a_3 & \cdots \\ \mu b_0 & \mu a_0 + \mu b_1 & \mu a_1 + \mu b_2 & \mu a_2 + \mu b_3 & \cdots \\ 0 & 2\mu b_0 & \mu a_0 + 2\mu b_1 & \mu a_1 + 2\mu b_2 & \cdots \\ 0 & 0 & 3\mu b_0 & \mu a_0 + 3\mu b_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{8.1.6}$$

The infinitesimal generator shows special interesting patterns:

(1) Every diagonal entries are linearly increasing. For example, the lower diagonal entries are $q_{i,i-1} = i\mu b_0$ $(i = 1, 2, 3, \ldots)$.

(2) All entries below the lower diagonal $q_{i,j}$ $(j \leq i - 2)$ are zeros.

Mimicking the **upwardly skip-free** processes which define $q_{i,j} = 0$ for all $j \geq i + 2$ (refer to Anderson [1991], Chapter 9), we may call the phenomena in the continuous-time GAR(1) process with non-negative integer margins as **downwardly skip-free**. They are opposite to the upwardly skip-free processes which are basically birth-death processes, but with downward jumps called "catastrophes". In our cases, there are upward jumps, so these processes extend birth-death processes by allowing multiple births in the next time instant.

A relevant question is when such a continuous-time GAR(1) process becomes a birth-death process. To be a birth-death process, all entries above the upper diagonal $q_{i,j}$ $(j \geq i + 2)$ should be zeros. This requires that

$$\mu a_k + i\mu b_{k+1} = 0, \qquad i = 0, 1, 2, 3, \ldots, \quad k = 2, 3, 4, \ldots.$$

Because $\mu > 0$, they are equivalent to

$$a_2 = a_3 = a_4 = \cdots = 0, \qquad b_3 = b_4 = b_5 = \cdots = 0,$$

i.e., $-H(s)$ is at most a second order polynomial of s and $\frac{G'_X(s)}{G_X(s)} H(s)$ is linear in s:

$$\frac{G'_X(s)}{G_X(s)} H(s) = a_0 + a_1 s, \qquad -H(s) = b_0 + b_1 s + b_2 s^2.$$

In this situation, the infinitesimal generator becomes

$$Q = \begin{pmatrix} \mu a_0 & \mu a_1 & 0 & 0 & \cdots \\ \mu b_0 & \mu a_0 + \mu b_1 & \mu a_1 + \mu b_2 & 0 & \cdots \\ 0 & 2\mu b_0 & \mu a_0 + 2\mu b_1 & \mu a_1 + 2\mu b_2 & \cdots \\ 0 & 0 & 3\mu b_0 & a_0 + 3\mu b_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{8.1.7}$$

which indicates that these birth-death processes within the class of the continuous-time GAR(1) process with non-negative integer margins are linear birth-death processes, because the birth-rates and death-rates are

$$q_{i,i+1} = \mu a_1 + i\mu b_2, \quad q_{i,i-1} = i\mu b_0.$$

To guarantee these birth-rates and death-rates are non-negative, it is necessary that $a_1,\ b_0,\ b_2 \geq 0$.

We've already met one birth-death process which comes from the continuous-time GAR(1) process with Poisson margins when the operation is binomial-thinning. In fact, this is the only process based on binomial thinning that is a birth-death process, because that $H(s) = s - 1$ and $G'_X(s)/G_X(s)$ can only be a constant. For K being from **I2**, since $H(s) = (1-\gamma s)(s-1)/(1-\gamma)$, it is possible to find a birth-death process with a GDSD(**I2**) stationary distribution. This example will be shown below. As for K being from **I4**, since $H(s)$ is no longer a second order polynomial of s, it is impossible to find a birth-death process with a GDSD(**I4**) stationary distribution. For K being from **I5**, $H(s)$ will become a second order polynomial of s iff $\theta = 1$, which in turn becomes **I2**. Hence, for $\theta \neq 1$, there is no birth-death process with a GDSD(**I5**) stationary distribution.

Example 8.3 *Suppose K is from* **I2** *with pgf*

$$G_K(s;\alpha) = \frac{(1-\alpha) + (\alpha - \gamma)s}{(1-\alpha\gamma) - (1-\alpha)\gamma s}, \qquad \text{where } \gamma \text{ is fixed and } 0 \leq \gamma < 1.$$

Consider the marginal distribution is NB(β, γ) with pgf $G_X(s) = \left(\frac{1-\gamma}{1-\gamma s}\right)^\beta$, where $\beta > 0$. Then

$$\begin{aligned} -H(s) &= -\frac{(1-\gamma s)(s-1)}{1-\gamma} = \frac{1}{1-\gamma} - \frac{1+\gamma}{1-\gamma}s + \frac{\gamma}{1-\gamma}s^2, \\ \frac{G'_X(s)}{G_X(s)}H(s) &= \frac{\beta\gamma}{1-\gamma s} \cdot \frac{(1-\gamma s)(s-1)}{1-\gamma} = -\frac{\beta\gamma}{1-\gamma} + \frac{\beta\gamma}{1-\gamma}s. \end{aligned}$$

Thus

$$a_0 = -\frac{\beta\gamma}{1-\gamma}, \quad a_1 = \frac{\beta\gamma}{1-\gamma}, \quad a_k = 0, \qquad k = 2, 3, \ldots,$$

and

$$b_0 = \frac{1}{1-\gamma}, \quad b_1 = -\frac{1+\gamma}{1-\gamma}, \quad b_2 = \frac{\gamma}{1-\gamma}, \quad b_k = 0, \qquad k = 3, 4, \ldots.$$

Hence

$$i\mu b_0 = \frac{i\mu}{1-\gamma}, \quad a_0 + i\mu b_1 = -\frac{\beta\gamma + i\mu(1+\gamma)}{1-\gamma}, \quad a_1 + i\mu b_2 = \frac{(\beta + i\mu)\gamma}{1-\gamma}, \qquad i = 0, 1, 2, \ldots.$$

These lead to a birth-death process with infinitesimal generator

$$Q = \begin{pmatrix} -\frac{\beta\gamma}{1-\gamma} & \frac{\beta\gamma}{1-\gamma} & 0 & 0 & \cdots \\ \frac{\mu}{1-\gamma} & -\frac{\beta\gamma+\mu(1+\gamma)}{1-\gamma} & \frac{(\beta+\mu)\gamma}{1-\gamma} & 0 & \cdots \\ 0 & \frac{2\mu}{1-\gamma} & -\frac{\beta\gamma+2\mu(1+\gamma)}{1-\gamma} & \frac{(\beta+2\mu)\gamma}{1-\gamma} & \cdots \\ 0 & 0 & \frac{3\mu}{1-\gamma} & -\frac{\beta\gamma+3\mu(1+\gamma)}{1-\gamma} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

This is another special case among linear birth-death processes (refer to Anderson [1991], Section 3.2, p. 103.). It has death-rates $q_{i,i-1} = i\frac{\mu}{1-\gamma}$ $(i = 1, 2, 3, \ldots)$ *and birth-rates* $q_{i,i+1} = \frac{\beta}{1-\gamma} + i\frac{\mu\gamma}{1-\gamma}$ $(i = 0, 1, 2, \ldots)$.

In practice, the transition probability approach, namely assigning the infinitesimal transition probabilities, is one effective method to construct continuous-time Markov process models, for example, in a medical study like tumor evolvement with several mutation states. The analyses in this section show the transition feature of the continuous-time GAR(1) process with non-negative integer margins. It also provides the interpretation from the view of transition for the continuous-time GAR(1) process with non-negative integer margins. By comparison of the infinitesimal generator, one can link the continuous-time GAR(1) process to a specific continuous-time Markov model constructed by the transition probability approach.

For the continuous-time GAR(1) process with non-negative integer margins, one byproduct of the transition analysis is that the process could be linked to queuing theory with unlimited number of servers, because of the downwardly skip-free pattern and linear death-rates in the infinitesimal generator. Such a model has application in the customer self-service system. Coincidently, the process in Example 8.1 is just $M/M/\infty$. See Taylor and Karlin [1998], p. 552-553. In general, these processes belong to $M/G/\infty$ queuing models (exponential distribution of service time /general distribution of inter-arrival time/infinite servers). The exponential distribution of service time comes from the linear pattern of the lower diagonal entries. Although the inter-arrival times have exponential distributions (see Section 8.3), the parameters of these exponential distributions are not the same, but depend on transition probabilities.

Lastly, we summarize the feature of the infinitesimal transition probabilities from new concepts: instantaneous change rate of the conditional pgf and instantaneous relative change rate of the conditional pgf. Conditioned on current state, the trivial conditional pgf of the current margin is $G_{X(t)|X(t)=i}(s) = \mathbf{E}\left(s^{X(t)}|X(t)=i\right) = s^i$. For the near future given the current state, we consider the change rate or relative change rate of its conditional pgf.

Definition 8.1 *Given $X(t)=i$, the instantaneous change rate of the conditional pgf is defined as*

$$O(s;t,i) = \lim_{h\to 0} \frac{G_{X(t+h)|X(t)=i}(s) - G_{X(t)|X(t)=i}(s)}{h} = \lim_{h\to 0} \frac{G_{X(t+h)|X(t)=i}(s) - s^i}{h},$$

and the instantaneous relative change rate of the conditional pgf is defined as

$$\begin{aligned} R(s;t,i) &= \left(\lim_{h\to 0} \frac{G_{X(t+h)|X(t)=i}(s) - G_{X(t)|X(t)=i}(s)}{h}\right) \Big/ G_{X(t)|X(t)=i}(s) \\ &= \frac{1}{s^i} \lim_{h\to 0} \frac{G_{X(t+h)|X(t)=i}(s) - s^i}{h} = \lim_{h\to 0} \frac{s^{-i}G_{X(t+h)|X(t)=i}(s) - 1}{h}. \end{aligned}$$

The relationship between $O(s)$ and $R(s)$ is

$$R(s;t,i) = s^{-i}O(s;t,i), \qquad \text{or} \qquad O(s;t,i) = s^i R(s;t,i).$$

Now we calculate them for the continuous-time GAR(1) process with non-negative integer margins.

$$\begin{aligned} O(s;t,i) &= \lim_{h\to 0} \frac{G_{X(t+h)|X(t)=i}(s) - s^i}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\left(s^i + s^i \frac{G'_X(s)}{G_X(s)} H(s)\mu h - i s^{i-1} H(s)\mu h + o(h) - s^i\right) \\ &= s^i \frac{G'_X(s)}{G_X(s)} H(s)\mu - i s^{i-1} H(s)\mu, \\ R(s;t,i) &= s^{-i}O(s;t,i) = \mu \frac{G'_X(s)}{G_X(s)} H(s) - i\mu \frac{H(s)}{s}. \end{aligned}$$

The interesting pattern we find in the instantaneous relative change rate of the conditional pgf is that $R(s;t,i)$ does not depend on time t and is linear in state i.

8.1.2 Positive-valued margin

Similarly, for the continuous-time GAR(1) process with positive margins, we can study its transition property by the **instantaneous change rate of the conditional LT** and **instantaneous relative change rate of the conditional LT**, because the conditional distribution of $X(t+h)$ given $X(t)$ can be governed by its conditional LT. The instantaneous change rate of the conditional LT is defined as

$$O(s;t,x) = \lim_{h\to 0} \frac{\phi_{X(t+h)|X(t)=x}(s) - \phi_{X(t)|X(t)=x}(s)}{h} = \lim_{h\to 0} \frac{\phi_{X(t+h)|X(t)=x}(s) - e^{-sx}}{h},$$

and the instantaneous relative change rate of the conditional LT is defined as

$$\begin{aligned} R(s;t,x) &= \left(\lim_{h\to 0} \frac{\phi_{X(t+h)|X(t)=x}(s) - \phi_{X(t)|X(t)=x}(s)}{h}\right) \Big/ \phi_{X(t)|X(t)=x}(s) \\ &= \frac{1}{e^{-sx}} \lim_{h\to 0} \frac{\phi_{X(t+h)|X(t)=x}(s) - e^{-sx}}{h} = \lim_{h\to 0} \frac{e^{sx}\phi_{X(t+h)|X(t)=x}(s) - 1}{h}. \end{aligned}$$

The relationship between $O(s)$ and $R(s)$ is

$$R(s;t,x) = \frac{O(s;t,x)}{e^{-sx}}, \qquad \text{or} \qquad O(s;t,x) = e^{-sx}R(s;t,x).$$

Assume the time difference h is small enough, then the continuous-time GAR(1) process can be expressed by the the stochastic difference equation

$$X(t+h) \stackrel{d}{=} (1-\mu h) \circledast X(t) + \Delta\epsilon(h), \qquad \Delta\epsilon(h) = \epsilon(t+h) - \epsilon(t).$$

Under steady state, the margins have the same LT as the stationary distribution, denoted as $\phi_X(s)$. Then $\Delta\epsilon(h)$ has LT of form

$$\phi_{\Delta\epsilon(h)}(s) = \exp\left\{\frac{\phi'_X(s)}{\phi_X(s)} H(s)\mu h\right\}, \qquad \text{where } H(s) = \frac{\partial}{\partial\alpha}\left[-\log\phi_K(s;\alpha)\right]\Big|_{\alpha=1}.$$

Given $X(t) = x$, the conditional LT of $X(t+h)$ is

$$\begin{aligned}
\phi_{X(t+h)|X(t)=x}(s) &= (\phi_K(s;1-\mu h))^x\,\phi_{\Delta\epsilon(h)}(s) \;=\; (\phi_K(s;1-\mu h))^x \exp\left\{\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h\right\} \\
&= e^{-x[-\log\phi_K(s;1-\mu h)]}\cdot\exp\left\{\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h\right\} \\
&= e^{-x(s-H(s)\mu h+o(h))}\cdot\left(1+\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h+o(h)\right) \\
&= e^{-sx}\cdot e^{xH(s)\mu h+o(h)}\cdot\left(1+\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h+o(h)\right) \\
&= e^{-sx}\left(1+xH(s)\mu h+o(h)\right)\cdot\left(1+\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h+o(h)\right) \\
&= e^{-sx}\left(1+xH(s)\mu h+\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h+o(h)\right) \\
&= e^{-sx}+xe^{-sx}H(s)\mu h+e^{-sx}\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h+o(h).
\end{aligned}$$

The instantaneous change rate is then

$$\begin{aligned}
O(s;t,x) &= \lim_{h\to 0}\frac{\phi_{X(t+h)|X(t)=x}(s)-e^{-sx}}{h} \\
&= \lim_{h\to 0}\frac{e^{-sx}+xe^{-sx}H(s)\mu h+e^{-sx}\frac{\phi'_X(s)}{\phi_X(s)}H(s)\mu h+o(h)-e^{-sx}}{h} \\
&= \mu xe^{-sx}H(s)+\mu e^{-sx}\frac{\phi'_X(s)}{\phi_X(s)}H(s),
\end{aligned}$$

and the instantaneous relative change rate is

$$R(s;t,x) \;=\; \frac{O(s)}{e^{-sx}} \;=\; \mu xH(s)+\mu\frac{\phi'_X(s)}{\phi_X(s)}H(s) \;=\; \mu\frac{\phi'_X(s)}{\phi_X(s)}H(s)+[\mu H(s)]x.$$

This discloses the pattern of the instantaneous relative change rate, $R(s;t,x)$, is constant in t and linear in x.

The expression of $R(s;t,x)$ only involves $\phi'_X(s)/\phi_X(s)$ and $H(s)$. Hence, it also shows that for steady state continuous-time GAR(1) process, the marginal distribution and the self-generalized distribution determine the whole process. This means that different continuous-time GAR(1) process will lead to different instantaneous relative change rate of the conditional pgf. For example, suppose two continuous-time GAR(1) processes share the same marginal distribution with LT $\phi(s)$. Then, their instantaneous relative change rates of the conditional pgf are

$$
\begin{aligned}
R_1(s;t,x) &= \mu\frac{\phi'(s)}{\phi(s)}H_1(s) + [\mu H_1(s)]x, \\
R_2(s;t,x) &= \mu\frac{\phi'(s)}{\phi(s)}H_2(s) + [\mu H_2(s)]x.
\end{aligned}
$$

They differ in $H_1(s)$ and $H_2(s)$, which come from their respective self-generalized distributions. If the instantaneous relative change rates of the conditional pgf for two continuous-time GAR(1) processes are the same, they must share the same marginal distribution and $H(s)$. Later in Section 8.2, we will point out there is one-to-one mapping of the self-generalized distribution to $H(s)$, the derivative at the boundary.

The transition approach offers another perspective for the continuous-time GAR(1) process modelling. The instantaneous change rate and relative change rate of the conditional LT are associated with current time t and current state x in the state space. Considering a time period $[t_1, t_2]$, if we know the instantaneous change rate or relative change rate of the conditional LT for any time t and any state x, we can obtain the conditional LT $\phi_{X(t_2)|X(t_1)=x_1}(s)$ of $X(t_2)$ given $X(t_1) = x_1$ by integration of them over the state space and the time period $[t_1, t_2]$.

8.2 Characteristic feature of the PDE of the conditional pgf or LT

We will deduce the form of the PDE of the conditional pgf or LT for the continuous-time GAR(1) process with non-negative integer and positive margins respectively. By solving the resulting PDE, we surprisingly find that there is one-to-one mapping between the self-generalized distribution and $H(s)$ which is defined by either $\left.\frac{\partial G_K(s;\alpha)}{\partial \alpha}\right|_{\alpha=1}$ or $\left.\frac{\partial}{\partial \alpha}\left[-\log \phi_K(s;\alpha)\right]\right|_{\alpha=1}$, the partial derivatives with respective to α at the boundary 1. This indicates that the relevant partial derivative at the boundary 1 determines the entire pgf or LT of the self-generalized distribution.

8.2.1 Non-negative integer margin: PDE of the conditional pgf

Kendall [1948, 1949] established the generating function method in the study of birth-death processes. This method is also effective in other kinds of continuous-time Markov processes. For example, Brockwell, Gani and Resnick [1982], Brockwell [1985, 1986], and Pakes [1986] applied it to upwardly skip-free processes. See the summary in Anderson [1991], Chapter 9. We will use the birth-death process to illustrate the generating function method.

Suppose $\{X(t) : t > 0\}$ is a homogeneous birth-death process with initial value $X(0) = I$. The birth-rates are $q_{i,i+1} = \lambda_i$ $(i = 0, 1, 2, \ldots)$ and death-rates are $q_{i,i-1} = \mu_i$ $(i = 1, 2, 3, \ldots)$. Denote the transition probabilities

$$p_{i,j}(h) = \Pr[X(t+h) = j|X(t) = i] = \Pr[X(h) = j|X(0) = i], \qquad i,\ j = 0, 1, 2, \ldots.$$

We are focusing on the conditional pgf of $X(t)$ given $X(0) = I$:

$$G_{X(t)|X(0)=I}(s) = \mathbf{E}(s^{X(t)}|X(0) = I) = \sum_{i=0}^{\infty} s^i p(i;t) \stackrel{\text{def}}{=} G(s;t),$$

where $p(i;t) = \Pr[X(t) = i|X(0) = I]$. Note that

$$G(s;0) = G_{X(0)|X(0)=I}(s) = \mathbf{E}\left(s^{X(0)}|X(0) = I\right) = s^I.$$

Then it follows that

$$\begin{aligned}
\frac{\partial G(s;t)}{\partial t} &= -\sum_{i=0}^{\infty}(\mu_i + \lambda_i)p(i;t)s^i + \sum_{i=0}^{\infty}\lambda_i p(i;t)s^{i+1} + \sum_{i=1}^{\infty}\mu_i p(i;t)s^{i-1} \\
&= (1-s)\sum_{i=1}^{\infty}\mu_i p(i;t)s^{i-1} + (s-1)\sum_{i=0}^{\infty}\lambda_i p(i;t)s^i.
\end{aligned}$$

This result comes from by applying the Chapman-Kolmogorov equations: for $h > 0$,

$$\begin{aligned}
p(0;t+h) &= p_{1,0}(h)p(1;t) + p_{0,0}(h)p(0;t) \\
&= p_{1,0}(h)p(1;t) + (1 - p_{0,1}(h))p(0;t), \\
p(j;t+h) &= p_{j-1,j}(h)p(j-1;t) + p_{j,j}(h)p(j;t) + p_{j+1,j}(h)p(j+1;t) \\
&= p_{j-1,j}(h)p(j-1;t) + (1 - p_{j,j-1}(h) - p_{j,j+1}(h))p(j;t) \\
&\quad + p_{j+1,j}(h)p(j+1;t), \qquad\qquad j \geq 1.
\end{aligned}$$

Hence,

$$
\begin{aligned}
\frac{p(0;t+h)-p(0;t)}{h} &= \frac{p_{1,0}(h)}{h}p(1;t) - \frac{p_{0,1}(h)}{h}p(0;t), \\
\frac{p(j;t+h)-p(j;t)}{h} &= \frac{p_{j-1,j}(h)}{h}p(j-1;t) - \frac{p_{j,j-1}(h)+p_{j,j+1}(h)}{h}p(j;t) \\
&\quad + \frac{p_{j+1,j}(h)}{h}p(j+1;t), \qquad j \geq 1.
\end{aligned}
$$

Let $h \to 0$, and note that

$$
\begin{aligned}
p_{j,j+1}(h) &= \lambda_j h + o(h), \qquad \lambda_j \geq 0, \quad j = 0, 1, \ldots, \\
p_{j,j-1}(h) &= \mu_j h + o(h), \qquad \mu_0 = 0, \ \mu_j \geq 0, \quad j = 1, 2, \ldots.
\end{aligned}
$$

We obtain the derivatives

$$
\begin{aligned}
p'(0;t) &= \mu_1 p(1;t) - \lambda_0 p(0;t) = \mu_1 p(1;t) - \mu_0 p(0;t) - \lambda_0 p(0;t), \\
p'(j;t) &= \lambda_{j-1} p(j-1;t) - (\mu_j + \lambda_j) p(j;t) + \mu_{j+1} p(j+1;t), \qquad j \geq 1.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{\partial G(s;t)}{\partial t} &= \frac{\partial}{\partial t}\left(\sum_{j=0}^{\infty} s^j p(j;t)\right) = \sum_{j=0}^{\infty} s^j p'(j;t) = p'(0,t) + \sum_{j=1}^{\infty} s^j p'(j;t) \\
&= \mu_1 p(1;t) - \mu_0 p(0;t) - \lambda_0 p(0;t) \\
&\quad + \sum_{j=1}^{\infty} s^j [\lambda_{j-1} p(j-1;t) - (\mu_j + \lambda_j) p(j;t) + \mu_{j+1} p(j+1;t)] \\
&= -\mu_0 p(0;t) - \lambda_0 p(0;t) - \sum_{j=1}^{\infty} s^j (\mu_j + \lambda_j) p(j;t) \\
&\quad + \sum_{j=1}^{\infty} s^j \lambda_{j-1} p(j-1;t) + \sum_{j=1}^{\infty} s^j \mu_{j+1} p(j+1;t) + \mu_1 p(1;t) \\
&= -\sum_{i=0}^{\infty} (\mu_i + \lambda_i) p(i;t) s^i + \sum_{i=0}^{\infty} \lambda_i p(i;t) s^{i+1} + \sum_{i=1}^{\infty} \mu_i p(i;t) s^{i-1} \\
&= (1-s) \sum_{i=1}^{\infty} \mu_i p(i;t) s^{i-1} + (s-1) \sum_{i=0}^{\infty} \lambda_i p(i;t) s^i.
\end{aligned}
$$

Choosing appropriate forms of birth-rates and death-rates, we can obtain a first order linear partial differential equation for the conditional pgf $G(s;t) = G_{X(t)|X(0)=I}(s)$. For example, consider the linear birth-death process with the birth-rates $\lambda_i = a + ib$ $(i = 0, 1, 2, \ldots)$ and death-rates $\mu_i = ic$ $(i = 1, 2, 3, \ldots)$, where $a,\ b,\ c \geq 0$. Then the resulting PDE is

$$\frac{\partial G(s;t)}{\partial t} = (s-1)(bs-c)\frac{\partial G(s;t)}{\partial s} + a(s-1)G(s;t).$$

Two special cases are the processes in Examples 8.1 and 8.3; they lead to the PDE

$$\frac{\partial G(s;t)}{\partial t} = \mu(1-s)\frac{\partial G(s;t)}{\partial s} + \lambda(s-1)G(s;t)$$

and

$$\frac{\partial G(s;t)}{\partial t} = \frac{\mu}{1-\gamma}(s-1)(\gamma s-1)\frac{\partial G(s;t)}{\partial s} + \frac{\beta\gamma}{1-\gamma}(s-1)G(s;t)$$

respectively.

The generating function method can be extended to upwardly skip-free processes. Hence, a natural question is: can we apply it to downwardly skip-free processes, or specifically the continuous-time GAR(1) process with non-negative integer margins? Our goal is to find the PDE of the conditional pgf $G_{X(t)|X(0)=I}(s)$. Essentially, this can be done through the infinitesimal generator matrix. However, the continuous-time GAR(1) process with non-negative integer margins is a special kind of process, which has a particular feature in its conditional pgf of the margins:

$$G_{X(t+h)|X(t)=i}(s) = s^i + s^i\frac{G'_X(s)}{G_X(s)}H(s)\mu h - is^{i-1}H(s)\mu h + o(h).$$

Note that it is conditioned on the current state, not the starting state; see Section 8.1. Thus,

$$\begin{aligned}
\frac{\partial G(s;t)}{\partial t} &= \lim_{h\to 0}\frac{G_{X(t+h)|X(0)=I}(s) - G_{X(t)|X(0)=I}(s)}{h} \\
&= \lim_{h\to 0}\frac{1}{h}\left\{\mathbf{E}\left(s^{X(t+h)}\middle|X(0)=I\right) - \mathbf{E}\left(s^{X(t)}\middle|X(0)=I\right)\right\} \\
&= \lim_{h\to 0}\frac{1}{h}\left\{\mathbf{E}\left(\mathbf{E}\left[s^{X(t+h)}\middle|X(t)\right]\middle|X(0)=I\right) - \mathbf{E}\left(s^{X(t)}\middle|X(0)=I\right)\right\} \\
&= \lim_{h\to 0}\frac{1}{h}\left\{\sum_{i=0}^{\infty}p(i;t)\mathbf{E}\left[s^{X(t+h)}\middle|X(t)=i\right] - \sum_{i=0}^{\infty}p(i;t)s^i\right\} \\
&= \sum_{i=0}^{\infty}p(i;t)\left\{\lim_{h\to 0}\frac{1}{h}\left(\mathbf{E}\left[s^{X(t+h)}|X(t)=i\right] - s^i\right)\right\} \\
&= \sum_{i=0}^{\infty}p(i;t)\left\{\lim_{h\to 0}\frac{1}{h}\left(G_{X(t+h)|X(t)=i}(s) - s^i\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=0}^{\infty} p(i;t)\left\{\lim_{h\to 0}\frac{1}{h}\left(s^i\frac{G'_X(s)}{G_X(s)}H(s)\mu h - is^{i-1}H(s)\mu h + o(h)\right)\right\}\\
&= \sum_{i=0}^{\infty} p(i;t)\left(s^i\frac{G'_X(s)}{G_X(s)}H(s)\mu - is^{i-1}H(s)\mu\right)\\
&= -\mu H(s)\sum_{i=0}^{\infty} ip(i;t)s^{i-1} + \mu\frac{G'_X(s)}{G_X(s)}H(s)\sum_{i=0}^{\infty}p(i;t)s^i,
\end{aligned}
$$

namely

$$\frac{\partial G(s;t)}{\partial t} = -\mu H(s)\frac{\partial G(s;t)}{\partial s} + \mu\frac{G'_X(s)}{G_X(s)}H(s)G(s;t),$$

or

$$\mu H(s)\frac{\partial G(s;t)}{\partial s} + \frac{\partial G(s;t)}{\partial t} = \mu\frac{G'_X(s)}{G_X(s)}H(s)G(s;t). \tag{8.2.1}$$

This is a special form of the first order linear partial differential equations, which is defined as

$$a(s,t)u_s(s,t) + b(s,t)u_t(s,t) = c(s,t)u(s,t) + d(s,t)$$

(u_s, u_t refer to partial derivatives). The technique of solution can be found in any introductory PDE textbook, say Fritz [1981], Chapter 1, Sections 4-6. Also one can refer to Anderson [1991], p. 104-105 for a quick review.

We now use the traditional approach, called the method of characteristics, to solve this particular form of PDE. The following is an outline of the procedure to find the solution for

$$\begin{cases} \mu H(s)\frac{\partial G(s;t)}{\partial s} + \frac{\partial G(s;t)}{\partial t} = \mu\frac{G'_X(s)}{G_X(s)}H(s)G(s;t), \\ G(s;0) = s^I. \end{cases} \tag{8.2.2}$$

Let $t = t(v)$, $s = s(v,w)$, $Z = G = Z(v,w)$, and

$$\begin{cases} \frac{dt}{dv} &= 1; \\ t(0) &= 0; \end{cases} \tag{A.1}$$

$$\begin{cases} \frac{ds}{dv} &= \mu H(s); \\ s(0,w) &= w; \end{cases} \tag{A.2}$$

$$\begin{cases} \frac{dZ}{dv} &= \mu\frac{G'_X(s)}{G_X(s)}H(s)Z; \\ Z(0,w) &= G(s;t)|_{v=0} = G(s(0,w),t(0)) = w^I. \end{cases} \tag{A.3}$$

From (A.1), we obtain $t = t(v) = v$. Hence, the above equations become

$$\begin{cases} \frac{ds}{dt} &= \mu H(s); \\ s(0,w) &= w; \end{cases} \tag{B.1}$$

$$\begin{cases} \frac{dZ}{dt} &= \mu \frac{G'_X(s)}{G_X(s)} H(s) Z; \\ Z(0,w) &= G(s(0,w),0) = w^I. \end{cases} \tag{B.2}$$

From (B.1), we get

$$\begin{cases} \int \frac{ds}{-H(s)} &= -\mu t, \\ s(0,w) &= w. \end{cases}$$

Denote $\int \frac{ds}{-H(s)} = -g(s) + c$. Then, $-g(s) + c = -\mu t$. By the initial condition, we have

$$-g(s(0,w)) + c = -g(w) + c = 0.$$

Thus, the solution of (B.1) is

$$g(s) = \mu t + g(w) \qquad \text{or} \qquad g(w) = -\mu t + g(s).$$

The latter can further lead to $w = g^{-1}(-\mu t + g(s))$ if the inverse function, $g^{-1}(\cdot)$, exists. From (B.2), we have

$$\frac{dZ}{Z} = \mu \frac{G'_X(s)}{G_X(s)} H(s) dt = \frac{G'_X(s)}{G_X(s)} ds = \frac{dG_X(s)}{G_X(s)},$$

which leads to $\log Z = \log G_X(s) + c$. By the boundary condition in (B.2), we obtain $c = \log \frac{w^I}{G_X(w)}$. This leads to the solution of the original PDE:

$$G(s;t) = Z(v,w) = \frac{w^I}{G_X(w)} G_X(s) = w^I \cdot \frac{G_X(s)}{G_X(w)}, \tag{8.2.3}$$

or furthermore in terms of s and t exclusively,

$$G(s;t) = \left[g^{-1}(-\mu t + g(s))\right]^I \cdot \frac{G_X(s)}{G_X(g^{-1}(-\mu t + g(s)))}.$$

For a continuous-time GAR(1) process with non-negative integer margins, the solution of the PDE of its conditional pgf discloses the fact that $G_K(s, \alpha)$, the pgf of the non-negative integer self-generalized rv K, is determined by its partial derivative with respect to the parameter α when $\alpha = 1$, namely $H(s) = \left.\frac{\partial G_K(s;\alpha)}{\partial \alpha}\right|_{\alpha=1}$.

Why? We can compare the solution from the PDE with the conditional pgf from the stochastic representation of the process. From the stochastic representation

$$X(t)\Big|_{X(0)=I} = \left(e^{-\mu t}\right)_K \circledast I + E(0;t),$$

we know that

$$G_{X(t)|X(0)=I} = \left(G_K(s;e^{-\mu t})\right)^I G_{E(0;t)}(s).$$

Hence, it should follow that

$$G_K(s;e^{-\mu t}) = g^{-1}(-\mu t + g(s)), \qquad \text{where } g(s) = \int \frac{ds}{H(s)} + c, \text{ and } c \text{ is a constant.}$$

Furthermore, it unveils one expression of the form of the pgf of a non-negative integer self-generalized rv, i.e.,

$$G_K(s;\alpha) = g^{-1}(\log \alpha + g(s)).$$

It is straightforward to verify that $g^{-1}(\log \alpha + g(s))$ is a self-generalized function with one boundary satisfying $g^{-1}(\log 1 + g(s)) = g^{-1}(g(s)) = s$. It may be possible to derive more pgf's of self-generalized distributions with non-negative integer support from this new self-generalized function.

Consequently, there arises an open question: **Under what kind of conditions on $g(s)$, is $g^{-1}(\log \alpha + g(s))$ a pgf?**

8.2.2 Positive-valued margin: PDE of the conditional LT

Based on the same reasoning, we can obtain the corresponding PDE of the conditional LT

$$\phi_{X(t)|X(0)=x}(s) = \mathbf{E}\left(e^{-sX(t)}\Big|X(0) = x\right) \stackrel{\text{def}}{=} \phi(s;t)$$

for the continuous-time GAR(1) process with positive margins. Note that it is conditioned on the starting state.

Recall from Section 8.1.2 that

$$O(s;t,x) = \lim_{h\to 0} \frac{\phi_{X(t+h)|X(t)=x}(s) - e^{-sx}}{h} = \mu x e^{-sx} H(s) + \mu e^{-sx} \frac{\phi'_X(s)}{\phi_X(s)} H(s).$$

We have

$$\begin{aligned}
\frac{\partial \phi(s;t)}{\partial t} &= \lim_{h\to 0} \frac{\phi_{X(t+h)|X(0)=x}(s) - \phi_{X(t)|X(0)=x}(s)}{h} \\
&= \lim_{h\to 0} \frac{1}{h} \left\{ \mathbf{E}\left(e^{-sX(t+h)}\Big|X(0)=x\right) - \mathbf{E}\left(e^{-sX(t)}\Big|X(0)=x\right)\right\} \\
&= \lim_{h\to 0} \frac{1}{h} \left\{ \mathbf{E}\left(\mathbf{E}\left[e^{-sX(t+h)} \mid X(t)\right]\Big|X(0)=x\right) - \mathbf{E}\left(e^{-sX(t)}\Big|X(0)=x\right)\right\} \\
&= \lim_{h\to 0} \frac{1}{h} \left\{ \mathbf{E}\left(\mathbf{E}\left[e^{-sX(t+h)} \mid X(t)\right] - e^{-sX(t)}\Big|X(0)=x\right)\right\} \\
&= \mathbf{E}\left(\lim_{h\to 0} \frac{1}{h} \left\{\mathbf{E}\left[e^{-sX(t+h)} \mid X(t)\right] - e^{-sX(t)}\right\}\Big|X(0)=x\right) \\
&= \mathbf{E}\left(O(s;t,X(t))\Big|X(0)=x\right) \\
&= \mathbf{E}\left(\mu X(t) e^{-sX(t)} H(s) + \mu e^{-sX(t)} \frac{\phi'_X(s)}{\phi_X(s)} H(s)\Big|X(0)=x\right) \\
&= \mu H(s) \mathbf{E}\left(X(t) e^{-sX(t)}\Big|X(0)=x\right) + \mu \frac{\phi'_X(s)}{\phi_X(s)} H(s) \mathbf{E}\left(e^{-sX(t)}\Big|X(0)=x\right).
\end{aligned}$$

Note that

$$\frac{\partial \phi(s;t)}{\partial s} = -\mathbf{E}\left(X(t) e^{-sX(t)}\Big|X(0)=x\right).$$

Therefore

$$\frac{\partial \phi(s;t)}{\partial t} = -\mu H(s) \frac{\partial \phi(s;t)}{\partial s} + \mu \frac{\phi'_X(s)}{\phi_X(s)} H(s) \phi(s;t),$$

or

$$\mu H(s) \frac{\partial \phi(s;t)}{\partial s} + \frac{\partial \phi(s;t)}{\partial t} = \mu \frac{\phi'_X(s)}{\phi_X(s)} H(s) \phi(s;t). \tag{8.2.4}$$

Comparing with (8.2.1), the PDE of the conditional pgf for the continuous-time GAR(1) process with non-negative integer margins, we find that both have essentially the same form. Hence, this new PDE of the conditional LT is also a special form of first order linear partial differential equations. From the outline of the solution of the PDE of the conditional pgf, we know the following

PDE with boundary condition:

$$\begin{cases} \mu H(s)\frac{\partial\phi(s;t)}{\partial s} + \frac{\partial\phi(s;t)}{\partial t} = \mu\frac{\phi'_X(s)}{\phi_X(s)}H(s)\phi(s;t), \\ \phi(s;0) = e^{-sx}, \end{cases} \tag{8.2.5}$$

will have solution (compare (8.2.3))

$$\phi(s;t) = \frac{e^{-wx}}{\phi_X(w)}\phi_X(s) = e^{-wx}\cdot\frac{\phi_X(s)}{\phi_X(w)},$$

where w satisfies

$$g(w) - g(s) = \int\frac{ds}{-H(s)} = -\mu t, \qquad \text{or} \qquad w = g^{-1}\left(-\mu t + g(s)\right),$$

that is, w is determined by the integration $g(s) = \int\frac{ds}{H(s)} + c$, where c is a constant.

On the other hand, the stochastic representation

$$X(t)\Big|_{X(0)=I} = \left(e^{-\mu t}\right)_K \circledast I + E(0;t),$$

shows that the conditional LT is

$$\phi_{X(t)|X(0)=I} = \left(\phi_K(s;e^{-\mu t})\right)^x\phi_{E(0;t)}(s).$$

By comparison, we know that

$$\phi_K(s;e^{-\mu t}) = e^{-w} = \exp\left\{-g^{-1}\left(-\mu t + g(s)\right)\right\},$$

which suggests another general expression for the LT of a positive self-generalized rv.

For this new expression form, it is straightforward to verify that

$$\phi(s;\alpha) = \exp\left\{-g^{-1}(\log\alpha + g(s))\right\}$$

satisfies

$$\phi(-\log\phi(s;\alpha');\alpha) = \phi(s;\alpha'\alpha).$$

In addition, $\phi(s;1) = e^{-g^{-1}(\log 1+g(s))} = e^{-s}$, the same as the boundary situation for any self-generalized LT. Comparing with the LT form of positive self-generalized distribution in Theorem 3.3.3, where $\phi(s;\alpha) = \exp\{G_K(1-s;\alpha)-1\}$, we find they are matched with each other, because

$1 - G_K(1-s;\alpha)$ is a self-generalized function too. By investigation of such form functions, we may likely to find new self-generalized LT's.

Similarly, it also raises the open question: **Under what kind of conditions on $g(s)$, is $\exp\left\{-g^{-1}(\log\alpha + g(s))\right\}$ a LT?**

These analyses also disclose the fact that for a continuous-time GAR(1) process with positive margins, the LT of the involved self-generalized distribution, $\phi_K(s,\alpha)$, is determined by its partial derivative of negative logarithm with respect to parameter α when $\alpha = 1$, namely $H(s) = \frac{\partial}{\partial\alpha}[-\log\phi_K(s;\alpha)]\Big|_{\alpha=1}$.

8.2.3 Summary: margins, self-generalized distribution and increment of innovation

In the previous two subsections, we uncovered the fact that the pgf or LT of a self-generalized distribution is determined by the boundary value of its relevant partial derivative with respect to parameter α when $\alpha = 1$, namely $H(s) = \frac{\partial G_K(s;\alpha)}{\partial\alpha}\Big|_{\alpha=1}$ or $H(s) = \frac{\partial}{\partial\alpha}[-\log\phi_K(s;\alpha)]\Big|_{\alpha=1}$. This is based on the solution of the PDE of the conditional pgf or LT. For a steady state continuous-time GAR(1) process, once we know the distribution of margins and the increment of innovation, we can determine the corresponding $H(s)$, and consequently determine the self-generalized distribution. This is because that the form of pgf or LT of the increment of innovation has a special form; refer to Theorem 6.3.1. Specifically, we can obtain the form of $g(\cdot)$ by integration: $g(s) = \int \frac{ds}{H(s)} + c$. Finally, we use the general expression form $g^{-1}(-\mu t + g(s))$ or $\exp\left\{-g^{-1}(-\mu t + g(s))\right\}$ to get the the pgf or LT of the self-generalized distribution.

It is feasible to obtain the PDE for the non-stationary situation, where the instantaneous change rate/relative change rate of the conditional pgf or LT are no longer independent of time t. These will be studied further.

Recalling the study in Chapters 5 and 6, we can give some brief comments on the margins, self-generalized distribution and increment of innovation of a steady state continuous-time GAR(1) process, which has the stochastic representation

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} \circledast X(t_1) + \int_0^{t_2-t_1} e^{-\mu t} \circledast d\epsilon(t),$$

or the stochastic difference equation:

$$X(t+h) \stackrel{d}{=} (1-\mu h)_K \circledast X(t) + \Delta\epsilon(h), \qquad \Delta\epsilon(h) = \epsilon(t+h) - \epsilon(t).$$

In Chapter 5, we fixed the self-generalized distribution and the increment of innovation, and deduced the representation of $X(t_2)$ at a future time. This will in turn determine the stationary distribution, and consequently all the marginal distributions in steady state.

In Chapter 6, the customizing approach shows how to find the increment of innovation by fixing the marginal distribution (in steady state) and the self-generalized distribution. Specifically, we obtain the form of pgf, or LT, or cf of the increment of innovation in terms of the pgf, or LT, or cf of the marginal distribution and the partial derivative $H(s) = \frac{\partial G_K(s;\alpha)}{\partial \alpha}\Big|_{\alpha=1}$ or $H(s) = \frac{\partial}{\partial \alpha}[-\log \phi_K(s;\alpha)]\Big|_{\alpha=1}$. These partial derivatives can be deduced from the specified self-generalized distribution.

Combining the study in the three chapters, we conclude that a steady state continuous-time GAR(1) process essentially consists of three elements: margins, self-generalized distribution, and increment of innovation.

From any two of the three elements, we can determine the third element. Thus, in principle, any two elements will determine the entire continuous-time GAR(1) process. These three approaches offer three different viewpoints for the researchers to build a continuous-time GAR(1) process model. One can start from an easier approach which may have a clearer interpretation to construct a reasonable model for a real problem. This indicates the framework of the continuous-time GAR(1) process is quite flexible for statisticians to build models.

For a continuous-time generalized AR(1) process not in steady state, one needs the self-generalized distribution, the increment of innovation and the distribution of $X(0)$.

8.3 Distributions of sojourn time

For the continuous-time GAR(1) process with non-negative integer margins, the process can stay in one state for a certain time, then jump to another state. This sojourn time is a random variable. However, for the continuous-time GAR(1) process with positive margins, since the state space consists of non-negative real values, the distribution of the near future conditioned on current state is still continuous, and consequently it is unlikely to have sojourn time because the probability mass on one point is usually zero for the continuous distribution. (In this case, an interesting question is whether or not the process has continuous sample path.)

Hence, in this section, we consider the continuous-time GAR(1) process with non-negative integer margins only. We are interested in the distribution of the sojourn time.

Conventionally, this can be obtained with the well known infinitesimal partition method we used in Chapter 4.

Suppose $\{X(t); t \geq 0\}$ is a continuous-time GAR(1) process with non-negative integer margins. Thus the discrete state space is $S = \{0, 1, 2, 3, \ldots\}$. Suppose the process is in state i at time t_1. Denote

$$T_{ii}=\{\text{waiting time since } t_1 \text{ in state } i \text{ until next jump occurs}\}.$$

We want to find the distribution of T_{ii}. Note that T_{ii} takes value in $(0, \infty)$.

Applying the infinitesimal partition method, we will obtain n equal subintervals $[t_1, t_1 + t]$, each with length $h = t/n$:

$$[t_1, t_1 + h), \quad [t_1 + h, t_1 + 2h), \quad \ldots, \quad [t_1 + (n-1)h, t_1 + nh] = [t_1 + (n-1)h, t_1 + t].$$

When n approaches infinity, each subinterval will become an infinitesimal interval. Note the following identity:

$$\Pr[T_{ii} > t] = \Pr[X(t_1 + u) = i;\ 0 \leq u \leq t \mid X(t_1) = i].$$

By the Markov property, we can decompose the right hand side as a limit of the product of n factors on those infinitesimal intervals:

$$\Pr[X(t_1 + u) = i;\ 0 \leq u \leq t \mid X(t_1) = i]$$

$$= \lim_{n\to+\infty} \Big(\Pr[X(t_1+h)=i \mid X(t_1)=i] \times \Pr[X(t_1+2h)=i \mid X(t_1+h)=i]$$
$$\times \cdots \times \Pr[X(t_1+t)=i \mid X(t_1+(n-1)h)=i] \Big)$$
$$= \lim_{n\to+\infty} \left(\prod_{k=1}^{n} \Pr[X(t_1+kh)=i \mid X(t_1+(k-1)h)=i] \right).$$

The probability on each infinitesimal interval can be approximated according to the infinitesimal generator matrix. Hence, we may find the limit when n goes to infinity. We shall discuss it for stationary and non-stationary situation respectively.

First, we consider the stationary situation, that is, the process is homogeneous over time. Assume that the infinitesimal transition probability of remaining in the same state is

$$\Pr[X(t+h)=i \mid X(t)=i] = 1 + q_{i,i}h + o(h),$$

where $q_{i,i} < 0$ ($i = 0, 1, 2, \ldots.$), and h is infinitesimal time increment. Then, it follows that

$$\begin{aligned}
\Pr[T_{ii} > t] &= \Pr[X(t_1+u)=i;\ 0 \le u \le t \mid X(t_1)=i] \\
&= \lim_{n\to+\infty} \left(\prod_{k=1}^{n} \Pr[X(t_1+kh)=i \mid X(t_1+(k-1)h)=i] \right) \\
&= \lim_{n\to+\infty} (\Pr[X(h)=i \mid X(0)=i])^n \\
&= \lim_{n\to+\infty} (1+q_{i,i}h+o(h))^n \;=\; \lim_{n\to+\infty} \left(1+q_{i,i}\frac{t}{n}+o(h)\right)^n \\
&= e^{q_{i,i}t} \;=\; e^{-|q_{i,i}|t}.
\end{aligned} \tag{8.3.1}$$

Therefore, $T_{ii} \sim \text{exponential}(|q_{i,i}|)$.

Secondly, we turn to non-stationary situation, in which the infinitesimal transition probability of remaining in the same state is

$$\Pr[X(t+h)=i \mid X(t)=i] = 1 + q_{i,i}(t)h + o(h),$$

where $q_{i,i}(t) < 0$ ($i = 1, 2, \ldots.$), and h is infinitesimal time increment. Suppose $q_{i,i}(t)$ is differentiable with bounded first order derivative. By Proposition 5.1.1, it then follows that

$$\Pr[T_{ii} > t] \;=\; \Pr[X(t_1+u)=i;\ 0 \le u \le t \mid X(t_1)=i]$$

$$
\begin{aligned}
&= \lim_{n\to+\infty}\left(\prod_{k=1}^{n}\Pr[X(t_1+kh)=i \mid X(t_1+(k-1)h)=i]\right)\\
&= \lim_{n\to+\infty}\prod_{k=1}^{n}[1+q_{i,i}(t_1+(k-1)h)h+o(h)]\\
&= \lim_{n\to+\infty}\exp\left\{\sum_{k=1}^{n}q_{i,i}(t_1+(k-1)h)h\right\}\\
&= \exp\left\{\lim_{n\to+\infty}\sum_{k=1}^{n}q_{i,i}(t_1+(k-1)h)h\right\}\\
&= \exp\left\{\int_{t_1}^{t}q_{i,i}(\tau)d\tau\right\}.
\end{aligned}
\tag{8.3.2}
$$

So, the distribution of T_{ii} is

$$
F_{T_{ii}}(t)=\Pr[T_{ii}\le t]=1-\Pr[T_{ii}>t]=1-\exp\left\{\int_{t_1}^{t}q_{i,i}(\tau)d\tau\right\}.
$$

Hence, T_{ii} need not be exponential distributed.

The study of sojourn time provides another perspective on the continuous-time GAR(1) processes with non-negative integer margins. For example, for the stationary situation, once we know the infinitesimal generator matrix, we can simulate the process by the embedding method, namely, simulate the waiting time in a state, then jump to another state based on the conditional probability mass function from a row of the infinitesimal generator matrix (excluding the current state, or diagonal entry), and so forth; see Section 12.5. Also for a continuous observation process like in queuing theory, the sojourn time can be observed, hence, it offers more information for the inference on the studied process.

Chapter 9

Conditional and joint distributions

In this chapter, we turn to study the conditional distributions and joint distributions resulting from the continuous-time GAR(1) processes. From the point of view of distribution theory, defining a discrete-time higher order Markov process is equivalent to defining a multivariate distribution for adjacent observations, in which the dependence structure stipulates the dynamic mechanism of the process. Conversely, the process provides an approach to construct multivariate distributions. This view will be discussed in Section 9.1.

We will also calculate the conditional mean and variance in Section 9.2, as well as the autocorrelation coefficient in Section 9.3. These statistics are very useful in the estimation of parameters with sample data. In addition, we study the bivariate and multivariate distributions resulting from the continuous-time GAR(1) processes in Section 9.3. Some of them are new compared with those existing in the literature.

9.1 Consistency in process construction: the view from distribution theory

Statistical inference is made based on a chance model. Diverse chance mechanisms are described through probability distributions. Essentially, statistical modelling consists of specifying an appropriate probabilistic framework for the practical problem.

A process (discrete-time or continuous-time) denoted as $\{X_t;\ t \in T\}$ is the collection of denumerable or innumerable random variables on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, which in turn leads to joint distributions for every finite subset of rv's like $(X_{t_1}, X_{t_2}, \ldots, X_{t_m})$. For a discrete-time process denoted as $\{X_n;\ n = 0, 1, 2, 3, \}$, the path is just the sequence $\{X_0(\omega),\ X_1(\omega),\ X_2(\omega), \ldots\}$, and the construction of such a process is equivalent to specifying all finite dimensional joint distributions:

$$F_{(X_{n_1}, X_{n_2}, \ldots, X_{n_m})}(x_1, x_2, \ldots, x_m) = \Pr[X_{n_1} \le x_1, X_{n_2} \le x_2, \ldots, X_{n_m} \le x_m], \qquad m \in \mathcal{N}.$$

For a continuous-time process denoted as $\{X(t);\ t \ge 0\}$, the path is a function of t, denoted as $X(\omega; t)$. Since for a time interval, there are innumerable rv's, the construction is a bit more complicated. In this situation, we need to resort to the infinitesimal partition method again when we evaluate an event over a continuous time period. For example, suppose we want to find

$$\Pr[X(t) \le f(t);\ A \le t \le B], \qquad f(t) \text{ is a function of } t.$$

Partition the interval $[A, B]$ by points $t_1, t_2, \ldots, t_n$ so that each piece $[t_i, t_{i+1}]$ is very small. As n goes to infinity, the length of each piece will go to zero. With the additional requirement like stochastic continuity: $\lim_{t \to t'} \Pr[|X(t) - X(t')| \ge \epsilon] = 0$ for every $\epsilon > 0$ and every t, we can obtain

$$\Pr[X(t) \le f(t);\ A \le t \le B] = \lim_{n \to \infty} \Pr[X(t_1) \le f(t_1), X(t_2) \le f(t_2), \ldots, X(t_n) \le f(t_n)].$$

This requires us to specify the joint distribution of $(X(t_1), X(t_2), \ldots, X(t_n))$ first; namely for any finite number of time points $t_1, t_2, \ldots, t_n$, we need to specify the joint distribution

$$F_{(X(t_1), X(t_2), \ldots, X(t_n))}(x_1, x_2, \ldots, x_n) = \Pr[X(t_1) \le x_1, X(t_2) \le x_2, \ldots, X(t_n) \le x_n], \qquad n \in \mathcal{N}.$$

For more details about defining a process, see Karlin and Taylor [1975], p. 32-33, Doob [1953], Chapter 2, and Breiman [1992], Chapter 12.

Basically, defining a process is equivalent to specifying all possible multivariate distributions with finite dimension no matter whether it is discrete-time or continuous-time.

However, those defined multivariate distributions can not be given arbitrarily. They should satisfy the consistency condition proposed by Kolmogorov in 1933. Based on this consistency condition, Kolmogorov's extension theorem asserts that there exists a probability space and a process on it such that any finite number of rv's from the process has the same joint distribution as that prescribed in the consistency condition. This theorem is also referred as the first fundamental theorem in the theory of stochastic processes; see Chung [1974], p. 60-61. The following states the consistency condition in terms of distribution functions: for each $m \geq 1$ and $(x_1, x_2, \ldots, x_m) \in \Re^m$ and $t_1, t_2, \ldots, t_n \in T$ which need not be ordered, if $n > m$, then

$$\lim_{\substack{x_{m+1}\to\infty \\ \cdots\cdots \\ x_n\to\infty}} F_{(X_{t_1},\ldots,X_{t_m},X_{t_{m+1}},\ldots,X_{t_n})}(x_1,\ldots,x_m,x_{m+1},\ldots,x_n) = F_{(X_{t_1},\ldots,X_{t_m})}(x_1,\ldots,x_m).$$

This means that the lower dimensional marginal distribution derived from a prescribed higher dimensional joint distribution should be the same as the corresponding prescribed joint distribution with lower dimension.

Now we look into the construction of a Markov process, and investigate whether it is well defined from the point of view of consistency. Assume the state space is S. The key feature of a Markov process is that the future depends only on the present, not the past.

First, we consider defining a discrete-time Markov process $\{X_n;\ n \geq 0\}$. In this situation, we specify the conditional probability structure for any two neighbors: $\Pr[X_{n+1} = y \mid X_n = x]$ for any $x,\ y \in S$. Once all neighborhood conditional structures are defined, and given the distribution of starting point, namely the distribution of X_0, then all finite dimensional joint distributions are stipulated. This is because that for any positive integer n, the joint distribution of $(X_0, X_1, \ldots, X_{n-1}, X_n)$ can be obtained by the following equation

$$f_{(X_0,X_1,\ldots,X_{n-1},X_n)} \quad = \quad f_{X_0} \times f_{X_1|X_0} \times \cdots \times f_{X_n|X_{n-1}}.$$

Here we employ the notations $f_{(X_0,X_1,\ldots,X_{n-1},X_n)}, f_{X_0}, f_{X_1|X_0}, \ldots, f_{X_n|X_{n-1}}$ to denote the pmf in the discrete case or pdf in the continuous case to avoid the tedious work of setting separate notation for the two cases. The joint distribution of any finite dimensional vector of random variables $(X_{m_1}, X_{m_2}, \ldots, X_{m_k})$ $(0 < m_1 < m_2 < \cdots < m_k)$ can then be derived by integrating (with respect to the appropriate measure, e.g., Lebesgue or counting) out irrelevant variables in the joint distribution of higher dimension of vector $(X_0, X_1, \ldots, X_{m_k})$. In this situation, the joint distribution of any finite dimensional vector of random variables is just prescribed in the way that the marginal distribution is obtained from a higher dimensional joint distribution. This means that consistency is automatically guaranteed. Therefore, Markov processes in the discrete-time situation are well defined if the conditional probability structures of all neighboring pairs of variables are prescribed. These conditional probability structures can be arbitrarily specified. We do not need to impose any conditions on such conditional probability structures.

The previous discussed Markov process is usually referred to as the first-order Markov process because the current state only depends on the last one. Similarly, for a higher order Markov process in which the current state depends on the last few states, we can have the same conclusion that the process is well defined if all conditional probability structures of the current state given certain previous neighbors, $X_n \mid X_{n-1}, \ldots, X_{n-k}$, are specified. No restrictions on these conditional probability structures are required.

Returning to modelling, we usually impose certain requirements on the marginal distributions of discrete sets of time points. For example, a common assumption is stationarity, which leads to all univariate marginal distributions being the same, namely the same as that of the starting point. In this specific situation, defining a first-order Markov process is equivalent to defining a bivariate distribution of (X_{n-1}, X_n) with the common univariate marginal distribution, because the distributions of X_{n-1} and X_n are the same, and the conditional probability structure of $X_n \mid X_{n-1}$ is the same for every n. In general, defining a stationary m^{th} order discrete-time Markov process is equivalent to defining a $(m+1)$-dimensional multivariate distribution of $(Y_1, Y_2, \ldots, Y_m, Y_{m+1})$ where $f_{Y_1} = f_{Y_2} = \cdots = f_{Y_{m+1}}$ and $f_{(Y_1,\ldots,Y_m)} = f_{(Y_2,\ldots,Y_{m+1})}$.

Let us look into Model (2.1) in Joe [1996] to check how it is well defined from the viewpoint of distribution theory. Let $\{F_\theta\}$ be a family of ID distributions, $G_{\alpha\theta,(1-\alpha)\theta,y}$ corresponds to conditional distribution of Z_1 given $Z_1 + Z_2 = y$, where $Z_1 \sim F_{\alpha\theta},\ Z_2 \sim F_{(1-\alpha)\theta}$. This model has stochastic representation

$$Y_t = A_t(Y_{t-1}) + \epsilon_t,$$

where Y_{t-1} have distribution F_θ, ϵ_t has distribution $F_{(1-\alpha)\theta}$, A is a random operator such that $A(Y)$ given $Y = y$ has distribution $G_{\alpha\theta,(1-\alpha)\theta,y}$ and $A(Y) \sim F_{\alpha\theta}$ when $Y \sim F_\theta$. $A_t(Y_{t-1})$ and ϵ_t are independent. Here $0 < \alpha < 1$ and $\theta > 0$. F_θ is an infinitely divisible convolution-closed parametric family such that $F_{\theta_1} * F_{\theta_2} = F_{\theta_1+\theta_2}$, where $*$ is the convolution operator. This is a typical Markov process. The conditional distribution of Y_t given $Y_{t-1} = y$ is $G_{\alpha\theta,(1-\alpha)\theta,y} * F_{(1-\alpha)\theta}$. Thus, the bivariate distribution of (Y_{t-1}, Y_t) is determined by the marginal distribution of Y_{t-1} and the conditional distribution of Y_t given $Y_{t-1} = y$. By induction, the marginal distribution of Y_t is also F_θ. This model is well defined because of the appropriate marginal distribution and conditional probability structure.

Conversely, if we define a bivariate distribution such that the two univariate margins are the same, and the conditional probability structure of one margin given another one is the same as before, we of course can obtain the same stochastic representation of one variable in terms of another one as in Model (2.1) in Joe [1996]. Such a model can be applied in count data or positive data time series, and it also unifies many models appeared in the literature like Lewis [1983], Lewis *et al.* [1989], McKenzie [1986], [1988], Al-Osh and Alzaid [1987], [1991] Al-Osh and Aly [1992], Barndorff-Nielsen and Jørgensen [1991], Alzaid and Al-Osh [1993]. Jørgensen and Song [1998] gives further study for this model. Note that a restriction on such models is that the innovation and the margins have distributions in the same family. Also in general, they do not extend to continuous time.

Next we turn to defining a continuous-time Markov process $\{X(t); t \geq 0\}$. We specify the conditional probability structure of the current state given the previous state. In discrete-time situation, the sequence of random variables are denumerable so that the neighbors are fixed. However, in continuous-time situation, the random variables are innumerable, and even worse, we can not fix the neighbor of time point t. We have to specify the conditional probability structure of $X(t)$ given $X(t') = x$ for any $t' < t$, i.e., $f_{X(t)|X(t')=x}(\cdot)$ or $f_{X(t)|X(t')}(\cdot \mid x)$.

With such a specification, for $t_1 < t_2 < \cdots < t_n$, we can obtain the joint distribution of $(X(t_1), X(t_2), \ldots, X(t_n))$ by the Markov property:

$$f_{(X(t_1),X(t_2),\ldots,X(t_n))}(x_1, x_2, \ldots, x_n) = f_{X(t_1)}(x_1) \times f_{X(t_2)|X(t_1)}(x_2 \mid x_1) \times \cdots \times f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1}).$$

We view this way as the prescription for the joint distributions. Now we check the consistency. Suppose $t_1 < t_2 < t_3$. Then the prescribed joint distribution of $(X(t_1), X(t_3))$ is

$$f_{(X(t_1),X(t_3))} = f_{X(t_1)} \times f_{X(t_3)|X(t_1)},$$

while the prescribed joint distribution of $(X(t_1), X(t_2), X(t_3))$ is

$$f_{(X(t_1),X(t_2),X(t_3))} = f_{X(t_1)} \times f_{X(t_2)|X(t_1)} \times f_{X(t_3)|X(t_2)}.$$

From the latter, we can derive the joint distribution of $(X(t_1), X(t_3))$ by integrating over possible values of $X(t_2)$:

$$f_{(X(t_1),X(t_3))}(x_1, x_3) = \int f_{X(t_1)}(x_1) f_{X(t_2)|X(t_1)}(x_2 \mid x_1) f_{X(t_3)|X(t_2)}(x_3 \mid x_2) d\nu(x_2).$$

Here $\nu(\cdot)$ is an appropriate measure. This raises the consistency problem, namely whether

$$f_{X(t_3)|X(t_1)} = \int f_{X(t_2)|X(t_1)}(x_2 \mid \cdot) \cdot f_{X(t_3)|X(t_2)}(\cdot \mid x_2) d\nu(x_2). \tag{9.1.1}$$

This means that the conditional probability structure $f_{X(t)|X(t')=x}(\cdot)$ is not arbitrary. It is totally unlike the discrete-time situation where the conditional probability structure $f_{X_n|X_{n-1}=x}(\cdot)$ can be arbitrary. Denote for $t' < t$

$$G_{X(t)|X(t')=x}(s) = \mathbf{E}\left(s^{X(t)} \mid X(t') = x\right), \quad \text{or} \quad G_{X(t)|X(t')}(s) = \mathbf{E}\left(s^{X(t)} \mid X(t')\right),$$

$$\phi_{X(t)|X(t')=x}(s) = \mathbf{E}\left(e^{-sX(t)} \mid X(t') = x\right), \quad \text{or} \quad \phi_{X(t)|X(t')}(s) = \mathbf{E}\left(e^{-sX(t)} \mid X(t')\right),$$

$$\varphi_{X(t)|X(t')=x}(s) = \mathbf{E}\left(e^{isX(t)} \mid X(t') = x\right), \quad \text{or} \quad \varphi_{X(t)|X(t')}(s) = \mathbf{E}\left(e^{isX(t)} \mid X(t')\right),$$

for non-negative integer, positive or real support respectively. Then the consistency requires that

$$G_{X(t_3)|X(t_1)=x}(s) = \mathbf{E}\left(\mathbf{E}\left(s^{X(t_3)} \middle| X(t_2)\right) \middle| X(t_1) = x\right) = \mathbf{E}_{X(t_2)}\left(G_{X(t_3)|X(t_2)}(s) \mid X(t_1) = x\right),$$

or

$$\phi_{X(t_3)|X(t_1)=x}(s) = \mathbf{E}\left(\mathbf{E}\left(e^{-sX(t_3)} \middle| X(t_2)\right) \middle| X(t_1) = x\right) = \mathbf{E}_{X(t_2)}\left(\phi_{X(t_3)|X(t_2)}(s) \mid X(t_1) = x\right),$$

or

$$\varphi_{X(t_3)|X(t_1)=x}(s) = \mathbf{E}\left(\mathbf{E}\left(e^{isX(t_3)} \middle| X(t_2)\right) \middle| X(t_1) = x\right) = \mathbf{E}_{X(t_2)}\left(\varphi_{X(t_3)|X(t_2)}(s) \mid X(t_1) = x\right)$$

respectively.

When we impose stationarity on a continuous-time Markov process, then for all t, $X(t)$ has the same distribution as $X(0)$, and the conditional probability structure of $X(t)$ given $X(t')$ only depends on the time difference $t - t'$. This may come from the practical consideration of modelling. In this situation, the continuous-time Markov process leads to a trivariate distribution of $(X(t_1), X(t_2), X(t_3))$ such that the conditional probability structures $f_{X(t_2)|X(t_1)=x}(\cdot)$, $f_{X(t_3)|X(t_2)=x}(\cdot)$ and $f_{X(t_3)|X(t_1)=x}(\cdot)$ have the same form and Equation 9.1.1 holds.

On the other hand, if there exists such a trivariate distribution, we can construct a stationary continuous-time Markov process based on it and the consistency is guaranteed by the feature of this trivariate distribution. Therefore, defining a stationary continuous-time Markov process is equivalent to defining a trivariate distribution with a special property satisfied by its conditional distributions.

This is the big gap between discrete-time and continuous-time Markov processes. Relatively, defining a required bivariate distribution is much easier than defining a required trivariate distribution. Hence, this could be a partial reason for the phenomena that discrete-time Markov processes were often developed earlier in the literature than the continuous-time Markov processes as we discussed in Section 6.4.

For higher order continuous-time Markov processes, the consistency conditions will be more complicated. However, essentially, defining a stationary higher order continuous-time Markov process (if possible) is equivalent to defining a higher dimensional multivariate distribution.

Sampling on equally-spaced time points from a continuous-time Markov process, we can obtain a discrete-time Markov process. Hence, if a stationary discrete-time Markov process is actually coming from a stationary continuous-time Markov process, then its conditional distributions of three neighboring points X_{n-2}, X_{n-1} and X_n will satisfy the consistency conditions automatically. We illustrate this issue by looking into the situation where the support of marginal distribution of X_n $(n \geq 0)$ is non-negative integer. Stationarity leads to the identity $G_{X_n|X_{n-1}=x}(s) = G_{X_{n-1}|X_{n-2}=x}(s)$. We can obtain the conditional pgf of X_n given $X_{n-2} = x$ as

$$G_{X_n|X_{n-2}=x}(s) = \mathbf{E}_{X_{n-1}}(G_{X_n|X_{n-1}}(s) \mid X_{n-2} = x).$$

The resulting $G_{X_n|X_{n-2}=x}(s)$ should have the same form as $G_{X_n|X_{n-1}=x}(s)$ or $G_{X_{n-1}|X_{n-2}=x}(s)$, because they have been defined so in a continuous-time Markov process.

This indicates that if we wish to extend a stationary discrete-time Markov process to a stationary continuous-time Markov process, the consistency among three neighboring points is a necessary condition. If a discrete-time Markov process doesn't satisfy this consistency condition, we can not expect to extend it to a continuous-time Markov process.

Finally, we end this section with two theorems to illustrate the construction of a continuous-time Markov process from the perspective of distribution theory.

Theorem 9.1.1 *Define a trivariate distribution of* $(X(t_1), X(t_2), X(t_3))$ *for any* $0 \le t_1 < t_2 < t_3$ *based on the following.*

(1) The distribution of $X(t_1)$ *is GDSD associated with a self-generalized distribution which has pgf* $G_K(s;\alpha)$ $(0 \le \alpha \le 1)$. *Assume the pgf of* $X(t_1)$ *is* $G(s)$.

(2) For $t' < t$, *the conditional pgf of* $X(t)$ *given* $X(t') = x$ *is*

$$G_{X(t)|X(t')=x}(s) = \mathbf{E}\left(s^{X(t)} \middle| X(t') = x\right) = G_K^x\left(s; e^{-\mu(t-t')}\right) \times \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t-t')}\right)\right)}.$$

Then the resulting trivariate distribution is consistent with a stationary continuous-time Markov process which has $G(s)$ *as the pgf of the univariate marginal distributions.*

Proof: The key step is to show that this trivariate distribution is well defined, or in another words, the conditional pgf's are consistent. It suffices to prove

$$G_{X(t_3)|X(t_1)=x}(s) = \mathbf{E}_{X(t_2)}\left(G_{X(t_3)|X(t_2)}(s) \mid X(t_1) = x\right).$$

From (2) in the definition of the trivariate distribution, we have

$$\begin{aligned}
G_{X(t_2)|X(t_1)=x}(s) &= G_K^x\left(s; e^{-\mu(t_2-t_1)}\right) \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t_2-t_1)}\right)\right)}, \\
G_{X(t_3)|X(t_2)=x}(s) &= G_K^x\left(s; e^{-\mu(t_3-t_2)}\right) \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)}, \\
G_{X(t_3)|X(t_1)=x}(s) &= G_K^x\left(s; e^{-\mu(t_3-t_1)}\right) \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t_3-t_1)}\right)\right)}.
\end{aligned}$$

By straightforward algebra, it follows that

$$
\begin{aligned}
\mathbf{E}_{X(t_2)}\left(G_{X(t_3)|X(t_2)}(s) \mid X(t_1)=x\right) &= \sum_{y=0}^{\infty} G_{X(t_3)|X(t_2)=y}(s)\,\Pr[X(t_2)=y \mid X(t_1)=x] \\
&= \sum_{y=0}^{\infty} G_K^y\left(s; e^{-\mu(t_3-t_2)}\right) \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)} \Pr[X(t_2)=y \mid X(t_1)=x] \\
&= \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)} \sum_{y=0}^{\infty} G_K^y\left(s; e^{-\mu(t_3-t_2)}\right) \Pr[X(t_2)=y \mid X(t_1)=x] \\
&= \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)} G_{X(t_2)|X(t_1)=x}\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right)\right) \\
&= \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)} \\
&\quad \times G_K^x\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right); e^{-\mu(t_2-t_1)}\right) \frac{G(G_K\left(s; e^{-\mu(t_3-t_2)}\right))}{G\left(G_K\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right); e^{-\mu(t_2-t_1)}\right)\right)} \\
&= G_K^x\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right); e^{-\mu(t_2-t_1)}\right) \frac{G(s)}{G\left(G_K\left(G_K\left(s; e^{-\mu(t_3-t_2)}\right); e^{-\mu(t_2-t_1)}\right)\right)} \\
&= G_K^x\left(s; e^{-\mu(t_3-t_1)}\right) \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t_3-t_1)}\right)\right)},
\end{aligned}
$$

which equals $G_{X(t_3)|X(t_1)=x}(s)$ obtained from (2) in the definition. Therefore, the trivariate distribution is well defined.

The trivariate distribution of $(X(t_1), X(t_2), X(t_3))$ can be written down in

$$f_{(X(t_1),X(t_2),X(t_3))} = f_{X(t_1)} \times f_{X(t_2)|X(t_1)} \times f_{X(t_3)|X(t_2)}.$$

This motivates us to define the finite dimensional joint distribution of $(X(t_1), X(t_2), \ldots, X(t_n))$ for $t_1 < t_2 < \cdots < t_n,\ n \geq 3$, in such a way:

$$f_{(X(t_1),X(t_2),\ldots,X(t_n))} = f_{X(t_1)} \times f_{X(t_2)|X(t_1)} \times \cdots \times f_{X(t_n)|X(t_{n-1})}.$$

Now we wish that this defined family is consistent. Because of the consistency feature of the trivariate distribution which satisfies:

$$f_{X(t)|X(t'')}(z \mid x) = \int f_{X(t')|X(t'')}(y \mid x) f_{X(t)|X(t')}(z \mid y) d\nu(y), \qquad \text{for any } t'' < t' < t,$$

we can take expectation over $X(t_{i+1}), \ldots, X(t_{j-1})$ for $f_{X(t_{i+1})|X(t_i)} \times \cdots \times f_{X(t_j)|X(t_{j-1})}$ and obtain

$$\begin{aligned}
&\int \cdots \int f_{X(t_{i+1})|X(t_i)}(x_{i+1} \mid x_i) \times \cdots \times f_{X(t_j)|X(t_{j-1})}(x_j \mid x_{j-1}) d\nu(x_{i+1}) \cdots d\nu(x_{j-1}) \\
&= \int \cdots \int f_{X(t_{i+1})|X(t_i)}(x_{i+1} \mid x_i) \times \cdots \times f_{X(t_{j-2})|X(t_{j-3})}(x_{j-2} \mid x_{j-3}) \\
&\quad \times \left(\int f_{X(t_{j-1})|X(t_{j-2})}(x_{j-1} \mid x_{j-2}) f_{X(t_j)|X(t_{j-1})}(x_j \mid x_{j-1}) d\nu(x_{j-1}) \right) d\nu(x_{i+1}) \cdots d\nu(x_{j-2}) \\
&= \int \cdots \int f_{X(t_{i+1})|X(t_i)}(x_{i+1} \mid x_i) \times \cdots \times f_{X(t_{j-2})|X(t_{j-3})}(x_{j-2} \mid x_{j-3}) \\
&\qquad\qquad \times f_{X(t_j)|X(t_{j-2})}(x_j \mid x_{j-2}) d\nu(x_{i+1}) \cdots d\nu(x_{j-2}) \\
&\vdots \\
&= \int f_{X(t_{i+1})|X(t_i)}(x_{i+1} \mid x_i) f_{X(t_j)|X(t_{i+1})}(x_j \mid x_{i+1}) d\nu(x_{i+1}) \\
&= f_{X(t_j)|X(t_i)}(x_j \mid x_i),
\end{aligned}$$

for $1 \leq i < j \leq n$. Consistency always obtains when integrating out $X(t_1)$ or $X(t_n)$. Hence, by induction, it follows that for any subset $(X(t_{m_1}), \ldots, X(t_{m_k}))$ of $(X(t_1), X(t_2), \ldots, X(t_n))$, the joint distribution of $(X(t_{m_1}), \ldots, X(t_{m_k}))$ deduced from the higher dimensional joint distribution of $(X(t_1), X(t_2), \ldots, X(t_n))$ by integrating out irrelevant arguments, is just the same as that from the direct definition. Therefore, the defined family is consistent.

According to Kolmogorov's extension theorem, there exists a corresponding process $\{X(t); t \geq 0\}$.

Now we calculate the pgf of $X(t)$ if the pgf of $X(t')$ is $G(s)$:

$$\begin{aligned}
G_{X(t)}(s) &= \mathbf{E}\left(\mathbf{E}\left(s^{X(t)} \middle| X(t')\right)\right) \\
&= \mathbf{E}\left(G_K^{X(t')}\left(s; e^{-\mu(t-t')}\right) \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t-t')}\right)\right)}\right) \\
&= \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t-t')}\right)\right)} \mathbf{E}\left(G_K^{X(t')}\left(s; e^{-\mu(t-t')}\right)\right) \\
&= \frac{G(s)}{G\left(G_K\left(s; e^{-\mu(t-t')}\right)\right)} G\left(G_K\left(s; e^{-\mu(t-t')}\right)\right) \\
&= G(s),
\end{aligned}$$

which is the same as the LT of $X(t')$. This fact indicates the distribution of $X(t_2)$ and $X(t_3)$ are the same as $X(t_1)$. Furthermore, the trivariate distribution corresponds to the stationary continuous-time Markov process with stochastic representation

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t,t'), \qquad 0 \le t' < t,$$

where the cumulative innovation $E(t,t')$ has the pgf $G(s)\Big/G\left(G_K\left(s; e^{-\mu(t-t')}\right)\right)$, and is independent of $\left(e^{-\mu(t-t')}\right)_K \circledast X(t')$. This comes from (2) in the definition of the trivariate distribution.

Theorem 9.1.2 *Define a trivariate distribution of* $(X(t_1), X(t_2), X(t_3))$ *for any* $0 \le t_1 < t_2 < t_3$ *based on the following.*

(1) The distribution of $X(t_1)$ *is GSD associated with a self-generalized distribution which has LT* $\phi_K(s;\alpha)$ $(0 \le \alpha \le 1)$*. Assume the LT of* $X(t_1)$ *is* $\phi(s)$*.*

(2) For $t' < t$*, the conditional LT of* $X(t)$ *given* $X(t') = x$ *is*

$$\phi_{X(t)|X(t')=x}(s) = \mathbf{E}\left(e^{sX(t)}\Big|X(t') = x\right) = \phi_K^x\left(s; e^{-\mu(t-t')}\right)\frac{\phi(s)}{\phi\left(-\log\phi_K\left(s; e^{-\mu(t-t')}\right)\right)}.$$

Then the resulting trivariate distribution is consistent with a stationary continuous-time Markov process which has $\phi(s)$ *as the LT of the univariate marginal distributions.*

Proof: This is similar to the proof of Theorem 9.1.1. We only check the consistency. From (2) in the definition of the trivariate distribution, it follows that

$$\begin{aligned}
\phi_{X(t_2)|X(t_1)=x}(s) &= \phi_K^x\left(s; e^{-\mu(t_2-t_1)}\right)\frac{\phi(s)}{\phi\left(-\log\phi_K\left(s; e^{-\mu(t_2-t_1)}\right)\right)},\\
\phi_{X(t_3)|X(t_2)=x}(s) &= \phi_K^x\left(s; e^{-\mu(t_3-t_2)}\right)\frac{\phi(s)}{\phi\left(-\log\phi_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)},\\
\phi_{X(t_3)|X(t_1)=x}(s) &= \phi_K^x\left(s; e^{-\mu(t_3-t_1)}\right)\frac{\phi(s)}{\phi\left(-\log\phi_K\left(s; e^{-\mu(t_3-t_1)}\right)\right)}.
\end{aligned}$$

By algebra, we have

$$\begin{aligned}
&\mathbf{E}_{X(t_2)}\left(\phi_{X(t_3)|X(t_2)}(s) \mid X(t_1) = x\right)\\
&= \mathbf{E}\left(\phi_K^{X(t_2)}\left(s; e^{-\mu(t_3-t_2)}\right)\frac{\phi(s)}{\phi\left(-\log\phi_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)}\Bigg|X(t_1) = x\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\phi(s)}{\phi\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)} \mathbf{E}\left(\phi_K^{X(t_2)}\left(s; e^{-\mu(t_3-t_2)}\right) \Big| X(t_1) = x\right) \\
&= \frac{\phi(s)}{\phi\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)} \times \phi_{X(t_2)|X(t_1)=x}\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right)\right) \\
&= \frac{\phi(s)}{\phi\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right)\right)} \times \phi_K^x\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right); e^{-\mu(t_2-t_1)}\right) \\
&\qquad\qquad \times \frac{\phi(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right))}{\phi\left(-\log \phi_K\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right); e^{-\mu(t_2-t_1)}\right)\right)} \\
&= \phi_K^x\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right); e^{-\mu(t_2-t_1)}\right) \times \frac{\phi(s)}{\phi\left(-\log \phi_K\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_2)}\right); e^{-\mu(t_2-t_1)}\right)\right)} \\
&= \phi_K^x\left(s; e^{-\mu(t_3-t_1)}\right) \frac{\phi(s)}{\phi\left(-\log \phi_K\left(s; e^{-\mu(t_3-t_1)}\right)\right)} \\
&= \phi_{X(t_3)|X(t_1)=x}(s).
\end{aligned}
$$

This completes the proof.

Remarks:

(1) For the stationary continuous-time GAR(1) process

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t,t'), \qquad 0 \le t' < t,$$

when $t - t' = h$ is very small, it will become a stochastic difference equation

$$X(t) \stackrel{d}{=} \left(1 - \mu(t-t')\right)_K \circledast X(t') + \Delta\epsilon(h),$$

where $\Delta\epsilon(h) = \epsilon(t) - \epsilon(t')$ is the increment of the innovation process $\{\epsilon(t); t \ge 0\}$. Since given $X(t') = x$, the pgf or LT of $(1 - \mu(t-t'))_K \circledast x$ is $G_K^x(s; 1-\mu(t-t'))$ or $\phi_K^x(s; 1-\mu(t-t'))$, by Properties 3.1 and 3.3, it follows that

$$G_K^x(s; 1-\mu(t-t')) \longrightarrow s^x, \quad \text{or} \quad \phi_K^x(s; 1-\mu(t-t')) \longrightarrow e^{-xs}, \qquad \text{as } t - t' \to 0.$$

This shows that

$$\left(1 - \mu(t-t')\right)_K \circledast x \xrightarrow{P} x, \qquad \text{as } t - t' \to 0.$$

Hence, as $t - t' \to 0$,

$$\left(1 - \mu(t-t')\right)_K \circledast X(t') - X(t') \xrightarrow{P} 0.$$

On the other hand, the innovation process is a Lévy process under stationarity for the generalized Ornstein-Uhlenbeck SDE. By the definition of Lévy process, we know

$$\Delta\epsilon(h) \xrightarrow{P} 0, \qquad \text{as } t - t' \to 0.$$

Thus,

$$X(t) - X(t') \xrightarrow{P} 0, \qquad \text{as } t - t' \to 0.$$

Therefore, the stationary continuous-time GAR(1) process is stochastically continuous.

(2) The stationary continuous-time GAR(1) process can be constructed by taking limit of a series of discrete-time processes:

$$X^{(k)} = \left\{ X(0),\ X\left(\frac{1}{2^k}\right),\ X\left(\frac{2}{2^k}\right),\ \ldots,\ X\left(\frac{j}{2^k}\right),\ \ldots \right\}, \qquad k = 0, 1, 2, 3, \ldots.$$

These discrete-time processes can be constructed by the trivariate distributions in Theorems 9.1.1 and 9.1.2. By the consistency of the trivariate distributions, the embedding of $X^{(k-1)}$ in $X^{(k)}$ is consistent. With the stochastic continuity, the limit exists in distribution:

$$\lim_{k\to\infty} X^{(k)} \xrightarrow{\mathrm{L}} X.$$

This limiting process is indexed by $t \in [0, \infty)$.

9.2 Conditional properties

In this section, we study some conditional properties of a continuous-time GAR(1) process. These properties include the conditional pmf or pdf, conditional mean and conditional variance. They are of particular interest because they are needed for statistical inference such as parameter estimation.

Recall that a continuous-time GAR(1) process has the stochastic representation

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1, t_2), \qquad t_1 < t_2,$$

where the two summands on the right hand side are independent. Therefore, given $X(t_1) = x$, $X(t_2)$ has representation

$$[X(t_2)|X(t_1) = x] \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast x + E(t_1, t_2).$$

Note that $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ is usually a rv. Only in special cases like K being from **P1**, i.e., the corresponding extended-thinning operation becomes the constant multiplier, does it degenerate to a constant. This conditional representation leads to the pgf or LT or cf of $X(t_2)$ given $X(t_1) = x$

$$\begin{aligned}
G_{X(t_2)|X(t_1)=x}(s) &= G_K^x\left(s; e^{-\mu(t_2-t_1)}\right) G_{E(t_1,t_2)}(s), \\
\phi_{X(t_2)|X(t_1)=x}(s) &= \phi_K^x\left(s; e^{-\mu(t_2-t_1)}\right) \phi_{E(t_1,t_2)}(s), \\
\varphi_{X(t_2)|X(t_1)=x}(s) &= \exp\left\{ixe^{-\mu(t_2-t_1)}s\right\} \varphi_{E(t_1,t_2)}(s).
\end{aligned}$$

The pgf, or LT, or cf of K and $E(t_1, t_2)$, and even the margin $X(t_1)$ and $X(t_2)$ are specified by the continuous-time GAR(1) process. In principle, the conditional pgf, or LT, or cf will determine the conditional pmf or pdf, conditional mean and conditional variance.

We first consider the conditional mean and conditional variance; assuming they exist:

$$\begin{aligned}
\mathbf{E}[X(t_2) \mid X(t_1) = x] &= \mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x\right] + \mathbf{E}\left[E(t_1, t_2)\right], \\
\mathbf{Var}[X(t_2) \mid X(t_1) = x] &= \mathbf{Var}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x\right] + \mathbf{Var}\left[E(t_1, t_2)\right].
\end{aligned}$$

The mean and variance of the cumulative innovation $E(t_1, t_2)$ can be obtained from its pgf, or LT, or cf by the general formula

$$\mathbf{E}(Y) = \begin{cases} G'_Y(1), \\ -\phi'_Y(0), \\ -i\varphi'_Y(0), \end{cases} \quad \text{and} \quad \mathbf{Var}(Y) = \begin{cases} G''_Y(1) + G'_Y(1) - (G'_Y(1))^2, \\ \phi''_Y(0) - (\phi'_Y(0))^2, \\ -\varphi''_Y(0) + (\varphi'_Y(0))^2. \end{cases}$$

They are independent of x. Now we investigate the mean and variance of $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$. Since

$$\begin{aligned}
\frac{\partial}{\partial s} G_K^x\left(s; e^{-\mu(t_2-t_1)}\right) &= xG_K^{x-1}\left(s; e^{-\mu(t_2-t_1)}\right) \frac{\partial}{\partial s} G_K\left(s; e^{-\mu(t_2-t_1)}\right), \\
\frac{\partial}{\partial s} \phi_K^x\left(s; e^{-\mu(t_2-t_1)}\right) &= x\phi_K^{x-1}\left(s; e^{-\mu(t_2-t_1)}\right) \frac{\partial}{\partial s} \phi_K\left(s; e^{-\mu(t_2-t_1)}\right),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial s}\exp\left\{ixe^{-\mu(t_2-t_1)}s\right\} &= ixe^{-\mu(t_2-t_1)}\exp\left\{ixe^{-\mu(t_2-t_1)}s\right\}, \\
\frac{\partial^2}{\partial s^2}G_K^x\left(s;e^{-\mu(t_2-t_1)}\right) &= x(x-1)G_K^{x-2}\left(s;e^{-\mu(t_2-t_1)}\right)\left[\frac{\partial}{\partial s}G_K\left(s;e^{-\mu(t_2-t_1)}\right)\right]^2 \\
&\quad +xG_K^{x-1}\left(s;e^{-\mu(t_2-t_1)}\right)\frac{\partial^2}{\partial s^2}G_K\left(s;e^{-\mu(t_2-t_1)}\right), \\
\frac{\partial^2}{\partial s^2}\phi_K^x\left(s;e^{-\mu(t_2-t_1)}\right) &= x(x-1)\phi_K^{x-2}\left(s;e^{-\mu(t_2-t_1)}\right)\left[\frac{\partial}{\partial s}\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)\right]^2 \\
&\quad +x\phi_K^{x-1}\left(s;e^{-\mu(t_2-t_1)}\right)\frac{\partial^2}{\partial s^2}\phi_K\left(s;e^{-\mu(t_2-t_1)}\right), \\
\frac{\partial^2}{\partial s^2}\exp\left\{ixe^{-\mu(t_2-t_1)}s\right\} &= -x^2e^{-2\mu(t_2-t_1)}\exp\left\{ixe^{-\mu(t_2-t_1)}s\right\},
\end{aligned}$$

and

$$G_K\left(1;e^{-\mu(t_2-t_1)}\right)=\phi_K\left(0;e^{-\mu(t_2-t_1)}\right)=\exp\left\{ixe^{-\mu(t_2-t_1)}\times 0\right\}=1,$$

We have

$$\begin{aligned}
\mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K\circledast x\right] &= \begin{cases}
\frac{\partial}{\partial s}G_K^x\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=1} = x\frac{\partial}{\partial s}G_K\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=1} \\
\frac{\partial}{\partial s}\phi_K^x\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=0} = x\frac{\partial}{\partial s}\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=0} \\
\frac{\partial}{\partial s}\varphi_K^x\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=0} = x\frac{\partial}{\partial s}\varphi_K\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=0}
\end{cases} \\
&= x\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right].
\end{aligned}$$

Here $K\left(e^{-\mu(t_2-t_1)}\right)$ denotes the self-generalized rv K associated with parameter $e^{-\mu(t_2-t_1)}$ (see Chapter 3). For the variance, we calculate it from the pgf, or LT, or cf respectively:

$$\begin{aligned}
&\frac{\partial^2}{\partial s^2}G_K^x\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=1}+\frac{\partial}{\partial s}G_K^x\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=1}-\left[\frac{\partial}{\partial s}G_K^x\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=1}\right]^2 \\
&=x(x-1)\left[\frac{\partial G_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s}\Big|_{s=1}\right]^2+x\frac{\partial^2 G_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s^2}\Big|_{s=1} \\
&\quad +x\frac{\partial G_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s}\Big|_{s=1}-x^2\left[\frac{\partial G_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s}\Big|_{s=1}\right]^2 \\
&=x\frac{\partial^2 G_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s^2}\Big|_{s=1}+x\frac{\partial G_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s}\Big|_{s=1}-x\left[\frac{\partial G_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s}\Big|_{s=1}\right]^2 \\
&=x\mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right],
\end{aligned}$$

$$
\begin{aligned}
&\frac{\partial^2}{\partial s^2}\phi_K^x\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=0}-\left[\frac{\partial}{\partial s}\phi_K^x\left(s;e^{-\mu(t_2-t_1)}\right)\Big|_{s=0}\right]^2\\
&= x(x-1)\left[\frac{\partial\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s}\Big|_{s=0}\right]^2+x\frac{\partial^2\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s^2}\Big|_{s=0}-x^2\left[\frac{\partial\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s}\Big|_{s=0}\right]^2\\
&= x\frac{\partial^2\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s^2}\Big|_{s=0}-x\frac{\partial\phi_K\left(s;e^{-\mu(t_2-t_1)}\right)}{\partial s}\Big|_{s=0}\\
&= x\mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{\partial^2}{\partial s^2}\exp\left\{ixe^{-\mu(t_2-t_1)}s\right\}\Big|_{s=0}+\left[\frac{\partial}{\partial s}\exp\left\{ixe^{-\mu(t_2-t_1)}s\right\}\Big|_{s=0}\right]^2\\
&= x^2e^{-2\mu(t_2-t_1)}+\left[ixe^{-\mu(t_2-t_1)}\right]^2\\
&= x\mathbf{Var}\left[e^{-\mu(t_2-t_1)}\right].
\end{aligned}
$$

In summary,

$$
\begin{aligned}
\mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K\circledast x\right] &= x\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right],\\
\mathbf{Var}\left[\left(e^{-\mu(t_2-t_1)}\right)_K\circledast x\right] &= x\mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right].
\end{aligned}
$$

These indicate that the conditional mean and variance of $\left(e^{-\mu(t_2-t_1)}\right)_K\circledast x$ are proportional to the last observation $X(t_1)=x$; the larger the value of x, the larger the conditional mean and variance of the dependence term. Table 9.1 summarizes the mean and variance of self-generalized rv's $K(\alpha)$ discussed in Chapter 3.

Consider a self-generalized distribution with finite mean. By Theorem 3.2.1, we have

$$
\mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K\circledast x\right]=x\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]=x\left(e^{-\mu(t_2-t_1)}\right)^r=xe^{-r\mu(t_2-t_1)},\qquad r>0.
$$

This shows that as the time difference t_2-t_1 increases to infinity, the mean of the dependent term decreases to zero. For the pair $\left(x,\left(e^{-\mu(t_2-t_1)}\right)_K\circledast x\right)$, or more generally $\left(X,\left(e^{-\mu(t_2-t_1)}\right)_K\circledast X\right)$, if K is a self-generalized rv from **P1**, then it becomes $\left(x,e^{-\mu(t_2-t_1)}x\right)$ or $\left(X,e^{-\mu(t_2-t_1)}X\right)$, a straight line through the origin with slope $e^{-\mu(t_2-t_1)}$, leading to a singular bivariate distribution in $\Re^2$. If K is a self-generalized rv other than from **P1** and $\mathbf{E}[K(\alpha)]=\alpha$, then the expectation line is the

Table 9.1: *Mean and variance of non-negative integer and positive self-generalized random variable $K(\alpha)$.*

$K(\alpha)$	Mean	Variance
I1	α	$\alpha(1-\alpha)$
I2	α	$\alpha(1-\alpha)(1+\gamma)/(1-\gamma)$
I3	∞	NA
I4	α	$\alpha(1-\alpha)e^{\theta}$
I5	α^{θ}	when $\theta = 1, \alpha(1-\alpha)(1+\gamma)/(1-\gamma)$ and ∞ $(\theta > 1)$
P1	α	0
P2	α	$2\alpha(1-\alpha)\gamma/(1-\gamma)$
P3	∞	NA
P4	α	$\alpha(1-\alpha)\left(e^{\theta}-1\right)$
P5	α^{θ}	when $\theta = 1, 2\alpha(1-\alpha)\gamma/(1-\gamma)$ and ∞ $(\theta > 1)$

same as in the case of **P1**. However, the second argument $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ or $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X$ is no longer a constant, but a random variable. This random variable will fluctuate around the expectation line $y = e^{-\mu(t_2-t_1)}x$, and form a cone shape, namely as x or X gets larger, the variation of $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ or $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X$ is proportionally larger.

For a stationary continuous-time GAR(1) process with marginal mean A and marginal variance V, we may find the mean and variance of the cumulative innovation $E(t_1, t_2)$, which further results in the the mean and variance of $X(t_2)$ given $X(t_1) = x$. This is to take advantage of the independence of the two terms in the right hand side of the stochastic representation

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1, t_2),$$

or

$$[X(t_2)|X(t_1) = x] \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast x + E(t_1, t_2).$$

From the first representation, applying the previous results on the dependence term, we obtain

$$\begin{aligned}
\mathbf{E}[X(t_2)] &= \mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\right] + \mathbf{E}[E(t_1, t_2)] \\
&= \mathbf{E}[X(t_1)]\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right] + \mathbf{E}[E(t_1, t_2)], \\
\mathbf{Var}[X(t_2)] &= \mathbf{Var}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\right] + \mathbf{Var}[E(t_1, t_2)]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{Var}\left(\mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\Big|X(t_1)\right]\right) \\
&\quad +\mathbf{E}\left(\mathbf{Var}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\Big|X(t_1)\right]\right) + \mathbf{Var}[E(t_1,t_2)] \\
&= \mathbf{Var}\left(X(t_1)\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\right) + \mathbf{E}\left(X(t_1)\mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\right) \\
&\quad +\mathbf{Var}[E(t_1,t_2)] \\
&= \mathbf{Var}[X(t_1)]\mathbf{E}^2\left[K\left(e^{-\mu(t_2-t_1)}\right)\right] + \mathbf{E}[X(t_1)]\mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right] \\
&\quad +\mathbf{Var}[E(t_1,t_2)].
\end{aligned}
$$

Thus

$$
\begin{aligned}
A &= A\cdot\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right] + \mathbf{E}[E(t_1,t_2)], \\
V &= V\cdot\mathbf{E}^2\left[K\left(e^{-\mu(t_2-t_1)}\right)\right] + A\cdot\mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right] + \mathbf{Var}[E(t_1,t_2)],
\end{aligned}
$$

which lead to

$$
\begin{aligned}
\mathbf{E}[E(t_1,t_2)] &= A\left(1-\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\right), \\
\mathbf{Var}[E(t_1,t_2)] &= V\cdot\left(1-\mathbf{E}^2\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\right) - A\cdot\mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right].
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
\mathbf{E}[X(t_2)\mid X(t_1)=x] &= \mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x\right] + \mathbf{E}\left[E(t_1,t_2)\right] \qquad (9.2.1) \\
&= A + (x-A)\cdot\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right], \\
\mathbf{Var}[X(t_2)\mid X(t_1)=x] &= \mathbf{Var}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x\right] + \mathbf{Var}\left[E(t_1,t_2)\right] \qquad (9.2.2) \\
&= V\cdot\left(1-\mathbf{E}^2\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\right) + (x-A)\cdot\mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right].
\end{aligned}
$$

This shows that the conditional mean and conditional variance only depend on the marginal mean and marginal variance specified in steady state, as well as the mean and variance of the self-generalized rv K in the dependent term. Thus, when we fix the stationary distribution of the process, the cumulative innovation seems to be dummy; it looks to have no influence on the conditional mean and conditional variance no matter whether the process is in steady state or not. Note that this approach can be extended to the non-stationary case where the stochastic representation is

$$
X(t_2) \stackrel{d}{=} \left(e^{-\int_{t_1}^{t_2}\mu(t)dt}\right)_K \circledast X(t_1) + E(t_1,t_2),
$$

or

$$X(t_2) \stackrel{d}{=} \left(e^{-\int_{t_1}^{t_2} \mu(t)dt}\right)_K \circledast x + E(t_1, t_2).$$

Assume that the margin $X(t)$ has the mean $A(t)$ and variance $V(t)$. Then it will follow that

$$\begin{aligned}
\mathbf{E}[X(t_2) \mid X(t_1) = x] &= \mathbf{E}\left[\left(e^{-\int_{t_1}^{t_2} \mu(t)dt}\right)_K \circledast x\right] + \mathbf{E}\left[E(t_1, t_2)\right] \\
&= A(t_2) + [x - A(t_1)] \cdot \mathbf{E}\left[K\left(e^{-\int_{t_1}^{t_2} \mu(t)dt}\right)\right], \\
\mathbf{Var}[X(t_2) \mid X(t_1) = x] &= \mathbf{Var}\left[\left(e^{-\int_{t_1}^{t_2} \mu(t)dt}\right)_K \circledast x\right] + \mathbf{Var}\left[E(t_1, t_2)\right] \\
&= V(t_2) - V(t_1) \cdot \mathbf{E}^2\left[K\left(e^{-\int_{t_1}^{t_2} \mu(t)dt}\right)\right] \\
&\quad + [x - A(t_1)] \cdot \mathbf{Var}\left[K\left(e^{-\int_{t_1}^{t_2} \mu(t)dt}\right)\right].
\end{aligned}$$

Next we turn to the conditional pmf and pdf. It is quite challenging to obtain them from the conditional pgf, or conditional LT, or conditional cf. Without loss of generality, we consider the stationary continuous-time GAR(1) process. Thus, given $X(t_1) = x$, $X(t_2)$ can be decomposed as the sum of two independent terms, namely the convolution of $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ and $E(t_1, t_2)$. The supports of $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ and $E(t_1, t_2)$ are of interest, because they will affect the expression form of conditional probability of $X(t_2)$ given $X(t_1) = x$.

(1) If K is a non-negative integer self-generalized rv, then the support of $\left(e^{-\mu(t_2-t_1)}\right)_K$ is either $\{0, 1\}$ (from **I1**) or $\mathcal{N}_0 = \{0, 1, 2, \ldots\}$ (other than **I1**). Hence, the support of $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ is either $\{0, 1, \ldots, x\}$ or $\mathcal{N}_0$. However, the support of cumulative innovation $E(t_1, t_2)$ is always $\mathcal{N}_0$.

In this situation, the general expression of conditional pmf is that for any $y \in \mathcal{N}_0$,

$$\Pr[X(t_2) = y \mid X(t_1) = x] = \Pr\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x + E(t_1, t_2) = y\right]$$

$$= \begin{cases} \sum_{i=0}^{\min(x,y)} \Pr\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x = i\right] \times \Pr[E(t_1, t_2) = y - i], & K \text{ from } \mathbf{I1}, \\ \sum_{i=0}^{y} \Pr\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x = i\right] \times \Pr[E(t_1, t_2) = y - i], & \text{otherwise.} \end{cases}$$

(2) If K is a self-generalized rv from **P1**, then $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ degenerates to the point $xe^{-\mu(t_2-t_1)}$, which could be positive or real depending on the data type of x. The cumu-

lative innovation $E(t_1, t_2)$ can be non-negative, positive or real. It could have non-zero mass on the point zero.

In this situation, the conditional distribution of $X(t_2)$ could have non-zero mass on the point $xe^{-\mu(t_2-t_1)}$, i.e.,

$$\Pr\left[X(t_2) = xe^{-\mu(t_2-t_1)}\right] = \Pr[E(t_1, t_2) = 0],$$

and conditional pdf

$$f_{X(t_2)|X(t_1)=x}(y) = f_{E(t_1,t_2)}\left(y - xe^{-\mu(t_2-t_1)}\right), \qquad y \neq xe^{-\mu(t_2-t_1)}.$$

(3) If K is a positive self-generalized rv from a distribution family other than **P1**, then the support of $\left(e^{-\mu(t_2-t_1)}\right)_K$ is either non-negative $\Re_0$ (with non-zero mass on point zero) or positive $\Re_+$ (without non-zero mass on point zero). This leads to the support of $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ being either non-negative $\Re_0$ or positive $\Re_+$. The support of $E(t_1, t_2)$ can be either $\Re_0$ or $\Re_+$ too.

In this situation, the conditional distribution of $X(t_2)$ could have non-zero mass on point zero if both have non-zero masses on 0:

$$\Pr\left[X(t_2) = 0 \mid X(t_1) = x\right] = \Pr\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x = 0\right] \Pr[E(t_1, t_2) = 0],$$

and have pdf on $y > 0$:

$$\begin{aligned} f_{X(t_2)|X(t_1)=x}(y) &= \int_0^y f_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}(z) \cdot f_{E(t_1,t_2)}(y-z)dz \\ &\quad + \Pr\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x = 0\right] \times f_{E(t_1,t_2)}(y) \\ &\quad + f_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}(y) \times \Pr[E(t_1, t_2) = 0]. \end{aligned} \tag{9.2.3}$$

The dependence term $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ is associated with the self-generalized rv K. When K is a non-negative integer rv, it is the sum of x iid rv's with pgf $G_K\left(s; e^{-\mu(t_2-t_1)}\right)$. When K is a positive rv (from distribution family other than **P1**), since $\phi_K\left(s; e^{-\mu(t_2-t_1)}\right)$ is of exponential form (see Section 8.2.2),

$$\phi_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}(s) = \left[\phi_K\left(s; e^{-\mu(t_2-t_1)}\right)\right]^x$$

is certainly of exponential form too, and consequently, $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ and K are in the same family. When K is from **P1**, it is trivial case. Essentially, this dependence term can be linked to the associated self-generalized distribution, which can help us to probe the pmf or pdf of this term.

Comparing with known self-generalized distributions, the cumulative innovation varies very much among different continuous-time GAR(1) processes. It is too general to be discussed in a simple way. Hence, we only focus on the self-generalized rv K to investigate its pmf or pdf.

First, we consider K being a non-negative integer self-generalized rv. Denote

$$\left[G_K\left(s; e^{-\mu(t_2-t_1)}\right)\right]^m = \sum_{i=0}^{\infty} p_i(m)s^i, \qquad m \geq 1. \tag{9.2.4}$$

When $m = 1$, $\{p_0(1), p_1(1), \ldots, p_i(1), \ldots\}$ is the pmf of K. By the property of convolution, we have the following recursive formula

$$p_i(m) = \sum_{j=0}^{i} p_j(m-1)p_{i-j}(1), \qquad m \geq 2.$$

This recursive formula is specially useful when we resort to computer to do the calculations. In general, we can not easily find the closed form of $p_i(m)$. However, when K from **I1**, the Bernoulli distribution family, the dependence term $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x = e^{-\mu(t_2-t_1)} * x$ is then distributed in Binomial$\left(x, e^{-\mu(t_2-t_1)}\right)$, which leads to

$$p_i(m) = \binom{m}{i} e^{-i\mu(t_2-t_1)}\left[1 - e^{-\mu(t_2-t_1)}\right]^{m-i}, \qquad \text{for } 0 \leq i \leq m \text{ and } m \geq 1.$$

For the pmf of other self-generalized distribution families, one can refer to Section 3.1.1.

Secondly, consider a positive self-generalized rv K. Now finding the pdf from the LT could be a tough task. In many cases, they are still open questions. Even for those discussed in Section 3.1.2, we don't know all of the pgf's. What we know is that **P2** is the compound Poisson with exponential distribution. Here we briefly give the pdf of the dependence term $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ when K is from **P2**. Since

$$\begin{aligned}\phi_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}(s) &= \left[\phi_K\left(s; e^{-\mu(t_2-t_1)}\right)\right]^x = \left[\exp\left\{-\frac{e^{-\mu(t_2-t_1)}(1-\gamma)s}{(1-\gamma)+(1-e^{-\mu(t_2-t_1)})\gamma s}\right\}\right]^x \\ &= \exp\left\{x \cdot \frac{e^{-\mu(t_2-t_1)}(1-\gamma)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\left[\left(1 + \frac{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}{1-\gamma}s\right)^{-1} - 1\right]\right\},\end{aligned}$$

the distribution of $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ is the compound Poisson with an exponential distribution. The means of the Poisson and exponential distributions are $\frac{xe^{-\mu(t_2-t_1)}(1-\gamma)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}$ and $\frac{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}{1-\gamma}$ respectively. This leads to the stochastic representation

$$\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x \;\stackrel{d}{=}\; \sum_{i=0}^{N} Y_i \stackrel{d}{=} Y, \tag{9.2.5}$$

where $N \sim \text{Poisson}\left(x\frac{e^{-\mu(t_2-t_1)}(1-\gamma)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right)$, $Y_0 = 0$ and $Y_i \stackrel{i.i.d.}{\sim} \text{Exponential}\left(\frac{1-\gamma}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right)$ $(i \geq 1)$. When $N = 0$, $Y = Y_0 = 0$, thus

$$\Pr[Y=0] \;=\; \Pr[N=0] \;=\; \exp\left\{-\frac{xe^{-\mu(t_2-t_1)}(1-\gamma)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right\} \;>\; 0;$$

that is, $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ has non-zero mass on point 0. If N takes value $n \geq 1$, the convolution $\sum_{i=0}^{n} Y_i \sim \text{Gamma}\left(n, \frac{1-\gamma}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right)$. In this situation, $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ has zero mass on point 0. For $n \geq 1$ and $y > 0$, the joint density of (Y, N) is

$$\begin{aligned}
f_{(Y,N)}(y,n) &= \frac{1}{\Gamma(n)}\left(\frac{1-\gamma}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right)^n y^{n-1}\exp\left\{-\frac{y(1-\gamma)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right\} \\
&\quad \times\frac{1}{n!}\left(\frac{xe^{-\mu(t_2-t_1)}(1-\gamma)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right)^n \exp\left\{-\frac{xe^{-\mu(t_2-t_1)}(1-\gamma)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right\} \\
&= \frac{1}{(n-1)!n!}\left(\frac{xe^{-\mu(t_2-t_1)}(1-\gamma)^2}{\left(1-e^{-\mu(t_2-t_1)}\right)^2\gamma^2}\right)^n y^{n-1}\exp\left\{-\frac{(1-\gamma)\left(xe^{-\mu(t_2-t_1)}+y\right)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right\},
\end{aligned}$$

which leads to the marginal pdf of Y as

$$\begin{aligned}
f_Y(y) &= \sum_{n=1}^{\infty} f_{(Y,N)}(y,n) \\
&= \sum_{n=1}^{\infty}\frac{1}{(n-1)!n!}\left(\frac{xe^{-\mu(t_2-t_1)}(1-\gamma)^2}{\left(1-e^{-\mu(t_2-t_1)}\right)^2\gamma^2}\right)^n y^{n-1}\exp\left\{-\frac{(1-\gamma)\left(xe^{-\mu(t_2-t_1)}+y\right)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right\} \\
&= \exp\left\{-\frac{(1-\gamma)\left(xe^{-\mu(t_2-t_1)}+y\right)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right\}\times\sum_{n=1}^{\infty}\frac{1}{(n-1)!n!}\left(\frac{xe^{-\mu(t_2-t_1)}(1-\gamma)^2}{\left(1-e^{-\mu(t_2-t_1)}\right)^2\gamma^2}\right)^n y^{n-1} \\
&= f_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}(y)
\end{aligned} \tag{9.2.6}$$

for $y > 0$. This calculation is tractable using the computer.

Finding the stochastic representation for the dependence term and cumulative innovation term is not only useful in obtaining their pmf's or pdf's, but also in simulation of the continuous-time GAR(1) process. Hence, they are of particular interest for many researchers. To conclude this section, we look at some specific examples.

Example 9.1 *Consider the continuous-time GAR(1) process obtained in Example 5.1:*

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} * X(t_1) + E(t_1,t_2),$$

where $E(t_1,t_2) \sim \text{Poisson}\left(\frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right]\right)$. *This process has stationary distribution Poisson*$\left(\frac{\lambda}{\mu}\right)$. *Since the operation is binomial-thinning, given* $X(t_1)=x$, $e^{-\mu(t_2-t_1)} * x \sim \text{Binomial}\left(x, e^{-\mu(t_2-t_1)}\right)$. *In this situation, we have*

$$\begin{aligned}
\mathbf{E}\left[X(t_2) \mid X(t_1)=x\right] &= xe^{-\mu(t_2-t_1)} + \frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right] \\
&= \frac{\lambda}{\mu} + \left(x-\frac{\lambda}{\mu}\right)e^{-\mu(t_2-t_1)}, \\
\mathbf{Var}\left[X(t_2) \mid X(t_1)=x\right] &= xe^{-\mu(t_2-t_1)}\left(1-e^{-\mu(t_2-t_1)}\right) + \frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right] \\
&= \frac{\lambda}{\mu} + \left(x-\frac{\lambda}{\mu}\right)e^{-\mu(t_2-t_1)} - xe^{-2\mu(t_2-t_1)},
\end{aligned}$$

and the conditional pmf is

$$\begin{aligned}
\Pr(X(t_2)=y \mid X(t_1)=x) &= \sum_{i=0}^{\min(x,y)} \binom{x}{i} e^{-i\mu(t_2-t_1)}\left(1-e^{-\mu(t_2-t_1)}\right)^{x-i} \\
&\quad \times \frac{1}{(y-i)!}\left(\frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right]\right)^{y-i} \exp\left\{-\frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right]\right\} \\
&= \sum_{i=0}^{\min(x,y)} \frac{1}{(y-i)!}\binom{x}{i}\left(\frac{\lambda}{\mu}\right)^{y-i} e^{-i\mu(t_2-t_1)}\left(1-e^{-\mu(t_2-t_1)}\right)^{x+y-2i} \\
&\quad \times \exp\left\{-\frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right]\right\} \\
&= \exp\left\{-\frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right]\right\}\left(1-e^{-\mu(t_2-t_1)}\right)^{x+y} \\
&\quad \times \sum_{i=0}^{\min(x,y)} \frac{1}{(y-i)!}\binom{x}{i}\left(\frac{\lambda}{\mu}\right)^{y-i}\left(\frac{e^{-\mu(t_2-t_1)}}{\left(1-e^{-\mu(t_2-t_1)}\right)^2}\right)^{i}
\end{aligned}$$

for $y \geq 0$.

Example 9.2 *Consider the continuous-time GAR(1) process obtained in Example 5.4:*

$$X(t_2) = \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1,t_2),$$

where K is from **I2** *and $E(t_1,t_2) \sim NB\left(\frac{\lambda(1-\gamma)}{\mu\gamma}, \frac{\gamma\left(1-e^{-\mu(t_2-t_1)}\right)}{1-\gamma e^{-\mu(t_2-t_1)}}\right)$. This process has stationary distribution $NB\left(\frac{\lambda(1-\gamma)}{\mu\gamma}, \gamma\right)$. The mean and variance of $E(t_1,t_2)$ are*

$$\frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right] \quad \textit{and} \quad \frac{\lambda}{\mu(1-\gamma)}\left[1-e^{-\mu(t_2-t_1)}\right]\left[1-\gamma e^{-\mu(t_2-t_1)}\right]$$

respectively. By Table 9.1, we have

$$\begin{aligned}
\mathbf{E}[X(t_2) \mid X(t_1)=x] &= xe^{-\mu(t_2-t_1)} + \frac{\lambda}{\mu}\left[1-e^{-\mu(t_2-t_1)}\right] = \frac{\lambda}{\mu} + \left(x-\frac{\lambda}{\mu}\right)e^{-\mu(t_2-t_1)}, \\
\mathbf{Var}[X(t_2) \mid X(t_1)=x] &= xe^{-\mu(t_2-t_1)}\left(1-e^{-\mu(t_2-t_1)}\right)\frac{1+\gamma}{1-\gamma} \\
&\quad + \frac{\lambda}{\mu(1-\gamma)}\left[1-e^{-\mu(t_2-t_1)}\right]\left[1-\gamma e^{-\mu(t_2-t_1)}\right] \\
&= \frac{\lambda}{\mu(1-\gamma)} + \frac{1+\gamma}{1-\gamma}\left(x-\frac{\lambda}{\mu}\right)e^{-\mu(t_2-t_1)} + \frac{\lambda\gamma/\mu - x(1+\gamma)}{1-\gamma}e^{-2\mu(t_2-t_1)}.
\end{aligned}$$

However, in this situation, we do not have explicit expression of conditional pmf, only a recursive form:

$$\begin{aligned}
&\Pr(X(t_2)=y \mid X(t_1)=x) \\
&= \sum_{i=0}^{y} p_i(y)\binom{\frac{\lambda(1-\gamma)}{\mu\gamma}+y-i-1}{y-i}\left(\frac{1-\gamma}{1-\gamma e^{-\mu(t_2-t_1)}}\right)^{\frac{\lambda(1-\gamma)}{\mu\gamma}}\left(\frac{\gamma\left(1-e^{-\mu(t_2-t_1)}\right)}{1-\gamma e^{-\mu(t_2-t_1)}}\right)^{y-i} \\
&= \left(\frac{1-\gamma}{1-\gamma e^{-\mu(t_2-t_1)}}\right)^{\frac{\lambda(1-\gamma)}{\mu\gamma}} \times \sum_{i=0}^{y} p_i(y)\binom{\frac{\lambda(1-\gamma)}{\mu\gamma}+y-i-1}{y-i}\left(\frac{\gamma\left(1-e^{-\mu(t_2-t_1)}\right)}{1-\gamma e^{-\mu(t_2-t_1)}}\right)^{y-i}
\end{aligned}$$

for $y \geq 0$, where $p_i(y)$ is defined in (9.2.4).

Example 9.3 *Consider the continuous-time GAR(1) process obtained in Example 5.12:*

$$X(t_2) = e^{-\mu(t_2-t_1)} \bullet X(t_1) + E(t_1,t_2),$$

where $E(t_1,t_2)$ is a positive rv with LT $\phi_{E(t_1,t_2)}(s) = \left(e^{-\mu(t_2-t_1)} + \left[1-e^{-\mu(t_2-t_1)}\right]\frac{\beta}{\beta+s}\right)^{\theta/\mu}$. This process will have stationary distribution $Gamma(\theta/\mu, \beta)$. By taking derivatives, we have

$$\phi'_{E(t_1,t_2)}(0) = -\frac{\theta}{\beta\mu}\left[1-e^{-\mu(t_2-t_1)}\right],$$

$$\phi''_{E(t_1,t_2)}(0) = \frac{2\theta}{\beta^2\mu}\left[1-e^{-\mu(t_2-t_1)}\right] + \frac{\theta}{\beta^2\mu}\left(\frac{\theta}{\mu}-1\right)\left[1-e^{-\mu(t_2-t_1)}\right]^2,$$

which lead to the mean and variance of $E(t_1,t_2)$

$$\mathbf{E}[E(t_1,t_2)] = \frac{\theta}{\beta\mu}\left[1-e^{-\mu(t_2-t_1)}\right] \quad \textit{and} \quad \mathbf{Var}[E(t_1,t_2)] = \frac{\theta}{\beta^2\mu}\left[1-e^{-2\mu(t_2-t_1)}\right].$$

Hence, given $X(t_1)=x$, *it follows that*

$$\begin{aligned}
\mathbf{E}[X(t_2)\mid X(t_1)=x] &= xe^{-\mu(t_2-t_1)} + \frac{\theta}{\beta\mu}\left[1-e^{-\mu(t_2-t_1)}\right] = \frac{\theta}{\beta\mu} + \left(x-\frac{\theta}{\beta\mu}\right)e^{-\mu(t_2-t_1)},\\
\mathbf{Var}[X(t_2)\mid X(t_1)=x] &= 0 + \frac{\theta}{\beta^2\mu}\left[1-e^{-2\mu(t_2-t_1)}\right] = \frac{\theta}{\beta^2\mu}\left[1-e^{-2\mu(t_2-t_1)}\right].
\end{aligned}$$

Walker [2000] proposed a better representation for a rv with the same LT as $E(t_1,t_2)$. *With that idea, we can write down the stochastic representation*

$$E(t_1,t_2) \stackrel{d}{=} \sum_{i=0}^{N} Y_i, \qquad Y_0 = 0,\ Y_i \stackrel{i.i.d.}{\sim} \text{Gamma}\left(1, \beta e^{\mu(t_2-t_1)}\right) = \text{Exponential}\left(\beta e^{\mu(t_2-t_1)}\right),$$

where N *is a rv resulting from a Gamma mixture of Poisson:*

$$N|Z=z \sim \text{Poisson}\left(z\left[e^{\mu(t_2-t_1)}-1\right]\right), \qquad Z \sim \text{Gamma}(\theta/\mu, 1).$$

This can be verified by algebra

$$\begin{aligned}
\phi_{\sum_{i=0}^N Y_i}(s) &= \mathbf{E}\left\{e^{-s\sum_{i=0}^N Y_i}\right\} = \mathbf{E}\left\{\mathbf{E}\left[e^{-s\sum_{i=0}^N Y_i}\Big|N\right]\right\} = \mathbf{E}\left\{\left(\frac{\beta e^{\mu(t_2-t_1)}}{\beta e^{\mu(t_2-t_1)}+s}\right)^N\right\}\\
&= \mathbf{E}\left\{\mathbf{E}\left[\left(\frac{\beta e^{\mu(t_2-t_1)}}{\beta e^{\mu(t_2-t_1)}+s}\right)^N \Big| Z\right]\right\}\\
&= \mathbf{E}\left\{\exp\left[Z\left(e^{\mu(t_2-t_1)}-1\right)\left(\frac{\beta e^{\mu(t_2-t_1)}}{\beta e^{\mu(t_2-t_1)}+s}-1\right)\right]\right\}\\
&= \mathbf{E}\left\{\exp\left[-\frac{Z\left(e^{\mu(t_2-t_1)}-1\right)s}{\beta e^{\mu(t_2-t_1)}+s}\right]\right\} = \phi_Z\left(\frac{\left(e^{\mu(t_2-t_1)}-1\right)s}{\beta e^{\mu(t_2-t_1)}+s}\right)\\
&= \left(\frac{1}{1+\frac{\left(e^{\mu(t_2-t_1)}-1\right)s}{\beta e^{\mu(t_2-t_1)}+s}}\right)^{\theta/\mu} = \left(\frac{\beta e^{\mu(t_2-t_1)}+s}{\beta e^{\mu(t_2-t_1)}+e^{\mu(t_2-t_1)}s}\right)^{\theta/\mu} = \left(\frac{\beta+e^{-\mu(t_2-t_1)}s}{\beta+s}\right)^{\theta/\mu}\\
&= \left(e^{-\mu(t_2-t_1)}+\left[1-e^{-\mu(t_2-t_1)}\right]\frac{\beta}{\beta+s}\right)^{\theta/\mu} = \phi_{E(t_1,t_2)}(s).
\end{aligned}$$

$E(t_1,t_2)$ has mass $\phi_{E(t_1,t_2)}(\infty) = e^{-\theta(t_2-t_1)}$ on the point zero. With such a representation, we can obtain $f_{E(t_1,t_2)}(y)$, the pdf of $E(t_1,t_2)$ for $y>0$:

$$\begin{aligned} f_{E(t_1,t_2)}(y) &= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \left[\beta e^{\mu(t_2-t_1)}\right]^n y^{n-1} \exp\left\{-y\beta e^{\mu(t_2-t_1)}\right\} \times \Pr[N=n] \\ &= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \beta^n e^{n\mu(t_2-t_1)} y^{n-1} \exp\left\{-y\beta e^{\mu(t_2-t_1)}\right\} \\ &\quad \times \int_0^\infty \left\{ \frac{1}{n!}\left[z\left(e^{\mu(t_2-t_1)}-1\right)\right]^n \exp\left\{-z\left(e^{\mu(t_2-t_1)}-1\right)\right\} \times \frac{1}{\Gamma\left(\frac{\theta}{\mu}\right)} z^{\theta/\mu-1} e^{-z} \right\} dz \\ &= \frac{1}{\Gamma\left(\frac{\theta}{\mu}\right)} \times \exp\left\{-y\beta e^{\mu(t_2-t_1)}\right\} \\ &\quad \times \sum_{n=1}^{\infty} \left(\frac{\beta^n e^{n\mu(t_2-t_1)} \left(e^{\mu(t_2-t_1)}-1\right)^n y^{n-1}}{(n-1)!n!} \int_0^\infty z^{\theta/\mu+n-1} \exp\left\{-z e^{\mu(t_2-t_1)}\right\} dz \right) \\ &= \frac{1}{\Gamma\left(\frac{\theta}{\mu}\right)} \times \exp\left\{-y\beta e^{\mu(t_2-t_1)}\right\} \\ &\quad \times \sum_{n=1}^{\infty} \left(\frac{\beta^n e^{n\mu(t_2-t_1)} \left(e^{\mu(t_2-t_1)}-1\right)^n y^{n-1}}{(n-1)!n!} e^{-(\theta+n\mu)(t_2-t_1)} \int_0^\infty z_1^{\theta/\mu+n-1} e^{-z_1} dz_1 \right) \\ &= \frac{1}{\Gamma\left(\frac{\theta}{\mu}\right)} \times \exp\left\{-y\beta e^{\mu(t_2-t_1)}\right\} e^{-\theta(t_2-t_1)} \sum_{n=1}^{\infty} \left[\frac{\beta^n \left(e^{\mu(t_2-t_1)}-1\right)^n y^{n-1}}{(n-1)!n!} \Gamma\left(\frac{\theta}{\mu}+n\right) \right] \\ &= \exp\left\{-y\beta e^{\mu(t_2-t_1)} - \theta(t_2-t_1)\right\} \sum_{n=1}^{\infty} \left[\frac{\prod_{i=0}^{n-1}\left(\frac{\theta}{\mu}+i\right)}{(n-1)!n!} \beta^n \left(e^{\mu(t_2-t_1)}-1\right)^n y^{n-1} \right]. \end{aligned}$$

Hence, given $X(t_1)=x$, $X(t_2)$ has mass $e^{-\theta(t_2-t_1)}$ on the point $xe^{-\mu(t_2-t_1)}$ and pdf

$$f_{X(t_2)|X(t_1)=x} = f_{E(t_1,t_2)}\left(y - xe^{-\mu(t_2-t_1)}\right), \qquad \text{for } y > xe^{-\mu(t_2-t_1)}.$$

The representation idea in Example 9.3 also leads to another representation of the cumulative innovation in the continuous-time GAR(1) process studied in Example 5.2; see Remark 2 in Walker [2000].

Example 9.4 *Consider the continuous-time GAR(1) process obtained in Example 5.2:*

$$X(t_2) = e^{-\mu(t_2-t_1)} * X(t_1) + E(t_1,t_2),$$

where $E(t_1, t_2)$ is a non-negative integer rv with pgf

$$G_{E(t_1,t_2)}(s) = \left(e^{-\mu(t_2-t_1)} + \left[1 - e^{-\mu(t_2-t_1)}\right] \frac{1-\gamma}{1-\gamma s}\right)^{\lambda/\mu}.$$

This process will have stationary distribution NB$(\lambda/\mu, \gamma)$. The mean and variance of $E(t_1, t_2)$ can be obtained by taking derivatives of its pgf when $s = 1$:

$$\begin{aligned}
G'_{E(t_1,t_2)}(1) &= \frac{\lambda\gamma}{\mu(1-\gamma)} \left[1 - e^{-\mu(t_2-t_1)}\right], \\
G''_{E(t_1,t_2)}(1) &= \frac{2\lambda\gamma^2}{\mu(1-\gamma)^2} \left[1 - e^{-\mu(t_2-t_1)}\right] + \frac{\lambda}{\mu}\left(\frac{\lambda}{\mu} - 1\right) \frac{\gamma^2}{(1-\gamma)^2} \left[1 - e^{-\mu(t_2-t_1)}\right]^2,
\end{aligned}$$

which lead to

$$\begin{aligned}
\mathbf{E}[E(t_1,t_2)] &= G'_{E(t_1,t_2)}(1) = \frac{\lambda\gamma}{\mu(1-\gamma)} \left[1 - e^{-\mu(t_2-t_1)}\right], \\
\mathbf{Var}[E(t_1,t_2)] &= G''_{E(t_1,t_2)}(1) + G'_{E(t_1,t_2)}(1) - \left(G'_{E(t_1,t_2)}(1)\right)^2 \\
&= \frac{\lambda\gamma}{\mu(1-\gamma)^2} \left[1 - e^{-\mu(t_2-t_1)}\right] + \frac{\lambda\gamma^2}{\mu(1-\gamma)^2} \cdot e^{-\mu(t_2-t_1)} \left[1 - e^{-\mu(t_2-t_1)}\right].
\end{aligned}$$

Hence, by Table 9.1, the conditional mean and variance of $X(t_2)$ given $X(t_1) = x$ are

$$\begin{aligned}
\mathbf{E}[X(t_2) \mid X(t_1) = x] &= x e^{-\mu(t_2-t_1)} + \frac{\lambda\gamma}{\mu(1-\gamma)} \left[1 - e^{-\mu(t_2-t_1)}\right] \\
&= \frac{\lambda\gamma}{\mu(1-\gamma)} + \left(x - \frac{\lambda\gamma}{\mu(1-\gamma)}\right) e^{-\mu(t_2-t_1)}, \\
\mathbf{Var}[X(t_2) \mid X(t_1) = x] &= x e^{-\mu(t_2-t_1)} \left[1 - e^{-\mu(t_2-t_1)}\right] + \frac{\lambda\gamma}{\mu(1-\gamma)^2} \left[1 - e^{-\mu(t_2-t_1)}\right] \\
&\quad + \frac{\lambda\gamma^2}{\mu(1-\gamma)^2} \cdot e^{-\mu(t_2-t_1)} \left[1 - e^{-\mu(t_2-t_1)}\right] \\
&= \left(x + \frac{\lambda\gamma^2}{\mu(1-\gamma)^2}\right) e^{-\mu(t_2-t_1)} \left[1 - e^{-\mu(t_2-t_1)}\right] \\
&\quad + \frac{\lambda\gamma}{\mu(1-\gamma)^2} \left[1 - e^{-\mu(t_2-t_1)}\right].
\end{aligned}$$

According to Remark 2 in Walker [2000], $E(t_1, t_2)$ can be represented as a rv of Poisson mixture:

$$E(t_1,t_2) \stackrel{d}{=} Y, \qquad Y|Z = z \sim \text{Poisson}(z) \text{ for } z > 0 \text{ and } [Y|Z = 0] \equiv 0,$$

where Z is a non-negative rv with LT $\phi_Z(s) = \left(e^{-\mu(t_2-t_1)} + \left[1-e^{-\mu(t_2-t_1)}\right]\frac{(1-\gamma)/\gamma}{(1-\gamma)/\gamma+s}\right)^{\lambda/\mu}$. This can be verified as follows:

$$\begin{aligned}
G_Y(s) &= \mathbf{E}\left(s^Y\right) = \mathbf{E}\left(\mathbf{E}\left[s^Y \mid Z\right]\right) = \mathbf{E}\left(e^{Z(s-1)}\right) = \phi_Z(1-s) \\
&= \left(e^{-\mu(t_2-t_1)} + \left[1-e^{-\mu(t_2-t_1)}\right]\frac{(1-\gamma)/\gamma}{(1-\gamma)/\gamma+1-s}\right)^{\lambda/\mu} \\
&= \left(e^{-\mu(t_2-t_1)} + \left[1-e^{-\mu(t_2-t_1)}\right]\frac{1-\gamma}{1-\gamma s}\right)^{\lambda/\mu} \\
&= G_{E(t_1,t_2)}(s).
\end{aligned}$$

Furthermore, from Example 9.3 with $\beta = (1-\gamma)/\gamma$ and $\theta = \lambda$, Z can be represented as

$$\begin{aligned}
&Z \stackrel{d}{=} \sum_{i=0}^{N} Z_i, \\
&Z_0 = 0,\ Z_i \stackrel{i.i.d.}{\sim} \text{Gamma}\left(1, \gamma^{-1}(1-\gamma)\cdot e^{\mu(t_2-t_1)}\right) = \text{Exponential}\left(\gamma^{-1}(1-\gamma)\cdot e^{\mu(t_2-t_1)}\right), \\
&N|W = w \sim \text{Poisson}\left(w\left[e^{\mu(t_2-t_1)}-1\right]\right), \qquad W \sim \text{Gamma}(\lambda/\mu, 1).
\end{aligned}$$

With this kind of representation, the pmf of $E(t_1,t_2)$ is tractable:

$$\Pr[E(t_1,t_2)=0] = G_{E(t_1,t_2)}(0) = \left(1-\gamma+\gamma e^{-\mu(t_2-t_1)}\right)^{\lambda/\mu},$$

and for $j > 0$

$$\begin{aligned}
\Pr[E(t_1,t_2)=j] &= \int_0^\infty \frac{z^j}{j!}e^{-z} f_Z(z)dz \\
&= \frac{1}{j!}e^{-\theta(t_2-t_1)}\int_0^\infty \Bigg(z^j \exp\left\{-z\gamma^{-1}(1-\gamma)e^{\mu(t_2-t_1)} - z\right\} \\
&\qquad \times \sum_{n=1}^{\infty}\left[\frac{\prod_{i=0}^{n-1}\left(\frac{\lambda}{\mu}+i\right)}{(n-1)!n!}\gamma^{-n}(1-\gamma)^n\left(e^{\mu(t_2-t_1)}-1\right)^n z^{n-1}\right]\Bigg)dz \\
&= \frac{1}{j!}e^{-\theta(t_2-t_1)}\sum_{n=1}^{\infty}\Bigg[\frac{\prod_{i=0}^{n-1}\left(\frac{\lambda}{\mu}+i\right)}{(n-1)!n!}\gamma^{-n}(1-\gamma)^n\left(e^{\mu(t_2-t_1)}-1\right)^n \\
&\qquad \times \int_0^\infty z^{j+n-1}\exp\left\{-z\gamma^{-1}(1-\gamma)e^{\mu(t_2-t_1)} - z\right\}dz\Bigg] \\
&= \frac{1}{j!}e^{-\theta(t_2-t_1)}\sum_{n=1}^{\infty}\left[\prod_{i=0}^{n-1}\left(\frac{\lambda}{\mu}+i\right)\times\frac{(j+n-1)!}{(n-1)!n!}\times\frac{\gamma^{-n}(1-\gamma)^n\left(e^{\mu(t_2-t_1)}-1\right)^n}{\left[1+\gamma^{-1}(1-\gamma)e^{\mu(t_2-t_1)}\right]^n}\right].
\end{aligned}$$

Applying the convolution formula in Example 9.2, we can obtain the conditional pmf of $X(t_2)$ given $X(t_1) = x$.

Note that there is another representation for $E(t_1, t_2)$:

$$E(t_1, t_2) \stackrel{d}{=} \sum_{i=0}^{N} \left(e^{-\mu(t_2-t_1)}\right)^{U_i} * V_i = \sum_{i=0}^{N} e^{-\mu(t_2-t_1)U_i} * V_i,$$

where $N \sim \mathrm{Poisson}(\lambda(t_2 - t_1))$, $U_0 = 0$, $V_0 = 0$, $U_i \stackrel{i.i.d.}{\sim} \mathrm{U}(0, 1)$, $V_i \stackrel{i.i.d.}{\sim} \mathrm{NB}(1, 1-\gamma)$ $(i \geq 1)$, and U_i, V_i are independent. This kind of representation can be found in McKenzie [1987], Sim and Lee [1989].

Another stationary continuous-time GAR(1) process with Gamma margins in Example 6.15 is also of special interest to us, because it is an alternative to the model in Example 9.3 with the same Gamma univariate margins.

Example 9.5 *Consider the second continuous-time GAR(1) process obtained in Example 6.15:*

$$X(t_2) = \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1, t_2),$$

where K is from **P2** *and $E(t_1, t_2)$ is a positive rv with LT*

$$\phi_{E(t_1,t_2)}(s) = \left(\frac{\beta + \frac{e^{-\mu(t_2-t_1)}(1-\gamma)s}{(1-\gamma)+(1-e^{-\mu(t_2-t_1)})\gamma s}}{\beta + s} \right)^{\delta}, \qquad 0 < \gamma \leq \frac{1}{1+\beta}.$$

We know this process has stationary distribution Gamma(δ, β), so that the marginal mean and variance are

$$A = \delta\beta^{-1} \qquad \text{and} \qquad V = \delta\beta^{-2}.$$

By employing Equations (9.2.1) and (9.2.2), as well as Table 9.1, we have

$$\begin{aligned}
\mathbf{E}[X(t_2) \mid X(t_1) = x] &= \delta\beta^{-1} + (x - \delta\beta^{-1})e^{-\mu(t_2-t_1)}, \\
\mathbf{Var}[X(t_2) \mid X(t_1) = x] &= \delta\beta^{-2}\left[1 - e^{-2\mu(t_2-t_1)}\right] + (x - \delta\beta^{-1})\frac{2\gamma}{1-\gamma}e^{-\mu(t_2-t_1)}\left[1 - e^{-\mu(t_2-t_1)}\right].
\end{aligned}$$

Here the fixed parameter γ associated with the self-generalized rv K appears: it does not affect the conditional mean, but affects the conditional variance. This may help the statistician to choose the appropriate γ among $\left(0, \frac{1}{1+\beta}\right]$ when modelling.

Inspired by Walker [2000], we now investigate the representation for the cumulative innovation $E(t_1, t_2)$. For the sake of simpler notation, we replace $e^{-\mu(t_2-t_1)}$ with α $(0 \le \alpha \le 1)$ in the LT of $E(t_1, t_2)$ and rewrite it as:

$$\begin{aligned}
\phi_{E(t_1,t_2)}(s) &= \left(\frac{\beta + \frac{e^{-\mu(t_2-t_1)}(1-\gamma)s}{(1-\gamma)+(1-e^{-\mu(t_2-t_1)})\gamma s}}{\beta+s}\right)^{\delta} = \left(\frac{\beta + \frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}}{\beta+s}\right)^{\delta} \\
&= \left(\frac{\beta(1-\gamma) + [\beta\gamma(1-\alpha) + \alpha(1-\gamma)]s}{(\beta+s)[(1-\gamma)+(1-\alpha)\gamma s]}\right)^{\delta} \\
&= \left(\frac{1}{1 + \frac{(1-\gamma)(1-\alpha)s + (1-\alpha)\gamma s^2}{\beta(1-\gamma)+[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]s}}\right)^{\delta} = \left(\frac{1}{1 + \frac{(1-\alpha)s[(1-\gamma)+\gamma s]}{\beta(1-\gamma)+[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]s}}\right)^{\delta},
\end{aligned}$$

and with further algebra,

$$\begin{aligned}
&\frac{(1-\alpha)s[(1-\gamma)+\gamma s]}{\beta(1-\gamma)+[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]s} = \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma} \cdot \frac{s\left(s+\frac{1-\gamma}{\gamma}\right)}{s + \frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}} \\
&= \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\left(s + \left[\frac{1-\gamma}{\gamma} - \frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\right] \cdot \frac{s}{s+\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}}\right) \\
&= \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\left(s + (1-\gamma)\left[\frac{1}{\gamma} - \frac{\beta}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\right] \cdot \frac{s}{s+\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}}\right) \\
&= \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\left(s + \frac{\alpha(1-\gamma)(1-\gamma-\beta\gamma)}{\gamma[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]} \cdot \frac{s}{s+\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}}\right). \qquad (9.2.7)
\end{aligned}$$

We propose the following representation for the cumulative innovation $E(t_1, t_2)$:

$$\begin{aligned}
&E(t_1,t_2) \stackrel{d}{=} Y, \qquad [Y|Z=z] \stackrel{d}{=} \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma} \cdot z + [W|Z=z], \\
&[W|Z=z] \stackrel{d}{=} \sum_{i=0}^{N} W_i, \qquad W_0 = 0, \quad W_i \stackrel{i.i.d.}{\sim} \text{Gamma}\left(1, \frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\right), \\
&N|Z=z \sim \text{Poisson}\left(z \cdot \frac{\alpha(1-\alpha)(1-\gamma)(1-\gamma-\beta\gamma)}{[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]^2}\right), \qquad Z \sim \text{Gamma}(\delta, 1).
\end{aligned}$$

We can verify this representation by checking the LT's:

$$\phi_Y(s) = \mathbf{E}\left(e^{-sY}\right) = \mathbf{E}\left(\mathbf{E}\left[e^{-sY}\middle|Z\right]\right)$$

$$
\begin{aligned}
&= \mathbf{E}\left(\mathbf{E}\left[\exp\left\{-s\left(\frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot Z + W\right)\right\}\Big| Z\right]\right) \\
&= \mathbf{E}\left(\exp\left\{-s\left(\frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot Z\right)\right\}\mathbf{E}\left[e^{-sW}\Big| Z\right]\right) \\
&= \mathbf{E}\left(\exp\left\{-s\left(\frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot Z\right)\right\}\right. \\
&\qquad \left.\times \exp\left\{Z\cdot\frac{\alpha(1-\alpha)(1-\gamma)(1-\gamma-\beta\gamma)}{[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]^2}\left(\frac{\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}}{s+\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}}-1\right)\right\}\right) \\
&= \mathbf{E}\left(\exp\left\{-Z\cdot\frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\left(s+\frac{\alpha(1-\gamma)(1-\gamma-\beta\gamma)}{\gamma[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]}\cdot\frac{s}{s+\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}}\right)\right\}\right) \\
&= \mathbf{E}\left(\exp\left\{-Z\cdot\frac{(1-\alpha)s[(1-\gamma)+\gamma s]}{\beta(1-\gamma)+[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]s}\right\}\right) \qquad \textit{[from (9.2.7)]} \\
&= \phi_Z\left(\frac{(1-\alpha)s[(1-\gamma)+\gamma s]}{\beta(1-\gamma)+[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]s}\right) = \left(\frac{1}{1+\frac{(1-\alpha)s[(1-\gamma)+\gamma s]}{\beta(1-\gamma)+[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]s}}\right)^{\delta} \\
&= \phi_{E(t_1,t_2)}(s).
\end{aligned}
$$

Note that W is the same kind of random variable as $\sum_{i=0}^{N} Y_i$ in Example 9.3. This can help us to find the pdf of Y in this example. Conditioned on $Z = z$, Y has non-zero mass on point $\frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot z$:

$$
\Pr\left[Y = \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot z \;\Big|\; Z = z\right] = \exp\left\{-z\cdot\frac{\alpha(1-\alpha)(1-\gamma)(1-\gamma-\beta\gamma)}{[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]^2}\right\},
$$

and the pdf for $y > \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot z$ is:

$$
\begin{aligned}
f_{Y|Z=z}(y) &= f_{W|Z=z}\left(y - \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot z\right) \\
&= \sum_{n=1}^{\infty}\frac{1}{(n-1)!}\left[\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\right]^n\left(y-\frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot z\right)^{n-1} \\
&\qquad\times\exp\left\{-\left(y-\frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\cdot z\right)\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma}\right\} \\
&\qquad\times\frac{1}{n!}\left[z\cdot\frac{\alpha(1-\alpha)(1-\gamma)(1-\gamma-\beta\gamma)}{[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]^2}\right]^n\exp\left\{-z\cdot\frac{\alpha(1-\alpha)(1-\gamma)(1-\gamma-\beta\gamma)}{[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]^2}\right\} \\
&= \exp\left\{-y\cdot\frac{\beta(1-\gamma)}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma} - z\cdot\frac{(1-\alpha)(1-\gamma)(\alpha-\alpha\gamma-\beta\gamma-\alpha\beta\gamma)}{[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]^2}\right\} \\
&\qquad\times\sum_{n=1}^{\infty}\left\{\frac{1}{(n-1)!n!}\left[z\cdot\frac{\alpha(1-\alpha)\beta(1-\gamma)^2(1-\gamma-\beta\gamma)}{[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]^3}\right]^n\right.
\end{aligned}
$$

$$\times \left(y - \frac{(1-\alpha)\gamma}{\alpha(1-\gamma-\beta\gamma)+\beta\gamma} \cdot z \right)^{n-1} \Bigg\}.$$

Hence, the unconditional pdf of Y is

$$\begin{aligned} f_Y(y) &= f_Z\left(\frac{y[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]}{(1-\alpha)\gamma}\right) \cdot \exp\left\{-\frac{y\alpha(1-\gamma)(1-\gamma-\beta\gamma)}{\gamma[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]}\right\} \\ &\quad + \int_0^{\frac{y[\alpha(1-\gamma-\beta\gamma)+\beta\gamma]}{(1-\alpha)\gamma}} f_{Y|Z=z}(y) f_Z(z) dz \end{aligned}$$

for $y > 0$. Note that Y has zero mass on point 0. This is unlike the cumulative innovation in Example 9.3. Thus, we obtain the pdf of $E(t_1, t_2)$. By Equation (9.2.3), we can obtain the conditional pdf of $X(t_2)$ given $X(t_1) = x$:

$$\begin{aligned} f_{X(t_2)|X(t_1)=x}(y) &= \int_0^y f_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}(z) \cdot f_{E(t_1,t_2)}(y-z) dz \\ &\qquad + \Pr\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x = 0\right] \times f_{E(t_1,t_2)}(y) \\ &= \int_0^y f_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}(z) \cdot f_{E(t_1,t_2)}(y-z) dz \\ &\qquad + \exp\left\{-\frac{xe^{-\mu(t_2-t_1)}(1-\gamma)}{\left(1-e^{-\mu(t_2-t_1)}\right)\gamma}\right\} \times f_{E(t_1,t_2)}(y), \end{aligned}$$

where $f_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}(z)$ can be found in Equation (9.2.6).

The conditional pdf does not have a closed form, but can be computed with numerical methods.

As an alternative to explicit stochastic representations, the numerical approach to calculate the conditional pmf or cdf via approximation of inversion of the characteristic function seems to be promising. This will be discussed in Section 10.1, where we study maximum likelihood estimation.

9.3 Joint properties

Developing multivariate distributions, in which every univariate margin has the same distribution, like multivariate Poisson, multivariate Gamma, etc, is very useful for modelling and quite challenging for researchers. Perhaps the most successful multivariate distribution is the multinormal, which has a correlation coefficient parameter for each bivariate margin. However, for the non-normal situation, there is in general no multivariate distribution with such nice properties. Construction approaches for multivariate distributions are very diverse. For references, recent books are Joe [1997], Johnson, Kotz and Balakrishnan [1997], Kocherlakota and Kocherlakota [1992], Hutchinson and Lai [1990].

Note that under steady state, the Markov process has the stationary distribution for each univariate marginal distribution. Hence, a steady state Markov process provides multivariate distribution with any finite dimensions, namely $(X(t_1), X(t_2), \ldots, X(t_n))$ is distributed in a n-dimensional multivariate distribution, where $n = 2, 3, \ldots$.

The multivariate distribution may be of interest in themselves, and potentially have applications to non-normal multivariate data. A byproduct of the construction of GAR(1) processes is a method of construction of multivariate distributions.

Since the continuous-time GAR(1) process is newly developed, we shall study the multivariate distributions resulting from margins of the stationary continuous-time GAR(1) process.

Also in this section, we shall study the covariance at two time points and the auto-correlation function; these are useful in describing the degree of dependence over time for the continuous-time GAR(1) process.

9.3.1 Bivariate distributions

We first investigate the auto-covariance and auto-correlation of a continuous-time GAR(1) process:

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1, t_2). \tag{9.3.1}$$

Assume that the mean function and variance function exist and are

$$A(t) = \mathbf{E}\,(X(t)) \qquad \text{and} \qquad V(t) = \mathbf{E}\,[(X(t) - A(t))^2] = \mathbf{E}\,[X^2(t)] - A^2(t).$$

By the independence property of two summands on the right hand side of (9.3.1), we have

$$\mathbf{Cov}\,[X(t_1), X(t_2)] \;=\; \mathbf{E}\,[(X(t_1) - A(t_1)) \cdot (X(t_2) - A(t_2))]$$

$$
\begin{aligned}
&= \mathbf{E}\left[X(t_1)\cdot X(t_2)\right] - \mathbf{E}[X(t_1)]\cdot\mathbf{E}[X(t_2)] \\
&= \mathbf{E}\left\{X(t_1)\cdot\left(\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1,t_2)\right)\right\} - A(t_1)A(t_2) \\
&= \mathbf{E}\left[X(t_1)\cdot\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\right] + \mathbf{E}\left[X(t_1)\cdot E(t_1,t_2)\right] - A(t_1)A(t_2) \\
&= \mathbf{E}\left\{\mathbf{E}\left[X(t_1)\cdot\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\middle|X(t_1)\right]\right\} + \mathbf{E}[X(t_1)]\mathbf{E}[E(t_1,t_2)] - A(t_1)A(t_2) \\
&= \mathbf{E}\left\{\mathbf{E}\left[X(t_1)\cdot\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\middle|X(t_1)\right]\right\} - A(t_1)\cdot\left(A(t_2) - \mathbf{E}[E(t_1,t_2)]\right) \\
&= \mathbf{E}\left\{\mathbf{E}\left[X(t_1)\cdot\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\middle|X(t_1)\right]\right\} - A(t_1)\cdot\mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\right] \\
&= \mathbf{E}\left\{X^2(t_1)\cdot\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\right\} - A(t_1)\cdot\mathbf{E}\left\{\mathbf{E}\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)\middle|X(t_1)\right]\right\} \\
&= \mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\cdot\mathbf{E}\left\{X^2(t_1)\right\} - A(t_1)\cdot\mathbf{E}\left\{X(t_1)\cdot\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\right\} \\
&= \mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\cdot\mathbf{E}\left\{X^2(t_1)\right\} - \mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\cdot A(t_1)\cdot\mathbf{E}\left\{X(t_1)\right\} \\
&= \mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\cdot\left(\mathbf{E}\left\{X^2(t_1)\right\} - A^2(t_1)\right) \\
&= \mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\cdot V(t_1).
\end{aligned}
$$

This shows that the auto-covariance is linear in the variance of earlier time point. If K has finite mean, then by Theorem 3.2.1,

$$
\mathbf{Cov}\left[X(t_1),X(t_2)\right] = e^{-r\mu(t_2-t_1)}\cdot V(t_1), \qquad \text{for } r>0.
$$

Hence, the auto-covariance decreases at an exponential rate in the time difference t_2-t_1.

Consequently, the auto-correlation function, $\rho(t_1,t_2)$, can be obtained as

$$
\begin{aligned}
\rho(t_1,t_2) &= \frac{\mathbf{Cov}\left[X(t_1),X(t_2)\right]}{\sqrt{\mathbf{Var}[X(t_1)]\mathbf{Var}[X(t_2)]}} = \frac{\mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\cdot V(t_1)}{\sqrt{V(t_1)V(t_2)}} \\
&= \mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\cdot\sqrt{\frac{V(t_1)}{V(t_2)}}.
\end{aligned}
$$

When the variance function $V(t)$ is a constant, the auto-correlation function will be

$$
\rho(t_1,t_2) = \mathbf{E}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right].
$$

Besides, if K has finite mean, the auto-correlation function is finite and in exponential form

$$
\rho(t_1,t_2) = e^{-r\mu(t_2-t_1)}, \qquad \text{where } r>0.
$$

Table 9.2: *Auto-covariance and auto-correlation function of the stationary continuous-time GAR(1) process associated with known self-generalized random variable* $K\left(e^{-\mu(t_2-t_1)}\right)$. *Here the variance function is* $V(t) = V$.

$K\left(e^{-\mu(t_2-t_1)}\right)$	$\mathbf{Cov}\left[X(t_1), X(t_2)\right]$	$\rho(t_1,t_2)$
I1	$e^{-\mu(t_2-t_1)} \cdot V$	$e^{-\mu(t_2-t_1)}$
I2	$e^{-\mu(t_2-t_1)} \cdot V$	$e^{-\mu(t_2-t_1)}$
I4	$e^{-\mu(t_2-t_1)} \cdot V$	$e^{-\mu(t_2-t_1)}$
I5	$e^{-\theta\mu(t_2-t_1)} \cdot V$	$e^{-\theta\mu(t_2-t_1)}$
P1	$e^{-\mu(t_2-t_1)} \cdot V$	$e^{-\mu(t_2-t_1)}$
P2	$e^{-\mu(t_2-t_1)} \cdot V$	$e^{-\mu(t_2-t_1)}$
P4	$e^{-\mu(t_2-t_1)} \cdot V$	$e^{-\mu(t_2-t_1)}$
P5	$e^{-\theta\mu(t_2-t_1)} \cdot V$	$e^{-\theta\mu(t_2-t_1)}$

Table 9.2 lists the auto-covariance and auto-correlation function of the stationary continuous-time GAR(1) process associated with known self-generalized random variable K discussed in Chapter 3. These continuous-time GAR(1) processes have constant variance function: $V(t) = V$.

This approach can be applied directly to the continuous-time GAR(1) process where $\mu(t)$ is a function, instead of a constant:

$$\begin{aligned}
\mathbf{Cov}\left[X(t_1), X(t_2)\right] &= \mathbf{E}\left[K\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)\right] \cdot V(t_1), \\
\rho(t_1,t_2) &= \mathbf{E}\left[K\left(e^{-\int_t^{t_2}\mu(\tau)d\tau}\right)\right] \cdot \sqrt{\frac{V(t_1)}{V(t_2)}}.
\end{aligned}$$

Next we consider the bivariate distribution of $(X(t_1), X(t_2))$. This is carried out by looking into the bivariate pgf, or bivariate LT, or bivariate cf depending if $X(t_1)$ and $X(t_2)$ are non-negative integer, or positive, or real-valued.

$$\begin{aligned}
G_{(X(t_1),X(t_2))}(s_1,s_2) &= \mathbf{E}\left[s_1^{X(t_1)} s_2^{X(t_2)}\right] = \mathbf{E}\left[s_1^{X(t_1)} s_2^{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1,t_2)}\right] \\
&= \mathbf{E}\left\{\mathbf{E}\left[s_1^{X(t_1)} s_2^{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)} s_2^{E(t_1,t_2)} \middle| X(t_1)\right]\right\} \\
&= \mathbf{E}\left[s_1^{X(t_1)} G_K^{X(t_1)}\left(s_2; e^{-\mu(t_2-t_1)}\right) \mathbf{E}\left(s_2^{E(t_1,t_2)}\right)\right] \\
&= \mathbf{E}\left[\left(s_1 G_K\left(s_2; e^{-\mu(t_2-t_1)}\right)\right)^{X(t_1)}\right] \cdot \mathbf{E}\left[s_2^{E(t_1,t_2)}\right]
\end{aligned}$$

$$
\begin{aligned}
&= G_{X(t_1)}\left(s_1 G_K\left(s_2; e^{-\mu(t_2-t_1)}\right)\right) \cdot G_{E(t_1,t_2)}(s_2), &(9.3.2)\\
\phi_{(X(t_1),X(t_2))}(s_1,s_2) &= \mathbf{E}\left[e^{-s_1X(t_1)-s_2X(t_2)}\right] = \mathbf{E}\left[e^{-s_1X(t_1)}e^{-s_2\left[\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)+E(t_1,t_2)\right]}\right]\\
&= \mathbf{E}\left\{\mathbf{E}\left[e^{-s_1X(t_1)}e^{-s_2\cdot\left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1)-s_2E(t_1,t_2)}\Big| X(t_1)\right]\right\}\\
&= \mathbf{E}\left[e^{-s_1X(t_1)}\phi_K^{X(t_1)}\left(s_2; e^{-\mu(t_2-t_1)}\right)\mathbf{E}\left(e^{-s_2E(t_1,t_2)}\right)\right]\\
&= \mathbf{E}\left[e^{-\left[s_1-\log\phi_K\left(s_2;e^{-\mu(t_2-t_1)}\right)\right]X(t_1)}\right]\cdot\mathbf{E}\left(e^{-s_2E(t_1,t_2)}\right)\\
&= \phi_{X(t_1)}\left(s_1-\log\phi_K\left(s_2;e^{-\mu(t_2-t_1)}\right)\right)\cdot\phi_{E(t_1,t_2)}(s_2), &(9.3.3)\\
\varphi_{(X(t_1),X(t_2))}(s_1,s_2) &= \mathbf{E}\left[e^{i(s_1X(t_1)+s_2X(t_2))}\right] = \mathbf{E}\left[e^{is_1X(t_1)+is_2e^{-\mu(t_2-t_1)}X(t_1)+is_2E(t_1,t_2)}\right]\\
&= \mathbf{E}\left[e^{i\left[s_1+e^{-\mu(t_2-t_1)}s_2\right]X(t_1)}\right]\cdot\mathbf{E}\left[e^{is_2E(t_1,t_2)}\right]\\
&= \varphi_{X(t_1)}\left(s_1+e^{-\mu(t_2-t_1)}s_2\right)\cdot\varphi_{E(t_1,t_2)}(s_2). &(9.3.4)
\end{aligned}
$$

Furthermore, if the continuous-time GAR(1) process is under steady state with the pgf $G_X(s)$, or LT $\phi_X(s)$, or cf $\varphi_X(s)$ of the marginal distribution, then the marginal distribution is DSD/GDSD, or SD/GDSD (see Chapter 7). Hence, $(X(t_1), X(t_2))$ is distributed in a bivariate DSD/GDSD or bivariate SD/GSD distribution.

Theorem 9.3.1 *Suppose the continuous-time GAR(1) process*

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1,t_2)$$

has stationary distribution with pgf $G_X(s)$, or LT $\phi_X(s)$, or cf $\varphi_X(s)$.

(1) If X is GDSD associated with self-generalized rv $K(\alpha)$, then

$$G_{(X(t_1),X(t_2))}(s_1,s_2) = \frac{G_X\left(s_1G_K\left(s_2;e^{-\mu(t_2-t_1)}\right)\right)\cdot G_X(s_2)}{G_X\left(G_K\left(s_2;e^{-\mu(t_2-t_1)}\right)\right)} \qquad (9.3.5)$$

is the pgf of a bivariate GDSD distribution whose marginal distributions are the same as that of X.

(2) If X is GSD associated with self-generalized rv $K(\alpha)$, then

$$\phi_{(X(t_1),X(t_2))}(s_1,s_2) = \frac{\phi_X\left(s_1-\log\phi_K\left(s_2;e^{-\mu(t_2-t_1)}\right)\right)\cdot\phi_X(s_2)}{\phi_X\left(-\log\phi_K\left(s_2;e^{-\mu(t_2-t_1)}\right)\right)} \qquad (9.3.6)$$

is the LT of a bivariate GSD distribution whose marginal distributions are the same as that of X.

(3) If X is a real SD rv, then

$$\varphi_{(X(t_1),X(t_2))}(s_1, s_2) = \frac{\varphi_X\left(s_1 + e^{-\mu(t_2-t_1)}s_2\right) \cdot \varphi_X(s_2)}{\varphi_X\left(e^{-\mu(t_2-t_1)}s_2\right)} \tag{9.3.7}$$

is the cf of a bivariate SD distribution whose marginal distributions are the same as that of X.

Proof: When the continuous-time GAR(1) process is under steady state, we know the form of the pgf, or LT, or cf of the cumulative innovation $E(t_1, t_2)$ is

$$\begin{aligned} G_{E(t_1,t_2)}(s_2) &= \frac{G_X(s_2)}{G_X\left(G_K\left(s_2; e^{-\mu(t_2-t_1)}\right)\right)}, \\ \phi_{E(t_1,t_2)}(s_2) &= \frac{\phi_X(s_2)}{\phi_X\left(-\log \phi_K\left(s_2; e^{-\mu(t_2-t_1)}\right)\right)}, \\ \varphi_{E(t_1,t_2)}(s_2) &= \frac{\varphi_X(s_2)}{\varphi_X\left(e^{-\mu(t_2-t_1)}s_2\right)}. \end{aligned}$$

Substituting in the previous equations (9.3.2) – (9.3.4) completes the proof.

This theorem indicates that the resulting bivariate distributions depend only on the univariate margin and the associated self-generalized random variable. After specifying the distributions for them, we can obtain a specific bivariate pgf, or bivariate LT, or bivariate cf, which will determine the resulting bivariate distribution. As to the joint pmf or pdf of $(X(t_1), X(t_2))$, in general, we can employ the equation:

$$f_{(X(t_1),X(t_2))}(x_1, x_2) = f_{X(t_1)}(x_1) \cdot f_{X(t_2)|X(t_1)}(x_2 \mid x_1),$$

where the conditional pmf or pdf $f_{X(t_2)|X(t_1)}(x_2 \mid x_1)$ has been discussed Section 9.2. Therefore, Theorem 9.3.1 shows one approach to constructing the bivariate distributions for GDSD and GSD univariate distributions.

In general, the bivariate distribution function of two adjacent time points is not symmetric in its arguments, i.e., $f_{(X(t_1),X(t_2))}(x_1, x_2) \neq f_{(X(t_1),X(t_2))}(x_2, x_1)$. Only a few special cases exists. This implies that generally the continuous-time generalized AR(1) process is not time reversible.

For DSD and SD, $G_K\left(s_2; e^{-\mu(t_2-t_1)}\right)$ and $\phi_K\left(s_2; e^{-\mu(t_2-t_1)}\right)$ are known. Hence, we have the following corollary.

Corollary 9.3.2 *When K is from* **I1** *or* **P1**, *Theorem 9.3.1 yields*

$$G_{(X(t_1),X(t_2))}(s_1,s_2) = \frac{G_X\left(\left[1-e^{-\mu(t_2-t_1)}\right]s_1 + e^{-\mu(t_2-t_1)}s_1s_2\right)\cdot G_X(s_2)}{G_X\left(\left[1-e^{-\mu(t_2-t_1)}\right] + e^{-\mu(t_2-t_1)}s_2\right)}, \tag{9.3.8}$$

$$\phi_{(X(t_1),X(t_2))}(s_1,s_2) = \frac{\phi_X\left(s_1 + e^{-\mu(t_2-t_1)}s_2\right)\cdot \phi_X(s_2)}{\phi_X\left(e^{-\mu(t_2-t_1)}s_2\right)}, \tag{9.3.9}$$

$$\varphi_{(X(t_1),X(t_2))}(s_1,s_2) = \frac{\varphi_X\left(s_1 + e^{-\mu(t_2-t_1)}s_2\right)\cdot \varphi_X(s_2)}{\varphi_X\left(e^{-\mu(t_2-t_1)}s_2\right)}. \tag{9.3.10}$$

They are the pgf of a bivariate DSD distribution and the LT or cf of a bivariate SD distribution respectively.

Next we look into some examples resulting from Theorem 9.3.1 and Corollary 9.3.2. For the sake of simpler notation, we denote $\alpha = e^{-\mu(t_2-t_1)}$. First, we study some bivariate DSD distributions.

Example 9.6 (Bivariate DSD distributions) *Here we assume the marginal distribution of a stationary continuous-time GAR(1) process associated with binomial-thinning has pgf $G_X(s)$. By (9.3.8), we can obtain the pgf of a bivariate DSD distribution.*

(1) **(Poisson margins)** *Let $G_X(s) = \exp\{\lambda(s-1)\}$ $(\lambda > 0)$. Then*

$$\begin{aligned}
G_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{G_X\left(s_1\left([1-\alpha]+\alpha s_2\right)\right)\cdot G_X(s_2)}{G_X\left([1-\alpha]+\alpha s_2\right)} \\
&= \exp\left\{\lambda\left(s_1\left([1-\alpha]+\alpha s_2\right) + s_2 - [1-\alpha] - \alpha s_2 - 1\right)\right\} \\
&= \exp\left\{\lambda\left(s_1[1-\alpha] + s_2[1-\alpha] + s_1s_2\alpha + \alpha - 2\right)\right\} \\
&= \exp\left\{\lambda(1-\alpha)(s_1-1)\right\}\times\exp\left\{\lambda(1-\alpha)(s_2-1)\right\}\times\exp\left\{\lambda\alpha(s_1s_2-1)\right\},
\end{aligned}$$

which leads to the following stochastic representation for bivariate $(X(t_1),X(t_2))$:

$$\begin{cases} X(t_1) \overset{d}{=} Z_1 + Z_{12}, \\ X(t_2) \overset{d}{=} Z_2 + Z_{12}, \end{cases}$$

where Z_1, $Z_2 \sim$ Poisson$(\lambda(1-\alpha))$, $Z_{12} \sim$ Poisson$(\lambda\alpha)$, and Z_1, Z_2, Z_{12} are independent. This representation is useful in finding the joint pmf of $(X(t_1),X(t_2))$:

$$\Pr[X(t_1)=x_1, X(t_2)=x_2] = \Pr[Z_1+Z_{12}=x_1, Z_2+Z_{12}=x_2]$$

$$= \sum_{i=0}^{\min(x_1,x_2)} \Pr[Z_{12} = i, Z_1 = x_1 - i, Z_2 = x_2 - i]$$

$$= \sum_{i=0}^{\min(x_1,x_2)} \Pr[Z_{12} = i] \cdot \Pr[Z_1 = x_1 - i] \cdot \Pr[Z_2 = x_2 - i]$$

$$= \sum_{i=0}^{\min(x_1,x_2)} \left[\frac{(\lambda\alpha)^i}{i!} e^{-\lambda\alpha} \times \frac{(\lambda [1-\alpha])^{x_1 - i}}{(x_1 - i)!} e^{-\lambda(1-\alpha)} \times \frac{(\lambda [1-\alpha])^{x_2 - i}}{(x_2 - i)!} e^{-\lambda(1-\alpha)} \right]$$

$$= [\lambda(1-\alpha)]^{x_1+x_2} \cdot e^{-\lambda(2-\alpha)} \times \sum_{i=0}^{\min(x_1,x_2)} \frac{1}{i!(x_1-i)!(x_2-i)!} \cdot \frac{\alpha^i}{\lambda^i (1-\alpha)^{2i}},$$

where $x_1,\ x_2 \geq 0$.

(2) **(NB margins)** *Let* $G_X(s) = \left(\frac{1-\gamma}{1-\gamma s}\right)^\delta$ $(0 < \gamma < 1, \delta > 0)$. *Then*

$$G_{(X(t_1),X(t_2))}(s_1, s_2) = \frac{G_X\left(s_1\left([1-\alpha] + \alpha s_2\right)\right) \cdot G_X(s_2)}{G_X\left([1-\alpha] + \alpha s_2\right)}$$

$$= \left(\frac{(1-\gamma) \cdot (1 - \gamma \cdot ([1-\alpha] + \alpha s_2))}{(1 - \gamma \cdot s_1\left([1-\alpha] + \alpha s_2\right)) \cdot (1 - \gamma s_2)} \right)^\delta$$

$$= \left(\frac{(1-\gamma)(1-\gamma+\alpha\gamma) - \alpha\gamma(1-\gamma)s_2}{1 - (1-\alpha)\gamma s_1 - \gamma s_2 + [(1-\alpha)\gamma - \alpha]\gamma s_1 s_2 + \alpha\gamma^2 s_1 s_2^2} \right)^\delta.$$

(3) **(generalized Poisson margins)** *Let* $G_X(s) = \exp\left\{\theta(\sum_{k=1}^\infty \eta(k\eta)^{k-1} e^{-k\eta} s^k / k! - 1)\right\}$, *where* $\theta > 0,\ 0 \leq \eta \leq 1$. *Then*

$$G_{(X(t_1),X(t_2))}(s_1, s_2) = \frac{G_X\left(s_1\left([1-\alpha] + \alpha s_2\right)\right) \cdot G_X(s_2)}{G_X\left([1-\alpha] + \alpha s_2\right)}$$

$$= \exp\left\{ \theta \left(\sum_{k=1}^\infty \frac{\eta(k\eta)^{k-1} e^{-k\eta}}{k!} \left[s_1^k \left([1-\alpha] + \alpha s_2\right)^k + s_2^k - \left([1-\alpha] + \alpha s_2\right)^k \right] - 1 \right) \right\}.$$

(4) **(power series margins)** *Let* $G_X(s) = s^{-1}[1 - (1-s)^\theta]$ $(0 < \theta \leq 1)$. *Then*

$$G_{(X(t_1),X(t_2))}(s_1, s_2) = \frac{G_X\left(s_1\left([1-\alpha] + \alpha s_2\right)\right) \cdot G_X(s_2)}{G_X\left([1-\alpha] + \alpha s_2\right)}$$

$$= \frac{s_1^{-1}(1-\alpha+\alpha s_2)^{-1}\left[1 - (1 - s_1(1-\alpha+\alpha s_2))^\theta\right] \cdot s_2^{-1}[1 - (1-s_2)^\theta]}{(1-\alpha+\alpha s_2)^{-1}\left[1 - (1 - (1-\alpha+\alpha s_2))^\theta\right]}$$

$$= \frac{\left(1 - [1 - (1-\alpha)s_1 - \alpha s_1 s_2)]^\theta\right) \cdot [1 - (1-s_2)^\theta]}{s_1 s_2 \left[1 - \alpha^\theta (1-s_2)^\theta\right]}.$$

(5) **(logarithmic series margins)** *Let* $G_X(s) = s^{-1}\log(1-cs)/\log(1-c)$, *where* $c = 1 - e^{-\theta}$, $\theta > 0$. *Then*

$$
\begin{aligned}
G_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{G_X\left(s_1\left([1-\alpha]+\alpha s_2\right)\right)\cdot G_X(s_2)}{G_X\left([1-\alpha]+\alpha s_2\right)} \\
&= \frac{s_1^{-1}(1-\alpha+\alpha s_2)^{-1}\log\left[1-cs_1(1-\alpha+\alpha s_2)\right]\cdot s_2^{-1}\log[1-cs_2]/\log(1-c)}{(1-\alpha+\alpha s_2)^{-1}\log\left[1-c(1-\alpha+\alpha s_2)\right]} \\
&= \frac{\log[1-(1-\alpha)cs_1-\alpha cs_1s_2]\cdot\log(1-cs_2)}{\log(1-c)\cdot s_1s_2\log(1-c+\alpha c-\alpha cs_2)}.
\end{aligned}
$$

(6) **(discrete stable margins)** *Let* $G_X(s) = e^{-\lambda(1-s)^\theta}$, *where* $\lambda > 0,\ 0 < \theta \le 1$. *Then*

$$
\begin{aligned}
G_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{G_X\left(s_1\left([1-\alpha]+\alpha s_2\right)\right)\cdot G_X(s_2)}{G_X\left([1-\alpha]+\alpha s_2\right)} \\
&= \exp\left\{-\lambda\left(\left[1-s_1\left(1-\alpha+\alpha s_2\right)\right]^\theta + (1-s_2)^\theta - \left[1-\left(1-\alpha+\alpha s_2\right)\right]^\theta\right)\right\} \\
&= \exp\left\{-\lambda\left(\left[1-(1-\alpha)s_1-\alpha s_1s_2\right]^\theta + (1-\alpha^\theta)(1-s_2)^\theta\right)\right\}.
\end{aligned}
$$

(7) **(discrete Mittag-Leffler margins)** *Let* $G_X(s) = \frac{1}{1+d(1-s)^\gamma}$, *where* $d > 0,\ 0 < \gamma \le 1$. *Then*

$$
\begin{aligned}
G_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{G_X\left(s_1\left([1-\alpha]+\alpha s_2\right)\right)\cdot G_X(s_2)}{G_X\left([1-\alpha]+\alpha s_2\right)} \\
&= \frac{1+d\left(1-\left([1-\alpha]+\alpha s_2\right)\right)^\gamma}{\left[1+d\left(1-s_1\left([1-\alpha]+\alpha s_2\right)\right)^\gamma\right]\cdot\left[1+d(1-s_2)^\gamma\right]} \\
&= \frac{1+d\alpha^\gamma(1-s_2)^\gamma}{\left[1+d\left(1-(1-\alpha)s_1-\alpha s_1s_2\right)^\gamma\right]\cdot\left[1+d(1-s_2)^\gamma\right]}.
\end{aligned}
$$

(8) **(GNBC margins)** *Let* $G_X(s) = \exp\left\{\int_{(0,1)}\log\left(\frac{1-q}{1-qs}\right)V(dq)\right\}$. *Then*

$$
\begin{aligned}
G_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{G_X\left(s_1\left([1-\alpha]+\alpha s_2\right)\right)\cdot G_X(s_2)}{G_X\left([1-\alpha]+\alpha s_2\right)} \\
&= \exp\left\{\int_{(0,1)}\left[\log\frac{1-q}{1-qs_1\left([1-\alpha]+\alpha s_2\right)} + \log\frac{1-q}{1-qs_2} - \log\frac{1-q}{1-q\left([1-\alpha]+\alpha s_2\right)}\right]V(dq)\right\} \\
&= \exp\left\{\int_{(0,1)}\log\left(\frac{(1-q)\cdot\left(1-q\cdot\left([1-\alpha]+\alpha s_2\right)\right)}{\left(1-q\cdot s_1\left([1-\alpha]+\alpha s_2\right)\right)\cdot(1-qs_2)}\right)V(dq)\right\} \\
&= \exp\left\{\int_{(0,1)}\log\left(\frac{(1-q)(1-q+\alpha q)-\alpha q(1-q)s_2}{1-(1-\alpha)qs_1-qs_2+[(1-\alpha)q-\alpha]qs_1s_2+\alpha q^2s_1s_2^2}\right)V(dq)\right\}.
\end{aligned}
$$

Note that this is the general form of the pgf of the bivariate GNBC distribution family (in the family of bivariate DSD distributions), which includes many bivariate DSD distributions with GNBC margins.

Now we turn to the bivariate SD distributions.

Example 9.7 (Bivariate SD distributions) *We assume the marginal distribution of a stationary continuous-time GAR(1) process associated with constant multiplier has LT $\phi_X(s)$ or cf $\varphi_X(s)$. By (9.3.9) and (9.3.10), we can obtain the LT or cf of a bivariate SD distribution.*

(1) **(Gamma margins)** *Let $\phi_X(s) = \left(\frac{\beta}{\beta+s}\right)^\delta$, where $\delta,\ \beta > 0$. Then*

$$\begin{aligned}\phi_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{\phi_X\left(s_1+\alpha s_2\right)\cdot\phi_X(s_2)}{\phi_X\left(\alpha s_2\right)}\\ = \left(\frac{\beta}{\beta+s_1+\alpha s_2}\right)^\delta\left(\frac{\beta}{\beta+s_2}\right)^\delta\left(\frac{\beta}{\beta+\alpha s_2}\right)^{-\delta} &= \left(\frac{\beta\cdot(\beta+\alpha s_2)}{(\beta+s_1+\alpha s_2)\cdot(\beta+s_2)}\right)^\delta\\ = \left(\frac{\beta^2+\alpha\beta s_2}{\beta^2+\beta s_1+(\alpha+1)\beta s_2+\alpha\beta s_1 s_2}\right)^\delta.\end{aligned}$$

(2) **(GGC margins)** *Let $\phi_X(s) = \exp\left\{\int_{(0,\infty)}\log\left(\frac{u}{u+s}\right)U(du)\right\}$, where the non-negative measure $U(du)$ on $(0,\infty)$ satisfies*

$$\int_{(0,1]}|\log u|U(du) < \infty \quad \textit{and} \quad \int_{(1,\infty)}u^{-1}U(du) < \infty.$$

Then

$$\begin{aligned}\phi_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{\phi_X\left(s_1+\alpha s_2\right)\cdot\phi_X(s_2)}{\phi_X\left(\alpha s_2\right)}\\ &= \exp\left\{\int_{(0,\infty)}\left[\log\frac{u}{u+s_1+\alpha s_2}+\log\frac{u}{u+s_2}-\log\frac{u}{u+\alpha s_2}\right]U(du)\right\}\\ &= \exp\left\{\int_{(0,\infty)}\log\left[\frac{u(u+\alpha s_2)}{(u+s_1+\alpha s_2)\cdot(u+s_2)}\right]U(du)\right\}\\ &= \exp\left\{\int_{(0,\infty)}\log\left[\frac{u^2+u\alpha s_2}{u^2+us_1+(\alpha+1)us_2+\alpha u s_1 s_2}\right]U(du)\right\}.\end{aligned}$$

(3) **(inverse Gaussian margins)** *Let* $\phi_X(s) = \exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 + \frac{2\mu^2}{\lambda}s\right)^{1/2}\right]\right\}$, *where* $\lambda,\ \mu > 0$. *Then*

$$\begin{aligned}
\phi_{(X(t_1),X(t_2))}(s_1, s_2) &= \frac{\phi_X(s_1 + \alpha s_2) \cdot \phi_X(s_2)}{\phi_X(\alpha s_2)} \\
&= \exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 + \frac{2\mu^2}{\lambda} \cdot (s_1 + \alpha s_2)\right)^{1/2}\right]\right\} \times \exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 + \frac{2\mu^2}{\lambda}s_2\right)^{1/2}\right]\right\} \\
&\quad \times \exp\left\{-\frac{\lambda}{\mu}\left[1 - \left(1 + \frac{2\mu^2}{\lambda} \cdot (\alpha s_2)\right)^{1/2}\right]\right\} \\
&= \exp\left\{\frac{\lambda}{\mu}\left[1 - \left(\left[1 + \frac{2\mu^2}{\lambda}s_1 + \frac{2\alpha\mu^2}{\lambda}s_2\right]^{1/2} + \left[1 + \frac{2\mu^2}{\lambda}s_2\right]^{1/2} - \left[1 + \frac{2\alpha\mu^2}{\lambda}s_2\right]^{1/2}\right)\right]\right\}.
\end{aligned}$$

(4) **(Mittag-Leffler margins)** *Let* $\phi_X(s) = \frac{1}{1+s^\gamma}$, *where* $0 < \gamma < 1$. *Then*

$$\begin{aligned}
\phi_{(X(t_1),X(t_2))}(s_1, s_2) &= \frac{\phi_X(s_1 + \alpha s_2) \cdot \phi_X(s_2)}{\phi_X(\alpha s_2)} \\
&= \frac{1}{1 + (s_1 + \alpha s_2)^\gamma} \cdot \frac{1}{1 + s_2^\gamma} \cdot \left(\frac{1}{1 + (\alpha s_2)^\gamma}\right)^{-1} = \frac{1 + \alpha^\gamma s_2^\gamma}{[1 + (s_1 + \alpha s_2)^\gamma] \cdot (1 + s_2^\gamma)}.
\end{aligned}$$

(5) **(logistic margins)** *Let* $\varphi_X(s) = \prod_{k=1}^{\infty} \frac{1}{(1+s^2/k^2)} = \frac{i\pi s}{\sin(i\pi s)}$. *Then*

$$\begin{aligned}
\varphi_{(X(t_1),X(t_2))}(s_1, s_2) &= \frac{\varphi_X(s_1 + \alpha s_2) \cdot \varphi_X(s_2)}{\varphi_X(\alpha s_2)} \\
&= \frac{i\pi(s_1 + \alpha s_2)}{\sin(i\pi(s_1 + \alpha s_2))} \cdot \frac{i\pi s_2}{\sin(i\pi s_2)} \cdot \frac{i\sin(i\pi\alpha s_2)}{i\pi\alpha s_2} = \frac{i\pi(s_1 + \alpha s_2) \cdot \sin(i\alpha\pi s_2)}{\alpha\sin(i\pi s_1 + i\alpha\pi s_2) \cdot \sin(i\pi s_2)}.
\end{aligned}$$

(6) **(symmetric EGGC margins)** *Let* $\varphi(s) = \exp\left\{-\frac{ds^2}{2} + \int_{(0,\infty)} \log\left(\frac{u^2}{u^2 + s^2}\right) U(du)\right\}$, *where* $d \geq 0$, $U(du)$ *is a symmetric non-negative measure on* $\Re \setminus \{0\}$ *satisfying*

$$\int_{\Re\setminus\{0\}} \frac{1}{1 + u^2} U(du) < \infty \quad \text{and} \quad \int_{|u|\leq 1} |\log u^2| U(du) < \infty.$$

Then

$$\begin{aligned}
\varphi_{(X(t_1),X(t_2))}(s_1, s_2) &= \frac{\varphi_X(s_1 + \alpha s_2) \cdot \varphi_X(s_2)}{\varphi_X(\alpha s_2)} \\
&= \exp\left\{-\frac{d \cdot (s_1 + \alpha s_2)^2}{2} + \int_{(0,\infty)} \log\left(\frac{u^2}{u^2 + (s_1 + \alpha s_2)^2}\right) U(du)\right\}
\end{aligned}$$

$$\times \exp\left\{-\frac{ds_2^2}{2}+\int_{(0,\infty)}\log\left(\frac{u^2}{u^2+s_2^2}\right)U(du)\right\}$$
$$\times \exp\left\{\frac{d\cdot(\alpha s_2)^2}{2}-\int_{(0,\infty)}\log\left(\frac{u^2}{u^2+(\alpha s_2)^2}\right)U(du)\right\}$$
$$=\exp\left\{-\frac{d\cdot(s_1^2+2\alpha s_1 s_2+s_2^2)}{2}+\int_{(0,\infty)}\log\left(\frac{u^2(u^2+\alpha^2 s_2^2)}{[u^2+(s_1+\alpha s_2)^2]\cdot(u^2+s_2^2)}\right)U(du)\right\}.$$

(7) **(stable margins)** *Let $\varphi_X(s)=\exp\{-\lambda|s|^\gamma\}$, where $\lambda>0$ and $0<\gamma\le 2$. Note that when $0<\gamma<1$, X is a positive rv, while for $1\le\gamma\le 2$, X is a real-valued rv. We deal with both situations in the unified form of a cf:*

$$\begin{aligned}\varphi_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{\varphi_X(s_1+\alpha s_2)\cdot\varphi_X(s_2)}{\varphi_X(\alpha s_2)}\\ &=\exp\{-\lambda|s_1+\alpha s_2|^\gamma\}\cdot\exp\{-\lambda|s_2|^\gamma\}\cdot\exp\{\lambda|\alpha s_2|^\gamma\}\\ &=\exp\{-\lambda[|s_1+\alpha s_2|^\gamma+(1-\alpha^\gamma)|s_2|^\gamma]\}.\end{aligned}$$

We end this subsection with two examples: a bivariate GDSD($\mathbf{I2}(\gamma)$) distribution, and a bivariate GSD($\mathbf{P2}(\gamma)$) distribution. They indicate that given marginal distributions, there may exist different families of bivariate distributions.

Example 9.8 *Consider K being from GDSD($\mathbf{I2}(\gamma)$) with pgf $G_K(s;\alpha)=\frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}$, where $0<\gamma<1$. Let the margins be NB(δ,γ) with pgf $G_X(s)=\left(\frac{1-\gamma}{1-\gamma s}\right)^\delta$ $(0<\gamma<1,\delta>0)$. By (9.3.5), we obtain*

$$\begin{aligned}G_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{G_X(s_1G_K(s_2;\alpha))\cdot G_X(s_2)}{G_X(G_K(s_2;\alpha))} = \frac{G_X\left(s_1\frac{(1-\alpha)+(\alpha-\gamma)s_2}{(1-\alpha\gamma)-(1-\alpha)\gamma s_2}\right)\cdot G_X(s_2)}{G_X\left(\frac{(1-\alpha)+(\alpha-\gamma)s_2}{(1-\alpha\gamma)-(1-\alpha)\gamma s_2}\right)}\\ &= \left[\frac{(1-\gamma)\cdot\left(1-\gamma\cdot\frac{(1-\alpha)+(\alpha-\gamma)s_2}{(1-\alpha\gamma)-(1-\alpha)\gamma s_2}\right)}{\left(1-\gamma\cdot s_1\frac{(1-\alpha)+(\alpha-\gamma)s_2}{(1-\alpha\gamma)-(1-\alpha)\gamma s_2}\right)\cdot(1-\gamma s_2)}\right]^\delta\\ &= \left[\frac{(1-\gamma)^2}{(1-\alpha\gamma)-\gamma(1-\alpha)s_1-\gamma(1-\alpha)s_2-\gamma(\alpha-\gamma)s_1s_2}\right]^\delta.\end{aligned}$$

This is different from (2) of Example 9.6.

Example 9.9 *Consider K being from GSD($\mathbf{P2}(\gamma)$) with LT* $\phi_K(s;\alpha) = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\}$, *where $0 \le \gamma < 1$. Let the margins be Gamma(δ, β) with LT* $\phi_X(s) = \left(\frac{\beta}{\beta+s}\right)^\delta$, *where δ, $\beta > 0$, $\gamma \le 1/(1+\beta)$. By (9.3.6), we obtain*

$$
\begin{aligned}
\phi_{(X(t_1),X(t_2))}(s_1,s_2) &= \frac{\phi_X\left(s_1 - \log\phi_K\left(s_2;\alpha\right)\right)\cdot\phi_X(s_2)}{\phi_X\left(-\log\phi_K\left(s_2;\alpha\right)\right)} \\
&= \frac{\phi_X\left(s_1 + \frac{\alpha(1-\gamma)s_2}{(1-\gamma)+(1-\alpha)\gamma s_2}\right)\cdot\phi_X(s_2)}{\phi_X\left(\frac{\alpha(1-\gamma)s_2}{(1-\gamma)+(1-\alpha)\gamma s_2}\right)} = \left[\frac{\beta\cdot\left(\beta + \frac{\alpha(1-\gamma)s_2}{(1-\gamma)+(1-\alpha)\gamma s_2}\right)}{\left(\beta + s_1 + \frac{\alpha(1-\gamma)s_2}{(1-\gamma)+(1-\alpha)\gamma s_2}\right)\cdot(\beta+s_2)}\right]^\delta \\
&= \left[\frac{\beta(1-\gamma) + \left[(1-\alpha)\gamma\beta + \alpha(1-\gamma)\right]\beta s_2}{\left[\beta(1-\gamma) + (1-\gamma)s_1 + \left[(1-\alpha)\beta\gamma + \alpha(1-\gamma)\right]s_2 + (1-\alpha)\gamma s_1 s_2\right]\cdot(\beta+s_2)}\right]^\delta \\
&= \left(\beta(1-\gamma) + \left[(1-\alpha)\gamma\beta + \alpha(1-\gamma)\right]\beta s_2\right)^\delta \\
&\quad \times\Big(\beta^2(1-\gamma) + \beta(1-\gamma)s_1 + \left[(1-\alpha)\beta\gamma + (\beta+\alpha)(1-\gamma)\right]s_2 \\
&\qquad + \left[(1-\alpha)\gamma + (1-\gamma)\right]s_1 s_2 + \left[(1-\alpha)\beta\gamma + \alpha(1-\gamma)\right]s_2^2 + (1-\alpha)\gamma s_1 s_2^2\Big)^{-\delta}.
\end{aligned}
$$

This resulting LT is different from (1) of Example 9.7.

9.3.2 Multivariate distributions

We can directly extend the construction of a bivariate GDSD/GSD distribution to a multivariate GDSD/GSD distribution via a stationary continuous-time GAR(1) process, where we incorporate the univariate margins of the multivariate distribution into the margins of the process at different time points, namely $(X(t_1), X(t_2), \cdots, X(t_n))$ for $t_1 < t_2 < \cdots < t_n$.

For the sake of simpler notation, we use the new notation

$$\alpha_{i,j} = e^{-\mu(t_j - t_i)}, \qquad i,j = 1,2,\ldots,n.$$

Note that $\alpha_{i,i} = 1$ $(i = 1,2,\ldots,n)$. For a self-generalized rv K with pgf $G_K(s;\alpha)$, or LT $\phi_K(s;\alpha)$, define the following recursive notations:

$$\begin{cases} s'_{n-1} &= s_{n-1} G_K(s_n; \alpha_{n-1,n}), \\ s'_{n-2} &= s_{n-2} G_K(s'_{n-1}; \alpha_{n-2,n-1}), \\ &\vdots \\ s'_1 &= s_1 G_K(s'_2; \alpha_{1,2}), \end{cases} \tag{9.3.11}$$

or

$$\begin{cases} s'_{n-1} &= s_{n-1} - \phi_K(s_n; \alpha_{n-1,n}), \\ s'_{n-2} &= s_{n-2} - \phi_K(s'_{n-1}; \alpha_{n-2,n-1}), \\ &\vdots \\ s'_1 &= s_1 - \phi_K(s'_2; \alpha_{1,2}). \end{cases} \tag{9.3.12}$$

When K is from **P1** with LT $\phi_K(s;\alpha) = e^{-\alpha s}$, definition (9.3.12) becomes

$$\begin{cases} s'_{n-1} &= s_{n-1} + \alpha_{n-1,n} s_n, \\ s'_{n-2} &= s_{n-2} + \alpha_{n-2,n-1} s'_{n-1}, \\ &\vdots \\ s'_1 &= s_1 + \alpha_{1,2} s'_2. \end{cases} \tag{9.3.13}$$

With these new notations, we give the following theorem.

Theorem 9.3.3 *Suppose the continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t', t), \qquad t' < t,$$

has stationary distribution with pgf $G_X(s)$, or LT $\phi_X(s)$, or cf $\varphi_X(s)$. $X(t_1), X(t_2), \ldots, X(t_n)$ are the rv's at time points $t_1 < t_2 < \cdots < t_n$ $(n \geq 3)$ respectively.

(1) If X is GDSD associated with self-generalized rv $K(\alpha)$, then

$$\begin{aligned} &G_{(X(t_1),\ldots,X(t_n))}(s_1, \ldots, s_n) \\ &= \frac{\prod_{i=1}^{n-1} G_X(s'_i)}{\prod_{j=2}^{n-1} G_X(G_K(s'_j; \alpha_{j-1,j}))} \times \frac{G_X(s_n)}{G_X(G_K(s_n; \alpha_{n-1,n}))} \end{aligned} \tag{9.3.14}$$

is the pgf of a multivariate GDSD distribution whose univariate marginal distributions are the same as that of X. Here $s'_1, \ldots, s'_{n-1}$ are defined in (9.3.11).

(2) If X is GSD associated with self-generalized rv $K(\alpha)$, then

$$\begin{aligned}
&\phi_{(X(t_1),\dots,X(t_n))}(s_1,\dots,s_n) \\
&= \frac{\prod_{i=1}^{n-1}\phi_X(s_i')}{\prod_{j=2}^{n-1}\phi_X(-\log\phi_K(s_j';\alpha_{j-1,j}))} \times \frac{\phi_X(s_n)}{\phi_X(-\log\phi_K(s_n;\alpha_{n-1,n}))}
\end{aligned} \tag{9.3.15}$$

is the LT of a multivariate GSD distribution whose univariate marginal distributions are the same as that of X. Here $s_1',\dots,s_{n-1}'$ are defined in (9.3.12).

(3) If X is a real SD rv, then

$$\varphi_{(X(t_1),\dots,X(t_n))}(s_1,\dots,s_n) = \frac{\prod_{i=1}^{n-1}\varphi_X(s_i')}{\prod_{j=2}^{n-1}\varphi_X(\alpha_{j-1,j}s_j')} \times \frac{\varphi_X(s_n)}{\varphi_X(\alpha_{n-1,n}s_n)} \tag{9.3.16}$$

is the cf of a multivariate SD distribution whose univariate marginal distributions are the same as that of X. Here $s_1',\dots,s_{n-1}'$ are defined in (9.3.13).

Proof: Similar to the proof of Theorem 9.3.1, we have:

$$\begin{aligned}
G_{E(t',t)}(s_2) &= \frac{G_X(s_2)}{G_X\left(G_K\left(s_2;e^{-\mu(t-t')}\right)\right)}, \\
\phi_{E(t',t)}(s_2 &= \frac{\phi_X(s_2)}{\phi_X\left(-\log\phi_K\left(s_2;e^{-\mu(t-t')}\right)\right)}, \\
\varphi_{E(t',t)}(s_2) &= \frac{\varphi_X(s_2)}{\varphi_X\left(e^{-\mu(t-t')}s_2\right)},
\end{aligned}$$

for any $t' < t$. Under steady state, the marginal distributions are the same as the stationary distribution of X. Hence, the joint distribution of $(X(t_1), X(t_2),\dots,X(t_n))$ has the same distribution for every margin. By the Markov property, we can obtain

$$\begin{aligned}
&G_{(X(t_1),\dots,X(t_n))}(s_1,\dots,s_n) = \mathbf{E}\left(s_1^{X(t_1)}s_2^{X(t_2)}\cdots s_n^{X(t_n)}\right) \\
&= \mathbf{E}\left(s_1^{X(t_1)}\cdots s_{n-1}^{X(t_{n-1})}\mathbf{E}\left(s_n^{X(t_n)}\middle| X(t_{n-1})\right)\right) \\
&= \mathbf{E}\left(s_1^{X(t_1)}\cdots s_{n-1}^{X(t_{n-1})}\cdot G_K^{X(t_{n-1})}(s_n;\alpha_{n-1,n})\cdot G_{E(t_{n-1},t_n)}(s_n)\right) \\
&= \mathbf{E}\left(s_1^{X(t_1)}\cdots (s_{n-1}')^{X(t_{n-1})}\right)\cdot G_{E(t_{n-1},t_n)}(s_n) \\
&= \mathbf{E}\left(s_1^{X(t_1)}\cdots (s_{n-2}')^{X(t_{n-2})}\right)\cdot G_{E(t_{n-2},t_{n-1})}(s_{n-1}')\cdot G_{E(t_{n-1},t_n)}(s_n) \\
&\vdots
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{E}\left((s_1')^{X(t_1)}\right) \cdot G_{E(t_1,t_2)}(s_2') \cdots G_{E(t_{n-2},t_{n-1})}(s_{n-1}') \cdot G_{E(t_{n-1},t_n)}(s_n) \\
&= G_X(s_1') \times \prod_{i=2}^{n-1} \frac{G_X(s_i')}{G_X(G_K(s_j'; \alpha_{j-1,j}))} \times \frac{G_X(s_n)}{G_X(G_K(s_n; \alpha_{n-1,n}))} \\
&= \frac{\prod_{i=1}^{n-1} G_X(s_i')}{\prod_{j=2}^{n-1} G_X(G_K(s_j'; \alpha_{j-1,j}))} \times \frac{G_X(s_n)}{G_X(G_K(s_n; \alpha_{n-1,n}))}.
\end{aligned}
$$

This completes the proof of case (1). Using the same reasoning, it is straightforward to prove cases (2) and (3).

Although we can obtain the pgf, or LT, or cf of a multivariate GDSD/GSD distribution by Theorem 9.3.3, the simplification of the resulting multivariate pgf, or LT, or cf to a direct expression in terms of $s_1, s_2, \ldots, s_n$ is very challenging in most situations. Sometimes, symbolic software like Maple can help us. Three examples are listed below, where tedious induction details are omitted.

Example 9.10 (Multivariate Poisson) *Consider Poisson margins, where $G_X(s) = e^{\lambda(s-1)}$ ($\lambda > 0$), and K is from* **I1** *with pgf $G_K(s;\alpha) = 1 - \alpha + \alpha s$. Define*

$$
\begin{aligned}
\beta_0 = \beta_n = 0, \quad \beta_i = \alpha_{i,i+1} = e^{-\mu(t_{i+1}-t_i)}, &\qquad i = 1, \ldots, n; \\
\lambda_{\{i\}} = \lambda(1-\beta_i)(1-\beta_{i-1}), &\qquad i = 1, \ldots, n; \\
\lambda_{\{i,\ldots,i+j\}} = \lambda \beta_i \cdots \beta_{i+j-1}(1-\beta_{i-1})(1-\beta_{i+j}), &\qquad i = 1, \ldots, n-j; \quad j = 1, \ldots, n-1.
\end{aligned}
$$

By (9.3.14) in Theorem 9.3.3, we can obtain

$$
\begin{aligned}
&G_{(X(t_1),\ldots,X(t_n))}(s_1, \ldots, s_n) \\
&= \exp\left\{ \sum_{i=1}^{n} \lambda_{\{i\}}(s_i - 1) + \cdots + \sum_{i=1}^{n-j} \lambda_{\{i,\ldots,i+j\}}(s_i \cdots s_{i+j} - 1) + \cdots + \lambda_{\{1,\ldots,n\}}(s_1 \cdots s_n - 1) \right\}.
\end{aligned}
$$

The stochastic representation is

$$
(X(t_1), \ldots, X(t_n)) \stackrel{d}{=} \Big(\sum_{S:1\in S} Y_S, \ldots, \sum_{S:n\in S} Y_S \Big),
$$

where the summations are over nonempty subsets S of $\{1, \ldots, n\}$, $Y_S \sim$ Poisson(λ_S), $\lambda_S = 0$ if S is not a subset of consecutive integers (corresponding $Y_S = 0$), and λ_S is defined as in the beginning otherwise. See Johnson, Kotz and Balakrishnan (1997) for this multivariate Poisson distribution.

Example 9.11 (Multivariate NB) *Consider the NB margins, where* $G_X(s) = \left(\frac{1-\gamma}{1-\gamma s}\right)^\delta$ $(0 < \gamma < 1,\ \delta > 0)$.

(1) K is from **I1** *with pgf* $G_K(s;\alpha) = 1 - \alpha + \alpha s$.

Define the new recursive notations:

$$\begin{aligned}
s_j(1) &= (1-\alpha_{j-1,j}) + \alpha_{j-1,j}s_j,\\
s_j(2) &= (1-\alpha_{j-2,j-1}) + \alpha_{j-2,j-1}s_j(1)\\
&= (1-\alpha_{j-2,j-1}) + \alpha_{j-2,j-1}[(1-\alpha_{j-1,j}) + \alpha_{j-1,j}s_j]\\
&= (1-\alpha_{j-2,j-1}\alpha_{j-1,j}) + \alpha_{j-2,j-1}\alpha_{j-1,j}s_j\\
&= (1-\alpha_{j-2,j}) + \alpha_{j-2,j}s_j,\\
s_j(3) &= (1-\alpha_{j-3,j-2}) + \alpha_{j-3,j-2}s_j(2)\\
&= (1-\alpha_{j-3,j-2}) + \alpha_{j-3,j-2}[(1-\alpha_{j-2,j}) + \alpha_{j-2,j}s_j]\\
&= (1-\alpha_{j-3,j}) + \alpha_{j-3,j}s_j,\\
&\ \vdots\\
s_j(l) &= (1-\alpha_{j-l,j-l+1}) + \alpha_{j-l,j-l+1}s_j(l-1)\\
&= (1-\alpha_{j-l,j}) + \alpha_{j-l,j}s_j.
\end{aligned}$$

Then by (9.3.14) in Theorem 9.3.3, we obtain

$$\begin{aligned}
G_{(X(t_1),\ldots,X(t_n))}(s_1,\ldots,s_n) &= \mathbf{E}[s_1^{X(t_1)}\cdots s_n^{X(t_n)}]\\
&= \mathbf{E}[s_1^{X(t_1)}s_2^{\alpha_{1,2}*X(t_1)+E(t_1,t_2)}\cdots s_n^{\alpha_{n-1,n}*X(t_{n-1})+E(t_{n-1},t_n)}]\\
&= \mathbf{E}[s_2^{E(t_1,t_2)}]\times\cdots\times\mathbf{E}[s_n^{E(t_{n-1},t_n)}]\times\mathbf{E}[s_1^{X(t_1)}s_2^{\alpha_{1,2}*X(t_1)}\cdots s_n^{\alpha_{n-1,n}*X(t_{n-1})}]\\
&= \mathbf{E}[s_2^{E(t_1,t_2)}]\times\cdots\times\mathbf{E}[s_n^{E(t_{n-1},t_n)}]\\
&\qquad\times\mathbf{E}\left\{s_1^{X(t_1)}[(1-\alpha_{1,2})+\alpha_{1,2}s_2]^{X(t_1)}\cdots[(1-\alpha_{n-1,n})+\alpha_{n-1,n}s_n]^{X(t_{n-1})}\right\}\\
&= \mathbf{E}[s_2^{E(t_1,t_2)}]\times\cdots\times\mathbf{E}[s_n^{E(t_{n-1},t_n)}]\mathbf{E}\left\{s_1^{X(t_1)}s_2(1)^{X(t_1)}\cdots s_n(1)^{X(t_{n-1})}\right\}\\
&= \mathbf{E}[s_2^{E(t_1,t_2)}]\times\cdots\times\mathbf{E}[s_n^{E(t_{n-1},t_n)}]\times\mathbf{E}[s_3(1)^{E(t_1,t_2)}]\cdots\mathbf{E}[s_n(1)^{E(t_{n-1},t_n)}]\\
&\qquad\times\mathbf{E}\left\{s_1^{X(t_1)}s_2(1)^{X(t_1)}s_3(2)^{X(t_1)}s_4(2)^{X(t_1)}\cdots s_n(2)^{X(t_1)}\right\}
\end{aligned}$$

$$
\begin{aligned}
&\vdots\\
&= \mathbf{E}\left\{[s_2 s_3(1)s_4(2)\cdots s_n(n-2)]^{E(t_1,t_2)}\right\} \times \mathbf{E}\left\{[s_3 s_4(1)s_5(2)\cdots s_n(n-3)]^{E(t_2,t_3)}\right\} \times \cdots\\
&\qquad \times \mathbf{E}\left\{[s_{n-1}s_n(1)]^{E(t_{n-2},t_{n-1})}\right\} \times \mathbf{E}\left\{s_n^{E(t_{n-1},t_n)}\right\}\\
&\qquad \times \mathbf{E}\left\{[s_1 s_2(1)s_3(2)\cdots s_n(n-1)]^{X(t_1)}\right\}\\
&= \mathbf{E}\left\{\left(s_1\prod_{i=2}^{n}[(1-\alpha_{1,i})+\alpha_{1,i}s_i]\right)^{X(t_1)}\right\} \times \mathbf{E}\left\{\left(s_2\prod_{i=3}^{n}[(1-\alpha_{2,i})+\alpha_{2,i}s_i]\right)^{E(t_1,t_2)}\right\}\\
&\qquad \times\cdots\times \mathbf{E}\left\{(s_{n-1}[(1-\alpha_{n-1,n})+\alpha_{n-1,n}s_n])^{E(t_{n-2},t_{n-1})}\right\} \times \mathbf{E}\left\{s_n^{E(t_{n-1},t_n)}\right\}\\
&= \left\{\frac{1-\gamma}{1-\gamma s_1\prod_{i=2}^{n}[(1-\alpha_{1,i})+\alpha_{1,i}s_i]}\right\}^{\delta}\\
&\qquad \times\left\{\frac{(1-\gamma+\alpha_{1,2}\gamma)-\alpha_{1,2}\gamma s_2\prod_{i=3}^{n}[(1-\alpha_{2,i})+\alpha_{2,i}s_i]}{1-\gamma s_2\prod_{i=3}^{n}[(1-\alpha_{2,i})+\alpha_{2,i}s_i]}\right\}^{\delta} \times\cdots\\
&\qquad \times\left\{\frac{(1-\gamma+\alpha_{n-2,n-1}\gamma)-\alpha_{n-2,n-1}\gamma s_{n-1}[(1-\alpha_{n-1,n})+\alpha_{n-1,n}s_n]}{1-\gamma s_{n-1}[(1-\alpha_{n-1,n})+\alpha_{n-1,n}s_n]}\right\}^{\delta}\\
&\qquad \times\left\{\frac{(1-\gamma+\alpha_{n-1,n}\gamma)-\alpha_{n-1,n}\gamma s_n}{1-\gamma s_n}\right\}^{\delta}.
\end{aligned}
$$

(2) K is from **I2** *with pgf* $G_K(s;\alpha) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}$ $(0 < \gamma < 1)$. *Here* γ *is the same as a parameter in the pgf of NB margin.*

Define

$$
\begin{aligned}
c_{i_1\cdots c_n} &= [I(i_1=0)-\gamma I(i_1=1)]\prod_{j=2}^{n}[(1-\gamma\alpha_{j-1,j})I(i_j=i_{j-1}=0)\\
&\quad +(1-\alpha_{j-1,j})I(i_{j-1}=1,i_j=0)-\gamma(1-\alpha_{j-1,j})I(i_{j-1}=0,i_j=1)\\
&\quad +(\alpha_{j-1,j}-\gamma)I(i_j=i_{j-1}=1)].
\end{aligned}
$$

By (9.3.14) in Theorem 9.3.3 and induction with the help of symbolic manipulation, we obtain

$$
G_{(X(t_1),\ldots,X(t_n))}(s_1,\ldots,s_n) = \mathbf{E}[s_1^{X(t_1)}\cdots s_n^{X(t_n)}] = \left[\frac{(1-\gamma)^n}{\sum_{i_1=0}^{1}\cdots\sum_{i_n=0}^{1}c_{i_1\cdots c_n}s_1^{i_1}\cdots s_n^{i_n}}\right]^{\delta}.
$$

This resulting multivariate distribution of $(X(t_1),\ldots,X(t_n))$ *is in the multivariate negative binomial family given in Doss (1979).*

Part IV

Statistical inference and applications

Chapter 10

Parameter estimation

In this chapter, we study parameter estimation based on observation of a GAR(1) process at a finite number of time points. Usually in a real problem, after we decide on the potential models for the observed data, the next step is to estimate the model parameter values based on the observations. This procedure is called parameter estimation. The estimation approaches for the parameters in the continuous-time GAR(1) process include: maximum likelihood, conditional least squares, empirical characteristic function, as well as the method of moments and miscellaneous for special cases. We wish to pursue some closed form estimates for easier computations. However, in general, there are no closed form expressions for the estimates, in which case we will use numerical methods to find the estimates.

We shall investigate maximum likelihood estimates (MLE) and conditional maximum likelihood estimates (CMLE) in Section 10.1. In Section 10.2, we discuss the conditional least squares estimates (CLS) and variations such as the conditional weighted least squares estimates (CWLS), quasi-conditional least squares estimates (QCLS) and conditional generalized least squares (CGLS). The empirical characteristic function (ECF) approach and variations are studied in Section 10.3. We shall discuss the method of moments, as well as miscellaneous methods for the continuous-time GAR(1) processes in Section 10.4. Numerical methods are mentioned in Section 10.5.

10.1 Maximum likelihood estimation

The maximum likelihood estimation approach is a conventional method, usually used in the models where the distribution of the sample $(X_1, X_2, \ldots, X_n)$ is clearly specified. One of the advantages of this approach is that it is often most efficient (assuming the model is correct).

Let vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ denote the data observed at time points $t_1 < t_2 < \cdots < t_n$ from a continuous-time GAR(1) process $\{X(t); t \geq 0\}$, namely

$$X(t_1) = x_1,\ X(t_2) = x_2,\ \ldots,\ X(t_n) = x_n.$$

Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$ be the vector of all parameters to be estimated in the continuous-time GAR(1) process model. One of the important parameters is the dependence intensity μ (see Chapter 4), which we put into the first argument position in $\boldsymbol{\theta}$, i.e., $\theta_1 = \mu$. Let Θ be the parameter space. Usually, the parameter space is the subset in $\Re^k$. For example, $\theta_1 = \mu > 0$. The sample $\mathbf{x}$ can be viewed from the joint distribution of $(X(t_1), X(t_2), \ldots, X(t_n))$. Then the likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta} \mid \mathbf{x}) &= f_{(X(t_1),X(t_2),\ldots,X(t_n))}(x_1, x_2, \ldots, x_n; \boldsymbol{\theta}) \\ &= f_{X(t_1)}(x_1; \boldsymbol{\theta}) f_{X(t_2)|X(t_1)}(x_2 \mid x_1; \boldsymbol{\theta}) \cdots f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1}; \boldsymbol{\theta}). \end{aligned} \tag{10.1.1}$$

Here f denotes the pmf in the discrete case or pdf in the continuous case. Because it is a product, we take logarithm and obtain the log-likelihood function:

$$\begin{aligned} \log L(\boldsymbol{\theta} \mid \mathbf{x}) &= \log f_{X(t_1)}(x_1; \boldsymbol{\theta}) + \log f_{X(t_2)|X(t_1)}(x_2 \mid x_1; \boldsymbol{\theta}) + \cdots \\ &\quad + \log f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1}; \boldsymbol{\theta}), \end{aligned} \tag{10.1.2}$$

The MLE of $\boldsymbol{\theta}$, denoted as $\widehat{\boldsymbol{\theta}}_{MLE}$, is then the value of parameter $\boldsymbol{\theta}$ where the likelihood (or log-likelihood) function reaches maximum, namely

$$\widehat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\theta \in \Theta} L(\boldsymbol{\theta} \mid \mathbf{x}) = \arg\max_{\theta \in \Theta} \log L(\boldsymbol{\theta} \mid \mathbf{x}).$$

By taking partial derivatives on log-likelihood function with respect to parameter $\boldsymbol{\theta}$, in general, we can obtain the score or MLE equations by equating them to zeros:

$$\begin{cases} \frac{\partial \log L(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_1} &= 0, \\ & \vdots \\ \frac{\partial \log L(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_k} &= 0. \end{cases}$$

Assume the MLE is not on the boundary of Θ. Solving these equations will lead to the MLE $\widehat{\boldsymbol{\theta}}_{MLE}$. By calculus, it follows that

$$\begin{aligned} \frac{\partial \log L(\boldsymbol{\theta} \mid \mathbf{x})}{\partial \theta_j} &= \frac{1}{f_{X(t_1)}(x_1;\boldsymbol{\theta})} \frac{\partial f_{X(t_1)}(x_1;\boldsymbol{\theta})}{\partial \theta_j} + \frac{1}{f_{X(t_2)|X(t_1)}(x_2 \mid x_1;\boldsymbol{\theta})} \frac{\partial f_{X(t_2)|X(t_1)}(x_2 \mid x_1;\boldsymbol{\theta})}{\partial \theta_j} \\ & \quad + \cdots + \frac{1}{f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1};\boldsymbol{\theta})} \frac{\partial f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1};\boldsymbol{\theta})}{\partial \theta_j} \\ &= 0, \end{aligned}$$

for $j = 1, 2, \ldots, k$.

Sometimes, the marginal distribution of $X(t_1)$ is not specified or is not of primary interest. In such a situation, we only focus on the conditional structure of the process, and view the sample $(x_2, x_3, \ldots, x_n)$ as a realization of $(Y_1, Y_2, \ldots, Y_{n-1})$, where

$$Y_i \stackrel{d}{=} [X(t_{i+1}) \mid X(t_i) = x_i], \qquad i = 1, 2, \ldots, n-1.$$

Note that these Y_i $(i = 1, 2, \ldots, n-1)$ are independent rv's. We can maximize the conditional likelihood function of $(X(t_2), X(t_3), \ldots, X(t_n))$ conditioned on $X(t_1) = x_1$, which is the likelihood function of $(Y_1, Y_2, \ldots, Y_{n-1})$,

$$\begin{aligned} L_1(\boldsymbol{\theta} \mid \mathbf{x}) &= f_{(X(t_2),\ldots,X(t_n))|X(t_1)}(x_2, \ldots, x_n \mid x_1;\boldsymbol{\theta}) \\ &= f_{X(t_2)|X(t_1)}(x_2 \mid x_1;\boldsymbol{\theta}) \cdots f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1};\boldsymbol{\theta}), \end{aligned} \tag{10.1.3}$$

or its logarithm

$$\log L_1(\boldsymbol{\theta} \mid \mathbf{x}) = \log f_{X(t_2)|X(t_1)}(x_2 \mid x_1;\boldsymbol{\theta}) + \cdots + \log f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1};\boldsymbol{\theta}), \tag{10.1.4}$$

to obtain the conditional maximum likelihood estimate (CMLE)

$$\widehat{\boldsymbol{\theta}}_{CMLE} = \arg\max_{\theta \in \Theta} L_1(\boldsymbol{\theta} \mid \mathbf{x}) = \arg\max_{\theta \in \Theta} \log L_1(\boldsymbol{\theta} \mid \mathbf{x}).$$

This will lead to

$$
\begin{aligned}
\frac{\partial \log L_1(\boldsymbol{\theta} \mid \mathbf{x})}{\partial \theta_j} &= \frac{1}{f_{X(t_2)|X(t_1)}(x_2 \mid x_1; \boldsymbol{\theta})} \frac{\partial f_{X(t_2)|X(t_1)}(x_2 \mid x_1; \boldsymbol{\theta})}{\partial \theta_j} + \cdots \\
&\quad + \frac{1}{f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1}; \boldsymbol{\theta})} \frac{\partial f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1}; \boldsymbol{\theta})}{\partial \theta_j} \\
&= 0,
\end{aligned}
$$

for $j = 1, 2, \ldots, k$. This simply drops off the term $f_{X(t_1)}(x_1)$ in (10.1.1) and (10.1.2). It is fine when the sample size n is large, because then the term $f_{X(t_1)}(x_1)$ has less influence in the MLE so that the difference between MLE and CMLE is very small.

The conditional pmf or pdf $f_{X(t_{i+1})|X(t_i)}$ $(i = 1, 2, \ldots, n-1)$ has been discussed in Section 9.2. Usually, they do not have explicit forms of expression. Hence, we can not obtain the explicit form of MLE or CMLE. Even the numerical solution of the MLE or CMLE could be a new challenge. Based on these difficulties, the maximum likelihood approach may not be the first choice in parameter estimation unless other approaches are not good enough or one is particularly attracted by the asymptotic efficiency of the MLE or CMLE.

The maximum likelihood approach can be used in either the stationary or non-stationary situation if the relevant distributions are known.

Lastly, we briefly discuss the numerical approach to obtain the MLE or CMLE when the explicit form of likelihood is not available, but a closed form exists for the conditional pgf or LT.

The key issue in this situation is how to calculate the pmf or pdf by the characteristic function. Theoretically, the cdf $F_Z(z)$ can be computed by the inversion of characteristic function $\varphi_Z(s)$. Lèvy's inversion theorem shows that

$$F_Z(z) - F_Z(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1 - e^{-isz}}{is} \cdot \varphi_Z(s) ds.$$

Gil-Pelaez [1951] gave a new version which has computational convenience:

$$F_Z(z) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{e^{isz}\varphi_Z(-s) - e^{-isz}\varphi_Z(s)}{is} ds. \tag{10.1.5}$$

Davies [1973] changed the form to

$$F_Z(z) = \frac{1}{2} - \int_{-\infty}^{\infty} \mathrm{Im}\left(\frac{e^{-isz}\varphi_Z(s)}{2\pi s}\right) ds \tag{10.1.6}$$

for the continuous real-valued case and to

$$F_Z(z) \quad = \quad \frac{1}{2} - \int_{-\pi}^{\pi} \mathrm{Re}\left(\frac{e^{-is(z+1)}\varphi_Z(s)}{2\pi\left[1-e^{-is(z+1)}\right]}\right) ds \tag{10.1.7}$$

for the discrete integer-valued case, then proposed a numerical approximation to the Gil-Pelaez's inversion theorem when the expectation exists. But, since only a one-dimension integral is involved, it is feasible and better to apply a numerical integration method and obtain the corresponding probability. Bohman [1970, 1972, 1975] gave a couple of approximations for different situations, especially for the case of non-negative support.

These numerical techniques will play a promising role in finding MLE or CMLE numerically. In addition, in a simulation study, we need to generate the samples from the continuous-time GAR(1) processes, the numerical inversion of characteristic function can work well when the explicit form of conditional cdf is not available, or when a simple stochastic representation has not been discovered.

10.2 Conditional least squares estimation and variations

The least squares approach is another conventional method. It does not depend on the full specification of distributions, instead, it just uses the means and/or variances. Hence, the estimates from the least squares approach can correspond to a class of distributions which have the same means and/or variances. Compared with the maximum likelihood approach, the least squares approach usually does not provide estimates as efficient as MLE. However, it usually has a computational advantage and a simple interpretation.

The conditional least squares approach is specifically suitable for Markov process. It focuses on the conditional mean and/or variance structure of the remaining observations given the first one like the treatment in conditional maximum likelihood estimation. Thus, the sample $x_2, x_3, \ldots, x_n$ are viewed from the independent rv $Y_1, Y_2, \ldots, Y_{n-1}$, where

$$Y_i \stackrel{d}{=} [X(t_{i+1}) \mid X(t_i) = x_i], \qquad i = 1, 2, \ldots, n-1.$$

It considers

$$R_{CLS}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \Big(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\Big)^2, \tag{10.2.1}$$

and the conditional least squares estimate (CLS) is defined as

$$\widehat{\boldsymbol{\theta}}_{CLS} = \arg\min_{\theta \in \Theta} R_{CLS}(\boldsymbol{\theta}).$$

By taking partial derivatives and equating them to zero, we can obtain estimating equations

$$\begin{cases} \sum_{i=1}^{n-1} \Big(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\Big) \frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_1} & = & 0, \\ & & \vdots \\ \sum_{i=1}^{n-1} \Big(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\Big) \frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_k} & = & 0. \end{cases}$$

The solution of these equations will be the conditional least squares estimate $\widehat{\boldsymbol{\theta}}_{CLS}$.

Consider the stationary continuous-time GAR(1) process with marginal mean function $A(\boldsymbol{\theta})$:

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t', t), \qquad t' < t.$$

Let

$$\alpha_i = \mathbf{E}\left[K\left(e^{-\mu(t_{i+1}-t_i)}\right)\right], \qquad \text{where } i = 1, 2, \ldots, n-1.$$

By (9.2.1), we have that for $i = 1, 2, \ldots, n-1$,

$$\begin{aligned} \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}] &= x_i \cdot \mathbf{E}\left[K\left(e^{-\mu(t_{i+1}-t_i)}\right)\right] + A(\boldsymbol{\theta}) \cdot \left(1 - \mathbf{E}\left[K\left(e^{-\mu(t_{i+1}-t_i)}\right)\right]\right), \\ &= x_i \alpha_i + A(\boldsymbol{\theta}) \left[1 - \alpha_i\right]. \end{aligned}$$

If K has finite mean, then, by Theorem 3.2.1,

$$\alpha_i = e^{-r\mu(t_{i+1}-t_i)}, \qquad \text{for some } r > 0.$$

In many situations, α_i is only related to the parameter μ. For example, when K is from **I1, I2, I4, P1, P2,** and **P4**, $r = 1$ and $\alpha_i = e^{-\mu(t_{i+1}-t_i)}$.

Now we consider that α_i involves only the parameter μ, namely θ_1 in $\boldsymbol{\theta}$. Thus,

$$\begin{aligned}
\frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_1} &= x_i \frac{\partial \alpha_i}{\partial \mu} + [1-\alpha_i] \frac{\partial A(\boldsymbol{\theta})}{\partial \mu} - A(\boldsymbol{\theta}) \frac{\partial \alpha_i}{\partial \mu}, \\
\frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_2} &= [1-\alpha_i] \frac{\partial A(\boldsymbol{\theta})}{\partial \theta_2}, \\
&\vdots \\
\frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_k} &= [1-\alpha_i] \frac{\partial A(\boldsymbol{\theta})}{\partial \theta_k},
\end{aligned}$$

and the CLS equations simplify to

$$\left\{ \begin{array}{ll} \sum\limits_{i=1}^{n-1} \Big(x_{i+1} - x_i \alpha_i - A(\boldsymbol{\theta})[1-\alpha_i] \Big) \left(x_i - A(\boldsymbol{\theta}) \right) \frac{\partial \alpha_i}{\partial \mu} & = 0, \\ \sum\limits_{i=1}^{n-1} \Big(x_{i+1} - x_i \alpha_i - A(\boldsymbol{\theta})[1-\alpha_i] \Big) [1-\alpha_i] & = 0. \end{array} \right. \tag{10.2.2}$$

When the data are observed at equally-spaced time points, a simpler result can be deduced. In this situation, let $\Delta = t_2 - t_1 = \cdots = t_n - t_{n-1}$, and all the α_i $(i = 1, 2, \ldots, n-1)$ are the same as $\alpha = \mathbf{E}\left[K\left(e^{-\mu\Delta}\right)\right]$. Then minimizing $\sum\limits_{i=1}^{n-1} \Big(x_{i+1} - x_i \alpha - A(\boldsymbol{\theta})\,[1-\alpha] \Big)^2$ is equivalent to find the regression line, hence, we obtain

$$\left\{ \begin{array}{lll} \widehat{\alpha}_{CLS} & = & \dfrac{\sum\limits_{i=1}^{n-1} x_i x_{i+1} - \frac{1}{n-1} \left(\sum\limits_{i=1}^{n-1} x_i \right) \left(\sum\limits_{i=1}^{n-1} x_{i+1} \right)}{\sum\limits_{i=1}^{n-1} x_i^2 - \frac{1}{n-1} \left(\sum\limits_{i=1}^{n-1} x_i \right)^2}, \\ \widehat{A(\boldsymbol{\theta})}_{CLS} & = & \dfrac{\frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_{i+1} - \left(\frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_i \right) \widehat{\alpha}}{1-\widehat{\alpha}}. \end{array} \right. \tag{10.2.3}$$

Note that $\widehat{\alpha}$ can be written

$$\widehat{\alpha}_{CLS} = \frac{\sum\limits_{i=1}^{n-1} \left(x_i - \frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_i \right) \left(x_{i+1} - \frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_{i+1} \right)}{\sum\limits_{i=1}^{n-1} \left(x_i - \frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_i \right)^2}.$$

If we arrange the generalized time series data $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_{n-1}, x_n)$ into lag-1 pairs:

$$(x_1, x_2),\ (x_2, x_3),\ \ldots,\ (x_{n-1}, x_n),$$

then we can calculated sample lag-1 auto-correlation coefficient by

$$\widehat{\rho}_1 = \frac{\sum\limits_{i=1}^{n-1} \left(x_i - \frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_i \right) \left(x_{i+1} - \frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_{i+1} \right)}{\sqrt{\sum\limits_{i=1}^{n-1} \left(x_i - \frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_i \right)^2} \sqrt{\sum\limits_{i=1}^{n-1} \left(x_{i+1} - \frac{1}{n-1} \sum\limits_{i=1}^{n-1} x_{i+1} \right)^2}}, \tag{10.2.4}$$

which is always within $[-1,1]$. Thus

$$\widehat{\alpha} = \frac{\sqrt{\sum_{i=1}^{n-1}\left(x_{i+1} - \frac{1}{n-1}\sum_{i=1}^{n-1} x_{i+1}\right)^2}}{\sqrt{\sum_{i=1}^{n-1}\left(x_i - \frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right)^2}} \times \widehat{\rho}_1.$$

When the sample size n is large, $\sum_{i=1}^{n-1}\left(x_{i+1} - \frac{1}{n-1}\sum_{i=1}^{n-1} x_{i+1}\right)^2$ will be close to $\sum_{i=1}^{n-1}\left(x_i - \frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right)^2$; thus, $\widehat{\alpha}$ will be close to $\widehat{\rho}_1$. This gives an explanation for $\widehat{\alpha}$: it is a modified sample lag-1 autocorrelation coefficient.

From the two equations in (10.2.3), we may obtain $\widehat{\boldsymbol{\theta}}_{CLS}$. However, sometimes the estimated value $\widehat{\alpha}$ or $\widehat{A(\boldsymbol{\theta})}$ may lie outside of the range of α or $A(\boldsymbol{\theta})$, then, the conditional least squares approach won't work. Note that for the discussed self-generalized rv K, the expectation form of α_i can be found in Table 9.1.

The following generalized time series examples are the applications of (10.2.3).

Example 10.1 (Poisson univariate margins) *Consider the time series from the stationary continuous-time GAR(1) process*

$$X(t) \overset{d}{=} e^{-\mu(t-t')} * X(t') + E(t',t), \qquad t' < t,$$

which have Poisson(λ) margin. In this case, the parameter vector $\boldsymbol{\theta} = (\mu, \lambda)'$, and marginal mean function of the process is $A(\boldsymbol{\theta}) = \lambda$. For binomial-thinning, $\alpha = e^{-\mu\Delta}$. According to (10.2.3), we have

$$\begin{cases} \widehat{\alpha}_{CLS} &= \dfrac{\sum_{i=1}^{n-1} x_i x_{i+1} - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)\left(\sum_{i=1}^{n-1} x_{i+1}\right)}{\sum_{i=1}^{n-1} x_i^2 - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)^2} \quad (= e^{-\widehat{\mu}\Delta}), \\ \widehat{\lambda}_{CLS} &= \dfrac{\frac{1}{n-1}\sum_{i=1}^{n-1} x_{i+1} - \left(\frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right)\widehat{\alpha}}{1-\widehat{\alpha}}. \end{cases}$$

Note that $0 < e^{-\mu\Delta} < 1$ and $\lambda > 0$. Hence, $\widehat{\alpha}$ should be in $(0,1)$ and $\widehat{\lambda} > 0$. When $0 < \widehat{\alpha} < 1$, we can further obtain $\widehat{\mu}_{CLS} = -\frac{\log \widehat{\alpha}}{\Delta}$.

Example 10.2 (Geometric univariate margins) *Consider the time series with Geometric(γ) univariate margins, where $0 < \gamma < 1$. They can be from either the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} e^{-\mu(t-t')} * X(t') + E(t', t), \qquad t' < t,$$

or the stationary continuous-time GAR(1) process

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t', t), \qquad t' < t,$$

where K is from **I2** *with pgf* $G_K(s;\alpha) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}$. *Here γ is the same as a parameter in the marginal distribution.*

Then $\boldsymbol{\theta} = (\mu, \gamma)'$, $A(\boldsymbol{\theta}) = \frac{\gamma}{1-\gamma}$. *For K being from either* **I1** *or* **I2**, $\alpha = e^{-\mu\Delta}$. *Therefore, according to (10.2.3), we have*

$$\begin{cases} \widehat{\alpha}_{CLS} & = \dfrac{\sum\limits_{i=1}^{n-1} x_i x_{i+1} - \frac{1}{n-1}\left(\sum\limits_{i=1}^{n-1} x_i\right)\left(\sum\limits_{i=1}^{n-1} x_{i+1}\right)}{\sum\limits_{i=1}^{n-1} x_i^2 - \frac{1}{n-1}\left(\sum\limits_{i=1}^{n-1} x_i\right)^2} \quad (= e^{-\widehat{\mu}\Delta}), \\ \widehat{A(\boldsymbol{\theta})}_{CLS} & = \dfrac{\frac{1}{n-1}\sum\limits_{i=1}^{n-1} x_{i+1} - \left(\frac{1}{n-1}\sum\limits_{i=1}^{n-1} x_i\right)\widehat{\alpha}}{1-\widehat{\alpha}} \quad (= \frac{\widehat{\gamma}}{1-\widehat{\gamma}}). \end{cases}$$

If $\widehat{\alpha}$ is in $(0,1)$ and $\widehat{A(\boldsymbol{\theta})} > 0$, then we can further obtain

$$\begin{cases} \widehat{\mu}_{CLS} & = -\frac{\log \widehat{\alpha}}{\Delta}, \\ \widehat{\gamma}_{CLS} & = \frac{\widehat{A(\boldsymbol{\theta})}_{CLS}}{1+\widehat{A(\boldsymbol{\theta})}_{CLS}}. \end{cases}$$

Example 10.3 (Exponential univariate margins) *Consider the time series with exponential(β) univariate margins, where $\beta > 0$. They can be from either the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} e^{-\mu(t-t')} \bullet X(t') + E(t', t), \qquad t' < t,$$

or the stationary continuous-time GAR(1) process

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t', t), \qquad t' < t,$$

where K is from **P2** *with LT* $\phi_K(s;\alpha) = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\}$. *Here* $\gamma = \frac{1}{1+\beta}$.

Then $\boldsymbol{\theta} = (\mu, \beta)'$, $A(\boldsymbol{\theta}) = \beta^{-1}$. *For* K *being from either* **P1** *or* **P2**, $\alpha = e^{-\mu\Delta}$. *Therefore, according to (10.2.3), we have*

$$\begin{cases} \widehat{\alpha}_{CLS} & = \dfrac{\sum_{i=1}^{n-1} x_i x_{i+1} - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)\left(\sum_{i=1}^{n-1} x_{i+1}\right)}{\sum_{i=1}^{n-1} x_i^2 - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)^2} \quad (= e^{-\widehat{\mu}\Delta}), \\ \widehat{A(\boldsymbol{\theta})}_{CLS} & = \dfrac{\frac{1}{n-1}\sum_{i=1}^{n-1} x_{i+1} - \left(\frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right)\widehat{\alpha}}{1-\widehat{\alpha}} \quad (= \frac{1}{\widehat{\beta}}). \end{cases}$$

If $\widehat{\alpha}$ *is in* $(0,1)$ *and* $\widehat{A(\boldsymbol{\theta})} > 0$, *then we can further obtain*

$$\begin{cases} \widehat{\mu}_{CLS} & = -\frac{\log \widehat{\alpha}}{\Delta}, \\ \widehat{\beta}_{CLS} & = \frac{1}{\widehat{A(\boldsymbol{\theta})}_{CLS}}. \end{cases}$$

By Table 9.1, for K from **I1**, **I2**, **I4**, **P1**, **P2** and **P4**, the form of α is $e^{-\mu\Delta}$. Hence, from the first equation of (10.2.3), we always obtain

$$\widehat{\alpha}_{CLS} = \frac{\sum_{i=1}^{n-1} x_i x_{i+1} - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)\left(\sum_{i=1}^{n-1} x_{i+1}\right)}{\sum_{i=1}^{n-1} x_i^2 - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)^2} = e^{-\widehat{\mu}\Delta}.$$

When $0 < \widehat{\alpha}_{CLS} < 1$, the CLS estimate of μ is

$$\widehat{\mu}_{CLS} = -\frac{1}{\Delta}\log\left(\frac{\sum_{i=1}^{n-1} x_i x_{i+1} - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)\left(\sum_{i=1}^{n-1} x_{i+1}\right)}{\sum_{i=1}^{n-1} x_i^2 - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)^2}\right). \tag{10.2.5}$$

However, if $\widehat{\alpha}_{CLS} \le 0$ or $\widehat{\alpha}_{CLS} \ge 1$, what can we do? In such situation, we may set $\widehat{\alpha} = 0$ or $\widehat{\alpha} = 1$ respectively. These two situations are extreme cases for the continuous-time GAR(1) processes: the first one corresponds to an iid situation, while the second one corresponds to a perfectly dependent situation. But if $\widehat{\alpha}$ is strongly negative and the sample size n is large, it may suggest that the specified continuous-time GAR(1) process model is not appropriate to use. Or other approaches should be considered. For the marginal mean function $A(\boldsymbol{\theta})$, if its estimate $\widehat{A(\boldsymbol{\theta})}_{CLS}$ exceeds the range of $A(\boldsymbol{\theta})$, one simple alternative estimate is $\widehat{A(\boldsymbol{\theta})} = \frac{1}{n}\sum_{i=1}^{n} x_i$, which obtains from the method of moments in Section 10.4.

Obviously, the advantage of CLS estimation is that it can offer closed form estimates for generalized time series. However, the disadvantages are clear too:

- it can only estimate two parameters, because the CLS estimating equations (10.2.2) or (10.2.3) include only two equations;
- it ignores the conditional variance information.

These motivate us to turn to conditional weighted least squares (CWLS) approach. It considers

$$R_{CWLS}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \frac{\left(x_{i+1} - \mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i\right]\right)^2}{\mathbf{Var}\left[X(t_{i+1}) \mid X(t_i) = x_i\right]}, \tag{10.2.6}$$

and the conditional least squares estimate (CLS) is defined as

$$\widehat{\boldsymbol{\theta}}_{CWLS} = \arg\min_{\theta\in\Theta} R_{CWLS}(\boldsymbol{\theta}).$$

Similarly, we can obtain the estimating equations by taking partial derivatives of $R_{CWLS}(\boldsymbol{\theta})$. For the stationary continuous-time GAR(1) process, let the marginal mean function and variance function be $A(\boldsymbol{\theta})$ and $V(\boldsymbol{\theta})$ respectively. Denote

$$\alpha_i = \mathbf{E}\left[K\left(e^{-\mu(t_{i+1}-t_i)}\right)\right], \qquad \nu_i = \mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right], \tag{10.2.7}$$

where $i = 1, 2, \ldots, n-1$. By (9.2.1) and (9.2.2), we have

$$\begin{aligned}
\mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i\right] &= x_i \cdot \mathbf{E}\left[K\left(e^{-\mu(t_{i+1}-t_i)}\right)\right] + A(\boldsymbol{\theta}) \cdot \left(1 - \mathbf{E}\left[K\left(e^{-\mu(t_{i+1}-t_i)}\right)\right]\right) \\
&= x_i\alpha_i + A(\boldsymbol{\theta})\left[1 - \alpha_i\right], && (10.2.8) \\
\mathbf{Var}\left[X(t_{i+1}) \mid X(t_i) = x_i\right] &= V(\boldsymbol{\theta})\left(1 - \mathbf{E}^2\left[K\left(e^{-\mu(t_2-t_1)}\right)\right]\right) \\
&\quad + [x_i - A(\boldsymbol{\theta})] \cdot \mathbf{Var}\left[K\left(e^{-\mu(t_2-t_1)}\right)\right] \\
&= [x_i - A(\boldsymbol{\theta})]\nu_i + V(\boldsymbol{\theta})\left[1 - \alpha_i^2\right]. && (10.2.9)
\end{aligned}$$

where $i = 1, 2, \ldots, n-1$. These lead to

$$R_{CWLS}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \frac{\left(x_{i+1} - x_i\alpha_i - A(\boldsymbol{\theta})\left[1 - \alpha_i\right]\right)^2}{[x_i - A(\boldsymbol{\theta})]\nu_i + V(\boldsymbol{\theta})\left[1 - \alpha_i^2\right]}. \tag{10.2.10}$$

It is straightforward to show that the number of estimating equations won't simplify to fewer than the number of parameters k. However, no more closed form estimates can be expected even in the generalized time series. Numerical methods have to be employed to obtain the CWLS estimates $\widehat{\boldsymbol{\theta}}_{CWLS}$.

An estimation method related to CWLS approach is the quasi-conditional least squares approach (QCLS), which is a modification of the quasi-least squares approach proposed by Chaganty [1997]. The estimating equations of CWLS can be obtained by taking partial derivatives of $R_{CWLS}(\boldsymbol{\theta})$ in (10.2.6) with respect to parameter $\boldsymbol{\theta}$. The partial derivatives will be the sum of two terms: one is regarding the partial derivatives of conditional expectation from the numerator in the summands of (10.2.6), one is regarding the partial derivatives of conditional variance from the denominator in the summands of (10.2.6). Further research shows that the CWLS estimator is not consistent for the true parameter value $\boldsymbol{\theta}^0$ (see the comments in the last part of Section 11.3). Thus, a consistent estimator is pursued.

The quasi-conditional least squares approach considers the sum

$$R_{QCLS}(\boldsymbol{\theta},\boldsymbol{\theta}^*) = \sum_{i=1}^{n-1} \frac{\Big(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\Big)^2}{\mathbf{Var}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}^*]}, \tag{10.2.11}$$

where $\boldsymbol{\theta}^*$ is a variable independent of $\boldsymbol{\theta}$. When $\boldsymbol{\theta}^* = \boldsymbol{\theta}$, $R_{QCLS}(\boldsymbol{\theta},\boldsymbol{\theta}) = R_{CWLS}(\boldsymbol{\theta})$. Taking partial derivatives of $R_{QCLS}(\boldsymbol{\theta},\boldsymbol{\theta}^*)$ with respect to $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ and equating to zero, we have

$$\sum_{i=1}^{n-1} \frac{\Big(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\Big)}{\mathbf{Var}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}^*]} \cdot \frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_j} = 0, \qquad j = 1, 2, \ldots, k.$$

Taking $\boldsymbol{\theta}^* = \boldsymbol{\theta}$, we obtain the following estimating equations

$$\sum_{i=1}^{n-1} \frac{\Big(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\Big)}{\mathbf{Var}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]} \cdot \frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_j} = 0, \tag{10.2.12}$$

where $j = 1, 2, \ldots, k$. These estimating equations are not obtained from $R_{CWLS}(\boldsymbol{\theta})$. Hence, they are not the same as those CWLS estimating equations. We call the solutions of (10.2.12) the **quasi-conditional least squares estimate,** $\widehat{\boldsymbol{\theta}}_{QCLS}$. Unlike other estimates, this estimate won't minimize any of the sums, at most it marginally minimizes $R_{QCLS}(\boldsymbol{\theta},\boldsymbol{\theta}^*)$.

Similar to (10.2.2), the estimating equations of the CLS, for the stationary continuous-time GAR(1) process with marginal mean function $A(\boldsymbol{\theta})$ and marginal variance function $V(\boldsymbol{\theta})$, (10.2.12) can simplify to

$$\begin{cases} \sum_{i=1}^{n-1} \frac{x_{i+1}-x_i\alpha_i-A(\boldsymbol{\theta})[1-\alpha_i]}{[x_i-A(\boldsymbol{\theta})]\nu_i+V(\boldsymbol{\theta})[1-\alpha_i^2]} \cdot [x_i - A(\boldsymbol{\theta})] \frac{\partial \alpha_i}{\partial \mu} & = 0, \\ \sum_{i=1}^{n-1} \frac{x_{i+1}-x_i\alpha_i-A(\boldsymbol{\theta})[1-\alpha_i]}{[x_i-A(\boldsymbol{\theta})]\nu_i+V(\boldsymbol{\theta})[1-\alpha_i^2]} \cdot [1 - \alpha_i] & = 0, \end{cases} \tag{10.2.13}$$

Here α_i, ν_i $(i = 1, 2, \ldots, n-1)$ are defined in (10.2.7). Since the number of parameters may exceed 2 which is the number of equations in (10.2.13), the application of QCLS approach is limited like the CLS approach.

Note that the conditional variance $\mathbf{Var}\left[X(t_{i+1}) \mid X(t_i) = x_i\right]$ is linear in x_i. This leads to a simple variation of the CWLS estimates for the continuous-time GAR(1) process. Instead of $R_{CWLS}(\boldsymbol{\theta})$ in (10.2.10), we shall consider

$$R_{CWLS2}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \frac{\left(x_{i+1} - x_i\alpha_i - A(\boldsymbol{\theta})\left[1 - \alpha_i\right]\right)^2}{cx_i + d}, \tag{10.2.14}$$

where $c,\ d \geq 0$, but both are not equal to zero. These two constants are introduced so that we can partially take advantage of the information from the conditional variance. Usually d is chosen to be positive so that the denominators are not zero. It could be a small number, say $d = 0.5$. The constant c may be chosen by borrowing information from other estimation sources, because ν_i could be estimated by the function of α_i which may be estimated by CLS estimates or method of moments estimate $\widehat{\rho_1}$. If $c = 0,\ d = 1$, $R_{CWLS2}(\boldsymbol{\theta})$ will simplify to $R(\boldsymbol{\theta})$ in (10.2.1) for the continuous-time GAR(1) process, which results to the conditional least squares estimates. By minimizing $R_{CWLS2}(\boldsymbol{\theta})$, we can obtain

$$\widehat{\boldsymbol{\theta}}_{CWLS2} = \arg\min_{\theta\in\Theta} R_{CWLS2}(\boldsymbol{\theta}).$$

The big advantage over (10.2.10) is that we may obtain the closed form of estimates $\widehat{\boldsymbol{\theta}}_{CWLS2}$ in the generalized time series situation, because the denominators do not depend on parameters; this is just like the situation of the CLS approach.

Suppose the generalized time series is from a continuous-time GAR(1) process; in other words, we make the equally-spaced assumption here. Let $a = \alpha$ and $b = A(\boldsymbol{\theta})(1-\alpha)$. Then by taking partial derivatives with respect to a and b for $R_{CWLS2}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \frac{\left(x_{i+1}-x_i\alpha-b\right)^2}{cx_i+d}$, we will have

$$\begin{cases} \sum_{i=1}^{n-1} \frac{x_{i+1}-x_i\alpha-b}{cx_i+d} x_i & = & 0, \\ \sum_{i=1}^{n-1} \frac{x_{i+1}-x_i\alpha-b}{cx_i+d} & = & 0. \end{cases}$$

Let

$$B_1 = \sum_{i=1}^{n-1} \frac{1}{cx_i+d},\ B_2 = \sum_{i=1}^{n-1} \frac{x_i}{cx_i+d},\ B_3 = \sum_{i=1}^{n-1} \frac{x_{i+1}}{cx_i+d},\ B_4 = \sum_{i=1}^{n-1} \frac{x_i^2}{cx_i+d},\ B_5 = \sum_{i=1}^{n-1} \frac{x_i x_{i+1}}{cx_i+d}.$$

Then the equations become

$$\begin{cases} B_1 b + B_2 a & = & B_3, \\ B_2 b + B_4 a & = & B_5, \end{cases}$$

which have solutions

$$a = \frac{B_1B_5 - B_2B_3}{B_1B_4 - B_2^2}, \quad b = \frac{B_2B_5 - B_3B_4}{B_2^2 - B_1B_4}.$$

These lead to

$$\begin{cases} \widehat{\alpha}_{CWLS2} & = & \frac{B_1B_5-B_2B_3}{B_1B_4-B_2^2}, \\ \widehat{A(\boldsymbol{\theta})}_{CWLS2} & = & \frac{1}{1-\widehat{\alpha}_{CWLS2}} \cdot \frac{B_2B_5-B_3B_4}{B_2^2-B_1B_4}. \end{cases} \tag{10.2.15}$$

From (10.2.15), we may obtain $\widehat{\boldsymbol{\theta}}_{CWLS2}$. If the estimate $\widehat{\boldsymbol{\theta}}_{CLS}$ is beyond the range of $\boldsymbol{\theta}$, it is possible to find an appropriate $\widehat{\boldsymbol{\theta}}_{CWLS2}$ to be the alternative estimate.

The idea of conditional least squares can be extended to more general situation. Instead of the original sample $x_1, x_2, \ldots, x_n$, we may consider applying a real-valued function g to get: $g(x_1), g(x_2), \ldots, g(x_n)$. Correspondingly, we will replace the conditional least squares

$$R_{CLS}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \Big(x_{i+1} - \mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}\right]\Big)^2,$$

with conditional generalized least squares (CGLS)

$$R_{CGLS}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \Big(g(x_{i+1}) - \mathbf{E}\left[g(X(t_{i+1})) \mid X(t_i) = x_i; \boldsymbol{\theta}\right]\Big)^2, \tag{10.2.16}$$

which results to the conditional generalized least squares estimate:

$$\widehat{\boldsymbol{\theta}}_{CGLS} = \arg\min_{\theta\in\Theta} R_{CGLS}(\boldsymbol{\theta}).$$

When g is the identity function, this reduces to conditional least squares.

Our goal is to find potential closed form estimates for the generalized time series. For this purpose, we choose $g(x) = x^2$. Then from (10.2.8) and (10.2.9),

$$\begin{aligned}
\mathbf{E}\left[X^2(t_{i+1}) \mid X(t_i) = x_i\right] &= (\mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i\right])^2 + \mathbf{Var}\left[X(t_{i+1}) \mid X(t_i) = x_i\right] \\
&= [x_i\alpha + A(\boldsymbol{\theta})(1-\alpha)]^2 + [x_i - A(\boldsymbol{\theta})]\nu + V(\boldsymbol{\theta})(1-\alpha^2) \\
&= \alpha^2 x_i^2 + [2\alpha(1-\alpha)A(\boldsymbol{\theta}) + \nu]x_i + \left[(1-\alpha)^2A^2(\boldsymbol{\theta}) - \nu A(\boldsymbol{\theta}) + (1-\alpha^2)V(\boldsymbol{\theta})\right],
\end{aligned}$$

a linear equation of x_i and x_i^2. Let

$$a = \alpha^2, \quad b = 2\alpha(1-\alpha)A(\boldsymbol{\theta}) + \nu, \quad c = (1-\alpha)^2A^2(\boldsymbol{\theta}) - \nu A(\boldsymbol{\theta}) + (1-\alpha^2)V(\boldsymbol{\theta}). \tag{10.2.17}$$

Then

$$R_{CGLS}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \Big(x_{i+1}^2 - a x_i^2 - b x_i - c\Big)^2, \tag{10.2.18}$$

and resulting estimating equations from differentiating with respect to a, b and c are

$$\left\{\begin{aligned}
\sum_{i=1}^{n-1}(x_{i+1} - a x_i^2 - b x_i - c)x_i^2 &= 0, \\
\sum_{i=1}^{n-1}(x_{i+1} - a x_i^2 - b x_i - c)x_i &= 0, \\
\sum_{i=1}^{n-1}(x_{i+1} - a x_i^2 - b x_i - c) &= 0.
\end{aligned}\right.$$

Let

$$C_1 = \sum_{i=1}^{n-1} x_i, \quad C_2 = \sum_{i=1}^{n-1} x_i^2, \quad C_3 = \sum_{i=1}^{n-1} x_i^3, \quad C_4 = \sum_{i=1}^{n-1} x_i^4,$$
$$C_5 = \sum_{i=1}^{n-1} x_{i+1}, \quad C_6 = \sum_{i=1}^{n-1} x_i x_{i+1}, \quad C_7 = \sum_{i=1}^{n-1} x_i^2 x_{i+1}.$$

We can rewrite the above equations as

$$\begin{cases} C_4 a + C_3 b + C_2 c = C_7, \\ C_3 a + C_2 b + C_1 c = C_6, \\ C_2 a + C_1 b + (n-1)c = C_5, \end{cases}$$

and the solutions are

$$\begin{cases} a = \frac{[(n-1)C_7 - C_2 C_5]\cdot[(n-1)C_2 - C_1^2] - [(n-1)C_6 - C_1 C_5]\cdot[(n-1)C_3 - C_1 C_2]}{[(n-1)C_4 - C_2^2]\cdot[(n-1)C_2 - C_1^2] - [(n-1)C_3 - C_1 C_2]^2}, \\ b = \frac{[(n-1)C_7 - C_2 C_5]\cdot[(n-1)C_3 - C_1 C_2] - [(n-1)C_6 - C_1 C_5]\cdot[(n-1)C_4 - C_2^2]}{[(n-1)C_3 - C_1 C_2]^2 - [(n-1)C_4 - C_2^2]\cdot[(n-1)C_2 - C_1^2]}, \\ c = \frac{C_5}{n-1} - \frac{C_2}{n-1} a - \frac{C_1}{n-1} b. \end{cases} \tag{10.2.19}$$

By (10.2.17) and (10.2.19), we may obtain the CGLS estimate $\widehat{\boldsymbol{\theta}}_{CGLS}$. Note that $\widehat{\alpha}_{CGLS} = \sqrt{a}$. Hence, it can be another alternative choice if $\widehat{\alpha}_{CLS}$ is outside of the range $[0, 1]$.

Example 10.4 (Generalized Poisson univariate margins) *Consider the time series from the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} e^{-\mu(t-t')} * X(t') + E(t', t), \qquad t' < t,$$

which have GP(θ, η) ($\theta > 0, 0 \le \eta \le 1$) margin. In this case, $\boldsymbol{\theta} = (\mu, \theta, \eta)'$, and the marginal mean function and variance function of the process are

$$A(\boldsymbol{\theta}) = \theta(1-\eta)^{-1}, \qquad V(\boldsymbol{\theta}) = \theta(1-\eta)^{-3}.$$

From Table 9.1, for binomial-thinning, $\alpha = e^{-\mu\Delta}$ and $\nu = \alpha(1-\alpha) = e^{-\mu\Delta}\left(1 - e^{-\mu\Delta}\right)$. According to (10.2.19), we have the values of a, b and c. If $0 \le a \le 1$, by (10.2.17), we can obtain the estimates

$$\widehat{\alpha} = \sqrt{a}, \quad \widehat{A(\boldsymbol{\theta})} = \frac{b - \widehat{\alpha}(1-\widehat{\alpha})}{2\widehat{\alpha}(1-\widehat{\alpha})}, \quad \widehat{V(\boldsymbol{\theta})} = \frac{c - (1-\widehat{\alpha})^2 \widehat{A(\boldsymbol{\theta})}^2 + \widehat{\alpha}(1-\widehat{\alpha})\widehat{A(\boldsymbol{\theta})}}{1 - \widehat{\alpha}^2}. \tag{10.2.20}$$

If $\widehat{A(\boldsymbol{\theta})} > 0$ and $\widehat{V(\boldsymbol{\theta})} > 0$, we can further obtain the CGLS estimates

$$\widehat{\mu}_{CGLS} = -\frac{\log \widehat{\alpha}}{\Delta}, \qquad \widehat{\theta}_{CGLS} = \sqrt{\frac{\widehat{A(\boldsymbol{\theta})}^3}{\widehat{V(\boldsymbol{\theta})}}}, \quad \widehat{\eta}_{CGLS} = 1 - \sqrt{\frac{\widehat{A(\boldsymbol{\theta})}}{\widehat{V(\boldsymbol{\theta})}}}.$$

Note that $\widehat{\theta}_{CGLS}$ should be positive and $\widehat{\eta}_{CGLS}$ should be in the range $[0, 1]$.

Example 10.5 (NB univariate margins) *Consider the time series from the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} e^{-\mu(t-t')} * X(t') + E(t', t), \qquad t' < t,$$

which have NB(δ, γ) ($\delta > 0, 0 < \gamma < 1$) margin. In this case, the parameter vector $\boldsymbol{\theta} = (\mu, \delta, \gamma)'$, and the marginal mean function and variance function of the process are

$$A(\boldsymbol{\theta}) = \delta\gamma(1-\gamma)^{-1}, \qquad V(\boldsymbol{\theta}) = \delta\gamma(1-\gamma)^{-2}.$$

Similar to Example 10.4, we can obtain the estimate $\widehat{\alpha}$, $\widehat{A(\boldsymbol{\theta})}$ and $\widehat{V(\boldsymbol{\theta})}$ by (10.2.20). If $\widehat{A(\boldsymbol{\theta})} > 0$ and $\widehat{V(\boldsymbol{\theta})} > 0$, we can further obtain the CGLS estimate

$$\widehat{\mu}_{CGLS} = -\frac{\log \widehat{\alpha}}{\Delta}, \qquad \widehat{\delta}_{CGLS} = \frac{\widehat{A(\boldsymbol{\theta})}^2}{\widehat{V(\boldsymbol{\theta})} - \widehat{A(\boldsymbol{\theta})}}, \qquad \widehat{\gamma}_{CGLS} = 1 - \frac{\widehat{A(\boldsymbol{\theta})}}{\widehat{V(\boldsymbol{\theta})}}.$$

Note that $\widehat{\delta}_{CGLS}$ should be positive and $\widehat{\gamma}_{CGLS}$ should be in the range $(0, 1)$.

Example 10.6 (Gamma univariate margins) *Consider the time series from the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} e^{-\mu(t-t')} \bullet X(t') + E(t', t), \qquad t' < t,$$

which have Gamma(δ, β) ($\delta, \beta > 0$) margin. In this case, the parameter vector $\boldsymbol{\theta} = (\mu, \delta, \beta)'$, and the marginal mean function and variance function of the process are

$$A(\boldsymbol{\theta}) = \delta\beta^{-1}, \qquad V(\boldsymbol{\theta}) = \delta\beta^{-2}.$$

For K from **P1**, *by Table 9.1, $\alpha = e^{-\mu\Delta}$ and $\nu = 0$. According to (10.2.19), we can calculate a, b and c. If $0 \le a \le 1$, we can obtain the following estimates by (10.2.17):*

$$\widehat{\alpha} = \sqrt{a}, \qquad \widehat{A(\boldsymbol{\theta})} = \frac{b}{2\widehat{\alpha}(1-\widehat{\alpha})}, \qquad \widehat{V(\boldsymbol{\theta})} = \frac{c - (1-\widehat{\alpha})^2 \widehat{A(\boldsymbol{\theta})}^2}{1-\widehat{\alpha}^2}.$$

If $\widehat{A(\boldsymbol{\theta})} > 0$ and $\widehat{V(\boldsymbol{\theta})} > 0$, we can further obtain the CGLS estimate

$$\widehat{\mu}_{CGLS} = -\frac{\log \widehat{\alpha}}{\Delta}, \qquad \widehat{\delta}_{CGLS} = \frac{\widehat{A(\boldsymbol{\theta})}^2}{\widehat{V(\boldsymbol{\theta})}}, \qquad \widehat{\gamma}_{CGLS} = \frac{\widehat{A(\boldsymbol{\theta})}}{\widehat{V(\boldsymbol{\theta})}}.$$

Note that $\widehat{\delta}_{CGLS}$ and $\widehat{\beta}_{CGLS}$ should be positive.

In Examples 10.2 and 10.3, we restrict the parameter γ associated with K from **I2** and **P2** to a specific value. This is simply because of the limitation of number of estimating equations in the CLS approach. Now we can loosen such a restriction in the CGLS approach.

Example 10.7 (Geometric univariate margins) *Consider the time series with Geometric(β) univariate margins, where $0 < \beta < 1$. This time series is from the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t', t), \qquad t' < t,$$

where K is from **I2** *with pgf $G_K(s;\alpha) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}$. By Example 7.15, we can choose $0 < \gamma < \beta$.*

Then $\boldsymbol{\theta} = (\mu, \gamma, \beta)'$, $A(\boldsymbol{\theta}) = \beta(1-\beta)^{-1}$ and $V(\boldsymbol{\theta}) = \beta(1-\beta)^{-2}$. For K from **I2**, *$\alpha = e^{-\mu\Delta}$ and $\nu = \alpha(1-\alpha)(1+\gamma)(1-\gamma)^{-1}$. Therefore, according to (10.2.19), we can compute a, b and c. Furthermore, from (10.2.17), we have*

$$\begin{cases} a &= \widehat{\alpha}^2, \\ b &= 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{\beta}(1-\widehat{\beta})^{-1} + \widehat{\alpha}(1-\widehat{\alpha})(1+\widehat{\gamma})(1-\widehat{\gamma})^{-1}, \\ c &= (1-\widehat{\alpha})^2\widehat{\beta}^2(1-\widehat{\beta})^{-2} - \widehat{\alpha}(1-\widehat{\alpha})(1+\widehat{\gamma})(1-\widehat{\gamma})^{-1}\widehat{\beta}(1-\widehat{\beta})^{-1} + (1-\widehat{\alpha}^2)\widehat{\beta}(1-\widehat{\beta})^{-2}. \end{cases}$$

Solving these equations, we can obtain $\widehat{\alpha}_{CGLS}$, $\widehat{\beta}_{CGLS}$ and $\widehat{\gamma}$. Since it is very tedious, we omit the details. If $0 \le a \le 1$, we can obtain $\widehat{\mu}_{CGLS} = -\frac{\log\widehat{\alpha}}{\Delta}$.

Example 10.8 (Exponential univariate margins) *Consider the time series with Exponential(β) univariate margins, where $\beta > 0$. This time series is from the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t', t), \qquad t' < t,$$

where K is from **P2** *with LT $\phi_K(s;\alpha) = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\}$. By Example 6.15, we can choose $0 < \gamma < (1+\beta)^{-1}$.*

Then $\boldsymbol{\theta} = (\mu, \gamma, \beta)'$, $A(\boldsymbol{\theta}) = \beta^{-1}$ and $V(\boldsymbol{\theta}) = \beta^{-2}$. For K from **P2**, *$\alpha = e^{-\mu\Delta}$ and $\nu = 2\alpha(1-\alpha)\gamma(1-\gamma)^{-1}$. Therefore, according to (10.2.19), we have a, b and c. Similarly, by*

(10.2.17), we obtain

$$\begin{cases} a &= \widehat{\alpha}^2, \\ b &= 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{\beta}^{-1} + 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{\gamma}(1-\widehat{\gamma})^{-1}, \\ c &= (1-\widehat{\alpha})^2\widehat{\beta}^{-2} - 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{\gamma}(1-\widehat{\gamma})^{-1}\widehat{\beta}^{-1} + (1-\widehat{\alpha}^2)\widehat{\beta}^{-2}. \end{cases}$$

The solutions will be $\widehat{\mu}_{CGLS}$, $\widehat{\gamma}_{CGLS}$ and $\widehat{\beta}_{CGLS}$. The tedious details are omitted too.

Since (10.2.19) only offers estimation for three parameters, we can not apply the CGLS approach to the continuous-time GAR(1) process associated with four or more parameters. For the time series from a four-parameter continuous-time GAR(1) process, we may combine (10.2.3) and (10.2.19) to obtain the estimate $\widehat{\boldsymbol{\theta}}$. The following is an example to illustrate this idea.

Example 10.9 (NB univariate margins) *Consider the time series with NB(δ, β) univariate margins, where $\delta > 0$ and $0 < \beta < 1$. This time series is from the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t',t), \qquad t' < t,$$

where K is from **I2** *with pgf $G_K(s;\alpha) = \frac{(1-\alpha)+(\alpha-\gamma)s}{(1-\alpha\gamma)-(1-\alpha)\gamma s}$ $(0 < \gamma < \beta)$.*

Then $\boldsymbol{\theta} = (\mu,\gamma,\beta)'$, $A(\boldsymbol{\theta}) = \delta\beta(1-\beta)^{-1}$ and $V(\boldsymbol{\theta}) = \delta\beta(1-\beta)^{-2}$. For K from **I2**, *$\alpha = e^{-\mu\Delta}$ and $\nu = \alpha(1-\alpha)(1+\gamma)(1-\gamma)^{-1}$. Therefore, according to (10.2.19) and (10.2.17), we have*

$$\begin{cases} a &= \widehat{\alpha}^2, \\ b &= 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{A(\boldsymbol{\theta})} + \widehat{\alpha}(1-\widehat{\alpha})(1+\widehat{\gamma})(1-\widehat{\gamma})^{-1}, \\ c &= (1-\widehat{\alpha})^2\widehat{A(\boldsymbol{\theta})}^2 - \widehat{\alpha}(1-\widehat{\alpha})(1+\widehat{\gamma})(1-\widehat{\gamma})^{-1}\widehat{A(\boldsymbol{\theta})} + (1-\widehat{\alpha}^2)\widehat{V(\boldsymbol{\theta})}. \end{cases}$$

On the other hand, by (10.2.3), we can obtain

$$\begin{cases} \widehat{\alpha} &= \dfrac{\sum_{i=1}^{n-1} x_i x_{i+1} - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)\left(\sum_{i=1}^{n-1} x_{i+1}\right)}{\sum_{i=1}^{n-1} x_i^2 - \frac{1}{n-1}\left(\sum_{i=1}^{n-1} x_i\right)^2}, \\ \widehat{A(\boldsymbol{\theta})} &= \dfrac{\frac{1}{n-1}\sum_{i=1}^{n-1} x_{i+1} - \left(\frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right)\widehat{\alpha}}{1-\widehat{\alpha}}. \end{cases}$$

Choosing one estimate of α from either set of equations, we can derive new equations

$$\begin{cases} \widehat{A(\boldsymbol{\theta})} &= \dfrac{\frac{1}{n-1}\sum_{i=1}^{n-1} x_{i+1} - \left(\frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right)\widehat{\alpha}}{1-\widehat{\alpha}}, \\ b &= 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{A(\boldsymbol{\theta})} + \widehat{\alpha}(1-\widehat{\alpha})(1+\widehat{\gamma})(1-\widehat{\gamma})^{-1}, \\ c &= (1-\widehat{\alpha})^2\widehat{A(\boldsymbol{\theta})}^2 - \widehat{\alpha}(1-\widehat{\alpha})(1+\widehat{\gamma})(1-\widehat{\gamma})^{-1}\widehat{A(\boldsymbol{\theta})} + (1-\widehat{\alpha}^2)\widehat{V(\boldsymbol{\theta})}. \end{cases}$$

Solving these equations, we can obtain

$$\widehat{\gamma} = \frac{b - 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{A(\boldsymbol{\theta})} - \widehat{\alpha}(1-\widehat{\alpha})}{b - 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{A(\boldsymbol{\theta})} + \widehat{\alpha}(1-\widehat{\alpha})}, \qquad \widehat{V(\boldsymbol{\theta})} = \frac{c + b\widehat{A(\boldsymbol{\theta})} - (1-\widehat{\alpha}^2)\widehat{A(\boldsymbol{\theta})}^2}{1-\widehat{\alpha}^2},$$

which finally lead to

$$\widehat{\delta} = \frac{\widehat{A(\boldsymbol{\theta})}^2}{\widehat{V(\boldsymbol{\theta})} - \widehat{A(\boldsymbol{\theta})}}, \qquad \widehat{\beta} = 1 - \frac{\widehat{A(\boldsymbol{\theta})}}{\widehat{V(\boldsymbol{\theta})}}.$$

Example 10.10 (Gamma univariate margins) *Consider the time series with Gamma(δ, β) univariate margins, where $\delta, \beta > 0$. This time series is from the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} \left(e^{-\mu(t-t')}\right)_K \circledast X(t') + E(t', t), \qquad t' < t,$$

where K is from **P2** *with LT* $\phi_K(s;\alpha) = \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\}$ $(0 < \gamma < (1+\beta)^{-1})$.

Then $\boldsymbol{\theta} = (\mu, \gamma, \beta)'$, $A(\boldsymbol{\theta}) = \delta\beta^{-1}$ and $V(\boldsymbol{\theta}) = \delta\beta^{-2}$. For K from **P2**, *$\alpha = e^{-\mu\Delta}$ and $\nu = 2\alpha(1-\alpha)\gamma(1-\gamma)^{-1}$. Similarly, choosing one estimate of α from either (10.2.3) or (10.2.19), and combining the remaining equations, we obtain new equations*

$$\begin{cases} \widehat{A(\boldsymbol{\theta})} &= \dfrac{\frac{1}{n-1}\sum_{i=1}^{n-1} x_{i+1} - \left(\frac{1}{n-1}\sum_{i=1}^{n-1} x_i\right)\widehat{\alpha}}{1-\widehat{\alpha}}, \\ b &= 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{A(\boldsymbol{\theta})} + 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{\gamma}(1-\widehat{\gamma})^{-1}, \\ c &= (1-\widehat{\alpha})^2\widehat{A(\boldsymbol{\theta})}^2 - 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{\gamma}(1-\widehat{\gamma})^{-1}\widehat{A(\boldsymbol{\theta})} + (1-\widehat{\alpha}^2)\widehat{V(\boldsymbol{\theta})}. \end{cases}$$

Solving these equations, we can obtain

$$\widehat{\gamma} = \frac{b - 2\widehat{\alpha}(1-\widehat{\alpha})\widehat{A(\boldsymbol{\theta})}}{b + 2\widehat{\alpha}(1-\widehat{\alpha})[1-\widehat{A(\boldsymbol{\theta})}]}, \qquad \widehat{V(\boldsymbol{\theta})} = \frac{c + b\widehat{A(\boldsymbol{\theta})} - (1-\widehat{\alpha}^2)\widehat{A(\boldsymbol{\theta})}^2}{1-\widehat{\alpha}^2}.$$

Thus,

$$\widehat{\delta} = \frac{\widehat{A(\boldsymbol{\theta})}^2}{\widehat{V(\boldsymbol{\theta})}}, \qquad \widehat{\beta} = \frac{\widehat{A(\boldsymbol{\theta})}}{\widehat{V(\boldsymbol{\theta})}}.$$

Now we look back the estimation approaches discussed in this section. The R_{CLS} and R_{CWLS2} functions are only related to conditional mean, hence, they are only applicable to estimate those parameters which are arguments of the conditional mean. The R_{CWLS}, R_{QCLS} and R_{CGLS} functions involve not only the conditional mean, but also the conditional variance. Thus, they can estimate more parameters than the previous two approaches. For the unequally-spaced time observations, in general, we do not have the closed form estimates. But, for equally-spaced time observations, we have derived closed form estimates in the CLS, CWLS2 and CGLS approach. However, we should always beware of the estimated parameter ranges in all approaches. If estimates are out of range, the model may not be appropriate.

In summary, the conditional least squares approach and its variations work only for finite conditional mean and/or conditional variance. This requires that at least the marginal mean function of a continuous-time GAR(1) process should be finite. However, there are some continuous-time GAR(1) process which have infinite marginal mean function, for example, the power series, logarithmic series, positive stable marginal distributions, etc.

10.3 Empirical characteristic function estimation approach and variations

There are some GDSD and GSD distributions which have infinite mean. These distributions include power series, logarithmic series, discrete stable, discrete Mittag-Leffler, Zeta (when $0 < \rho \leq 1$), Mittag-Leffler, stable (most of them), etc. Hence, the conditional least squares approach is not applicable in the parameter estimation for the continuous-time GAR(1) process with marginal distributions which have infinite mean. In principle, the maximum likelihood approach can handle an infinite mean, but the conditional pdf or pmf should be in closed or simple form. Although we usually know the conditional pgf, or LT, or cf form for those continuous-time GAR(1) processes, it is very difficult to find the corresponding pdf's or pmf's. Thus, in most cases, the maximum likelihood approach may not realistic.

If we just consider the equally-spaced time observation from a stationary continuous-time GAR(1) process with infinite marginal mean, we may adopt the empirical characteristic function estimation approach. For references, one can see Paulson, Holcomb and Leitch [1975], Feuerverger and McDunnough [1981a, 1981b], Feuerverger [1990], Ushakov [1999] and references therein.

In such a stationary setting of time series, we view the sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ as a realization of $(X(t_1), X(t_2), \ldots, X(t_n))$. Hence, the pairs

$$(X(t_1), X(t_2)), \quad (X(t_2), X(t_3)), \quad \ldots, \quad (X(t_{n-1}), X(t_n))$$

have the same bivariate distribution with cf $\varphi_{(X_1,X_2)}(s_1, s_2; \boldsymbol{\theta}) = \mathbf{E}\left[e^{i(s_1 X_1 + s_2 X_2)}\right]$, although they are not independent. Unlike in Feuerverger [1990], for the time series from a stationary continuous-time GAR(1) process, we only need to consider the bivariate marginal cf and its corresponding empirical bivariate cf.

Denote

$$\varphi_n(s_1, s_2) = \frac{1}{n-1} \sum_{i=1}^{n-1} e^{i(s_1 x_i + s_2 x_{i+1})},$$

the empirical bivariate characteristic function of the stationary time series. Note that this is called the poly-cf in Feuerverger and McDunnough [1981b] where they wanted to distinguish it from the iid sample case. This is not necessary in our context. Intuitively, as sample size n goes to infinity, the sample function $\varphi_n(s_1, s_2)$ will tend to be the theoretical function $\varphi_{(X_1,X_2)}(s_1, s_2; \boldsymbol{\theta})$. Hence, we wish this empirical bivariate cf to be close to the corresponding bivariate cf. By some procedures to minimize the difference between the function $\varphi_n(s_1, s_2)$ and $\varphi_{(X_1,X_2)}(s_1, s_2; \boldsymbol{\theta})$, we may use the empirical characteristic function (ECF) estimate $\widehat{\boldsymbol{\theta}}_{ECF}$ for the parameter $\boldsymbol{\theta}$.

In the earlier papers for estimation in the stable distributions such as Paulson, Holcomb and Leitch [1975], the goal is to minimize

$$I = \int_{-\infty}^{\infty} ||\varphi_n(s) - \varphi_X(s; \boldsymbol{\theta})||^2 e^{-s^2} ds = \int_{-\infty}^{\infty} \lambda(\boldsymbol{\theta}) e^{-s^2} ds$$

to obtain the estimate for univariate case, where

$$\varphi_n(s) = \frac{1}{n} \sum_{i=1}^{n} e^{isx_i}, \quad \varphi_X(s; \boldsymbol{\theta}) = \mathbf{E}\left[e^{isX}\right], \quad \text{and} \quad \lambda(\boldsymbol{\theta}) = ||\varphi_n(s) - \varphi_X(s; \boldsymbol{\theta})||^2.$$

Note that here $||\cdot||^2$ is the modulus of a complex number. For the specific stable distribution case, one can approximate the integration by 20 point Hermitian quadrature and then find the minimum (see Paulson, et al. [1975]). However, this may not work for other distributions. Feuerverger and McDunnough [1981b] summarized four estimation procedures. Instead of the difference between two functions on every point, one may consider the difference between two functions on their finite grid points. Then consider some kind of quadratic form of these finite differences to substitute the overall difference I, and finally minimize the quadratic to obtain the parameter estimates. The last stage is similar to the least squares approach. Let $\mathbf{s} = (s_1, s_2)'$, and set the grid points as

$$\mathbf{s}_1 = (s_{11}, s_{21})', \quad \mathbf{s}_2 = (s_{12}, s_{22})', \quad \ldots, \quad \mathbf{s}_m = (s_{1m}, s_{2m})',$$

where m is a positive integer. Define

$$\mathbf{z}_n = \Big(\text{Re } \varphi_n(\mathbf{s}_1), \ldots, \text{Re } \varphi_n(\mathbf{s}_m), \text{Im } \varphi_n(\mathbf{s}_1), \ldots, \text{Im } \varphi_n(\mathbf{s}_m)\Big)',$$

and

$$\mathbf{z}_{\boldsymbol{\theta}} = \Big(\text{Re } \varphi_{(X_1,X_2)}(\mathbf{s}_1;\boldsymbol{\theta}), \ldots, \text{Re } \varphi_{(X_1,X_2)}(\mathbf{s}_m;\boldsymbol{\theta}), \text{Im } \varphi_{(X_1,X_2)}(\mathbf{s}_1;\boldsymbol{\theta}), \ldots, \text{Im } \varphi_{(X_1,X_2)}(\mathbf{s}_m;\boldsymbol{\theta})\Big)'.$$

Consider the quadratic form

$$R_{ECF}(\boldsymbol{\theta}) = (\mathbf{z}_n - \mathbf{z}_{\boldsymbol{\theta}})'\mathbf{Q}(\mathbf{z}_n - \mathbf{z}_{\boldsymbol{\theta}}),$$

where $\mathbf{Q}$ is a $2m$ by $2m$ positive definite matrix. This quadratic somehow measures the closeness of the empirical bivariate cf and the theoretical bivariate cf. Thus, the empirical characteristic function estimate of $\boldsymbol{\theta}$ is then defined as

$$\widehat{\boldsymbol{\theta}}_{ECF} = \arg\min_{\theta\in\Theta} R_{ECF}(\boldsymbol{\theta}).$$

Feuerverger and McDunnough [1981b] considered four selections for the matrix $\mathbf{Q}$ which yield consistent estimators with the same asymptotic normal distribution. They also claimed that the asymptotic variances of these four estimators can be arbitrarily close to the Cramer-Rao bound by choosing the grid $\{s_j\}$ to be sufficiently fine and extended. However, how to choose those grids in the practice is not clear.

Feuerverger and McDunnough [1981b] applied the empirical characteristic function approach to a stationary Markov emigration-immigration process $\{X_t;\ t = 0, \pm 1, \ldots\}$, which has bivariate marginal pgf

$$G_{(X_1,X_2)}(s_1, s_2; \nu, \rho) = \mathbf{E}\left[s_1^{X_1} s_2^{X_2}\right] = \exp\left\{\nu[(s_1 - 1) + (s_2 - 1) + \rho(s_1 - 1)(s_2 - 1)]\right\},$$

where $\nu = \mathbf{E}(X_1) = \mathbf{Var}(X_1)$ and $\rho = \mathbf{Cor}(X_1, X_2)$. Hence, the bivariate marginal cf is

$$\varphi_{(X_1,X_2)}(s_1, s_2; \nu, \rho) = \exp\left\{\nu\left[\left(e^{is_1} - 1\right) + \left(e^{is_2} - 1\right) + \rho\left(e^{is_1} - 1\right)\left(e^{is_2} - 1\right)\right]\right\}.$$

This process is coincidently the discrete process sampling from a stationary continuous-time GAR(1) process with Poisson margins. See the case of bivariate DSD distributions in Example 9.6, where $\nu = \lambda$ and $\rho = \alpha$.

Most of the marginal distributions of the stationary continuous-time GAR(1) processes have non-negative integer or positive real support. This motivates us to consider the variations of the empirical characteristic function approach. Instead of the characteristic function, we may use the probability generating function for non-negative integer margins and the Laplace transformation for the positive real margins respectively. Since

$$G_{(X_1,X_2)}(s_1, s_2; \boldsymbol{\theta}) = \varphi_{(X_1,X_2)}(-i \log s_1, -i \log s_2; \boldsymbol{\theta}), \quad \phi_{(X_1,X_2)}(s_1, s_2; \boldsymbol{\theta}) = \varphi_{(X_1,X_2)}(is_1, is_2; \boldsymbol{\theta}),$$

both are compound functions of the characteristic functions. Therefore, in principle, the new estimates from the empirical pgf or LT approach will inherit the consistency, but change the efficiency.

Define the empirical bivariate pgf and LT as below:

$$G_n(s_1, s_2) = \frac{1}{n-1} \sum_{i=1}^{n-1} s_1^{x_i} s_2^{x_{i+1}}, \quad \phi_n(s_1, s_2) = \frac{1}{n-1} \sum_{i=1}^{n-1} e^{-(s_1 x_i + s_2 x_{i+1})}.$$

Choose the grid points

$$\mathbf{s}_1 = (s_{11}, s_{21})', \quad \mathbf{s}_2 = (s_{12}, s_{22})', \quad \ldots, \quad \mathbf{s}_m = (s_{1m}, s_{2m})'.$$

Since the pgf and LT are real functions, we do not need to consider the imaginary part. Hence, we obtain

$$\mathbf{z}_n = \Big(G_n(\mathbf{s}_1), \ldots, G_n(\mathbf{s}_m)\Big)', \quad \text{or} \quad \mathbf{z}_n = \Big(\phi_n(\mathbf{s}_1), \ldots, \phi_n(\mathbf{s}_m)\Big)',$$

and

$$\mathbf{z}_{\boldsymbol{\theta}} = \Big(G_{(X_1,X_2)}(\mathbf{s}_1;\boldsymbol{\theta}),\ldots,G_{(X_1,X_2)}(\mathbf{s}_m;\boldsymbol{\theta})\Big)', \quad \text{or} \quad \mathbf{z}_{\boldsymbol{\theta}} = \Big(\phi_{(X_1,X_2)}(\mathbf{s}_1;\boldsymbol{\theta}),\ldots,\phi_{(X_1,X_2)}(\mathbf{s}_m;\boldsymbol{\theta})\Big)'.$$

Consider the quadratic form of $\mathbf{z}_n - \mathbf{z}_{\boldsymbol{\theta}}$ for both cases:

$$R_{EPGF}(\boldsymbol{\theta}) = (\mathbf{z}_n - \mathbf{z}_{\boldsymbol{\theta}})'\mathbf{Q}(\mathbf{z}_n - \mathbf{z}_{\boldsymbol{\theta}}), \quad \text{or} \quad R_{ELT}(\boldsymbol{\theta}) = (\mathbf{z}_n - \mathbf{z}_{\boldsymbol{\theta}})'\mathbf{Q}(\mathbf{z}_n - \mathbf{z}_{\boldsymbol{\theta}}),$$

where $\mathbf{Q}$ is an m by m positive matrix. We can derive the empirical probability generating function (EPGF) estimate or empirical Laplace transformation (ELT) estimate of $\boldsymbol{\theta}$ by minimizing this quadratic:

$$\widehat{\boldsymbol{\theta}}_{EPGF} = \arg\min_{\theta\in\Theta} R_{EPGF}(\boldsymbol{\theta}), \quad \text{or} \quad \widehat{\boldsymbol{\theta}}_{ELT} = \arg\min_{\theta\in\Theta} R_{ELT}(\boldsymbol{\theta}).$$

To give a rough estimation, one can naively choose $\mathbf{Q} = \mathbf{I}$, the identity matrix. For bivariate pgf, since the domain is $[0,1]^2$, a closed set in $\Re^2$, we may simply choose the uniform grids:

$$\mathbf{s}_{ij} = (i/l, j/l)', \qquad i,\ j = 0,1,\ldots,l; \quad l \text{ is a positive integer.}$$

Hence, in such a situation, the number of grid points $m = (l+1)^2$.

In general, there are no closed form estimates from the empirical characteristic function approach and its variations. We need to employ numerical methods to find $\widehat{\boldsymbol{\theta}}_{EPGF}$, $\widehat{\boldsymbol{\theta}}_{ELT}$ and $\widehat{\boldsymbol{\theta}}_{ECF}$. However, if we take an appropriate transformation for the pgf, or LT, or cf, and define the corresponding sample counterpart, it may be possible to obtain a closed form estimate. This is the generalization of the empirical characteristic function approach and its variations. One application of such estimates is to use them as the initial values of numerical methods.

We remark that the empirical characteristic function approach and variations work not only for the stationary continuous-time GAR(1) processes with infinite marginal mean, but also with finite marginal mean. However, they are only applicable in the equally-spaced time observations. For the unequally-spaced time observations, we can't use these approaches to estimate the parameters because we can't define a reasonable empirical characteristic function for bivariate margins or multivariate margins. This is a disadvantage of the empirical characteristic function approach.

10.4 Other estimation approaches

The maximum likelihood, the conditional least squares and the empirical characteristic function approaches are classical recipes for parameter estimation. However, for models with special features, one may use common sense or imagination to construct a reasonable estimate. In this section, we shall discuss some special approaches.

Methods of moment approach. In this approach, we view the sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ as a realization of $X(t_1), X(t_2), \ldots, X(t_n)$. Define

$$
\begin{aligned}
R_1 &= \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \quad R_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad R_{12} = \frac{1}{n-1}\sum_{i=1}^{n-1} x_i x_{i+1}, \\
R_{13} &= \frac{1}{n-2}\sum_{i=1}^{n-2} x_i x_{i+2}, \quad \ldots, \quad R_{1l} = \frac{1}{n-l+1}\sum_{i=1}^{n-l+1} x_i x_{i+l-1},
\end{aligned}
$$

where l is a positive integer. Assume that the corresponding stationary continuous-time GAR(1) process has marginal mean function $A(\boldsymbol{\theta})$ and marginal variance function $V(\boldsymbol{\theta})$. By algebra, we have

$$
\begin{aligned}
\mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X(t_i)\right] &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\left[X(t_i)\right] = A(\boldsymbol{\theta}), \\
\mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X^2(t_i)\right] &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\left[X^2(t_i)\right] = A^2(\boldsymbol{\theta}) + V(\boldsymbol{\theta}).
\end{aligned}
$$

By equating them to the corresponding sample averages, we obtain two estimating equations:

$$
A(\boldsymbol{\theta}) = R_1, \qquad A^2(\boldsymbol{\theta}) + V(\boldsymbol{\theta}) = R_2. \tag{10.4.1}
$$

If the sampling is based on the equally-spaced time scheme, we can calculate the expectation of $\frac{1}{n-j+1}\sum_{i=1}^{n-j+1} X(t_i)X(t_{i+j-1})$. Let $\alpha = \mathbf{E}\left[K\left(e^{-\mu\Delta}\right)\right]$, where Δ is the common time difference. Since

$$
\mathbf{Cov}\left(X(t_i), X(t_j)\right) = \alpha^{j-i} V(\boldsymbol{\theta}), \qquad j > i,
$$

we can derive

$$
\mathbf{E}\left[\frac{1}{n-j+1}\sum_{i=1}^{n-j+1} X(t_i)X(t_{i+j-1})\right] = \frac{1}{n-j+1}\sum_{i=1}^{n-j+1} \mathbf{E}\left[X(t_i)X(t_{i+j-1})\right]
$$

$$= \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} \left\{ \mathbf{Cov}\left(X(t_i), X(t_{i+j-1})\right) + \mathbf{E}\left[X(t_i)\right] \mathbf{E}\left[X(t_{i+j-1})\right] \right\}$$
$$= \alpha^{j-1} V(\boldsymbol{\theta}) + A^2(\boldsymbol{\theta}),$$

for $j = 1, \ldots, l$. These lead to estimating equations:

$$\alpha^{j-1} V(\boldsymbol{\theta}) + A^2(\boldsymbol{\theta}) \quad = \quad R_{1j}, \qquad j = 1, \ldots, l. \tag{10.4.2}$$

From (10.4.1) and (10.4.2), we can derive the method of moment estimates for $A(\boldsymbol{\theta})$, $V(\boldsymbol{\theta})$ and α:

$$\widehat{A(\boldsymbol{\theta})}_M \; = \; R_1, \quad \widehat{V(\boldsymbol{\theta})}_M \; = \; R_2 - R_1^2, \quad \widehat{\alpha}_M \; = \; \left(\frac{R_{1r} - R_1^2}{R_2 - R_1^2} \right)^{1/(r-1)}, \tag{10.4.3}$$

where $2 \le r \le l$ such that $R_{1r} - R_1^2 \ge 0$. To keep $n - r + 1$ as large as possible, we choose the smallest r. Since $R_2 - R_1^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \ge 0$, $\widehat{V(\boldsymbol{\theta})}_M \ge 0$. Hence, the method of moments approach guarantees that the estimate $\widehat{A(\boldsymbol{\theta})}_M$ and $\widehat{V(\boldsymbol{\theta})}_M$ are always positive for the stationary continuous-time GAR(1) process with non-negative margins.

We briefly discuss the situation when $r = 2$. Let

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{x}_{(-1)} = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i, \quad \bar{x}_{(-n)} = \frac{1}{n-1} \sum_{i=1}^{n-1} x_{i+1}.$$

We shall have

$$\begin{aligned}
\widehat{\alpha}_M \;&=\; \frac{R_{12} - R_1^2}{R_2 - R_1^2} \;=\; \frac{\frac{1}{n-1} \sum_{i=1}^{n-1} x_i x_{i+1} - \bar{x}_n^2}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2} \\
&= \frac{n}{n-1} \cdot \frac{\sum_{i=1}^{n-1} \left(x_i - \bar{x}_{(-1)}\right)\left(x_{i+1} - \bar{x}_{(-n)}\right)}{\sum_{i=1}^{n} \left(x_i - \bar{x}_n\right)^2} + \frac{n}{n-1} \cdot \frac{(n-1)\bar{x}_{(-1)} \cdot \bar{x}_{(-n)} - (n-1)\bar{x}_n^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}_n\right)^2} \\
&= \frac{n}{n-1} \widehat{\alpha}_{CLS} + \frac{\bar{x}_{(-1)} \cdot \bar{x}_{(-n)} - \bar{x}_n^2}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2}.
\end{aligned}$$

When n is large, $\frac{1}{n-1} \sum_{i=1}^{n-1} x_i \approx \frac{1}{n-1} \sum_{i=1}^{n-1} x_{i+1} \approx \frac{1}{n} \sum_{i=1}^{n} x_i$; thus $\widehat{\alpha}_M$ is close to $\widehat{\alpha}_{CLS}$.

From (10.4.3), we can further obtain the method of moments estimates of the model parameters associated with $A(\boldsymbol{\theta})$, $V(\boldsymbol{\theta})$ and α. Typically, this approach allows us to handle three-parameter models, because we only have three estimating equations. Of course, these equations can be combined with estimating equations from other approaches.

Note that this approach only works in the continuous-time GAR(1) processes with finite marginal mean and variance functions, not in the infinite case.

Now we study Example 10.4 again.

Example 10.11 (Generalized Poisson univariate margins) *Consider the time series from the stationary continuous-time GAR(1) process*

$$X(t) \stackrel{d}{=} e^{-\mu(t-t')} * X(t') + E(t', t), \qquad t' < t,$$

which have GP(θ,η) ($\theta > 0, 0 \leq \eta \leq 1$) margin. Thus, the parameter vector $\boldsymbol{\theta} = (\mu, \theta, \eta)'$, and the marginal mean function and variance function of the process are

$$A(\boldsymbol{\theta}) = \theta(1-\eta)^{-1}, \qquad V(\boldsymbol{\theta}) = \theta(1-\eta)^{-3}.$$

For the sake of convenience, we assume $R_{12} - R_1^2 \geq 0$. According to (10.4.3), we have

$$\widehat{A(\boldsymbol{\theta})}_M = R_1, \quad \widehat{V(\boldsymbol{\theta})}_M = R_2 - R_1^2, \quad \widehat{\alpha}_M = \frac{R_{12} - R_1^2}{R_2 - R_1^2}.$$

If $0 \leq \alpha \leq 1$, we can further obtain the moment estimates

$$\widehat{\mu}_M = -\frac{1}{\Delta} \log\left(\frac{R_{12} - R_1^2}{R_2 - R_1^2}\right), \qquad \widehat{\theta}_M = \sqrt{\frac{R_1^3}{R_2 - R_1^2}}, \quad \widehat{\eta}_M = 1 - \sqrt{\frac{R_1}{R_2 - R_1^2}}.$$

Note that $\widehat{\eta}_M$ should be in the range $[0, 1]$.

Ratio approach. This approach only focuses on the discrete-time process

$$X_{i+1} = \alpha X_i + \epsilon_i,$$

with positive real margins, and estimates the autoregressive coefficient α. See Bell and Smith [1986], and Anděl [1989]. It is based on the following inequality

$$\frac{X_{i+1}}{X_i} = \alpha + \frac{\epsilon_i}{X_i} \geq \alpha.$$

The proposed ratio estimate is

$$\widehat{\alpha}_R = \min\left\{\frac{x_2}{x_1}, \frac{x_3}{x_2}, \ldots, \frac{x_n}{x_{n-1}}\right\}.$$

This non-parametric estimate was called "quick and dirty" in Bell and Smith [1986]. This ratio estimate $\widehat{\alpha}_R$ is always positive, which is not always the case in other approaches. Anděl [1989] showed that it is a strongly consistent estimator for α, and in the simulation study it is better than the least squares estimator when the marginal distribution is the exponential distribution.

Here the process is a particular case, where K is from **P1**. This approach can not be generalized to other self-generalized distribution. It seems that this approach also works in the continuous-time GAR(1) processes with infinite marginal mean functions.

Marginal estimating (ME) approach. Like the ratio approach, this approach can help us to estimate the parameter μ too. It works in not only the equally-spaced case but also the unequally-spaced case. The idea is to estimate those parameters in marginal distribution first, then use them to estimate the marginal mean or variance, and substitute these estimated mean and variance in the sum of conditional least squares or conditional weighted least squares. In this way, we will obtain an objective function with only parameter μ. Thus, we get strength from the information of marginal distribution. For illustration, we consider the sum of conditional least squares:

$$\begin{aligned} R_{CLS} &= \sum_{i=1}^{n-1}\left(x_{i+1} - e^{-\mu(t_{i+1}-t_i)}x_i - A(\boldsymbol{\theta})\left[1 - e^{-\mu(t_{i+1}-t_i)}\right]\right)^2 \\ &= \sum_{i=1}^{n-1}\left([x_{i+1} - A(\boldsymbol{\theta})] - e^{-\mu(t_{i+1}-t_i)}[x_i - A(\boldsymbol{\theta})]\right)^2. \end{aligned}$$

Reparametrize μ as $\alpha_0 = e^{-\mu}$ and estimate $A(\boldsymbol{\theta})$ by $R_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$. Then, we obtain

$$R_{CLS(ME)}(\alpha_0) = \sum_{i=1}^{n-1}\left([x_{i+1} - R_1] - \alpha_0^{t_{i+1}-t_i}[x_i - R_1]\right)^2. \tag{10.4.4}$$

Now it is relatively easy to find the minimum point of α_0, because $R_{CLS(ME)}(\alpha_0)$ is a univariate function on the bounded domain $(0,1)$. We can draw the plot of function $R_{CLS(ME)}(\alpha_0)$ in $(0,1)$ which can even allow us to identify a rough estimation by eye.

This method can be combined with other methods to find the estimates of the entire parameter vector $\boldsymbol{\theta}$. Such an estimate of $\boldsymbol{\theta}$ can in turn serve as the initial value for other approaches like maximum likelihood which optimize a non-linear objective function.

Robust approach. Sometimes outliers may have a big influence on parameter estimates. To reduce such influence, we may consider other convex objective function other than quadratic forms. One common choice is to change the L_2 function like R_{CLS}, the sum of conditional least squares, to L_1:

$$R_{CL_1}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \Big| x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}] \Big|. \tag{10.4.5}$$

Minimizing such L_1 objective function, we shall obtain a more robust estimate:

$$\widehat{\boldsymbol{\theta}}_{CL_1} = \arg\min_{\theta \in \Theta} R_{CL_1}(\boldsymbol{\theta}).$$

However, to obtain the robust estimates, numerical methods are inevitable.

Diagonal probability least squares (DPLS) approach. This is a new estimating approach inspired by the diagnostic technique developed in Section 12.3.1. It considers the bivariate cumulative distribution function along the diagonal line through the first quadrant, namely $F_{12}(x, x)$. For simplicity, we take the equally-spaced time series as the example and consider the lag-1 pairs $(X(t_{i+1}), X(t_i))$ $(i = 1, 2, \ldots, n-1)$. Let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. Denote

$$\widehat{F}_{12}(x_{(j)}, x_{(j)}) = \frac{\text{the number of } (X(t_i), X(t_{i-1}) \text{ where } X(t_i) \leq x_{(j)} \text{ and } X(t_{i-1}) \leq x_{(j)}}{n-1},$$

the empirical bivariate cdf at point $(x_{(j)}, x_{(j)})$, and

$$\begin{aligned} F_{12}(x_{(j)}, x_{(j)}) &= \Pr[X(0) \leq x_{(j)},\ X(t_2 - t_1) \leq x_{(j)}] \\ &= \Pr[X(0) \leq x_{(j)},\ (\alpha)_K \circledast X(0) + E(0, t_2 - t_1) \leq x_{(j)}], \end{aligned}$$

the theoretical bivariate cdf at point $(x_{(j)}, x_{(j)})$. We consider the sum of diagonal probability least squares

$$R_{DPLS}(\boldsymbol{\theta}) = \sum_{j=1}^{n-1} \Big[\widehat{F}_{12}(x_{(j)}, x_{(j)}) - F_{12}(x_{(j)}, x_{(j)})\Big]^2, \tag{10.4.6}$$

which measures the closeness of the observations with the model in the sense of diagonal probability. Minimizing $R_{DPLS}(\boldsymbol{\theta})$, we can obtain the diagonal PLS estimate:

$$\widehat{\boldsymbol{\theta}}_{DPLS} = \arg\min_{\theta\in\Theta} R_{DPLS}(\boldsymbol{\theta}).$$

This estimate depends on the choice of time difference between the pairs. Numerical methods are needed to find $F_{12}(x,y)$, as well as the solution of minimum.

Subset-observation approach. In practice, we may encounter the zero-inflated situation for count data. This motivates us to consider the subset-observations which can be viewed as independent innovation samples. For a stationary continuous-time GAR(1) process, conditioned on $X(t_i) = x_i$, it follows that

$$[X(t_{i+1})|X(t_i) = x_i] \stackrel{d}{=} \left(e^{-\mu(t_{i+1}-t_i)}\right)_K \circledast x_i + E(t_i, t_{i+1}).$$

If $x_i = 0$, then $\left(e^{-\mu(t_{i+1}-t_i)}\right)_K \circledast x_i = 0$ and $[X(t_{i+1})|X(t_i) = x_i] \stackrel{d}{=} E(t_i, t_{i+1})$. This implies that the observation x_{i+1} is an outcome of the cumulative innovation $E(t_i, t_{i+1})$. These cumulative innovations are independent each other. Hence, such subset-observations can be considered as independent replications. This feature may allow us to simplify the estimation.

Let $\{y_1, y_2, \dots, y_l\}$ being the subset-observations whose previous observations are zeros. Usually, we know the pgf, or LT, or cf of the cumulative innovation, even the pmf or pdf in some special cases. Then the maximum likelihood, or least squares, or empirical characteristic function approach can be based on the subset-observations $\{y_1, y_2, \dots, y_l\}$.

This subset-observation idea can be extended to other cases where x_i is a fixed number other than 0, because conditioned on a fixed number, say $x_i = l$,

$$[X(t_{i+1})|X(t_i) = l] \stackrel{d}{=} \sum_{j=0}^{l} K_j \left(e^{-\mu(t_{i+1}-t_i)}\right) + E(t_i, t_{i+1}),$$

are still independent of one another in the subset $\{X(t_j) : X(t_{j-1}) = l\}$.

In summary, these approaches seem to provide rough estimates. They can be the initial values of numerical solutions for the better estimates. There are some other estimation approaches for specific models in the literature. People should be aware of the features of specific models. This may help us to develop special parameter estimation methods for them.

10.5 Numerical solution of optimization

In previous sections, we often encounter the function maximization or minimization in parameter estimation, such as maximizing the log-likelihood function or minimizing a quadratic form. Usually, we can not obtain closed form solutions. Hence, numerical methods have to be employed. A good reference on various methods of optimization of functions is Press, Teukolsky, Vetterling and Flannery [1996].

Maximizing a function is equivalent to minimizing the negative of the function. For the sake of simplicity, we use function minimization to unify the optimization issue. Among those optimization approaches, we favor the variable metric algorithms which are also known as quasi-Newton algorithms, especially if it is tedious to obtain derivatives of the function to be minimized. A good introduction to this method can be found in Nash [1990], Section 15.3. This method also provides the numerical evaluation of the asymptotic covariance matrix.

Suppose $R(\boldsymbol{\theta})$ is a real function with argument $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$. Now our task is to minimize $R(\boldsymbol{\theta})$. Let the gradient of the function $R(\boldsymbol{\theta})$ be $\mathbf{g}(\boldsymbol{\theta}) = (g_1, \ldots, g_k)'$, where

$$g_i = \frac{\partial R(\boldsymbol{\theta})}{\partial \theta_i}, \qquad i = 1, \ldots, k,$$

and the Hessian matrix be $H(\boldsymbol{\theta}) = (H_{ij})_{k\times k}$, where

$$H_{ij} = \frac{\partial g_i(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial^2 R(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}, \qquad i,\ j = 1, \ldots, k.$$

Then all the variable metric methods seek to minimize the function $R(\boldsymbol{\theta})$ by means of a sequence of steps

$$\boldsymbol{\theta}' = \boldsymbol{\theta} - k\mathbf{B}\mathbf{g}(\boldsymbol{\theta}),$$

where k is a step length, $\mathbf{B}$ is an approximation of the inverse of Hessian matrix. This means that the search direction at each iteration step is $-\mathbf{B}\mathbf{g}(\boldsymbol{\theta})$. $\mathbf{B}$ is obtained iteratively and does not require the analytical form of $H(\boldsymbol{\theta})$. Also $\mathbf{g}(\boldsymbol{\theta})$ can be computed as a numerical derivative rather than in analytical form. Different approximation methods for $\mathbf{B}$ lead to different variable metric algorithms.

Chapter 11

Asymptotic study of estimators

Asymptotic properties of estimators of parameters have been always a key topic in statistical inference. These basically mean the consistency and asymptotic normality of the parameter estimators, namely the convergence, in probability and in distribution, when sample size n goes to infinity, of the estimates to the true values of parameters. An asymptotic analysis can help us not only in choosing better estimators (in the sense of small asymptotic variances), but also in obtaining asymptotic confidence intervals or regions of the parameters.

In our study, we consider the data $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ from a stationary continuous-time GAR(1) process $\{X(t); t \geq 0\}$. They are observed at either equally-spaced time points for which $t_2 - t_1 = \cdots = t_n - t_{n-1}$, or unequally-spaced time points for which the time differences $t_2 - t_1, \ldots, t_n - t_{n-1}$ are not all equal. In principle, both cases can be equivalently seen as samples from a discrete-time process $\{X(t); t = 0, 1, 2, \ldots\}$. However, there is an obvious difference between the two cases: the resulting process from sampling at equally-spaced time points has constant transition probabilities, while the resulting process from sampling at unequally-spaced time points has time-varying transition probabilities. One common feature for both types of observations is that the marginal distributions are the same if the continuous-time process is in steady state.

In Chapter 10, we have studied estimation methods based on maximum likelihood, conditional least squares and the empirical characteristic function, etc. The asymptotic study of these estimators for the stationary discrete-time process case has been well studied: Billingsley [1961a] first discussed the results for the MLE, and gave a fundamental Central Limit theorem for martingale; Klimko and Nelson [1978] investigated CLS estimation; Feuerverger and McDunnough [1981b], Feuerverger [1990] studied the estimation based on empirical characteristic function. Basawa and Prakasa Rao [1980], Nanthi [1983], Nanthi and Wasan [1987] summarized and studied the asymptotic properties of many estimators (except for the ECF estimator) for various processes, while Ushakov [1999] provided a rich collection of results for the ECF estimator. These results for ML, or CLS or ECF estimators are obtained for general stationary process families. For the specific process in Section 10.4, Bell and Smith [1986] proved strong consistency of the ratio estimator. Chaganty [1997] showed consistency and asymptotic normality for the quasi-least squares estimator in a multivariate setting.

Therefore, for the stationary continuous-time GAR(1) process, a special case in the stationary process family, we can directly adopt existing results for the case of equally-spaced time observations. What we should do is to investigate the asymptotic properties of the estimators for unequally-spaced time observations. For such a situation, the maximum likelihood estimator and the conditional least squares estimator are applicable. Hence, our task will focus on these two kinds of estimators for the unequally-spaced time observations.

In Section 11.1, we propose a random sampling scheme and some assumptions, as well as the fundamental results needed for the proof of asymptotic properties. Sections 11.2 and 11.3 have results for the MLE and CLS estimator respectively.

11.1 Random sampling scheme, assumptions and fundamental theorem

First, we discuss why the unequally-spaced time observations happen in reality. This will help us to propose a plausible random sampling scheme from a continuous-time Markov process.

Usually, for a study which requires repeated measurements over time, the experiment will be arranged to make observations at equally-spaced time points. For example, in a clinical trial study, patients may be asked to visit the clinic every six weeks. However, due to various reasons, some subjects may not appear at scheduled times. They may come earlier or later, or even do not appear for a specific scheduled test(missing values!). Such an occurrence can not be controlled in advance. It is somehow random. Hence, instead of planned equally-spaced time observations, a random sampling scheme may happen, yielding the unequally-spaced time observations. This random sampling scheme is equivalent to a waiting time process in a recurrent dynamic system. But it is usually not observable.

Based on such investigation, we propose a random sampling scheme which results in unequally-spaced time observations. Let $T_1, T_2, \ldots, T_n, \ldots$ be iid positive random variables with distribution function $F_T(t)$, where $t > 0$. Suppose for a fixed n, the values of T_i $(i = 1, 2, \ldots, n)$ are

$$T_1 = t_1, \quad T_2 = t_2 - t_1, \quad \ldots, \quad T_n = t_n - t_{n-1}.$$

Observations are made at time points $t_1, t_2, \ldots, t_n$ of a continuous-time Markov process $\{X(t); t \geq 0\}$, yielding

$$X(t_1) = x_1, \quad X(t_2) = x_2, \quad \ldots, \quad X(t_n) = x_n.$$

Here the T_i's can be seen as waiting time between two successive events. A special case is that when all T_n $(n \geq 1)$ have a degenerate distribution with mass 1 on a single point Δt. Then, the waiting times are common, leading to a discrete-time process sample of equally-spaced time points from the underlying continuous-time process $\{X(t); t \geq 0\}$.

We pursue the consistency and asymptotic normality of MLE's and CLS estimators. For this purpose, we need some assumptions regarding such the random sampling scheme for each estimation method.

The goal of maximum likelihood estimation is to maximize

$$\begin{aligned} \log L(\boldsymbol{\theta} \mid \mathbf{x}) \;&=\; \log f_{X(t_1)}(x_1; \boldsymbol{\theta}) + \log f_{X(t_2)|X(t_1)}(x_2 \mid x_1; \boldsymbol{\theta}) + \cdots \\ &\quad + \log f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1}; \boldsymbol{\theta}), \end{aligned}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)'$. For large sample theory, as the sample size n goes to infinity, the influence of the first term, $\log f_{X(t_1)}(x_1; \boldsymbol{\theta})$, will reduce to zero, so we can ignore it. This leads to the maximization of

$$\log L_1(\boldsymbol{\theta} \mid \mathbf{x}) = \log f_{X(t_2)|X(t_1)}(x_2 \mid x_1; \boldsymbol{\theta}) + \cdots + \log f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1}; \boldsymbol{\theta}),$$

the logarithm of conditional likelihood function. Denote

$$g(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}) = \log f_{X(t_{i+1})|X(t_i)}(x_{i+1} \mid x_i; \boldsymbol{\theta}), \qquad i = 1, 2, \ldots, n-1.$$

To associate the log-likelihood function with sample size n, we rewrite it as

$$\begin{aligned}\log L_n(\boldsymbol{\theta} \mid \mathbf{x}) &= \log f_{X(t_2)|X(t_1)}(x_2 \mid x_1; \boldsymbol{\theta}) + \cdots + \log f_{X(t_n)|X(t_{n-1})}(x_n \mid x_{n-1}; \boldsymbol{\theta}) \\ &= \sum_{i=1}^{n-1} g(x_i, x_{i+1}; \boldsymbol{\theta}, t_1, t_2). \qquad (11.1.1)\end{aligned}$$

If the MLE is not on the boundary of the parameter space, then it is obtained from the estimating equations

$$\frac{\partial}{\partial \theta_j} \log L_n(\boldsymbol{\theta} \mid \mathbf{x}) = 0, \qquad i = 1, 2, \ldots, k. \qquad (11.1.2)$$

In the asymptotic study of the MLE estimator, the classical technique is to expand $\frac{\partial}{\partial \theta_j} \log L_n(\boldsymbol{\theta} \mid \mathbf{x})$ $(j = 1, 2, \ldots, k)$ around the true value $\boldsymbol{\theta}^0$. Hence, the following assumption is required and plays an important role in the proof of asymptotic properties.

Assumption 11.1.1 *Suppose the parameter space Θ is an open set in $\Re^k$, and $\boldsymbol{\omega}$ is a small neighborhood of true parameter value $\boldsymbol{\theta}^0$: $\boldsymbol{\omega} = \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}^0|| < \delta, \delta > 0\}$. For $i = 1, 2, \ldots, n-1$, $g(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1})$ is thrice continuously differentiable with respect to $\boldsymbol{\theta}$. Denote*

$$\begin{aligned} g'_j(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}) &= \frac{\partial}{\partial \theta_j} g(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}), \\ g''_{j_1 j_2}(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}) &= \frac{\partial^2}{\partial \theta_{j_1} \partial \theta_{j_2}} g(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}), \\ g'''_{j_1 j_2 j_3}(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}) &= \frac{\partial^3}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3}} g(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}), \\ G(x_i, x_{i+1}; t_i, t_{i+1}) &= \sup_{\boldsymbol{\theta} \in \boldsymbol{\omega}} |g'''_{j_1 j_2 j_3}(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1})|, \qquad (11.1.3) \end{aligned}$$

where $j, j_1, j_2, j_3 = 1, 2, \ldots, k$. As n goes to infinity, assume that

(1) $\quad (n-1)^{-1} \sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \xrightarrow{\mathrm{P}} \int_0^\infty \mathbf{E}\left(g_j'(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\right) dF_T(t),$

(2) $\quad (n-1)^{-1} \sum_{i=1}^{n-1} g_{j_1 j_2}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \xrightarrow{\mathrm{P}}$

$$\int_0^\infty \mathbf{E}\left(g_{j_1 j_2}''(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\right) dF_T(t),$$

(3) $\quad (n-1)^{-1} \sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}) \xrightarrow{\mathrm{P}} \int_0^\infty \mathbf{E}\left(G(X(t_0), X(t_0+t); t_0, t_0+t)\right) dF_T(t),$

where F_T is the "distribution" of $\{t_{i+1} - t_i\}$, $1 \le j,\ j_1,\ j_2 \le k$. Also assume that all integrals on the right hand sides are finite.

Remark: Assumption 11.1.1 is reasonable under the random sampling scheme designed for unequally-spaced time observations. For equally-spaced time observations which is a special case in the random sampling scheme, the sample forms a discrete-time process. Under ergodicity, (1)–(3) hold as facts, not assumptions, namely it follows that

$$(n-1)^{-1} \sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \xrightarrow{\mathrm{P}} \mathbf{E}\left(g_j'(X(t_0), X(t_0+\Delta t); \boldsymbol{\theta}^0, t_0, t_0+\Delta t)\right),$$

$$(n-1)^{-1} \sum_{i=1}^{n-1} g_{j_1 j_2}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \xrightarrow{\mathrm{P}} \mathbf{E}\left(g_{j_1 j_2}''(X(t_0), X(t_0+\Delta t); \boldsymbol{\theta}^0, t_0, t_0+\Delta t)\right),$$

$$(n-1)^{-1} \sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}) \xrightarrow{\mathrm{P}} \mathbf{E}\left(G(X(t_0), X(t_0+\Delta t); t_0, t_0+\Delta t)\right),$$

as n goes to infinity, where Δt is the common time difference between two successive observations. Here the expectations are taken with respect to the random sampling scheme.

A rough interpretation of Assumption 11.1.1 for unequally-spaced time observations is given below. For the sake of simplicity, we suppose the underlying process is a discrete-time Markov process: $\{X(t); t = 0, 1, 2, \ldots\}$. The sample $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ is observed under the random sampling scheme. Hence, we can arrange $n-1$ successive pairs:

$$(X(t_1), X(t_2)), \quad (X(t_2), X(t_3)), \quad \ldots, \quad (X(t_{n-1}), X(t_n)).$$

Let n_j be the number of pairs with time difference equal to j, i.e., $t_{i+1} - t_i = j$, where $j \in \mathcal{N}$. As n and all the nonzero n_j's go to infinity, we have

$$\frac{n_j}{n-1} \xrightarrow{\text{P}} F_T(j) - F_T(j-1), \qquad j = 1, 2, \ldots.$$

Hence, for the average of a summation with summand being the function of the successive pairs like $h(X(t_i), X(t_{i+1}))$, it follows that

$$(n-1)^{-1} \sum_{i=1}^{n-1} h(X(t_i), X(t_{i+1})) \;=\; \frac{n_1}{n-1} \cdot n_1^{-1} S_1 + \frac{n_2}{n-1} \cdot n_2^{-1} S_2 + \cdots + \frac{n_j}{n-1} \cdot n_j^{-1} S_j + \cdots,$$

where S_j ($j = 1, 2, \ldots$) is the sum consisting of summands of function of successive pairs with time lag j. Assume ergodicity holds. Then,

$$n_j^{-1} S_j \xrightarrow{\text{P}} \mathbf{E}[h(X(t_0), X(t_0 + j))], \qquad j = 1, 2, \ldots.$$

Thus,

$$\begin{aligned}(n-1)^{-1} \sum_{i=1}^{n-1} h(X(t_i), X(t_{i+1})) &\xrightarrow{\text{P}} \sum_{j=1}^{\infty} \mathbf{E}[h(X(t_0), X(t_0 + j))][F_T(j) - F_T(j-1)] \\ &= \int_0^{\infty} \mathbf{E}[h(X(t_0), X(t_0 + t))] dF_T(t).\end{aligned}$$

A continuous-time process can be approximated by a sequence of discrete-time processes. Thus, for the continuous-time underlying process, this limit can be expected to hold.

The above assumptions just try to generalize the facts which hold for stationary and ergodic discrete-time processes to the unequally-spaced case based on random sampling scheme from a stationary continuous-time process. We don't know if they hold as facts under certain conditions. This is left as an open question.

The conditional least squares estimator is obtained by minimizing

$$R_n(\boldsymbol{\theta}) = R_{CLS}(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{n-1} \Big(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}] \Big)^2.$$

For $i = 1, 2, \ldots, n-1$, let

$$\begin{aligned} g(x_i; \boldsymbol{\theta}, t_i, t_{i+1}) &= \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}], \\ u(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}) &= x_{i+1} - g(x_i; \boldsymbol{\theta}, t_i, t_{i+1}). \end{aligned}$$

Then

$$R_n(\boldsymbol{\theta}) = \sum_{i=1}^{n-1}\Big(x_{i+1} - g(x_i;\boldsymbol{\theta},t_i,t_{i+1})\Big)^2 = \sum_{i=1}^{n-1} u^2(x_i,x_{i+1};\boldsymbol{\theta},t_i,t_{i+1}). \tag{11.1.4}$$

For an asymptotic analysis, the traditional approach is to expand $R_n(\boldsymbol{\theta})$ around the true parameter value $\boldsymbol{\theta}^0$. Hence, the following assumptions play a critical role in the asymptotic properties of the CLS estimator.

Assumption 11.1.2 *Suppose the parameter space Θ is an open set in $\Re^k$, and $\boldsymbol{\omega}$ is a small neighborhood of true parameter value $\boldsymbol{\theta}^0$: $\boldsymbol{\omega} = \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}^0|| < \delta, \delta > 0\}$. For $i = 1, 2, \ldots, n-1$, $g(x_i;\boldsymbol{\theta},t_i,t_{i+1})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$. Denote*

$$\begin{aligned}
g'_j(x_i;\boldsymbol{\theta},t_i,t_{i+1}) &= \frac{\partial}{\partial\theta_j} g(x_i;\boldsymbol{\theta},t_i,t_{i+1}),\\
g''_{j_1j_2}(x_i;\boldsymbol{\theta},t_i,t_{i+1}) &= \frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}} g(x_i;\boldsymbol{\theta},t_i,t_{i+1}),\\
\mathbf{V}_n &= \left(\frac{\partial^2 R_n(\boldsymbol{\theta}^0)}{\partial\theta_{j_1}\partial\theta_{j_2}}\right)_{k\times k},\\
\mathbf{W}_n(\boldsymbol{\theta}^*) &= (W_{j_1j_2}(\boldsymbol{\theta}^*))_{k\times k} = \left(\frac{\partial^2 R_n(\boldsymbol{\theta}^*)}{\partial\theta_{j_1}\partial\theta_{j_2}}\right)_{k\times k} - \mathbf{V}_n, \qquad \boldsymbol{\theta}^* \in \boldsymbol{\omega},
\end{aligned}$$

where $j, j_1, j_2 = 1, 2, \ldots, k$. Assume that as $n \to \infty$,

(1) $\displaystyle \lim_{n\to\infty}\sup_{\delta\to 0}\left(\frac{|W_{j_1j_2}(\boldsymbol{\theta}^*)|}{(n-1)\delta}\right) < \infty,$ *a.s.*

(2) $\displaystyle (n-1)^{-1}\sum_{i=1}^{n-1} u(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1})g'_j(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1}) \xrightarrow{\text{a.s.}} \int_0^\infty \mathbf{E}\left(u(X(t_0),X(t_0+t);\boldsymbol{\theta}^0,t_i,t_{i+1})g'_j(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right) dF_T(t),$

(3) $\displaystyle (n-1)^{-1}\sum_{i=1}^{n-1} g'_{j_1}(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})g'_{j_2}(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1}) \xrightarrow{\text{a.s.}} \int_0^\infty \mathbf{E}\left(g'_{j_1}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)g'_{j_2}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right) dF_T(t),$

(4) $\displaystyle (n-1)^{-1}\sum_{i=1}^{n-1} u(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1})g''_{j_1j_2}(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1}) \xrightarrow{\text{a.s.}} \int_0^\infty \mathbf{E}\left(u(X(t_0),X(t_0+t);\boldsymbol{\theta}^0,t_i,t_{i+1})g''_{j_1j_2}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right) dF_T(t),$

where F_T is the "distribution" of $\{t_{i+1} - t_i\}$, $1 \leq j,\ j_1,\ j_2 \leq k$. Also assume that all integrals on the right hand sides are finite.

Remark: (1) in Assumption 11.1.2 is inherited from conventional regularity conditions. Similar to Assumption 11.1.1, (2), (3) and (4) in Assumption 11.1.2 are reasonable under the random sampling scheme designed for unequally-spaced time observations. For equally-spaced time observations which is a special case in the random sampling scheme, the sample forms a discrete-time process. Under ergodicity, as well as other conditions (see Klimko and Nelson [1978], section 3), (2), (3) and (4) are facts, not assumptions. Unlike Assumption 11.1.1, here we require convergence almost surely, not in probability. Corresponding, the CLS estimator will be strongly consistent.

The asymptotic normality of both estimators makes use of the central limit theorem for martingales, which was given by Billingsley [1961a]. We refer to this theorem as the fundamental theorem for the asymptotic normality of an estimator in a Markov process.

Theorem 11.1.1 (Central Limit Theorem for Martingales)

Let $u_1, u_2, \ldots$ be random variables with moments of order $2+d$ $(d>0)$, and let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ be a filtration of Borel fields such that

$$\mathbf{E}\left(u_n \mid \mathcal{F}_{n-1}\right) \;=\; 0, \qquad n = 1, 2, \ldots$$

with probability one. Here $\mathcal{F}_{n-1}$ is the σ-algebra generated by $u_1, u_2, \ldots, u_{n-1}$. Suppose that

(1) $$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbf{E}\left(u_i^2 \mid \mathcal{F}_{i-1}\right) \;=\; D, \qquad D \geq 0,$$

(2) $$\lim_{n\to\infty} n^{-1-d/2} \sum_{i=1}^{n} \mathbf{E}\left(u_i^{2+d} \mid \mathcal{F}_{i-1}\right) \;=\; 0,$$

with probability one. Then

$$n^{-1/2} \sum_{i=1}^{n} u_i \;\xrightarrow{\mathrm{L}}\; \mathrm{N}\left(0, D\right).$$

This result can be generalized to the multivariate situation where $\mathbf{u}_n = (u_{n1}, u_{n2}, \ldots, u_{nk})^T$. Each of the components has moment of order $2+d$ $(d>0)$ and

$$\mathbf{E}\left(u_{nj} \mid \mathcal{F}_{n-1}\right) \;=\; 0, \qquad j = 1, 2, \ldots, k; \;\; n = 1, 2, \ldots$$

with probability one. Suppose that

(1) $\quad \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \mathbf{E}\left(u_{ij}u_{il} \mid \mathcal{F}_{i-1}\right) = D_{jl}, \qquad j,\ l = 1, 2, \ldots, k,$

(2) $\quad \lim_{n\to\infty} n^{-1-d/2} \sum_{i=1}^{n} \mathbf{E}\left(u_{ij}^{2+d} \mid \mathcal{F}_{i-1}\right) = 0, \qquad j = 1, 2, \ldots, k,$

with probability one, where $\mathbf{D} = (D_{jl})_{k\times k}$ *is a non-negative definite matrix. Then*

$$n^{-1/2} \sum_{i=1}^{n} \mathbf{u}_i \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}_{k\times 1}, \mathbf{D}_{k\times k}\right).$$

For the univariate situation, since $\mathbf{E}\left(u_n \mid \mathcal{F}_{n-1}\right) = 0$ for all n, the sequence of partial sums $\{S_n = \sum_{i=1}^{n} u_i; n = 1, 2, \ldots\}$ forms a martingale because

$$\mathbf{E}\left(S_n \mid S_{n-1}\right) = S_{n-1} + \mathbf{E}\left(u_n \mid S_{n-1}\right) = S_{n-1} + 0 = S_{n-1}, \qquad n = 2, 3, \ldots.$$

Here we assume that $\mathcal{F}_{n-1}$ is the σ-algebra generated by $u_1, u_2, \ldots, u_{n-1}$. For the multivariate situation, the sequence of partial sums of any linear transformation $\{S_n = \sum_{i=1}^{n} \mathbf{a}^T \mathbf{u}_i; n = 1, 2, \ldots\}$ forms a martingale, where $\mathbf{a} = (a_1, a_2, \ldots, a_k)^T$. For the details of the proof, see Billingsley [1961a], Theorem 9.1, p. 52, and Theorem 1.2, p. 6 and p. 61. With these preparations, we shall proceed to the asymptotic study of the MLE estimator and CLS estimator in the next two sections.

11.2 Asymptotic properties of MLE

As mentioned before, the classical technique to investigate the asymptotic properties of MLE is to take a Taylor expansion for

$$\frac{\partial}{\partial\theta_j} \log L_n(\boldsymbol{\theta} \mid \mathbf{x}) = \sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}), \qquad j = 1, 2, \ldots, k \tag{11.2.1}$$

around the true parameter value $\boldsymbol{\theta}^0$. By the Mean Value Theorem, if $\boldsymbol{\theta} \in \boldsymbol{\omega}$, then

$$\begin{aligned} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}) &= g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) + \sum_{l=1}^{k} (\theta_l - \theta_l^0) g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \\ &\quad + c\|\boldsymbol{\theta} - \boldsymbol{\theta}^0\|^2 G(x_i, x_{i+1}; t_i, t_{i+1}), \qquad |c| \le k^2/2, \end{aligned}$$

where G is defined in (11.1.3). Hence, it follows that for $j = 1, 2, \ldots, k$,

$$\begin{aligned}(n-1)^{-1}\frac{\partial}{\partial\theta_j}\log L_n(\boldsymbol{\theta} \mid \mathbf{x}) &= (n-1)^{-1}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \\ &\quad + \sum_{l=1}^{k}(\theta_l - \theta_l^0)\left[(n-1)^{-1}\sum_{i=1}^{n-1} g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right] \\ &\quad + c||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2 (n-1)^{-1}\sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}). \end{aligned} \tag{11.2.2}$$

By controlling the behavior of

$$(n-1)^{-1}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}), \qquad (n-1)^{-1}\sum_{i=1}^{n-1} g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}),$$
$$(n-1)^{-1}\sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}),$$

we may further obtain a simpler approximation as n goes to infinity, and obtain the consistency and asymptotic normality of the MLE. For this purpose, we investigate Assumption 11.1.1, and proceed to the regularity conditions for the asymptotic properties.

First, for fixed t and $j = 1, 2, \ldots, k$, it follows that

$$\begin{aligned}&\mathbf{E}\left\{g_j'(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\right\} \\ &= \mathbf{E}\left\{\mathbf{E}\left[\frac{\partial}{\partial\theta_j}\log f_{X(t_0+t)|X(t_0)}(X(t_0+t)|X(t_0); \boldsymbol{\theta})\Big| X(t_0) = x\right]\right\} \\ &= \mathbf{E}\left\{\int \frac{\partial f_{X(t_0+t)|X(t_0)}(y|X(t_0); \boldsymbol{\theta})}{\partial\theta_j}[f_{X(t_0+t)|X(t_0)}(y|X(t_0); \boldsymbol{\theta})]^{-1} f_{X(t_0+t)|X(t_0)}(y|X(t_0); \boldsymbol{\theta}) dy\right\} \\ &= \mathbf{E}\left\{\int \frac{\partial f_{X(t_0+t)|X(t_0)}(y|X(t_0); \boldsymbol{\theta})}{\partial\theta_j} dy\right\}.\end{aligned}$$

Traditionally, one imposes that the differentiation with respect to parameter $\boldsymbol{\theta}$ can be carried out equivalently both inside and outside of integral sign. If so, it will yield

$$\begin{aligned}\mathbf{E}\left\{g_j'(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\right\} &= \mathbf{E}\left\{\int \frac{\partial f_{X(t_0+t)|X(t_0)}(y|X(t_0); \boldsymbol{\theta})}{\partial\theta_j} dy\right\} \\ = \mathbf{E}\left\{\frac{\partial}{\partial\theta_j}\int f_{X(t_0+t)|X(t_0)}(y|X(t_0); \boldsymbol{\theta})\, dy\right\} &= \mathbf{E}\left\{\frac{\partial(1)}{\partial\theta_j}\right\} = 0.\end{aligned} \tag{11.2.3}$$

This further leads to

$$\int_0^\infty \mathbf{E}\left(g_j'(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\right) dF_T(t) \;=\; \int_0^\infty 0\, dF_T(t) \;=\; 0. \tag{11.2.4}$$

Secondly, for fixed t and $j = 1, 2, \ldots, k$,

$$\begin{aligned}
&\mathbf{E}\left\{g''_{j_1 j_2}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\right\} \\
&= \mathbf{E}\left\{\mathbf{E}\left[\frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}} \log f_{X(t_0+t)|X(t_0)}\left(y|X(t_0);\boldsymbol{\theta}\right) \Big| X(t_0) = x\right]\right\} \\
&= \mathbf{E}\Bigg\{\int \frac{\partial^2 f_{X(t_0+t)|X(t_0)}\left(y|X(t_0);\boldsymbol{\theta}\right)}{\partial\theta_{j_1}\partial\theta_{j_2}} dy \\
&\qquad - \int \frac{\partial f_{X(t_0+t)|X(t_0)}\left(y|X(t_0);\boldsymbol{\theta}\right)}{\partial\theta_{j_1}} \cdot \frac{\partial f_{X(t_0+t)|X(t_0)}\left(y|X(t_0);\boldsymbol{\theta}\right)}{\partial\theta_{j_2}} \\
&\qquad\qquad \times [f_{X(t_0+t)|X(t_0)}\left(y|X(t_0);\boldsymbol{\theta}\right)]^{-2} f_{X(t_0+t)|X(t_0)}\left(y|X(t_0);\boldsymbol{\theta}\right) dy\Bigg\} \\
&= \mathbf{E}\left\{\frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}} \int f_{X(t_0+t)|X(t_0)}\left(y|X(t_0);\boldsymbol{\theta}\right) dy\right\} \\
&\qquad - \mathbf{E}\left\{\mathbf{E}\left[g'_{j_1}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t) g'_{j_2}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t) \Big| X(t_0) = x\right]\right\} \\
&= -\mathbf{E}\left\{\mathbf{Cov}\left[g'_{j_1}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t), g'_{j_2}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t) \Big| X(t_0) = x\right]\right\}.
\end{aligned} \tag{11.2.5}$$

The last step is due to (11.2.3). Since

$$\left(\mathbf{Cov}\left[g'_{j_1}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t), g'_{j_2}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t) \Big| X(t_0) = x\right]\right)_{k\times k}$$

is the covariance matrix of the random vector

$$\left(g_1'(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t),\ \ldots,\ g_k'(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\right)^T$$

conditioned on $X(t_0) = x$, it is a non-negative definite matrix. Consequently, it follows that

$$\left(\mathbf{E}\left\{\mathbf{Cov}\left[g'_{j_1}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t), g'_{j_2}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t) \Big| X(t_0)\right]\right\}\right)_{k\times k}$$

and

$$\boldsymbol{\Sigma}_{k\times k} = (\sigma_{j_1 j_2})_{k\times k} = -\left(\int_0^\infty \mathbf{E}\left(g''_{j_1 j_2}(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\right) dF_T(t)\right)_{k\times k} \tag{11.2.6}$$

are also non-negative definite matrices.

Denote

$$\frac{\partial h(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial h(\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial h(\boldsymbol{\theta})}{\partial \theta_2}, \dots, \frac{\partial h(\boldsymbol{\theta})}{\partial \theta_k} \right)^T, \qquad \frac{\partial^2 h(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \left(\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_{j_1} \partial \theta_{j_2}} \right)_{k \times k}.$$

Then (11.2.2) can be rewritten in a vector form

$$\begin{aligned} (n-1)^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta} \mid \mathbf{x}) &= (n-1)^{-1} \sum_{i=1}^{n-1} \frac{\partial}{\partial \boldsymbol{\theta}} g(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \\ &+ \left((n-1)^{-1} \sum_{i=1}^{n-1} g''_{jl}(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \right)_{k \times k} (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\ &+ ||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2 \left[(n-1)^{-1} \sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}) \right] \mathbf{C}, \end{aligned} \tag{11.2.7}$$

where $\mathbf{C}_{k\times 1} = (c_1, c_2, \dots, c_k)^T$ is a constant vector with $|c_j| \le k^2/2$ for $j = 1, 2, \dots, k$. According to Assumption 11.1.1, as $n \to \infty$,

$$(n-1)^{-1} \sum_{i=1}^{n-1} \frac{\partial}{\partial \boldsymbol{\theta}} g(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \xrightarrow{\mathrm{P}} \mathbf{0}_{k\times 1}, \tag{11.2.8}$$

$$\left((n-1)^{-1} \sum_{i=1}^{n-1} g''_{jl}(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \right)_{k \times k} \xrightarrow{\mathrm{P}} -\boldsymbol{\Sigma}_{k\times k}. \tag{11.2.9}$$

This implies that $(n-1)^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta} \mid \mathbf{x})$ will be dominated by

$$\left((n-1)^{-1} \sum_{i=1}^{n-1} g''_{jl}(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}) \right)_{k \times k} (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \tag{11.2.10}$$

and

$$||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2 (n-1)^{-1} \sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}) \mathbf{C} \tag{11.2.11}$$

when n goes to infinity. By imposing non-singularity on the matrix $\boldsymbol{\Sigma}$, we can bound the vector (11.2.10) away from the zero vector $\mathbf{0}$ in probability. This is because that all eigenvalues of $\boldsymbol{\Sigma}$ are bigger than 0 and there exists a non-random function $\lambda(\boldsymbol{\theta}^0) > 0$ such that

$$\mathbf{z}^T \boldsymbol{\Sigma} \mathbf{z} \ge \lambda(\boldsymbol{\theta}^0), \qquad ||\mathbf{z}|| = \mathbf{z}^T \mathbf{z} = 1. \tag{11.2.12}$$

Note that $(n-1)^{-1} \sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1})$ is bounded in probability (see (3) in Assumption 11.1.1), i.e.,

$$(n-1)^{-1} \sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}) \xrightarrow{\mathrm{P}} M < \infty, \tag{11.2.13}$$

and $||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2$ is of one higher order than $(\boldsymbol{\theta} - \boldsymbol{\theta}^0)$. We can choose $\boldsymbol{\theta}$ close enough to $\boldsymbol{\theta}^0$ so that every component in vector (11.2.11) has smaller absolute value than the one in the vector (11.2.10). This means that in probability, $(n-1)^{-1}\frac{\partial}{\partial\theta}\log L_n(\boldsymbol{\theta} \mid \mathbf{x})$ will be dominated by (11.2.10) in a small neighborhood of $\boldsymbol{\theta}^0$. Such a feature determines the consistency and asymptotic normality of the MLE.

With the above discussion, we now give further regularity conditions for the consistency of MLE.

Assumption 11.2.1

(1) Conditioned on $X(t_0) = x$, differentiation with respect to $\boldsymbol{\theta}$ for the integration of

$$\int g(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)dF_T(t)$$

is equivalent outside and inside of the integral sign;

(2) The limiting matrix $\boldsymbol{\Sigma}$ in (11.2.6) is non-singular.

Assumption 11.2.1 together with Assumption 11.1.1 lead to the following consistency theorem. For this theorem, we use the technique of proof given by Billingsley [1961a], Theorem 2.1, p. 10. The next two lemmas will be needed in the proof.

Lemma 11.2.1 *If* $\Pr[|U_1| > c_1] \le \epsilon, \ \ldots, \ \Pr[|U_m| > c_m] \le \epsilon$, *then*

$$\Pr\left[|U_1 + \cdots + U_m| > \sum_{i=1}^{m} c_i\right] \le m\epsilon,$$

or equivalently,

$$\Pr\left[|U_1 + \cdots + U_m| \le \sum_{i=1}^{m} c_i\right] > 1 - m\epsilon.$$

Proof: Since $|U_1 + \cdots + U_m| \le |U_1| + \cdots + |U_m|$, we can claim that if event $\left\{|U_1 + \cdots + U_m| > \sum_{i=1}^{m} c_i\right\}$ occurs, then at least one of the following events occurs:

$$\{|U_1| > c_1\}, \quad \ldots, \quad \{|U_m| > c_m\}.$$

Otherwise, event $\left\{|U_1 + \cdots + U_m| \le \sum_{i=1}^{m} c_i\right\}$ must happen. Thus,

$$\Pr\left[|U_1 + \cdots + U_m| > \sum_{i=1}^{m} c_i\right] \;\le\; \Pr\left[\cup_{i=1}^{m}\{|U_i| > c_i\}\right] \;\le\; \sum_{i=1}^{m} \Pr\left[|U_i| > c_i\right] \;\le\; m\epsilon.$$

Lemma 11.2.2 *If $h(\boldsymbol{\theta})$ is a continuous function mapping $\Re^k$ into itself with the property that, for every $\boldsymbol{\theta}$ such that $||\boldsymbol{\theta}|| = 1$, $\boldsymbol{\theta}^T h(\boldsymbol{\theta}) < 0$, then there exists a point $\widehat{\boldsymbol{\theta}}$ such that $||\widehat{\boldsymbol{\theta}}|| < 1$ and $h(\widehat{\boldsymbol{\theta}}) = \mathbf{0}_{k\times 1}$.*

This is Lemma 2 in Aitchison and Silvey [1958]. A nice proof by contradiction can be found there.

Theorem 11.2.3 *Under Assumptions 11.1.1 and 11.2.1, the MLE $\widehat{\boldsymbol{\theta}}_{MLE}$ is consistent for $\boldsymbol{\theta}^0$.*

Proof: As discussed previously, (11.2.8), (11.2.9), (11.2.12) and (11.2.13) will hold under Assumptions 11.1.1 and 11.2.1. Now for any $\epsilon > 0$, we can choose a small $\delta = \delta(\epsilon) > 0$ in such a way that

$$\delta < \epsilon, \qquad \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}^0|| \le \delta\}, \qquad \delta < \lambda(\boldsymbol{\theta}^0)/3k^2(M+1). \tag{11.2.14}$$

After choosing δ, we choose $n_0(\epsilon)$ large enough so that for $n \ge n_0(\epsilon)$,

$$\begin{aligned}
\Pr\left[\left|(n-1)^{-1}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right| < \delta^2\right] &\ge 1 - \epsilon/3, \qquad j = 1, 2, \ldots, k, \\
\Pr\left[0 \le (n-1)^{-1}\sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}) < M + 1\right] &\ge 1 - \epsilon/3, \\
\Pr\left[\left|(n-1)^{-1}\sum_{i=1}^{n-1} g_{j_1 j_2}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) + \sigma_{j_1 j_2}\right| < \delta\right] &\ge 1 - \epsilon/3, \qquad j_1, j_2 = 1, 2, \ldots, k.
\end{aligned}$$

By (11.2.2) and Lemma 11.2.1, if $n \geq n_0(\epsilon)$ and $||\boldsymbol{\theta} - \boldsymbol{\theta}^0|| \leq \delta$, we have

$$\begin{aligned}
&\left|(n-1)^{-1}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1}) + \sum_{l=1}^{k} \sigma_{jl}(\theta_l - \theta_l^0)\right| \\
&= \left|(n-1)^{-1}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right. \\
&\quad + \sum_{l=1}^{k}(\theta_l - \theta_l^0)\left[(n-1)^{-1}\sum_{i=1}^{n-1} g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) + \sigma_{jl}\right] \\
&\quad \left. + c||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2 (n-1)^{-1}\sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1})\right| \\
&\leq \left|(n-1)^{-1}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right| \\
&\quad + \left|\sum_{l=1}^{k}(\theta_l - \theta_l^0)\left[(n-1)^{-1}\sum_{i=1}^{n-1} g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) + \sigma_{jl}\right]\right| \\
&\quad + \left|c||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2 (n-1)^{-1}\sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1})\right| \\
&\leq \delta^2 + k\delta||\boldsymbol{\theta} - \boldsymbol{\theta}^0|| + k^2||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2(M+1)/2 \\
&\leq \delta^2 + k\delta^2 + k^2||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2(M+1)/2 \ \leq \ \left(1 + k + k^2\frac{M+1}{2}\right)\delta^2 \\
&\leq \left(1 + k + k^2/2\right)(M+1)\delta^2 \ \leq \ 3k^2(M+1)\delta^2
\end{aligned}$$

with probability exceeding $1 - \epsilon$. Thus, by (11.2.14), if $||\boldsymbol{\theta} - \boldsymbol{\theta}^0|| = \delta$, we have

$$\begin{aligned}
&\sum_{j=1}^{k}\left[(n-1)^{-1}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}, t_i, t_{i+1})\right](\theta_j - \theta_j^0) \\
&\leq \ -\sum_{j=1}^{k} \sigma_{jl}(\theta_j - \theta_j^0)(\theta_l - \theta_l^0) + k \cdot 3k^2(M+1)\delta^2 \\
&\leq \ -\lambda(\boldsymbol{\theta}^0)||\boldsymbol{\theta} - \boldsymbol{\theta}^0||^2 + 3k^3(M+1)\delta^2 \ = \ -\lambda(\boldsymbol{\theta}^0)\delta^2 + 3k^3(M+1)\delta^2 \ < \ 0
\end{aligned}$$

with probability exceeding $1 - \epsilon$. According to Lemma 11.2.2, there exists a value $\widehat{\boldsymbol{\theta}}_{MLE}$ such that $||\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0|| < \delta < \epsilon$ and

$$(n-1)^{-1}\sum_{i=1}^{n-1} \frac{\partial}{\partial \boldsymbol{\theta}} g(x_i, x_{i+1}; \widehat{\boldsymbol{\theta}}_{MLE}, t_i, t_{i+1}) \ = \ \mathbf{0}_{k\times 1}$$

with probability exceeding $1-\epsilon$. This completes the proof of the consistency of MLE.

With the consistency of MLE, we now can study its asymptotic normality. First, we give a rough analysis. The maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_{MLE}$ is obtained by equating (11.2.7) to the zero vector. When n goes to infinity, $\widehat{\boldsymbol{\theta}}_{MLE} \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^0$. Ignoring the term of $||\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0||$ with the second order in (11.2.7) and dividing them by $(n-1)^{-1/2}$, we then have

$$(n-1)^{-1/2}\sum_{i=1}^{n-1}\frac{\partial}{\partial\boldsymbol{\theta}}g(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1})-\boldsymbol{\Sigma}_{k\times k}\left[(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE}-\boldsymbol{\theta}^0\right)\right] = \mathbf{0}_{k\times 1}, \qquad \text{as } n\to\infty,$$

or

$$(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE}-\boldsymbol{\theta}^0\right) = \boldsymbol{\Sigma}^{-1}\left[(n-1)^{-1/2}\sum_{i=1}^{n-1}\frac{\partial}{\partial\boldsymbol{\theta}}g(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1})\right], \qquad \text{as } n\to\infty.$$

Because of (11.2.3), the sequence of partial sums of partial derivatives with respect to θ_j ($j = 1,2,\ldots,k$), $\left\{\sum_{i=1}^{n-1} g_j'(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1}) : n=2,3,\ldots\right\}$ is a martingale with respect to $\mathcal{F}_n$, where $\mathcal{F}_n$ = σ-algebra generated by $\{X(t_1),\ldots,X(t_{n-1})\}$. By Theorem 11.1.1 and same conditions, $(n-1)^{-1/2}\sum_{i=1}^{n-1}\frac{\partial}{\partial\boldsymbol{\theta}}g(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1})$ converges in law to the multivariate normal distribution $\mathrm{N}\left(\mathbf{0}_{k\times 1},\boldsymbol{\Sigma}_{k\times k}\right)$. This leads to the result that $(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE}-\boldsymbol{\theta}^0\right)$ converges in law to a multivariate normal distribution. The conditions in Theorem 11.1.1 lead to the following additional regularity conditions in asymptotic normality of MLE.

Assumption 11.2.2 $\frac{\partial}{\partial\theta_j}g(X(t_0),X(t_0+t);\boldsymbol{\theta}^0,t_0,t_0+t)$ ($j=1,2,\ldots,k$) *has moment of order* $2+d$ ($d>0$) *for any* $t_0,t\geq 0$. *Also it satisfies that*

$$\lim_{n\to\infty}(n-1)^{-1-d/2}\sum_{i=1}^{n-1}\mathbf{E}\left(\left[\frac{\partial}{\partial\theta_j}g(x_i,X(t_{i+1});\boldsymbol{\theta}^0,t_i,t_{i+1})\right]^{2+d}\right) = 0, \qquad j=1,2,\ldots,k.$$

The following lemma guarantees that the second order term $||\widehat{\boldsymbol{\theta}}_{MLE}-\boldsymbol{\theta}^0||^2$ in (11.2.7) is negligible.

Lemma 11.2.4 *Suppose* $\mathbf{u}_1,\mathbf{u}_2,\ldots$ *are random vectors in* $\Re^k$ *satisfying*

$$\mathbf{u}_n \xrightarrow{\mathrm{L}} F_0, \qquad \textit{as } n\to\infty,$$

where F_0 is a multivariate distribution in $\Re^k$. Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots$ are random vectors in $\Re^k$ satisfying either

$$||\mathbf{u}_n - \mathbf{v}_n|| \leq \epsilon_n ||\mathbf{u}_n||, \quad \text{and} \quad \epsilon_n \xrightarrow{\mathrm{P}} 0, \quad \text{as } n \to \infty,$$

or

$$||\mathbf{u}_n - \mathbf{v}_n|| \leq \epsilon_n' ||\mathbf{v}_n||, \quad \text{and} \quad \epsilon_n' \xrightarrow{\mathrm{P}} 0, \quad \text{as } n \to \infty.$$

Then, $\mathbf{u}_n - \mathbf{v}_n \xrightarrow{\mathrm{P}} \mathbf{0}_{k\times 1}$*, so that* $\mathbf{v}_n \xrightarrow{\mathrm{L}} F_0$ *as* $n \to \infty$*.*

This is Theorem 10.1 in Billingsley [1961a].

Theorem 11.2.5 *Assume $\widehat{\boldsymbol{\theta}}_{MLE}$ is a root of (11.1.2). Under Assumptions 11.1.1, 11.2.1 and 11.2.2,*

$$(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0\right) \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}_{k\times k}^{-1}\right),$$

where $\boldsymbol{\Sigma}_{k\times k}$ is defined in (11.2.6).

Proof: By Theorem 11.1.1, it follows that

$$(n-1)^{-1/2}\sum_{i=1}^{n-1}\frac{\partial}{\partial\boldsymbol{\theta}}g(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}_{k\times 1}, \boldsymbol{\Sigma}_{k\times k}\right).$$

Since $\widehat{\boldsymbol{\theta}}_{MLE}$ is consistent for $\boldsymbol{\theta}^0$, by (11.2.7), we have

$$\begin{aligned}
\mathbf{0}_{k\times 1} &= (n-1)^{-1/2}\sum_{i=1}^{n-1}\frac{\partial}{\partial\boldsymbol{\theta}}g(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \\
&\quad + \left((n-1)^{-1}\sum_{i=1}^{n-1} g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right)_{k\times k} (n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0\right) \\
&\quad + ||\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0|| \cdot \left|\left|(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0\right)\right|\right| \left[(n-1)^{-1}\sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1})\right] \mathbf{C}.
\end{aligned}$$

Noticing that $||\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0|| \xrightarrow{\mathrm{P}} 0$, we can then obtain

$$\begin{aligned}
&\left|\left|(n-1)^{-1/2}\sum_{i=1}^{n-1}\frac{\partial}{\partial\boldsymbol{\theta}}g(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) - \boldsymbol{\Sigma}_{k\times k}\left[(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0\right)\right]\right|\right| \\
&\leq \epsilon_n \left|\left|(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0\right)\right|\right|, \qquad \text{where } \epsilon_n \xrightarrow{\mathrm{P}} 0. \qquad (11.2.15)
\end{aligned}$$

According to Lemma 11.2.4, we have

$$\mathbf{\Sigma}_{k\times k}\left[(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE}-\boldsymbol{\theta}^0\right)\right] \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}_{k\times 1}, \mathbf{\Sigma}_{k\times k}\right), \qquad \text{as } n\to\infty,$$

which yields

$$(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE}-\boldsymbol{\theta}^0\right) \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}_{k\times 1}, \mathbf{\Sigma}_{k\times k}^{-1}\right), \qquad \text{as } n\to\infty.$$

Remark: In practice, the asymptotic covariance matrix $\mathbf{\Sigma}$ is estimated from the data. A natural estimator is

$$\widehat{\mathbf{\Sigma}} = \left(\widehat{\sigma}_{j_1 j_2}\right)_{k\times k},$$

where

$$\widehat{\sigma}_{j_1 j_2} = -(n-1)^{-1}\sum_{i=1}^{n-1} g''_{j_1 j_2}(x_i, x_{i+1}; \widehat{\boldsymbol{\theta}}_{MLE}, t_i, t_{i+1}), \qquad j_1,\ j_2 = 1,2,\ldots,k. \tag{11.2.16}$$

This asymptotic normality will help us to obtain confidence intervals or regions and hypothesis tests regarding parameters.

Note that the martingale feature of the sequence of partial sums comes from (1) of Assumption 11.2.1, not from the Markov property of the underlying process.

A byproduct is the following theorem, which is also useful in hypothesis testing.

Theorem 11.2.6 $\log L_n(\boldsymbol{\theta} \mid \mathbf{x})$ *is defined as in (11.1.1). Assume* $\widehat{\boldsymbol{\theta}}_{MLE}$ *is a root of (11.1.2). Under Assumptions 11.1.1, 11.2.1 and 11.2.2,*

$$2\left[\max_{\boldsymbol{\theta}\in\boldsymbol{\omega}} \log L_n(\boldsymbol{\theta} \mid \mathbf{x}) - \log L_n(\boldsymbol{\theta}^0 \mid \mathbf{x})\right] \xrightarrow{\mathrm{L}} \chi_k^2, \qquad \text{as } n\to\infty.$$

Proof: Let $\widehat{\boldsymbol{\theta}}_{MLE} = (\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_k)^T$. Continuing from Theorem 11.2.5, by the Mean Value Theorem, we have

$$\begin{aligned} g(x,y;\widehat{\boldsymbol{\theta}}_{MLE},t_0,t_0+t) &= g(x,y;\boldsymbol{\theta}^0,t_0,t_0+t) + \sum_{j=1}^{k}(\widehat{\theta}_j-\theta_j^0)g'_j(x,y;\boldsymbol{\theta}^0,t_0,t_0+t) \\ &\quad + \frac{1}{2}\sum_{j,l=1}^{k}(\widehat{\theta}_j-\theta_j^0)(\widehat{\theta}_l-\theta_l^0)g''_{jl}(x,y;\boldsymbol{\theta}^0,t_0,t_0+t) \\ &\quad + c\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0\|^3 G(x,y;t_0,t_0+t), \end{aligned}$$

where $|c| \leq k^3/6$. Therefore, when $n \to \infty$,

$$
\begin{aligned}
2\left[\max_{\boldsymbol{\theta}\in\boldsymbol{\omega}} \log L_n(\boldsymbol{\theta} \mid \mathbf{x}) - \log L_n(\boldsymbol{\theta}^0 \mid \mathbf{x})\right] &= 2\sum_{j=1}^{k}(\widehat{\theta}_j - \theta_j^0)\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \\
&\quad + \sum_{j,l=1}^{k}(\widehat{\theta}_j - \theta_j^0)(\widehat{\theta}_l - \theta_l^0)\sum_{i=1}^{n-1} g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) \\
&\quad + 2c\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|^3 \sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}).
\end{aligned}
$$

When $n \to \infty$,

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|^3 \sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1}) \\
&= \left\|(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0\right)\right\|^3 \cdot (n-1)^{-1/2} \cdot \left[(n-1)^{-1}\sum_{i=1}^{n-1} G(x_i, x_{i+1}; t_i, t_{i+1})\right] \\
&\xrightarrow{\mathrm{P}} 0.
\end{aligned}
$$

This implies that as $n \to \infty$,

$$
\begin{aligned}
&2\left[\max_{\boldsymbol{\theta}\in\boldsymbol{\omega}} \log L_n(\boldsymbol{\theta} \mid \mathbf{x}) - \log L_n(\boldsymbol{\theta}^0 \mid \mathbf{x})\right] \\
&\quad -2\sum_{j=1}^{k}(n-1)^{1/2}(\widehat{\theta}_j - \theta_j^0)\left[(n-1)^{-1/2}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right] \\
&\quad -\sum_{j,l=1}^{k}(n-1)^{1/2}(\widehat{\theta}_j - \theta_j^0)\cdot(n-1)^{1/2}(\widehat{\theta}_l - \theta_l^0)\left[(n-1)^{-1}\sum_{i=1}^{n-1} g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right] \\
&\xrightarrow{\mathrm{P}} 0.
\end{aligned}
$$

Under all the given assumptions, and using (11.2.15), as $n \to \infty$,

$$
\begin{aligned}
&2\sum_{j=1}^{k}(n-1)^{1/2}(\widehat{\theta}_j - \theta_j^0)\left[(n-1)^{-1/2}\sum_{i=1}^{n-1} g_j'(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right] \\
&\quad +\sum_{j,l=1}^{k}(n-1)^{1/2}(\widehat{\theta}_j - \theta_j^0)\cdot(n-1)^{1/2}(\widehat{\theta}_l - \theta_l^0)\left[(n-1)^{-1}\sum_{i=1}^{n-1} g_{jl}''(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\right] \\
&-\left[(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0\right)\right]^T \boldsymbol{\Sigma}\left[(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE} - \boldsymbol{\theta}^0\right)\right] \\
&\xrightarrow{\mathrm{P}} 0.
\end{aligned}
$$

By Theorem 11.2.5,

$$\left[(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE}-\boldsymbol{\theta}^0\right)\right]^T \boldsymbol{\Sigma}\left[(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{MLE}-\boldsymbol{\theta}^0\right)\right] \xrightarrow{\text{L}} \chi_k^2.$$

Thus,

$$2\left[\max_{\boldsymbol{\theta}\in\boldsymbol{\omega}} \log L_n(\boldsymbol{\theta} \mid \mathbf{x}) - \log L_n(\boldsymbol{\theta}^0 \mid \mathbf{x})\right] \xrightarrow{\text{L}} \chi_k^2,$$

when $n \to \infty$.

All of the techniques of proofs of theorems in this section are credited to Billingsley [1961a].

11.3 Asymptotic properties of conditional least squares estimator

The conditional least squares estimator is obtained by minimizing the sum in (11.1.4). Note that for $i = 1, 2, \ldots$,

$$u_{i+1} = X(t_{i+1}) - \mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}\right] = X(t_{i+1}) - g(x_i; \boldsymbol{\theta}, t_i, t_{i+1})$$

has zero expectation conditioned on $X(t_i) = x_i$. This feature is totally determined by the definition, not the Markov property of the underlying process. The zero conditional expectation feature implies that the partial sums sequences constructed by u_i such as $\left\{\sum_{i=1}^{n-1} u_i; n = 2, 3, \ldots\right\}$ form martingales with respect to $\mathcal{F}_n$, the σ-algebra generated by $\{X(t_1), \ldots, X(t_{n-1})\}$. Furthermore, the relevant expectations in Assumption 11.1.2 are zero, because

$$\begin{aligned}
&\mathbf{E}\left(u(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_i, t_{i+1}) g_j'(X(t_0); \boldsymbol{\theta}^0, t_0, t_0+t)\right) \\
&= \mathbf{E}\left[\mathbf{E}\left(u(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_i, t_{i+1}) g_j'(X(t_0); \boldsymbol{\theta}^0, t_0, t_0+t) \mid X(t_0) = x\right)\right] \\
&= \mathbf{E}\left[0\right] \;=\; 0, \\
&\mathbf{E}\left(u(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_i, t_{i+1}) g_{j_1 j_2}''(X(t_0); \boldsymbol{\theta}^0, t_0, t_0+t)\right) \\
&= \mathbf{E}\left[\mathbf{E}\left(u(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_i, t_{i+1}) g_{j_1 j_2}''(X(t_0); \boldsymbol{\theta}^0, t_0, t_0+t) \mid X(t_0) = x\right)\right] \\
&= 0,
\end{aligned}$$

which imply that the limits on the right hand side of (2) and (4) in Assumption 11.1.2 are zero, namely,

$$(n-1)^{-1}\sum_{i=1}^{n-1} u(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) g'_j(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1}) \quad \xrightarrow{\text{a.s.}} \quad 0, \tag{11.3.1}$$

$$(n-1)^{-1}\sum_{i=1}^{n-1} u(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1}) g''_{j_1 j_2}(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1}) \quad \xrightarrow{\text{a.s.}} \quad 0. \tag{11.3.2}$$

Note that the sum in (11.1.4) also doesn't require Markov property. Hence, the conditional least squares estimating approach can be applicable in processes other than Markov processes.

To pursue the consistency of the CLS estimator, we expand $R_n(\boldsymbol{\theta})$ in (11.1.4) around the true parameter value $\boldsymbol{\theta}^0$, not the partial derivative functions $\frac{\partial}{\partial \boldsymbol{\theta}} R_n(\boldsymbol{\theta})$. This is unlike the previous section where we prove the consistency of the MLE. However, to obtain the asymptotic normality, we shall expand the partial derivative functions $\frac{\partial}{\partial \boldsymbol{\theta}} R_n(\boldsymbol{\theta})$, which is similar to the MLE situation.

From (11.1.4), in a neighborhood $\boldsymbol{\omega}$ of $\boldsymbol{\theta}^0$, we have by the Mean Value Theorem that

$$\begin{aligned}
R_n(\boldsymbol{\theta}) &= R_n(\boldsymbol{\theta}^0) + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \frac{\partial^2 R_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\
&= R_n(\boldsymbol{\theta}^0) + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \frac{\partial^2 R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\
&\qquad + \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \left[\frac{\partial^2 R_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} - \frac{\partial^2 R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right] \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\
&= R_n(\boldsymbol{\theta}^0) + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \mathbf{V}_n \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\
&\qquad + \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \mathbf{W}_n(\boldsymbol{\theta}^*) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}^0),
\end{aligned} \tag{11.3.3}$$

where $\boldsymbol{\theta}^* \in \boldsymbol{\omega}$, and $\mathbf{V}_n$, $\mathbf{W}_n$ are defined in Assumption 11.1.2. We shall control the asymptotic behavior of the first-order and second-order terms. Since for $j, l = 1, 2, \ldots, k$,

$$\begin{aligned}
\frac{\partial u_{i+1}^2}{\partial \theta_j} &= -2u_{i+1} \frac{\partial g(x_i; \boldsymbol{\theta}, t_i, t_{i+1})}{\partial \theta_j}, \\
\frac{\partial^2 u_{i+1}^2}{\partial \theta_j \partial \theta_l} &= 2\frac{\partial g(x_i; \boldsymbol{\theta}, t_i, t_{i+1})}{\partial \theta_j} \frac{\partial g(x_i; \boldsymbol{\theta}, t_i, t_{i+1})}{\partial \theta_l} - 2u_{i+1} \frac{\partial^2 g(x_i; \boldsymbol{\theta}, t_i, t_{i+1})}{\partial \theta_j \partial \theta_l},
\end{aligned}$$

by (11.3.1) and (11.3.2), we obtain that as $n \to \infty$,

$$(n-1)^{-1}\frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} \quad = \quad -2(n-1)^{-1}\sum_{i=1}^{n-1} u_{i+1} \frac{\partial g(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1})}{\partial \boldsymbol{\theta}} \ \xrightarrow{\text{a.s.}} \ \mathbf{0}_{k\times 1},$$

$$
\begin{aligned}
(n-1)^{-1}\frac{\partial^2 R_n(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} &= 2\left((n-1)^{-1}\sum_{i=1}^{n-1} g'_{j_1}(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})g'_{j_2}(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})\right)_{k\times k} \\
&\quad -2\left((n-1)^{-1}\sum_{i=1}^{n-1} u_{i+1}g''_{j_1j_2}(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})\right)_{k\times k} \\
&\xrightarrow{\text{a.s.}} 2\left(\int_0^\infty \mathbf{E}\left(g'_{j_1}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)g'_{j_2}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)dF_T(t)\right)_{k\times k}.
\end{aligned}
$$

Because the matrix $\left(\mathbf{E}\left(g'_{j_1}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)g'_{j_2}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)\right)_{k\times k}$ is the covariance matrix of random vector $\frac{\partial}{\partial\boldsymbol{\theta}}g(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)$, it is non-negative definite. This yields that the matrix

$$
\mathbf{V} = \left(\int_0^\infty \mathbf{E}\left(g'_{j_1}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)g'_{j_2}(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)dF_T(t)\right)_{k\times k} \tag{11.3.4}
$$

is also non-negative definite. By (1) in Assumption 11.1.2, we know that $(\boldsymbol{\theta}-\boldsymbol{\theta}^0)^T\cdot\mathbf{W}_n(\boldsymbol{\theta}^*)\cdot(\boldsymbol{\theta}-\boldsymbol{\theta}^0)$ is dominated by δ^3 if $||\boldsymbol{\theta}^*-\boldsymbol{\theta}^0||<\delta$, $||\boldsymbol{\theta}-\boldsymbol{\theta}^0||<\delta$ as $n\to\infty$. Choosing δ small enough, we shall see that the right hand side of (11.3.3) multiplying through $(n-1)^{-1}$ will be dominated by $(n-1)^{-1}R_n(\boldsymbol{\theta}^0)+\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^0)^T\cdot\mathbf{V}\cdot(\boldsymbol{\theta}-\boldsymbol{\theta}^0)$.

To guarantee that $(\boldsymbol{\theta}-\boldsymbol{\theta}^0)^T\cdot\mathbf{V}\cdot(\boldsymbol{\theta}-\boldsymbol{\theta}^0)$ is positive, we should impose the condition that $\mathbf{V}$ is non-singular. To satisfy (1) in Assumption 11.1.2, we can require that $g(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)$ has partial derivatives up to the third order, which also satisfy certain conditions. All these analyses lead to the following regularity conditions for the consistency of the CLS estimator.

Assumption 11.3.1

(1) $\frac{\partial}{\partial\theta_j}g(X(t_0);\boldsymbol{\theta},t_0,t_0+t)$, $\frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}}g(X(t_0);\boldsymbol{\theta},t_0,t_0+t)$, *and* $\frac{\partial^3}{\partial\theta_{j_1}\theta_{j_2}\theta_{j_3}}g(X(t_0);\boldsymbol{\theta},t_0,t_0+t)$ *exist and are continuous in* Θ *for* $j,j_1,j_2,j_3=1,2,\ldots,k$;

(2) *For* $j,j_1,j_2=1,2,\ldots,k$, *and* $t_0,t\ge 0$,

$$
\begin{aligned}
\mathbf{E}\left[\left(X(t_0+t)-g(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)\cdot\left(\frac{\partial}{\partial\theta_j}g(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)\right] &< \infty, \\
\mathbf{E}\left[\left(X(t_0+t)-g(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)\cdot\left(\frac{\partial^2}{\partial\theta_{j_1}\partial\theta_{j_2}}g(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)\right] &< \infty, \\
\mathbf{E}\left[\left(\frac{\partial}{\partial\theta_{j_1}}g(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)\cdot\left(\frac{\partial}{\partial\theta_{j_2}}g(X(t_0);\boldsymbol{\theta}^0,t_0,t_0+t)\right)\right] &< \infty;
\end{aligned}
$$

(3) *For $j, j_1, j_2, j_3 = 1, 2, \ldots, k$, and $t_0, t \geq 0$, there exist functions*

$$H^{(0)}(X(t_0); t_0, t_0 + t), \qquad H_j^{(1)}(X(t_0); t_0, t_0 + t),$$
$$H_{j_1 j_2}^{(2)}(X(t_0); t_0, t_0 + t), \qquad H_{j_1 j_2 j_3}^{(3)}(X(t_0); t_0, t_0 + t)$$

such that

$$\begin{aligned}
|g(X(t_0); \boldsymbol{\theta}, t_0, t_0 + t)| &\leq H^{(0)}(X(t_0); t_0, t_0 + t), \\
\left| \frac{\partial}{\partial \theta_j} g(X(t_0); \boldsymbol{\theta}, t_0, t_0 + t) \right| &\leq H_j^{(1)}(X(t_0); t_0, t_0 + t), \\
\left| \frac{\partial^2}{\partial \theta_{j_1} \partial \theta_{j_2}} g(X(t_0); \boldsymbol{\theta}, t_0, t_0 + t) \right| &\leq H_{j_1 j_2}^{(2)}(X(t_0); t_0, t_0 + t), \\
\left| \frac{\partial^3}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3}} g(X(t_0); \boldsymbol{\theta}, t_0, t_0 + t) \right| &\leq H_{j_1 j_2 j_3}^{(3)}(X(t_0); t_0, t_0 + t)
\end{aligned}$$

for all $\boldsymbol{\theta} \in \Theta$, and

$$\begin{aligned}
\mathbf{E}\left[\left| X(t_0 + t) H_{j_1 j_2 j_3}^{(3)}(X(t_0); t_0, t_0 + t) \right| \right] &< \infty, \\
\mathbf{E}\left[\left| H^{(0)}(X(t_0); t_0, t_0 + t) \cdot H_{j_1 j_2 j_3}^{(3)}(X(t_0); t_0, t_0 + t) \right| \right] &< \infty, \\
\mathbf{E}\left[\left| H_j^{(1)}(X(t_0); t_0, t_0 + t) \cdot H_{j_1 j_2}^{(2)}(X(t_0); t_0, t_0 + t) \right| \right] &< \infty.
\end{aligned}$$

(4) *The limiting matrix (11.3.4) is non-singular.*

The proof of the strong consistency of the CLS estimator requires Egoroff's theorem, which deals with almost uniform convergence:

Theorem 11.3.1 (Egoroff's Theorem) *Suppose h and $\{h_n\}$ are measurable complex-valued functions on measure space $(\Omega, \mathcal{F}, \mu)$ with $\mu(\Omega) < \infty$ such that $h_n \xrightarrow{\text{a.s.}} h$. Then for every $\epsilon > 0$, there exists $E \subset \Omega$ such that $\mu(E) < \epsilon$ and $h_n \longrightarrow h$ uniformly on E^c.*

For reference, see Folland [1984], p. 60.

Denote $\widehat{\boldsymbol{\theta}}_{CLS}(n)$ as the CLS estimator when the sample size is n. Then, we have the following theorem on the strong consistency of this estimator.

Theorem 11.3.2 *Under Assumptions 11.1.2 and 11.3.1,*

$$\widehat{\boldsymbol{\theta}}_{CLS}(n) \xrightarrow{\text{a.s.}} \boldsymbol{\theta}^0, \qquad \text{as } n \to \infty.$$

Proof: Let $\delta > 0$, and $\boldsymbol{\omega}(\delta)$ be the neighborhood of $\boldsymbol{\theta}^0$ with radius δ.

Regularity conditions (1), (2) and (3) in Assumption 11.3.1 lead to condition (1) in Assumption 11.1.2 being satisfied. Then for any $\epsilon > 0$, under all the conditions in Assumption 11.1.2 and (4) in Assumption 11.3.1, we can find by Egoroff's theorem an event E with $\Pr(E) > 1 - \epsilon$, a constant δ^* $(0 < \delta^* < \delta)$, $M > 0$ and $n_0 > 0$ such that on E, for any $n > n_0$, $\boldsymbol{\theta} \in \boldsymbol{\omega}(\delta^*)$, the following three conditions hold:

(1) $\quad \left|(\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}}\right| < (n-1)\delta^3, \quad$ (refer to (11.3.1).)

(2) $\quad \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \cdot \mathbf{W}_n(\boldsymbol{\theta}^*) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}^0) < (n-1)M\delta^3, \quad$ (refer to (1) in Assumption 11.1.2.)

(3) $\quad$ the minimum eigenvalue of $\frac{1}{2(n-1)}\mathbf{V}_n$ is greater than some $\lambda_0 > 0$.

Thus, by (11.3.3), for $\boldsymbol{\theta}$ on the boundary of $\boldsymbol{\omega}(\boldsymbol{\theta}^*)$,

$$R_n(\boldsymbol{\theta}) \;\geq\; R_n(\boldsymbol{\theta}^0) + (n-1)(-\delta^3 + \delta^2\lambda_0 - M\delta^3) \;=\; R_n(\boldsymbol{\theta}^0) + (n-1)\delta^2[\lambda_0 - (M+1)\delta].$$

Since δ can be chosen small enough such that $\lambda_0 - (M+1)\delta > 0$, $R_n(\boldsymbol{\theta})$ must attain a minimum at some $\widehat{\boldsymbol{\theta}}_{CLS}(n) \in \boldsymbol{\omega}(\delta^*)$.

Let $\epsilon_l = 2^{-l}$ and $\delta_l = 1/l$, where $l = 1, 2, \ldots$. Then they will determine a sequence of events $\{E_l\}$ and an increasing sequence $\{n_l\}$ having the above properties. For $n_l < n \leq n_{l+1}$, define $\widehat{\boldsymbol{\theta}}_{CLS}(n)$ on E_l to be the point within $\boldsymbol{\omega}(\delta_l)$ where $R_n(\boldsymbol{\theta})$ attains a relative minimum, and define $\widehat{\boldsymbol{\theta}}_{CLS}(n)$ to be zero on E_l^c. This will yield that $\widehat{\boldsymbol{\theta}}_{CLS}(n) \longrightarrow \boldsymbol{\theta}^0$ on $\liminf E_l = \bigcup_{m=1}^{\infty} \bigcap_{l=m}^{\infty} E_l$. Furthermore, since for any $m > 1$, it holds that

$$\Pr\left[\limsup E_l^c\right] \;=\; \Pr\left[\bigcap_{m=1}^{\infty} \bigcup_{l=m}^{\infty} E_l^c\right] \;\leq\; \Pr\left[\bigcup_{l=m}^{\infty} E_l^c\right] \;\leq\; \sum_{l=m}^{\infty} \Pr\left[E_l^c\right] \;\leq\; \sum_{l=m}^{\infty} 2^{-l} \;=\; 2^{m-1},$$

which implies that $\Pr\left[\limsup E_l^c\right] = 0$. Therefore, we have $\Pr[\liminf E_l] = 1 - \Pr\left[\limsup E_l^c\right] = 1$. This completes the proof.

To obtain the asymptotic normality, we need to expand the partial derivative function $\frac{\partial}{\partial \boldsymbol{\theta}} R_n(\boldsymbol{\theta})$ around $\boldsymbol{\theta}^0$:

$$\frac{\partial}{\partial \boldsymbol{\theta}} R_n(\boldsymbol{\theta}) \;=\; \frac{\partial}{\partial \boldsymbol{\theta}} R_n(\boldsymbol{\theta}^0) + (\mathbf{V}_n + \mathbf{W}_n(\boldsymbol{\theta}^*)) \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}^0), \qquad ||\boldsymbol{\theta}^* - \boldsymbol{\theta}^0|| \leq ||\boldsymbol{\theta} - \boldsymbol{\theta}^0|| \leq \delta, \; \delta > 0.$$

Multiplying through $(n-1)^{-1/2}$ and evaluating at the point $\widehat{\boldsymbol{\theta}}_{CLS}$, we have

$$\begin{aligned}
\mathbf{0}_{k\times 1} &= (n-1)^{-1/2}\frac{\partial}{\partial\boldsymbol{\theta}}R_n(\widehat{\boldsymbol{\theta}}_{CLS}) \\
&= (n-1)^{-1/2}\frac{\partial}{\partial\boldsymbol{\theta}}R_n(\boldsymbol{\theta}^0) + (n-1)^{-1}\left(\mathbf{V}_n + \mathbf{W}_n(\boldsymbol{\theta}^*)\right)\cdot(n-1)^{1/2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0) \\
&= -2(n-1)^{-1/2}\sum_{i=1}^{n-1} u(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\frac{\partial g(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1})}{\partial\boldsymbol{\theta}} \\
&\quad + (n-1)^{-1}\left(\mathbf{V}_n + \mathbf{W}_n(\boldsymbol{\theta}^*)\right)\cdot(n-1)^{1/2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0). \qquad (11.3.5)
\end{aligned}$$

For each $j = 1, 2, \ldots, k$, $\left\{\sum_{i=1}^{n-1} u(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\frac{\partial g(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})}{\partial\theta_j}\right\}$ is a martingale with respect to $\mathcal{F}_n$. Under appropriate regularity conditions, $(n-1)^{-1/2}\sum_{i=1}^{n-1} u(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\frac{\partial g(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})}{\partial\boldsymbol{\theta}}$ converges in law to a multivariate normal distribution as $n \to \infty$. Note that under previous regularity conditions for strong consistency, $(n-1)^{-1}\left(\mathbf{V}_n + \mathbf{W}_n(\boldsymbol{\theta}^*)\right)$ goes to $2\mathbf{V}_{k\times k}$. Hence, $(n-1)^{1/2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0)$ will converge in law to a multivariate normal distribution as $n \to \infty$.

We tailor the following additional regularity conditions for the asymptotic normality of the CLS estimator.

Assumption 11.3.2 *$u(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)\frac{\partial}{\partial\theta_j}g(X(t_0); \boldsymbol{\theta}^0, t_0, t_0+t)$ $(j = 1, 2, \ldots, k)$ has moment of order $2+d$ $(d > 0)$ for any $t_0, t \geq 0$. Also for $j = 1, 2, \ldots, k$,*

$$\lim_{n\to\infty}(n-1)^{-1-d/2}\sum_{i=1}^{n-1}\mathbf{E}\left(\left[u(x_i, X(t_{i+1}); \boldsymbol{\theta}^0, t_i, t_{i+1})\frac{\partial}{\partial\theta_j}g(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1})\right]^{2+d}\right) = 0.$$

In addition, for $j_1, j_2 = 1, 2, \ldots, k$,

$$(n-1)^{-1}\sum_{i=1}^{n-1} u^2(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\cdot g'_{j_1}(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1})\cdot g'_{j_2}(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1}) \xrightarrow{\mathrm{P}} \sigma_{j_1 j_2} =$$

$$\int_0^\infty \mathbf{E}\left[u^2(X(t_0), X(t_0+t); \boldsymbol{\theta}^0, t_0, t_0+t)g'_{j_1}(X(t_0); \boldsymbol{\theta}^0, t_0, t_0+t)g'_{j_2}(X(t_0); \boldsymbol{\theta}^0, t_0, t_0+t)\right]dF_T(t).$$

namely,

$$\left((n-1)^{-1}\sum_{i=1}^{n-1} u^2(x_i, x_{i+1}; \boldsymbol{\theta}^0, t_i, t_{i+1})\cdot g'_{j_1}(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1})\cdot g'_{j_2}(x_i; \boldsymbol{\theta}^0, t_i, t_{i+1})\right) \xrightarrow{\mathrm{P}} \left(\sigma_{j_1 j_2}\right)_{k\times k} = \boldsymbol{\Sigma},$$

as $n \to \infty$.

Theorem 11.3.3 *Assume $\widehat{\boldsymbol{\theta}}_{CLS}$ is a root of $\frac{\partial R_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$. Under Assumptions 11.1.2, 11.3.1 and 11.3.2,*

$$(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0\right) \xrightarrow{\text{L}} \text{N}\left(\mathbf{0}_{k\times 1},\ \mathbf{V}_{k\times k}^{-1}\boldsymbol{\Sigma}_{k\times k}\mathbf{V}_{k\times k}^{-1}\right),$$

where $\boldsymbol{\Sigma}_{k\times k}$ is defined in Assumption 11.3.2 and $\mathbf{V}_{k\times k}$ is defined in (11.3.4).

Proof: We continue the analysis from (11.3.5). Under Assumption 11.3.1, there is a $\boldsymbol{\theta}^*$ such that

$$\mathbf{0}_{k\times 1} = (n-1)^{-1/2}\frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + (n-1)^{-1}\left(\mathbf{V}_n + \mathbf{W}_n(\boldsymbol{\theta}^*)\right)\cdot (n-1)^{1/2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0),$$

where $\mathbf{W}_n(\boldsymbol{\theta}^*)$ can be written as

$$\mathbf{W}_n(\boldsymbol{\theta}^*) = \left(2^{-1}\sum_{j_3=1}^{k}\left(\widehat{\theta}_{CLS}(n,j_3) - \theta_{j_3}\right)\frac{\partial^3 R_n(\boldsymbol{\theta}^*)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}}\right)_{k\times k}.$$

Here $\widehat{\boldsymbol{\theta}}_{CLS} = \widehat{\boldsymbol{\theta}}_{CLS}(n) = \left(\widehat{\theta}_{CLS}(n,1),\ldots,\widehat{\theta}_{CLS}(n,k)\right)^T$. Since $\widehat{\boldsymbol{\theta}}_{CLS}$ is strongly consistent for $\boldsymbol{\theta}$, we can conclude that

$$(n-1)^{-1}\mathbf{W}_n(\boldsymbol{\theta}^*) \xrightarrow{\text{a.s.}} \mathbf{0}_{k\times k}, \qquad n\to\infty.$$

Thus,

$$(n-1)^{-1}\left(\mathbf{V}_n + \mathbf{W}_n(\boldsymbol{\theta}^*)\right) \xrightarrow{\text{a.s.}} 2\mathbf{V}, \qquad n\to\infty.$$

This implies that $(n-1)^{1/2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0)$ will have the same limiting distribution as

$$\begin{aligned}&(2\mathbf{V})^{-1}\left[2(n-1)^{-1/2}\sum_{i=1}^{n-1}u(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1})\frac{\partial g(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})}{\partial\boldsymbol{\theta}}\right]\\ &= \mathbf{V}^{-1}\left[(n-1)^{-1/2}\sum_{i=1}^{n-1}u(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1})\frac{\partial g(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})}{\partial\boldsymbol{\theta}}\right].\end{aligned}$$

By Theorem 11.1.1, it follows that

$$(n-1)^{-1/2}\sum_{i=1}^{n-1}u(x_i,x_{i+1};\boldsymbol{\theta}^0,t_i,t_{i+1})\frac{\partial g(x_i;\boldsymbol{\theta}^0,t_i,t_{i+1})}{\partial\boldsymbol{\theta}} \xrightarrow{\text{L}} \text{N}\left(\mathbf{0}_{k\times 1},\boldsymbol{\Sigma}_{k\times k}\right), \qquad n\to\infty.$$

Therefore,

$$(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0\right) \xrightarrow{\text{L}} \text{N}\left(\mathbf{0}_{k\times 1},\ \mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{V}^{-1}\right).$$

Remark: Usually, it is intractable to find the analytic forms of the matrices $\mathbf{V}$ and $\boldsymbol{\Sigma}$. However, they can be estimated from the data. The natural estimators for them are

$$\widehat{\boldsymbol{\Sigma}} = \left((n-1)^{-1} \sum_{i=1}^{n-1} u^2(x_i, x_{i+1}; \widehat{\boldsymbol{\theta}}_{CLS}, t_i, t_{i+1}) \cdot g'_{j_1}(x_i; \widehat{\boldsymbol{\theta}}_{CLS}, t_i, t_{i+1}) \cdot g'_{j_2}(x_i; \widehat{\boldsymbol{\theta}}_{CLS}, t_i, t_{i+1}) \right)_{k\times k}, \tag{11.3.6}$$

$$\widehat{\mathbf{V}} = \left((n-1)^{-1} \sum_{i=1}^{n-1} g'_{j_1}(x_i; \widehat{\boldsymbol{\theta}}_{CLS}, t_i, t_{i+1}) g'_{j_2}(x_i; \widehat{\boldsymbol{\theta}}_{CLS}, t_i, t_{i+1}) \right)_{k\times k}. \tag{11.3.7}$$

They are useful in construction of asymptotic confidence intervals or regions and hypothesis testing.

The following theorem is the analogue of Theorem 11.2.6.

Theorem 11.3.4 *Let $Y_1, Y_2, \ldots, Y_k$ be iid rv's χ_1^2. Under Assumptions 11.1.2, 11.3.1 and 11.3.2,*

$$R_n(\boldsymbol{\theta}^0) - R_n\left(\widehat{\boldsymbol{\theta}}_{CLS}\right) \xrightarrow{\mathrm{L}} \sum_{j=1}^{k} \lambda_j Y_j, \qquad n \to \infty,$$

where λ_j $(j = 1, 2, \ldots, k)$ are the (non-negative) eigenvalues of $\mathbf{V}^{-1}\boldsymbol{\Sigma}$.

Proof: From (11.3.3), we have

$$\begin{aligned} R_n(\boldsymbol{\theta}^0) - R_n\left(\widehat{\boldsymbol{\theta}}_{CLS}\right) &= -(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0)^T \cdot \frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} - \tfrac{1}{2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0)^T \cdot \mathbf{V}_n \cdot (\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0) \\ &\quad -\tfrac{1}{2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0)^T \cdot \mathbf{W}_n(\boldsymbol{\theta}^*) \cdot (\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0). \end{aligned}$$

Furthermore, from (11.3.5),

$$-\frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} = (\mathbf{V}_n + \mathbf{W}_n(\boldsymbol{\theta}^*)) \cdot (\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0).$$

Thus we obtain

$$\begin{aligned} R_n(\boldsymbol{\theta}^0) - R_n\left(\widehat{\boldsymbol{\theta}}_{CLS}\right) &= \tfrac{1}{2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0)^T \cdot (\mathbf{V}_n + \mathbf{W}_n(\boldsymbol{\theta}^*)) \cdot (\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0) \\ = \left[(n-1)^{1/2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0)^T\right] &\cdot \tfrac{1}{2}\left[\mathbf{V}_n/(n-1) + \mathbf{W}_n(\boldsymbol{\theta}^*)/(n-1)\right] \cdot \left[(n-1)^{1/2}(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0)\right]. \end{aligned}$$

According to Theorem 11.3.3, as $n \to \infty$,

$$\tfrac{1}{2}\left[\mathbf{V}_n/(n-1) + \mathbf{W}_n(\boldsymbol{\theta}^*)/(n-1)\right] \xrightarrow{\text{a.s.}} \mathbf{V}, \qquad (n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}^0\right) \xrightarrow{\mathrm{L}} \mathbf{Z},$$

where $\mathbf{Z}_{k\times 1} \sim \mathrm{N}\left(\mathbf{0}_{k\times 1},\ \mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{V}^{-1}\right)$. This yields that the limiting distribution of $R_n(\boldsymbol{\theta}^0) - R_n\left(\widehat{\boldsymbol{\theta}}_{CLS}\right)$ is the same as the distribution of $\mathbf{Z}^T\mathbf{V}\mathbf{Z}$, which has mgf

$$M_{\mathbf{Z}^T\mathbf{V}\mathbf{Z}}(s) \;=\; \mathbf{E}\left[e^{s\cdot\mathbf{Z}^T\mathbf{V}\mathbf{Z}}\right] \;=\; \left|1 - 2s\mathbf{V}^{-1}\boldsymbol{\Sigma}\right|^{-1/2} \;=\; \left[\prod_{j=1}^{k}(1-2s\lambda_j)\right]^{-1/2}.$$

It is also the mgf of $\sum_{j=1}^{k}\lambda_j Y_j$. Therefore, this completes the proof.

The techniques of proof of asymptotic properties of the CLS estimator in this section are from Klimko and Nelson [1978]. However, we deal with these asymptotic properties under different regularity conditions.

Finally, we comment on the asymptotic properties of other estimators of variations of the CLS estimating approach. Because of similar techniques to previous results, we will not discuss in detail.

Similar to the CLS estimator, the CWLS2 and CGLS estimator have consistency and asymptotic normality under appropriate regularity conditions like those for the CLS estimator, because the expansion of

$$R_{CWLS2}(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{n-1}\frac{\Big(x_{i+1} - \mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}\right]\Big)^2}{cx_i + d}$$

and

$$R_{CGLS}(\boldsymbol{\theta}) \;=\; \sum_{i=1}^{n-1}\Big(g(x_{i+1}) - \mathbf{E}\left[g(X(t_{i+1})) \mid X(t_i) = x_i; \boldsymbol{\theta}\right]\Big)^2$$

are similar to the expansion of $R_{CLS}(\boldsymbol{\theta})$. The expansion of a function $R_n(\boldsymbol{\theta})$ around $\boldsymbol{\theta}^0$ usually has the following form

$$R_n(\boldsymbol{\theta}) \;\approx\; R_n(\boldsymbol{\theta}^0) + (\boldsymbol{\theta}-\boldsymbol{\theta}^0)^T\cdot\frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}} + \tfrac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^0)^T\cdot\frac{\partial^2 R_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\cdot(\boldsymbol{\theta}-\boldsymbol{\theta}^0).$$

The essential requirements for the strong consistency of the CLS estimator are that

$$(n-1)^{-1}\frac{\partial R_n(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}} \xrightarrow{\text{a.s.}} \mathbf{0}_{k\times 1}$$

and $(n-1)^{-1}\frac{\partial^2 R_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}$ converges a.s. to a non-negative definite matrix as $n \to \infty$. Now we check the first-order and second-order partial derivatives of $R_{CWLS2}(\boldsymbol{\theta})$ and $R_{CGLS}(\boldsymbol{\theta})$. They are

$$\begin{aligned}
\frac{\partial R_{CWLS2}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= -2\sum_{i=1}^{n-1}\frac{\left(x_{i+1}-\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]\right)}{cx_i+d}\cdot\frac{\partial\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\boldsymbol{\theta}},\\
\frac{\partial^2 R_{CWLS2}(\boldsymbol{\theta})}{\partial\theta_j\partial\boldsymbol{\theta}_l} &= 2\sum_{i=1}^{n-1}\frac{1}{cx_i+d}\cdot\frac{\partial\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_j}\cdot\frac{\partial\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_l}\\
&\quad -2\sum_{i=1}^{n-1}\frac{\left(x_{i+1}-\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]\right)}{cx_i+d}\cdot\frac{\partial^2\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_j\partial\theta_l},
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial R_{CGLS}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\\
&= -2\sum_{i=1}^{n-1}\left(g(x_{i+1})-\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]\right)\cdot\frac{\partial\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\boldsymbol{\theta}},\\
&\frac{\partial^2 R_{CGLS}(\boldsymbol{\theta})}{\partial\theta_j\partial\boldsymbol{\theta}_l}\\
&= 2\sum_{i=1}^{n-1}\frac{\partial\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_j}\cdot\frac{\partial\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_l}\\
&\quad -2\sum_{i=1}^{n-1}\left(g(x_{i+1})-\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]\right)\cdot\frac{\partial^2\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_j\partial\theta_l},
\end{aligned}$$

where $j,l=1,2,\ldots,k$. Since the sequences of the partial sums

$$\left\{\sum_{i=1}^{n-1}\frac{\left(x_{i+1}-\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]\right)}{cx_i+d}\cdot\frac{\partial\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_j}, n=2,3,\ldots\right\},$$

$$\left\{\sum_{i=1}^{n-1}\frac{\left(x_{i+1}-\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]\right)}{cx_i+d}\cdot\frac{\partial^2\mathbf{E}[X(t_{i+1}) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_j\partial\theta_l}, n=2,3,\ldots\right\},$$

$$\left\{\sum_{i=1}^{n-1}\left(g(x_{i+1})-\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]\right)\cdot\frac{\partial\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_j}, n=2,3,\ldots\right\},$$

$$\left\{\sum_{i=1}^{n-1}\left(g(x_{i+1})-\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]\right)\cdot\frac{\partial^2\mathbf{E}[g(X(t_{i+1})) \mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\theta_j\partial\theta_l}, n=2,3,\ldots\right\}$$

are martingales with respect to $\mathcal{F}_n$ for $j,\ l=1,2,\ldots,k$, by ergodicity, their averages will converge to zero. Applying the same techniques of proof for the CLS estimator, under appropriate regularity conditions, we can derive the consistency and asymptotic normality for the CWLS2 and CGLS estimators. Like the CLS estimator, their asymptotic normal distributions are of the form $\mathrm{N}\left(\mathbf{0}_{k\times 1},\ \mathbf{V}_{k\times k}^{-1}\boldsymbol{\Sigma}_{k\times k}\mathbf{V}_{k\times k}^{-1}\right)$. For CWLS2, the estimated asymptotic matrices are

$$\widehat{\boldsymbol{\Sigma}} = \left((n-1)^{-1}\sum_{i=1}^{n-1} u^2(x_i, x_{i+1}; \widehat{\boldsymbol{\theta}}_{CWLS2}, t_i, t_{i+1}) \cdot \tilde{g}'_{j_1}(x_i; \widehat{\boldsymbol{\theta}}_{CWLS2}, t_i, t_{i+1}) \cdot \tilde{g}'_{j_2}(x_i; \widehat{\boldsymbol{\theta}}_{CWLS2}, t_i, t_{i+1})\right), \tag{11.3.8}$$

$$\widehat{\mathbf{V}} = \left((n-1)^{-1}\sum_{i=1}^{n-1} \frac{1}{cx_i+d} \cdot \tilde{g}'_{j_1}(x_i; \widehat{\boldsymbol{\theta}}_{CWLS2}, t_i, t_{i+1})\tilde{g}'_{j_2}(x_i; \widehat{\boldsymbol{\theta}}_{CWLS2}, t_i, t_{i+1})\right), \tag{11.3.9}$$

where

$$\begin{aligned} u(x_i, x_{i+1}; \widehat{\boldsymbol{\theta}}_{CWLS2}, t_i, t_{i+1}) &= \frac{\left(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\right)}{cx_i + d}, \\ \tilde{g}'_j(x_i; \widehat{\boldsymbol{\theta}}_{CWLS2}, t_i, t_{i+1}) &= \frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_j}, \qquad i = 1, 2, \ldots, n-1,\ j = 1, \ldots, k; \end{aligned}$$

and for CGLS, they are

$$\widehat{\boldsymbol{\Sigma}} = \left((n-1)^{-1}\sum_{i=1}^{n-1} u^2(x_i, x_{i+1}; \widehat{\boldsymbol{\theta}}_{CGLS}, t_i, t_{i+1}) \cdot \tilde{g}'_{j_1}(x_i; \widehat{\boldsymbol{\theta}}_{CGLS}, t_i, t_{i+1}) \cdot \tilde{g}'_{j_2}(x_i; \widehat{\boldsymbol{\theta}}_{CGLS}, t_i, t_{i+1})\right), \tag{11.3.10}$$

$$\widehat{\mathbf{V}} = \left((n-1)^{-1}\sum_{i=1}^{n-1} \tilde{g}'_{j_1}(x_i; \widehat{\boldsymbol{\theta}}_{CGLS}, t_i, t_{i+1})\tilde{g}'_{j_2}(x_i; \widehat{\boldsymbol{\theta}}_{CGLS}, t_i, t_{i+1})\right), \tag{11.3.11}$$

where

$$\begin{aligned} u(x_i, x_{i+1}; \widehat{\boldsymbol{\theta}}_{CGLS}, t_i, t_{i+1}) &= g(x_{i+1}) - \mathbf{E}[g(X(t_{i+1})) \mid X(t_i) = x_i; \boldsymbol{\theta}], \\ \tilde{g}'_j(x_i; \widehat{\boldsymbol{\theta}}_{CGLS}, t_i, t_{i+1}) &= \frac{\partial \mathbf{E}[g(X(t_{i+1})) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \theta_j}, \quad i = 1, 2, \ldots, n-1,\ j = 1, \ldots, k. \end{aligned}$$

However, the CWLS estimator is no longer consistent for $\boldsymbol{\theta}^0$, because in the expansion of $R_{CWLS}(\boldsymbol{\theta})$, the first-order partial derivative is

$$\begin{aligned} \frac{\partial R_{CWLS}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -2\sum_{i=1}^{n-1} \frac{\left(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\right)}{\mathbf{Var}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]} \cdot \frac{\partial \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}} \\ &\quad -2\sum_{i=1}^{n-1} \frac{\left(x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]\right)^2}{\mathbf{Var}^2[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]} \cdot \frac{\partial \mathbf{Var}[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}]}{\partial \boldsymbol{\theta}}, \end{aligned}$$

under ergodicity,

$$(n-1)^{-1}\sum_{i=1}^{n-1}\frac{\Big(x_{i+1}-\mathbf{E}[X(t_{i+1})\mid X(t_i)=x_i;\boldsymbol{\theta}]\Big)}{\mathbf{Var}[X(t_{i+1})\mid X(t_i)=x_i;\boldsymbol{\theta}]}\cdot\frac{\partial\mathbf{E}[X(t_{i+1})\mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\boldsymbol{\theta}}\xrightarrow{\text{a.s.}}\mathbf{0}_{k\times 1},$$

but

$$(n-1)^{-1}\sum_{i=1}^{n-1}\frac{\Big(x_{i+1}-\mathbf{E}[X(t_{i+1})\mid X(t_i)=x_i;\boldsymbol{\theta}]\Big)^2}{\mathbf{Var}^2[X(t_{i+1})\mid X(t_i)=x_i;\boldsymbol{\theta}]}\cdot\frac{\partial\mathbf{Var}[X(t_{i+1})\mid X(t_i)=x_i;\boldsymbol{\theta}]}{\partial\boldsymbol{\theta}}$$

doesn't. Hence, $(n-1)^{-1}\frac{\partial R_{CWLS}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ doesn't converge to a zero vector.

This pitfall can be overcome with the quasi-conditional least squares estimator, because it omits the second term with the partial derivative of the conditional variance. Following the techniques of proof for the MLE estimator, under appropriate regularity conditions, we can show that the QCLS estimator is consistent and asymptotically normal. Here the asymptotic normality comes from Theorem 11.1.1.

Chapter 12

Autocorrelation detection, model selection, testing, diagnosis, forecasting and process simulation

In this chapter, we shall consider some practical problems with non-normal time series data. When we apply a specific model to data, a natural question is why should we use this model, not the others? Such a question motivates people to scrutinize the data in all possible aspects. For example, when we try to apply an auto-regressive model to a stationary time series, we should first check the auto-correlation function plot to see if there exists any correlation pattern over time. Similarly, for count data or positive data observations over time, when we consider fitting a continuous-time GAR(1) process, we should investigate if there is any auto-correlation in the data. If the dependence over time appears to be geometrically decreasing, then applying the continuous-time GAR(1) process model may be appropriate.

Usually, for a real problem, there could be several possible models which can be applied to the problem. Since the family of the continuous-time GAR(1) processes is so abundant, we have to think about which models are most suitable. This raises the model selection issue.

Once we select a model, we can estimate the parameters in the model using the data. Later, we can use this model for inferences such as hypothesis testing, forecasting, etc. These steps essentially cover the entire procedure for an applied statistical problem.

In Section 12.1, we will discuss visual and analytical detection of serial dependence. Section 12.2 deals with the model selection issue. We study diagnosis and hypothesis testing in Section 12.3, and forecasting in Section 12.4. Finally, we discuss the approaches of process simulation in Section 12.5.

12.1 Assessing autocorrelation

When modelling, a necessary step is to examine the features of the data so that proper models can be chosen. These examinations include graphical and analytical investigations. For the data sampled from a dynamic system, or from a subject over time, an interesting question is whether there exists any dependence structure over time, commonly called serial dependence or serial association. Because we are focusing on stationary time series, stationarity should be checked before modelling. This can be done with a time series plot to check for a trend or pattern, or an ACF plot to check for periodicity or seasonality, or other advanced techniques like smoothing to check for potentially non-stationary patterns. If the series can be considered stationary with geometric serial dependence, then it is reasonable to model the data with a continuous-time GAR(1) process. In the following two subsections, we will discuss some visual detection techniques for serial dependence.

Serial dependence, if it exists, is essentially hidden in pairs of data points. For a sample with size n, we can obtain $\binom{n}{2} = n(n-1)/2$ pairs. We can group them according to their lag lengths or time differences as: Group 1, Group 2, $\ldots$, Group m. Each group consists of pairs with equal or roughly equal lag lengths.

For equally-spaced time series, grouping these pairs is very easy. The lag lengths are very regular: $1, 2, 3, \ldots$. Hence, we can obtain $n-1$ pairs with lag one, $n-2$ pairs with lag 2, and so on. But for the unequally-spaced time series, the lag lengths and numbers of pairs may not be as regular as the equally-spaced case.

For the sake of simplicity, we concentrate on the data sampled at the equally-spaced time points. This is partially because that in many studies, the data $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ are scheduled to be observed at equally-spaced time points. The results from equally-spaced time series can then be easily generalized to unequally-spaced time series.

Traditionally, for the stationary Gaussian time series $\{X_0, X_1, X_2, \ldots\}$, the scatterplot of successive lag-j pairs such as (X_i, X_{i+j}) $(i = 1, 2, \ldots)$ is very helpful in recognizing any potential linear association patterns visually.

In the context of the continuous-time GAR(1) processes, if the self-generalized rv K in the extended-thinning operation is from **P1** so that the observations are real-valued, the traditional scatterplot can still display the linear pattern among successive pairs $(X(t_i), X(t_{i+1}))$, where $i = 1, 2, \ldots, n-1$. Hence, it works well in this case and should be kept as a basic graphical tool. However, if the observations are non-negative integer-valued or positive-valued, the marginal distribution is no longer symmetric. Instead, it is most likely to be skewed. This will cause the association pattern not to be a linear pattern with an ellipsoidal cloud of points. Besides, for the count data, there may be many coincidences in the scatterplot because of the discreteness. Thus, some other graphical tools should be introduced in such kinds of situations.

In this subsection, we mainly study three kinds of graphical methods: the sunflower plot, the diagonal P-P plot and the randomized quantile transformation scatterplot. We illustrate them by considering the lag-1 pairs in an equally-spaced setting. These tools, of course, will be applied in other groups of pairs to check for the serial dependence of the equally-spaced or unequally-spaced time series.

(1) Sunflower plot.

First, we turn to the sunflower plot introduced by Cleveland and McGill [1984]. This tool was later improved by Schilling and Watkins [1994] to overcome some disadvantages. The sunflower plot is in fact a type of two-dimensional histogram or contour plot without equal altitude curves. It is designed to display bivariate data with coincident points. This is the typical phenomenon when the data are discrete. Even for the positive-valued data, when we discretize them, the coincidences will likely occur. These coincidences reflect the dependent information between two random variables in a bivariate situation. However, they are not shown in the traditional scatterplot. Thus, displaying the occurrence of coincidences is meaningful in understanding the dependence structure of two variables in bivariate data.

A **sunflower plot** displays the bivariate data on a plane with points labeled by integers showing the number of coincidences at each location. For count time series data, we illustrate it by the $n-1$ successive lag-1 pairs:

$$(X(t_1), X(t_2)), \quad (X(t_2), X(t_3)), \quad \ldots, \quad (X(t_{n-1}), X(t_n)).$$

We count the coincidences on different locations, and plot them with those counts (as labels) on each location. This will give us the sunflower plot for time series data, which provides more information than the scatterplot. For positive time series data, there may not be any coincidences. In such a situation, we can discretize the positive-valued data. By taking discretization, we may expect more coincidences. Then we follow the steps for the count time series data to give the sunflower plot. The integer number at each point is the frequency at that point. These frequencies contain the serial dependence information. Usually, if there is an auto-correlation between the successive pairs, it often forms some kind of ridge shape in the sunflower plot. Hence, we should probe any potential ridge shape in the sunflower plot for evidence of serial dependence for the time series data.

Figure 12.1 illustrates the sunflower plots for two count time series data. The first one is from model

$$X_{i+1} \stackrel{d}{=} 0.65 * X_i + E_i, \qquad i = 1, 2, \ldots, 500, \tag{12.1.1}$$

where the marginal distribution is Poisson(5), and $E_i \stackrel{i.i.d.}{\sim}$ Poisson(1.75) $(i = 1, 2, \ldots, 500)$, and the second one is just an iid sequence of Poisson(5) with sample size 500. The ridge shape of the first plot is different from the second one because of serial dependence.

(2) P-P plot and diagonal P-P plot.

Sometimes it may not be easy to judge the association pattern in the sunflower plot. Hence, we develop the diagonal P-P plot to detect the potential dependence between two variables in bivariate data. The diagonal P-P plot is a special P-P plot, which is the inversion of the Q-Q plot.

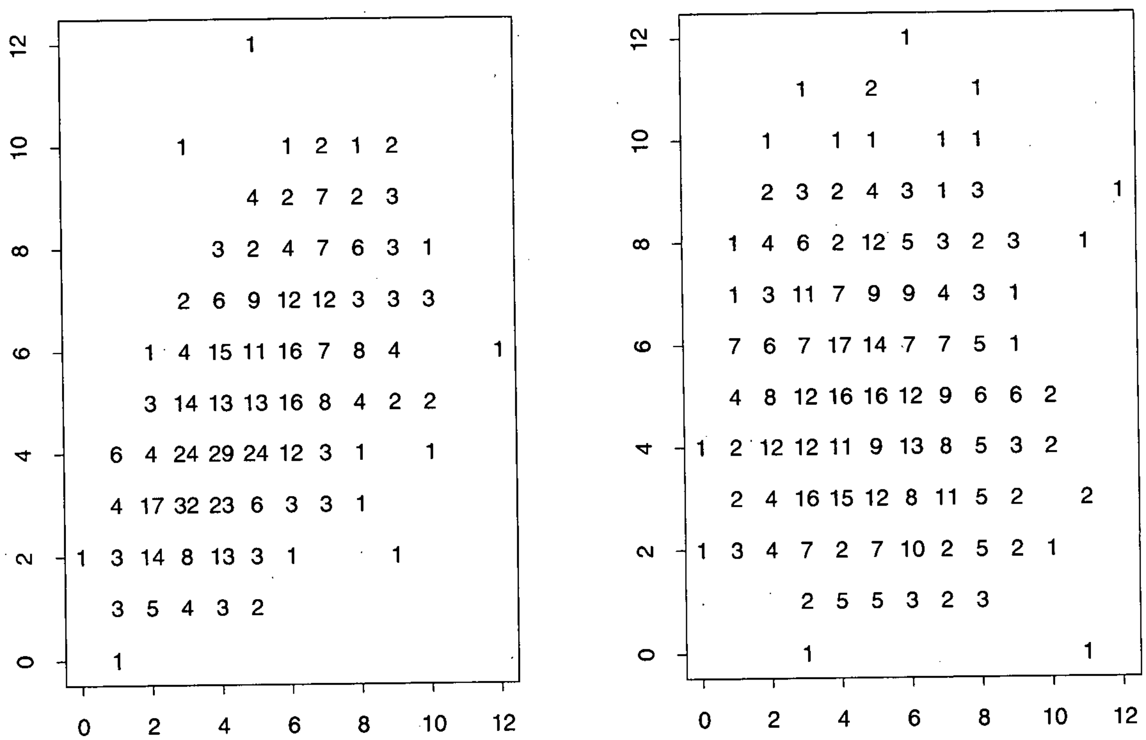


Figure 12.1: *Sunflower plots of two time series count data. The left one is from the model in (12.1.1), while the right one is from an independent Poisson*(5) *series.*

To compare the cdf's of two univariate rv's: $F_X(\cdot)$ and $F_Y(\cdot)$, the common Q-Q plot (quantile vs. quantile) will display the corresponding quantile pairs of

$$\left(F_X^{-1}(p_1),\ F_Y^{-1}(p_1)\right),\quad \left(F_X^{-1}(p_2),\ F_Y^{-1}(p_2)\right),\quad \ldots,\quad \left(F_X^{-1}(p_n),\ F_Y^{-1}(p_n)\right),\qquad 0 \le p_i \le 1,$$

in a plane. Here F^{-1} means the inverse function of F or the quantile function. If the two distributions are the same, the points on the Q-Q plot will roughly locate around the 45° diagonal line. Otherwise, the points will show a pattern deviating from the diagonal line.

The P-P plot takes an inverse approach. It compares the probabilities of two distributions. For the corresponding pairs of cdf values calculated at cut points $c_1, c_2, \ldots, c_n$:

$$(F_X(c_1),\ F_Y(c_1)),\quad (F_X(c_2),\ F_Y(c_2)),\quad \ldots,\quad (F_X(c_n),\ F_Y(c_n));$$

we plot them in a plane. Because the range of a distribution function is from 0 to 1, these points are displayed in a unit square. If the two distributions are the same, then the points will roughly lie around the 45° diagonal line. Otherwise, they will deviate from the line somehow. Therefore, the P-P plot also has the ability to check if two distributions can be considered the same or not.

It is equivalent to plot the survival probabilities, $(1 - F_X(c_i), 1 - F_Y(c_i))$, of the two distributions. As to which form, it is up to user's preference.

To investigate the independence or dependence of two random variables in bivariate data, say

$$(x_1, y_1),\quad (x_2, y_2),\quad \ldots,\quad (x_n, y_n),$$

we can borrow the idea of the P-P plot for univariate distributions. If the rv X is independent of the rv Y, then it holds that

$$\Pr[X \le a, Y \le b] \;=\; F_{(X,Y)}(a,b) \;=\; \Pr[X \le a]\cdot\Pr[Y \le b] \;=\; F_X(a)\cdot F_Y(b),$$

for any a, b in the support. Thus, we choose the grid points (a_j, b_j), where $j = 1, 2, \ldots, N$, and plot $\left(\widehat{F}_{(X,Y)}(a_j, b_j),\ \widehat{F}_X(a_j)\widehat{F}_Y(b_j)\right)$ for all j. Here

$$\begin{aligned}
\widehat{F}_X(a_j) &= \frac{\text{the number of } x_i \text{ being smaller than or equal to } a_j}{n},\\
\widehat{F}_Y(b_j) &= \frac{\text{the number of } y_i \text{ being smaller than or equal to } b_j}{n},\\
\widehat{F}_{(X,Y)}(a_j, b_j) &= \frac{\text{the number of } (x_i, y_i) \text{ where } x_i \le a_j \text{ and } y_i \le b_j}{n}, \qquad j = 1, 2, \ldots, N.
\end{aligned}$$

These grid points are in fact the two-dimensional cut points. In the resulting P-P plot, if X and Y are independent, then the points will lie around the diagonal line, if they are dependent, then $\Pr[X \leq a, Y \leq b] \neq \Pr[X \leq a] \cdot \Pr[Y \leq b]$, and the points will tend to deviate from the diagonal line.

A simplified version of P-P plot to diagnose the independence or dependence of two random variables in bivariate data is to choose the points (a_j, b_j) $(j = 1, 2, \ldots, N)$ on a line $y = cx + d$. This will reduce the burden of selecting grid points. In the stationary time series framework, this line will be chosen as the diagonal line of the first and third quadrants, namely $y = x$. We do so because $X(t_i)$ and $X(t_{i+1})$ have the same marginal distribution, hence, we should put equal weight on the two elements of the pairs $(X(t_i), X(t_{i+1}))$ which lead to this diagonal line. Due to the feature of choosing cut points on the diagonal line, we call this special graphical tool the **diagonal P-P plot**. Suppose the observations $\{X(t_1), \ldots, X(t_n)\}$ are arranged in increasing order $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, then we can choose $a_j = b_j = x_{(j)}$, where $j = 1, 2, \ldots, n$, and plot

$$\left(\widehat{F}_{12}(x_{(j)}, x_{(j)}),\ \widehat{F}^2(x_{(j)})\right), \qquad j = 1, 2, \ldots, n.$$

Here

$$\begin{aligned} \widehat{F}(x_{(j)}) &= \frac{\text{the number of } X(t_i) \text{ being smaller than or equal to } x_{(j)}}{n}, \\ \widehat{F}_{12}(x_{(j)}, x_{(j)}) &= \frac{\text{the number of } (X(t_i), X(t_{i-1}) \text{ where } X(t_i) \leq x_{(j)} \text{ and } X(t_{i-1}) \leq x_{(j)}}{n-1}, \end{aligned}$$

for $j = 1, 2, \ldots, n$. If the diagonal P-P plot shows that there exists a pattern deviating from the diagonal line, then it suggests that the serial dependence exists in the time series.

We briefly discuss the pattern of diagonal P-P plot in positively and negatively correlated bivariate distribution. We illustrate the patterns by bivariate normal distribution with standard normal margins. The correlation coefficients are chosen to be 0.5 and -0.5. Setting the X-axis being the bivariate cdf of independent margins and Y-axis being the empirical bivariate cdf of 500 samples, we obtain Figure 12.2. From these plots, we see that positive correlation leads to fish back pattern (a curve above the diagonal), while the negative correlation leads to fish belly pattern (a curve below the diagonal).

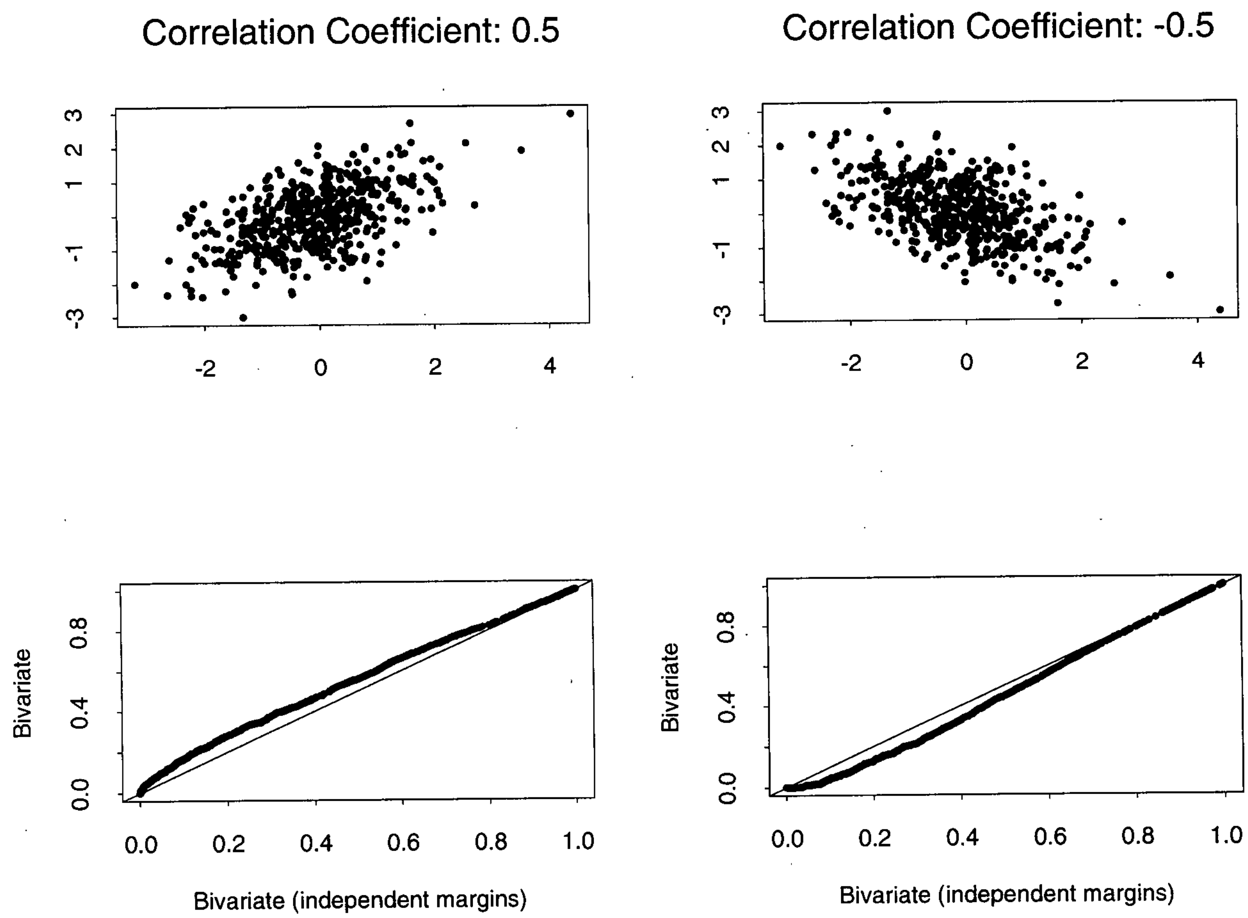


Figure 12.2: *Scatterplots and diagonal P-P plots of positively correlated and negatively correlated bivariate normal data. The left side corresponds to positive correlation, while the right side corresponds to negative one.*

The diagonal P-P plot can be applied to check whether the samples are from a specific bivariate distribution $F_{12}(x, y)$. It is just necessary to substitute $\widehat{F}^2(x_{(j)})$ in the case of independent margins by $F_{12}(x_{(j)}, x_{(j)})$, and check the diagonal P-P plot. If there is any obvious deviation, it suggests that the samples may not come from the bivariate distribution $F_{12}(x, y)$.

Note that the P-P plot for bivariate data can be easily generalized to higher dimensions which allows us to compare two multivariate distributions. We invent this P-P plot because it is easy to apply this graphical tool in high dimensions while the Q-Q plot doesn't exist in dimensions greater than or equal to 2.

Figure 12.3 illustrates the diagonal P-P plots for two count time series data. The first one is from the model in (12.1.1), and the second one is from an iid sequence of Poisson(5). One can see the pattern of deviation in the first plot, and the pattern of closeness to the diagonal line in the second plot. Therefore, they match the theory.

(3) Randomized quantile transformation plot.

The randomized quantile concept was introduced by Dunn and Smyth [1996]. It transforms skewed data $\{x_1, x_2, \ldots, x_n\}$ into symmetric data $\{r_1, r_2, \ldots, r_n\}$ to please our eyes so that we can get a more intuitive impression. This is because that our eyes handle symmetric data more easily than non-symmetric data. Empirically, it is quite complicated to understand some features from skewed distributions.

Let F be the cdf of the sampled population and Φ be the standard normal cdf. Then the randomized quantile transformation is defined as

$$r_i = \Phi^{-1}(u_i), \qquad i = 1, 2, \ldots, n,$$

where $u_i = F(x_i)$ if F is continuous at x_i, and u_i is a uniform random number on the interval $[F(x_i^-), F(x_i^+)]$ if F is not continuous at x_i. That is, it first transforms the raw data into a roughly uniform random numbers, then transforms again to standard normal random numbers. The standard normal distribution is symmetric around the origin.

We can borrow this idea for the count and positive-valued time series data to obtain the **randomized quantile transformation scatterplot**, which is third useful graphical tool to por-

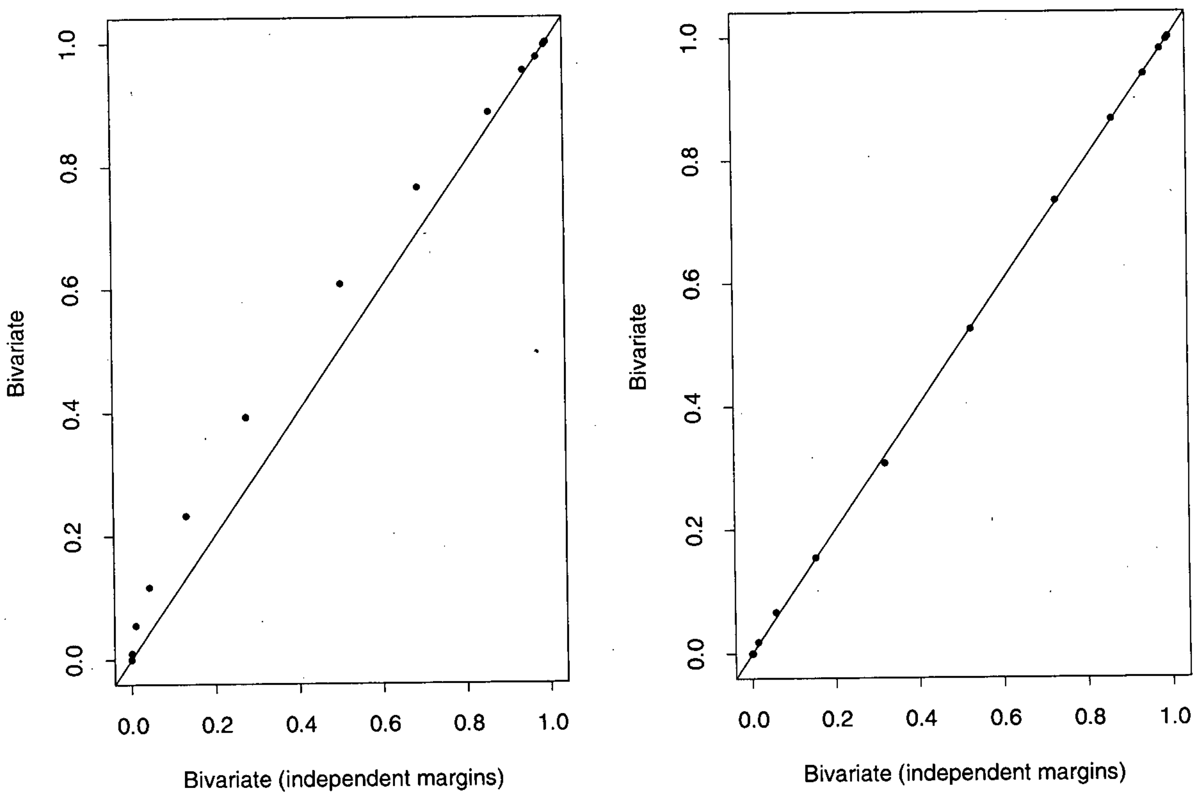


Figure 12.3: *Diagonal P-P plots of two time series count data. The left one is from the model in (12.1.1), while the right one is from an independent Poisson*(5) *series.*

tray the dependence of two random variables in bivariate data. This is because almost all the marginal distributions in such kinds of time series are skewed. We can first calculate the randomized quantile transformation for the stationary time series data $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ to obtain $\{r_1, r_2, \ldots, r_n\}$, then we plot the scatterplot for the pairs

$$(r_1, r_2), \quad (r_2, r_3), \quad \ldots, \quad (r_{n-1}, r_n),$$

namely the traditional scatterplot for the pairs (r_i, r_{i+1}) $(i = 1, 2, \ldots, n-1)$. Here F, the marginal distribution, can be estimated by the data either parametrically or non-parametrically. If the original time series data are from an iid sequence, then the randomized quantile transformations are also iid. Thus, the randomized quantile transformation scatterplot won't show any association pattern. If the original data are serially correlated, then the randomized quantile transformations are also serially correlated, which leads to some kind of association pattern in the randomized quantile transformation scatterplot. Thus, if there is any association pattern in the randomized quantile transformation plot, we can conclude that there exists serial correlation in the time series data.

Figure 12.4 illustrates the randomized quantile transformation scatterplots for two count time series data. The first one is from the model in (12.1.1), and the second one is the same iid sequence as before. The distribution F in both cases is estimated parametrically, namely, we assume that two distributions are from the Poisson family, and estimate the parameters from the data. We can use the empirical distributions in both cases too. From the two scatterplots, we see different patterns. The first one shows serial dependence, while the second one suggests independence.

Finally, we discuss the **ACF plot**. The widely used ACF plot is a sophisticated tool to detect the serial dependence in the equally-spaced or unequally-spaced time series. It plots the auto-correlation coefficients against the lag lengths. It works well in Gaussian time series, as well as non-normal time series. For example, we can draw the ACF plot of the count time series from the model in (12.1.1). See Figure 12.5. In addition, it can identify one kind of non-stationarity: seasonality.

One may also want to try some analytical tests for the temporal dependence. Dependence

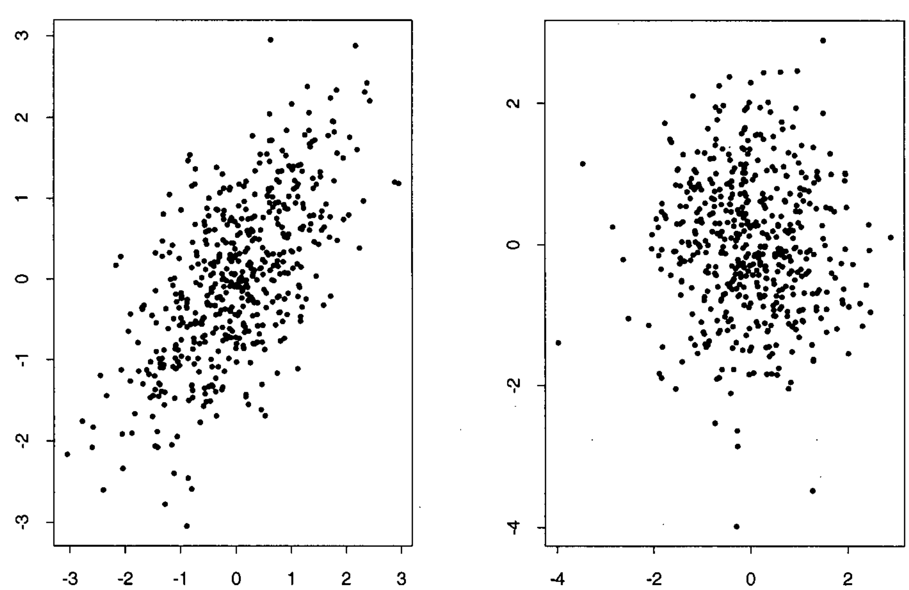


Figure 12.4: *Randomized quantile transformation scatterplots of two time series count data. The left one is from the model in (12.1.1), while the right one is from an independent Poisson*(5) *series.*

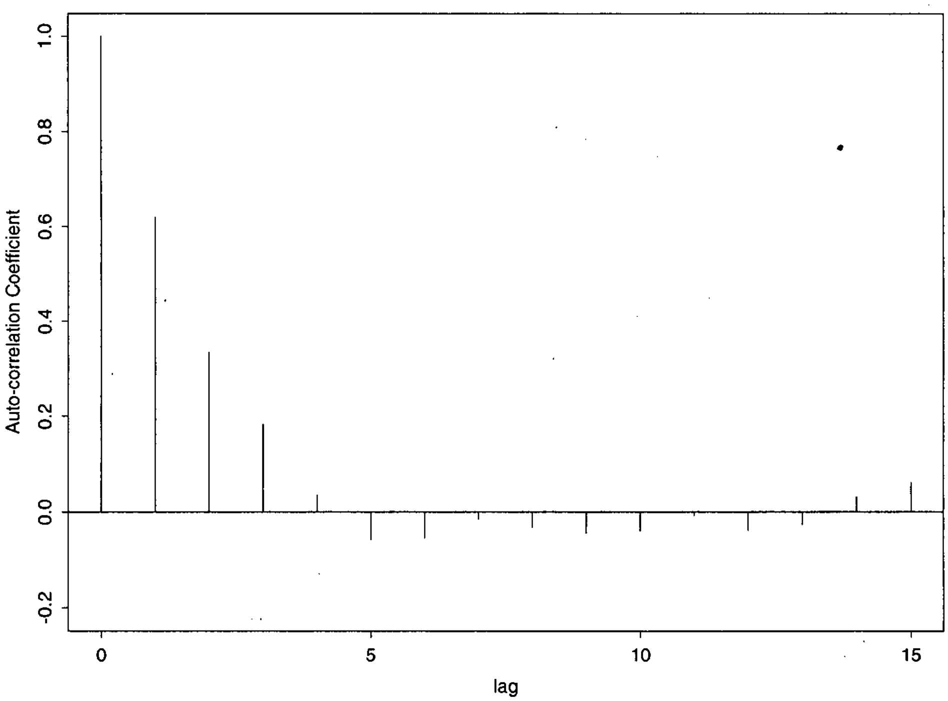


Figure 12.5: *The Auto-correlation function (ACF) plot of the count time series from the model in (12.1.1).*

over time in a time series data is often called serial correlation or serial dependence. See Anderson [1988a, 1988b]. Testing the serial dependence or correlation has been studied since the 1940's. Usually, the null hypothesis H_0 is "randomness" meaning "no serial dependence" or "independence", the alternative hypothesis H_A is "serial dependence" or "serial correlation" under some underlying stochastic process. There are many articles in the literature under the key word "serial correlation" or "serial dependence". Since the marginal distributions and dependence structures in the context of the continuous-time GAR(1) processes are quite diverse, one may apply some non-parametric tests for the serial dependence. For example, the contingency table test and Goodman's simplified runs test are two of those methods. The contingency table test is applicable for unequally-spaced time series data. But Goodman's simplified runs test is not applicable for unequally-spaced time series; see Goodman [1958] and Granger [1963].

12.2 Model selection

If we detect serial dependence in the stationary equally-spaced or unequally-spaced time series data, the next step is to find appropriate models for them. The continuous-time GAR(1) process models will be naturally considered if the observations are positive or non-negative integer-valued.

Usually, we will first investigate the observations to see what kinds of distributions could be the possible marginal distributions. For example, if the data are non-negative integer-valued, we may use the Poisson distribution for the marginal distribution if the sample mean and variance of $\{X(t_1) = x_1, \ldots, X(t_n) = x_n\}$ (or $\mathbf{x} = (x_1, \ldots, x_n)^T$) are roughly equal, or we may try the negative binomial or generalized Poisson distribution if the sample variance is much larger than the sample mean. The possible family of marginal distribution to be considered may not be unique.

Next we will select the continuous-time GAR(1) process models which have the marginal distributions under consideration. It is common that the processes with different extended-thinning operations may have the same stationary distribution. Therefore, for a specific time series, we may have several continuous-time GAR(1) process models to consider.

For the possible models, we first apply them to fit the data. Then we do the diagnostic check for each model to see if this model is suitable for the data. If a model is not suitable for the data, we will remove it from the model list. The diagnosis techniques will be studied in Section 12.3. Hence, we finally could obtain more than one suitable model.

Then a natural question arises: which one is the best to model the data? To this end, we should compare how well these models fit the data. Here we present a couple of approaches for the model selection.

AIC approach. The Akaike information criterion (AIC) is widely used in model selection. It is useful in either nested models case or non-nested models. Suppose the number of parameters $\boldsymbol{\theta}$ in a model is k, and the log-likelihood is $L(\boldsymbol{\theta}|\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n)^T$ is the vector of observations. Then, the Akaike information criterion is defined as

$$AIC = -2L(\boldsymbol{\theta}|\mathbf{x}) + 2k.$$

For the fitted model, the AIC will be evaluated at the estimates of the parameters. The models will be judged according to their AIC values; the smaller, the better. We will choose the model which has the smallest AIC value. Joe [1997] commented that this approach is in fact a penalized log-likelihood method if we look at $L(\boldsymbol{\theta}|\mathbf{x}) - k$, which takes the number of parameters as the penalty. In this equivalent criterion, we will choose the model with the largest value of $L(\boldsymbol{\theta}|\mathbf{x}) - k$.

Comparison of fit approach. This idea had appeared in Joe [1997], Section 11.5, p. 365-367. The sum of conditional least squares

$$R_{CLS}(\boldsymbol{\theta}) = \sum_{i=1}^{n-1} \Big(x_{i+1} - \mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}\right] \Big)^2$$

measures the closeness of the model with the real data. This sum will be evaluated under each suitable model with estimated value of parameter $\boldsymbol{\theta}$ from the data, namely the conditional expectation $\mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i; \boldsymbol{\theta}\right]$ is calculated under each suitable model where $\boldsymbol{\theta}$ is estimated from $\mathbf{x}$. The smaller this value is, the closer the fitted model is to the data in a prediction sense. This will lead us to choose the model with the smallest value of the sum of conditional least squares.

Prediction comparison (cross-validation) approach. If the sample size n is adequately large, we may use first part of the observations, say first half, to build each suitable model.

Then we apply each built model to predict the remaining observations. In this way, we can see which model has the best predictive ability, and such a model with the best prediction for the remaining observations will be selected. The model prediction or forecasting will be discussed in Section 12.4.

12.3 Model diagnostics and hypothesis testing

After building a continuous-time GAR(1) process model by choosing its form and estimating related parameters, we must check if the model fits the data well. This is called model diagnosis, and can avoid the naive subjective mistake in the model choice. From the view of applied statistics, we wish the subjective model to be as close to the reality as possible. If the built model doesn't fit the data well, then it implies that there is a gap between the subjectiveness and the reality, and the built model could be a wrong model. Thus, building a good or suitable model is very important, because a good model can summarize the information from the data and allow us to make correct inference. We will present a graphical diagnostic technique, the diagonal P-P plot, in Section 12.3.1.

Based on the suitable model built from the data, we may test some kinds of practical questions such as if a drug is effective, or if one treatment is better than another. These questions are part of hypothesis testing. Sometimes, they simply test the parameters in the built model, and sometimes they may test more complex composite hypotheses regarding model parameters. We will narrow the topic of hypothesis testing to the simple parameter test, and give a brief discussion in Section 12.3.2.

12.3.1 Graphical diagnostic method

Traditionally, model diagnosis involves checking some kind of residuals. See Lindsey [1997], p. 223-225 for a short summary of categories of residuals. The two widely used kinds of residuals are fitted value residuals (observed value minus fitted value) or the variations (like studentized residuals), and deviance residuals. These different types of residuals will then be displayed in the scatterplot against a variety of statistics like fitted values, or in a Q-Q plot against a specific distribution like the standard normal, to check for any obvious departures from the specified model. If no departures are observed, we will accept the fitted model as a suitable model.

For example, in linear regression, we display the scatterplot of fitted value residuals against the fitted response values or one of the covariates to check if the residuals symmetrically lie around the horizontal line within certain range, and/or show the Q-Q plot of studentized residuals against the standard normal quantiles to check the normality; in the generalized linear model, we usually display the deviance residual plot to check if the fitted model is close to the data. The residuals could be plotted against their lagged values, which is common in the normal time series.

However, in the continuous-time GAR(1) process models, it is not easy to use these two types of residuals. This is because both residuals are applicable in certain types of models or distributions. For the fitted value residual, it is very useful in the structure model like linear model where conditioned on covariates, the model can be represented by two terms: one is fixed, one is random. In this situation, the fixed term is estimated at each covariate value, and the random term is obtained by subtracting the fitted value from the observed response value. This estimated random term is in fact the fitted value residual. Checking these residuals by residual scatterplot and/or Q-Q plot, we can find if the random term matches the assumptions imposed in the specified model.

For the deviance residual, it is usually used in the stochastic model where no fixed term can be decomposed out, only random term(s). In fact, they are defined according to the specific exponential form of the pdf of the exponential family, and measure the difference of log-likelihood between a saturated model and the fitted model for each observation. See Lindsey [1997], p. 210-211, or Venables and Ripley [1994], Chapter 7 for a quick reference. Due to the discrete feature or skewed feature of some distributions in the exponential family, the deviance residual plot may show some special pattern which is hard to be understood through our eyes. That is why the randomized quantile transformation scatterplot is proposed.

The continuous-time GAR(1) process family is very rich in the stationary distributions, and varies according to the extended-thinning operation. Many of them do not belong to the exponential family specified in the generalized linear model. From the representation

$$X(t_2) \stackrel{d}{=} (\alpha)_K \circledast X(t_1) + E(t_1, t_2), \qquad t_1 < t_2,$$

we know that they are stochastic model (two random terms) unless K is from **P1** which corresponds to the constant multiplier and leads to a structure model like linear model. Hence, neither the fitted value residual nor the deviance residual can meet the diagnostic need of the continuous-time GAR(1) process model. To this end, developing a new diagnostic technique is necessary.

We propose the diagonal P-P plot introduced in Section 12.1 to diagnose the built continuous-time GAR(1) process model. This method doesn't use any kind of residuals. The idea is that any model specifies the theoretical distribution, which can then be compared with the empirical distribution obtained from the data by the P-P plot. By comparison, we can find if they are close to each other or not.

The continuous-time GAR(1) process specifies the multivariate marginal distribution for any number of adjacent margins. We choose the bivariate marginal distribution in our consideration because it is not likely to be the same for two different models in the continuous-time GAR(1) process family. Hence, for the equally-spaced time series data $\{X(t_1), \ldots, X(t_n)\}$, we can estimate the bivariate empirical distribution of lag 1 from the $n-1$ successive pairs $(X(t_i), X(t_{i+1}))$. Setting the cut points as those observations $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$, we obtain the estimate of the bivariate cdf

$$\widehat{F}_{12}(x_{(j)}, x_{(j)}) \;=\; \frac{\text{the number of } (X(t_i), X(t_{i-1}) \text{ where } X(t_i) \le x_{(j)} \text{ and } X(t_{i-1}) \le x_{(j)}}{n-1},$$

Note that these $n-1$ successive pairs are not independent. Hence, ergodicity is necessary to guarantee the consistent estimation of the joint distribution. The theoretical bivariate distribution of two adjacent margins with the same time difference as that of the data can then be calculated from the fitted model, namely

$$\begin{aligned} F_{12}(x_{(j)}, x_{(j)}) &= \Pr[X(0) \le x_{(j)},\ X(t_2 - t_1) \le x_{(j)}] \\ &= \Pr[X(0) \le x_{(j)},\ (\alpha)_K \circledast X(0) + E(0, t_2 - t_1) \le x_{(j)}]. \end{aligned}$$

Here the model parameters are estimated by the data, thus, they are known so that we can theoretically calculate the required probabilities. Then, we can draw the diagonal P-P plot of the successive pairs against the fitted model, namely plot the points

$$\left(\widehat{F}_{12}(x_{(j)}, x_{(j)}),\ F_{12}(x_{(j)}, x_{(j)})\right), \qquad j = 1, 2, \ldots, n.$$

Repeat it for lag $2, 3, \ldots, m$, where m is an adequate integer that depends on the length of the series.

In these diagonal P-P plots, if the points lie around the diagonal line, then we will accept the fitted model as suitable. Otherwise, there is an obvious departure between the data and the fitted model, which suggests that the fitted model is not suitable.

The calculation of the theoretical bivariate distribution may employ the stochastic representation for the bivariate margins, or numerical inversion of bivariate characteristic function. In practice, at the two ends of the plot, there might occur deviation from the diagonal straight line at 45°, because there are too few observed pairs at the lower end which could lead to inaccurate estimates, and the calculation of theoretical bivariate cdf is cut off at the upper end so that it is always less than 1 while the empirical cdf reaches 1.

For unequally-spaced time series data, we first divide the $n(n-1)/2$ pairs into different groups. Each group consists of pairs with common or roughly common time difference. Apply the diagonal P-P plot to the groups with adequately large number of pairs. This is because for a group with too few pairs, it's hard to obtain the empirical bivariate distribution. If all of the diagonal P-P plots show the pattern of a straight line at 45°, we will accept the fitted model as suitable. Otherwise, if any of them doesn't show this kind pattern, we will reject the fitted model because the empirical bivariate distribution doesn't match the theoretical calculated from the fitted model.

Note that the fitted value residual plot can still be applied in the continuous-time GAR(1) process model with extended-thinning operation **P1**, but one should check the histogram and Q-Q plot of the residuals because they are usually distributed in a special distribution. The deviance residual plot can also be applicable in the models where the margins have specific exponential family required in the generalized linear model, if one favors it.

12.3.2 Parameter testing

In this section, we simply test the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}^0$ under a continuous-time GAR(1) process model. We don't set up the alternative hypothesis, because it's quite subjective to choose such an alternative hypothesis. This alternative one could be the same structure model with different parameter values, or another structure model. Here we want to have a general discussion, thus, we do not have particular reason to choose one of them as the specific alternative hypothesis. Due to this lack, we can not obtain the power function.

All testing approaches we will discussed depend on the asymptotic distribution of the test statistic. Among these approaches, some depend on the estimation approach or model, and others do not.

Asymptotic normality testing approach. Most estimators have an asymptotic normal distribution, i.e.,

$$(n-1)^{1/2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}_{k\times 1},\ \boldsymbol{\Sigma}_{k\times k}\right), \qquad n \to \infty,$$

where $\boldsymbol{\Sigma}_{k\times k}$ is estimated by the data. Thus, we can use the confidence region for a test. Under H_0, the $100(1-\alpha)\%$ confidence region for the mean vector is

$$(n-1)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)^T \cdot \boldsymbol{\Sigma} \cdot \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) \ \le\ \frac{k(n-2)}{(n-k-1)} F_{k,n-k-1}(\alpha). \tag{12.3.1}$$

See Johnson and Wichern [1998], p. 236. If the null value $\boldsymbol{\theta}^0$ is in this ellipsoid, we accept H_0, otherwise we reject it.

This method depends on the estimation approach, but doesn't depend on the model.

Log-likelihood testing approach. For the MLE estimator, it follows that

$$2\left[\log L_n(\widehat{\boldsymbol{\theta}}_{MLE} \mid \mathbf{x}) - \log L_n(\boldsymbol{\theta}^0 \mid \mathbf{x})\right] \xrightarrow{\mathrm{L}} \chi^2_k, \qquad n \to \infty,$$

where L_n is the likelihood function under the model. Refer to Theorem 11.2.6. This fact provides another testing method. We first calculate the value of $c = 2\left[\log L_n(\widehat{\boldsymbol{\theta}}_{MLE} \mid \mathbf{x}) - \log L_n(\boldsymbol{\theta}^0 \mid \mathbf{x})\right]$ under the assumed $\boldsymbol{\theta}^0$, then we obtain the p-value $p = \Pr[Y > c]$ where $Y \sim \chi^2_k$. We accept the null hypothesis if $p \le \alpha$, and reject it otherwise.

This method depends on both the estimation approach and model.

Conditional least squares testing approach. Similar to the MLE, the CLS estimator has an analogous result:

$$R_n(\boldsymbol{\theta}^0) - R_n\left(\widehat{\boldsymbol{\theta}}_{CLS}\right) \xrightarrow{\mathrm{L}} \sum_{j=1}^{k} \lambda_j Y_j, \qquad n \to \infty,$$

where R_n is the sum of conditional least squares defined in (11.1.4), λ_j $(j = 1, 2, \ldots, k)$ are the (non-negative) eigenvalues of $\mathbf{V}^{-1}\boldsymbol{\Sigma}$, and $Y_1, Y_2, \ldots, Y_k$ be independent and identically distributed in χ_1^2. Here $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are estimated by (11.3.6) and (11.3.7). See Theorem 11.3.4. This result also provides another testing method for the CLS estimator. Under the assumed $\boldsymbol{\theta}^0$, we calculate the value of the difference $c = R_n(\boldsymbol{\theta}^0) - R_n\left(\widehat{\boldsymbol{\theta}}_{CLS}\right)$, and then the p-value $p = \Pr[\sum_{j=1}^{k} \lambda_j Y_j > c]$. This probability consists of a high dimensional integration or inversion of a characteristic function. It can be computed by the method proposed by Imhof [1961]. We will accept H_0 if $p \leq \alpha$, and reject it otherwise.

This method depends on the estimation approach, but doesn't depend on the model (other than the conditional expectation).

There exists some other approaches which use non-parametric statistics as the test statistics such as the contingency table test, Goodman's simplified run test, etc. These tests are also applicable in testing the Markov process.

12.4 Forecasting

In some fields like economics or actuarial science, people need to make decisions for the future. This leads to the issue of forecasting. In the following, we shall discuss three forecasting approaches to meet such a need. Freeland [1998] studied all these three methods in a special GAR(1) time series with Poisson margins.

Suppose we have a continuous-time GAR(1) process model, which has observation $X(t_1) = x_1$. Our task is to forecast the future value of $X(t_2)$ where $t_2 > t_1$. This is equivalent to predict the conditional random variable $X(t_2)$, $\widehat{X(t_2)}$, under certain criterion. The common criteria are minimum mean squared error, minimum mean absolute error, maximum likelihood, and so on.

Conditional mean. The criterion of minimum mean squared error will yield the estimate

$$\widehat{X(t_2)} = \mathbf{E}[X(t_2) \mid X(t_1) = x_1]. \tag{12.4.1}$$

This value usually can be obtained easily by an explicit formula. However, it is in general a real number, not an integer. We may use it in the model with positive-valued margins. For models with non-negative integer-valued margins, this forecast may not be natural.

Conditional median. The criterion of minimum mean absolute error leads to the estimate of $X(t_2)$ conditioned on $X(t_1) = x_1$ to be the median $Q(0.5)$, of the conditional distribution $F_{X(t_2)|X(t_1)}(x_2 \mid x_1)$,

$$\widehat{X(t_2)} = Q(0.5). \tag{12.4.2}$$

This value is in the support of the marginal distribution. Hence, it can be used in models with any type of margin. However, due to the lack of explicit formulas, we may have to pay the computational price.

Conditional mode. By the criterion of maximum likelihood, we mean that the random variable will have the largest chance. This will lead to the conditional mode, Q, of the distribution $F_{X(t_2)|X(t_1)}(x_2 \mid x_1)$ to be the estimate:

$$\widehat{X(t_2)} = Q. \tag{12.4.3}$$

Similar to the conditional median, this value is in the support of the marginal distribution. Hence, it also can be used in models with any type of margins.

Sometimes we may be interested in interval prediction rather than point prediction. Based on the known conditional cdf, we can construct a prediction interval with given probability by cutting two sides with half of the given probability. This approach is called **conditional prediction interval** method, and is independent to any criterion mentioned before.

12.5 Simulation of the continuous-time GAR(1) processes

The simulation of the continuous-time GAR(1) process is the basis for further simulation study of methods of parameter estimation. For the general need, we want to simulate the equally-spaced or unequally-spaced time series $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ from a stationary continuous-time continuous-time GAR(1) process $\{X(t); t \geq 0\}$. If the margins of the continuous-time GAR(1) process are non-negative integer-valued, we even can simulate the whole path in any time range of the continuous-time process by simulating the jump points. However, for the continuous margin case, we can't simulate the continuous path of the continuous-time process, instead, what we can do is to set the time increment very small and simulate the discrete-time process as an approximation.

There are two simulation methods: the conditional and the embedding approaches. The conditional approach will simulate the next observation conditioned on the current observation. It takes advantage of the stochastic representation of the conditional random variable and the conditional distribution. By repeating the steps in the conditional approach, we can simulate the equally-spaced or unequally-spaced time series. The embedding approach works for the discrete state space only, where we will simulate every sojourn time and successive jumping state alternately. In this way, we can obtain the continuous-time path of the continuous-time GAR(1) process. But both methods need the starting state at time $t = 0$. This starting point can be simulated from the stationary distribution of the process.

First, we consider the conditional approach. Without loss of generality, we just consider the simulation of $X(t_2)$ given $X(t_1) = x_1$ where $t_1 < t_2$, namely a conditional random variable. This conditional random variable is denoted as $[X(t_2) \mid X(t_1) = x_1]$. According to the model,

$$[X(t_2) \mid X(t_1) = x_1] \stackrel{d}{=} (\alpha)_K \circledast x_1 + E(t_1, t_2), \qquad \alpha = e^{-\mu(t_2 - t_1)}, \tag{12.5.1}$$

where all parameters such as μ are known. Thus, this conditional rv is the sum of two rv's: $(\alpha)_K \circledast x_1$ and $E(t_1, t_2)$. The simulation of K from **I1** to **I5** is straightforward because their pgf's have closed form and Taylor expansions can be taken to obtain the pmf's. For **I1** and **I2**, K has a simple stochastic representation. For non-negative integer-valued K, $(\alpha)_K \circledast x_1$ is just the summation of the iid copies of K, thus, its simulation is rather straightforward. However, for positive-valued

K, it's not as clear as that for the non-negative integer-valued case. For **P1**, it's trivial because K is constant and the extended-thinning operation is just the constant multiplication. For **P2**, it's compound Poisson with exponential, thus, $(\alpha)_K \circledast x_1$ is another rv of compound Poisson with exponential. For **P3** to **P5**, we don't know of a simple approach for simulation. But we guess that they may be compound Poisson or Poisson mixture. This is under further study. Also $E(t_1, t_2)$ is another concern in simulation. Recalling from Section 9.2, we find some examples where $E(t_1, t_2)$ has a stochastic representation. One advantage of stochastic representation is that it leads to easy simulation. If both $(\alpha)_K \circledast x_1$ and $E(t_1, t_2)$ have stochastic representations, then we can simulate $[X(t_2) \mid X(t_1) = x_1]$ easily. Thus, it provides a direct and fast way to simulate the conditional random variable. The following are two examples.

Example 12.1 (Poisson margins) *Consider the stationary continuous-time GAR(1) process with representation*

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} * X(t_1) + E(t_1, t_2), \qquad t_1 < t_2, \tag{12.5.2}$$

where $E(t_1, t_2) \sim \text{Poisson}\left(\frac{\lambda}{\mu}\left(1 - e^{-\mu(t_2-t_1)}\right)\right)$, $\lambda,\ \mu > 0$. *This process has Poisson*$\left(\frac{\lambda}{\mu}\right)$ *as the marginal distribution. Since the dependent term* $e^{-\mu(t_2-t_1)} * X(t_1)$ *and innovation term* $E(t_1, t_2)$ *can be simulated directly, we can obtain the simulation of* $[X(t_2) \mid X(t_1) = x_1]$ *easily.*

We illustrate the simulation by generating a time series from this model on the equally-spaced time points: $t_i = i$ *for* $i = 1, \ldots, 100$, *with* $\lambda = 2.15$ *and* $\mu = 0.43$. *Hence, the marginal distribution is Poisson*(5). *Figure 12.6 shows one simulation.*

Example 12.2 (Gamma margins) *Consider the stationary continuous-time GAR(1) process with representation*

$$X(t_2) \stackrel{d}{=} e^{-\mu(t_2-t_1)} \bullet X(t_1) + E(t_1, t_2), \qquad \text{where } \phi_{E(t_1,t_2)}(s) = \left(\frac{\beta + e^{-\mu(t_2-t_1)} s}{\beta + s}\right)^{\delta}, \tag{12.5.3}$$

and the process with representation

$$X(t_2) \stackrel{d}{=} \left(e^{-\mu(t_2-t_1)}\right)_K \circledast X(t_1) + E(t_1, t_2), \tag{12.5.4}$$

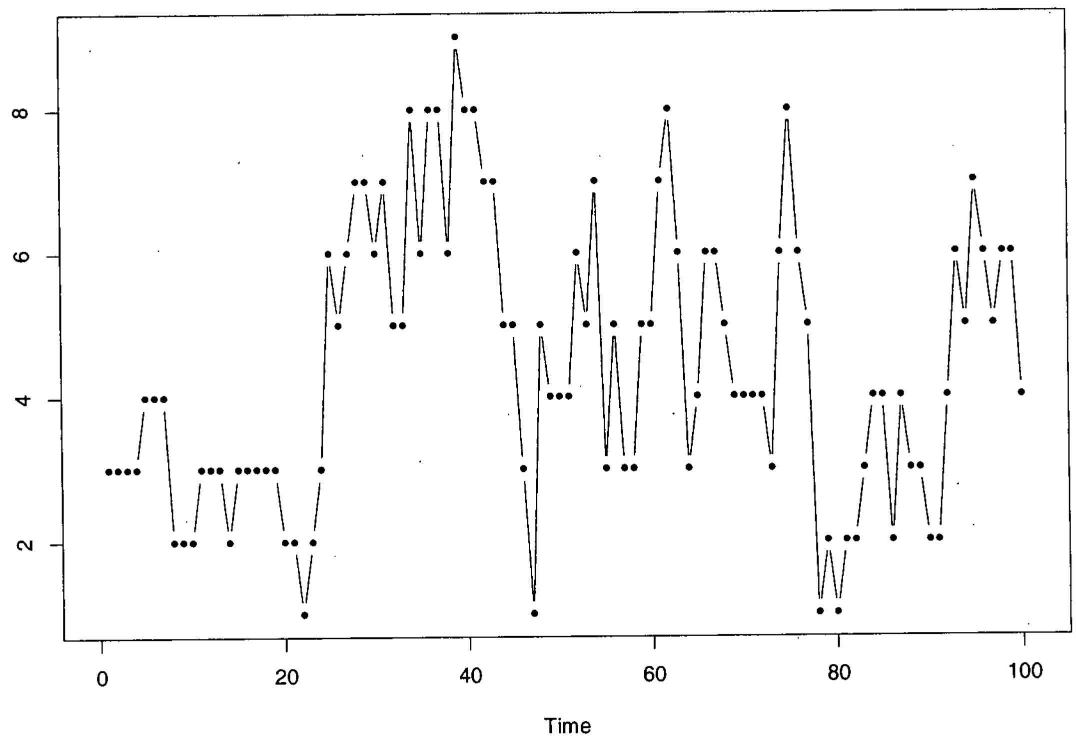


Figure 12.6: Simulation of a time series with length 100 *from (12.5.2) with* $\lambda = 2.15$ *and* $\mu = 0.43$.

where $\phi_{E(t_1,t_2)}(s) = \left(\frac{\beta}{\beta+\left(1-e^{-\mu(t_2-t_1)}\right)s}\right)^{\delta}$, *and the LT of* $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ *is*

$$\phi_{\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x}\left(s; e^{-\mu(t_2-t_1)}\right) = \exp\left\{-\frac{e^{-\mu(t_2-t_1)}s}{\beta+\left(1-e^{-\mu(t_2-t_1)}\right)s}\right\}.$$

Here $\mu, \delta, \beta > 0$. *Both models lead to Gamma*(δ, β) *margins. In (12.5.3), we need only to simulate* $E(t_1, t_2)$. *Its stochastic representation can be found in Example 9.3 which helps us to simulate this innovation. In (12.5.4),* $E(t_1, t_2) \sim \mathrm{Gamma}\left(\delta, \beta/\left(1-e^{-\mu(t_2-t_1)}\right)\right)$, *and* $\left(e^{-\mu(t_2-t_1)}\right)_K \circledast x$ *can be simulated by the stochastic representation in (9.2.5).*

Similarly, we simulate the time series from both models at time points: $t_i = i$ *where* $i = 1, \ldots, 100$, *with* $\mu = 0.54$, $\delta = 8.1$ *and* $\beta = 0.17$. *The marginal distribution is Gamma*$(8.1, 0.17)$ *with mean* 47.65. *Figure 12.7 shows the simulations.*

However, in many cases, we do not have a stochastic representation for $E(t_1, t_2)$, and thus for $[X(t_2) \mid X(t_1) = x_1]$. For example, for most GAR(1) processes with the binomial thinning operation, we only know the pgf forms of $E(t_1, t_2)$, but don't know the stochastic representations. In this situation, we turn to the conditional distribution of $[X(t_2) \mid X(t_1) = x_1]$, $F_{X(t_2)|X(t_1)}(x_2 \mid x_1)$. Specifically, we can first simulate $(\alpha)_K \circledast x_1$ and $E(t_1, t_2)$ according to their pmf or pdf separately, and then sum these two random numbers to get the desired simulation of $[X(t_2) \mid X(t_1) = x_1]$. For the stationary GAR(1) process with the binomial-thinning operation, once the pmf of marginal distribution is known, we can obtain the pmf of innovation term and consequently the pmf of $[X(t_2) \mid X(t_1) = x_1]$. This enables us to simulate the stationary GAR(1) process with binomial thinning operation, in which the pmf of marginal distribution is known. As to simulation methods for $(\alpha)_K \circledast x_1$ and $E(t_1, t_2)$, refer to Rubinstein [1981] or other books on simulation.

For the GAR(1) processes with the binomial thinning operation, once we know the marginal distribution and the parameter involved in binomial thinning, we can obtain the pmf of the innovation term $E(t_1, t_2)$, as well as the conditional random variable $[X(t_2) \mid X(t_1) = x_1]$ in (12.5.1). These will allow us to simulate the innovation $E(t_1, t_2)$ or the conditional random variable $[X(t_2) \mid X(t_1) = x_1]$ directly.

Secondly, we study the embedding approach. This approach is valid for the continuous-time

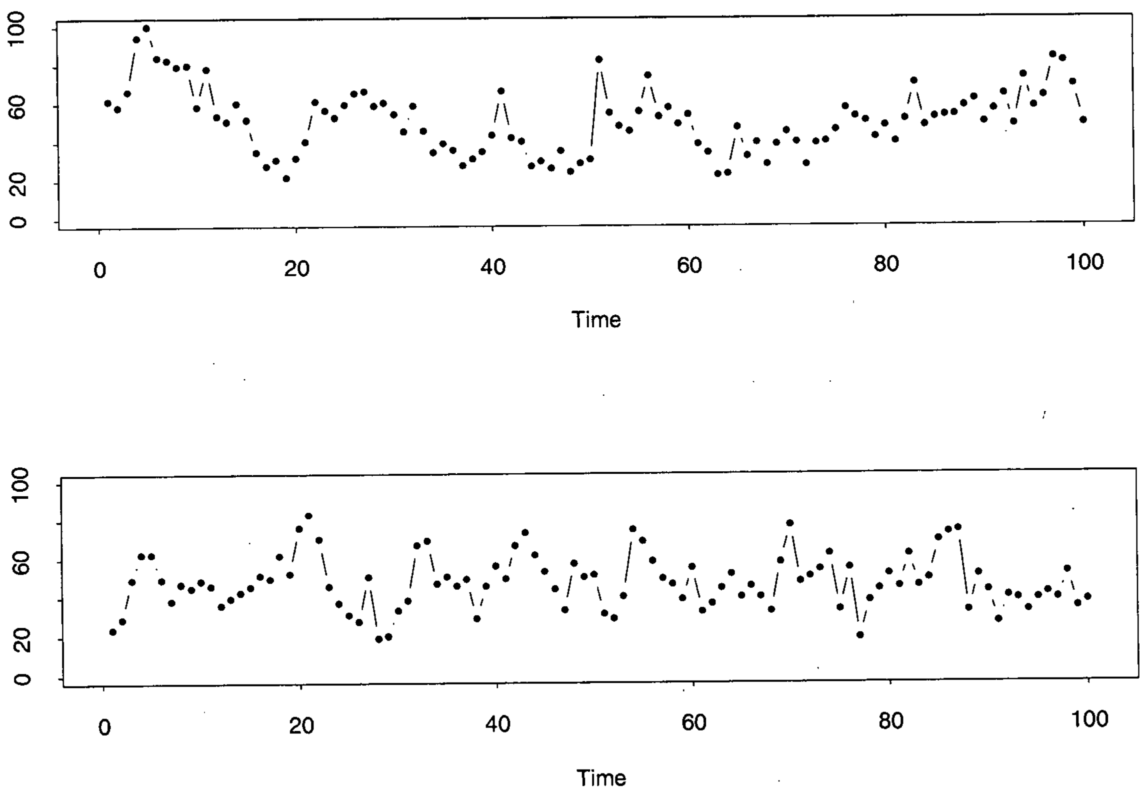


Figure 12.7: *Simulation of time series with length* 100 *from (12.5.3) and (12.5.4). Both processes have* $Gamma(8.1, 0.17)$ *margins.*

Markov processes with discrete states. The continuous-time GAR(1) processes with non-negative integer margins belong to this family. The feature of the path or realization of this kind of process is a random step function. It will stay at one state for certain time, then jump to another state, and so on so forth. The key points in embedding approach are

- the distribution of sojourn time;
- the probabilities of jumps to other states.

These two probability structures are governed by the infinitesimal generator.

For a continuous-time GAR(1) process model with non-negative integer margins, we can obtain its infinitesimal generator $Q = (q_{i,j})$ by (8.1.6), where $i, j = 0, 1, 2, \ldots$. This infinitesimal generator Q is downwardly skip-free, namely $q_{i,j} = 0$ if $j < i - 1$. We say the process is in state i if the path is taking value i. Count the starting state as the first jump, then denote the sojourn time of the l^{th} jump since beginning as

$$T_i^{(l)}=\{\text{waiting time since the } l^{th} \text{ jump to state } i \text{ until next jump happens}\}.$$

By (8.3.1), $T_i^{(l)} \sim \text{Exponential}(|q_{i,i}|)$. Hence, we simulate $T_i^{(l)}$ by an exponential random number with parameter $|q_{i,i}|$. Then, the process jumps to a state j other than i according to probability $q_{i,j}/|q_{i,i}|$. However, by the property of downwardly skip-free, we only need to consider the states $\{i-1, i+1, i+2, \ldots\}$. Repeating these two steps, we will obtain two sequences: one is state sequence $\{S_1, S_2, \ldots, S_l, \ldots\}$ and one is sojourn time sequence $\{T_{S_1}^{(1)}, T_{S_2}^{(2)}, \ldots, T_{S_l}^{(l)}, \ldots\}$. Plotting two sequences on a plane as a step function, we will obtain the graph of a path of the continuous-time continuous-time GAR(1) process model.

From this path, we can get the equally-spaced or unequally-spaced time series at time points $\{t_1, t_2, \ldots, t_n\}$. Let

$$\tau_j = \sum_{l=1}^{j} T_{(S_l)}^{(l)}, \qquad j = 1, 2, \ldots.$$

If $\tau_j \le t_i < \tau_{j+1}$, then $X(t_i) = S_j$, where $i = 1, 2, \ldots, n$. The following is an example with the same model as Example 12.1.

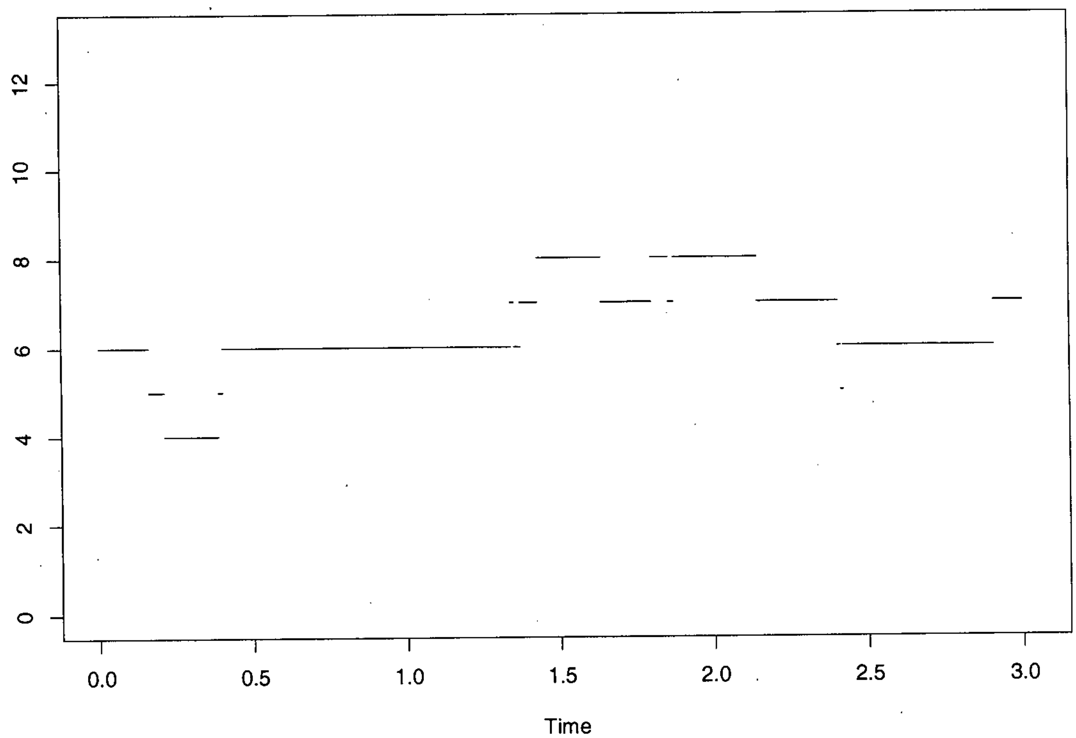


Figure 12.8: Simulation of a continuous-time path from (12.5.2) with $\lambda = 2.15$ *and* $\mu = 0.43$.

Example 12.3 (Poisson margins) *Consider the GAR(1) process in Example 12.1 with Poisson*(λ/μ) *margins. According to Example 8.1,*

$$q_{i,i} = -(\lambda + i\mu), \quad q_{i,i+1} = \lambda, \qquad i = 0, 1, 2, \ldots; \quad q_{i,i-1} = i\mu, \qquad i = 1, 2, 3, \ldots.$$

Hence, from i*, the process will jump to* $i+1$ *with probability* $\lambda/(\lambda+i\mu)$*, and to* $i-1$ *with probability* $i\mu/(\lambda + i\mu)$.

Choosing $\lambda = 2.15$ *and* $\mu = 0.43$*, we simulate a continuous-time path for such a model. See Figure 12.8, which shows a simulation up to* $t = 3$.

Chapter 13

Applications and data analyses

In this chapter, we analyze some real time series data to see the capability of the continuous-time generalized AR(1) process as the model. These real cases have equally-spaced or unequally-spaced time series data which are non-negative integers or positive real numbers. In practice, one approach for such kinds of data, is to transform them into real values, say logarithmic transformation, and then apply Gaussian time series models. By virtue of the generalized continuous-time AR(1) process, we can model the count or positive time series data directly. This brings a new approach to the real cases. Such a transition of methodology is analogous to that of linear model to generalized linear model approach.

This new approach also brings new thinking for modelling time series data; that is, how we choose a model from the family of the continuous-time generalized AR(1) process. We give a brief discussion in Section 13.1 before we proceed to the real data in the subsequent sections. In Sections 13.2 to 13.4, we analyze real data and illustrate the modelling theory.

13.1 Introduction to modelling procedure

In this section, we summarize the procedure to model stationary count or positive-valued time series (equally-spaced or unequally-spaced) with the GAR(1) process, because it is different from

Gaussian time series modelling. Some special features arise when we use the GAR(1) process as the model.

First, we need to investigate the marginal distribution. This is the significant difference from Gaussian time series modelling where we never have this question due to the normal assumption on the marginal distribution. However, for count or positive-valued time series, we need to know what kind of distribution could be the marginal distribution. This is important information to guide the model choice. To this end, we will check the histogram, mean, variance and skewness of the observations so that we can find the proper choices, e.g. Poisson, negative binomial or Gamma, for the margins. No doubt, the support of the marginal distribution of the time series data is one of the major factors to motivate us to use the GAR(1) process model. For example, it is not appropriate to use Gaussian time series model for the count time series if the observed values are not large; also it is not appropriate for positive-valued time series if the variation is large. In such situations, we may try the GAR(1) models with non-negative integer support margins. However, this is not the absolute rule. Sometimes, for the positive-valued time series, say the daily price series in economics, if the variation is small, we may be satisfied with the Gaussian time series model, because the normal marginal distribution can cover that range with probability near one. However, if the variation is large, the normal marginal distribution is no longer convincing. Thus, besides the support of the marginal distribution, the variation is also a major factor to influence us to use the GAR(1) process model or not. Empirically, we would try the GAR(1) process model for time series with large variation compared with its mean. With the information on suitable marginal distributions, we can find the corresponding choices of GAR(1) processes.

Secondly, we check the serial dependence or auto-correlation. This can be done by the ACF plot, sunflower plot, randomized quantile transformation plot, diagonal P-P plot, or other analytical methods. If time series data can be modelled by a GAR(1) process, then its auto-correlation coefficients will be positive and geometrically decreasing as the time lag increases. This can be seen by rough calculations: suppose for an equally-spaced time series, the lag one auto-correlation coefficient is 0.8, then the lag 2 to lag 6 auto-correlation coefficients will be

$$0.8^2 = 0.64, \quad 0.8^3 = 0.512, \quad 0.8^4 = 0.41, \quad 0.8^5 = 0.328 \quad \text{and} \quad 0.8^6 = 0.262.$$

Hence, time series data with such high lag one auto-correlation coefficient and medium sample size will show the geometrical decrease in the first few lags, then fluctuate within a small range in its ACF plot. Therefore, for the sample ACF plot, we only focus on the first few lags. The evidence of fast decrease in the ACF plot would suggest the GAR(1) process models may be appropriate. In general, the ACF plot can also help us to check for seasonality or non-stationarity in the time series.

If the evidence of serial dependence is strong enough, we could fit a GAR(1) process model to the data. Because sometimes different GAR(1) processes can have the same marginal distribution, we may have many choices of models for the data. To find the proper models, we need to study the mechanism or behavior of the underlying process for real problems. Such information will help us to pin down the reasonable models. If no such information is available, we can try those models with the required marginal distributions found in the first step. In this situation, if the involved self-generalized rv K is not unique, then the preferred model is the simpler one.

These fitted GAR(1) models will be diagnosed with the diagonal P-P plot proposed in Section 12.1. This is another difference from Gaussian time series modelling where we usually check the ACF plot for the estimated residuals. But now we think that any residual concept is not universally appropriate in the diagnosis of GAR(1) models, although it might be still valid for some specific models. In this situation, we will draw the diagonal P-P plots for lag 1 to lag k where the positive integer k is chosen to be adequate, say 6.

The data to be studied in the following sections don't involve any covariates, although these often exist in longitudinal studies. We will be developing methods in the future to incorporate covariates. The examples in this chapter are mainly illustrative and on trial to obtain some necessary experience with the use of the GAR(1) model.

13.2 Manuscript data study

In this section, we shall analyze an unequally-spaced count time series which records the number of manuscripts in the refereeing queue of Prof. H. Joe.

Table 13.1: *Summary of the frequencies of the number of manuscripts in refereeing queue.*

Number	0	1	2	3	4	5	6	7	8	9
Frequency	8	17	25	19	6	8	0	0	1	1
Proportion	0.094	0.200	0.294	0.224	0.071	0.094	0.000	0.000	0.012	0.012

Prof. Joe has been serving as a referee for many academic journals for many years. Every year he receives a certain number of manuscripts now and then. Upon receiving a manuscript, he will immediately decide to be a referee or not. The time series data (unequally-spaced) since January 1, 1990 is based on retrospective construction from dates of correspondence. The recording dates are given as the first day of a month because the exact month but not the day can be recovered. After March 1999, the record keeping was better and the data are monthly. Here we have the data until December 2000 so that the total length is 11 years. Prof. Joe feels that his refereeing process is relatively smooth with no obvious increase or decrease, or other non-stationary patterns in this period. The data are given in Appendix A.1.

We plot the manuscript number against date (in month) in Figure 13.1, where the starting time 0 corresponds to January 1, 1990. Here we choose month, rather than day, as the time unit because of the data recording feature and the sake of simplicity. Next, we treat this series data as univariate data, and check its histogram. The sample size $n = 85$, sample mean is 2.412 and sample variance is 2.793. The frequencies and proportions (rounded) of the observed number of manuscripts are summarized in Table 13.1. Since the sample distribution is skewed, and sample mean and variance are quite close, it may suggest the Poisson for the marginal distribution. In fact, ignoring dependence, Poisson(2.412) does fit the manuscript data very well. See Figure 13.2 for the evidence of the skewed distribution and Poisson fit.

Is this data set just an independent series, or to the contrary, a serially dependent series? If there is no serial dependence, we will be happy to end with the modelled distribution of Poisson(2.412). Otherwise, further modelling is needed. Because these time series data are unequally-spaced, we have to carefully select these pairs with a given monthly lag. The number of these pairs are summarized in Table 13.2. These groups of pairs lead to the ACF plot in Figure

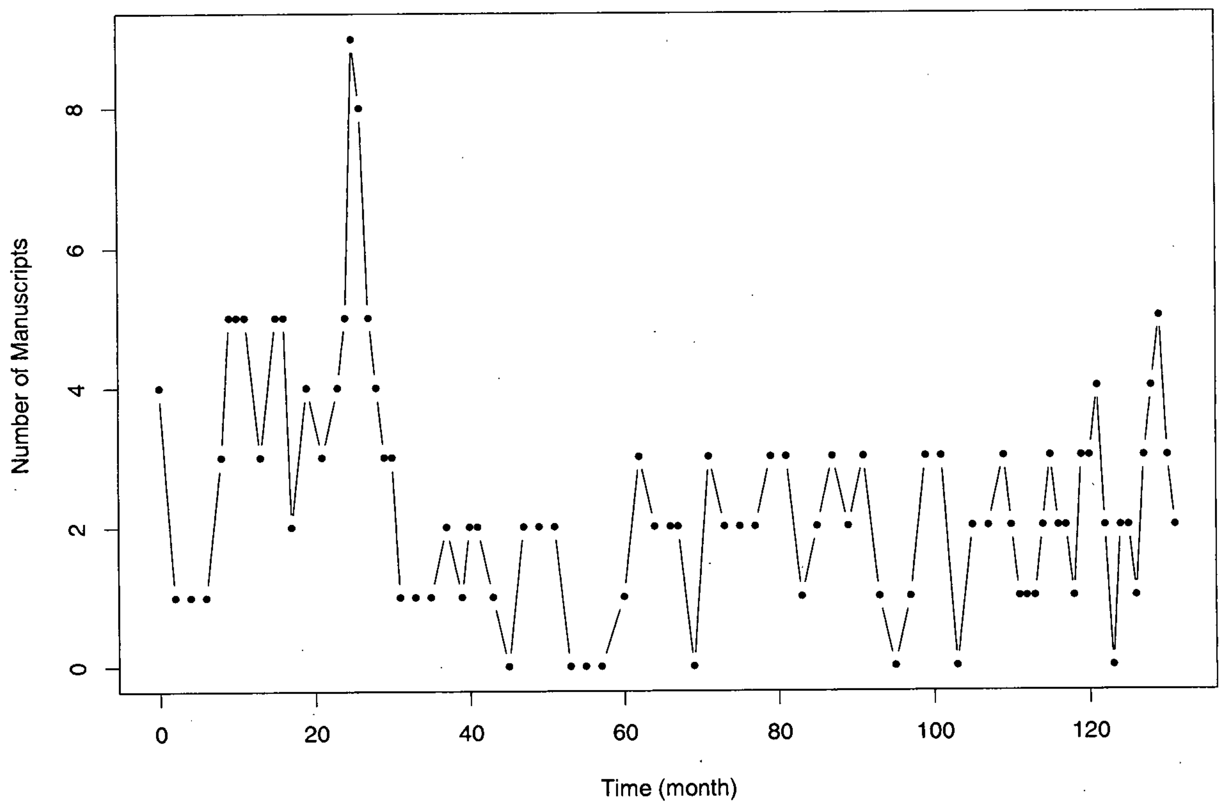


Figure 13.1: *The time series plot of refereeing queue length of manuscripts.*

Table 13.2: *Summary of the number of pairs by lag for the manuscripts data.*

Lag month	1	2	3	4	5	6	7	8	9	10	11	12
Number of pairs	38	76	38	70	40	65	42	60	42	58	41	55

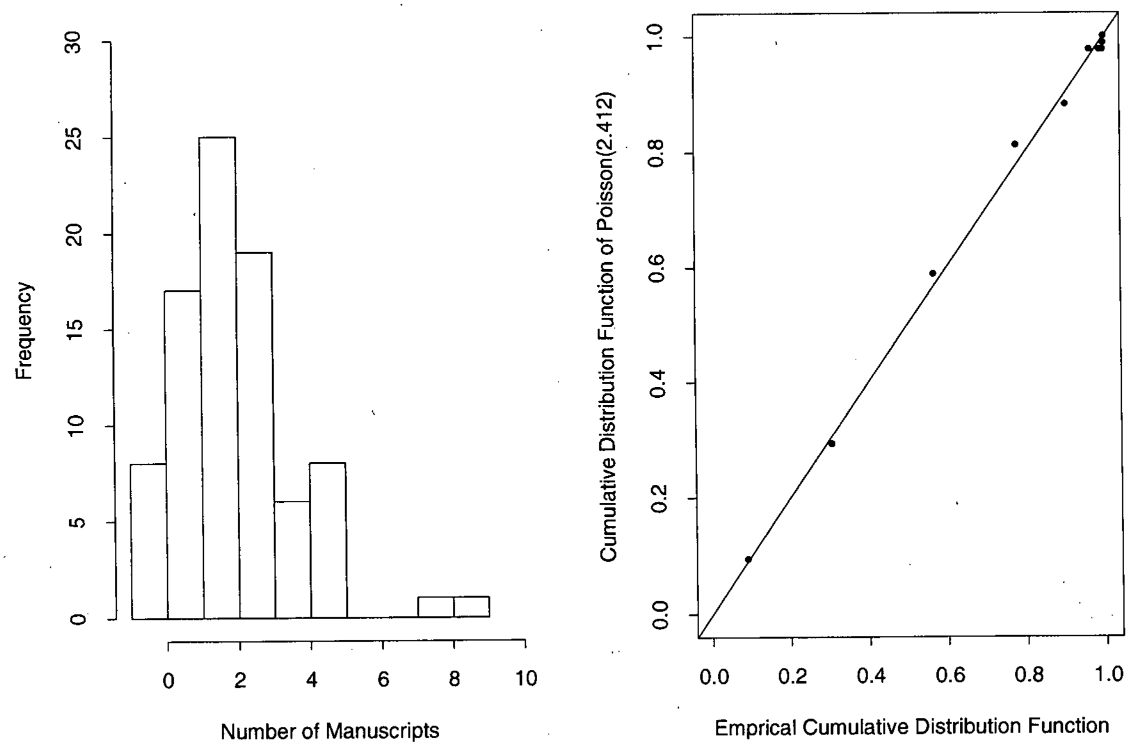


Figure 13.2: *The histogram of the manuscript data, and its P-P plot against* Poisson(2.412).

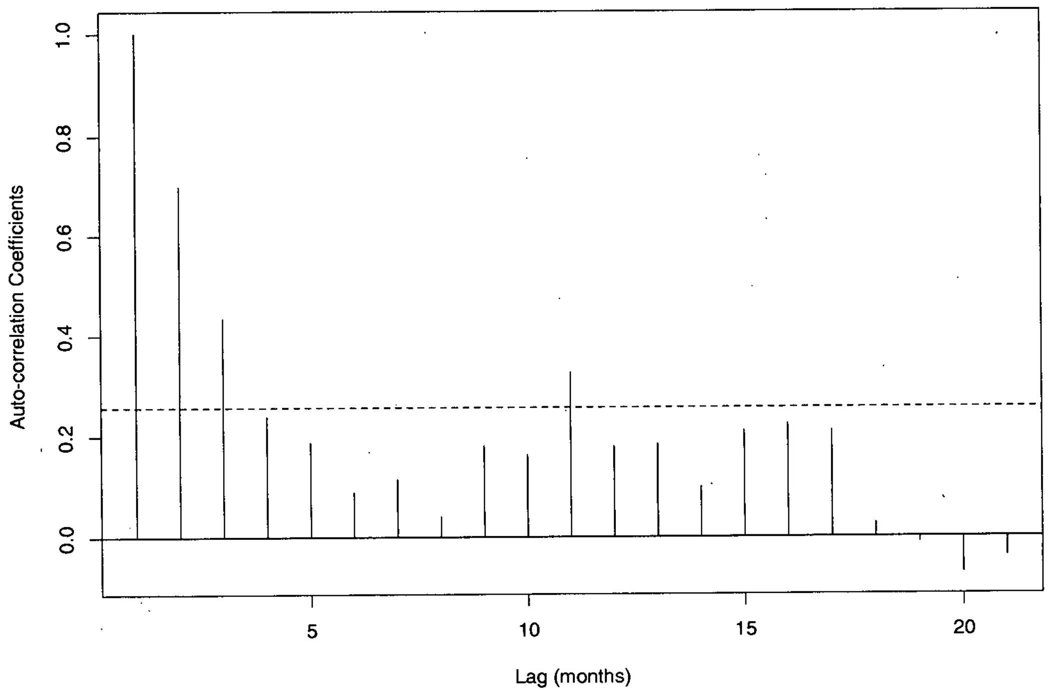


Figure 13.3: *The ACF plot of the manuscript data. The dotted horizontal line indicates the* 95% *boundary of the estimate of correlation coefficient for* 85 *pairs of independent Poisson*(2.412) *random variables; the boundary is obtained by simulation.*

13.3. The ACF plot shows positive auto-correlation in the first couple of points, and then decreases quickly to the 95% upper limit of the estimated correlation coefficient of 85 pairs of independent Poisson(2.412) random variables. The 95% critical value is based on 10,000 simulations. We choose the sample size 85 which is bigger than the number in each lagged pair group. Thus, it will lead to a conservative boundary for all lagged pair groups. Besides, we do not observe any seasonality or trend pattern on the ACF plot. This phenomena suggests that there is a strong serial dependence and the series could be modelled by a GAR(1) process. The serial dependence is also disclosed by the sunflower plot, randomized quantile transformation plot, and diagonal P-P plot. To save space, we only show them for lag 1 to 3 months; see Figure 13.4. All plots for lag one month pairs show

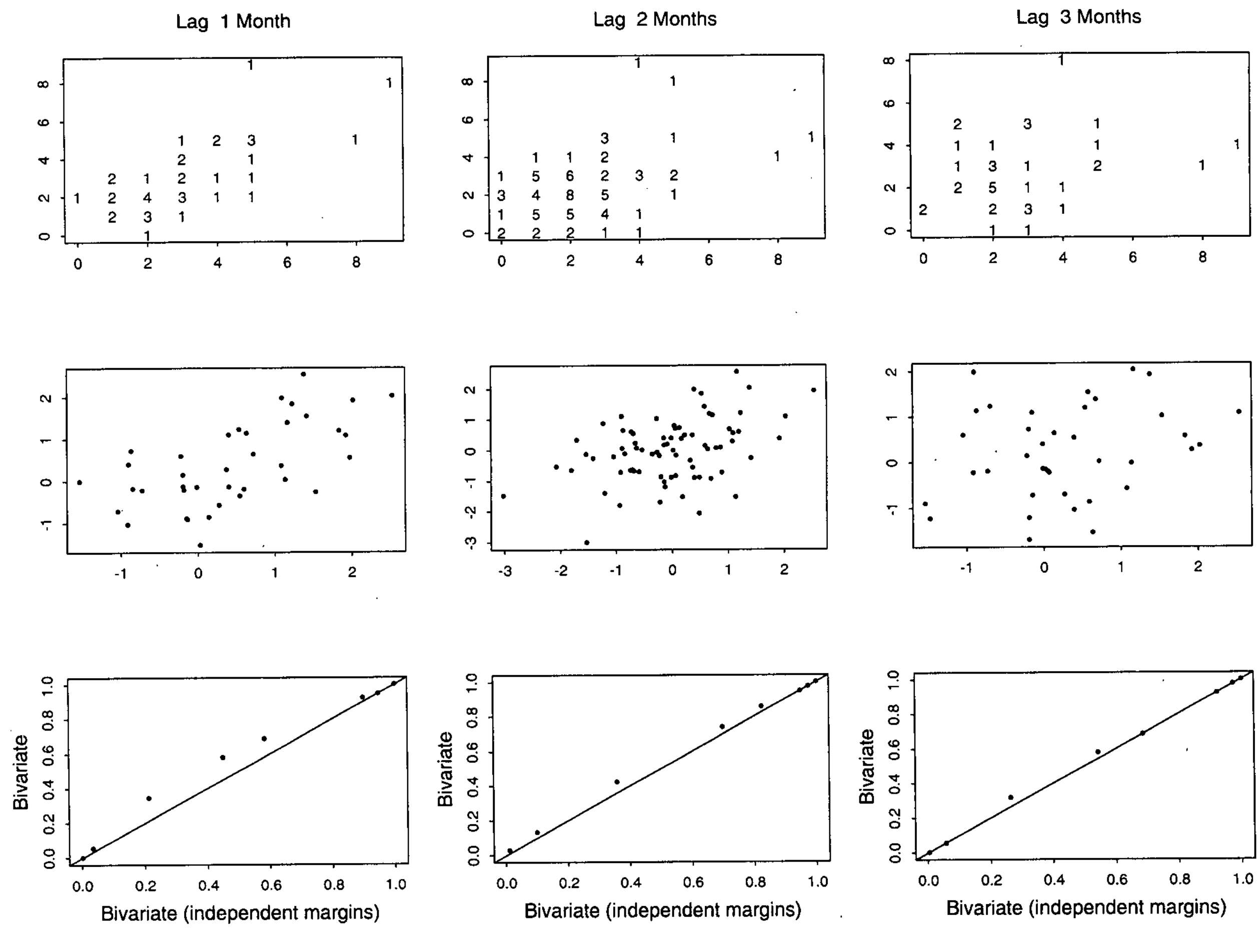


Figure 13.4: *The sunflower, randomized quantile transformation, and diagonal P-P plots for the pairs with lag* 1, 2 *and* 3 *months from the manuscript data.*

strong auto-correlation. As the lag increases, the correlation pattern gradually diminishes.

Now we face the modelling work for this unequally-spaced count time series. Actually, the underlying refereeing process is a continuous-time process $\{X(t); t \geq 0\}$. We may use a simple birth-death process to approximate the refereeing process, where the infinitesimal transition probabilities are supposed to be

$$\begin{aligned}
&\Pr[X(t+h) = x+1 \mid X(t) = x] = \lambda h + o(h), && x \geq 0,\ \lambda > 0,\\
&\Pr[X(t+h) = x-1 \mid X(t) = x] = \mu x h + o(h), && x \geq 1,\ \mu > 0,\\
&\Pr[|X(t+h) - x| > 1 \mid X(t) = x] = o(h), && x \geq 0.
\end{aligned}$$

Here h is small. The arrival process is assumed to have constant intensity λ, while the intensity for the leaving, μx, is assumed to be proportional to the number of current manuscripts in the queue. The later is because of the naive assumption that Prof. Joe would speed up if more manuscripts are accumulated. Such a model has stationary distribution Poisson(λ/μ), and has GAR(1) form representation (with binomial-thinning operation). See Example 8.1, (8.1.7) and relevant discussion in Section 8.2. Therefore, the background information of the underlying refereeing process leads to the following GAR(1) process to be the approximation of reality:

$$X(t_{i+1}) \overset{d}{=} e^{-\mu(t_{i+1}-t_i)} * X(t_i) + E(t_i, t_{i+1}), \qquad i = 1, 2, \ldots, 84. \tag{13.2.1}$$

where $E(t_i, t_{i+1}) \sim$ Poisson $\left(\frac{\lambda}{\mu}\left[1 - e^{-\mu(t_{i+1}-t_i)}\right]\right)$, and the marginal distribution is Poisson(λ/μ).

Next, we turn to estimating the parameters λ and μ. For an estimation approach like MLE, CLS, CWLS2, as well as diagonal PLS, since no explicit forms of estimates, we need initial values of the parameters to find the solutions when minimizing or maximizing the non-linear objective functions. It's better to find a good initial point which is close to the true parameter vector. In this case, we have a simple way to find the initial point. By virtue of the marginal estimating approach (refer to Section 10.4.), we can first estimate $\alpha_0 = e^{-\mu}$ or μ,

$$R_{CLS(ME)}(\alpha_0) = \sum_{i=1}^{84}\left([x_{i+1} - \bar{x}] - \alpha_0^{t_{i+1}-t_i}[x_i - \bar{x}]\right)^2, \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{85} x_i, \quad \alpha_0 \in (0,1).$$

The plot of the function $R_{CLS(ME)}(\alpha_0)$ is shown in Figure 13.5. This method provides the estimate

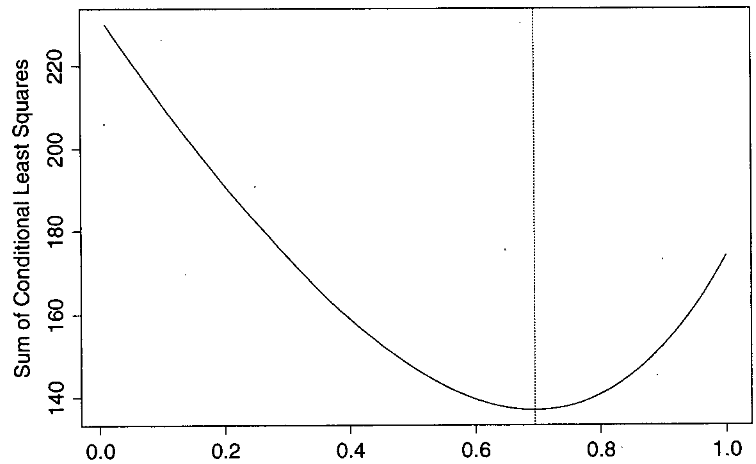


Figure 13.5: *The plot of function* $R_{CLS(ME)}(\alpha_0)$ *of the manuscript data.*

of α_0: 0.695, or equivalently, $-\log 0.695 = 0.384$ for μ. Since the mean of marginal distribution $A(\mu, \lambda)$ equals λ/μ, which we have already estimated as $\bar{x} = 2.412$ in above $R_{CLS(ME)}(\alpha_0)$, thus, the estimate for λ is $2.412 \times 0.384 = 0.878$. This type of estimate is called the CLS(ME) estimate. The point $(\mu, \lambda) = (0.384, 0.878)$ is now the initial point for solving other types of estimates.

With this initial point, we use numerical iterative methods to obtain estimates of the ML, CLS, CWLS2 and diagonal PLS. They are listed in Table 13.3. For the CWLS2 estimate, we choose $c = 1$ and $d = 0.5$ in target function

$$R_{CWLS2} = \sum_{i=1}^{84} \frac{\left([x_{i+1} - \lambda/\mu] - \alpha_0^{t_{i+1}-t_i}[x_i - \lambda/\mu]\right)^2}{cx_i + d}.$$

For the diagonal PLS estimate, we choose those pairs with lag 1 month. From this table, we can roughly see that μ is likely in $(0.4, 0.5)$, λ is likely in $(0.8, 1.4)$, and the marginal mean is likely in the range $(2.0, 3.0)$.

For the MLE and diagonal PLS approach, conditional probabilities are involved. We use

Table 13.3: *Summary of different estimates of parameter μ and λ in the GAR(1) model for the manuscript data.*

Estimation Type	$\widehat{\mu}$	$\widehat{\lambda}$	$\widehat{A(\mu,\lambda)} = \widehat{\lambda}/\widehat{\mu}$
CLS(ME)	0.384	0.878	2.286
MLE	0.448	1.002	2.237
CLS	0.378	0.793	2.098
CWLS2	0.492	1.027	2.087
Diagonal PLS	0.481	1.405	2.921

the following recursion to compute them:

$$\begin{aligned}\Pr[X(t_{i+1}) = x_{i+1} \mid X(t_i) = x_i] &= e^{-\mu(t_{i+1}-t_i)} \cdot \Pr[X(t_{i+1}) = x_{i+1} - 1 \mid X(t_i) = x_i - 1] \\ &\quad + \left(1 - e^{-\mu(t_{i+1}-t_i)}\right) \cdot \Pr[X(t_{i+1}) = x_{i+1} \mid X(t_i) = x_i - 1],\end{aligned}$$

where $x_i = 1, 2, \ldots$ and $x_{i+1} = 0, 1, 2, \ldots$. Hence, with

$$\Pr[X(t_{i+1}) = l \mid X(t_i) = 0] = \Pr[E(t_i, t_{i+1}) = l], \qquad l = 0, 1, \ldots, x_{i+1},$$

we can determine $\Pr[X(t_{i+1}) = x_{i+1} \mid X(t_i) = x_i]$ for any $i > 0$. This recursion formula is also helpful in computing the bivariate cumulative distribution function of the model GAR(1) process, and hence for the diagonal P-P plots in model diagnosis.

We have obtained a few estimates of parameters for the GAR(1) model. A natural question is that how good is the fit of the model plugged in with different estimates for the manuscript data. To diagnose the fitted models, we resort to the diagonal P-P plot. For each estimating method, we will draw the diagonal P-P plot for the pairs of manuscript data with lag $1, 2, \ldots, l$ month (l is an adequate positive integer). If the model fits the data well, then all diagonal P-P plots will display the ideal pattern that all points fluctuate around the diagonal straight line at 45°. Otherwise, the fitted model is not good. To save space, we only draw lag 1 to 3 month plots for the five estimating methods; see Figures 13.6 and 13.7.

From these diagnostic plots, we see that none of the fitted models by five estimating methods is ideal. Each model shows some minor discrepancies The first four methods (except for the diagonal

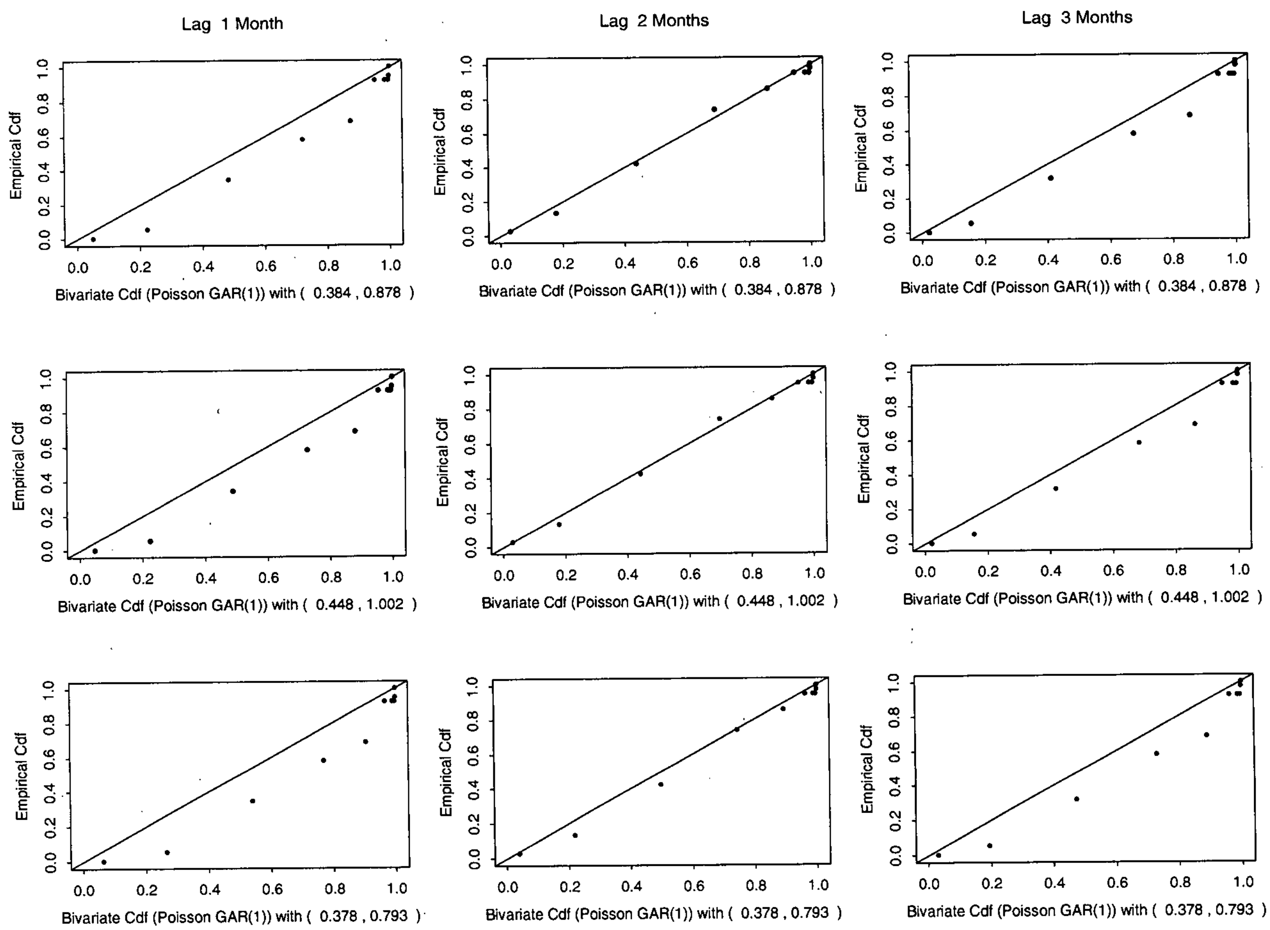


Figure 13.6: *Model diagnosis for manuscript data: diagonal P-P plots for estimates of the CLS(ME) (top row), MLE (middle row) and CLS (bottom row).*

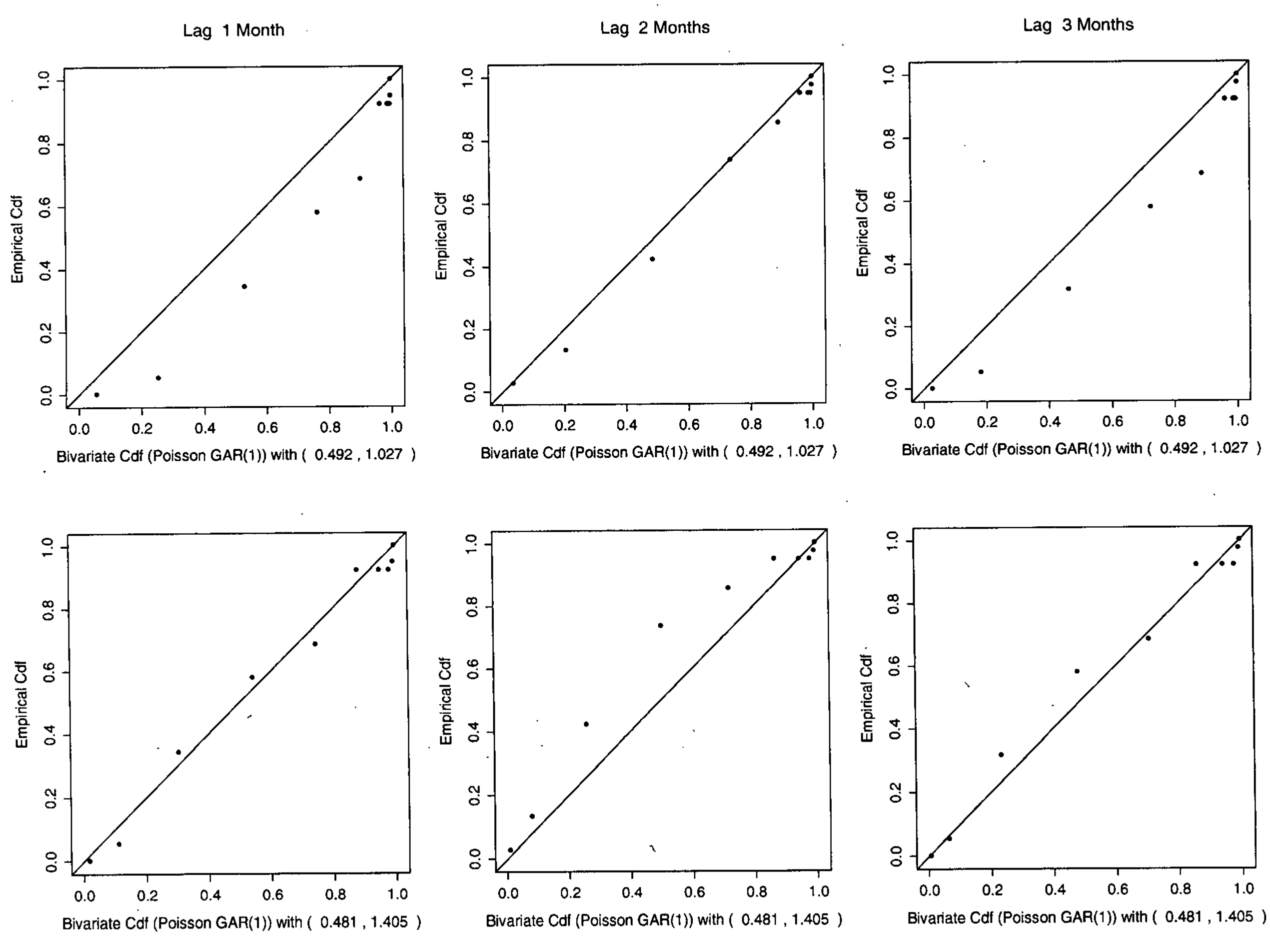


Figure 13.7: *Model diagnosis for manuscript data: diagonal P-P plots for estimates of the CWLS (top row) and diagonal PLS (bottom row).*

PLS method) seem to overestimate the autocorrelation coefficient of lag 1 month. There could be a couple of possible reasons. First, the number of pairs with lag 1 month is small, only 38, which may lead to an inaccurate estimate of bivariate cumulative distribution function. Secondly, the total sample size 85 may not be big enough to obtain more accurate estimates of parameters for the GAR(1) model. Thirdly, the specified GAR(1) model may not approximate reality very well. If this is the case, we have to figure out a better model.

For the possible reason of inaccurate estimates of parameters, we have tried a robust method. It seems not to be helpful in this problem. Thus, we set up grid points for $(\mu, \lambda) \in (0.4, 0.5) \times (0.8, 1.4)$ and find a better one by looking into the diagonal P-P plots. It seems that $(\mu, \lambda) = (0.433, 1.04)$ leads to a better GAR(1) model from the view of diagonal P-P plot. See Figure 13.8. Note that estimates based on a graphical plot are not asymptotically efficient.

This study shows that the diagonal P-P plot is an intuitive and useful graphical tool in diagnosing or building GAR(1) process models.

13.3 WCB claims data study

The WCB claims data was originally studied by Freeland [1998] in his Ph.D thesis. Dr. Freeland applied the Poisson GAR(1) model, i.e., the discrete-time version of the model in Section 13.2, to the data and made predictions based on the fitted models. These data are given in Appendix A.2.

The data record the monthly claim number of workers who got injured during work time and collect the short-term disability benefit (STWLB) from the Workers' Compensation Board (WCB) of the province of British Columbia, Canada. These data are reported from one city center, the Richmond claims center of WCB from the years 1985 to 1994. According to the industry category and injury type, the claim counts are classified into six time series: C0 (heavy manufacturing, burn related injury), C1 to C5 (logging corresponding to five different types of injury). Other series (C1a, C1A to C5A) in the dataset are relevant information for C1 to C5. It is well known that in British Columbia, logging is a seasonal industry while heavy manufacturing is not.

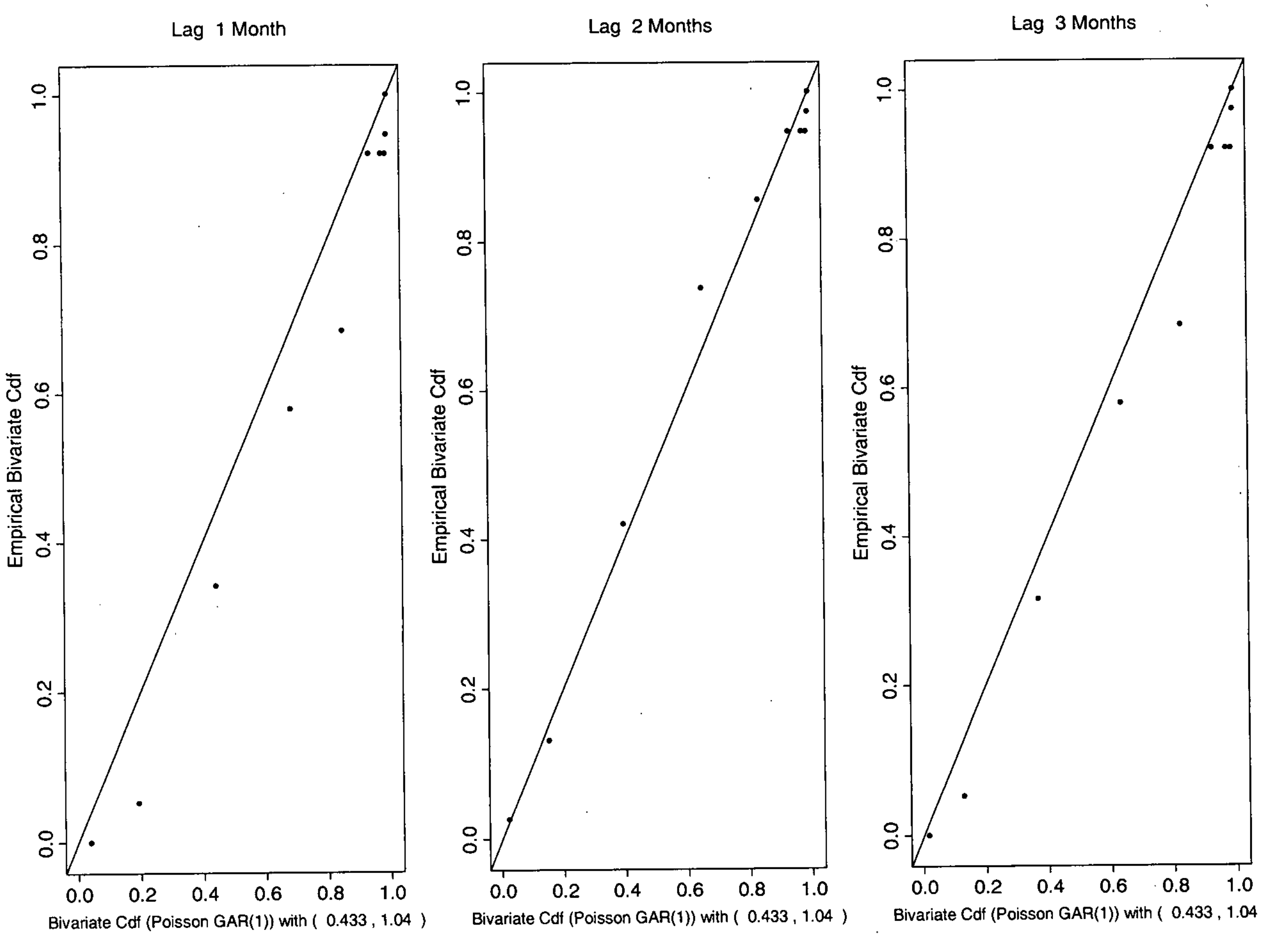


Figure 13.8: *Model diagnosis for manuscript data: diagonal P-P plots for the estimate* $(\mu, \lambda) = (0.433, 1.04)$.

Unlike the manuscript data in previous section, all series in this data set are equally-spaced. This feature brings convenience for data analysis. Freeland [1998] applied the stationary Poisson GAR(1) model for the series C0, and non-stationary Poisson GAR(1) models for C1 to C5 with a seasonal covariate adjustment. However, only C3 was modelled by the Poisson GAR(1) model with Poisson innovations whose means depend on the season. We investigate the seasonality of C3 again. We think that the seasonality might not be large enough to cause the dispersion of the marginal distribution for the series C3. Hence, we want to try other stationary GAR(1) models which have negative binomial or generalized Poisson marginal distributions for the series C3. We do so to see whether these new stationary GAR(1) models are good enough for the data. If they are adequate, they are simpler than non-stationary models.

As pointed out by Freeland [1998], the claim counts of each month can be decomposed as two parts: one is from the claimants from previous month, one is from arrival of new claimants. Both parts are random, and seem to be independent. It is reasonable to think that the number of continuing claimants depends on the number of claimants in the previous month. Thus, this leads to the type of Galton-Watson process with immigration (see Nanthi [1983], p. 180-181 for the definition) as the model, with a branching term and an immigration term. For each claimant, we can make a simple assumption that this person continues to collect the STWLB in next month in probability α. Hence, the resulting model based on this simplification is the GAR(1) model with the binomial thinning operation, namely

$$X(t_{i+1}) \stackrel{d}{=} \alpha * X(t_i) + E_i, \qquad i = 1, \dots, n. \tag{13.3.1}$$

This is the model of generalized AR(1) time series. If E_i is distributed as Poisson, then it leads to Poisson GAR(1) model as in Section 13.2. Freeland [1998] had chosen the Poisson GAR(1) model with constant α for all series C0, C1 to C5, and Poisson innovation E_i whose means are exponential with sinusoid exponents, to account for the influence of season for C1 to C5.

Now we investigate the series C3. The summary statistics regarding this series are given in Table 13.4. The histogram of C3 is plotted in (a) of Figure 13.9. From the histogram, we see it is skewed. However, the variance of 11.8 is quite a bit larger than the mean of 6.13, thus, leading to a big coefficient of variation of 1.92. Hence, the Poisson may not be appropriate for the marginal

Table 13.4: *Summary of the series C3 in WCB claims data.*

Sample size	Minimum	Maximum	Mean	Variance	D
120	1	21	6.13	11.8	1.92

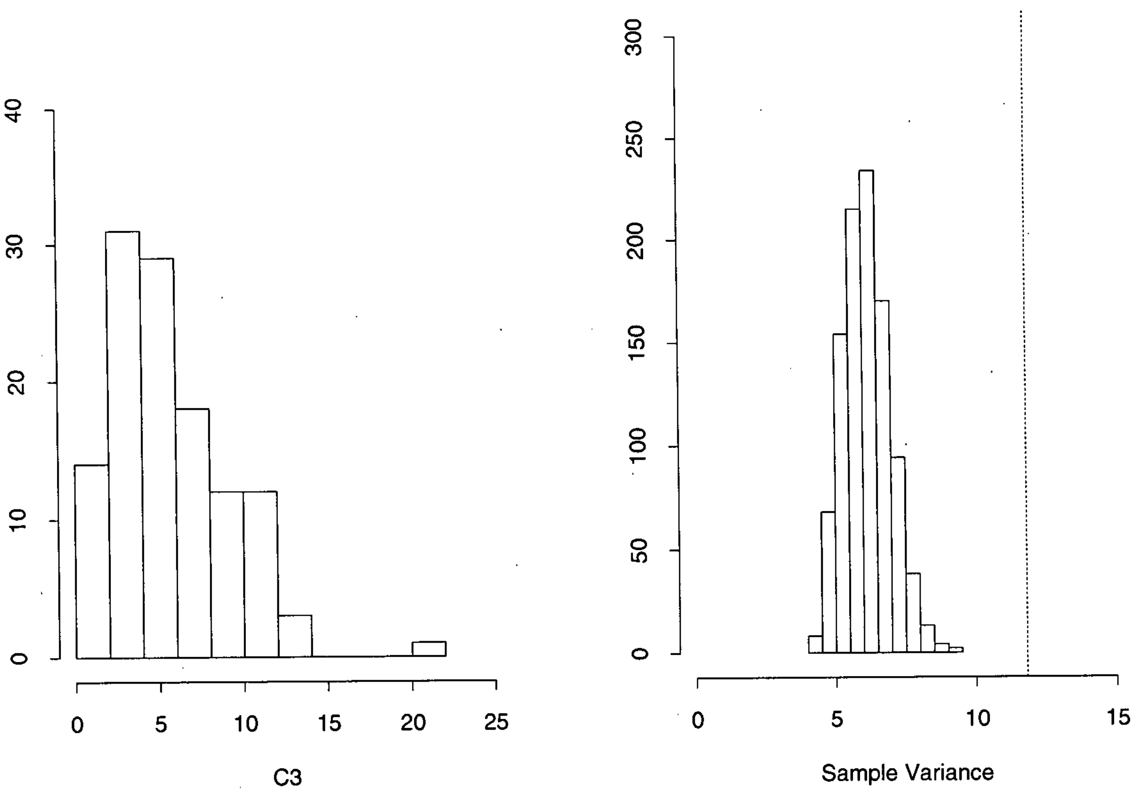


Figure 13.9: *The histogram of the series C3 (left), and* 1000 *simulated sample variances from Poisson*(6.13)*; the dotted vertical line is the sample variance of C3 (right).*

distribution. This is verified by (b) of Figure 13.9, where the variance of C3, 11.8, is far away from the simulated sample variances of Poisson(6.13) with sample size 120. Such a dispersion index leads us to consider the negative binomial and generalized Poisson distributions. Next we treat the series C3 to be univariate data, and try to fit them by the NB(β,γ) and GP(θ,η) distributions. For the NB(β,γ) distribution, the mean A and variance V are

$$A = \beta\gamma(1-\gamma)^{-1}, \qquad V = \beta\gamma(1-\gamma)^{-2}, \tag{13.3.2}$$

and for the GP(θ,η) distribution, they are

$$A = \theta(1-\eta)^{-1}, \qquad V = \theta(1-\eta)^{-3}. \tag{13.3.3}$$

By the method of moments, we obtain their estimates

$$\widehat{\beta} = 6.64, \quad \widehat{\gamma} = 0.48; \qquad \widehat{\theta} = 4.42, \quad \widehat{\eta} = 0.28,$$

as well as the P-P plots of the series C3 against NB(6.64, 0.48) and GP(4.42, 0.28). See Figure 13.10. These P-P plots show that the Poisson distribution is not suitable, but that the negative binomial and generalized Poisson distributions are fairly good univariate fits.

Next we turn to check the autocorrelation in series C3. In Figure 13.11, (a) shows the time series plot and (b) shows the ACF plot. From the ACF plot, the geometrical decrease is very obvious, and indicates that serial dependence exists. The serial dependence is also detected by sunflower plots, randomized quantile transformation plots and diagonal P-P plots. For the sake of space, we only show them for lag 1 month to lag 3 months; see Figure 13.12. The ACF plot shows that there exists seasonality in the series C3, as pointed out by Freeland [1998]. There is a yearly period, but the yearly dependence is not strong. Hence, using the stationary model to approximate non-stationary reality may work in this case. Thus, we finally decide to try two GAR(1) models with NB(β,γ) and GP(θ,η) margins respectively. This means that we choose two kinds of innovations in (13.3.1) which do not have an explicit form for the pmf.

Because each of the two models has three parameters, the CLS approach is not suitable, and thus, we try the CGLS approach. By (10.2.19), this method leads to

$$a = -0.023, \qquad b = 0.929, \qquad c = 1.587.$$

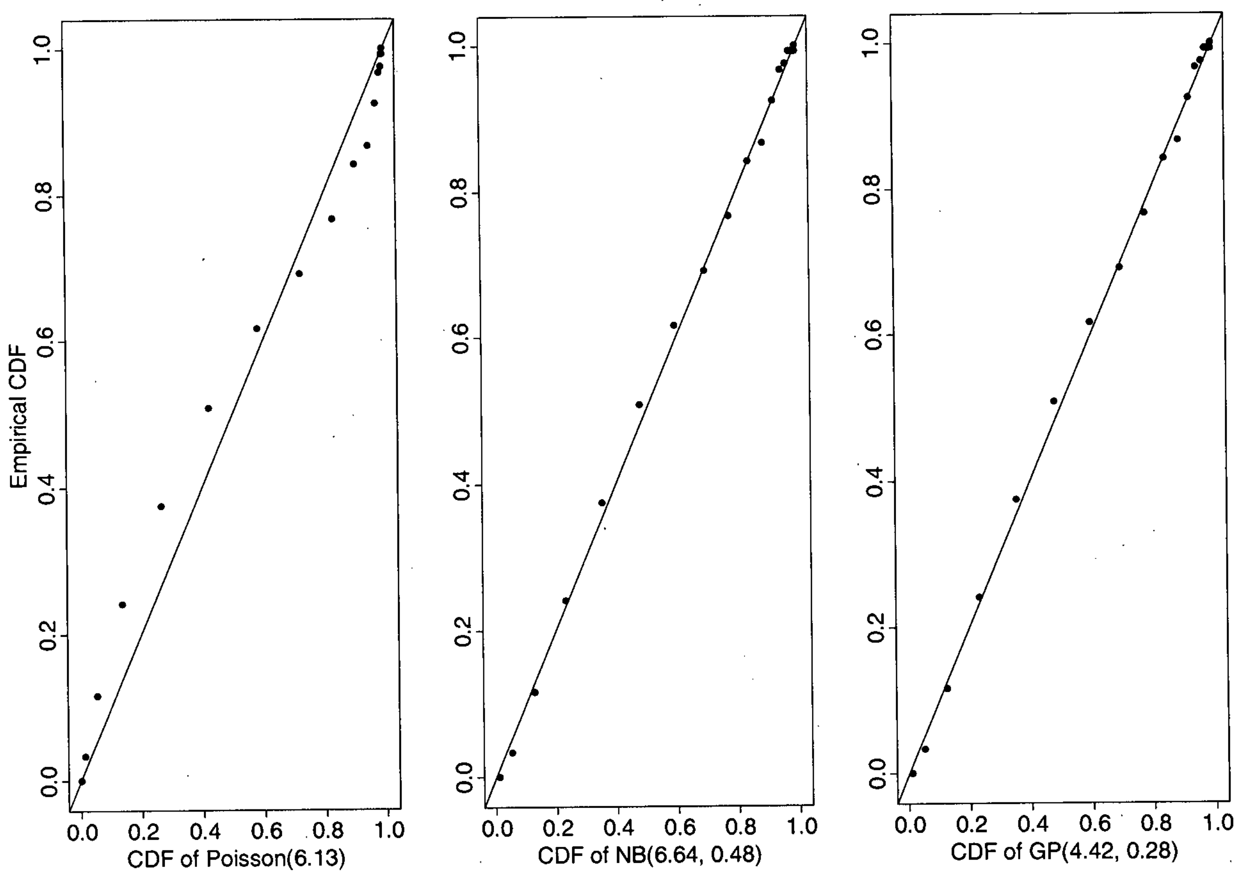


Figure 13.10: *The P-P plots of C3 in WCB claims data against Poisson*(6.13) *(left), NB*(6.64, 0.48) *(middle) and GP*(4.42, 0.28) *(right).*

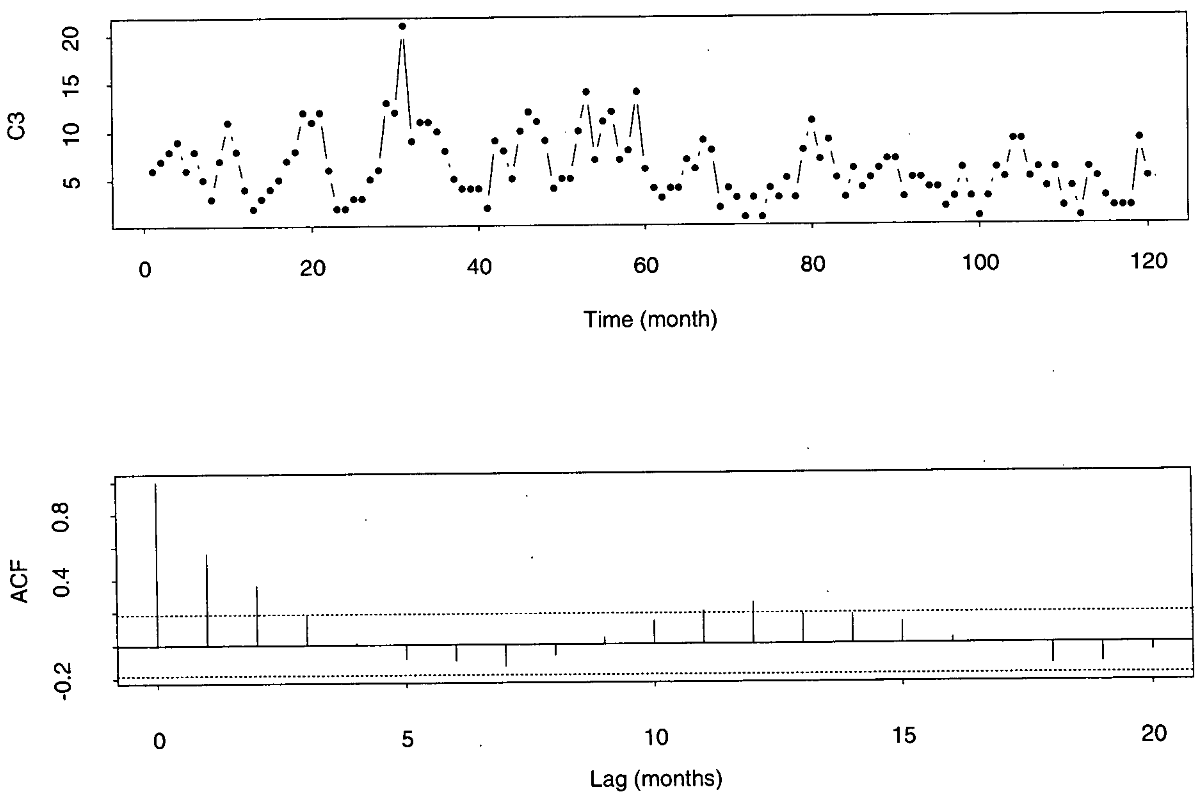


Figure 13.11: *The time series plot (top) and ACF plot (bottom) of C3 in WCB claims data.*

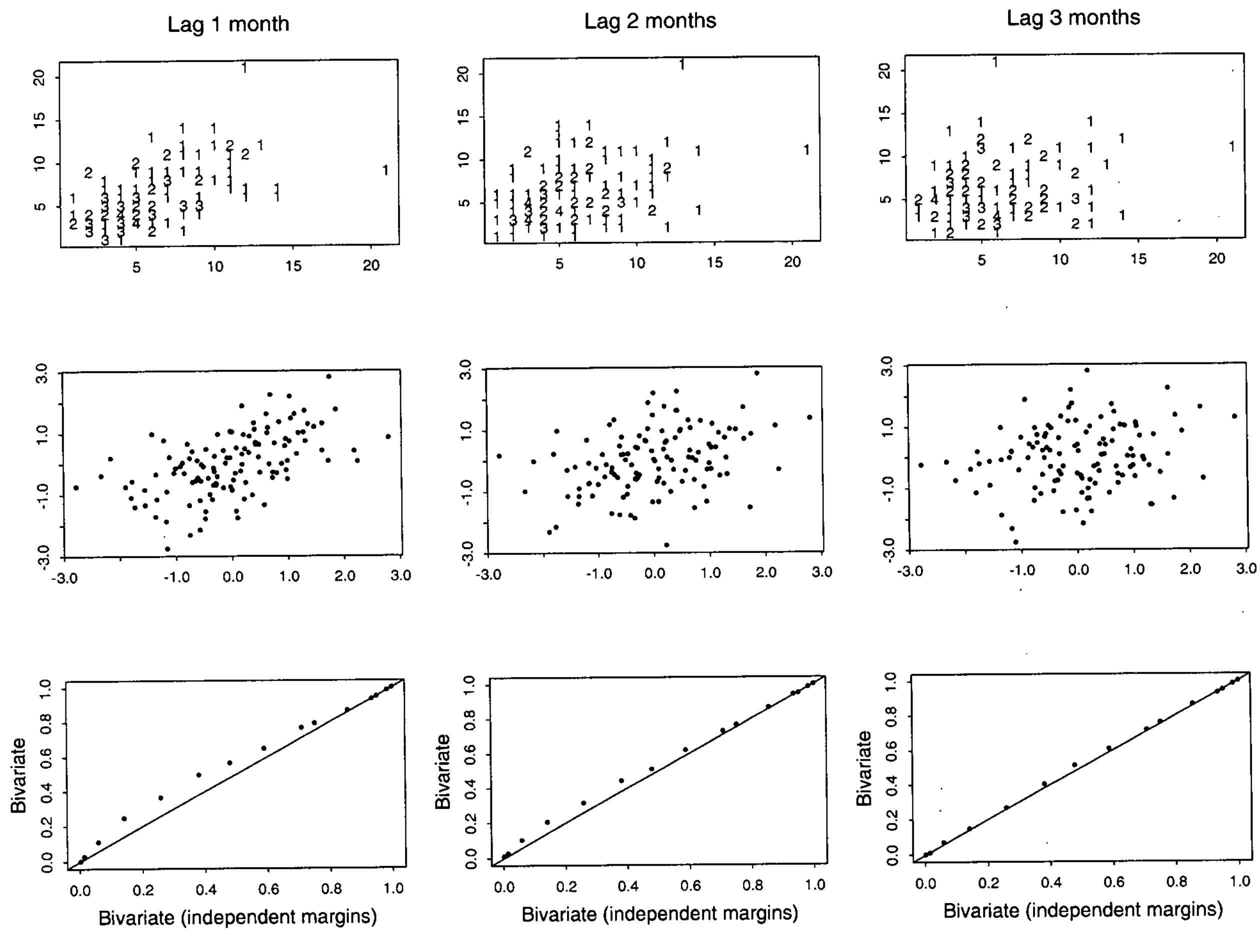


Figure 13.12: *Serial dependence: sunflower plots (1st row), randomized quantile transformation plots (2nd row) and diagonal P-P plots (3rd row) of pairs in C3 with lag 1, 2 and 3 months.*

However, it is impossible to estimate the positive parameter α $(= \sqrt{a})$ by a negative number. Hence, the CGLS approach fails for this data set. We then try the method of moments approach. According to (10.4.3),

$$R_1 = \frac{1}{120}\sum_{i=1}^{120} x_i = 6.13, \quad R_2 = \frac{1}{120}\sum_{i=1}^{120} x_i^2 = 49.32, \quad R_{12} = \frac{1}{119}\sum_{i=1}^{119} x_i x_{i+1} = 44.27,$$

which lead to estimates of the marginal mean A, marginal variance V and α of

$$\widehat{A}_M = R_1 = 6.13, \qquad \widehat{V}_M = R_2 - R_1^2 = 11.7, \qquad \widehat{\alpha}_M = \frac{R_{12} - R_1^2}{R_2 - R_1^2} = 0.57.$$

By (13.3.2) and (13.3.3), we obtain

$$\widehat{\beta}_M = \widehat{A}_M^2/(\widehat{V}_M - \widehat{A}_M) = 6.76, \qquad \widehat{\gamma}_M = 1 - \widehat{A}_M/\widehat{V}_M = 0.48;$$
$$\widehat{\theta}_M = \sqrt{\widehat{A}_M^3/\widehat{V}_M} = 4.44, \qquad \widehat{\eta}_M = 1 - \sqrt{\widehat{A}_M/\widehat{V}_M} = 0.28.$$

For the fitted GAR(1) models, we need to know the bivariate cdf so that we can draw diagonal P-P plot for model diagnosis. Suppose $t < t'$. Then,

$$\begin{aligned}\Pr[X(t) \le x, X(t') \le y] &= \sum_{i \le x, j \le y} \Pr[X(t) = i, X(t') = j] \\ &= \sum_{i=0}^{x}\left(\sum_{j=0}^{y} \Pr[X(t') = j \mid X(t) = i]\right)\Pr[X(t) = i].\end{aligned}$$

In Section 13.2, we have noticed that the pmf of innovation $E(t,t')$ will determine all conditional probabilities $\Pr[X(t') = j \mid X(t) = i]$ $(i, j \ge 0)$. Can we find the pmf of innovation by the marginal distribution for the GAR(1) process with binomial thinning operation? The answer is yes. To obtain the pmf of $E(t,t')$, we will take advantage of the stochastic representation for the GAR(1) process:

$$X(t') \stackrel{d}{=} \alpha * X(t) + E(t,t').$$

By this representation, we have

$$\Pr[X(t') = 0] = \Pr[\alpha * X(t) + E(t,t') = 0] = \Pr[\alpha * X(t) = 0] \cdot \Pr[E(t,t') = 0],$$

$$
\begin{aligned}
\Pr[X(t')=j] &= \Pr[\alpha * X(t) + E(t,t') = j] \\
&= \Pr[\alpha * X(t) = 0] \cdot \Pr[E(t,t') = j] + \sum_{l=1}^{j} \Pr[\alpha * X(t) = l] \cdot \Pr[E(t,t') = j-l],
\end{aligned}
$$

where $j > 0$. Thus,

$$
\begin{aligned}
\Pr[E(t,t')=0] &= \frac{\Pr[X(t')=0]}{\Pr[\alpha * X(t)=0]}, \\
\Pr[E(t,t')=j] &= \frac{\Pr[X(t')=j] - \sum_{l=1}^{j} \Pr[\alpha * X(t)=l] \cdot \Pr[E(t,t')=j-l]}{\Pr[\alpha * X(t)=0]},
\end{aligned}
$$

for $j = 1, 2, \ldots$. Because the stochastic operation is binomial thinning, it is easy to find that

$$
\begin{aligned}
\Pr[\alpha * X(t) = 0] &= \Pr[X(t)=0] + \sum_{k=1}^{\infty} (1-\alpha)^k \Pr[X(t)=k], \\
\Pr[\alpha * X(t) = l] &= \sum_{k=l}^{\infty} \binom{k}{l} \alpha^l (1-\alpha)^{k-l} \Pr[X(t)=k], \qquad l > 0.
\end{aligned}
$$

Figures 13.13 and 13.14 show the diagonal P-P plots of the series C3 against the fitted NB GAR(1) and GP GAR(1) models. They are shown for lags 1 month to 6 months. From these plots, it seems that both models fit the data well. Compared with the non-stationary GAR(1) model in Freeland [1998], they have a simpler structure.

Since the previous numerical method of probability calculation allows us to compute the conditional probability $\Pr[X(t') = j \mid X(t) = i]$ $(i, j \geq 0)$, we can also try the MLE method with the initial value being the estimates from the method of moments approach. This leads to

$$
\widehat{\alpha}_{MLE} = 0.50, \qquad \widehat{\beta}_{MLE} = 8.82, \qquad \widehat{\gamma}_{MLE} = 0.41
$$

for the NB GAR(1) model, and

$$
\widehat{\alpha}_{MLE} = 0.50, \qquad \widehat{\theta}_{MLE} = 4.71, \qquad \widehat{\eta}_{MLE} = 0.23
$$

for the GP GAR(1) model. Their diagonal P-P plots are shown in Figures 13.15 and 13.16. These plots are roughly ok, but show slight discrepancies with fitted models.

Now we have two kinds of fitted models at hand: NB GAR(1) model and GP GAR(1). Which one is better? We compare them with the AIC, and follow the convention to check two

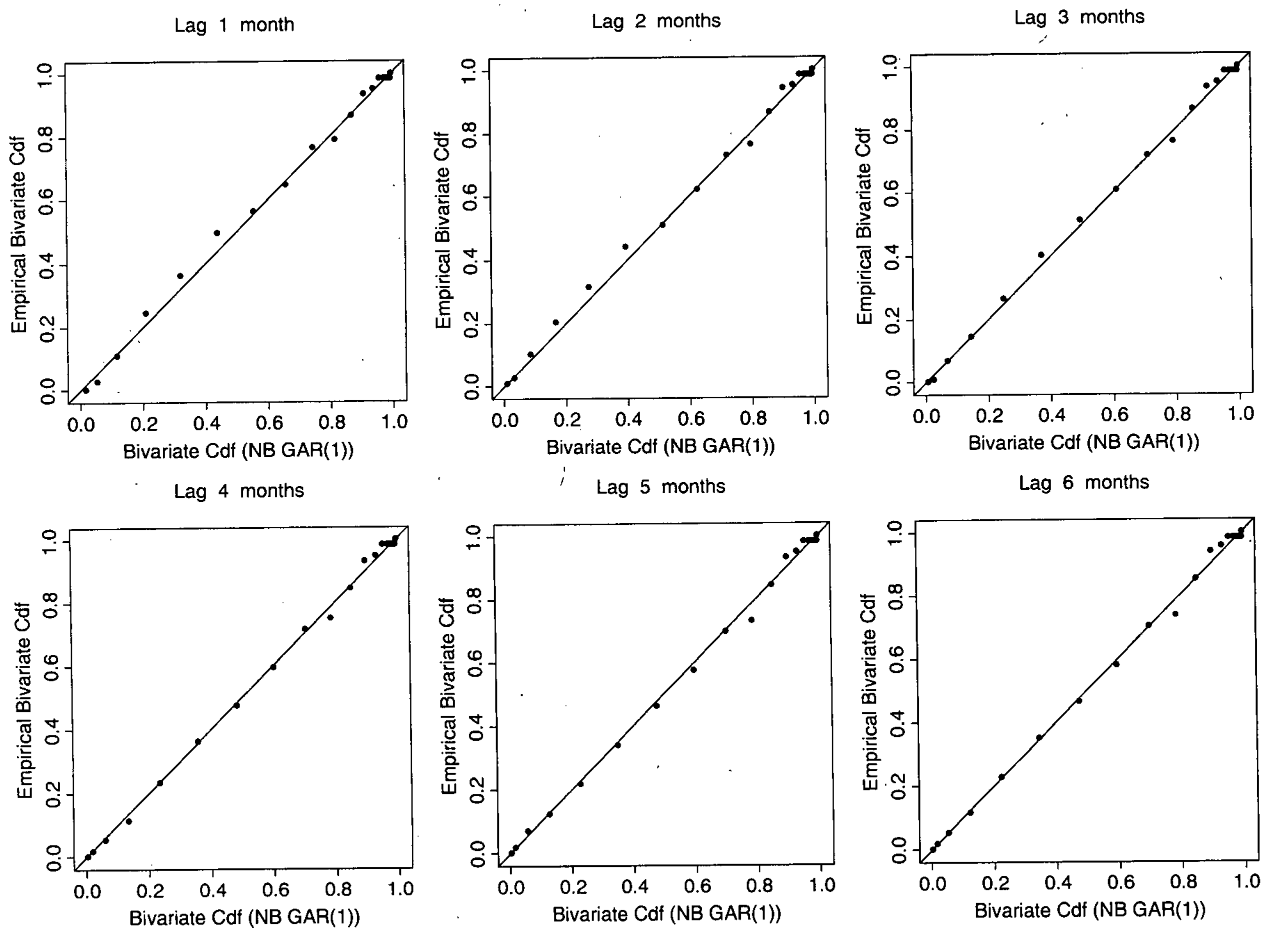


Figure 13.13: *Model diagnosis for WCB claims data: diagonal P-P plots of C3 against the fitted (method of moments) NB GAR*(1) *model.*

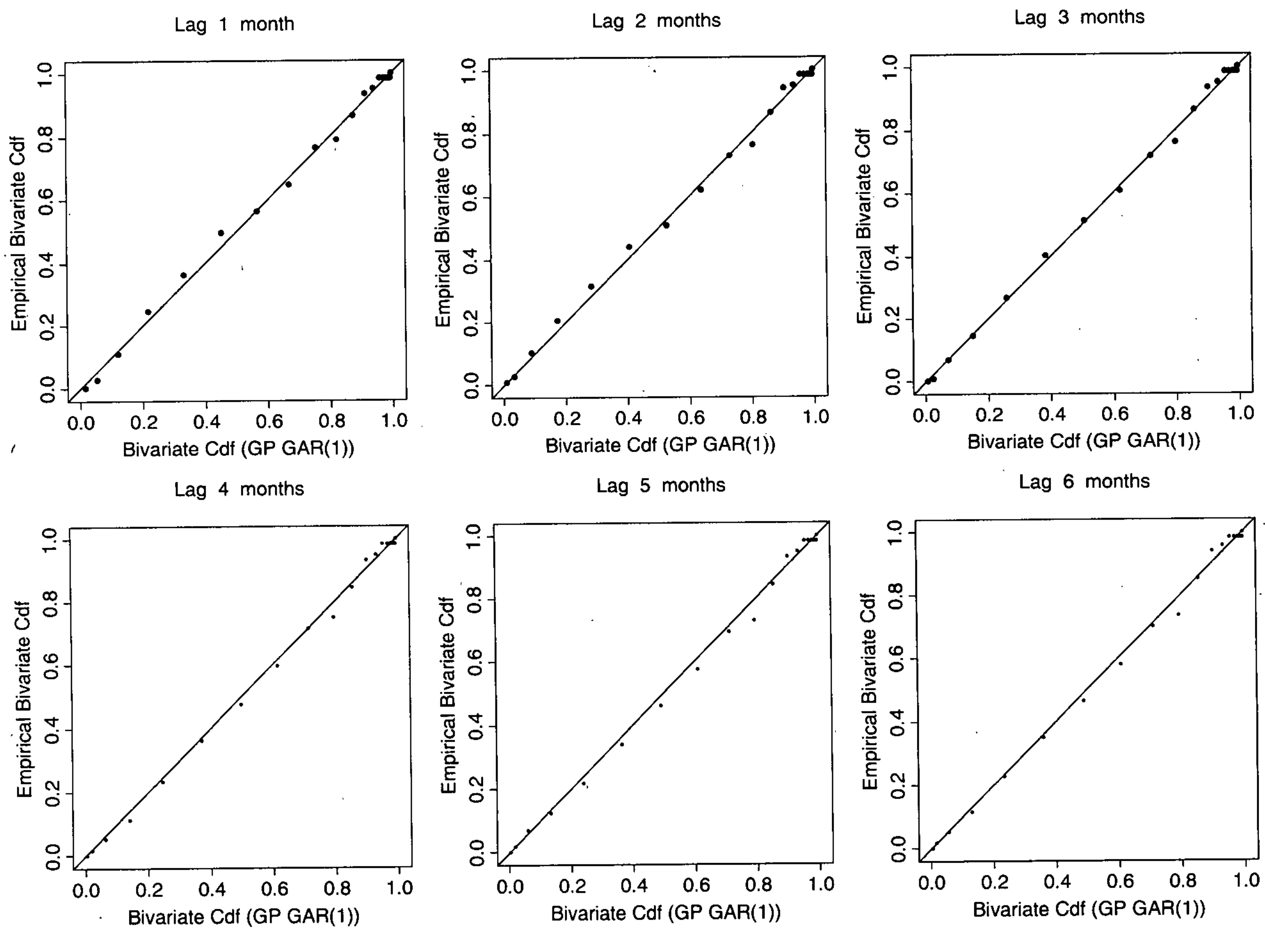


Figure 13.14: *Model diagnosis for WCB claims data: diagonal P-P plots of C3 against the fitted (method of moments) GP GAR*(1) *model.*

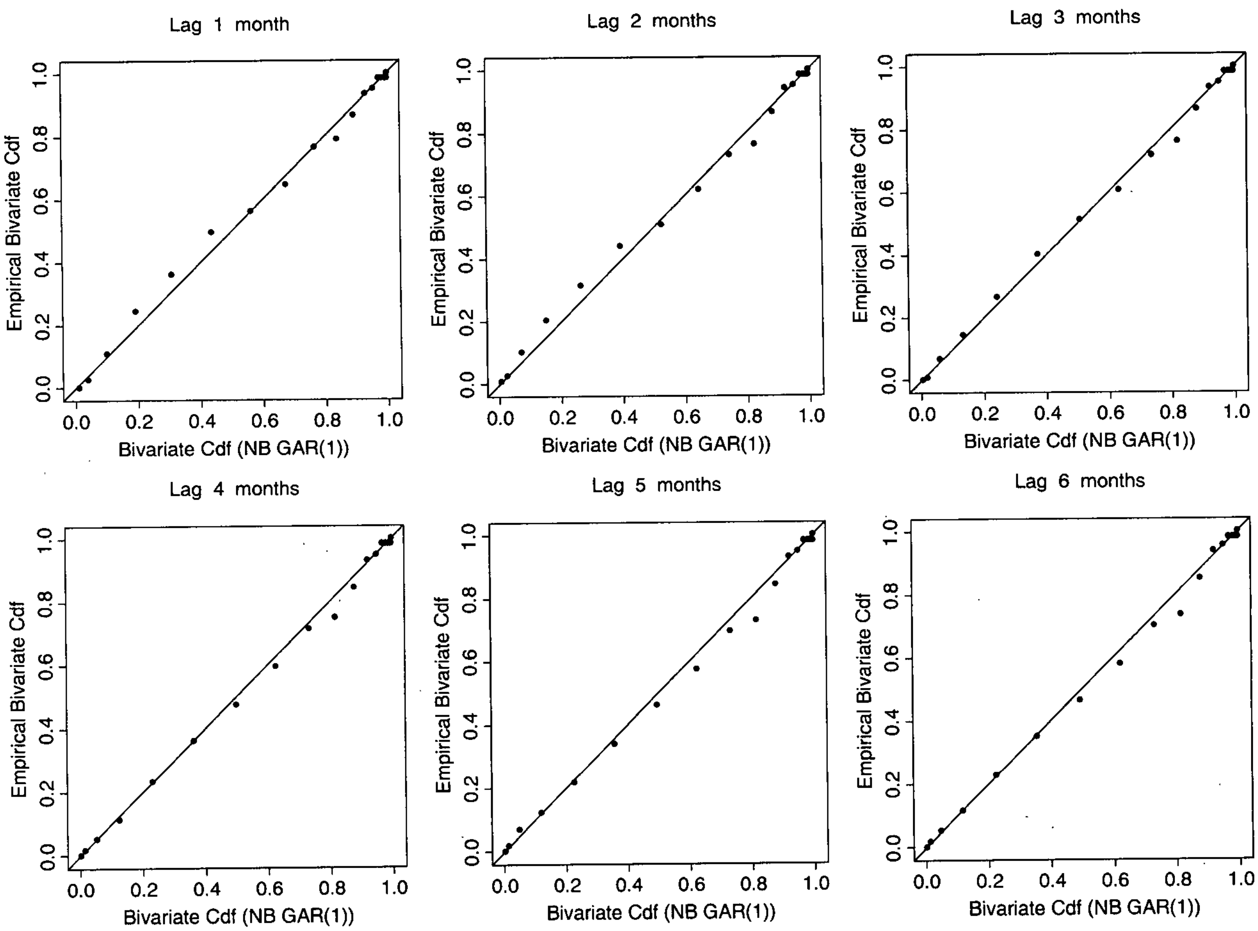


Figure 13.15: *Model diagnosis for WCB claims data: diagonal P-P plots of C3 against the fitted (MLE) NB GAR*(1) *model.*

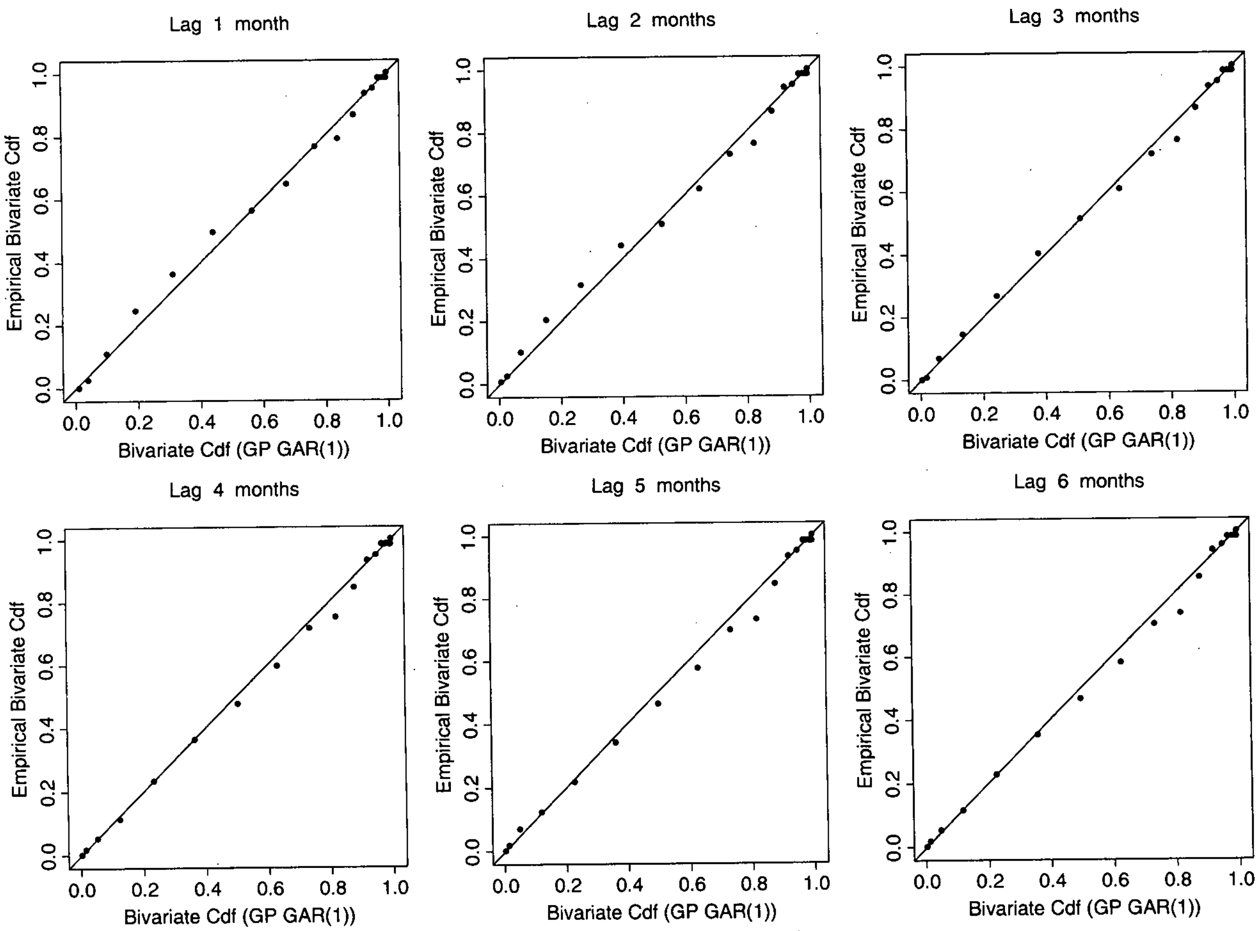


Figure 13.16: *Model diagnosis for WCB claims data: diagonal P-P plots of C3 against the fitted (MLE) GP GAR*(1) *model.*

fitted models with MLE. We obtain

$$\text{AIC for NB GAR(1) model} = 576.6, \qquad \text{AIC for GP GAR(1) model} = 578.4.$$

These AIC values are only slightly different. Thus, fairly speaking, the two models are equally good. It may depend on user's preference or other considerations to choose one of them.

Forecasting is one of the concerns in this study. We now give a brief discussion based on the fitted GP GAR(1) model by method of moments approach. We want to forecast the number of claimants in the next month conditioned on the current month for the year 1994. Why do we just make one month prediction? This is because that the autocorrelation is geometrically decreasing which implies short memory. Hence, a fairly long time prediction may not be useful.

The year 1994 has observations:

$$6, \quad 2, \quad 4, \quad 1, \quad 6, \quad 5, \quad 3, \quad 2, \quad 2, \quad 2, \quad 9, \quad 5.$$

The corresponding preceding observations are

$$4, \quad 6, \quad 2, \quad 4, \quad 1, \quad 6, \quad 5, \quad 3, \quad 2, \quad 2, \quad 2, \quad 9.$$

We calculate the conditional pmf and cdf of $X(t')$ given $X(t) = x$, where the time difference $t' - t$ equals one month. They are given in Tables 13.5 and 13.6 respectively. Based on these two tables, we make a one month prediction of the number of claimants by the approach of conditional mode, conditional median and 50% conditional prediction interval (PI). The 50% conditional prediction interval may not be exactly 50% because of discreteness. In fact, we cut both sides with probability being less or equal to 25%. Thus, the real probability to construct the prediction intervals is bigger than or equal to 50%. The predicted results are given in Table 13.7. For the conditional mode predictions, the absolute error is

$$|6-3|+|2-4|+|4-1|+|1-3|+|6-1|+|5-4|+|3-4|+|2-2|+|2-1|+|2-1|+|9-1|+|5-6| = 28,$$

while for the conditional median predictions, the absolute error is

$$|6-4|+|2-5|+|4-3|+|1-4|+|6-2|+|5-5|+|3-4|+|2-3|+|2-3|+|2-3|+|9-3|+|5-7| = 25.$$

Table 13.5: *Estimated conditional probabilities:* $\Pr[X(t') = y \mid X(t) = x]$. *The highest probability in each column is highlighted with an asterisk.*

	x=1	x=2	x=3	x=4	x=5	x=6	x=9
y=0	0.203	0.087	0.037	0.016	0.007	0.003	0.000
y=1	0.204*	0.203*	0.137	0.080	0.044	0.023	0.003
y=2	0.171	0.190	0.197*	0.163	0.116	0.075	0.014
y=3	0.132	0.154	0.175	0.188*	0.174	0.141	0.045
y=4	0.096	0.116	0.138	0.159	0.175*	0.174*	0.094
y=5	0.067	0.083	0.102	0.122	0.143	0.161	0.141
y=6	0.045	0.057	0.072	0.089	0.108	0.128	0.160*
y=7	0.030	0.039	0.049	0.062	0.078	0.095	0.148
y=8	0.020	0.025	0.033	0.042	0.054	0.067	0.119
y=9	0.013	0.017	0.022	0.028	0.036	0.046	0.088

Table 13.6: *Estimated conditional cdf:* $\Pr[X(t') \leq y \mid X(t) = x]$. *The median in each column is highlighted with an asterisk.*

	x=1	x=2	x=3	x=4	x=5	x=6	x=9
y=0	0.203	0.087	0.037	0.016	0.007	0.003	0.000
y=1	0.406	0.290	0.175	0.096	0.051	0.026	0.003
y=2	0.578*	0.480	0.372	0.259	0.166	0.100	0.017
y=3	0.709	0.634*	0.546*	0.447	0.340	0.241	0.062
y=4	0.805	0.751	0.684	0.606*	0.515*	0.415	0.156
y=5	0.872	0.834	0.786	0.728	0.658	0.577*	0.297
y=6	0.917	0.891	0.858	0.817	0.766	0.705	0.457
y=7	0.947	0.930	0.908	0.880	0.844	0.800	0.605*
y=8	0.966	0.955	0.941	0.922	0.898	0.867	0.724
y=9	0.979	0.972	0.962	0.950	0.934	0.913	0.813

Table 13.7: *One month predictions:* $\hat{y}_{mode}$, $\hat{y}_{median}$ and $\hat{y}_{PI}$.

x	4	6	2	4	1	6	5	3	2	2	2	9
$\hat{y}_{mode}$	3	4	1	3	1	4	4	2	1	1	1	6
$\hat{y}_{median}$	4	5	3	4	2	5	4	3	3	3	3	7
$\hat{y}_{PI}$	[2,5]	[4,6]	[1,4]	[2,5]	[1,3]	[4,6]	[3,5]	[2,4]	[1,4]	[1,4]	[1,4]	[5,8]
y	6	2	4	1	6	5	3	2	2	2	9	5

Table 13.8: *Summary of the series of daily maximum ozone concentration.*

Sample size	Minimum	Maximum	Mean	Variance	Variance/Mean
110	19.5	106.7	48.1	288.5	6.0

For the conditional prediction intervals, we count the number of intervals which contain the real observations. This number is 7, hence the successful prediction rate is $7/12 = 58.3\%$, close to the actual probability we use to construct these prediction intervals.

13.4 Ozone data study

In this section, we study a positive-valued time series from a project on tropospheric ozone forecasting in the Lower Fraser Valley, British Columbia, Canada.

These data are daily maximum ozone concentrations (thus, positive-valued) collected at the Abbortsford ozone station in the summer of 1985 from May 1 to August 18 inclusively, and can be considered roughly stationary in this interval. See Appendix A.3. They are just part of a large data set in this environmental study.

The summary statistics of these ozone data are presented in Table 13.8. The ratio of the variance to the mean, 6.0, is large, which suggests a marginal distribution with large dispersion. In addition, it is expected to see the skewed pattern of the distribution of daily maximum ozone concentration because they are maxima. The histogram verifies the skewness; see the left subplot in Figure 13.17. The Gamma distribution is often applied in modelling skewed positive data, and can have large dispersion. Thus, we fit the margins of the process by Gamma(δ, β) with

$$f_X(x; \delta, \beta) = \frac{\beta^\delta}{\Gamma(\delta)} x^{\delta-1} e^{-\beta x}, \qquad x,\ \delta,\ \beta > 0;$$

$$\phi_X(s) = \left(\frac{\beta}{\beta + s}\right)^\delta, \qquad \mathbf{E}(X) = \delta\beta^{-1}, \qquad \mathbf{Var}(X) = \delta\beta^{-2}.$$

By the method of moments, we can obtain the estimates

$$\widehat{\delta} = 8.03, \qquad \widehat{\beta} = 0.17.$$

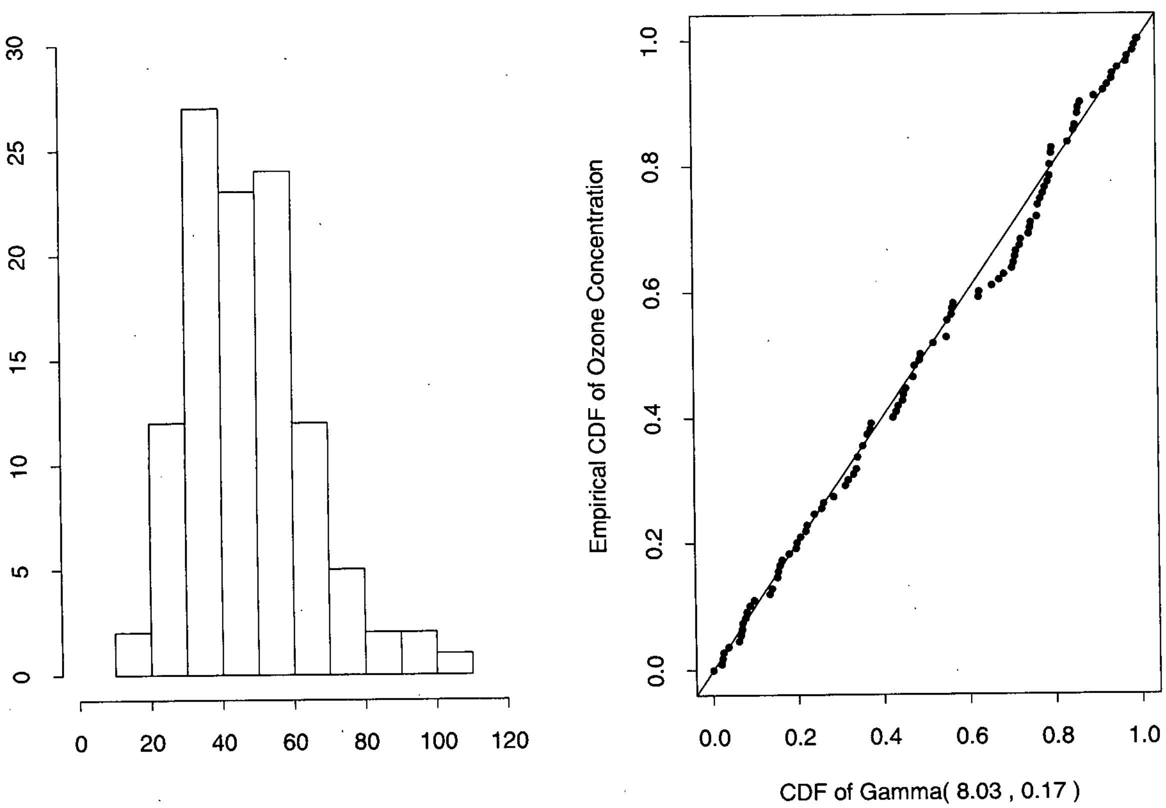
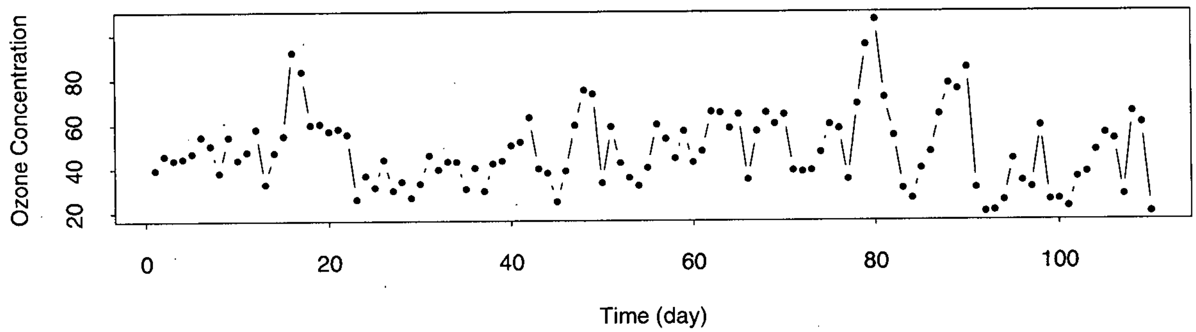


Figure 13.17: *The histogram of the daily maximum ozone concentration, and the P-P plot against* Gamma(8.03, 0.17).



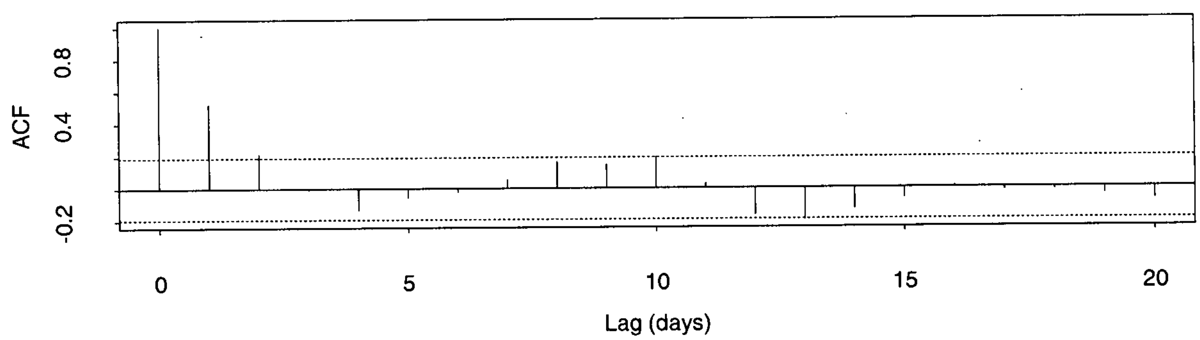


Figure 13.18: *The time series plot and ACF plot for the daily maximum ozone concentration.*

This Gamma(8.03, 0.17) distribution fits the data well. See the right subplot in Figure 13.17.

Our next concern is whether there exists serial dependence in the series of daily maximum ozone concentration. The time series plot and ACF plot are shown in Figure 13.18. The ACF plot shows an obvious pattern of geometrical decrease. The serial dependence is also confirmed by the scatterplot and diagonal P-P plot with lagged days. We show them for lag one day to lag three days in Figure 13.19.

Why does there exist serial dependence? Is there any scientific explanation for such a phenomena? To this end, we study the mechanism of formation and decomposition of ozone in

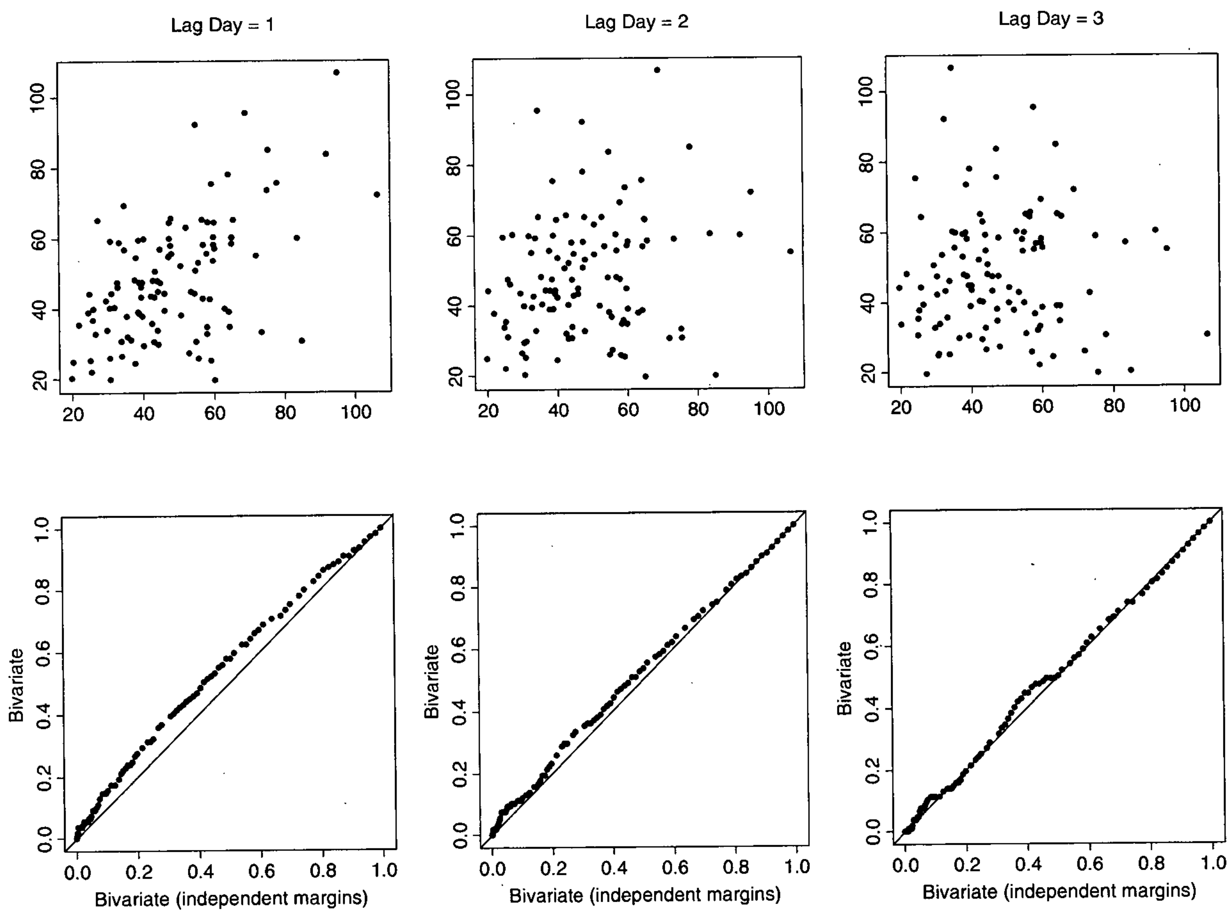


Figure 13.19: *The scatterplots and diagonal P-P plots of lag one day to three days for the daily maximum ozone concentration.*

troposphere. Nitrogen dioxide (NO_2) is photodissociated by solar radiation to be nitric oxide (NO) and ground state oxygen atoms, $O(^3P)$:

$$NO_2 \; + \; h \cdot \; \nu \longrightarrow \; NO \; + \; O(^3P),$$

where $h \cdot \nu$, the product of Planck's constant h and the frequency ν of the electromagnetic wave of solar radiation, presents the energy from solar radiation. Then oxygen atoms combine with molecular oxygen to form ozone:

$$O(^3P) \; + \; O_2 \; + \; \xrightarrow{M} \; O_3 \; + \; M.$$

Ozone will be photodissociated by near-ultraviolet solar radiation to form an excited oxygen atom, $O(^1D)$:

$$O_3 \; + \; h \cdot \; \nu \longrightarrow \; O_2 \; + \; O(^1D).$$

On the other hand, the nitric oxide can react with peroxy ($RO_2\cdot$) to form nitrogen dioxide:

$$RO_2 \cdot \; + \; NO \; + \; \longrightarrow \; NO_2 \; + \; RO \cdot .$$

This process is a chain reaction. The solar radiation plays a key role in this process. Due to the alternating of day and night, the daily ozone concentration curve against hour is cyclic. It increases from a low value at midnight, and reaches a maximum in the afternoon, then decreases in the night. For more details, see NRC (National Research Council) [1992], p. 24-37.

Based on the photochemical mechanism of ozone, we can make up a simple reasoning. The amount of today's NO_2 consists of two parts: one is the newly formed NO_2 from NO reacting with $RO_2\cdot$, one is emitted NO_2 from other sources. The whole NO in the troposphere, of course, includes the NO decomposed from NO_2 yesterday, which roughly accounts for the amount of yesterday's ozone. Roughly, the daily maximum is positively associated with the daily amount of ozone. Thus, the newly formed NO_2 from NO links today's maximum with yesterday's maximum, a positive association. Thus, today's maximum can be expressed in two terms: one is dependent on yesterday's maximum, and one is innovation. The emitted NO_2 and part of the NO which is not involved in yesterday chemical reaction, can be accounted for the innovation.

Note that the daily maximum ozone concentration series is not a continuous-time process. It is discrete-time, artificially divided by day. But from the continuous-time GAR(1) process, we can obtain the discrete-time GAR(1) process which may be appropriate to model such data. Because of the feature of Gamma margins as we have considered, we try the following GAR(1) model with Gamma(δ, β) margins:

$$X(t_{i+1}) \stackrel{d}{=} (\alpha)_K \circledast X(t_i) + E_i, \tag{13.4.1}$$

where the self-generalized rv K is from **P2** with LT

$$\phi_K(s;\alpha) \quad = \quad \exp\left\{-\frac{\alpha(1-\gamma)s}{(1-\gamma)+(1-\alpha)\gamma s}\right\}, \qquad 0 \leq \gamma \leq \frac{1}{1+\beta}.$$

When $\gamma = 0$, K becomes the self-generalized rv from **P1**, and (13.4.1) is

$$X(t_{i+1}) \stackrel{d}{=} \alpha \bullet X(t_i) + E_i. \tag{13.4.2}$$

(13.4.1) is a big model family as γ changes in $\left[0, (1+\beta)^{-1}\right]$. Although the process has different dependence structure for different γ, each process of this family has the common conditional mean

$$\mathbf{E}\left[X(t_{i+1}) \mid X(t_i) = x_i\right] = \delta\beta^{-1}(1-\alpha) + \alpha x_i$$

no matter what the value of γ. In addition, γ is a fixed parameter in **P2** definition. Hence, we first ignore it by assuming it being the boundary value 0 or $(1+\beta)^{-1}$. Applying the method of moments, we obtain the estimates

$$\widehat{\delta} = 8.10, \qquad \widehat{\beta} = 0.17, \qquad \widehat{\alpha} = 0.58,$$

for all the models in (13.4.1).

To diagnose the fitted models, we need to calculate the bivariate cdf. However, it involves two-dimension integration, and the integrand formula is very complicated. We can estimate the bivariate cdf by simulating the process in Model 13.4.1 for any $\gamma \in \left[0, (1+\beta)^{-1}\right]$. We choose

$$\gamma = 0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 1/1.17,$$

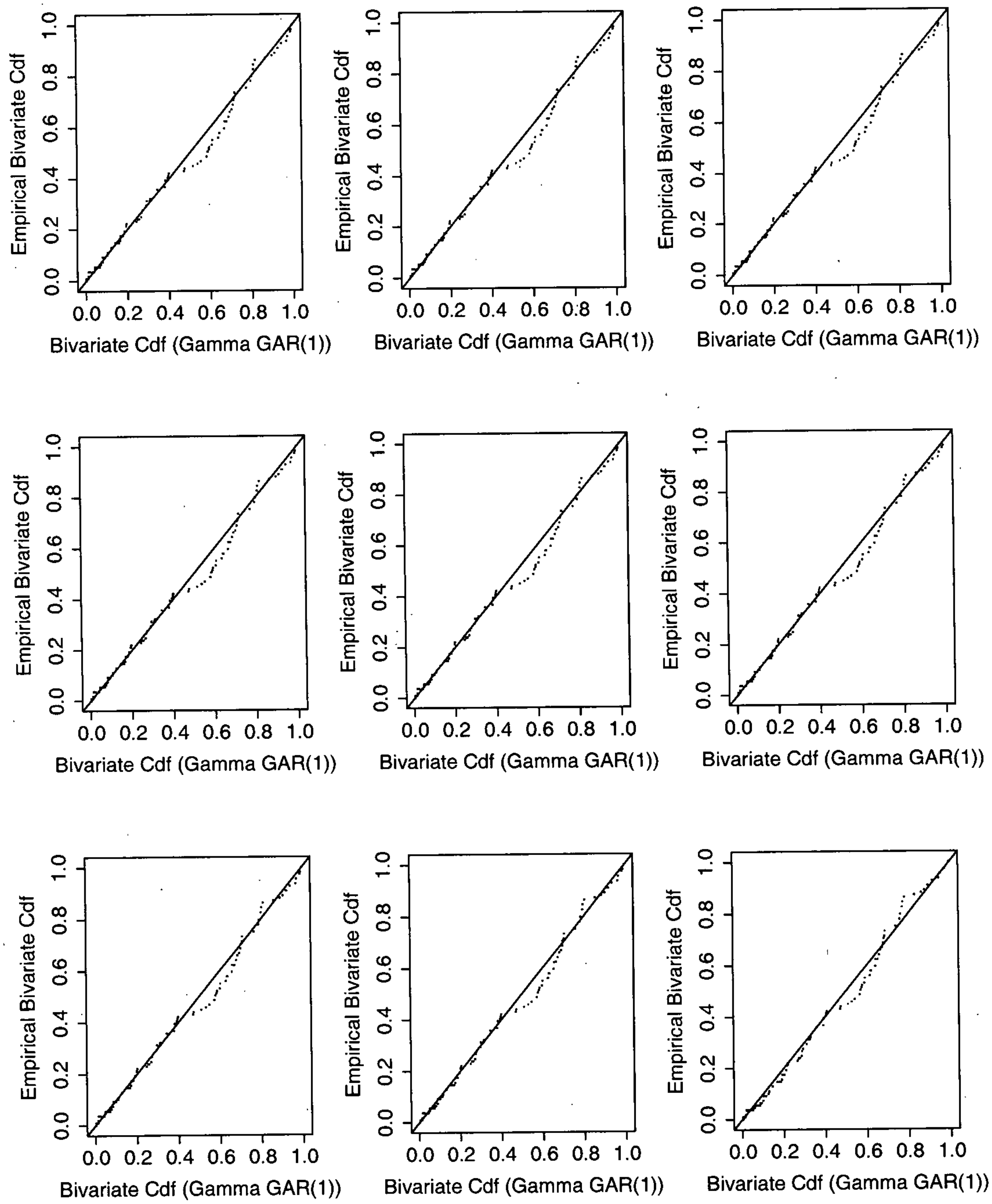


Figure 13.20: *Model diagnosis: diagonal P-P plots of lag one day for* $\gamma = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 *and* 1/1.17 *in the ozone data study.*

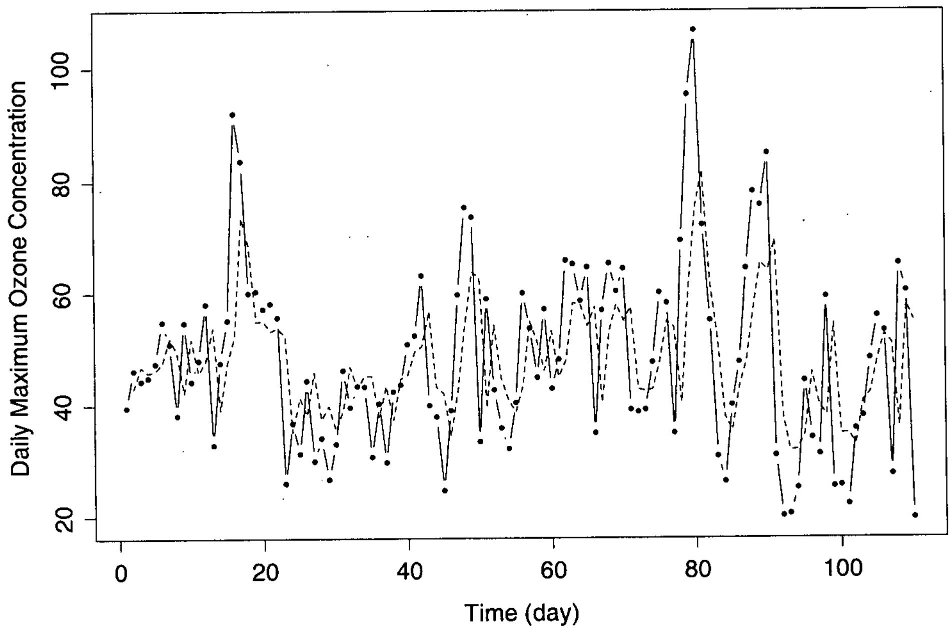


Figure 13.21: *One step ahead predictions (dotted line) by the conditional mean for the daily maximum ozone concentrations (solid line).*

and simulate each process with 1000,000 samples. The diagonal P-P plots are given in Figure 13.20. These diagonal P-P plots show that the discrepancy gradually decrease as γ increases. This suggests that the model with $\gamma = 1/1.17$ seems to fit the data better.

There is additional evidence that doesn't support (13.4.2) where γ reaches the lower boundary 0. If the true model is (13.4.2), then by "quick and dirty" method, we can obtain a upper bound of α: $\widehat{\alpha}_R = \min\{X(t_{i+1})/X(t_i)\} = 0.32$. However, this upper bound is quite a bit smaller than the estimated lag-1 autocorrelation coefficient of 0.53.

Now we consider one step ahead prediction by using the conditional mean

$$\mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i] = \delta\beta^{-1}(1-\alpha) + \alpha x_i = 0.58x_i + 20.01, \qquad i = 1, 2, \ldots, 109,$$

to forecast $X(t_2), X(t_3), \ldots, X(t_{110})$. This leads to Figure 13.21, where the dotted line denotes the predictions. It seems that the forecasting captures the main fluctuation. Furthermore, we investigate the differences, $x_{i+1} - \mathbf{E}[X(t_{i+1}) \mid X(t_i) = x_i]$ $(i = 2, 3, \ldots, 110)$, between the observations and predictions. They can be viewed as residuals from (13.4.2). We can expect the skewness of their distributions. Figure 13.22 presents the histogram, time series plot and ACF plot of these residuals. They have a skewed distribution, fluctuating around zero with no obvious serial dependence. It shows from another aspect that our model fitting is successful in getting rid of the autocorrelation.

Lastly, we restrict the models in (13.4.1) which are associated with parameter $0 \leq \gamma \leq 1/(\beta+1)$ to the upper boundary case $\gamma = 1/(\beta+1)$. This is because the diagnostic analysis from the diagonal P-P plots. Thus, the model is

$$X(t_{i+1}) \stackrel{d}{=} (\alpha)_K \circledast X(t_i) + E_i, \tag{13.4.3}$$

where

$$\phi_K(s;\alpha) = \exp\left\{-\frac{\alpha\beta s}{\beta + (1-\alpha)s}\right\} = \exp\left\{\frac{\alpha\beta}{1-\alpha}\left(\frac{\beta/(1-\alpha)}{\beta/(1-\alpha)+s} - 1\right)\right\}$$

and

$$\phi_{E_i}(s;\alpha) = \left(\frac{\beta}{\beta + (1-\alpha)s}\right)^{\delta} = \left(\frac{\beta/(1-\alpha)}{\beta/(1-\alpha)+s}\right)^{\delta}.$$

Note that K is a rv of compound Poisson$(\alpha\beta/(1-\alpha))$ with exponential$(\beta/(1-\alpha))$, and E_i is a rv of Gamma$(\delta, \beta/(1-\alpha))$. Therefore, conditioned on $X(t_i) = x_i$, $(\alpha)_K \circledast x_i$ is a rv of compound Poisson$(\alpha\beta x_i/(1-\alpha))$ with exponential$(\beta/(1-\alpha))$, leading to a stochastic representation

$$(\alpha)_K \circledast x_i \stackrel{d}{=} \sum_{j=0}^{N} Y_j, \quad N \sim \text{Poisson}(\alpha\beta x_i/(1-\alpha)), \quad Y_0 = 0,\ Y_j \stackrel{i.i.d.}{\sim} \text{exponential}(\beta/(1-\alpha)).$$

This representation will help us to find the closed form of pdf for the conditional rv $[X(t_{i+1})|X(t_i) = x_i]$ because that conditioned on $N = n$,

$$\sum_{j=0}^{n} Y_j \sim \text{Gamma}(n, \beta/(1-\alpha)) \quad \text{and} \quad \sum_{j=0}^{n} Y_j + E_i \sim \text{Gamma}(n+\delta, \beta/(1-\alpha)).$$

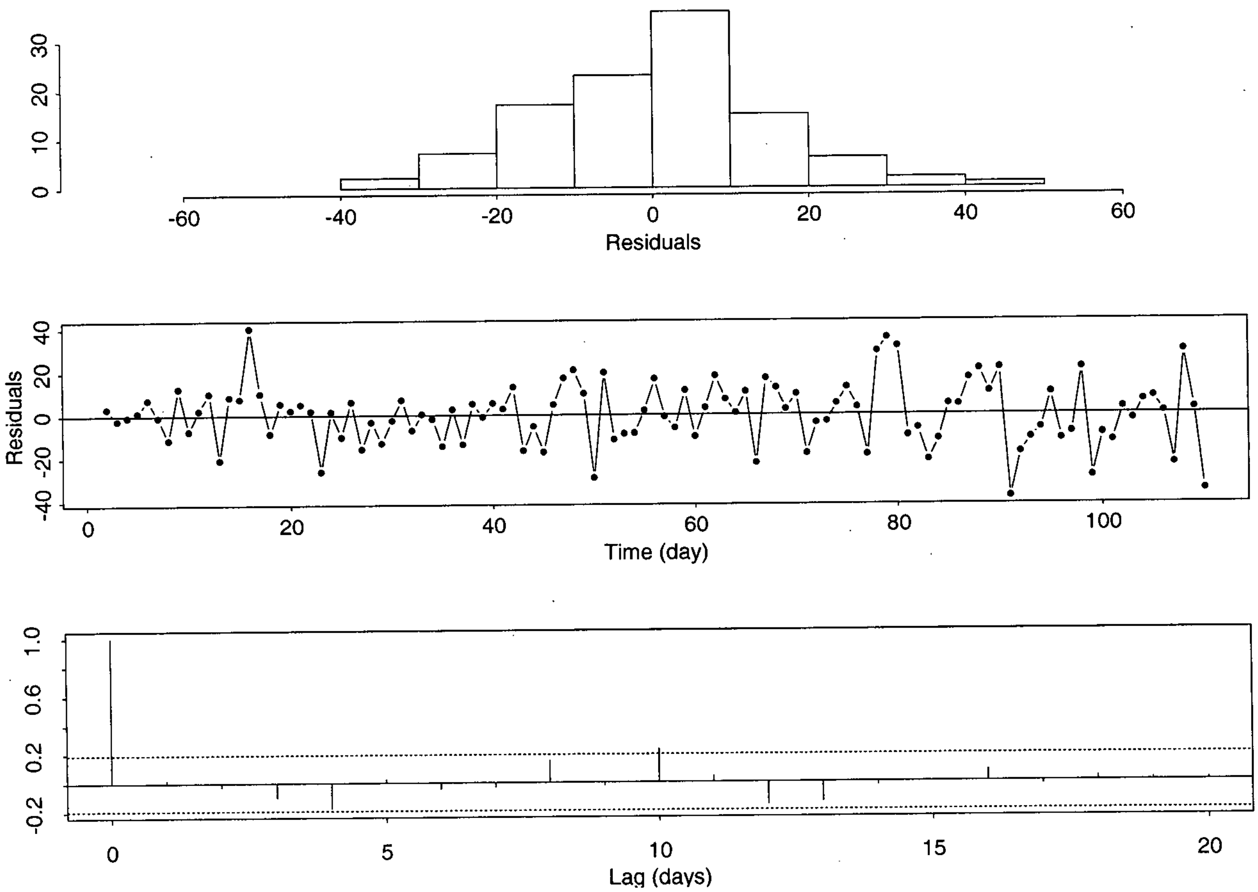


Figure 13.22: *The analysis of differences between observations and one step ahead predictions by the conditional mean for the daily maximum ozone concentration.*

Hence, the conditional pdf of $X(t_{i+1})$ given $X(t_i) = x_i$ is

$$\begin{aligned} f_{X(t_{i+1})|X(t_i)}(x|x_i) &= \sum_{n=0}^{\infty} f_{\sum_{j=0}^{n} Y_j + E_i}(x) \Pr(N=n) \\ &= \sum_{n=0}^{\infty} \frac{(\frac{\beta}{1-\alpha})^{n+\delta}}{\Gamma(n+\delta)} x^{n+\delta-1} e^{-\beta x/(1-\alpha)} \times \frac{1}{n!}\left(\frac{\alpha\beta x_i}{1-\alpha}\right)^n e^{-\alpha\beta x_i/(1-\alpha)} \quad (13.4.4) \end{aligned}$$

Generally, statistical software can calculate the Gamma density and the Poisson pmf. Thus, the pdf of $[X(t_{i+1})|X(t_i) = x_i]$ can be easily computed. With (13.4.4) and numerical optimization of the log-likelihood using a quasi-Newton routine, we can obtain the MLE:

$$\widehat{\delta}_{MLE} = 8.31, \qquad \widehat{\beta}_{MLE} = 0.17, \qquad \widehat{\alpha}_{MLE} = 0.51,$$

for Model (13.4.3). The estimate of β is still 0.17.

The calculation of the joint cdf for $(X(t_i), X(t_{i+1}))$ requires one-dimensional integration.

$$\begin{aligned} \Pr[X(t_i) < x, X(t_{i+1}) < y] &= \int_0^x \Pr[(\alpha)_K \circledast x + E_i < y] f_{X(t_i)}(x)dx \\ &= \int_0^x \Pr\left[\sum_{j=0}^{N} Y_j + E_i < y\right] \frac{\beta^\delta}{\Gamma(\delta)} x^{\delta-1} e^{-\beta x} dx, \quad (13.4.5) \end{aligned}$$

where

$$\Pr\left[\sum_{j=0}^{N} Y_j + E_i < y\right] = \sum_{n=0}^{\infty} \Pr\left[\sum_{j=0}^{n} Y_j + E_i < y\right] \times \frac{1}{n!}\left(\frac{\alpha\beta x_i}{1-\alpha}\right)^n e^{-\alpha\beta x_i/(1-\alpha)},$$

$$\sum_{j=0}^{n} Y_j + E_i \sim \text{Gamma}(n+\delta, \beta/(1-\alpha)).$$

Usually, the gamma cdf is available in statistical software. With one-dimensional numerical integration function, such a joint probability can be obtained easily. Based on (13.4.5), we can draw the diagonal P-P plot for Model (13.4.3). See Figure 13.23. For $\gamma = 1/0.17$, comparing the model with estimates by the method of moments, it seems there is an improvement for the model with maximum likelihood estimates.

If $0 < \gamma < 1/(\beta+1)$, however, E_i is no longer a Gamma rv. Thus, finding the pdf of $\sum_{j=0}^{n} Y_j + E_i$ may be a problem. If we know the pdf of E_i, we can still use the previous method to numerically obtain the pdf of $\sum_{j=0}^{n} Y_j + E_i$, because it will involve a one-dimensional integration. Such a calculation, of course, is much messier than that for γ being the upper boundary.

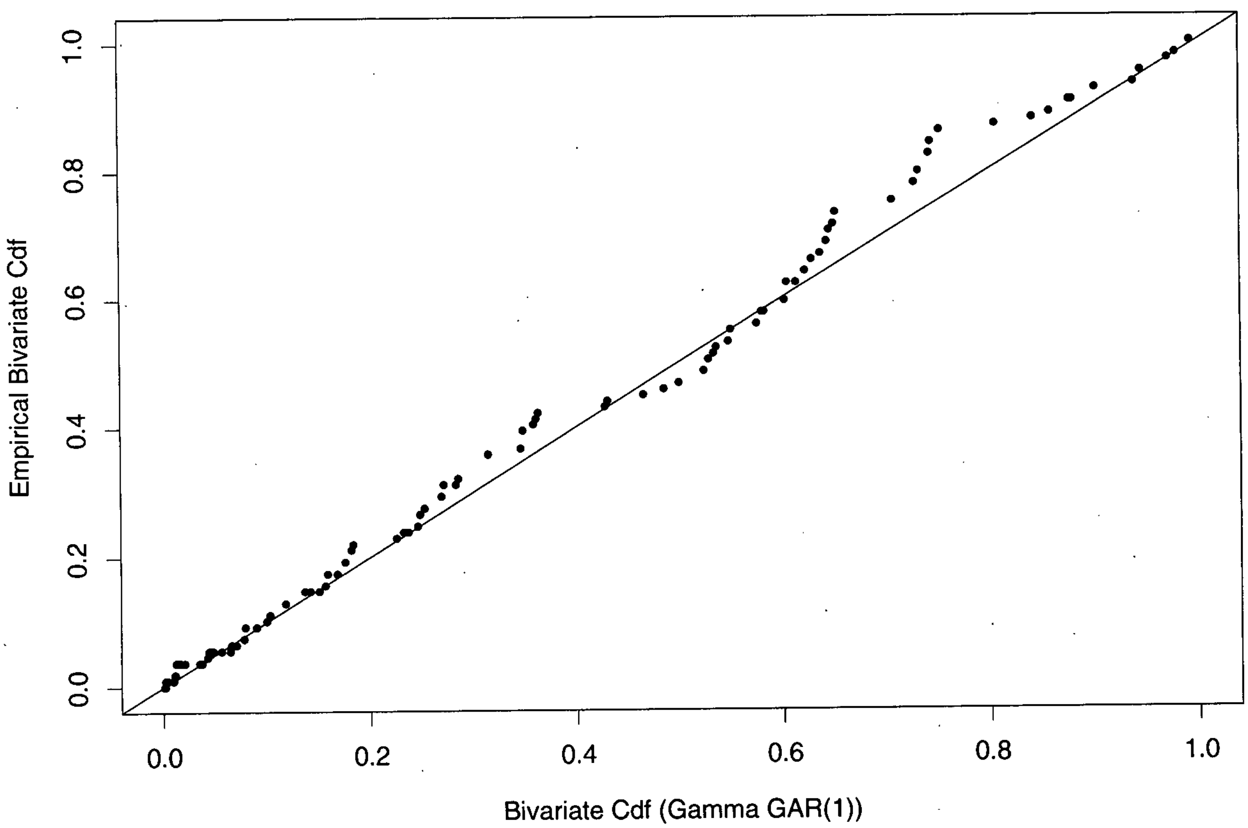


Figure 13.23: *Model diagnosis: diagonal P-P plot of lag one day for Model (13.4.3) at* $\widehat{\delta}_{MLE} = 8.31$, $\widehat{\beta}_{MLE} = 0.17$ *and* $\widehat{\alpha}_{MLE} = 0.51$.

Part V

Discussion

Chapter 14

Conclusions and further research topics

Research in the area of non-normal time series does not have a long history. People have been working on it from various perspectives. The theory of continuous-time generalized AR(1) processes is developed to model problems from this area, especially with unequally-spaced time series data. We of course want to know how good and how flexible it is for modelling practical problems. In addition, we are also concerned with its application scope and limitations. Based on this new theory, we may further develop other complex (continuous-time or discrete-time) stochastic processes to handle those problems which can't be suitably modelled by the continuous-time generalized AR(1) processes. There remains much research for the new theory; there is interest from the viewpoint of either probability or statistics.

In Section 14.1, we summarize some advantages and disadvantages of the theory of continuous-time generalized AR(1) processes. In Section 14.2, we also discuss some ideas of construction of stochastic processes for more complex problems. Lastly, in Section 14.3, we present some topics of future research in the theory of continuous-time generalized AR(1) processes.

14.1 Discussion of continuous-time generalized AR(1) processes

In the theory of generalized linear model (GLM), the distributions like Poisson, Gamma, etc, for count and positive-valued responses are well developed; only a few situations like zero-inflation need further development of distributions. However, to handle count or positive-valued time series, we need models which haven't been well developed. This is why we spend much time and effort to develop the probabilistic foundation. Thus, unlike GLM, developing appropriate models in the stochastic framework is one of the key pursuits.

The theory of continuous-time generalized AR(1) processes is designed to model equally-spaced or unequally-spaced time series with count or positive-valued observations. From the sense of data type, it is quite similar to generalized linear model theory, which handles the non-negative integer or positive-valued response. Our theory presents a systematic way to construct the continuous-time Markov processes. From the continuous-time processes, we then can easily obtain the discrete-time processes by sampling or observing at equally-spaced time points. The strengths of the GAR(1) processes are:

- simple decomposition of the dependent and innovation terms;
- parametric families for the various probabilistic components;
- flexible choice among abundant models.

These features result in the model having a simple interpretation, a desired marginal distribution and a reasonable dependence for real problems. The case studies in Chapter 13 illustrate these capabilities.

Of course, the theory of continuous-time generalized AR(1) processes has some weaknesses:

- only a positive geometric autocorrelation function is possible;
- marginal or stationary distribution is restricted to the infinitely divisible class;
- no explicit expression of the conditional pmf or pdf for many processes;

- support range $[0, \infty)$ or $(-\infty, \infty)$ or $\{0, 1, 2, \ldots\}$;
- computational complexity.

These disadvantages motivate us to think about new stochastic processes to handle more general situations.

14.2 Some thoughts on model construction

Stochastic operators play an important role in the construction of stochastic processes, or even more generally, multivariate distributions. The extended-thinning operators concretely provide the positive association between two non-negative random variables; this enlarges our scope beyond the constant multiplier operator. Exploration of new stochastic operators is very meaningful in developing new multivariate distributions, which consequently could lead to new stochastic processes (in either continuous-time or discrete-time). This could help us to construct bivariate distribution where two margins have negative correlation; such negatively correlated bivariate distributions may help us to build discrete-time stochastic processes with negative lag-1 correlation.

Since construction of discrete-time processes only requires specifying the bivariate distribution of two adjacent time points, we may create a discrete-time stochastic process with more than one stochastic operator We can apply one operator on one time period, and apply another one on the next time period, and so on. If we properly choose the stochastic operators, we may obtain a stationary discrete-time process with the same correlation coefficient between any two adjacent time points, because different extended-thinning operators can lead to the same correlation coefficient.

It's possible to encounter time series data which can have a slow decrease in the ACF. This may suggest higher order autoregressive processes which can be defined in discrete-time. [Note that there are no known higher-order autoregressive or Markov processes in continuous-time.] Thus, it is necessary to build the higher order Markov processes. One idea is described in (2.2) and (2.4) of Lawrance and Lewis [1980]; it leads to the similar autocorrelation structure to the Gaussian AR(p) model, for non-normal discrete-time time series.

Non-stationary process development is another practical concern, because there often arises trends or periodicities. We could adjust the parameter μ from constant to time-varying, which leads to a time-varying correlation structure for possible application in growth curve studies. We could also change the mean of the innovation to be time-varying, and allow a trend. Of course, we can modify both to obtain more complex non-stationary processes.

14.3 Future research

In this section, we briefly discuss some further research topics relating to the theory of continuous-time generalized AR(1) processes.

First, we could search for more families of self-generalized distributions leading to new extended-thinning operations. This in turn leads to new GSD and GDSD classes. Although we fortunately find ten families of self-generalized distributions, it is not enough. We wish there could exist a representation form for the pgf or LT of all self-generalized distributions. But this may not be true, and thus, remains an open question.

We also need to study the property of different GSD or GDSD classes of distributions. This leads to new continuous-time generalized AR(1) processes with different dependent structures and marginal distributions of interest. These developments will meet the potential needs of real problems from the point of view of either dependence or marginal distribution.

Further probabilistic study on the continuous-time generalized AR(1) processes is required. This will be useful in probability calculations (conditional and bivariate), as well as in simulation studies.

Asymptotic study of estimators other than MLE and variations of CLS are also needed, because they will help us to construct confidence intervals or regions, and do hypothesis testing.

In addition, incorporating covariates to the model should be considered in the framework of stochastic processes. Most of the models in both the GLM and GAR(1) theory are stochastic models, hence, the P-P plot is a sophisticated diagnostic tool, different from residual plots which work well in structure models. When we develop non-stationary processes, we also need to develop new diagnostic graphical tools.

Appendix A

Data sets

A.1 Manuscripts data

Description: The following irregular time series consists of the number of manuscripts in refereeing queue of Prof. H. Joe.

Date	Manuscripts	Date	Manuscripts	Date	Manuscripts
1990-01-01	4	1992-03-01	8	1994-06-01	0
1990-03-01	1	1992-04-01	5	1994-08-01	0
1990-05-01	1	1992-05-01	4	1994-10-01	0
1990-07-01	1	1992-06-01	3	1995-01-01	1
1990-09-01	3	1992-07-01	3	1995-03-01	3
1990-10-01	5	1992-08-01	1	1995-05-01	2
1990-11-01	5	1992-10-01	1	1995-07-01	2
1990-12-01	5	1992-12-01	1	1995-08-01	2
1991-02-01	3	1993-02-01	2	1995-10-01	0
1991-04-01	5	1993-04-01	1	1995-12-01	3
1991-05-01	5	1993-05-01	2	1996-02-01	2
1991-06-01	2	1993-06-01	2	1996-04-01	2
1991-08-01	4	1993-08-01	1	1996-06-01	2
1991-10-01	3	1993-10-01	0	1996-08-01	3
1991-12-01	4	1993-12-01	2	1996-10-01	3
1992-01-01	5	1994-02-01	2	1996-12-01	1
1992-02-01	9	1994-04-01	2	1997-02-01	2

Date	Manuscripts	Date	Manuscripts	Date	Manuscripts
1997-04-01	3	1999-03-01	2	2000-03-01	2
1997-06-01	2	1999-04-01	1	2000-04-01	0
1997-08-01	3	1999-05-01	1	2000-05-01	2
1997-10-01	1	1999-06-01	1	2000-06-01	2
1997-12-01	0	1999-07-01	2	2000-07-01	1
1998-02-01	1	1999-08-01	3	2000-08-01	3
1998-04-01	3	1999-09-01	2	2000-09-01	4
1998-06-01	3	1999-10-01	2	2000-10-01	5
1998-08-01	0	1999-11-01	1	2000-11-01	3
1998-10-01	2	1999-12-01	3	2000-12-01	2
1998-12-01	2	2000-01-01	3		
1999-02-01	3	2000-02-01	4		

Source: Prof. H. Joe. Dept. of Statistics, UBC, Vancouver, B.C. V6T 1Z2, Canada.

A.2 WCB claims data

Description: The following data are monthly counts of claims by workers collecting short-term disability benefit (STWLB) from the Richmond claims center of the Workers' Compensation Board (WCB) of British Columbia, Canada.

Each column forms a time series. C0 denotes the series of counts of claims by workers in heavy manufacturing industry who are male and between age of 25 and 34. The injury is burn related. The claimants in C1, C2, C3, C4, C5 are male and between age of 35 and 54 who work in the logging industry. The difference among them is the nature of injury: C1 indicates burn related injury; C2 indicates soft tissue injury such as contusions and bruises; C3 indicates cuts, lacerations or punctures; C4 indicates dermatitis; C5 indicates dislocations.

C1a is obtained from C1 by removing one claimant who has a ten year claim. C1A to C5A are the arrival data corresponding to C1 to C5, with the counts of new claims for each month.

Date	C0	C1	C1a	C2	C3	C4	C5	C1A	C2A	C3A	C4A	C5A
Jan-85	NA	0	0	9	6	0	0	0	2	2	0	0
Feb-85	NA	0	0	6	7	1	0	0	0	3	1	0

Mar-85	NA	0	0	6	8	0	1	0	3	4	0	1
Apr-85	NA	0	0	7	9	0	1	0	3	5	0	0
May-85	NA	0	0	10	6	1	1	0	8	1	1	0
Jun-85	NA	0	0	8	8	0	1	0	2	4	0	0
Jul-85	NA	0	0	14	5	0	1	0	10	4	0	0
Aug-85	NA	0	0	8	3	0	1	0	4	1	0	0
Sep-85	NA	0	0	7	7	0	0	0	5	4	0	0
Oct-85	NA	0	0	10	11	0	1	0	8	8	0	1
Nov-85	NA	0	0	10	8	1	1	0	9	5	1	0
Dec-85	NA	0	0	12	4	0	2	0	6	3	0	2
Jan-86	NA	0	0	8	2	0	0	0	6	1	0	0
Feb-86	NA	0	0	8	3	0	0	0	4	2	0	0
Mar-86	NA	1	1	8	4	0	0	1	5	3	0	0
Apr-86	NA	1	1	8	5	1	0	0	4	4	1	0
May-86	NA	1	1	13	7	1	1	1	8	2	1	0
Jun-86	NA	1	1	12	8	0	0	1	8	5	0	0
Jul-86	NA	0	0	14	12	0	0	0	7	8	0	0
Aug-86	NA	0	0	13	11	0	1	0	6	6	0	1
Sep-86	NA	0	0	13	12	0	1	0	6	7	0	1
Oct-86	NA	0	0	8	6	1	1	0	3	5	1	0
Nov-86	NA	0	0	13	2	1	1	0	8	1	1	0
Dec-86	NA	1	1	10	2	0	1	1	3	0	0	0
Jan-87	6	1	1	12	3	0	0	1	5	2	0	0
Feb-87	11	0	0	12	3	0	0	0	7	1	0	0
Mar-87	5	0	0	9	5	0	0	0	4	2	0	0
Apr-87	5	0	0	8	6	0	1	0	4	2	0	1
May-87	5	0	0	13	13	2	0	0	10	9	2	0
Jun-87	2	0	0	9	12	0	0	0	4	6	0	0
Jul-87	7	0	0	8	21	0	0	0	3	15	0	0
Aug-87	4	0	0	6	9	0	0	0	3	3	0	0
Sep-87	5	0	0	7	11	1	0	0	4	6	1	0
Oct-87	4	0	0	10	11	0	0	0	8	7	0	0
Nov-87	6	1	1	17	10	0	2	1	11	5	0	2
Dec-87	8	0	0	11	8	0	1	0	6	2	0	1
Jan-88	7	1	1	13	5	0	0	1	8	2	0	0
Feb-88	7	0	0	10	4	0	0	0	4	2	0	0
Mar-88	9	1	1	9	4	0	2	0	5	3	0	2
Apr-88	9	2	2	15	4	0	2	2	7	2	0	1
May-88	13	0	0	13	2	0	2	0	8	1	0	0
Jun-88	12	0	0	12	9	0	2	0	9	8	0	1
Jul-88	11	0	0	8	8	0	1	0	2	5	0	0
Aug-88	13	1	1	8	5	0	0	1	5	3	0	0
Sep-88	16	0	0	9	10	0	0	0	2	8	0	0
Oct-88	8	0	0	9	12	1	1	0	4	4	1	1
Nov-88	14	0	0	12	11	1	1	0	5	7	0	0
Dec-88	10	0	0	9	9	0	1	0	4	3	0	0
Jan-89	6	0	0	5	4	0	1	0	2	1	0	0
Feb-89	6	0	0	9	5	0	1	0	4	2	0	0

Mar-89	15	0	0	10	5	0	2	0	5	2	0	1
Apr-89	9	0	0	6	10	0	2	0	2	6	0	0
May-89	15	0	0	8	14	0	2	0	5	5	0	0
Jun-89	13	0	0	17	7	1	2	0	11	2	1	0
Jul-89	11	0	0	16	11	1	2	0	11	4	0	1
Aug-89	14	0	0	17	12	1	3	0	12	9	0	0
Sep-89	11	0	0	16	7	1	1	0	7	2	0	0
Oct-89	17	0	0	8	8	1	2	0	4	6	0	2
Nov-89	8	0	0	10	14	1	2	0	8	7	0	2
Dec-89	10	0	0	7	6	1	1	0	2	2	0	1
Jan-90	11	0	0	8	4	1	1	0	3	0	0	0
Feb-90	13	0	0	7	3	1	0	0	1	1	0	0
Mar-90	10	0	0	4	4	1	1	0	1	3	0	1
Apr-90	8	0	0	5	4	1	3	0	3	3	0	1
May-90	8	0	0	4	7	0	1	0	2	3	0	0
Jun-90	6	0	0	4	6	0	1	0	2	4	0	0
Jul-90	9	1	1	10	9	0	1	0	4	3	0	0
Aug-90	12	0	0	9	8	0	1	0	2	5	0	0
Sep-90	11	1	1	12	2	0	2	1	6	0	0	1
Oct-90	9	0	0	12	4	0	1	0	4	3	0	0
Nov-90	11	1	1	11	3	0	1	1	6	0	0	0
Dec-90	7	0	0	9	1	0	2	0	3	0	0	1
Jan-91	9	0	0	8	3	0	2	0	0	2	0	0
Feb-91	11	0	0	9	1	0	1	0	3	0	0	0
Mar-91	6	0	0	8	4	1	1	0	3	2	1	0
Apr-91	4	0	0	6	3	0	3	0	3	1	0	1
May-91	6	0	0	8	5	0	4	0	3	4	0	2
Jun-91	6	0	0	13	3	0	3	0	6	1	0	1
Jul-91	12	0	0	13	8	0	2	0	5	6	0	0
Aug-91	10	0	0	10	11	0	1	0	3	4	0	0
Sep-91	12	0	0	7	7	0	1	0	2	4	0	0
Oct-91	8	1	1	17	9	0	0	1	11	6	0	0
Nov-91	6	0	0	14	5	0	0	0	2	1	0	0
Dec-91	1	0	0	10	3	1	0	0	0	1	1	0
Jan-92	3	0	0	12	6	0	1	0	3	2	0	0
Feb-92	5	0	0	6	4	0	2	0	0	2	0	1
Mar-92	5	0	0	4	5	0	0	0	1	4	0	0
Apr-92	10	0	0	7	6	0	0	0	4	3	0	0
May-92	12	0	0	8	7	1	0	0	4	3	1	0
Jun-92	9	0	0	10	7	1	0	0	4	4	0	0
Jul-92	7	0	0	16	3	1	0	0	9	2	0	0
Aug-92	9	0	0	15	5	1	0	0	8	4	0	0
Sep-92	12	0	0	10	5	2	0	0	1	2	1	0
Oct-92	14	0	0	14	4	0	1	0	6	2	0	1
Nov-92	11	0	0	16	4	0	1	0	5	3	0	1
Dec-92	9	0	0	12	2	0	1	0	3	0	0	0
Jan-93	3	0	0	10	3	0	1	0	2	1	0	0
Feb-93	4	0	0	11	6	0	1	0	2	1	0	0

Mar-93	10	0	0	10	3	0	1	0	2	1	0	0
Apr-93	2	0	0	8	1	0	1	0	5	0	0	0
May-93	7	1	0	9	3	0	0	1	6	2	0	0
Jun-93	9	1	0	10	6	1	0	1	6	3	1	0
Jul-93	9	2	1	13	5	0	0	1	8	3	0	0
Aug-93	3	1	0	6	9	0	1	0	1	5	0	1
Sep-93	6	1	0	8	9	0	1	0	4	6	0	0
Oct-93	9	1	0	9	5	0	1	0	5	1	0	0
Nov-93	9	1	0	6	6	0	1	0	1	4	0	0
Dec-93	9	1	0	9	4	0	1	0	5	2	0	0
Jan-94	6	2	1	12	6	0	0	1	4	3	0	0
Feb-94	5	1	0	8	2	0	0	0	2	0	0	0
Mar-94	6	1	0	9	4	0	0	0	4	3	0	0
Apr-94	5	1	0	5	1	1	0	0	2	0	1	0
May-94	9	1	0	6	6	0	0	0	5	4	0	0
Jun-94	7	1	0	9	5	0	1	0	2	3	0	1
Jul-94	11	1	0	9	3	0	0	0	4	1	0	0
Aug-94	12	1	0	13	2	0	0	0	6	1	0	0
Sep-94	11	2	1	12	2	0	1	1	5	1	0	0
Oct-94	12	2	1	10	2	1	2	0	4	1	0	1
Nov-94	7	2	1	9	9	0	3	1	2	8	0	1
Dec-94	11	1	0	7	5	0	2	0	3	3	0	0

Source: Freeland [1998], Appendix. The original data description is scattered throughout Freeland's thesis, mainly in Chapter 8.

A.3 Abbotsford daily maximum ozone concentrations data

Description: The following data are daily maximum ozone concentration collected at the Abbortsford (British Columbia, Canada) ozone station in the summer of 1985 from May 1 to August 18 inclusively. The rows from the first to the eleventh are records corresponding to May 1 to 10, May 11 to 20, May 21 to 30, May 31 to June 9, June 10 to June 19, June 20 to June 29, June 30 to July 9, July 10 to July 19, July 20 to July 29, July 30 to August 8, and August 9 to August 18. This is a part of a data set from a consulting project Prof. Joe conducted at the Statistical Consulting and Research Laboratory (SCARL), UBC.

39.5	46.1	44.3	44.9	47.4	54.8	50.8	38.1	54.6	44.2

47.9	58	32.8	47.5	55.1	92.2	83.7	60	60.3	57.1
58.1	55.6	25.9	36.7	31.2	44.3	29.9	34	26.6	32.9
46.1	39.5	43.3	43.2	30.6	40.2	29.6	42.3	43.5	50.7
52.2	63.1	39.9	37.8	24.6	38.9	59.6	75.3	73.5	33.3
58.9	42.6	35.7	32	40.3	59.9	53.6	44.8	57	42.8
48	65.7	65.1	58.5	64.5	34.8	56.8	65.2	60.2	64.3
39	38.6	39	47.5	60	58.1	34.8	69.2	95.4	106.7
72	55	30.6	26.1	39.9	47.5	64.3	78.1	75.6	85
30.7	19.9	20.3	25	44.2	33.9	30.9	59.2	25.2	25.4
22	35.5	37.8	48.1	55.7	53	27.4	65.1	60.2	19.5

Bibliography

[1] Adke, S.R. and Balakrishna, N. (1992). Markovian chi-square and gamma processes. *Statistics & Probability Letters.* **15**, 349-356.

[2] Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution with Special Reference to its Uses in Economics.* Cambridge University Press, London.

[3] Aitchison, J. and Silvey, S.D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29**, 813-828.

[4] Allen, O.B. (1983). Asymptotic properties of the maximum-likelihood estimator for a class of birth-and-death processes admitting a unique stationary distribution. *Canad. J. Statist.*,**11**, 109-118.

[5] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Series Anal.*, **8**, 61-275.

[6] Al-Osh, M.A. and Alzaid, A.A. (1991). Binomial autoregressive moving average models. *Commun. Statist. Stoch. Models* **7**, 261-282.

[7] Al-Osh, M.A. and Aly, E.A.A. (1992). First order autoregressive time series with negative binomial and geometric marginals. *Commun. Statist.-Theor. Meth.* **21**, 2483-2492.

[8] Aly, E.A.A. and Bouzar, N. (1994). Explicit stationary distributions for some Galton-Watson processes with immigration. *Commun. Statist. -Stochastic Models,* **10(2)**, 499-517.

[9] Alzaid, A.A. and Al-Osh, M.A. (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions *Ann. Int. Statist. Math.* **45**, 223-232.

[10] Anděl, J. (1988). On AR(1) processes with exponential white noise. *Comm. Statist. Theory Methods,* **17**, 1481-1495.

[11] Anděl, J. (1989a). Nonnegative autoregressive processes. *J. Time Ser. Anal.,* **10**, 1-11.

[12] Anděl, J. (1989b). Nonlinear nonnegative AR(1) processes. *Comm. Statist. Theory Methods,* **18**, 4029-4037.

[13] Anderson, O.D. (1988a). Serial correlation. In *Encyclopedia of Statistical Sciences.* Vol. **8**, p. 411-415. Chief-editor: Kotz, S. & Johnson, N. L. Wiley, New York.

[14] Anderson, O.D. (1988b). Serial dependence. In *Encyclopedia of Statistical Sciences.* Vol. **8**, p. 415-417. Chief-editor: Kotz, S. & Johnson, N. L. Wiley, New York.

[15] Anderson, W.J. (1991). *Continuous-Time Markov Chains: An Applications-Oriented Approach.* Springer-Verlag, New York.

[16] Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions. *Biometrics,* **41**, 875-885.

[17] Azencott, R. and Dacunha-Castelle, D. (1986). *Series of Irregular Observations. Forecasting and Model Building.* Springer-Verlag, New York.

[18] Barndorff-Nielsen, O.E. and Jørgensen, B. (1991). Some parametric models on the simplex. *J. Multivariate Anal.* **39**, 106-116.

[19] Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1993). A statistical model for the streamwise component of a turbulent velocity field. *Ann. Geophys.* **11**, 99-103.

[20] Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1998a). Some stationary processes in discrete and continuous time. *Adv. Appl. Probab.* **30**, 989–1007.

[21] Barndorff-Nielsen, O.E. (1998b). Processes of normal inverse Gaussian type. *Finance and Stochastics* **2**, 41-68.

[22] Barndorff-Nielsen, O.E. (1998c). Probability and Statistics: self-decomposability, finance and turbulence. In *Probability Towards 2000*, ed. L. Accardi and C.C. Heyde (Lecture Notes in Statist. **128**). Springer, New York, pp.45-57.

[23] Bartlett, M.S. (1960). *An Introduction to Stochastic Processes with Special Reference to Methods and Applications.* Cambridge University Press, London.

[24] Basawa, I.V. and Prakasa Rao, B.L.S. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London-New York.

[25] Bell, C.B. and Smith, E.P. (1986). Inference for non-negative autoregressive schemes. *Commun. Statist.-Theor. Meth.,* **15(8)**, 2267-2293.

[26] Bertoin, J. (1996). *Lévy Processes.* Cambridge University Press, London.

[27] Bharucha-Reid, A. T. (1960). *Elements of Theory of Markov Processes and Their Applications.* McGraw-Hill, New York.

[28] Billard, L. and Mohamed, F.Y. (1991). Estimation of the parameters of an EAR(p) process. *J. Time Series Analysis,* **12**, 179-192.

[29] Billingsley, P. (1961a). *Statistical Inference for Markov Processes.* Statistical Research Monographs, Vol. II. The University of Chicago Press, Chicago, Ill.

[30] Billingsley, P. (1961b). Statistical methods in Markov chains. *Ann. Math. Statist.,* **32**, 12-40.

[31] Bloch, D. (1966). A note on the estimation of the location parameter of the Cauchy distribution. *J. Amer. Statist. Assoc.,* **61**, 852-855.

[32] Bohman, H. (1970). A method to calculate the distribution function when the characteristic function is known. *Nordisk Tidskr. Informationsbehandling (BIT),* **10**, 237-242.

[33] Bohman, H. (1972). From characteristic function to distribution function via Fourier analysis. *Nordisk Tidskr. Informationsbehandling (BIT),* **12**, 279-283.

[34] Bohman, H. (1975). Numerical inversions of characteristic functions. *Scand. Actuar. J.,* **2**, 121-124.

[35] Bondesson, L. (1992). *Generalized Gamma Convolutions and Related Classes of Distributions and Densities.* Springer-Verlag, New York.

[36] Breiman, L. (1992). *Probability.* SIAM, Philadelphia.

[37] Brockwell, P.J., Gani, J., and Resnick, S.I. (1982). Birth, immigration and catastrophe processes. *Adv. Appl. Prob.* **14**, 709-731.

[38] Brockwell, P.J. (1985). The extinction time of a birth, death and catastrophe process and of a related diffusion model. *J. Appl. Prob.* **23**, 851-858.

[39] Brockwell, P.J. (1986). The extinction time of a general birth and death process with catastrophes. *Adv. Appl. Prob.* **17**, 42-52.

[40] Brockwell, P.J. and Davis, R.A. (1996). *Introduction to Time Series and Forecasting.* Springer-Verlag, New York.

[41] Casella G. and Berger R. L. (1990). *Statistical Inference.* Duxbury Press, Belmont, CA.

[42] Chaganty, N. R. (1997). An alternative approach to the analysis of longitudinal data via generalized estimating equations. *Journal of Statistical Planning and Inference.* **63**, 39-54.

[43] Chambers, J.M., Cleveland, W.S., Kleiner, B., and Tukey, P.A. (1983). *Graphical Methods for Data Analysis.* Duxbury Press, Boston.

[44] Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics.* Wiley, New York.

[45] Chung, K.L. (1974). *A Course in Probability Theory.* Second edition. Probability and Mathematical Statistics, Vol. 21. Academic Press, New York-London.

[46] Chung, K.L. and Williams, R.J. (1990). *Introduction to Stochastic Integration.* Birkhäuser, Boston.

[47] Cleveland, W.S. and McGill, R. (1984). The many faces of a scatterplot. *J. Amer. Statist. Assoc.,* **79**, 807-822.

[48] Consul, P.C. (1989). *Generalized Poisson Distributions: Properties and Applications.* Marcel Dekker, New York.

[49] Cumberland, W.G. and Sykes, Z.M. (1982). Weak convergence of an autoregressive process used in modeling population growth. *J. Appl. Probab.* **19**, 450-455.

[50] Davies, R.B. (1973). Numerical inversion of a characteristic function. *Biometrika,* **60**, 415-417.

[51] Dewald, L.S. and Lewis, P.A.W. (1985). A new Laplace second-order autoregressive time series model-NLAR(2). *IEEE trans. Information Theory,* **31**, 645-651.

[52] Doob, J.L. (1953). *Stochastic Processes.* Wiley, New York.

[53] Doss, D.C. (1979). Definition and characterization of multivariate negative binomial distributions. *J. Multivariate Anal.*, **9**, 460-464.

[54] DuMouchel, W.H. (1973). Stable distributions in statistical inference. I. Symmetric stable distributions compared to other symmetric long-tailed distributions. *J. Amer. Statist. Assoc.,* **68**, 469-477.

[55] DuMouchel, W.H. (1975). Stable distributions in statistical inference. II. Information from stably distributed samples. *J. Amer. Statist. Assoc.,* **70**, 386-393

[56] Dunlop, D.D. (1994). Regression for longitudinal data: a bridge from least squares regression. *American Statistician,* **48,** *No.* **4.**, 299-303.

[57] Dunn, K.P. and Smyth, G.K. (1996). Randomized quantile residuals. *J. Comput. Graph. Statist.,* **5**, 1-10.

[58] Feller, W. (1966a). *An Introduction to Probability Theory and Its Applications.* Volume **I**, 2nd ed. Wiley, New York.

[59] Feller, W. (1966b). *An Introduction to Probability Theory and Its Applications.* Volume **II**. Wiley, New York.

[60] Ferguson, T.S. (1996). *A Course in Large Sample Theory.* Chapman & Hall, London.

[61] Feuerverger, A. and Mureika, R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.,* **5**, 88-97.

[62] Feuerverger, A. and McDunnough, P. (1981a). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc., Ser. B,* **43**, 20-27.

[63] Feuerverger, A. and McDunnough, P. (1981b). On some Fourier methods for inference. *J. Amer. Statist. Assoc.,* **76**, 379-387.

[64] Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *Canad. J. Statist.,* **18**, 155-161.

[65] Folland, G.B. (1984). *Real Analysis – Modern Techniques and Their Applications.* Wiley, New York.

[66] Freeland, R.K. (1998). *Statistical Analysis of Discrete Time series with Application to the Analysis of Workers' Compensation Claims Data.* Ph.D Thesis. Management Science Division, Faculty of Commerce and Business Administration, University of British Columbia.

[67] Fritz, J. (1981). *Partial Differential Equations.* Springer-Verlag, New York.

[68] Gaver, D.P. and Lewis, P.A.W. (1980). First-order autoregressive gamma sequences and point processes. *Adv. Appl. Probab.* **12**, 727–745.

[69] Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika,* **38**, 481-482.

[70] Goodman, L.A. (1958). Simplified runs tests and likelihood ratio tests for Markoff chains. *Biometrika,* **45**, 181-197.

[71] Graham, R.L., Knuth, D.E. and Patashnik, O. (1994). *Concrete Mathematics: A Foundation for Computer Science.* Second edition. Addison-Wesley Publishing Company, MA.

[72] Granger, C.W.J. (1963). A quick test for serial correlation suitable for use with non-stationary time series. *J. Amer. Statist. Assoc.,* **58**, 728-736.

[73] Griffiths, R.C., Milne, R.K. and Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Austral. J. Statist.* **21**, 238–255.

[74] Griffiths, R.C. and Milne, R.K. (1987). A class of infinitely divisible multivariate negative binomial distributions. *J. Mult. Anal.* **22**, 13–23.

[75] Grimmett, G.R. and Stirzaker, D.R. (1992). *Probability and Random Processes.* Clarendon Press, Oxford.

[76] Harvey, A.C. and Fernandes, C. (1989). Time series models for count data or qualitative observations. *J. Business Economic Statistics,* **7**, 407-422.

[77] Hsu, Y. and Park, W.J. (1988). Ornstein-Uhlenbeck process. In *Encyclopedia of Statistical Sciences.* Vol. **6**, p. 518-521. Chief-editor: Kotz, S. & Johnson, N. L. Wiley, New York.

[78] Hunter, J.J. (1983). *Mathematical techniques of Applied Probability. Vol. 1. Discrete Time Models: Basic Theory.* Academic Press, New York.

[79] Hutchinson, T.P. and Lai, C.D. (1990). *Continuous Bivariate Distributions, Emphasising Applications.* Rumsby Scientific Publishing, Adelaide, South Australia.

[80] Hutton, J.L. (1990). Non-negative time series models for dry river flow. *J. Appl. Prob.* **27**, 171-182.

[81] Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419-426.

[82] Iosifescu, M. and Tăutu, P. (1973). *Stochastic Processes and Applications in Biology and Medicine (II), Models.* Springer-Verlag, Berlin.

[83] Jacod, J. and Protter, P. (2000). *Probability Essentials.* Marcel Dekker, New York.

[84] Jagerman, D.L. (2000). *Difference Equations with Applications to Queues.* Springer, New York.

[85] Jayakumar, K. and Pillai, R.N. (1993). The first-order autoregressive Mittag-Leffler process. *J. Appl. Prob.* **30**, 462-466.

[86] Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *J. Appl. Probab.* **33**, 664–677.

[87] Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

[88] Johnson, N.L. and Kotz, S. (1969). *Distribution in Statistics, Discrete Distributions.* Wiley, New York.

[89] Johnson, N.L. and Kotz, S. (1970a). *Distribution in Statistics, Continuous Univariate Distributions-I.* Wiley, New York.

[90] Johnson, N.L. and Kotz, S. (1970b). *Distribution in Statistics, Continuous Univariate Distributions-II.* Wiley, New York.

[91] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions.* Wiley, New York.

[92] Johnson, R.A. and Wichern, D.W. (1998). *Applied Multivariate Statistical Analysis.* Fourth edition. Prentice Hall, Inc., Upper Saddle River, NJ.

[93] Jones, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics,* **22**, 389-395.

[94] Jones, R.H. (1993). *Longitudinal Data with Serial Correlation: A State–Space Approach.* Chapman & Hall, London.

[95] Jørgensen, B. (1986). Some properties of exponential dispersion models. *Scand. J. Statist.* **13**, 187-198.

[96] Jørgensen, B. (1987). Exponential dispersion models (with discussion). *J. Roy. Statist. Soc. B* **49**, 127-162.

[97] Jørgensen, B., Seshadri, V. and Whitmore, G.A. (1991). On the mixture of the inverse Gaussian distribution with its complementary reciprocal. *Scand. J. Statist.* **18**, 77-89.

[98] Jørgensen, B. (1992). Exponential dispersion models and extensions: a review. *Int. Statist. Rev.* **60**, 5-20.

[99] Jørgensen, B. (1997). *The Theory of Dispersion Models.* Chapman & Hall, London.

[100] Jørgensen, B. and Song, P. (1998). Stationary time series models with exponential dispersion model margins. *J. Appl. Probab.* **35**, 78-92.

[101] Jurek, Z.J. and Vervaat, W. (1983). An integral representation for self-decomposable Banach space valued random variables. *Z. Wahrscheinlichkeitsth.* **62**, 247-262.

[102] Karlin, S. and Taylor, H.M. (1975). *A First Course in Stochastic Processes*, 2nd edition. Academic Press, San Diego.

[103] Karlin, S. and Taylor, H.M. (1981). *A Second Course in Stochastic Processes.* Academic Press, San Diego.

[104] Taylor, H.M. and Karlin, S. (1998). *An Introduction to Stochastic Modeling*, Third edition. Academic Press, San Diego.

[105] Kendall, D.G. (1948). On generalised "birth and death" processes. *Ann. Math. Statist.* **19**, 1-15.

[106] Kendall, D.G. (1949). Stochastic process and population growth. *J. Roy. Statist. Soc., B* **11**, 230-264.

[107] Kendall, D.G. (1960). Birth and death processes and the theory of carcinogenesis. *Biometrika* **47**, 13-21.

[108] Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 3, 629-642.

[109] Kocherlakota, S., and Kocherlakota, K. (1992). *Bivariate Discrete Distributions.* Marcel Dekker, New York.

[110] Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). *Continuous Multivariate Distributions. Volume 1: Models and Applications.* Second edition. Wiley, New York.

[111] Küchler, U. and Sørensen M. (1997). *Exponential Families of Stochastic Processes.* Springer, New York.

[112] Lancaster, P. (1969). *Theory of Matrices.* Academic Press, New York.

[113] Lawrance, A.J. and Lewis, P.A.W. (1980). The exponential autoregressive-moving average EARMA(p,q) process. *J. R. Statist. Soc.* B **42**, 150–161.

[114] Lawrance, A.J. (1982). The innovation distribution of a gamma distributed autoregressive process. *Scand. J. Statist.* **9**, 234-236.

[115] Leitch, R.A. and Paulson, A.S. (1975). Estimation of stable law parameters: stock price behavior application. *J. Amer. Statist. Assoc.,* **70**, 690-697.

[116] Lewis, P.A.W. (1983). Generating negatively correlated gamma variates using the beta-gamma transform. In *Proc. 1983 Winter Simulation Conf.* ed. S. Roberts, J. Banks and B. Schmeiser. IEEE Press, New York. pp. 175-176.

[117] Lewis, P.A.W., McKenzie, E. and Hugus, D.K. (1989). Gamma processes. *Commun. Statist. Stoch. Models* **5**, 1-30.

[118] Lindsey, J.K. (1997). *Applying Generalized Linear Models.* Springer, New York.

[119] Lukacs, E. (1975). *Stochastic Convergence.* Academic Press, New York.

[120] Lukacs, E. (1970). *Characteristic Functions.* Hafner, New York.

[121] MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series.* Chapman & Hall, London.

[122] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* Chapman & Hall, London.

[123] McDunnough, P. (1979). Estimating the law of randomly moving particles by counting. *J. Appl. Probab.,* **16**, no. 1, 25-35.

[124] McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin* **21**, 645-650.

[125] McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Adv. Appl. Probab.* **18**, 679-705.

[126] McKenzie, E. (1987). Innovation distributions for gamma and negative binomial autoregressions. *Scand. J. Statist.* **14**, 79-85.

[127] McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Probab.* **20**, 822-835.

[128] Mitrinović, D.S. and Vasić, P.M. (1970). *Analytic Inequalities.* Springer-Verlag, Berlin, New York.

[129] Nanthi, K. (1983). *Statistical Estimation for Stochastic Processes.* Queen's Papers in Pure and Applied Mathematics, 62. Queen's University, Kingston, ON.

[130] Nanthi, K. and Wasan, M.T. (1987). *Statistical Estimation for Stochastic Processes.* Queen's Papers in Pure and Applied Mathematics, 78. Queen's University, Kingston, ON.

[131] Nash, J.C. (1990). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation.* 2nd ed., Hilger, Bristol, New York.

[132] Nelson, E. (1967). *Dynamical theories of Brownian motion.* Princeton University Press, Princeton, N.J.

[133] Neftci, S.N. (1996). *An Introduction to the Mathematics of Financial Derivatives.* Academic Press, San Diego.

[134] NRC (National Research Council) (1992). *Rethinking the Ozone Problem in Urban and Regional Air Pollution.* National Academy Press, Washington, District of Columbia.

[135] Øksendal, B. (1995). *Stochastic Differential Equations: An Introduction with Applications.* Springer, Berlin.

[136] Ornstein, L.S., Uhlenbeck, G.E. (1930). On the theory of Brownian motion. *Physical Review,* **36**, 823-841. Reprinted in: Selected papers on noise and stochastic processes, ed. N. Wax. Dover, New York, 1954.

[137] Pakes, A.G. (1986). The Markov branching-catastrophe process. *Stochastic Proc. Appl.,* **23**, 1-33.

[138] Patankar, V.N. (1954). The goodness of fit of frequency distributions obtained from stochastic processes. *Biometrika,* **41**, 450-462.

[139] Paulson, A.S., Holcomb, E.W. and Leitch, R.A. (1975). The estimation of the parameters of the stable laws. *Biometrika,* **62**, 163-170.

[140] Petkovšek, M., Wilf, H.S. and Zeilberger, D. (1996). $A = B$. A K Peters, Wellesley, MA.

[141] Phatarfod, R.M. and Mardia, K.V. (1973). Some results for dams with Markovian inputs. *J. Appl. Probab.* **10**, 166-180.

[142] Pillai, R.N. (1990). On Mittag-Leffler functions and related distributions. *Ann. Inst. Statist. Math.* **42,** 157-161.

[143] Pillai, R.N. and Jayakumar, K. (1995). Discrete Mittag-Leffler distributions. *Statistics & Probability Letters,* **23**, 271-274.

[144] Prabhu, N.U. (1980). *Stochastic Storage Processes: Queues, Insurance Risk, and Dams.* Springer-Verlag, New York.

[145] Press, S.J. (1972). Estimation in univariate and multivariate stable distributions. *J. Amer. Statist. Assoc.,* **67**, 842-846.

[146] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1996). *Numerical Recipes in Fortran 77 and Fortran 90. The Art of Scientific and Parallel Computing.* Second edition. Cambridge University Press, Cambridge.

[147] Priestley, M.B. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

[148] Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach.* Springer, New York.

[149] Rao, P.S. and Johnson, D.H. (1988). A first-order AR model for non-Gaussian time series. *Proceedings of IEEE International Conference on ASSP,* **3**, 1534-1537.

[150] Rubinstein, R.Y. (1981). *Simulation and the Monte Carlo method.* Wiley, New York.

[151] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance.* Chapman & Hall, New York.

[152] Sato, K. and Yamazato, M. (1983). Stationary processes of Ornstein-Uhlenbeck type. In *Probability Theory and Mathematical Statistics: Fourth USSR-Japan Symposium Proceedings, 1982*, edited by K. Itô and J.V. Prokhorov. (Lecture Notes in Mathematics. **1021**). Springer-Verlag, Berlin. 1983.

[153] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, London.

[154] Schiff, J.L. (1999). *The Laplace Transform: Theory and Applications.* Springer, New York.

[155] Schilling, M.F. and Watkins, A.E. (1994). A suggestion for sunflower plots. *American Statistician,* **48**, 303-305.

[156] Schuss, Z. (1988). Stochastic differential equations. In *Encyclopedia of Statistical Sciences.* **8**, p. 801-812. Chief-editors: Kotz, S. & Johnson, N. L. Wiley., New York.

[157] Sen, P.K. and Singer, J.M. (1993). *Large Sample Methods in Statistics. An Introduction with Applications.* Chapman & Hall, London.

[158] Seshadri V. (1993). *The Inverse Gaussian Distribution: A Case Study in Exponential Families.* Clarendon Press, Oxford.

[159] Seshadri V. (1999). *The Inverse Gaussian Distribution: Statistical Theory and Applications.* Springer, New York.

[160] Sheehan, D.P. (1983). Approximating estimators of the first-order autoregression. *J. Statist. Comput. Simul.* **18**, 15-43.

[161] Shults, J. and Chaganty, N.R. (1998). Analysis of serially correlated data using quasi-least squares. *Biometrics.* **54**, 1622-1630.

[162] Sim, C.H. and Lee, P.A. (1989). Simulation of negative binomial processes. *J. Statist. Simul. Computation* **34**, 29–42.

[163] Sim, C.H. (1990). First-order autoregressive models for gamma and exponential processes. *J. Appl. Prob.* **27**, 325-332.

[164] Sim, C.H. (1993). First-order autoregressive logistic processes. *J. Appl. Prob.* **30**, 467-470.

[165] Sim, C.H. (1994). Modelling non-normal first-order autoregressive time series. *J. Forecasting, Vol.* **13**, 369-381.

[166] Song, P. (1996). *Some Statistical Models for the Multivariate Analysis of Longitudinal Data.* Ph.D Thesis. Department of Statistics, University of British Columbia.

[167] Stefanski, L.A. (1991). A normal scale mixture representation of the logistic distribution. *Statist. Probab. Lett.,* **11**, 69-70.

[168] Steutel, F.W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7**, 893-899.

[169] Steutel, F.W. (1979). Infinite divisibility in theory and practice. *Scand. J. Statist.* **6**, 57-64.

[170] Steutel, F.W., Vervaat, W. and Wolfe, S.J. (1983). Integer-valued branching processes with immigration. *Adv. Appl. Prob.* **15**, 713-725.

[171] Tavaré, Simon (1983). Serial dependence in contingency tables. *J. Roy. Statist. Soc., Ser. B,* **45**, 100-106.

[172] Thorin, O. (1977a). On the infinite divisibility of the Pareto distribution. *Scand. Actuarial J.* **1977**, 31-40.

[173] Thorin, O. (1977b). On the infinite divisibility of the lognormal distribution. *Scand. Actuarial J.* **1977**, 121-148.

[174] Thorin, O. (1978). An extension of the notion of a generalized Γ- convolution. *Scand. Actuarial J.* **1978**, 141-149.

[175] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Clarendon Press Oxford, New York.

[176] Tweedie, M.C.K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proc. Cambridge Phil. Soc.* **49**, 41-49.

[177] Urbanik, K. (1972). Limit laws for sequences of normed sums satisfying some stability conditions. In *Multivariate Analysis* III. Ed. P.R. Krishnaiah, 225-237. Academic Press, New York.

[178] Ushakov, N.G. (1999). *Selected Topics in Characteristic Functions.* VSP BV, Utrecht, the Netherlands.

[179] van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

[180] van Harn, K., Steutel, F.W. and Vervaat, W. (1981). Self-decomposable discrete distribution and branching processes. In *Analytical Methods in Probability Theory (Oberwolfach, 1980), pp. 60-64, Lecture Notes in Math., 861*, Springer, Berlin-New York.

[181] van Harn, K., Steutel, F.W. (1985). Integer-valued self-similar processes. *Comm. Statist. Stochastic Models,* **1**, 191-208.

[182] Venables, W.N. and Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

[183] Vervaat, W. (1979). On a stochastic difference equation and a representation of positive infinitely divisible random variables. *Adv. Appl. Probab.* **11**, 750-783.

[184] Walker, S.G. (2000). A note on the innovation distribution of a gamma distributed autoregressive process. *Scand. J. Statist.,* **27**, 575-576.

[185] Wei, C.Z. and Winnicki, J. (1990). Estimation of the means in the branching process with immigration. *Ann. Statist.*, Vol **18**, 1757-1773.

[186] Whittaker, E.T. and Watson, G.N. (1927). *A Course of Modern Analysis.* Cambridge University Press.

[187] Winkelmann, R. (1995). Duration dependence and dispersion in count data. *J. Bus. Economic Statistics.* **13**, 467-474.

[188] Winkelmann, R. (1996a). A count data model for gamma waiting times. *Statistical Papers.* **37**, 177-187.

[189] Winkelmann, R. (1997). *Econometric Analysis of Count Data.* Springer, Berlin.

[190] Wolfe, S.J. (1982). On a continuous analogue of the stochastic difference equation $X_n = \rho X_{n-1} + B_n$. *Stoch. Proc. Appl.* **12**, 301-312.