

**Statistical Modelling and Inference for
Discrete and Censored Familial Data**

by

Yinshan Zhao

B.Sc, USTC 1993

M.A., York University 1998

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

Mar 2004

© Yinshan Zhao, 2004

Abstract

Analysis of familial data with quantitative traits based on the multivariate normal distribution has been well studied. However, little attention has been devoted to traits which do not have a multivariate normal distribution, such as traits with discrete or censored values. In this thesis, we devote our effort to (1) construct models for familial data when the trait value is discrete and/or censored, and (2) study alternative estimation methods when maximum likelihood estimation is infeasible. We discuss two existing classes of models: models with random effects which are multivariate normally distributed, and models constructed from the multivariate normal copula. These two classes include a variety of models which can be applied to familial data. We also propose another class of models which we call conditional independence models. This type of model is based on a conditional independence assumption: for a trait variable, we assume independence of a pair of non-sibling relatives conditional on their parents, so that the dependence structure is built on the Markov property.

Maximum likelihood estimates are generally difficult to obtain for random effect models and copula models when there are large families involved. We propose two estimation procedures based on composite likelihoods: the first is a two-stage method in which univariate marginal parameters are estimated based on univariate marginal distributions and the dependence parameters are estimated separately based on bivariate marginal distributions with the marginal parameters treated as known; whereas in the second, all the parameters are estimated using the likelihoods of bivariate marginal distributions. The composite likelihood methods can greatly reduce computation in parameter estimation, but with a price of efficiency loss. In this thesis, extensive investigations based on asymptotic covariance matrices and simulations were carried out to compare the asymptotic efficiency of these two procedures with the maximum likelihood method. In our efficiency

comparisons, we investigate the multivariate normal model for a continuous trait, the multivariate probit model for a binary trait, the multivariate Poisson-lognormal mixture model for a count trait and multivariate lognormal model for a censored variable. We found that when the dependence is strong, the first approach is inefficient for the regression parameters; whereas when the dependence is weak, the second approach is inefficient for the dependence parameters.

In many familial analyses, quantifying familial association is of great interest. For a binary trait, the odds ratio may be used as a measure of association between a parent-offspring pair or a sibling pair. We develop theories so that the asymptotic variance of an odds ratio can be computed from a 2×2 contingency table formed by dependent pairs.

Contents

Abstract	ii
Contents	iv
List of Tables	viii
List of Figures	x
Glossary	xii
Acknowledgements	xiii
Dedication	xiv
1 Introduction	1
1.1 Familial Data and Notation	1
1.2 Genetics Background	5
1.3 Neurofibromatosis Type 1 and 2	6
1.4 Motivation and Background for Statistical Models	6
1.5 Outline of the Thesis	9
2 Models for Non-normal Familial Data	11
2.1 General Introduction of Model Classes	12
2.1.1 Models with Multinormal Random Effects (Mixture Models)	12
2.1.2 Multinormal Copula Models	14
2.1.3 Conditional Independence (CI) Model	16

2.2	Binary Traits	21
2.2.1	Multivariate Probit (MVP) Model	22
2.2.2	Two-Component Model	22
2.3	Count Traits	31
2.3.1	Poisson-Lognormal Mixture	31
2.3.2	Multinormal Copula Model for Count Traits	32
2.3.3	Conditional Independence Models for Count Traits	33
2.4	Survival Data	39
2.4.1	Multivariate Lognormal Models	40
2.4.2	Multivariate Lognormal Frailty	40
2.4.3	Multinormal Copula Model	41
3	Estimating Procedures Generated from Composite Likelihood Functions	42
3.1	Notation	44
3.2	Estimation Procedures based on Composite Likelihood	46
3.2.1	General Properties	46
3.2.2	A Two-stage Estimating Procedure (CL1)	48
3.2.3	Estimating Equations Based on BCL (CL2)	60
3.3	Estimating the Asymptotic Covariance Matrix of $\hat{\theta}$	65
3.4	Remarks	65
4	Efficiency Comparisons	68
4.1	Continuous Response: MVN model	69
4.1.1	MLE	69
4.1.2	Efficiency of the Two-Stage Estimating Approach (CL1)	70
4.1.3	Estimates Based on Bivariate Composite Likelihood (CL2)	82
4.2	Binary Response: MVP Model	89
4.2.1	Likelihood and Composite Likelihood Estimating Functions	89
4.2.2	Efficiency Comparison	91
4.2.3	Conclusion	101
4.3	Count Response: Multivariate Poisson-Lognormal Mixture Model	101

4.3.1	ML and CL estimating functions	101
4.3.2	Efficiency Comparison	102
4.4	Survival Data Subject to Right Censoring: Multivariate Lognormal Model	104
4.5	Conclusion	106
5	Inferences for Log Odds Ratio with Dependent Pairs	108
5.1	Interclass Odds Ratio	110
5.1.1	Estimator of γ_{po}	111
5.1.2	Asymptotic Variance of $\hat{\gamma}_{po}$	111
5.1.3	Methods to Estimate the Asymptotic Variance of $\hat{\gamma}_{po}$ in (5.5)	116
5.1.4	Asymptotic Relative Efficiency of $\hat{\gamma}_{po}$	117
5.2	Intraclass Odds Ratio	120
5.2.1	The Estimator of γ_{ss}	120
5.2.2	Asymptotic Variance of $\hat{\gamma}_{ss}$	121
5.2.3	Weighted Estimator of γ_{ss}	124
5.2.4	Methods to Estimate the Asymptotic Variance of $\hat{\gamma}_{ss}$	125
5.2.5	Efficiency Comparison	126
5.3	Discussion	129
6	Applications to NF1 and NF2 Data	130
6.1	Intertriginous Freckling in NF1 Patients	131
6.1.1	Interclass Odds Ratio	131
6.1.2	Intraclass Odds Ratio	132
6.2	Presence of Subcutaneous Neurofibromas in NF1 Patients	133
6.2.1	Summary of Patients	134
6.2.2	Probit Model	134
6.2.3	Two-Component Series Model	136
6.2.4	Comparison of the Probit Model and the Two-Component Series Model . . .	137
6.3	Age at First Symptom of NF2	139
6.3.1	Summary of Patients	139
6.3.2	Model	139

6.3.3	Results	140
6.4	Number of Intracranial Meningiomas in NF2 Patients	141
6.4.1	Summary of Patients	141
6.4.2	Multinormal Copula Model with Negative Binomial Margin for Number of Meningiomas	142
6.4.3	Multivariate Poisson-Lognormal Mixture Model for Number of Meningiomas	144
6.4.4	Comparison of the Two Models	144
6.5	Discussion	146
7	Conclusion and Future Work	147
7.1	Conclusion	147
7.2	Future Work	147
Appendix A	Dependence Structure of Quantitative Traits	150
A.1	Mendelian Models	151
A.2	Polygenic Models	153
A.3	Major Gene Models	154
A.4	Models for Autosomal-Dominant Diseases	155
A.5	Comments	156
Appendix B	Appendix for Chapter 4	157
B.1	General Results	157
B.2	Asymptotic Properties of the Covariates	158
B.3	Maple Code	160

List of Tables

4.1	AREs of CL1 for Type-3 families with equal number of offspring (MVN model) . . .	82
4.2	Range of $\Pr(Y = 1)$ for each combination of β_0 and β_1 (MVP model).	92
4.3	ARE of the CL1 and CL2 estimators for the exchangeable case at different levels of β_1 and ρ (MPLN model).	103
4.4	ARE of the CL1 and CL2 estimators for the exchangeable case at different levels of σ^2 and ρ (MVPL model).	104
4.5	ARE of the CL1 and CL2 estimators for the exchangeable case with family size varying from 1 to 4 (MVN model with right censoring).	105
5.1	Contingency table formed by independent parent-offspring pairs	109
6.1	Interclass odds ratio in intertriginous freckling	132
6.2	Intraclass odds ratio of intertriginous freckling	133
6.3	Number of affected pairs included in the analysis of presence of subcutaneous neurofibromas in NF1 patients	134
6.4	Parameter estimates of the Probit model for the presence of subcutaneous neurofibromas in NF1 patients. The three dependence parameters are: ρ_{ss} (sib-sib correlation), ρ_{po} (parent-offspring correlation) and ρ_{2+} (correlation between any relative pair higher than 2nd degree).	135
6.5	Parameter estimates of the Probit model with CI structure for the presence of subcutaneous neurofibromas in NF1 patients.	135
6.6	The ML estimates of the two-component series model for presence of subcutaneous neurofibromas in NF1 patients	137

6.7	Probabilities and correlations of the heritable component in the two-component series model for the presence of subcutaneous neurofibromas in NF1 patients. Index l is the index of a parent and j, j' are the indices of l 's two children.	137
6.8	Number of affected pairs included in the analysis of age at first symptom of NF2 . .	139
6.9	Parameter estimates of the MVN model for age at first symptom of NF2	140
6.10	Number of affected pairs included in the analysis of number of meningiomas in NF2 patients.	142
6.11	Summary of meningioma counts in NF2 patients	142
6.12	Parameter estimates of the multinormal copula model with negative binomial margin for number of meningiomas in NF2 patients	143
6.13	Parameter estimates of the multivariate Poisson-lognormal mixture model for number of meningiomas in NF2 patients	144
6.14	Expected number of meningiomas based on the multinormal copula model with negative binomial margin (NBMC) and the multivariate Poisson-lognormal mixture (MPLN) model	145
A.1	Covariances between relatives in a simple Mendelian model	152

List of Figures

1.1	Diagram of a Type-I pedigree	3
1.2	Diagram of a Type-II pedigree	4
2.1	Path from i to j . If j_1 and j_2 are full siblings, the path is $(i, \dots, j_1, j_2, \dots, j)$; whereas if j_1 and j_2 are half siblings, the path is $(i, \dots, j_1, j_0, j_2, \dots, j)$	20
4.1	Efficiency of CL1 for exchangeable dependence structure. (MVN model)	77
4.2	Efficiency of CL1 for Type-3 families with $k^* = 3$ offspring (MVN model); $\tilde{\rho}_2 = (\rho_2 - \rho_1^2)/(1 - \rho_1^2)$ is the sib-sib correlation conditional on their parent.	78
4.3	ARE of $\hat{\beta}_{1,CL1}$ against average number of offspring, \bar{k}^* (MVN model).	80
4.4	Efficiency of $\hat{\beta}_{0,CL1}$, $\hat{\sigma}_{CL1}^2$, $\hat{\rho}_{1,CL1}$ and $\hat{\rho}_{2,CL1}$ (MVN model).	81
4.5	Efficiency of CL1 and BCL for the exchangeable case. Families are of size two and six with proportion 3:1 (MVN model).	83
4.6	Efficiency of $\hat{\rho}_{CL1}$, $\hat{\rho}_{BCL}$ and $\hat{\rho}_{CL2}$ (MVN model); family sizes are 2 or 6 with proportion 3:1. The CL2 estimate is weighted by family size; whereas the BCL estimate is not.	87
4.7	Efficiency CL1 and CL2 for Type-3 families. The family size varies; the number of offspring is from 1 to 5 with equal proportion (MVN model).	88
4.8	ARE of $\hat{\beta}_{0,CL1}$ and $\hat{\beta}_{1,CL1}$ when family size $k = 2$. (MVP model)	93
4.9	ARE of $\hat{\beta}_{CL1}$, $\hat{\beta}_{WCL1}$ and $\hat{\beta}_{CL2}$ for the exchangeable case with family size varying from 1 to 4 (MVP model).	95
4.10	ARE of $\hat{\rho}_{CL1}$, $\hat{\rho}_{WCL1}$ and $\hat{\rho}_{CL2}$ for the exchangeable case with family size varying from 1 to 4. (MVP model)	96

4.11	ARE of $\hat{\beta}_{CL1}$, $\hat{\beta}_{WCL1}$ and $\hat{\beta}_{CL2}$ for Type-3 families, at $\rho_1 = 0.3$ (MVP model). Solid line: CL1, dashed line: WCL1, dotted line: CL2.	97
4.12	ARE of $\hat{\rho}_{CL1}$, $\hat{\rho}_{WCL1}$ and $\hat{\rho}_{CL2}$ at $\rho_1 = 0.3$ for Type-3 families (MVP model). Solid line: CL1, dashed line: WCL1, dotted line: CL2.	98
4.13	ARE of $\hat{\beta}_{CL1}$, $\hat{\beta}_{WCL1}$ and $\hat{\beta}_{CL2}$ for Type-3 families, at $\rho_1 = 0.6$ (MVP model). Solid line: CL1, dashed line: WCL1, dotted line: CL2.	99
4.14	ARE of $\hat{\rho}_{CL1}$, $\hat{\rho}_{WCL1}$ and $\hat{\rho}_{CL2}$ for Type-3 families, at $\rho_1 = 0.6$ (MVP model). Solid line: CL1, dashed line: WCL1, dotted line: CL2.	100
5.1	ARE of $\hat{\gamma}_{po}$ based on a multivariate binary beta-binomial model when $\alpha_1 = \beta_0 =$ $\alpha, \alpha_0 = \beta_1 = \beta$	118
5.2	ARE of $\hat{\gamma}_{po}$ based on the multivariate Probit model.	119
5.3	Efficiency of $\hat{\gamma}_{po}$ versus the variance-mean ratio (V/M) of family size	120
5.4	Efficiencies of γ_{ss} estimators (multivariate binary beta-binomial model). Solid line: no weights, dotted line: optimal weights, dashed line: simple weights.	127
5.5	Efficiencies of γ_{ss} estimators (multivariate Probit model). Solid line: no weights, dotted line: optimal weights, dashed line: simple weights.	128
5.6	Efficiency comparison of γ_{ss} estimators with varied proportion of families with size 2 and 6. Solid line: no weights, dotted line with "+": optimal weights, dashed line: simple weights. In the figure on the right hand side, the dotted line with "+" and the dashed line are coincident.	129
6.1	Region of $P_{11 3}$ and $P_{11 4}$ (intraclass odds ratio).	133
6.2	Predictions of the prevalence of subcutaneous neurofibromatosis in NF1 patients by the probit model and the two-component model.	138

Glossary

1. pmf: probability mass function

pdf: probability density function

cdf: cumulative distribution function

MLE: maximum likelihood estimation (or estimator)

CL: composite likelihood

MVN: multivariate normal

i.i.d.: independent and identically distributed

AVar: asymptotic variance

2. ϕ , Φ : the pdf and cdf of univariate standard normal distribution

$\phi_k(\cdot; \mu, \Sigma)$, $\Phi_k(\cdot; \mu, \Sigma)$: the pdf and cdf of k -variate normal distribution with mean μ and covariance matrix Σ

3. \vee : maximum

\wedge : minimum

\perp : independent

4. \mathfrak{R}^k k -dimensional real space

5. \xrightarrow{p} : converge in probability \xrightarrow{d} converge in distribution

6. \succeq_{pd} : positive definite ordering, i.e. if $\mathbf{A} \succeq_{pd} \mathbf{B}$, then $\mathbf{A} - \mathbf{B}$ is non-negative definite.

Acknowledgements

I would like to express my deep gratitude to my supervisor, Professor Harry Joe. Without his support and guidance, this thesis would not have been completed. Very special thanks are due the other members of my supervisory committee, Professor John Petkau and Professor Nancy Heckman, for their invaluable advice and encouragement. Also, I thank Professor Jan Friedman for his helpful advice and for providing us with the NF data used in this thesis. Additionally, I would like to thank Dr. Gareth Evans and Dr. Michael Baser for providing the family information of the NF2 data. I am thankful for the simulating environment of the department. It is a pleasure to acknowledge helpful discussions with my fellow graduate students: Jochen Brumm, Rachel MacKay, Weiliang Qiu and Rong Zhu and all of the others. Finally, I thank Steven Ruuth for his expertise in numerical methods.

YINSHAN ZHAO

The University of British Columbia

Mar 2004

To my parents and my brother

Chapter 1

Introduction

This thesis is concerned with models and estimation methods for familial data that may be discrete or censored. The major contributions of this thesis are: (1) a new class of models called conditional independence models, (2) estimating approaches for models which are limited in practice due to computational difficulties, and the asymptotic properties of these approaches, and (3) a nonparametric method to estimate familial odds ratios for binary traits. In this chapter, we first introduce the structure of familial data being considered, followed by a brief overview of some genetic background and two genetic disorders which are used as examples in this thesis. Then, we address the challenges in analyses of discrete or censored familial data. An outline of this thesis is given at the end of this chapter.

1.1 Familial Data and Notation

Many genetic studies seek to understand how different genetic and environmental factors affect a trait or condition. These studies examine a sample of a large number of families. The data obtained from such a study come from measuring trait variables and other variables of interest, such as age and treatments, from one or more members of each family, and therefore the data are called familial data. This thesis presents only models for a single response measurement on each individual. This measurement can be any one of several different types: (1) quantitative, such as body size or blood pressure; (2) survival, i.e. a realization of nonnegative random variable, with possible censoring, such as life time or onset age of a disease, normally subject to right censoring;

(3) count (non-negative integer), such as number of tumors; (4) binary, such as absence/presence of certain feature; (5) ordinal, such as level of severity of a disease.

We denote the number of families in the sample by n and the number of members in the i th family by k_i . Then $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ik_i})'$ is the vector of measured values of a single trait on the members of the i th family and $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{ik_i})$ is a $k_i \times m$ covariate matrix containing variables which may influence the trait. When we refer to a family in general, the family index i will be omitted.

For familial data, trait values for individuals in different families are considered independent. Members in the same family are expected to resemble each other more than members in different families. Sometimes, the familial resemblance is the primary interest of a study. Other times, the primary interest rests on other factors. However, in order to obtain correct statistical inferences, the familial dependence has to be taken into account. From the statistical point of view, analyses of these data involve two tasks: first, constructing a statistical model which (1) describes the relation between the trait and the covariates; (2) quantifies the degree of association between different types of relatives, and making statistical inferences based on the model. In this thesis, we focus on non-normal responses, more particularly, survival, binary and count responses. In the rest of this thesis, the term “non-normal responses” refers to these three types of data.

Before we proceed, we define the family types considered in this thesis. In familial analysis, the sample unit is a family. In the literature, “familial data” sometimes only refers to data with nuclear families as the sample units, i.e., to data obtained by sampling independent families containing two parents and their offspring. The familial data considered in this thesis are not limited to nuclear families. The data structure under consideration varies according to the sample unit. Here we define the types of families which may be considered as sample units. The first four types are small families with one or two generations:

Type-1 a pair of relatives: typically twins, parent-offspring pair or sibling pair.

Type-2 a group of full siblings: the size of each group could vary from family to family.

Type-3 one parent and his/her offspring.

Type-4 a nuclear family: two parents and their common offspring.

The remaining two types of sample units are pedigrees with multiple generations. In general, we only consider pedigrees without twins or half siblings. Discussion will be given later on how to generalize our models to include those individuals.

Type-I pedigree: a pedigree formed by family members who share a common ancestor. The spouse of a family member is not included. Thus, only one parent of each individual is included in the pedigree. Such a pedigree can be illustrated by a tree as shown in Figure 1.1. The root of the tree is called the founder of the pedigree who is the ancestor of all the other members. We define the following notation to describe this type of pedigree. For individual i , let $A_i = (i_1, i_2, \dots)$ be the set of i 's direct ancestors with its elements indexed so that i_1 is i 's parent, i_2 is i_1 's parent, and so on. For the founder, $A_i = \emptyset$. For two individuals i and j , we define $A_{i \setminus j} = A_i \setminus A_j$, namely, $A_{i \setminus j}$ is the collection of all the individuals who are the ancestors of i , but not j . If i and j are siblings, $A_{i \setminus j}$ is empty. Otherwise $A_{i \setminus j} = (i_1, \dots, i_{g_{i,j}})$, where $i_{g_{i,j}}$ is the most distant ancestor among all the ancestors of i who are not shared by j . Also we define $A'_{i \setminus j}$ as the same collection but in reverse order. Let D_i be i 's descendants set. Clearly, (i, D_i) is also a tree with i being the root. Let S_i be a sibling group whose parent is i and $T = \{i : S_i \neq \emptyset\}$, i.e. the set of indices corresponding to parents. The parent i combined with his/her offspring (S_i) is called a family unit, the basic unit of a pedigree. Each family unit is a Type-3 family. We index a family unit by the index of the parent and denote it by $U_i = (i, S_i)$. Two units overlap, if and only if an individual is an offspring in one of the units and the parent in another unit. In Figure 1, there are three family units and unit 1 overlaps with both units 2 and 3.

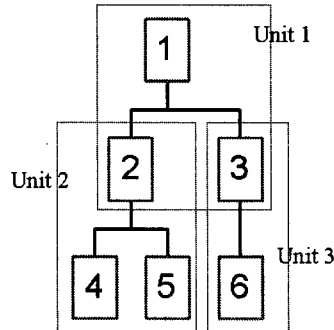


Figure 1.1: Diagram of a Type-I pedigree

Type-II pedigree: a pedigree formed by family members of a common ancestor, plus possibly their spouses. The spouses are called marry-in's and do not have their ancestors in the pedigree. To describe such a pedigree, we need to modify the previous notation. Let m_i be the index of the spouse of individual i . The family unit is (i, m_i, S_i) , a Type-4 family. Such a pedigree is illustrated in Figure 1.2. Suppose individual 1 is the founder. Individual 2, 3 and 6 are marry-ins.

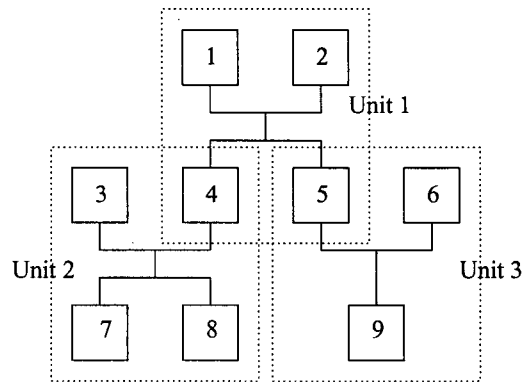


Figure 1.2: Diagram of a Type-II pedigree

The type of sample unit involved in a study depends on the purpose of the study. For example, data with Type-I pedigrees are often collected from a population affected by an autosomal dominant disorder, caused by alteration or mutation of an autosome (i.e. chromosome 1 – 22, non-sex chromosome). Within a family, affected individuals normally share the same mutation. However, commonly even within the same family there is marked variation in the expression of the disorder. The focus of the study would be the reason for the variable expression. On the other hand, data with Type-II pedigrees are collected from a general population. Therefore, Type-I pedigree and Type-II pedigree data require different model assumptions. A model for Type-I pedigrees is not necessarily a special case of a model for Type-II pedigrees.

In data with Type-I pedigrees as sample units, information is sometimes available only on a few individuals, for example a sibling group. In this case, even though the family only contains siblings, it should be considered as a pedigree.

1.2 Genetics Background

To comprehend the genetic basis of familial resemblance, it is crucial to understand the transmission of genetic information from generation to generation both at the individual and population level.

Mendelian laws of transmission tell us how a gene is passed from the parents to their offspring:

During gamete formation, the two alleles of a gene separate from each other and go into different gametes. The segregation of one gene pair is independent of other gene pairs.

(The second part is not true if two genes are closely located on the same chromosome.)

According to Mendel's laws, an offspring receives half of its genetic materials from the father and the other half from the mother. In genetics, the specific allelic composition is called a genotype. Given the parents' genotype, the offspring's genotype is not dependent on other ancestors. The more closely related are two individuals, the more likely they share a gene by descent.

On the population level, we consider a population which is infinite in size and randomly mating with no inbreeding. We will further assume that the effects of migration, selection and mutation on the population are insignificant. Although such an idealized situation is never realized perfectly, it is close enough to the truth for practical purposes. With these assumptions, the gene frequencies and the genotype frequencies remain constant from generation to generation. This characteristic is described by the **Hardy-Weinberg** equilibrium (Lange, 1997).

Genotypes often cannot be easily observed, but their effects are reflected on an observable character or characters (i.e., trait), such as eye-color, or fingerprints. The form of the observable character (or the value of the trait) is called a phenotype. It is important to understand how genotypic difference affects phenotypic difference. There are different mechanisms. The simplest case is that a trait is completely determined by a single gene. Then the phenotypic difference can be explained solely by the genotypic difference. More often, the trait is not just affected by genetic factors, but also affected by random environmental factors. Then there are two sources of variation: genetic and environmental. Normally these two sources are treated independently in a statistical model. In reality, there are few traits determined by a single gene. More generally, a trait can be affected by several genes jointly, or by a large number of genes, each of them with a small effect and assumed to be acting independently. It also can be a combination of the two cases: the trait is affected by one or more genes with major effects and a large number of genes with small effects which are assumed to be additive. Each of these mechanisms leads to different resemblance patterns

among relatives. The basic idea is that if a trait is determined by genes and random environmental factors, the resemblance between relatives is determined by (1) the possibility of their sharing of the responsible genes and (2) the proportion of variation due to genetic factors against that due to environmental factors.

1.3 Neurofibromatosis Type 1 and 2

This thesis is motivated by problems that arise from several studies on two genetic disorders, neurofibromatosis type 1 (NF1) and type 2 (NF2). The methodologies developed in this thesis are illustrated in analyses of different features in NF1 and NF2 patients. Therefore we give a brief introduction of these two disorders. Both NF1 and NF2 are autosomal dominant genetic disorders. Common features of both are nerve tumors. Although both are called neurofibromatosis, their clinical manifestations and genetic origins are distinct. The gene for NF1 is located on chromosome 17. Features, such as intertriginous freckling, neurofibromas of different types, Lisch nodules and learning disabilities are common in NF1 patients. The gene for NF2 is located on chromosome 22. The hallmark of NF2 is a nerve tumor called bilateral vestibular schwannoma, but other tumors such as meningiomas and nonvestibular schwannomas are also common.

Both disorders have highly variable expressivity manifested in many different ways, including variation in age at onset, types and numbers of clinical features, overall disease severity, and rate of progression. Clinical studies indicate that the expression tends to be similar within a family. But substantial phenotypic differences are also observed within families despite the fact that family members share an identical mutant allele. The reason for the variability in the disease expression is unknown. Clearly, variation in the mutant allele itself does not account for all of the variability. Many different genetic and nongenetic causes may exist and act alone or in combination. Comparison of the strength of association between different types of relatives will help us to identify the sources of the variation.

1.4 Motivation and Background for Statistical Models

The statistical analysis of familial association started early in the 20th century. In 1918, Fisher published his first and also one of his most important papers in genetics, *The Correlation Between*

Relatives on the Supposition of Mendelian Inheritance. In this paper, he showed how the phenotypic variation could be partitioned into heritable and non-heritable components, and therefore established the link between the association observed between relatives and the underlying genetic mechanism. This paper along with the work of Haldane (1932) and Wright (1921) laid the foundation of quantitative genetics. Since then, sophisticated models for quantitative traits (continuous variables) have been developed, mainly based on the normal distribution. (A detailed discussion is given in Appendix A.) In addition, there are papers on estimation procedures for interclass and intraclass correlations for familial data, for example, Donner and Koval (1980), Kleffe (1993) and Donner et al. (1998). These estimation procedures are based on the MVN distribution assumption without adjustment for covariates and mainly apply to data with Type-2, Type-3 or Type-4 families. Cases with covariates are also considered by Srivastava and Ng (1990).

As for traits of other kinds, some work is scattered in the literature, but in terms of both quantity and maturity of this work there is no comparison with the well established normal models for quantitative traits. Many open questions remain. Difficulties in analysis of non-normal familial data mostly lie in two aspects. The first aspect is the lack of models which can provide a flexible dependence structure able to reflect the nature of familial resemblance, i.e., the degree to which the dependence of two relatives depends on their biological relationship. Generally, that dependence is stronger for two closely related individuals and weaker for two distantly related individuals. When the families under consideration are arbitrary, the dependence structure varies from family to family. Many models existing in the literature can only be applied to certain types of families. Other models may provide a flexible dependence structure, but give results with no natural interpretation. The second aspect concerns the computational problems in parameter estimation. Many of those multivariate models do not have a simple closed-form likelihood. Therefore, the maximum likelihood estimators (MLEs) are generally difficult to obtain. Even though there are numerical techniques and approximation methods available, those methods may not be stable when the family size becomes large. Below are some examples for binary, count and survival familial data.

The latent threshold method was introduced to analyze a binary trait (Falconer, 1965, 1967). Normality is assumed for the latent variables, leading to a multivariate probit model. A pleasing aspect of this model is that it is well connected to quantitative genetics and therefore provides a meaningful interpretation of the parameters. It also admits a wide range of dependence structures.

Moreover, it can be extended to analyze an ordinal response (Manatunga and Williamson, 2001). However, its drawback is clear as well. When family size is large, likelihood evaluation involves high dimensional numerical integrations. Other approaches such as regressive logistic models (Bonney, 1986) and conditional logistic regression models (Connolly and Liang, 1988; Tosteson et al., 1991; Abel et al., 1993) have been developed. However, as Trégouët et al. (1999) pointed out, the results from these models may depend on the size of the family or the ordering of the family members. They propose a model using a parametric copula which overcomes this problem, but their model only applies to nuclear families.

A general approach to modeling correlated count data is the Poisson mixture model, in which the Poisson parameters are modelled by fixed and random effects (Cameron and Trivedi, 1998). If the random effects are assumed to have a gamma distribution, the joint distribution has a closed form, but with only one level of dependence. To achieve a more flexible dependence structure, multivariate log normally distributed random effects have been considered in modelling animal breeding (Tempelman, 1996; Tempelman and Gianola, 1999). Again, computational difficulties arise in parameter estimation for such a model. The multivariate Poisson model is also used to model bivariate count data (Xu, 1996; Joe, 1997). In such a model, each count variable is the sum of two independent Poisson random variables, one of which is shared within the sample unit. The dependence comes from the shared component. Similar models can be formed with the multivariate negative binomial or the generalized Poisson distributions. However, this type of model is difficult to extend to accommodate the dependence structure of a complex pedigree.

In recent years substantial research has been devoted to developing methodologies for clustered survival data. An example is the univariate frailty model (Vaupel et al., 1979) where the frailties induce an exchangeable dependence structure. Clearly, it is limited by its unduly simple dependence structure. It is possible to allow more than one frailty in the sample unit, so that a nested or hierarchical dependence structure can be obtained (Bandein-Roche and Liang, 1996; Joe, 1993). However, this can only be applied to certain types of familial data, such as Type-3 families. Pickles et al. (1994) as well as Yashin and Iachine (1995) propose a correlated frailty model. In this model, instead of sharing a common frailty, each individual has his own frailty correlated with frailties in the same family. As in the multivariate probit model and log-normal Poisson mixture model, the likelihood function is difficult to evaluate. Later, Korsgaard and Andersen (1998) proposed an

additive genetic gamma frailty model. The frailties are constructed as a summation of independent gamma variables. This model provides a dependence structure similar to the polygenic model (see Appendix A.2). But there is no natural explanation of the decomposition.

In this thesis, we devote our efforts to overcoming these challenges. We consider different modelling approaches which can lead to models that are both practical and interpretable, including random effects models, multivariate normal copula models and a family of new models called conditional independence models. Conditional independence models are constructed based on the assumption that the trait values of two non-sibling relatives are independent conditional on their parents. This approach reduces the task of modelling a complex pedigree to modelling a family unit containing only the parents and their offspring and simplifies the computation in estimation. For models where the MLE method is difficult to apply, we propose two alternative estimating procedures based on composite likelihoods and investigate the asymptotic relative efficiency of the estimates. We also propose a nonparametric method to estimate the interclass and intraclass odds ratios for a binary trait. The next section is an overview of the work included in this thesis.

1.5 Outline of the Thesis

The rest of this thesis is organized as follows.

In Chapter 2, we discuss three different modelling approaches for familial data. The first approach uses random effects to construct the dependence structure, and the second uses the multivariate normal copula. These two existing approaches have been applied to familial data or other multivariate data. In this thesis we provide a comprehensive summary of these models in the context of familial data analysis. Chapter 2 also serves as background for Chapter 3, in which estimation procedures are addressed. We then present the conditional independence models and study the properties of this model family. After the general introduction of the modelling approaches, we study some specific models for binary, count and survival responses.

In Chapter 3, we propose two likelihood-based estimation methods. The first approach is a two-stage method in which univariate marginal parameters are estimated in the first stage based on the likelihoods of the univariate marginal distributions, whereas the dependence parameters are estimated in the second stage based on the bivariate marginal distributions with the marginal

parameters fixed at their estimated values from the first stage. In the second approach, all the parameters are estimated simultaneously based on the likelihoods of the bivariate marginal distributions. Both methods yield asymptotically consistent parameter estimates. In Chapter 4, we investigate the performance of the two methods. The asymptotic efficiencies of the estimates are compared with that of the MLEs when the MLEs can be obtained. Our results show that both methods behave well except that the two-stage method can be inefficient for the estimation of the regression coefficients when the dependence is strong.

Chapter 5 deals with estimation of intraclass and interclass odds ratios for a binary trait. The problem arises from a study in NF1 patients. We propose a nonparametric method to estimate intraclass and interclass odds ratios using relative pairs and derive the asymptotic variances. Asymptotic efficiency is compared with that of the MLE based on a parametric model.

In Chapter 6, the models and inferential approaches studied in the previous chapters are applied to datasets of NF1 and NF2 patients. In Chapter 7, the final chapter, we discuss some future research topics related to this thesis.

Chapter 2

Models for Non-normal Familial Data

In this chapter, we introduce models for non-normal familial data. We will focus on three different types of traits: binary, count and survival. A general assumption underlying the models is the “common univariate margin” assumption: the univariate marginal distribution of each individual belongs to the same family of distributions, say, $Y_{ij} \sim F_1(\cdot; \theta_{ij})$. In familial data analyses, it is common, but not necessary, to assume that the dependence structure between siblings is exchangeable and the order of the siblings is irrelevant to the distribution.

The models introduced in this chapter can be classified into three categories: models with random effects which are multivariate normally distributed, multinormal copula models and conditional independence models. The first two classes of models were originally developed for multivariate data under different circumstances. They are feasible for familial data analysis because they provide a wide range of choices of distributions and a flexible dependence structure. One can find a considerable amount of research on the properties of these models in the literature. Therefore, we provide only a brief summary of those properties which are important to familial data analysis. Models of the first two classes are limited in practice due to computational difficulties. So we develop a third class of models based on a conditional independence assumption. The dependence structure generated by this approach is an approximation to the dependence structure under polygenic effects. The conditional independence assumption enables one to easily carry out the task of model fitting. Models from this class have not been studied in the literature. Therefore, the properties of this class will be discussed in detail.

The first part of this chapter is an overview of the three classes of models. General prop-

erties such as dependence structure and likelihood are discussed for each class. Then we study some specific models for each type of trait we consider, including: multivariate probit (MVP) and two-component models for binary traits; Poisson-lognormal mixture, multinormal copula and conditional independence models for count traits; multivariate lognormal, multivariate lognormal frailty and multivariate copula models for survival data.

2.1 General Introduction of Model Classes

2.1.1 Models with Multinormal Random Effects (Mixture Models)

Definition

A random effects model has a hierarchical structure. The first level specifies the distribution of the observed response vector, \mathbf{Y} , conditional on an unobservable random vector $\mathbf{\Lambda} = (\Lambda_1, \dots, \Lambda_k)$. Given $\mathbf{\Lambda} = \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)$, we assume that Y_1, \dots, Y_k are conditionally independent with pdf (if Y_i is continuous) or pmf (if Y_i is discrete) $g(y_i; \lambda_i, \boldsymbol{\alpha}_i)$. Note that the $\boldsymbol{\alpha}_i$ represents other parameters, if they exist, which are not random. The second level specifies the distribution of $\mathbf{\Lambda}$. We assume that $h(\mathbf{\Lambda}) = \mathbf{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, for some elementwise function h .

Applications and Examples

Such models include the models Tempelman (1996) proposed for animal breeding, in which multivariate normal random effects were incorporated into a negative binomial distribution, and the multivariate lognormal frailty model applied to animal data by Korsgaard, Madsen and Jensen (1998). We describe two examples in the later sections: the multivariate Poisson-lognormal (MPLN) (Section 2.3.1), and the multivariate lognormal frailty models (Section 2.4.2). In both cases, $h(\mathbf{\Lambda}) = (\log \Lambda_1, \dots, \log \Lambda_k)'$. The parameter λ_i is linked to the conditional mean in the MPLN models and to the hazard in the frailty models.

Dependence Structure

The dependence among the Y_i s comes entirely from the dependence among the Z_i s. Let $\nu_i(Z_i) = E(Y_i|\mathbf{Z}) = E(Y_i|Z_i)$. Then

$$\text{Cov}(Y_i, Y_j) = \text{Cov}(E(Y_i|\mathbf{Z}), E(Y_j|\mathbf{Z})) + E(\text{Cov}(Y_i, Y_j|\mathbf{Z})) = \text{Cov}(\nu_i(Z_i), \nu_j(Z_j)),$$

since $E(\text{Cov}(Y_i, Y_j | \mathbf{Z})) = 0$ from the conditional independence assumption. Intuitively, the correlation between Y_i and Y_j increases with the correlation between Z_i and Z_j . However, the range of the correlation between Y_i and Y_j is generally smaller than $[-1, 1]$.

Likelihood

The unconditional joint pdf/pmf of $\mathbf{Y} = \mathbf{y}$ is then given by

$$f(\mathbf{y}) = f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\mathcal{R}^k} \left\{ \prod_{i=1}^k g(y_i; \lambda_i(z_i), \boldsymbol{\alpha}_i) \right\} \phi_k(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{z}. \quad (2.1)$$

Unfortunately, there is no closed form for the above integration, except for a few special cases. With an arbitrary $\boldsymbol{\Sigma}$, use of multi-dimensional numerical integration is inevitable in evaluating the likelihood.

Advantages and Disadvantages

The advantage of normal distributions for the random effects is clear — its great flexibility in dependence structure. It also establishes a link to traditional quantitative genetics. This provides a meaningful interpretation of the models. We might consider Z_i as a latent quantitative trait under the influence of polygenic and multiple environmental factors. The mean vector of \mathbf{Z} , $\boldsymbol{\mu}$, can be specified as a linear function of the covariates, i.e. $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$. The covariance matrix of \mathbf{Z} , $\boldsymbol{\Sigma}$, can be partitioned into different correlation components according to the assumptions on genetic and environmental influences, as described in (A.3).

The major disadvantage is the computational difficulty in parameter estimation. Even with today's powerful computers and numerical methods, multi-dimensional integration is formidable when the dimension is higher than four. To overcome this obstacle, we need to consider estimating methods which are less computationally demanding than ML. Some likelihood-type approaches and their performance are investigated in Chapters 3 and 4, and the use of those approaches is illustrated in Chapter 6.

Another disadvantage of using normal random effects is the limitation in the type of the distributions. When we assume a parameter of the conditional distribution is random, the shape of the resulting marginal distribution differs from that of the conditional distribution. For example, the Poisson lognormal mixture has a marginal distribution which is much more skewed than the Poisson distribution. In the next section, we introduce the multinormal copula model which is more

flexible in modelling the marginal distributions.

2.1.2 Multinormal Copula Models

Definition

The theory of copulas provides a means of constructing various multivariate distributions. The joint distribution is formed by combining the marginal distributions through a copula, a multivariate distribution function with $U(0, 1)$ univariate margins. (Properties and parametric families of copulas can be found in Joe (1997) or Nelsen (1999).) Among the extensive parametric families of copulas, the multinormal (or Gaussian) copula provides a wide range of dependence and therefore is suitable for familial data analysis.

Suppose Y_i has univariate marginal cdf G_i . The joint distribution of \mathbf{Y} constructed using a multinormal copula is

$$G(\mathbf{y}) = \Phi_k(\Phi^{-1}[G_1(y_1)], \dots, \Phi^{-1}[G_k(y_k)]; \mathbf{R}(\boldsymbol{\alpha})), \quad (2.2)$$

where $\mathbf{R} = (\rho_{ij})$ is a k -dimensional correlation matrix. The cdf G_i can be chosen as any family of univariate distributions. For example, for count data, we can choose the Poisson or negative binomial as the marginal distribution; for survival data, we can choose the Weibull distribution.

Applications and Examples

Applications of multinormal copulas can be found in longitudinal data analysis (Xu, 1996; Song, 2000), but not in familial data. Examples of multinormal copulas are given in Sections 2.2.1, 2.3.2 and 2.4.3.

Dependence Structure

Since $\rho_{ij} = \text{Corr}(\Phi^{-1}[G_i(Y_i)], \Phi^{-1}[G_j(Y_j)])$ for fixed G_i and G_j , $\text{Corr}(Y_i, Y_j)$ is a monotone transformation of ρ_{ij} . In fact, that $\text{Corr}(Y_i, Y_j)$ and ρ_{ij} are very close to each other (Song, 2000). When $\rho_{ij} = 0$, we have independence between Y_i and Y_j . For the bivariate case, when $\rho_{ij} = 1$ or -1 , the most extreme positive (or negative) dependence between Y_i and Y_j is achieved (see discussion of the upper and lower Fréchet bounds in Joe, 1997, Chapter 3). For continuous G_i , when $\rho_{ij} = 1$, Y_i and Y_j are perfectly dependent. Moreover, if $G_i = G_j$ or G_i and G_j are related

by a location/scale shift, then the dependence is linear and $\text{Corr}(Y_i, Y_j) = 1$ as well. For discrete G_i , assuming that Y_i and Y_j have same support, $\text{Corr}(Y_i, Y_j) = 1$ only if $\rho_{ij} = 1$ and $G_i = G_j$.

Likelihood

The response Y_i can be treated as a transformed standard normal random variable. Let $\mathbf{Z} = (Z_1, \dots, Z_k)' \sim N(\mathbf{0}, \mathbf{R})$. When G_i is continuous with corresponding density function g_i , $Y_i = G_i^{-1}(\Phi(Z_i))$. Let $z_i = \Phi^{-1}(G_i(y_i))$ and $\mathbf{z} = (z_1, \dots, z_k)'$. The joint density function of \mathbf{Y} can be expressed as

$$f(\mathbf{y}) = \phi_k(\mathbf{z}; \mathbf{R}) \prod_i \frac{g_i(y_i)}{\phi(z_i)}.$$

When G_i is discrete, the transformation from Z to Y is discrete. The joint pmf of \mathbf{Y} involves an integral of the multinormal density over a k -dimensional rectangle. Multidimensional integration is also required to evaluate the likelihood function when censoring is present in survival data. Thus, the application of multinormal copula models is limited in practice.

Advantages and Disadvantages

One advantage of the multinormal copula model is that the marginal distributions and the joint distribution can be modelled separately. Combining the copula and a variety of univariate marginal models, a large number of multivariate distributions can be constructed, including both continuous and discrete distributions. Moreover, in models formed by a copula, the parameters which describe the association between the variables can be easily separated from those of the univariate marginals. This allows us to partition parameters into two groups: one group only related to the univariate marginals and the other only related to the copula.

Another advantage of this class of model is its flexible dependence structure. The correlation matrix $\mathbf{R} = (\rho_{ij})$ can adopt the structures described for quantitative traits in Chapter A. Since the correlations between the responses are close to those between the Z s, the dependence structure generated from such a model is similar to the dependence structure of a quantitative trait.

As mentioned before, the major disadvantage of the multinormal copula model is its computational infeasibility.

2.1.3 Conditional Independence (CI) Model

In this section, we present a new method to construct models for traits influenced by many genes and other factors. We consider this class of models when the primary interest is to determine if there is a familial correlation and how strong it is, but not to identify all the variance components. The motivation of this model is to simplify the task of modelling and computing the likelihood. There is a practical reason to consider this class of models as well. In data analysis, the information provided by the data might not be enough to identify every variance component. For example, suppose most families in a dataset are small containing no more than two generations, while only a few families are large with multiple generations. Under this scenario, it is impossible to obtain satisfactory estimation of the associations between distant relatives. So we focus on modelling the two most important dependencies, parent-offspring and sib-sib dependence, while approximating the association between other relatives.

The main idea of the conditional independence models is that the dependence between two relatives of second degree or higher comes through their parents. Conditionally on the status of their parents, the two individuals are assumed to be independent. However full siblings and monozygotic twins need to be treated differently. Because of the polygenic and environmental effects, they tend to resemble each other more than other types of relatives because (1) they share more genetic material and (2) they are more likely to share a common environment. For these reasons, we allow stronger dependence among them.

We begin with data formed by Type-I pedigrees without monozygotic twins or half siblings. We first state the conditional independence assumption. We then derive the dependence structure and the likelihood of a model under the CI assumption. Finally we extend the model to include monozygotic twins and half siblings and Type-II pedigrees. Examples for binary and count traits are given in Sections 2.2.2 and 2.3.3, respectively.

The notation used in the following sections was introduced in Section 1.1.

Conditional Independence Assumption for Type-I Pedigree

The **conditional independence assumption** is stated as follows:

If a sibling group S_i is not empty, then conditionally on the value of the parent, Y_i , (1) $(Y_j : j \in S_i)$ are non-negatively dependent; (2) $(Y_j : j \in S_i)$ and $(Y_{j'} : j' \notin D_i)$ are mutually

independent.

Under this assumption, the Y s corresponding to a group of siblings are allowed to be positively dependent on each other conditional on the value of their parent. Normally we assume an exchangeable dependence structure among siblings.

From this assumption, two important properties follow and the likelihood and dependence structure of a CI model can be derived from these two properties.

Properties of the CI Model for Type-I Pedigree

Property 1 *The joint pdf/pmf of \mathbf{Y} can be written as the product of the marginal pdf of the founder and the conditional probabilities of all the sibling groups given the value of their parent. Let 1 be the index of the founder. The joint pdf/pmf of \mathbf{Y} is*

$$f(\mathbf{y}; \mathbf{X}, \boldsymbol{\theta}) = f_1(y_1; \mathbf{X}_1, \boldsymbol{\theta}) \prod_{i \in T} f_{S_i|Y_i}(\mathbf{y}_{S_i}; y_i, \mathbf{X}_{S_i}, \boldsymbol{\theta}), \quad (2.3)$$

where f_1 is the univariate marginal pdf/pmf of Y_1 and $f_{S_i|Y_i}$ is the conditional pdf/pmf of the sibling group S_i , \mathbf{X}_1 is a vector of covariate values of the founder and \mathbf{X}_{S_i} is a matrix of covariate values of the sibling group S_i .

As a result, once we specify the joint distribution of the family units, we can derive the likelihood of the whole pedigree. Assuming exchangeable dependence among siblings, a family unit has two levels of dependence, parent-offspring dependence and sib-sib dependence, and can be modelled with a nested dependence structure — the sibling group is nested within a family unit, since generally the sib-sib dependence is stronger than the parent-offspring dependence.

Property 2 *For two individuals j and j' , if j is individual i 's descendant, i.e., $j \in D_i$, while j' is not i 's descendant, i.e., $j' \notin D_i$, then $Y_j \perp Y_{j'} | Y_i$.*

Proof: For discrete Y_i , we need to show that $\Pr(Y_j = y_j | Y_i = y_i, Y_{j'} = y_{j'}) = \Pr(Y_j = y_j | Y_i = y_i)$. This is true when j is i 's child. Now suppose j is not i 's child. Then let $(j_1, \dots, j_L) = A_j \cap D_i$, i.e., the collection of j 's ancestors who were descended from i . From the way j s are indexed, j_1 is

j 's parent, j_2 is j_1 's parent and so on. So the last element j_L is i 's child.

$$\begin{aligned}
& \Pr(Y_j = y_j | Y_i = y_i, Y_{j'} = y_{j'}) \\
&= \sum_{y_{j_1}} \cdots \sum_{y_{j_L}} \Pr(Y_j = y_j, Y_{j_1} = y_{j_1}, \dots, Y_{j_L} = y_{j_L} | Y_i = y_i, Y_{j'} = y_{j'}) \\
&= \sum_{y_{j_1}} \cdots \sum_{y_{j_L}} \Pr(Y_j = y_j | Y_{j_1} = y_{j_1}) \Pr(Y_{j_1} = y_{j_1} | Y_{j_2} = y_{j_2}) \cdots \Pr(Y_{j_L} = y_{j_L} | Y_i = y_i) \\
&= \sum_{y_{j_2}} \cdots \sum_{y_{j_L}} \Pr(Y_j = y_j | Y_{j_2} = y_{j_2}) \cdots \Pr(Y_{j_L} = y_{j_L} | Y_i = y_i) \\
&\quad \dots \\
&= \Pr(Y_j = y_j | Y_i = y_i) \text{ by induction.}
\end{aligned}$$

An analogous derivation holds for continuous \mathbf{Y} if we replace summations by integrations and probabilities by densities in the proof.

The second property enables us to determine the dependence between any two family members once the model for family units is specified. If we take two individuals, i and j , from the family tree then there are two possibilities: (1) both are on the same branch, i.e., one is the other's ancestor; (2) they are on two different branches, i.e., they are not ancestor-descendant.

Case 1: Consider the first case. Suppose individual i is individual j 's descendant. We define the path from i to j as $(i, A_{i \setminus j})$ (note that $i_{g_{i,j}} = j$). It follows directly from Property 2 that the path satisfies the Markov property. We can, therefore, view Y_i and Y_j as two points on a Markov process separated by $g_{i,j}$ steps.

Case 2: Now consider the second case when individual i and j are on two branches. Suppose the two branches meet at k . Then k is the first common ancestor shared by i and j . To go from i to j , we can follow the path described below: start from i and follow i 's ancestors up to the one who is k 's child, then go to another k 's children who is j 's ancestor, from there descend to j . Expressed in our notation, this path is $(i, A_{i \setminus j}, A'_{j \setminus i}, j)$. Next we show that it satisfies the Markov property. There are three possibilities: (1) Both are k 's children. Then $A_{i \setminus j}$ and $A'_{j \setminus i}$ are both empty and i is directly connected to j . (2) Only one is k 's child. Suppose it is i . Then $A_{i \setminus j}$ is empty. The path is $(i, A'_{j \setminus i}, j) = (Y_i, Y_{j_{g_{j,i}}}, \dots, Y_{j_1}, Y_j)$. From Property 2, the stretch from $Y_{j_{g_{j,i}}}$ to j satisfies the Markov property and $Y_i \perp Y_{j_{g_{j,i}-1}} | Y_{j_{g_{j,i}}}$ (since i is not $j_{g_{j,i}}$ descendant, while $j_{g_{j,i}-1}$ is). Therefore the whole path satisfies the Markov property. (3) Neither is k 's child. Then the path is $(Y_i, Y_{i_1}, \dots, Y_{i_{g_{i,j}}}, Y_{j_{g_{j,i}}}, \dots, Y_{j_1}, Y_j)$, where $i_{g_{i,j}}$ and $j_{g_{j,i}}$ are k 's children. The two

stretches from Y_i to $Y_{i_{g_i,j}}$ and from $Y_{j_{g_j,i}}$ to Y_j satisfy the Markov property. If we write $i_0 = i$, we have $Y_{i_{g_i,j-1}} \perp Y_{j_{g_j,i}} | Y_{i_{g_i,j}}$. Therefore the Markov property is satisfied everywhere.

The dependence between any two individuals in a pedigree can be derived based on the Markov property. The correlation structure will depend on the distribution under consideration.

Example 2.1 Suppose that the Y s are normally distributed. Suppose that the parent-offspring and sib-sib correlations are specified as ρ_{po} and ρ_{ss} , respectively. For case 1, variables on the same branch can be thought of an autocorrelated process with lag one correlation ρ_{po} . The correlation between individuals i and j is $\rho_{po}^{g_{i,j}}$. For case 2, the variables on the path from i to j are also autocorrelated but with correlation ρ_{ss} between $Y_{i_{g_i,j}}$ and $Y_{j_{g_j,i}}$, while ρ_{po} elsewhere. The correlation is $\rho_{po}^{g_{i,j}+g_{j,i}} \rho_{ss}$ when i and j are on two branches. Thus, the correlation between grandparent and grandchild is ρ_{po}^2 , the correlation between uncle and nephew is $\rho_{po}\rho_{ss}$, and the correlation between first cousins is $\rho_{po}^2\rho_{ss}$.

From the above example, we see that if ρ_1 and ρ_2 are positive, the dependence between two family members is always positive, but declines by generation. The same pattern will be shown in our other examples later.

Extension to Include Monozygotic Twins and Half Siblings

To extend the ideas to families with monozygotic twins, we can consider three levels of dependence within a family unit. In this case, each family unit can still be considered as a nested cluster, but with three levels—the twins are nested in the full siblings group. Again, let us use the normal distribution as an example. Suppose j_1 and j_2 are twins and the correlation between them is ρ_3 . Individual i is the g_i th descendant of j_1 and j is the g_j th descendant of j_2 . Then the correlation between i and j is $\rho_1^{g_i+g_j} \rho_3$. The correlation between other individuals remain the same.

It is also straightforward to extend the model to include half siblings. We assume that the half siblings are independent conditionally on their common parent. Under this assumption, half siblings are treated as second degree relatives. Suppose j_1 and j_2 are half siblings with common parent j_0 , i is the g_i th descendant of j_1 and j is the g_j th descendant of j_2 . Unlike the full siblings, the path between i and j will go through j_0 as shown in Figure 2.1. So $\text{Corr}(Y_i, Y_j) = \rho_1^{g_i+g_j+2}$ under the normal distribution.

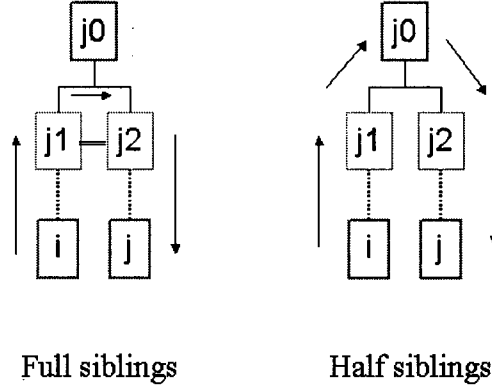


Figure 2.1: Path from i to j . If j_1 and j_2 are full siblings, the path is $(i, \dots, j_1, j_2, \dots, j)$; whereas if j_1 and j_2 are half siblings, the path is $(i, \dots, j_1, j_0, j_2, \dots, j)$.

Extension to Type-II Pedigree

In a Type-II pedigree, marry-in's are included. To extend the CI model to Type-II pedigree, we assume **Independence between spouses**:

Y_i and the spouse Y_{m_i} are independent. In the extension, Y_{m_i} and Y_j are independent if j is not one of m_i 's descendants, i.e. $j \notin D_{m_i}$.

We also modify the **conditional independence assumption** as follows:

The parent-offspring dependence is the same for both parents. Conditional on the value of Y_i and/or Y_{m_i} , Y_j and $Y_{j'}$ are non-negatively dependent if $j, j' \in S_i$; conditional on the value of Y_i , Y_j and $Y_{j'}$ are independent if j in S_i and j' not in D_i .

Based on the new assumptions, (2.3) can be modified as

$$f(\mathbf{y}) = \prod_{i \in B} f(y_i; \mathbf{X}_i, \boldsymbol{\theta}) \prod_{i \in RT} f_{S_i|Y_i, Y_{m_i}}(\mathbf{y}_{S_i}; y_i, y_{m_i}, \mathbf{X}_{S_i}, \boldsymbol{\theta}), \quad (2.4)$$

where B is the set of the founder and all the marry-in's.

Discussion

The issue is the appropriateness of the conditional independence assumption. It is similar to the Mendelian genetics assumption that the genotype is only determined by the genotypes of the parents. Here we assume that the values of the phenotype have the same property. It seems an

oversimplification of the reality. But we should regard it as an approximation of reality. The biggest challenge of models for familial data is to accommodate the dependence structure of a complex pedigree. The two properties of the dependence structure generated from this assumption suggest it is reasonable for familial data. First, the dependence between two relatives is weaker as they are more distantly related, which agrees with the structures discussed in Appendix A. Second, there are no strong assumptions made on the two most important dependencies — the parent-offspring and sib-sib dependence — therefore they can be modelled with reasonable flexibility.

There are two major advantages of the CI models. First, the task of modelling a complex pedigree reduces to the much easier task of modelling family units. The choices of marginal distributions are much broader as only two levels of dependence need to be considered. Second, CI models can be handled computationally. The conditional independence assumption allows us to write the likelihood as a product of univariate marginal likelihoods and conditional likelihoods which are computable. Even though a CI model may look like a choice of convenience, the simplicity of this model does make it appealing when identifying the variance components is not of interest.

2.2 Binary Traits

A binary trait is characterized by two states, the presence and absence of a certain feature. For each individual there is an indicator random variable Y such that $Y = 0$ if the trait is absent and $Y = 1$ if the trait is present. The feature may have both genetic and environmental determinants. In this section, we introduce two models to describe their effects: the multivariate probit model and two-component model. In the latter, the trait is assumed to be affected by two independent components, a heritable component and a non-heritable component. The probit model is a well accepted method to analyze multivariate binary responses. Therefore our discussion of this model is brief. On the other hand, the two-component model is new. Therefore we give a detailed discussion of the properties of this model.

2.2.1 Multivariate Probit (MVP) Model

The multivariate probit model was introduced to familial data analysis in the 1960's (Falconer, 1965). The idea is that Y_i is determined by an unobserved continuous variable Z_i . When the value of Z_i exceeds the fixed threshold 0, $Y_i = 1$; otherwise, $Y_i = 0$. It is assumed that $\mathbf{Z} = (Z_1, \dots, Z_k)'$ has distribution $N(\boldsymbol{\mu}, \mathbf{R})$, where $\boldsymbol{\mu}$ is usually specified as a linear function of the covariates, i.e., $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$. Unit variance is assumed for each Z_i since variance is not identifiable when only binary data are observed.

It is easy to see that $\Pr(Y_i = 1) = \Pr(Z_i > 0) = \Phi(\mu_i)$ and $\Pr(Y_i = 1, Y_j = 1) = \Phi_2(\mu_i, \mu_j; \rho_{ij})$. It follows that $\text{Corr}(Y_i, Y_j) = \Phi_2(\mu_i, \mu_j; \rho_{ij}) - \Phi(\mu_i)\Phi(\mu_j)$, which is an increasing function of ρ_{ij} .

The pmf of \mathbf{Y} is given by

$$\Pr(\mathbf{Y} = \mathbf{y}) = \int_{\mathfrak{A}} \phi_k(\mathbf{z}; \boldsymbol{\mu}, \mathbf{R}) d\mathbf{z}, \quad (2.5)$$

where $\mathfrak{A} = A_1 \times \dots \times A_k$ with $A_i = (-\infty, 0]$ when $y_i = 0$ and $A_i = (0, \infty)$ when $y_i = 1$.

The MVP model also belongs to the multinormal copula family. It carries the advantages and disadvantages of this class of models.

2.2.2 Two-Component Model

The idea of a two-component model is to assume that the presence of a binary trait is caused by two independent components: a heritable component and a non-heritable component. The heritable component includes more than genetic factors. It includes non-genetic aspects which one might also inherit from the family, such as life style or personality. The non-heritable component could be any non-familial factors, like age, gender, treatment. Observed genetic information, such as mutation type, can be treated as a non-heritable factor.

We propose two different two-component models: parallel and series. In a parallel two-component model, we assume that either of the components alone can cause the presence of the trait. If the presence of the trait is due to heritable factors, we say that event A_H has occurred, on the other hand, if it is due to non-heritable factors, we say the event A_N has occurred. Let A be the event that the trait is present, that is $A = A_H \cup A_N$. Let Y , Y_H and Y_N be the indicators for A , A_H and A_N , respectively. Then $Y = Y_H \vee Y_N$. In a series two-component model, we assume that the

trait is only present when both components are present. In this case, mathematically, $A = A_H \cap A_N$ and $Y = Y_H \wedge Y_N$. In this model, the event A_H means a person is genetically prone to develop the trait, but will not become affected unless he/she is exposed to unhealthy environmental factors, or grows old. The parallel and series models stand for different mechanisms, but one can be obtained from the other by a change of notation. Specifically, suppose $Y = Y_H \vee Y_N$. Let $\bar{Y} = 1 - Y$, $\bar{Y}_H = 1 - Y_H$ and $\bar{Y}_N = 1 - Y_N$, then $\bar{Y} = \bar{Y}_H \wedge \bar{Y}_N$.

In the following, we first discuss how to model the two components, then explore the dependence pattern generated by a two-component model and, finally, show how to evaluate the likelihood of an extended pedigree.

Models for Y_N and Y_H

Let \mathbf{Y} , \mathbf{Y}_H and \mathbf{Y}_N be the indicator vectors of a family. We assume that \mathbf{Y}_H and \mathbf{Y}_N are independent and model them separately. Y_{Ni} is determined by the characteristics of individual i which are not shared by others. Therefore the Y_{Ni} s are independent both within and between families. Y_{Ni} may depend on a set of covariates \mathbf{X}_i , which characterize the individual exposure, and can be modelled by a parametric model such as

$$\begin{aligned} \text{logit}[\Pr(Y_{Ni} = 1|\mathbf{X}_i)] &= \mathbf{X}_i\boldsymbol{\beta} \\ \text{or } \Phi^{-1}[\Pr(Y_{Ni} = 1|\mathbf{X}_i)] &= \mathbf{X}_i\boldsymbol{\beta}. \end{aligned}$$

\mathbf{Y}_H is determined by the heritable factors. Therefore, \mathbf{Y}_{Hi} s are dependent within families, but independent across families. We will construct a \mathbf{Y}_H that follows a conditional independence model. In addition to the CI assumption, we also make the following two assumptions on \mathbf{Y}_H :

A-1 Identical univariate margins: the probability of $Y_{Hi} = 1$, denoted by p_H , does not vary among individuals.

A-2 Exchangeability among full siblings: let \mathbf{Y}_{Hs} be the vector of indicators for k full siblings, then $\Pr(\mathbf{Y}_{Hs} = \mathbf{y}) = \Pr(\mathbf{Y}_{Hs} = \mathbf{y}')$, where \mathbf{y}' is an arbitrary permutation of \mathbf{y} .

First consider models for \mathbf{Y}_H for a sample containing only Type-2 families. Exchangeability is assumed within each family. Since we also assume that siblings are positively dependent, one

way to model \mathbf{Y}_H is to use an exchangeable mixture model. In such a model,

$$\begin{aligned} Y_{H1}, \dots, Y_{Hk} | P = \pi &\stackrel{iid}{\sim} \text{Bernoulli}(\pi) \\ P &\sim G(\cdot; \boldsymbol{\theta}), \end{aligned}$$

where G is a distribution function with support $[0,1]$. The unconditional probability of \mathbf{Y}_H is given by

$$\Pr(Y_{Hi} = y_i, i = 1, \dots, k) = \int_0^1 \pi^{y_+} (1 - \pi)^{k - y_+} dG(\pi; \boldsymbol{\theta}), \quad (2.6)$$

where $y_+ = \sum_i y_i$. A simple example of a choice of G is the Beta(α, β) distribution which yields the multivariate binary Beta-binomial(α, β) distribution of \mathbf{Y}_H . In this example, (2.6) becomes

$$\Pr(Y_{Hi} = y_i, i = 1, \dots, k) = \frac{B(\alpha + y_+, \beta + k - y_+)}{B(\alpha, \beta)},$$

where $B(\alpha, \beta)$ is the beta function. Another example for a choice of G is the probit model: $P = \Phi(Z)$ and $Z \sim N(\mu_0, \sigma^2)$. Then

$$G(\pi; \mu_0, \sigma^2) = \Pr(Z \leq \Phi^{-1}(\pi)) = \Phi\left(\frac{\Phi^{-1}(\pi) - \mu_0}{\sigma}\right).$$

An alternative to the mixture model is a copula which provides an exchangeable dependence. An example is the multivariate extension of the Frank copula (Joe, 1997, Section 7.1). The cdf of \mathbf{Y}_H is given by

$$F_k(\mathbf{y}) = \frac{-1}{\alpha} \log \left\{ 1 - (1 - e^{-\alpha}) \prod_{i=1}^k \left[\frac{1 - e^{-\alpha F_1(y_i)}}{1 - e^{-\alpha}} \right] \right\}, \quad \alpha \geq 0,$$

where F_1 is the univariate marginal pdf of Y_i . For example, one joint probability from this is

$$\Pr(Y_{Hi} = 0, i = 1, \dots, k) = \frac{-1}{\alpha} \log \left\{ 1 - (1 - e^{-\alpha}) \left[\frac{1 - e^{-\alpha(1 - p_H)}}{1 - e^{-\alpha}} \right]^k \right\}.$$

We need to mention that Y_H and Y_N may sometimes not be identifiable. For example, if \mathbf{Y}_H follows a beta-binomial distribution with parameters α and β and $\Pr(Y_{Ni} = 1) = p_N$ does not vary from person to person, then the three parameters α , β and p_N can not be identified in the joint distribution of \mathbf{Y} if we only have data containing Type-1 families. When data contain families of different sizes and structures or there are covariates present, the parameters are normally identifiable.

Next consider models for Y_H for Type-3 families. Let 1 be the index of the parent and $\mathbf{Y}_{H,-1}$ be the indicator vector for the siblings. Then

$$\mathbf{Y}_{H,-1} | Y_{H1} = y_1 \sim F_k(\cdot; \boldsymbol{\theta}(y_1)),$$

where k is the number of siblings. To keep p_H constant for both parent and siblings, $\boldsymbol{\theta}(y_1)$ needs to satisfy the following constraint:

$$\begin{aligned} \Pr(Y_{Hi} = 1) &= \Pr(Y_{H1} = 1) \Pr(Y_{Hi} = 1 | Y_{H1} = 1; \boldsymbol{\theta}(1)) + \Pr(Y_{H1} = 0) \Pr(Y_{Hi} = 1 | Y_{H1} = 0; \boldsymbol{\theta}(0)) \\ &= p_H \Pr(Y_{Hi} = 1 | Y_{H1} = 1; \boldsymbol{\theta}(1)) + (1 - p_H) \Pr(Y_{Hi} = 1 | Y_{H1} = 0; \boldsymbol{\theta}(0)) \\ &= p_H, \quad i \neq 1. \end{aligned}$$

For example, if $\mathbf{Y}_{H,-1}$ follows a multivariate binary beta-binomial distribution with parameters (α_1, β_1) or (α_0, β_0) , given $Y_{H1} = 1$ or 0, respectively, then this constraint is

$$p_H \frac{\alpha_1}{\alpha_1 + \beta_1} + (1 - p_H) \frac{\alpha_0}{\alpha_0 + \beta_0} = p_H.$$

The above models for Type-3 families are applicable to family units in a Type-I pedigree, in which each family unit is a Type-3 family.

Dependence Structure

The strength of dependence between two binary variables can be indexed by the so called risk ratio (R), defined by the conditional probability

$$R_{ij} = \Pr(Y_i = 1 | Y_j = 1) = \frac{\Pr(Y_i = 1, Y_j = 1)}{\Pr(Y_j = 1)}.$$

It measures the increased risk of having the trait for an individual when the trait is observed on a relative. When Y_i and Y_j are positively dependent, $R_{ij} > \Pr(Y_j = 1)$ and the closer R_{ij} is to 1, the stronger is the dependence.

First, let us assume that the non-heritable component does not exist. In this case $\mathbf{Y} = \mathbf{Y}_H$ and $R_{ij} = R_{ji}$ since $\Pr(Y_i = 1) = \Pr(Y_j = 1) = p_H$. To determine the dependence between any two family members, we need to know the univariate marginal probabilities, the dependence between parent and offspring, and the dependence between siblings. Let $p = \Pr(Y_i = 1)$, $q = 1 - p$. For a parent-offspring pair (i, j) , let

$$p_{1,(m|l)} = \Pr(Y_j = m | Y_i = l) \quad \text{and} \quad p_{1,(ml)} = \Pr(Y_j = m, Y_i = l),$$

with $l, m = 0$ or 1 . Then we specify the parent-offspring conditional probability matrix (CPM):

$$\mathbf{P}_1 = \begin{pmatrix} p_{1,(0|0)} & p_{1,(1|0)} \\ p_{1,(0|1)} & p_{1,(1|1)} \end{pmatrix}.$$

Similarly, let $\mathbf{P}_2 = (p_{2,(m|l)})$ be the CPM for a sibling pair and $p_{2,(ml)}$ be the joint probabilities. The role of \mathbf{P}_1 and \mathbf{P}_2 is similar to a Markov chain transition matrix. With assumption A-1, we can write $\mathbf{P}_l = \mathbf{A} + \Delta_l \mathbf{B}$ for $l = 1, 2$, where

$$\mathbf{A} = \begin{pmatrix} q & p \\ q & p \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} p & -p \\ -q & q \end{pmatrix},$$

and $\Delta_l = 1 - p_{l,(0|1)} - p_{l,(1|0)} = p_{l,(1|1)} - p_{l,(1|0)} = p_{l,(0|0)} - p_{l,(0|1)}$. This can be shown as follows:

$$\begin{aligned} \mathbf{A} + \Delta_l \mathbf{B} &= \begin{pmatrix} q & p \\ q & p \end{pmatrix} + \begin{pmatrix} p[p_{l,(0|0)} - p_{l,(0|1)}] & -p[p_{l,(1|1)} - p_{l,(1|0)}] \\ -q[p_{l,(0|0)} - p_{l,(0|1)}] & q[p_{l,(1|1)} - p_{l,(1|0)}] \end{pmatrix} \\ &= \begin{pmatrix} q - p_{l,(01)} + pp_{l,(0|0)} & p - p_{l,(11)} + pp_{l,(1|0)} \\ q - p_{l,(00)} + qp_{l,(0|1)} & p - p_{l,(10)} + qp_{l,(1|1)} \end{pmatrix} \\ &= \begin{pmatrix} p_{l,(00)} + pp_{l,(0|0)} & p_{l,(10)} + pp_{l,(1|0)} \\ p_{l,(01)} + qp_{l,(0|1)} & p_{l,(11)} + qp_{l,(1|1)} \end{pmatrix} \\ &= \begin{pmatrix} qp_{l,(0|0)} + pp_{l,(0|0)} & qp_{l,(1|0)} + pp_{l,(1|0)} \\ pp_{l,(0|1)} + qp_{l,(0|1)} & pp_{l,(1|1)} + qp_{l,(1|1)} \end{pmatrix} \\ &= \begin{pmatrix} p_{l,(0|0)} & p_{l,(1|0)} \\ p_{l,(0|1)} & p_{l,(1|1)} \end{pmatrix} = \mathbf{P}_l. \end{aligned}$$

Once p , \mathbf{P}_1 and \mathbf{P}_2 are specified, for any two individuals i and j in the family, we can derive their CPM, denoted by $\mathbf{P}^{(i,j)}$, and this will immediately give us their risk ratio (R_{ij}). First, consider individuals i and j on the same branch. Then the following result holds.

Lemma 2.1 *When i and j are on the same branch, the conditional probability matrix of Y_j given Y_i is the $g_{i,j}$ -step transition matrix*

$$\mathbf{P}^{(i,j)} = \mathbf{P}_1^{g_{i,j}} = \mathbf{A} + \Delta_1^{g_{i,j}} \mathbf{B} \quad (2.7)$$

Proof: Since each branch of the family tree is a Markov chain with stationary transition probability matrix \mathbf{P}_1 , we have $\mathbf{P}^{(i,j)} = \mathbf{P}_1^{g_{i,j}}$. To verify the second equality we use an induction proof. We have shown that this is true when $g_{i,j} = 1$. Now assume that it is true for $g_{i,j}$. Then

$$\begin{aligned}\mathbf{P}^{(i,j)}\mathbf{P}_1 &= (\mathbf{A} + \Delta_1^{g_{i,j}}\mathbf{B})(\mathbf{A} + \Delta_1\mathbf{B}) \\ &= \mathbf{A}^2 + \Delta_1^{g_{i,j}}\mathbf{B}\mathbf{A} + \Delta_1\mathbf{A}\mathbf{B} + \Delta_1^{g_{i,j}+1}\mathbf{B}^2 \\ &= \mathbf{A} + \Delta_1^{g_{i,j}+1}\mathbf{B}.\end{aligned}$$

The last step follows because $\mathbf{A}^2 = \mathbf{A}$, $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = 0$ and $\mathbf{B}^2 = \mathbf{B}$. We just verified that the formula holds for $g_{i,j} + 1$. It thereby is established for all g . \square

(This proof is provided in “An Introduction to Stochastic Modeling”, Taylor (1994), page 127-128.)

The conditional probability matrix gives us $R_{ij} = p + \Delta_1^{g_{i,j}}q$. When the dependence between parent and offspring is positive, the case we consider in a family setting, then $0 < \Delta_1 < 1$. Consequently, the dependence between two individuals on the same branch is positive but declines by generation.

From Lemma 2.1 we also can derive the CPM for two individuals on different branches. When i and j are on different branches, the path between them is a Markov chain with transition matrix \mathbf{P}_2 between $Y_{ig_{i,j}}$ and $Y_{jg_{j,i}}$, and \mathbf{P}_1 elsewhere. The conditional probability matrix for i and j is given by

$$\begin{aligned}\mathbf{P}^{(i,j)} &= \mathbf{P}^{(i,ig_{i,j})}\mathbf{P}_2\mathbf{P}^{(jg_{j,i},j)} \\ &= (\mathbf{A} + \Delta_1^{g_{i,j}}\mathbf{B})(\mathbf{A} + \Delta_2\mathbf{B})(\mathbf{A} + \Delta_1^{g_{j,i}}\mathbf{B}) \\ &= \mathbf{A} + \Delta_2\Delta_1^{g_{i,j}+g_{j,i}}\mathbf{B}.\end{aligned}\tag{2.8}$$

(2.8) holds when $A_{i \setminus j}$ or $A_{j \setminus i}$ is empty, since

$$\mathbf{A} + \Delta_1^0\mathbf{B} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The expression for $\mathbf{P}^{(i,j)}$ yields $R_{ij} = p + \Delta_2\Delta_1^{g_{i,j}+g_{j,i}}q$. Provided $\Delta_2, \Delta_1 > 0$, the dependence is positive and decreases when individuals i and j are more distantly related.

Actually, Δ_1 is equal to the parent-offspring correlation:

$$\begin{aligned}\Delta_1 &= p_{1,(1|1)} - p_{1,(1|0)} = \frac{qp_{1,(11)} - pp_{1,(10)}}{pq} \\ &= \frac{p_{1,(11)} - p(p_{1,(11)} + p_{1,(10)})}{pq} \\ &= \frac{p_{1,(11)} - p^2}{pq}.\end{aligned}$$

Similarly, Δ_2 is equal to the sib-sib correlation. Therefore,

$$\text{Corr}(Y_i, Y_j) = (pR_{ij} - p^2)/(pq) = \begin{cases} \Delta_1^{g_{i,j}} & \text{when } i \text{ and } j \text{ are on the same branch,} \\ \Delta_2 \Delta_1^{g_{i,j} + g_{j,i}} & \text{when } i \text{ and } j \text{ are on different branches.} \end{cases}$$

This is the same correlation structure for the CI model as in Example 2.1.

Now, consider that the non-heritable component exists. To make two individuals comparable, let $\Pr(Y_{Ni} = 1) = \Pr(Y_{Nj} = 1) = p_N$. If $Y_i = Y_{Hi} \vee Y_{Ni}$, then

$$\Pr(Y_i = 1) = p_H + p_N - p_H p_N = p_H + q_H p_N$$

and

$$\begin{aligned}\Pr(Y_i = 1, Y_j = 1) &= p_{11}^{(i,j)} + 2p_{10}^{(i,j)} p_N + p_{00}^{(i,j)} p_N^2 \\ &= p_{11}^{(i,j)} + 2(p_H - p_{11}^{(i,j)}) p_N + (1 - 2p_H + p_{11}^{(i,j)}) p_N^2 \\ &= p_{11}^{(i,j)} (1 - p_N)^2 + 2p_H p_N + (1 - 2p_H) p_N^2,\end{aligned}$$

where the $p^{(i,j)}$ s are the joint probabilities of Y_{Hi} and Y_{Hj} . With p_H and p_N fixed, $\Pr(Y_i = 1, Y_j = 1)$ reaches its upper limits $p_H + q_H p_N^2$ when $p_{11}^{(i,j)} = p_H$, i.e., Y_{Hi} and Y_{Hj} are perfectly dependent. So the upper bound of R_{ij} is

$$R_{ij} \leq \frac{p_H + q_H p_N^2}{p_H + q_H p_N}.$$

When $Y_i = Y_{Hi} \wedge Y_{Ni}$, $\Pr(Y_i = 1) = p_H p_N$ and $\Pr(Y_i = 1, Y_j = 1) = p_{11}^{(i,j)} p_N^2$. In this case, the expression for the risk ratio is simpler: $R_{ij} = p_{11}^{(i,j)} p_N$, where $p_{11}^{(i,j)} = p_{11}^{(i,j)} / p_H$. Obviously, R_{ij} achieves its upper bound p_N when $p_{11}^{(i,j)} = 1$.

Remarks:

1. With the involvement of the independent Y_{Ni} and Y_{Nj} , the dependence between Y_i and Y_j is weaker than that between Y_{Hi} and Y_{Hj} , provided that Y_{Hi} and Y_{Hj} are positively associated with each other.

2. The dependence between Y_i and Y_j decreases with the dependence between Y_{Hi} and Y_{Hj} .

Likelihood

In this section we derive the likelihood. Since the parallel and series models can be obtained by changing notation, we only consider the case of $Y = Y_H \vee Y_N$.

For $\mathbf{Y}_H = \mathbf{y}_H$ and $\mathbf{Y}_N = \mathbf{y}_N$, $\mathbf{y} = \mathbf{y}_H \vee \mathbf{y}_N$ (the element wise maximum). The likelihood of $\mathbf{Y} = \mathbf{y}$ is $L(\mathbf{y}) = \Pr(\mathbf{Y}_H = \mathbf{y}_H) \prod_i \Pr(Y_{Ni} = y_{Ni})$. However, \mathbf{Y}_H and \mathbf{Y}_N are not directly observable. Given the value of Y_i , there are different possible values of Y_{Hi} and Y_{Ni} . For $Y_i = 0$, we must have $Y_{Hi} = Y_{Ni} = 0$. But, for $Y_i = 1$, there are two possibilities: (1) $Y_{Hi} = 1$, in which case the value of Y_{Ni} does not matter; (2) $Y_{Hi} = 0$, in which case Y_{Ni} has to be 1. For a family with k_+ members having the feature, there are 2^{k_+} different combinations to consider. The likelihood is the summation of the probability over all of the combinations. Let $D = \{i : Y_i = 1\}$ be the set of affected individuals and $\bar{D} = \{i : Y_i = 0\}$ be the set of unaffected individuals; $\mathbf{Y}_H^d = \{Y_{Hi} : i \in D\}$, $\mathbf{Y}_H^{\bar{d}} = \{Y_{Hi} : i \in \bar{D}\}$, $\mathbf{Y}_N^d = \{Y_{Ni} : i \in D\}$, $\mathbf{Y}_N^{\bar{d}} = \{Y_{Ni} : i \in \bar{D}\}$. Then $\mathbf{Y}_H^{\bar{d}} = \mathbf{Y}_N^{\bar{d}} = \mathbf{0}$. Given a possible value of $\mathbf{Y}_H^d = \mathbf{y}$, the likelihood is

$$\Pr(\mathbf{Y}_H^d = \mathbf{y}, \mathbf{Y}_H^{\bar{d}} = \mathbf{0}) \prod_{j \in D} \left[\Pr(Y_{Nj}^d = 1) \right]^{1-y_j} \prod_{i \in \bar{D}} \Pr(Y_{Ni}^{\bar{d}} = 0).$$

Let C be the collection of all possible values of \mathbf{Y}_H^d . Then the likelihood is

$$L = \prod_{i \in \bar{D}} \Pr(Y_{Ni}^{\bar{d}} = 0) \sum_{\mathbf{y} \in C} \left(\Pr(\mathbf{Y}_H^d = \mathbf{y}, \mathbf{Y}_H^{\bar{d}} = \mathbf{0}) \prod_{j \in D} \left[\Pr(Y_{Nj}^d = 1) \right]^{1-y_j} \right).$$

The calculation of the probabilities involving the Y_{Ni} s is straightforward based on the model specified for this variable. Here we discuss how to calculate the joint probabilities of \mathbf{Y}_H . For Type-2 families, the probability is directly given by the specified joint distribution of \mathbf{Y}_H . For example, if a mixture model is specified, then the probability is given by (2.6) and it only depends on y_+ . For an Type-I pedigree (including a Type-3 family as a special case), the probability is given by (2.3) and can be written as

$$\Pr(\mathbf{Y}_H = \mathbf{y}) = \Pr(Y_{H1} = y_1) \prod_{i \in RT} \Pr(\mathbf{Y}_{HS_i} = \mathbf{y}_{HS_i} | Y_{Hi} = y_i).$$

Here we use the family shown in Figure 1.1 as an example to demonstrate the algorithm. With

$Y_{H1} = Y_{H2} = Y_{H4} = Y_{H6} = 1$ and $Y_{H3} = Y_{H5} = 0$, the joint probability of \mathbf{Y}_H can be written as

$$\Pr(Y_{H1} = 1) \Pr(Y_{H2} = 1, Y_{H3} = 0 | Y_{H1} = 1) \Pr(Y_{H4} = 1, Y_{H5} = 0 | Y_{H2} = 1) \Pr(Y_{H6} = 1 | Y_{H3} = 0).$$

A family tree may not be complete. Values of some parents might be missing. For example, if the value of individual 3 in Figure 1.1 is not observed, the joint probability of the remaining individuals has to be calculated by summing the full joint probability over Y_{H3} .

Even though for a big family with many affected members, 2^{k+} can be quite large, it is manageable since the individual conditional and marginal probabilities are very easy to evaluate.

Two-Component Models for Type-4 Families and Type-II Pedigrees

For a Type-4 family, let Y_1 and Y_2 be the indicators of the parents and \mathbf{Y}_s be those of their offspring. We assume that Y_{H1} and Y_{H2} are independent. Consider the following model for $\mathbf{Y}_{H,s}$:

$$\mathbf{Y}_{H,s} | Y_{H1} + Y_{H2} = l \sim F_k(\cdot; \boldsymbol{\theta}_l), \quad l = 0, 1, 2$$

where F_k is an exchangeable mixture model with parameters $\boldsymbol{\theta}_l$. To keep p_H constant for all i , $\boldsymbol{\theta}_l$ needs to satisfy the following constraint:

$$\begin{aligned} \Pr(Y_{Hi} = 1) &= \Pr(Y_{H1} + Y_{H2} = 0) \Pr(Y_{Hi} = 1 | Y_{H1} + Y_{H2} = 0; \boldsymbol{\theta}_0) + \\ &\quad + 2 \Pr(Y_{H1} + Y_{H2} = 1) \Pr(Y_{Hi} = 1 | Y_{H1} + Y_{H2} = 1; \boldsymbol{\theta}_1) \\ &\quad + \Pr(Y_{H1} + Y_{H2} = 2) \Pr(Y_{Hi} = 1 | Y_{H1} + Y_{H2} = 2; \boldsymbol{\theta}_2) \\ &= (1 - p_H)^2 \Pr(Y_{Hi} = 1 | Y_{H1} + Y_{H2} = 0; \boldsymbol{\theta}_0) \\ &\quad + 2(1 - p_H)p_H \Pr(Y_{Hi} = 1 | Y_{H1} + Y_{H2} = 1; \boldsymbol{\theta}_1) \\ &\quad + p_H^2 \Pr(Y_{Hi} = 1 | Y_{H1} + Y_{H2} = 2; \boldsymbol{\theta}_2) = p_H, \quad \text{for } i > 2 \end{aligned} \tag{2.9}$$

For a special case, in which the indicators for two siblings are independent given the values of those for their two parents, the model is specified by four probabilities: p_H and

$$p_{H,l} = \Pr(Y_{Hi} = 1 | Y_{H1} + Y_{H2} = l), \quad \text{for } i > 2 \text{ and } l = 0, 1, 2,$$

where $p_{H,l}$ satisfies the constraint in (2.9).

These models can be applied to family units in a Type-II pedigree, since each family unit in such a pedigree is a Type-4 family.

Discussion

Here we would like to discuss the two component assumption. It is plausible to consider the heritable and non-heritable factors as two components of the binary trait. However, are they separable and independent like two components in an electrical system? It is a rather doubtful assumption. Does this not oversimplify reality? This is not an unusual question in statistical practice. The idea of the two-component model is similar to that of the well accepted random effects model. We can not always verify the assumption that the mean structure can be separated into random effects and fixed effects either. Models are an approximate reflection of the reality. It might work well with certain binary traits and might not with others.

2.3 Count Traits

In this section we consider non-negative integer random variables representing the results of counts, such as the number of meningiomas in a patient with NF2. Often analysis of univariate count data is based on the Poisson distribution, or a Poisson mixture distribution when the data are overdispersed relative to the Poisson. In this section we study parametric models for multivariate counts, including the Poisson-lognormal mixture, the multinormal copula model and a new model developed based on the conditional independence assumption. The first two models have been studied in the literature and their strengths and weaknesses have been discussed in Section 2.1. Therefore, our effort here is mainly focused on the development of the CI model and its properties.

2.3.1 Poisson-Lognormal Mixture

This is an example of a random effects model discussed in Section 2.1. Given $\Lambda_i = \lambda_i$, it is assumed that $Y_i|\Lambda_i = \lambda_i \sim \text{Poisson}(\lambda_i)$. The random vector $\log \mathbf{\Lambda}$ itself follows a MVN distribution with mean $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \mathbf{R}(\boldsymbol{\alpha})$. Here \mathbf{R} is the correlation matrix which depends on dependence parameters $\boldsymbol{\alpha}$.

The moments of \mathbf{Y} can be easily obtained through conditional expectation and properties

of the Poisson and lognormal distributions. Let $\rho_{ij} = (\mathbf{R}(\boldsymbol{\alpha}))_{ij} = \text{Corr}(\log \lambda_i, \log \lambda_j)$.

$$\begin{aligned} E(Y_i) &= \exp(\mu_i + \frac{1}{2}\sigma^2) \equiv \nu_i, \\ \text{Var}(Y_i) &= \nu_i + \nu_i^2(\exp(\sigma^2) - 1), \\ \text{Cov}(Y_i, Y_j) &= \nu_i \nu_j (\exp(\sigma^2 \rho_{ij}) - 1), \quad i \neq j, \\ \text{Corr}(Y_i, Y_j) &= \frac{\exp(\sigma^2 \rho_{ij}) - 1}{[(\exp(\sigma^2) - 1 + \nu_i^{-1})(\exp(\sigma^2) - 1 + \nu_j^{-1})]^{1/2}}. \end{aligned}$$

Since $E(Y_i) \leq \text{Var}(Y_i)$, the marginal distribution of Y_i is overdispersed relative to the Poisson. The correlation between Y_i and Y_j depends directly on ρ_{ij} and $|\text{Corr}(Y_i, Y_j)| < |\rho_{ij}|$. The range of the correlation between Y_i and Y_j is not as wide as for the lognormal distribution. When $\nu_i = \nu_j = \nu$, the correlation is bounded by

$$\frac{\nu[\exp(-\sigma^2) - 1]}{1 + \nu[\exp(\sigma^2) - 1]} \leq \text{Corr}(Y_i, Y_j) \leq \frac{\nu[\exp(\sigma^2) - 1]}{1 + \nu[\exp(\sigma^2) - 1]}$$

(Joe, 1997, Section 7.2). The upper bound goes to 1 as either ν or σ^2 goes to infinity.

The joint probability function of \mathbf{Y} is given by

$$\begin{aligned} \Pr(Y_i = y_i, i = 1, \dots, k; \boldsymbol{\mu}, \sigma^2, \mathbf{R}) &= \int_{\mathfrak{A}^k} \prod_i \Pr(Y_i = y_i | \Lambda_i) f(\boldsymbol{\lambda}; \boldsymbol{\mu}, \sigma^2 \mathbf{R}) d\boldsymbol{\lambda} \\ &= \int_{\mathfrak{R}^k} \prod_i \Pr(Y_i = y_i | \Lambda_i = e^{z_i}) \phi_k(\mathbf{z}; \boldsymbol{\mu}, \sigma^2 \mathbf{R}) d\mathbf{z} \end{aligned} \quad (2.10)$$

where \mathfrak{A}^k is the positive orthant of the k -dimensional Euclidean space and f is the density function of the multivariate lognormal distribution.

The distribution of the counts based on a Poisson-lognormal mixture model is highly skewed. Data generated from such a model normally have a few extremely large counts whereas the majority are around the median. For data with a high frequency of zero counts, we can consider the zero-inflated Poisson-lognormal mixture distribution (see Yau and Lee (2001) for an example). Another way to model multivariate counts is to incorporate the random effects with the negative binomial distribution (see Tempelman (1996) as an example).

2.3.2 Multinormal Copula Model for Count Traits

To construct a multinormal copula model for a count trait, the marginal distribution in (2.2), G_i , can be any univariate distribution for count data, such as Poisson, negative binomial, or generalized

Poisson. The parameters of the marginal distribution can be related to a linear function of the covariate \mathbf{X} through a link function as in a univariate regression model. For example, if G_i is the $\text{Poisson}(\lambda_i)$ cdf, we can specify that $\log(\lambda_i) = \mathbf{X}_i\boldsymbol{\beta}$.

Let $\mathbf{Z} = (Z_1, \dots, Z_k)' \sim N(\mathbf{0}, \mathbf{R})$, then $\Pr(Y_i = y_i) = \Pr(z_i(y_i - 1) < Z_i < z_i(y_i))$, where $z_i(l) = \Phi^{-1}(G_i(l))$ for $l = 0, 1, \dots$, and $z(-1) = -\infty$. The joint pmf of \mathbf{Y} is given by

$$\Pr(\mathbf{Y} = \mathbf{y}) = \Pr(z_i(y_i) < Z_i < z_i(y_i + 1), i = 1, \dots, k).$$

2.3.3 Conditional Independence Models for Count Traits

In this section, we construct conditional independence models for count data and study the model properties, such as dependence structure and likelihood. The models we considered are based on a univariate distribution belonging to the convolution-closed infinitely divisible class and a random operator \mathcal{A} , which are introduced in the following section.

Convolution-Closed Infinitely Divisible Distributions and Random Operator \mathcal{A}

Consider a class of distributions $\{F_\theta, \theta > 0\}$, which is convolution-closed infinitely divisible (CCID), i.e., $F_{\theta_1} * F_{\theta_2} = F_{\theta_1 + \theta_2}$, where $*$ is the convolution operator. Such distributions for count data include the Poisson, negative binomial (θ, p) with fixed p , and generalized Poisson (θ, η) with fixed η . The negative binomial distribution, denoted by $\text{NB}(\theta, p)$, has pmf

$$f(y) = \frac{\Gamma(\theta + y)p^\theta q^y}{\Gamma(\theta)y!}, \quad 0 < p \leq 1 \text{ and } q = 1 - p, \theta > 0.$$

The mean and variance are given by $\theta q/p$ and $\theta q/p^2$, respectively. The generalized Poisson distribution, denoted by $\text{GP}(\theta, \eta)$, has pmf

$$f(y) = \frac{\theta(\theta + \eta y)^{y-1}}{e^{\theta + \eta y} y!}, \quad \theta > 0, 0 \leq \eta < 1.$$

The mean and variance are $\theta/(1 - \eta)$ and $\theta/(1 - \eta)^3$.

For two independent random variables Z_1 and Z_2 with CCID distributions F_{θ_1} and F_{θ_2} , respectively, $Z_1 + Z_2 \sim F_{\theta_1 + \theta_2}$. Let $G_{\alpha, y}$ be the conditional distribution of Z_1 given $Z_1 + Z_2 = y$. $\mathcal{A}(Y, \alpha)$, $0 < \alpha \leq 1$, is a random operator such that given $Y = y$, $\mathcal{A}(y, \alpha) \sim G_{\alpha, y}$ and $\mathcal{A}(Y, \alpha) \sim F_{\alpha\theta}$ if $Y \sim F_\theta$.

Examples of the $\mathcal{A}(Y, \alpha)$ are:

1. Poisson: $\mathcal{A}(y, \alpha) \sim \text{Binomial}(y, \alpha)$;
2. negative binomial (NB): $\mathcal{A}(y, \alpha) \sim \text{Beta-binomial}(y, \alpha\theta, (1 - \alpha)\theta)$;
3. generalized Poisson (GP): $\mathcal{A}(y, \alpha) \sim \text{quasi-binomial}(y, \alpha, \zeta)$, which has the following pmf

$$f(k) = \binom{y}{k} \frac{\alpha(1 - \alpha)}{1 + \zeta k} \left[\frac{\alpha + \zeta k}{1 + \zeta y} \right]^{k-1} \left[\frac{1 - \alpha + \zeta(y - k)}{1 + \zeta y} \right]^{y-k-1},$$

where $\zeta = \eta/\theta$.

In all the three examples, $E\mathcal{A}(y, \alpha) = \alpha y$ (see Joe (1997) Section 8.4 for details).

Models for Type-3 Families

Let \mathbf{Y} be the vector of count variables of a Type-3 family where Y_1 is the variable corresponding to the parent. Define

$$Y_1 = Z_1 + Z_{12} + Z_{13},$$

and for $i > 1$,

$$Y_i = Z_1 + Z_{22} + Z_{i3},$$

where Z_1 , Z_{12} , Z_{22} and Z_{i3} are independent variables with distribution $Z_1 \sim F_{\theta_1}$, Z_{12} , $Z_{22} \sim F_{\theta_2}$ and $Z_{i3} \sim F_{\theta_{i3}}$.

The interpretation of the model is that Y_i is formed by latent components Z_1 , Z_{12} or Z_{22} and Z_{i3} ; among the latent components, Z_1 is shared by all the family members while Z_{22} is only shared by siblings. Z_1 , Z_{12} and Z_{22} together are the heritable (or familial) components. Z_{i3} is the non-heritable component determined by individual characteristics and could be modelled as depending on a set of covariates \mathbf{X}_i through a link function h , i.e., $h(\theta_{i3}) = \mathbf{X}_i\boldsymbol{\beta}$.

If F_θ is the Poisson, NB or GP distribution, for $Y \sim F_\theta$, $\text{Var}(Y) = \theta\gamma$, where

$$\gamma = \begin{cases} 1 & \text{for Poisson,} \\ (1 - p)/p^2 & \text{for NB,} \\ 1/(1 - \eta)^3 & \text{for GP.} \end{cases}$$

Therefore, for these three distributions, it is straightforward to derive the parent-offspring and

sib-sib correlations. Let $\theta_i = \theta_1 + \theta_2 + \theta_{i3}$. Then

$$\text{Corr}(Y_i, Y_j) = \begin{cases} \frac{\text{Var}Z_1}{\sqrt{(\text{Var}Y_i)(\text{Var}Y_j)}} = \frac{\theta_1}{\sqrt{\theta_i\theta_j}} & \text{if } i \text{ and } j \text{ are parent and offspring} \\ \frac{(\text{Var}Z_1 + \text{Var}Z_{22})}{\sqrt{(\text{Var}Y_i)(\text{Var}Y_j)}} = \frac{(\theta_1 + \theta_2)}{\sqrt{\theta_i\theta_j}} & \text{if } i \text{ and } j \text{ are siblings} \end{cases} \quad (2.11)$$

Model for Type-I Pedigrees

Now we apply this type of model to family units in a Type-I pedigree. For the family unit U_i , $i \in RT$, we specify

$$\begin{aligned} Y_i &= Z_1^{(i)} + Z_{12}^{(i)} + Z_{i3} \\ Y_j &= Z_1^{(i)} + Z_{22}^{(i)} + Z_{j3}, \text{ for } j \in S_i, \end{aligned} \quad (2.12)$$

where $Z_1^{(i)} \sim F_{\theta_1}$, $Z_{12}^{(i)}$, $Z_{22}^{(i)} \sim F_{\theta_2}$ and $Z_{l3} \sim F_{\theta_{l3}}$, $l = i$ or j .

To link the family units together, we need the stochastic representation:

$$Y_j = \mathcal{A}(Y_i, \alpha_i) + Z_{22}^{(i)} + Z_{j3}, \text{ for } j \in S_i, \quad (2.13)$$

where $\alpha_i = \theta_1/(\theta_1 + \theta_2 + \theta_{i3})$. By definition, $\mathcal{A}(Y_i, \alpha_i) \sim F_{\theta_1}$ and $\mathcal{A}(y_i, \alpha_i)$ has the same distribution as $Z_1^{(i)}|Y_i = y_i$. Also, since Y_i is independent of $Z_{22}^{(i)}$ and Z_{j3} , $\mathcal{A}(Y_i, \alpha_i)$ is independent of them as well.

Example: The stochastic representation for the pedigree in Figure 1.1:

$$\begin{aligned} Y_1 &= Z_1^{(1)} + Z_{12}^{(1)} + Z_{13}, \\ Y_2 &= \mathcal{A}(Y_1, \alpha_1) + Z_{22}^{(1)} + Z_{23}, \\ Y_3 &= \mathcal{A}(Y_1, \alpha_1) + Z_{22}^{(1)} + Z_{33}, \\ Y_4 &= \mathcal{A}(Y_2, \alpha_2) + Z_{22}^{(2)} + Z_{43}, \\ Y_5 &= \mathcal{A}(Y_2, \alpha_2) + Z_{22}^{(2)} + Z_{53}, \\ Y_6 &= \mathcal{A}(Y_3, \alpha_3) + Z_{22}^{(3)} + Z_{63}. \end{aligned}$$

Based on such a model, the path between individuals i and j is a discrete Markov series (Joe, 1997, section 8.4) with varying lag one correlation. Suppose $(Y_{l_1}, \dots, Y_{l_m})$ is the path between

i and j , with $l_1 = i$ and $l_m = j$. Thus, for $i' = 1, \dots, m-1$, the path has the following stochastic representation:

$$Y_{l_{i'+1}} = \mathcal{A}(Y_{l_{i'}}, \alpha_{i'}) + Z'_{l_{i'+1}}.$$

where $\alpha_{i'}$ and $Z'_{l_{i'+1}}$ depend on the relationship between $l_{i'+1}$ and $l_{i'}$. If $l_{i'+1}$ and $l_{i'}$ are parent and offspring, $\alpha_{i'} = \theta_1/(\theta_1 + \theta_2 + \theta_{l_{i'}3})$ and $Z'_{l_{i'+1}}$ is the summation of the second and third components in (2.13), hence $Z'_{l_{i'+1}} \sim F_{\theta_2 + \theta_{l_{i'}3}}$; if $l_{i'+1}$ and $l_{i'}$ are siblings, $\alpha_{i'} = (\theta_1 + \theta_2)/(\theta_1 + \theta_2 + \theta_{l_{i'}3})$ and $Z'_{l_{i'+1}}$ is the third component in 2.13, hence $Z'_{l_{i'+1}} \sim F_{\theta_{l_{i'}3}}$.

For the Poisson, NB or GP distribution, the correlation between any two family members is given in the result below.

Result 2.1 *For a model specified in (2.12), when F is the Poisson, NB or GP distribution, we have*

$$\text{Corr}(Y_i, Y_j) = \prod_{j'=1}^{m-1} \rho_{j', j'+1}, \quad (2.14)$$

where $\rho_{j', j'+1} = \text{Corr}(Y_{l_{j'}}, Y_{l_{j'+1}})$.

Proof: When Y has a Poisson, NB or GP distribution, given $Y = y$, $E[\mathcal{A}(y, \alpha)] = \alpha y$. Therefore $E(Y_{l_{i'+1}} | Y_{l_{i'}}) = \alpha_{i'} Y_{l_{i'}} + E Z'_{l_{i'+1}}$ and

$$\begin{aligned} \text{Cov}(Y_{l_{i'}}, Y_{l_{i'+1}}) &= E Y_{l_{i'}} Y_{l_{i'+1}} - E Y_{l_{i'}} E Y_{l_{i'+1}} \\ &= E \{ Y_{l_{i'}} E(Y_{l_{i'+1}} | Y_{l_{i'}}) \} - E Y_{l_{i'}} E \{ E(Y_{l_{i'+1}} | Y_{l_{i'}}) \} \\ &= E \alpha_{i'} Y_{l_{i'}}^2 + E Y_{l_{i'}} E Z'_{l_{i'+1}} - E Y_{l_{i'}} (E \alpha_{i'} Y_{l_{i'}} + E Z'_{l_{i'+1}}) \\ &= \alpha_{i'} \text{Var} Y_{l_{i'}}. \end{aligned}$$

Let $\sigma_{l_{i'}}^2 = \text{Var} Y_{l_{i'}}$, then $\rho_{i', i'+1} = \text{Corr}(Y_{l_{i'}}, Y_{l_{i'+1}}) = \alpha_{i'} \sigma_{l_{i'}} / \sigma_{l_{i'+1}}$. Moreover,

$$\begin{aligned} E(Y_{l_{i'+1}} | Y_{l_{i'-1}}) &= E \{ E(Y_{l_{i'+1}} | Y_{l_{i'}}) | Y_{l_{i'-1}} \} \\ &= E(\alpha_{i'} Y_{l_{i'}} | Y_{l_{i'-1}}) + E Z'_{l_{i'+1}} \\ &= \alpha_{i'} \alpha_{i'-1} Y_{l_{i'-1}} + \alpha_{i'} E Z'_{l_{i'}} + E Z'_{l_{i'+1}}. \end{aligned}$$

By induction,

$$E(Y_{l_{i'+1}} | Y_{l_1}) = Y_{l_1} \prod_{j'=1}^{i'} \alpha_{j'} + E Z'_{l_1} \prod_{j'=2}^{i'} \alpha_{j'} + \dots + E Z'_{l_{i'+1}}.$$

It follows for the preceding covariance calculation that

$$\begin{aligned}
\text{Cov}(Y_{l_1}, Y_{l_{i'+1}}) &= \text{Var}Y_{l_1} \prod_{j'=1}^{i'} \alpha_{j'} = \sigma_{l_1}^2 \prod_{j'=1}^{i'} \alpha_{j'} \\
&= \sigma_{l_1}^2 \prod_{j'=1}^{i'} \frac{\sigma_{l_{j'+1}}}{\sigma_{l_{j'}}} \rho_{j', j'+1} \\
&= \sigma_{l_1} \sigma_{l_{i'+1}} \prod_{j'=1}^{i'} \rho_{j', j'+1}.
\end{aligned}$$

Therefore $\text{Corr}(Y_{l_1}, Y_{l_{i'+1}}) = \prod_{j'=1}^{i'} \rho_{j', j'+1}$. The equation (2.14) is obtained by taking $i' + 1 = m$.

□

When θ_{i3} is a constant, say θ_3 , let $\theta = \theta_1 + \theta_2 + \theta_3$. Then the parent-offspring and sib-sib correlations are $\rho_1 = \theta_1/\theta$ and $\rho_2 = (\theta_1 + \theta_2)/\theta$, respectively. In this case, for the Poisson, NB or GP distribution,

$$\text{Corr}(Y_i, Y_j) = \begin{cases} \rho_1^{g_{i,j}} & i \text{ and } j \text{ on the same branch,} \\ \rho_1^{g_{i,j}+g_{j,i}} \rho_2 & i \text{ and } j \text{ on different branches.} \end{cases}$$

Likelihood

As shown in (2.3), the likelihood involves the conditional likelihoods of the sibling groups. Here we show how to compute those conditional likelihoods for the model proposed in the preceding section. Let $Z_1^{(i)*} = \mathcal{A}(y_i, \alpha_i)$. The conditional likelihood of \mathbf{Y}_{S_i} given $Y_i = y_i$ is:

$$\sum_{l=0}^y \sum_{l_0=0}^l \Pr(Z_1^{(i)*} = l_0) \Pr(Z_{22}^{(i)} = l - l_0) \prod_{j \in S_i} \Pr(Z_{j3} = y_j - l),$$

where $y = \min(y_j, j \in S_i)$. We know that $Z_1^{(i)*} \sim G_{\alpha_i, y_i}$.

Extension to Type-4 Families

To extend the model to a Type-4 family, we consider four independent latent components for an individual: $Y_i = Z_1 + Z_2 + Z_3 + Z_4$. Among these components, Z_1 is related to parent 1 while Z_2 is related to parent 2 and these two components have the same distribution. Z_3 is shared by siblings. The last component is the non-heritable component.

If we let the siblings share the first two components, this forces the parent-offspring correlation to be less than half of the sib-sib correlation which is unusual under a polygenic effect. For this reason, we construct the model as follows:

$$Y_i = Z_{i1} + Z_{i2} + Z_{i3} + Z_{i4}, \quad i = 1, 2 \text{ (parents)}, \quad (2.15)$$

$$Y_i = \mathcal{A}_{i1}(Y_1, \alpha_1) + \mathcal{A}_{i2}(Y_2, \alpha_2) + Z_{33} + Z_{i4}, \quad i > 2 \text{ (offspring)}. \quad (2.16)$$

where $Z_{i1}, Z_{i2} \sim F_{\theta_1}$, $Z_{j3} \sim F_{\theta_3}$, $j = 1, 2, 3$ and $Z_{i4} \sim F_{\theta_{i4}}$; $\alpha_l = \theta_1 / (2\theta_1 + \theta_3 + \theta_{l4})$, $l = 1, 2$. \mathcal{A}_{i1} and $\mathcal{A}_{i'1}$ are two independent realizations of \mathcal{A} , as are \mathcal{A}_{i2} and $\mathcal{A}_{i'2}$. Z_{33} is shared by siblings representing additional dependence from other sources. All the Z -components are independent. So, the sib-sib correlation is partly from the \mathcal{A} operators and partly from Z_{33} .

Both the parents and offspring have a marginal distribution $F_{2\theta_1 + \theta_3 + \theta_{i4}}$, for all i . For the parents, this is directly from (2.16). By the definition of \mathcal{A} , we have \mathcal{A}_{i1} , $\mathcal{A}_{i'1}$, \mathcal{A}_{i2} and $\mathcal{A}_{i'2}$ have the same distribution F_{θ_1} . \mathcal{A}_{i1} and \mathcal{A}_{i2} are independent since Y_1 and Y_2 are independent. Therefore, for $i > 2$, $Y_i \sim F_{2\theta_1 + \theta_3 + \theta_{i4}}$ as well.

Next, we use the Poisson distribution as an example to illustrate the parent-offspring and sib-sib correlations for such a model. In this case, Y_i has mean $\lambda_i = 2\theta_1 + \theta_3 + \theta_{i4}$. The correlation between parent i and offspring j is

$$\text{Corr}(Y_i, Y_j) = \frac{\text{Var}(\mathcal{A}_{ji}(Y_i, \alpha_i))}{\sqrt{(\text{Var} Y_i)(\text{Var} Y_j)}} = \frac{\theta_1}{\sqrt{\lambda_i \lambda_j}}.$$

Since $\theta_1 \leq \lambda_i/2$, for all i , the parent-offspring correlation is always less than or equal to 0.5 in this model. This agrees with a polygenic model for quantitative traits when there is no household effect.

To obtain the correlation between offspring j and j' , we first derive the covariance:

$$\text{Cov}(Y_j, Y_{j'}) = \text{Cov}(\mathcal{A}_{j1}(Y_1, \alpha_1), \mathcal{A}_{j'1}(Y_1, \alpha_1)) + \text{Cov}(\mathcal{A}_{j2}(Y_2, \alpha_2), \mathcal{A}_{j'2}(Y_2, \alpha_2)) + \text{Var} Z_{33}.$$

Based on the fact that $\mathcal{A}_{jl}(Y_l, \alpha_l) \sim \text{Poisson}(\alpha_l \lambda_l)$ and $\mathcal{A}_{jl}(y_l, \alpha_l) \sim \text{Binomial}(y_l, \alpha_l)$, we have

$$\begin{aligned} \text{Cov}(\mathcal{A}_{jl}(Y_l, \alpha_l), \mathcal{A}_{j'l}(Y_l, \alpha_l)) &= \text{E}[\mathcal{A}_{jl}(Y_l, \alpha_l) \mathcal{A}_{j'l}(Y_l, \alpha_l)] - \text{E}[\mathcal{A}_{jl}(Y_l, \alpha_l)] \text{E}[\mathcal{A}_{j'l}(Y_l, \alpha_l)] \\ &= \text{E}\{\text{E}[\mathcal{A}_{jl}(y_l, \alpha_l) \mathcal{A}_{j'l}(y_l, \alpha_l)] | Y_l = y_l\} - \alpha_l^2 \lambda_l^2 \\ &= \text{E} \alpha_l^2 Y_l^2 - \alpha_l^2 \lambda_l^2 = \alpha_l^2 \lambda_l. \end{aligned}$$

Hence,

$$\text{Cov}(Y_j, Y_{j'}) = \alpha_1^2 \lambda_1 + \alpha_2^2 \lambda_2 + \theta_3,$$

and it follows that

$$\text{Corr}(Y_j, Y_{j'}) = \frac{\alpha_1^2 \lambda_1 + \alpha_2^2 \lambda_2 + \theta_3}{\sqrt{\lambda_j \lambda_{j'}}}.$$

If $\lambda_i = \lambda$ for all i , then the parent-offspring and sib-sib correlations do not vary from relative pair to pair. The parent-offspring correlation becomes $\rho_1 = \theta_1/\lambda$. Recall that $\alpha_l = \theta_l/\lambda_l$, therefore $\alpha_1 = \alpha_2 = \rho_1$. Consequently, the sib-sib correlation becomes $\rho_2 = 2\rho_1^2 + \theta_3/\lambda = (\Delta\theta_1 + \theta_3)/\lambda$, where $\Delta = 2\theta_1/\lambda$.

2.4 Survival Data

This section is concerned with data representing times to the occurrence of some specified event, such as death, onset of a disease or a clinical feature. The time to an event, denoted by T_i , is a positive continuous random variable. The special feature of survival data is the presence of censoring. There are three types of censoring: left, right and interval. In general, if the survival time of family member i is censored, we observe a time interval $A_i = (0, R_i)$, (L_i, ∞) or $(L_i, R_i]$, corresponding to left, right and interval censored, respectively. We make the following assumptions about the mechanisms behind the censoring, regardless of its type: the censoring times of a family are independent of the survival times.

Suppose a full parametric model is specified for the distribution of $\mathbf{T} = (T_1, \dots, T_k)'$. Let S_l and f_l denote the l -dimensional marginal survival and density function of \mathbf{T} . When the censoring times are independent of the survival times, the model can be estimated based on the partial likelihood of the survival times. Without loss of generality, assume that the first k_1 individuals in the family are not censored and have observed failure time t_1, \dots, t_{k_1} , and the rest are censored with censoring intervals A_{k_1+1}, \dots, A_k . Then the partial likelihood of this family is:

$$L = f_{k_1}(t_i, i = 1, \dots, k_1) \int_{\mathfrak{A}} f_{k_2|k_1}(t_j, j = k_1 + 1, \dots, k \mid T_i = t_i, i = 1, \dots, k_1) dt_{k_1+1} \dots dt_k,$$

where $k_2 = k - k_1$, $f_{k_2|k_1}$ is the conditional density function of T_{k_1+1}, \dots, T_k given T_1, \dots, T_{k_1} and $\mathfrak{A} = A_{k_1+1} \times \dots \times A_k$.

In particular, if only right censoring is present, Let C_i be the censoring time and $Y_i = T_i \vee C_i$, either the time of event or the time of censoring, whichever is observed first. Then the partial likelihood becomes:

$$L = f_{k_1}(y_i, i = 1, \dots, k_1) S_{k_2|k_1}(y_j, j = k_1 + 1, \dots, k \mid T_i = t_i, i = 1, \dots, k_1),$$

where $S_{k_2|k_1}$ is the conditional survival function.

2.4.1 Multivariate Lognormal Models

Here we assume that $\log \mathbf{T} = (\log T_1, \dots, \log T_k)' \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$. We model $\log \mathbf{T}$ as a usual quantitative trait. The only difference in the analysis is that censoring may be present.

Suppose right censoring is present. Let $\mathbf{Z} = \log \mathbf{Y}$ and \mathbf{z}_1 be the vector of the observed times of event with length k_1 and \mathbf{z}_2 the vector of censored times with length $k_2 = k - k_1$. The partial likelihood is given by

$$L = \phi_{k_1}(\mathbf{z}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \Phi_{k_2}(-\mathbf{z}_2; -\boldsymbol{\mu}_2^*, \boldsymbol{\Sigma}_2^*),$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are the mean and covariance matrix of \mathbf{Z}_1 ; $\boldsymbol{\mu}_2^*$ and $\boldsymbol{\Sigma}_2^*$ are the conditional mean and covariance matrix of \mathbf{Z}_2 given $\mathbf{Z}_1 = \mathbf{z}_1$.

2.4.2 Multivariate Lognormal Frailty

In a frailty model, the hazard of T_i conditionally on $\Lambda_i = \lambda_i$ has the form

$$h_i(t|\Lambda_i = \lambda_i) = \lambda_i h_i^*(t)$$

for some $h_i^*(t)$. If $h_i^*(t) = 1$, then the conditional distribution of T_i is exponential(λ_i), while if $h_i^*(t) = \gamma t^{\gamma-1}$, it is Weibull(λ_i, γ). Λ_i is called the frailty representing the underlying risk. The observed survival function (also called the marginal survival function) of T_i can be obtained by integrating out Λ_i :

$$S_i(t) = \int \exp(-\lambda_i H_i^*(t)) f_i(\lambda_i) d\lambda_i = E \{ \exp(-\Lambda_i H_i^*(t)) \},$$

where $H_i^*(t) = \int_0^t h_i^*(t) dt$ and f_i is the density function of Λ_i . We see that $S_i(t)$ is the Laplace transform of Λ_i , evaluated at $H_i^*(t)$.

Under the familial setting, it is natural to assume that each family member has a separate frailty and these are correlated with each other. Again we choose the multivariate lognormal as the distribution of Λ , i.e.,

$$\log \Lambda = \mathbf{Z} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{R}(\boldsymbol{\alpha})).$$

The distribution of \mathbf{T} is difficult to obtain as there is no simple expression for the Laplace transform of lognormal random variables.

Let $S_i(t|\lambda_i)$ and $f_i(t|\lambda_i)$ be the conditional survival and density function of T_i . Then with right censoring, the contribution to the partial likelihood of this family is

$$L = \int_{\mathfrak{R}^k} \prod_i [f_i(t_i|e^{z_i})]^{d_i} [S_i(t_i|e^{z_i})]^{1-d_i} \phi_k(\mathbf{z}; \boldsymbol{\mu}, \sigma^2 \mathbf{R}) d\mathbf{z},$$

where $d_i = 1$ if the time of the event is observed, 0 if it is censored.

2.4.3 Multinormal Copula Model

Suppose that T_i has univariate marginal survival function S_i . In this section we construct the joint survival function of \mathbf{T} using a multinormal copula. Since it is more convenient to use the survival functions in survival analysis, we replace the cdf G_i in (2.2) with the survival function S_i . The two definitions are equivalent. Then we have

$$S(\mathbf{t}) = \Phi_k(\Phi^{-1}(S_1(t_1)), \dots, \Phi^{-1}(S_k(t_k)); \mathbf{R}(\boldsymbol{\alpha})). \quad (2.17)$$

Suppose S_i is fully specified with a density function f_i . Consider only right censoring. Let $z_i = \Phi^{-1}(S_i(y_i))$, \mathbf{z}_1 be the vector of uncensored times and \mathbf{z}_2 the vector of censored ones. The partial likelihood is given by

$$L = \Phi_{k_1}(\mathbf{z}_1; \mathbf{0}, \mathbf{R}_1) \prod_{i=1}^{k_1} \frac{f_i(y_i)}{\phi(z_i)} \Phi_{k_2}(\mathbf{z}_2; \boldsymbol{\mu}_2^*, \mathbf{R}_2^*),$$

where $\boldsymbol{\mu}_2^*$, \mathbf{R}_2^* are the conditional mean and covariance matrix of \mathbf{Z}_2 given $\mathbf{Z}_1 = \mathbf{z}_1$.

Chapter 3

Estimating Procedures Generated from Composite Likelihood Functions

As mentioned in Chapter 2, the major difficulty in implementing models with multinormal random effects and multivariate normal copula is parameter estimation when data sets involve large families. The MLE is generally computationally difficult to obtain since it involves high dimensional integration. We will view the MVN model as a special case, which is used for quantitative traits or survival data (if on the log scale). Computational difficulty also occurs with the MVN model when censoring is present. Therefore, developing estimation procedures that are less computationally demanding is important.

The models we mentioned share a common property: the parameters which specify the models can be classified as univariate marginal parameters and dependence parameters. The former characterizes the univariate margins, such as the means and variances in the MVN model, while the latter, joint with the univariate marginal parameters, fully specify the multivariate law, such as the correlations in the MVN model. This property allows us to develop estimation methods based on composite likelihood from the univariate and bivariate margins. A composite likelihood (CL), as named by Lindsay (1988), is formed by adding together individual component log likelihoods, each of which is a log likelihood or conditional log likelihood of a marginal distribution of a multivariate model. We consider two CL approaches for familial data: the first is based on both CL of the univariate margins and bivariate margins and estimates the marginal parameters and the dependent

parameters in two steps while the second only uses the bivariate CL and estimates the parameters simultaneously. Weighting schemes are also considered to improve the efficiency of these approaches.

To illustrate the main ideas, we use a simple MVP model as an example. Suppose

$$Y_{ij} = I(Z_{ij} > 0),$$

where the latent variables Z_{ij} have a normal distribution with common mean μ and variance 1. Moreover, suppose $\text{Cov}(Z_{ij}, Z_{ij'}) = \rho$ for all i and $j \neq j'$. In this model, there is one marginal parameter, μ , and one dependence parameter, ρ . In the first approach, we estimate μ using the univariate marginal log likelihood of $Y_{ij} = y_{ij}$:

$$l_1(y_{ij}; \mu) = y_{ij} \log \Phi_1(\mu) + (1 - y_{ij}) \log(1 - \Phi_1(\mu)).$$

We form a CL function by summing $l_1(y_{ij}; \mu)$ over all i and j : $\Psi_1(\mu) = \sum_i \sum_j l_1(y_{ij}; \mu)$, then maximize $\Psi_{UCL}(\mu)$ with respect to μ . This leads to $\hat{\mu} = \Phi_1^{-1}(\bar{y})$, where \bar{y} is the grand average of y_{ij} . We then estimate ρ using the bivariate marginal log likelihood. For a pair $(Y_{ij}, Y_{ij'}) = (y_{ij}, y_{ij'})$, the bivariate log likelihood is:

$$\begin{aligned} l_2(y_{ij}, y_{ij'}; \rho, \mu) &= \log \left\{ \Phi_2(\mu, \mu; \rho)^{(1-y_{ij})(1-y_{ij'})} [\Phi_1(\mu) - \Phi_2(\mu, \mu; \rho)]^{y_{ij}+y_{ij'}-2y_{ij}y_{ij'}} \right. \\ &\quad \left. [1 - 2\Phi_1(\mu) + \Phi_2(\mu, \mu; \rho)]^{y_{ij}y_{ij'}} \right\} \\ &= (1 - y_{ij})(1 - y_{ij'}) \log \Phi_2(\mu, \mu; \rho) + (y_{ij} + y_{ij'} - 2y_{ij}y_{ij'}) \log[\Phi_1(\mu) - \Phi_2(\mu, \mu; \rho)] \\ &\quad + y_{ij}y_{ij'} \log[1 - 2\Phi_1(\mu) + \Phi_2(\mu, \mu; \rho)]. \end{aligned}$$

We form the second CL function by summing $l_2(y_{ij}, y_{ij'}; \mu, \rho)$ over all possible pairs:

$$\Psi_{BCL}(\mu, \rho) = \sum_i \sum_{j > j'} l_2(y_{ij}, y_{ij'}; \mu, \rho).$$

The estimate of ρ is obtained by replacing μ by $\hat{\mu}$ and then maximizing $\Psi_{BCL}(\hat{\mu}, \rho)$ with respect to ρ .

In the second approach, we form a CL function

$$\Psi_{BCL}^*(\mu, \rho) = \sum_i w_i \sum_{j > j'} l_2(y_{ij}, y_{ij'}; \mu, \rho),$$

where w_i is a weight depending on the family size. We maximize $\Psi_{BCL}^*(\mu, \rho)$ and thus estimate μ and ρ in one step. The motivation and choice of the weights will be given later.

The same ideas work for models with covariates although the notation is more complicated. In Section 3.1, we introduce notation to reflect the structure of the parameters. In Section 3.2, we first give a general introduction to estimating procedures based on composite likelihood. We then present the two approaches described above and consider their performance under independence. Section 3.3 deals with methods to estimate the covariance matrix of the estimates and Section 3.4 contains some remarks. In the next chapter, we examine the efficiency of these procedures by comparing them with the ML method based on some special cases for which the comparison is actually possible.

3.1 Notation

Let $F(\cdot; \gamma_{i1}, \dots, \gamma_{iM})$ be the joint cdf of \mathbf{Y}_i , a vector of dimension k_i , where the vector γ_{i1} consists of all the parameters appearing in the univariate margins of \mathbf{Y}_i ; γ_{i2} consists of those not in the univariate margins, but in the bivariate margins, and so on. We assume that margins of order higher than M , including the full joint distribution, are fully determined by $\gamma_{i1}, \dots, \gamma_{iM}$. The following discussion deals with models having $M = 2$, but the arguments can be extended to models having $M > 2$.

Let $\gamma_{i1}^{(j)}$ be the elements of γ_{i1} pertaining to Y_{ij} 's marginal distribution. That is, suppose the marginal distribution of Y_{ij} is $F_1(\cdot; \gamma_{i1}^{(j)})$. Suppose that $\gamma_{i1}^{(j)}$ is of length d_1 :

$$\gamma_{i1}^{(j)} = (\gamma_{i1,1}^{(j)}, \dots, \gamma_{i1,d_1}^{(j)})'.$$

Let $\gamma_{i1,l} = (\gamma_{i1,l}^{(1)}, \dots, \gamma_{i1,l}^{(k_i)})'$, $l = 1, \dots, d_1$, and $\gamma_{i1} = (\gamma'_{i1,1}, \dots, \gamma'_{i1,d_1})'$. Similarly, suppose the bivariate marginal distribution of Y_{ij} and $Y_{ij'}$ is $F_2(\cdot, \cdot; \gamma_{i1}^{(j)}, \gamma_{i1}^{(j')}, \gamma_{i2}^{(j,j')})$, where

$$\gamma_{i2}^{(j,j')} = (\gamma_{i2,1}^{(j,j')}, \dots, \gamma_{i2,d_2}^{(j,j')})'.$$

Let $\gamma_{i2,l} = (\gamma_{i2,l}^{(2,1)}, \gamma_{i2,l}^{(3,1)}, \gamma_{i2,l}^{(3,2)}, \dots, \gamma_{i2,l}^{(k_i,1)}, \dots, \gamma_{i2,l}^{(k_i,k_i-1)})'$, $l = 1, \dots, d_2$ and $\gamma'_{i2} = (\gamma'_{i2,1}, \dots, \gamma'_{i2,d_2})'$.

Furthermore, suppose γ_{im} , $m = 1, 2$, can be modelled as a function of a set of parameters θ_m , i.e. $\gamma_{im} = \mathbf{h}_m(\mathbf{X}_i^{(m)}; \theta_m, k_i)$ for a parametric family $\mathbf{h}_m(\cdot; \theta_m, k)$, $m = 1, 2$. One particular case is $\gamma_{im} = \mathbf{X}_i^{(m)} \theta_m$. Here $\mathbf{X}_i^{(m)}$ is a matrix of covariates, not necessarily linked to the mean of the distribution. The values in $\mathbf{X}_i^{(m)}$ are treated as known constants. Suppose that the vectors θ_1

and θ_2 are two distinct sets of parameters. Let $\gamma'_i = (\gamma'_{i1}, \gamma'_{i2})$, $\theta' = (\theta'_1, \theta'_2)$, $\mathbf{h}' = (\mathbf{h}'_1, \mathbf{h}'_2)$, and

$$\tilde{\mathbf{X}}_i = \begin{pmatrix} \mathbf{X}_i^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i^{(2)} \end{pmatrix}.$$

Then $\gamma_i = \mathbf{h}(\tilde{\mathbf{X}}_i; \theta, k)$. In the linear case $\gamma_i = \tilde{\mathbf{X}}_i \theta$.

To illustrate the notation, we list a few examples.

Example 3.1 MVN model. For $k_i = 3$, $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})$ with $E(\mathbf{Y}_i) = \boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$; $\text{Var}(Y_{ij}) = \sigma_{ij}^2 = \sigma^2$ for all j . For an exchangeable dependence model, $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho_{ijj'} = \alpha$ for $j > j'$. Then, $\gamma_{i1} = \begin{pmatrix} \boldsymbol{\mu}_i \\ \boldsymbol{\sigma}_i^2 \end{pmatrix}$ with $\boldsymbol{\sigma}_i^2 = (\sigma_{i1}^2, \sigma_{i2}^2, \sigma_{i3}^2)'$. We can write $\gamma_{i1} = \mathbf{X}_i^{(1)} \boldsymbol{\theta}_1$ with $\boldsymbol{\theta}_1 = \begin{pmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{pmatrix}$ and

$$\mathbf{X}_i^{(1)} = \begin{pmatrix} \mathbf{X}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix},$$

where $\mathbf{0}/\mathbf{1}$ is a vector or matrix of 0s/1s with dimensions matching those of \mathbf{X}_i , γ_{i1} and $\boldsymbol{\theta}_1$. Also, $\gamma_{i2} = (\rho_{i21}, \rho_{i31}, \rho_{i32})'$ with $\theta_2 = \alpha$. Since γ_{i2} can be written as $\mathbf{1}\alpha$, $\mathbf{X}_i^{(2)} = \mathbf{1}$.

As an example of a non-exchangeable dependence model, suppose $\rho_{i21} = \rho_{i31} = \alpha_1$ and $\rho_{i32} = \alpha_2$. Then

$$\boldsymbol{\theta}_2 = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \text{ and } \mathbf{X}_i^{(2)} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example 3.2 MVP model. For $k_i = 3$, $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})'$, $Y_{ij} = I(Z_{ij} > 0)$ where Z_{ij} is the latent variable and $\mathbf{Z} = (Z_{i1}, Z_{i2}, Z_{i3})'$ has a multivariate normal distribution with mean $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$ and a common variance 1. For an exchangeable dependence model, $\text{Corr}(Z_{ij}, Z_{ij'}) = \rho_{ijj'} = \alpha$ for $j > j'$. Then, $\gamma_{i1} = \boldsymbol{\mu}_i$, $\boldsymbol{\theta}_1 = \boldsymbol{\beta}$ and $\mathbf{X}_i^{(1)} = \mathbf{X}_i$; $\gamma_{i2} = (\rho_{i21}, \rho_{i31}, \rho_{i32})'$, $\boldsymbol{\theta}_2 = \alpha$ and $\mathbf{X}_i^{(2)} = \mathbf{1}$.

Example 3.3 PLNM model with exchangeable dependence and $k_i = 3$. $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})'$ and $\boldsymbol{\lambda}_i = (\lambda_{i1}, \lambda_{i2}, \lambda_{i3})'$ has mean $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta}$, $\text{Var}(\boldsymbol{\lambda}_i) = \boldsymbol{\sigma}_i^2 = \sigma^2$ and $\text{Corr}(\lambda_{ij}, \lambda_{ij'}) = \rho_{ijj'} = \alpha$ for $j > j'$. Then, the parameterization is same as Example 1.

Example 3.4 Gaussian copula with Weibull margins. The marginal density function of the failure time T_{ij} is:

$$f(t; \mathbf{X}_{ij}) = \frac{b_{ij}}{a_{ij}} \left(\frac{t}{a_{ij}} \right)^{b_{ij}-1} \exp \left[- \left(\frac{t}{a_{ij}} \right)^{b_{ij}} \right],$$

where $\log a_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}$ and $b_{ij} = b$. The joint survival function of \mathbf{T}_i is specified as in (2.17). Suppose each of the off-diagonal entries of \mathbf{R}_i , $\rho_{ijj'}$, is a linear combination of $\boldsymbol{\alpha}$. Let $\boldsymbol{\rho}_i$ be the collection of all the $\rho_{ijj'}$ s. Let $\mathbf{a}_i = (a_{ij})$ and $\mathbf{b}_i = (b_{ij})$. Then the marginal parameters are specified as $\boldsymbol{\gamma}'_{i1} = (\mathbf{a}'_i, \mathbf{b}'_i)$, $\boldsymbol{\theta}'_1 = (\boldsymbol{\beta}', b)$; the dependence parameters are specified as $\boldsymbol{\gamma}_{i2} = \boldsymbol{\rho}_i$ and $\boldsymbol{\theta}_2 = \boldsymbol{\alpha}$.

3.2 Estimation Procedures based on Composite Likelihood

3.2.1 General Properties

Composite likelihood methods are estimation procedures that involve maximizing a function based on the summation of individual component log likelihoods from the marginal distributions. The CL method is appealing for the following reasons: firstly, it inherits some properties of the ordinary likelihood. In particular, under regularity conditions, the estimates based on CL are consistent and asymptotically unbiased. Secondly, under many circumstances the estimates are much easier to compute than the ML estimates.

With the growth of research in areas involving multivariate data, there is an increasing use of CL. The following are some recent examples: Xu (1996) proposed using inference functions formed by univariate marginal likelihood for multivariate discrete data; Heagerty and Lele (1998) and Currier and Lele (1999) considered pairwise composite likelihood estimation for binary spatial data; Parner (2001) used pairwise likelihood contributions to analyze familial survival data; Jöreskog and Moustaki (2001) compared a CL approach with two other approaches in factor analysis of ordinal variables; Lele and Taper (2002) considered a CL approach to estimate the variance components of a MVN random effects model.

Let $\Psi_{(n)} = \sum_{i=1}^n \Psi_i(\boldsymbol{\theta})$ be a CL function, where $\Psi_i(\boldsymbol{\theta})$ is the contribution of the i th family. As with the ML method, the problem of maximizing $\Psi_{(n)}$ usually becomes the problem of solving the equations:

$$\left. \frac{\partial \Psi_{(n)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} = \mathbf{0}.$$

Let $\boldsymbol{\psi}_i(\boldsymbol{\theta}) = \partial \Psi_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$. Then $\boldsymbol{\psi}_{(n)}(\boldsymbol{\theta}) \equiv \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta})$, called the composite score function, is the inference function generated from $\Psi_{(n)}$. This can be directly applied to the one-step procedure based on the bivariate CL function. The two-stage procedure involve two CL functions: the univariate CL function, $\Psi_{UCL(n)}$, and the bivariate CL function, $\Psi_{BCL(n)}$. The procedure becomes to solve

two sets of equations:

$$\begin{aligned} \left. \frac{\partial \Psi_{UCL(n)}(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} &= \mathbf{0}. \\ \left. \frac{\partial \Psi_{BCL(n)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} &= \mathbf{0}. \end{aligned}$$

Next, we state the asymptotic properties of the estimators from CL methods. Under similar regularity conditions applied to the MLE (Serfling, 1980), $E_{\boldsymbol{\theta}} \boldsymbol{\psi}_i(\boldsymbol{\theta}) = \mathbf{0}$ (Lindsay, 1988), therefore $\boldsymbol{\psi}_{(n)}$ is unbiased. Assume that M , the size of the largest family, is bounded. The standard theory for inference functions (Godambe, 1991) can be applied to derive the asymptotic properties of $\hat{\boldsymbol{\theta}}_n$, the estimate of $\boldsymbol{\theta}$ obtained from solving $\boldsymbol{\psi}_{(n)}(\boldsymbol{\theta}) = \mathbf{0}$ hp.

Theorem 3.1 *Assume that the size of the largest family is bounded and the families are from a finite mixture of family structures. Under the usual regularity conditions for log-likelihood of univariate and bivariate margins, as $n \rightarrow \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_{\boldsymbol{\psi}}^{-1}(\boldsymbol{\theta})),$$

where $\mathbf{G}_{\boldsymbol{\psi}}(\boldsymbol{\theta})$ is the Godambe information matrix defined as:

$$\mathbf{G}_{\boldsymbol{\psi}}(\boldsymbol{\theta}) = \left[\lim_{n \rightarrow \infty} \mathbf{D}_{\boldsymbol{\psi}}(\boldsymbol{\theta}) \right]' \left[\lim_{n \rightarrow \infty} \mathbf{M}_{\boldsymbol{\psi}}(\boldsymbol{\theta}) \right]^{-1} \left[\lim_{n \rightarrow \infty} \mathbf{D}_{\boldsymbol{\psi}}(\boldsymbol{\theta}) \right], \quad (3.1)$$

with

$$\mathbf{D}_{\boldsymbol{\psi}}(\boldsymbol{\theta}) = -\frac{1}{n} E_{\boldsymbol{\theta}} \left\{ \frac{\partial \boldsymbol{\psi}_{(n)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right\}$$

and

$$\mathbf{M}_{\boldsymbol{\psi}}(\boldsymbol{\theta}) = \frac{1}{n} E_{\boldsymbol{\theta}} \{ \boldsymbol{\psi}_{(n)}(\boldsymbol{\theta}) \boldsymbol{\psi}_{(n)}'(\boldsymbol{\theta}) \}.$$

A proof of the above theorem can be found in Godambe (1991).

An approximation of the covariance matrix of $\hat{\boldsymbol{\theta}}_n$ is taken as

$$\mathbf{V} = n^{-1} \mathbf{D}_{\boldsymbol{\psi}}^{-1} \mathbf{M}_{\boldsymbol{\psi}} (\mathbf{D}_{\boldsymbol{\psi}}^{-1})', \quad (3.2)$$

provided the inverse of $\mathbf{D}_{\boldsymbol{\psi}}$ exists.

In the next two sections, we introduce two estimating procedures based on CL. In short, we call the two methods CL1 and CL2.

3.2.2 A Two-stage Estimating Procedure (CL1)

Inference Functions

In this approach, θ_1 and θ_2 are estimated in two separate steps. The idea is: for each set of parameters, consider the CL from the marginal distribution with the lowest order in which the parameters are identifiable.

More specifically, θ_1 is identifiable from the univariate marginal distributions of \mathbf{Y} , and therefore we consider estimating it by maximizing the CL function of the univariate margins (UCL):

$$\Psi_{UCL}(\theta_1) = \sum_{i=1}^n \sum_{j=1}^{k_i} l_1(Y_{ij}; \gamma_{i1}^{(j)}(\theta_1)) \quad (3.3)$$

where l_1 is the log likelihood based on the univariate marginal density of Y_{ij} . Note that Ψ_{UCL} is a log likelihood if Y_{i1}, \dots, Y_{ik_i} are independent random variables for each i . Differentiating Ψ_{UCL} with respect to θ_1 leads to the composite score function:

$$\psi_{UCL}(\theta_1) = \frac{\partial \Psi_{UCL}(\theta_1)}{\partial \theta_1} = \sum_i (\mathbf{H}_i^{(1)}(\mathbf{X}_i^{(1)}))' \mathbf{g}_{UCL}(\mathbf{Y}_i, \gamma_{i1}), \quad (3.4)$$

where $\mathbf{H}_i^{(1)}(\mathbf{X}_i^{(1)}) = \partial \mathbf{h}_1(\mathbf{X}_i^{(1)}; \theta_1, k_i) / \partial \theta_1$ and

$$\begin{aligned} \mathbf{g}_{UCL}(\mathbf{Y}_i, \gamma_{i1}) &= \sum_j \frac{\partial l_1(Y_{ij})}{\partial \gamma_{i1}} \\ &= \left(\frac{\sum_j \partial l_1(Y_{ij})}{\partial \gamma_{i1,1}^{(1)}}, \dots, \frac{\sum_j \partial l_1(Y_{ik_i})}{\partial \gamma_{i1,1}^{(k_i)}}, \frac{\sum_j \partial l_1(Y_{i1})}{\partial \gamma_{i1,2}^{(1)}}, \dots, \frac{\sum_j \partial l_1(Y_{ik_i})}{\partial \gamma_{i1,d_1}^{(k_i)}} \right)' \\ &= \left(\frac{\partial l_1(Y_{i1})}{\partial \gamma_{i1,1}^{(1)}}, \dots, \frac{\partial l_1(Y_{ik_i})}{\partial \gamma_{i1,1}^{(k_i)}}, \frac{\partial l_1(Y_{i1})}{\partial \gamma_{i1,2}^{(1)}}, \dots, \frac{\partial l_1(Y_{ik_i})}{\partial \gamma_{i1,d_1}^{(k_i)}} \right)'. \end{aligned}$$

In the second step, the first element $\sum_j \partial l_1(Y_{ij}) / \partial \gamma_{i1,1}^{(1)} = \partial l_1(Y_{i1}) / \partial \gamma_{i1,1}^{(1)}$ since $\gamma_{i1,1}^{(1)}$ is only related to Y_{i1} . It is the same for other elements. When $\gamma_{i1}^{(j)} = \mathbf{X}_i^{(1)} \theta_1$, $\mathbf{H}_i^{(1)}(\mathbf{X}_i^{(1)}) = \mathbf{X}_i^{(1)}$.

θ_2 can be identified by the bivariate marginal distribution of \mathbf{Y} , so we consider estimating it by maximizing the composite likelihood function of the bivariate margins (BCL). The contribution of \mathbf{Y}_i is the summation of the log likelihoods of all the possible pairs $(Y_{ij}, Y_{ij'})$, $j > j'$. Let l_2 denote the bivariate log likelihood. The BCL function is

$$\Psi_{BCL}(\theta_1, \theta_2) = \sum_{i=1}^n \sum_{j > j'} l_2(Y_{ij}, Y_{ij'}; \gamma_{i1}^{(j)}(\theta_1), \gamma_{i1}^{(j')}(\theta_1), \gamma_{i2}^{(j,j')}(\theta_2)). \quad (3.5)$$

The estimating function for θ_2 is

$$\psi_{2BCL}(\theta_1, \theta_2) = \frac{\partial \Psi_{BCL}(\theta_1, \theta_2)}{\partial \theta_2} = \sum_i (\mathbf{H}_i^{(2)}(\mathbf{X}_i^{(2)})) \mathbf{g}_{2BCL}(\mathbf{Y}_i, \gamma_i), \quad (3.6)$$

where

$$\mathbf{g}_{2BCL}(\mathbf{Y}_i, \gamma_i) = \sum_{j>j'} \frac{\partial l_2(Y_{ij}, Y_{ij'})}{\partial \gamma_{i2}} = \left(\frac{\partial l_2(Y_{ij}, Y_{ij'})}{\partial \gamma_{i2}^{(j,j')}} \right), j > j'$$

and $\mathbf{H}_i^{(2)}(\mathbf{X}_i^{(2)}) = \partial \mathbf{h}_2(\mathbf{X}_i^{(2)}; \theta_2, k_i) / \partial \theta_2$.

Let $\psi_{CL1} = (\psi_{UCL}, \psi_{2BCL})'$ and $\mathbf{g}_{CL1} = (\mathbf{g}_{UCL}, \mathbf{g}_{2BCL})'$. Then

$$\psi_{CL1} = \sum_i (\mathbf{H}_i(\tilde{\mathbf{X}}_i))' \mathbf{g}_{CL1}(\mathbf{Y}_i, \gamma_i). \quad (3.7)$$

Implementation

Computationally, the estimating procedure can be carried out in two steps:

Step 1: estimate θ_1 by $\hat{\theta}_{1,CL1}$, a solution of $\psi_{UCL}(\theta_1) = \mathbf{0}$ (usually equivalent to maximizing $\Psi_{UCL}(\theta_1)$).

Step 2: estimate θ_2 by $\hat{\theta}_{2,CL1}$, a solution of $\psi_{2BCL}(\hat{\theta}_{1,CL1}, \theta_2) = \mathbf{0}$ (usually equivalent to maximizing $\Psi_{BCL}(\hat{\theta}_{1,CL1}, \theta_2)$).

Asymptotic Variance under Linearity

We first derive the Godambe information matrix of ψ_{CL1} assuming a linear relationship between $\tilde{\mathbf{X}}_i$ and γ_i , that is, $\gamma_i = \tilde{\mathbf{X}}_i \theta$. Then \mathbf{H}_i can be replaced by $\tilde{\mathbf{X}}_i$. Firstly, $\mathbf{D}_\psi = \mathbf{D}_{CL1} = n^{-1} \sum_i \mathbf{X}_i' \Delta_i \mathbf{X}_i$ with

$$\Delta_i = -E \frac{\partial \mathbf{g}_{CL1}(\mathbf{Y}_i, \gamma_i)}{\partial \gamma_i'} = \begin{pmatrix} \Delta_i^{(1,1)} & \mathbf{0} \\ \Delta_i^{(2,1)} & \Delta_i^{(2,2)} \end{pmatrix} = \begin{pmatrix} -E \frac{\partial \mathbf{g}_{UCL}(\mathbf{Y}_i, \gamma_i)}{\partial \gamma_{i1}} & \mathbf{0} \\ -E \frac{\partial \mathbf{g}_{2BCL}(\mathbf{Y}_i, \gamma_i)}{\partial \gamma_{i1}} & -E \frac{\partial \mathbf{g}_{2BCL}(\mathbf{Y}_i, \gamma_i)}{\partial \gamma_{i2}} \end{pmatrix}.$$

So

$$\mathbf{D}_\psi = \mathbf{D}_{CL1} = \frac{1}{n} \sum_i \begin{pmatrix} \mathbf{X}_{i1}' \Delta_i^{(1,1)} \mathbf{X}_{i1} & \mathbf{0} \\ \mathbf{X}_{i2}' \Delta_i^{(2,1)} \mathbf{X}_{i1} & \mathbf{X}_{i2}' \Delta_i^{(2,2)} \mathbf{X}_{i2} \end{pmatrix}. \quad (3.8)$$

Also, $\mathbf{M}_\psi = \mathbf{M}_{CL1} = n^{-1} \sum_i \mathbf{X}'_i \boldsymbol{\Omega}_i \mathbf{X}_i$ with

$$\boldsymbol{\Omega}_i = \text{Var}(\mathbf{g}_{CL1}(\mathbf{Y}_i)) = \begin{pmatrix} \boldsymbol{\Omega}_i^{(1,1)} & \boldsymbol{\Omega}_i^{(1,2)} \\ \boldsymbol{\Omega}_i^{(2,1)} & \boldsymbol{\Omega}_i^{(2,2)} \end{pmatrix} = \begin{pmatrix} \text{Var}(\mathbf{g}_{UCL}(\mathbf{Y}_i)) & \text{Cov}(\mathbf{g}_{UCL}(\mathbf{Y}_i), \mathbf{g}_2(\mathbf{Y}_i)) \\ \text{Cov}(\mathbf{g}_{2BCL}(\mathbf{Y}_i), \mathbf{g}_1(\mathbf{Y}_i)) & \text{Var}(\mathbf{g}_{2BCL}(\mathbf{Y}_i)) \end{pmatrix}.$$

Then

$$\mathbf{M}_\psi = \mathbf{M}_{CL1} = \frac{1}{n} \sum_i \begin{pmatrix} \mathbf{X}'_{i1} \boldsymbol{\Omega}_i^{(1,1)} \mathbf{X}_{i1} & \mathbf{X}'_{i1} \boldsymbol{\Omega}_i^{(1,2)} \mathbf{X}_{i2} \\ \mathbf{X}'_{i2} \boldsymbol{\Omega}_i^{(2,1)} \mathbf{X}_{i1} & \mathbf{X}'_{i2} \boldsymbol{\Omega}_i^{(2,2)} \mathbf{X}_{i2} \end{pmatrix}. \quad (3.9)$$

The approximation of the covariance matrix of $\hat{\boldsymbol{\theta}}_{CL1}$ is $\mathbf{V}_{CL1} = n^{-1} \mathbf{D}_{CL1}^{-1} \mathbf{M}_{CL1} (\mathbf{D}_{CL1}^{-1})'$. Write

$$\mathbf{D}_{CL1} = \frac{1}{n} \begin{pmatrix} \mathbf{D}_{1,11} & \mathbf{0} \\ \mathbf{D}_{1,21} & \mathbf{D}_{1,22} \end{pmatrix} \text{ and } \mathbf{M}_{CL1} = \frac{1}{n} \begin{pmatrix} \mathbf{M}_{1,11} & \mathbf{M}_{1,12} \\ \mathbf{M}_{1,21} & \mathbf{M}_{1,22} \end{pmatrix} \quad (3.10)$$

where $\mathbf{D}_{1,l'l'} = \sum_i \mathbf{X}'_{il} \boldsymbol{\Delta}_i^{(l,l')} \mathbf{X}_{il'}$ and $\mathbf{M}_{1,l'l'} = \sum_i \mathbf{X}'_{il} \boldsymbol{\Omega}_i^{(l,l')} \mathbf{X}_{il'}$. We have

$$\begin{pmatrix} \mathbf{D}_{1,11} & \mathbf{0} \\ \mathbf{D}_{1,21} & \mathbf{D}_{1,22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{D}_{1,11}^{-1} & \mathbf{0} \\ -\mathbf{D}_{1,22}^{-1} \mathbf{D}_{1,21} \mathbf{D}_{1,11}^{-1} & \mathbf{D}_{1,22}^{-1} \end{pmatrix}.$$

It follows that

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\theta}}_{1,CL1}) &\approx \mathbf{D}_{1,11}^{-1} \mathbf{M}_{1,11} (\mathbf{D}_{1,11}^{-1})' \\ &= \left(\sum_i \mathbf{X}'_{i1} \boldsymbol{\Delta}_i^{(1,1)} \mathbf{X}_{i1} \right)^{-1} \left(\sum_i \mathbf{X}'_{i1} \boldsymbol{\Omega}_i^{(1,1)} \mathbf{X}_{i1} \right) \left(\sum_i \mathbf{X}'_{i1} \boldsymbol{\Delta}_i^{(1,1)} \mathbf{X}_{i1} \right)^{-1}. \end{aligned} \quad (3.11)$$

$$\text{Var}(\hat{\boldsymbol{\theta}}_{2,CL1}) \approx \begin{pmatrix} -\mathbf{D}_{1,22}^{-1} \mathbf{D}_{1,21} \mathbf{D}_{1,11}^{-1} & \mathbf{D}_{1,22}^{-1} \end{pmatrix} \mathbf{M}_{CL1} \begin{pmatrix} -\mathbf{D}_{1,22}^{-1} \mathbf{D}_{1,21} \mathbf{D}_{1,11}^{-1} \\ \mathbf{D}_{1,22}^{-1} \end{pmatrix}. \quad (3.12)$$

$$\text{Cov}(\hat{\boldsymbol{\theta}}_{1,CL1}, \hat{\boldsymbol{\theta}}_{2,CL1}) \approx \begin{pmatrix} \mathbf{D}_{1,11}^{-1} & \mathbf{0} \end{pmatrix} \mathbf{M}_{CL1} \begin{pmatrix} -\mathbf{D}_{1,22}^{-1} \mathbf{D}_{1,21} \mathbf{D}_{1,11}^{-1} \\ \mathbf{D}_{1,22}^{-1} \end{pmatrix}. \quad (3.13)$$

Performance of CL1 Estimates in Independent Case

In this two-stage approach, in the first step, Y_{i1}, \dots, Y_{ik_i} are treated as if they are independent and in the second step, the pairs $(Y_{ij}, Y_{ij'})$ are treated as if they are independent. If Y_{i1}, \dots, Y_{ik_i} are actually independent, is this approach equivalent to the MLE method in terms of asymptotic variance? To answer this question, we compare the Godambe information matrix of ψ_{CL1} and Fisher's information matrix of the full likelihood score function. The Fisher information matrix is defined by

$$\bar{\mathbf{F}}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i \mathbf{E} \left[\frac{\partial l_{k_i}(\mathbf{Y}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \left[\frac{\partial l_{k_i}(\mathbf{Y}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]', \quad (3.14)$$

where l_{k_i} is the log likelihood function of \mathbf{Y}_i . $\bar{\mathbf{F}}(\boldsymbol{\theta})$ represents the average Fisher information per family.

Theorem 3.2 *Assume the same conditions in Theorem 3.1. Let $\mathbf{G}_{CL1}(\boldsymbol{\theta})$ be the Godambe information matrix defined in (3.1) with D_ψ and M_ψ as in (3.8) and (3.9) and $\bar{\mathbf{F}}(\boldsymbol{\theta})$ be the Fisher information matrix as defined in (3.14). Suppose $\boldsymbol{\theta}'_0 = (\boldsymbol{\theta}'_{0,1}, \boldsymbol{\theta}'_{0,2})$ is the true value of $\boldsymbol{\theta}$. When Y_{i1}, \dots, Y_{ik_i} are independent, i.e. $f_k(\mathbf{y}_i; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) = \prod_j f_1(Y_{ij}; \boldsymbol{\theta}_{0,1})$, we have (a)*

$$\mathbf{G}_{CL1}(\boldsymbol{\theta}_0) = \bar{\mathbf{F}}(\boldsymbol{\theta}_0)$$

when $\boldsymbol{\theta}_{0,2}$ is in the interior of the parameter space; and (b)

$$\lim_{\boldsymbol{\theta}_2 \rightarrow \boldsymbol{\theta}_{0,2}} \mathbf{G}_{CL1}(\boldsymbol{\theta}) = \lim_{\boldsymbol{\theta}_2 \rightarrow \boldsymbol{\theta}_{0,2}} \bar{\mathbf{F}}(\boldsymbol{\theta})$$

when $\boldsymbol{\theta}_{0,2}$ is on the boundary of the parameter space and $\bar{\mathbf{F}}(\boldsymbol{\theta})$ and $\mathbf{G}_{CL1}(\boldsymbol{\theta})$ are continuous at $\boldsymbol{\theta}_{0,2}$.

Before we prove Theorem 3.2, we first introduce some notation used in this proof. Here \mathbf{Y} is the response vector of a family in general, so the family index i is suppressed. Suppose h is a function of \mathbf{Y} and $\boldsymbol{\theta}' = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)$. We use a dot on top of h to denote the first partial derivative, i.e.,

$$\dot{h}(\mathbf{Y}) = \frac{\partial h(\mathbf{Y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \dot{h}^{(l)}(\mathbf{Y}) = \frac{\partial h(\mathbf{Y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_l}.$$

So,

$$\dot{h}(\mathbf{Y}) = \begin{bmatrix} \dot{h}^{(1)}(\mathbf{Y}) \\ \dot{h}^{(2)}(\mathbf{Y}) \end{bmatrix}.$$

We use $\dot{h}(\mathbf{Y}; \boldsymbol{\theta}_0)$ to represent the derivative evaluated at $\boldsymbol{\theta}_0$, i.e.,

$$\dot{h}(\mathbf{Y}; \boldsymbol{\theta}_0) = \left. \frac{\partial h(\mathbf{Y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad \text{and} \quad \dot{h}^{(l)}(\mathbf{Y}; \boldsymbol{\theta}_0) = \left. \frac{\partial h(\mathbf{Y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_l} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

We use double dots to denote the second partial derivative, i.e.,

$$\ddot{h}^{(l,l')}(\mathbf{Y}) = \frac{\partial^2 h(\mathbf{Y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_l \partial \boldsymbol{\theta}_{l'}} \quad \text{and} \quad \ddot{h}^{(l,l')}(\mathbf{Y}; \boldsymbol{\theta}_0) = \left. \frac{\partial^2 h(\mathbf{Y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_l \partial \boldsymbol{\theta}_{l'}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

where $l, l' = 1, 2$.

Let f_k and l_k be the pdf/pmf and log likelihood function of \mathbf{Y} , respectively, and ν be a Lebesgue measure when \mathbf{Y} is continuous or a counting measure when \mathbf{Y} is discrete. The regularity conditions (Serfling, 1980) ensure that

$$\frac{\partial \int f_k(\mathbf{y}; \boldsymbol{\theta}) d\nu(\mathbf{y})}{\partial \boldsymbol{\theta}} = \int \frac{\partial f_k(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\nu(\mathbf{y})$$

and

$$\frac{\partial \int l_k(\mathbf{y}; \boldsymbol{\theta}) d\nu(\mathbf{y})}{\partial \boldsymbol{\theta}} = \int \frac{\partial l_k(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\nu(\mathbf{y})$$

Thus, we have the following three results which are needed in the proof of Theorem 3.2. Results 3.1 and 3.2 are standard results of likelihood theory.

Result 3.1

$$\int \frac{\partial f_k(\mathbf{y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_l} d\nu(\mathbf{y}) = \frac{\partial}{\partial \boldsymbol{\theta}_l} \int f_k(\mathbf{y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\nu(\mathbf{y}) = 0.$$

This leads to $E_{\boldsymbol{\theta}_0} \dot{l}_k^{(l)}(\boldsymbol{\theta}_0) = 0$.

Result 3.2 $E_{\boldsymbol{\theta}_0} \ddot{l}_k^{(l,l')}(\mathbf{Y}; \boldsymbol{\theta}_0) = -E_{\boldsymbol{\theta}_0} \dot{l}_k^{(l)}(\mathbf{Y}; \boldsymbol{\theta}_0) [\dot{l}_k^{(l')}(\mathbf{Y}; \boldsymbol{\theta}_0)]'$.

Result 3.3 $E_{\boldsymbol{\theta}_0} \dot{l}_1^{(1)}(Y_j; \boldsymbol{\theta}_{0,1}) \dot{l}_k^{(2)}(\mathbf{Y}; \boldsymbol{\theta}_0) = 0$.

Proof: By definition,

$$\begin{aligned} \dot{l}_1^{(1)}(Y_j; \boldsymbol{\theta}_{0,1}) &= \dot{f}_1^{(1)}(Y_j; \boldsymbol{\theta}_{0,1}) / f_1(Y_j; \boldsymbol{\theta}_{0,1}) \\ \dot{l}_k^{(2)}(\mathbf{Y}; \boldsymbol{\theta}_0) &= \dot{f}_k^{(2)}(\mathbf{Y}; \boldsymbol{\theta}_0) / f_k(\mathbf{Y}; \boldsymbol{\theta}_0). \end{aligned}$$

Since $f_k(\mathbf{Y}; \boldsymbol{\theta}_0) = f_1(Y_j; \boldsymbol{\theta}_{0,1})f_{k-1}^*(\mathbf{Y}_{-j}|Y_j; \boldsymbol{\theta}_0)$, where \mathbf{Y}_{-j} is \mathbf{Y} without Y_j and f_{k-1}^* is the conditional pdf/pmf of \mathbf{Y}_{-j} ,

$$i_k^{(2)}(\mathbf{Y}; \boldsymbol{\theta}_0) = \frac{f_1(y_j; \boldsymbol{\theta}_{0,1})}{f_k(\mathbf{Y}; \boldsymbol{\theta}_0)} \dot{f}_{k-1}^{*(2)}(\mathbf{Y}_{-j}|Y_j; \boldsymbol{\theta}_0).$$

Therefore,

$$i_1^{(1)}(Y_j; \boldsymbol{\theta}_{0,1}) i_k^{(2)}(\mathbf{Y}; \boldsymbol{\theta}_0) = \frac{\dot{f}_1^{(1)}(Y_j; \boldsymbol{\theta}_{0,1}) \dot{f}_{k-1}^{*(2)}(\mathbf{Y}_{-j}|Y_j; \boldsymbol{\theta}_0)}{f_k(\mathbf{Y}; \boldsymbol{\theta}_0)}.$$

Thus,

$$\begin{aligned} E_{\boldsymbol{\theta}_0} i_1^{(1)}(Y_j; \boldsymbol{\theta}_{0,1}) i_k^{(2)}(\mathbf{Y}; \boldsymbol{\theta}_0) &= E_{\boldsymbol{\theta}_0} \frac{\dot{f}_1^{(1)}(Y_j; \boldsymbol{\theta}_{0,1}) \dot{f}_{k-1}^{*(2)}(\mathbf{Y}_{-j}|Y_j; \boldsymbol{\theta}_0)}{f_k(\mathbf{Y}; \boldsymbol{\theta}_0)} \\ &= \int \dot{f}_1^{(1)}(y_j; \boldsymbol{\theta}_0) \dot{f}_{k-1}^{*(2)}(\mathbf{y}_{-j}|y_j; \boldsymbol{\theta}_0) d\nu(\mathbf{y}) \\ &= \int \dot{f}_1^{(1)}(y_j; \boldsymbol{\theta}_0) \left[\int \dot{f}_{k-1}^{*(2)}(\mathbf{y}_{-j}|y_j; \boldsymbol{\theta}_0) d\nu(\mathbf{y}_{-j}) \right] d\nu(y_j). \end{aligned}$$

Similar to Result 3.1, $\int \dot{f}_{k-1}^{*(2)}(\mathbf{y}_{-j}|y_j; \boldsymbol{\theta}_0) d\nu(\mathbf{y}_{-j}) = 0$. Therefore, $E_{\boldsymbol{\theta}_0} i_1^{(1)}(Y_j; \boldsymbol{\theta}_{0,1}) i_k^{(2)}(\mathbf{Y}; \boldsymbol{\theta}_0) = 0$.

Proof of Theorem 3.2:

The asymptotic properties of the CL estimate and the MLE estimate are both derived under the condition that the parameters are in the interior of the parameter space. In (a), $\boldsymbol{\theta}_{0,2}$ satisfies this condition. We first prove that $\mathbf{M}_{CL1}(\boldsymbol{\theta}_0) = \mathbf{D}_{CL1}(\boldsymbol{\theta}_0)$, so that

$$G_{CL1}(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} \mathbf{M}_{CL1}(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} \mathbf{D}_{CL1}(\boldsymbol{\theta}_0).$$

As defined in (3.10), \mathbf{D}_{CL1} and \mathbf{M}_{CL1} are partitioned into submatrices $\mathbf{D}_{1,l'}$ and $\mathbf{M}_{1,l'}$, where $l, l'=1$ or 2 .

(1) Claim $\mathbf{D}_{1,11}(\boldsymbol{\theta}_0) = \mathbf{M}_{1,11}(\boldsymbol{\theta}_0)$.

By Result 3.2,

$$\begin{aligned} \mathbf{D}_{1,11}(\boldsymbol{\theta}_0) &= -E \left\{ \dot{\boldsymbol{\psi}}_{UCL}^{(1)}(\boldsymbol{\theta}_0) \right\} \\ &= - \sum_i \sum_j E \ddot{l}_1^{(1,1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \\ &= \sum_i \sum_j E \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1})'. \end{aligned}$$

By definition,

$$\begin{aligned}
\mathbf{M}_{1,11}(\boldsymbol{\theta}_0) &= \mathbb{E} \{ \boldsymbol{\psi}_{UCL}(\boldsymbol{\theta}_0) \boldsymbol{\psi}_{UCL}'(\boldsymbol{\theta}_0) \} \\
&= \mathbb{E} \left(\sum_i \sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right) \left(\sum_i \sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right)' \\
&= \sum_i \mathbb{E} \left(\sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right) \left(\sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right)'.
\end{aligned}$$

The last equality holds since $\mathbb{E} \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) [\dot{l}_1^{(1)}(Y_{i'j'}; \boldsymbol{\theta}_{0,1})]' = 0$ for $i \neq i'$ under the assumption that individuals from different families are independent. Under independence of Y_{ij} and $Y_{ij'}$, $\mathbb{E} \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) [\dot{l}_1^{(1)}(Y_{ij'}; \boldsymbol{\theta}_{0,1})]' = 0$ for $j \neq j'$. Therefore

$$\mathbf{M}_{1,11}(\boldsymbol{\theta}_0) = \sum_i \sum_j \mathbb{E} \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1})' = \mathbf{D}_{1,11}(\boldsymbol{\theta}_0).$$

(2) Claim $\mathbf{D}_{1,21}(\boldsymbol{\theta}_0) = \mathbf{0}$.

$$\begin{aligned}
\mathbf{D}_{1,21}(\boldsymbol{\theta}_0) &= -\mathbb{E} \{ \dot{\boldsymbol{\psi}}_{2BCL}^{(1)}(\boldsymbol{\theta}_0) \} \\
&= -\sum_i \sum_{j>j'} \mathbb{E} \dot{l}_2^{(2,1)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \\
&= \sum_i \sum_{j>j'} \mathbb{E} \left(\dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\dot{l}_2^{(1)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right)',
\end{aligned}$$

by Result 3.2. Under independence,

$$\dot{l}_2^{(1)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) = \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) + \dot{l}_1^{(1)}(Y_{ij'}; \boldsymbol{\theta}_{0,1}).$$

By Result 3.3 (with $k = 2$),

$$\mathbb{E} \left(\dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) + \dot{l}_1^{(1)}(Y_{ij'}; \boldsymbol{\theta}_{0,1}) \right)' = \mathbf{0};$$

therefore $\mathbf{D}_{1,21}(\boldsymbol{\theta}_0) = \mathbf{0}$.

(3) Claim $\mathbf{M}_{1,21}(\boldsymbol{\theta}_0) = \mathbf{0}$.

$$\begin{aligned}
\mathbf{M}_{1,21}(\boldsymbol{\theta}_0) &= \mathbf{E} \psi_{2BCL}(\boldsymbol{\theta}_0) \psi'_{UCL}(\boldsymbol{\theta}_0) \\
&= \mathbf{E} \left(\sum_i \sum_{j>j} \dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\sum_i \sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right)' \\
&= \sum_i \mathbf{E} \left(\sum_{j>j} \dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right)' \\
&= \sum_i \sum_{j>j'} \sum_{j''} \mathbf{E} \left(\dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\dot{l}_1^{(1)}(Y_{ij''}; \boldsymbol{\theta}_{0,1}) \right)'.
\end{aligned}$$

Let (A) = $\mathbf{E} \left(\dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\dot{l}_1^{(1)}(Y_{ij''}; \boldsymbol{\theta}_{0,1}) \right)'$, an arbitrary term in the above summation. If $j \neq j''$ and $j' \neq j''$, under independence,

$$(A) = \left(\mathbf{E} \dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\mathbf{E} \dot{l}_1^{(1)}(Y_{ij''}; \boldsymbol{\theta}_{0,1}) \right)' = \mathbf{0},$$

since by Result 3.1, $\mathbf{E} \dot{l}_1^{(1)}(Y_{ij''}; \boldsymbol{\theta}_{0,1}) = \mathbf{0}$. If $j = j''$ or $j' = j''$, (A) = $\mathbf{0}$ by Result 3.3 with $k = 2$.

Therefore $\mathbf{M}_{1,21}(\boldsymbol{\theta}_0) = \mathbf{0}$. By symmetry, $\mathbf{M}_{1,12}(\boldsymbol{\theta}_0) = \mathbf{0}$ as well.

(4) Claim $\mathbf{D}_{1,22}(\boldsymbol{\theta}_0) = \mathbf{M}_{1,22}(\boldsymbol{\theta}_0)$.

$$\begin{aligned}
\mathbf{D}_{1,22}(\boldsymbol{\theta}_0) &= -\mathbf{E} \left\{ \dot{\psi}_{2BCL}^{(2)}(\boldsymbol{\theta}_0) \right\} \\
&= -\sum_i \sum_{j>j} \mathbf{E} \ddot{l}_2^{(2,2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \\
&= \sum_i \sum_{j>j} \mathbf{E} \left(\dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right)',
\end{aligned}$$

by Result 3.2. Meanwhile,

$$\begin{aligned}
\mathbf{M}_{1,22}(\boldsymbol{\theta}_0) &= \mathbf{E} \left\{ \psi_{2BCL}(\boldsymbol{\theta}_0) \psi'_{2BCL}(\boldsymbol{\theta}_0) \right\} \\
&= \mathbf{E} \left(\sum_i \sum_{j>j'} \dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\sum_i \sum_{j>j'} \dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right)' \\
&= \sum_i \mathbf{E} \left(\sum_{j>j'} \dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\sum_{l>l'} \dot{l}_2^{(2)}(Y_{il}, Y_{il'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right)'.
\end{aligned}$$

Let (B) = $\mathbf{E} \left(\dot{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\dot{l}_2^{(2)}(Y_{il}, Y_{il'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right)'$, an arbitrary term in the above summation. If (j, j') and (l, l') are two pairs with no common individual, (B) = $\mathbf{0}$. On the other hand,

if (j, j') and (l, l') share one common individual, say $j = l$, but $j' \neq l'$, then

$$(B) = \iiint \left(\frac{f_2^{(2)}(y_{ij}, y_{ij'}; \theta_{0,1}, \theta_{0,2})}{f_1(y_{ij}, \theta_{0,1}) f_1(y_{ij'}, \theta_{0,1})} \right) \left(\frac{f_2^{(2)}(y_{ij}, y_{il'}; \theta_{0,1}, \theta_{0,2})}{f_1(y_{ij}, \theta_{0,1}) f_1(y_{il'}, \theta_{0,1})} \right)' \\ f_1(y_{ij}, \theta_{0,1}) f_1(y_{ij'}, \theta_{0,1}) f_1(y_{il'}, \theta_{0,1}) d\nu(y_{ij}) d\nu(y_{ij'}) d\nu(y_{il'}).$$

Since $f_2^{(2)}(y_{ij}, y_{ij'}; \theta_{0,1}, \theta_{0,2}) = f_1(y_{ij}, \theta_{0,1}) f_1^*(y_{ij'} | Y_{ij} = y_{ij}; \theta_{0,1}, \theta_{0,2})$, where f^* is the conditional pdf/pmf,

$$(B) = \iiint [f_2^{*(2)}(Y_{ij}, Y_{ij'}; \theta_{0,1}, \theta_{0,2})] [f_2^{*(2)}(Y_{ij}, Y_{il'}; \theta_{0,1}, \theta_{0,2})]' f_1(y_{ij}, \theta_{0,1}) d\nu(y_{ij}) d\nu(y_{ij'}) d\nu(y_{il'}) \\ = \int \left[\int f_2^{*(2)}(Y_{ij}, Y_{ij'}; \theta_{0,1}, \theta_{0,2}) d\nu(y_{ij'}) \right] \left[\int f_2^{*(2)}(Y_{ij}, Y_{il'}; \theta_{0,1}, \theta_{0,2}) d\nu(y_{il'}) \right]' f_1(y_{ij}, \theta_{0,1}) d\nu(y_{ij}) \\ = 0,$$

since by Result 3.1 the integrals inside the square brackets are zero. It follows that

$$\mathbf{M}_{1,22}(\theta_0) = \sum_i \sum_{j>j} E(i_2^{(2)}(Y_{ij}, Y_{ij'}; \theta_{0,1}, \theta_{0,2})) (i_2^{(2)}(Y_{ij}, Y_{ij'}; \theta_{0,1}, \theta_{0,2}))' = \mathbf{D}_{1,22}(\theta_0).$$

Based on (1) - (4), $\mathbf{M}_{CL1}(\theta_0) = \mathbf{D}_{CL1}(\theta_0)$, therefore

$$\mathbf{G}_{CL1}(\theta_0) = \lim_{n \rightarrow \infty} \mathbf{M}_{CL1}(\theta_0) = \lim_{n \rightarrow \infty} \mathbf{D}_{CL1}(\theta_0).$$

Next we prove that $\bar{\mathbf{F}}(\theta_0) = \mathbf{M}_{CL1}(\theta_0)$. We write

$$\bar{\mathbf{F}}(\theta_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \begin{pmatrix} \mathbf{F}_{11}(\theta_0) & \mathbf{F}_{12}(\theta_0) \\ \mathbf{F}_{21}(\theta_0) & \mathbf{F}_{22}(\theta_0) \end{pmatrix},$$

with

$$\mathbf{F}_{ll'}(\theta_0) = \sum_i E i_{k_i}^{(l)}(\mathbf{Y}_i; \theta_0) i_{k_i}^{(l')}(\mathbf{Y}_i; \theta_0)', \quad l, l' = 1, 2$$

(5) Claim $\mathbf{F}_{11}(\theta_0) = \mathbf{M}_{1,11}(\theta_0)$.

$$i_{k_i}^{(1)}(\mathbf{Y}_i; \theta_0) = \left. \frac{\partial l_{k_i}(\mathbf{Y}_i; \theta_1, \theta_2)}{\partial \theta_1} \right|_{\substack{\theta_1 = \theta_{0,1} \\ \theta_2 = \theta_{0,2}}} = \left. \frac{\partial l_{k_i}(\mathbf{Y}_i; \theta_1, \theta_{0,2})}{\partial \theta_1} \right|_{\theta_1 = \theta_{0,1}}. \quad (3.15)$$

Under independence, $l_{k_i}(\mathbf{Y}_i; \theta_1, \theta_{0,2}) = \sum_j l_1(Y_{ij}; \theta_{0,1})$. Therefore,

$$i_{k_i}^{(1)}(\mathbf{Y}_i; \theta_0) = \sum_j i_1^{(1)}(Y_{ij}; \theta_{0,1}).$$

It follows that

$$\mathbf{F}_{11}(\boldsymbol{\theta}_0) = \sum_i \mathbf{E} \left(\sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right) \left(\sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right)' = \mathbf{M}_{1,11}(\boldsymbol{\theta}_0).$$

(6) Claim $\mathbf{F}_{12}(\boldsymbol{\theta}_0) = \mathbf{0}$.

$$\begin{aligned} \mathbf{F}_{12}(\boldsymbol{\theta}_0) &= \sum_i \mathbf{E} \dot{l}_{k_i}^{(1)}(\mathbf{Y}_i, \boldsymbol{\theta}_0) [\dot{l}_{k_i}^{(2)}(\mathbf{Y}_i, \boldsymbol{\theta}_0)]' \\ &= \sum_i \mathbf{E} \left(\sum_j \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right) \left(\dot{l}_{k_i}^{(2)}(\mathbf{Y}_i, \boldsymbol{\theta}_0) \right)' \\ &= \sum_i \sum_j \mathbf{E} \dot{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) [\dot{l}_{k_i}^{(2)}(\mathbf{Y}_i, \boldsymbol{\theta}_0)]'. \end{aligned}$$

By Result 3.3, $\mathbf{F}_{12}(\boldsymbol{\theta}_0) = \mathbf{0}$.

(7) Claim $\mathbf{F}_{22}(\boldsymbol{\theta}_0) = \mathbf{M}_{1,22}(\boldsymbol{\theta}_0)$.

$$\mathbf{F}_{22}(\boldsymbol{\theta}_0) = \sum_i \mathbf{E} \dot{l}_{k_i}^{(2)}(\mathbf{Y}_i, \boldsymbol{\theta}_0) [\dot{l}_{k_i}^{(2)}(\mathbf{Y}_i, \boldsymbol{\theta}_0)]'.$$

As defined before, l_{k_i} is a function of γ_2 which in turn is a function of $\boldsymbol{\theta}_2$. So for a vector \mathbf{Y} of length k , we can write

$$l_k(\mathbf{Y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = l_k(\mathbf{Y}; \gamma_1(\boldsymbol{\theta}_1), \gamma_2(\boldsymbol{\theta}_2)).$$

To prove (7), it suffices to show that

$$\dot{l}_k^{(2)}(\mathbf{Y}; \gamma_1(\boldsymbol{\theta}_{0,1}), \gamma_2(\boldsymbol{\theta}_{0,2})) = \sum_{j>j'} \dot{l}_2^{(2)}(Y_j, Y_{j'}; \gamma_1^{(j)}(\boldsymbol{\theta}_{0,1}), \gamma_1^{(j')}(\boldsymbol{\theta}_{0,1}), \gamma_2^{(j,j')}(\boldsymbol{\theta}_{0,2})),$$

since

$$\mathbf{M}_{1,22}(\boldsymbol{\theta}) = \sum_i \mathbf{E} \left(\sum_{j>j'} \dot{l}_2^{(2)}(Y_j, Y_{j'}; \gamma_2^{(j,j')}(\boldsymbol{\theta}_{0,2})) \right) \left(\sum_{j>j'} \dot{l}_2^{(2)}(Y_j, Y_{j'}; \gamma_2^{(j,j')}(\boldsymbol{\theta}_{0,2})) \right)'.$$

(For a simpler notation, we omit γ_1 in the likelihood.) We can write

$$\dot{l}_k^{(2)}(\mathbf{Y}; \boldsymbol{\theta}_0) = \sum_{j>j'} \frac{\partial l_k(\mathbf{Y}; \gamma_2(\boldsymbol{\theta}_2))}{\partial \gamma_2^{(j,j')}} \frac{\partial \gamma_2^{(j,j')}(\boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Let $\gamma_2^{-(j,j')}$ be the vector of parameters in γ_2 without $\gamma_2^{(j,j')}$. We have

$$\begin{aligned} \frac{\partial l_k(\mathbf{Y}; \gamma_2(\boldsymbol{\theta}_2))}{\partial \gamma_2^{(j,j')}} \Big|_{\boldsymbol{\theta}_2=\boldsymbol{\theta}_{0,2}} &= \frac{\partial l_k(\mathbf{Y}; \gamma_2^{(j,j')}, \gamma_2^{-(j,j')}(\boldsymbol{\theta}_{0,2}))}{\partial \gamma_2^{(j,j')}} \Big|_{\gamma_2^{(j,j')}=\gamma_2^{(j,j')}(\boldsymbol{\theta}_{0,2})}, \\ &= \frac{\partial l_2(Y_j, Y_{j'}; \gamma_2^{(j,j')})}{\partial \gamma_2^{(j,j')}} \Big|_{\boldsymbol{\theta}_2=\boldsymbol{\theta}_{0,2}}. \end{aligned}$$

Under independence,

$$l_k(\mathbf{Y}; \gamma_2^{(j,j')}, \gamma_2^{-(j,j')}(\theta_{0,2})) = l_2(Y_j, Y_{j'}; \gamma_2^{(j,j')}) + \sum_{l \neq j, j'} l_1(Y_l).$$

Therefore,

$$\begin{aligned} i_k^{(2)}(\mathbf{Y}; \theta_0) &= \sum_{j > j'} \frac{\partial l_2(Y_j, Y_{j'}; \gamma_2^{(j,j')}(\theta_2))}{\partial \gamma_2^{(j,j')}} \frac{\partial \gamma_2^{(j,j')}(\theta_2)}{\partial \theta_2} \bigg|_{\theta_2 = \theta_{0,2}} \\ &= \sum_{j > j'} i_2^{(2)}(Y_j, Y_{j'}; \gamma_2^{(j,j')}(\theta_{0,2})) \end{aligned}$$

Based on (3) and (5) - (7), we have $\bar{\mathbf{F}}(\theta_0) = \mathbf{M}_{CL1}(\theta_0)$. The earlier results (1) - (4) then yield $\mathbf{G}_{CL1}(\theta_0) = \bar{\mathbf{F}}(\theta_0)$.

In (b), $\theta_{0,2}$ is on the boundary of the parameter space. As long as $\bar{\mathbf{F}}(\theta)$ and $\mathbf{G}_{CL1}(\theta)$ are continuous at $\theta_{0,2}$,

$$\lim_{\theta_2 \rightarrow \theta_{0,2}} \mathbf{G}(\theta_0) = \lim_{\theta_2 \rightarrow \theta_{0,2}} \bar{\mathbf{F}}(\theta_0). \quad \square$$

Weights for the Linear Case

Theorem 3.2 tells us that the CL1 estimates are asymptotically as efficient as the ML estimates when the data are independent within each family. The next question is how well this method performs when the data are correlated. Later investigations show that it can be inefficient compared with the MLE when the data are highly correlated. To improve the efficiency of θ_{CL1} , we consider adding weights in the estimating function. For the linear case, consider

$$\psi_{WCL1} = \sum_i \mathbf{X}_i' \mathbf{W}_i' \mathbf{g}_{CL1}(\mathbf{Y}_i, \gamma_i).$$

The idea can be generalized to nonlinear cases. For any fixed \mathbf{W}_i , $E(\psi_{WCL1}) = 0$. Therefore, ψ_{WCL1} are unbiased estimating functions which would provide a consistent estimate of θ . With \mathbf{W}_i in the estimating equations, we have

$$\begin{aligned} \mathbf{D}_\psi &= \mathbf{D}_{WCL1} = n^{-1} \sum_i \mathbf{X}_i' \mathbf{W}_i' \Delta_i \mathbf{W}_i \mathbf{X}_i \\ \mathbf{M}_\psi &= \mathbf{M}_{WCL1} = n^{-1} \sum_i \mathbf{X}_i' \mathbf{W}_i' \Omega_i \mathbf{W}_i \mathbf{X}_i. \end{aligned}$$

From Chaganty (1997), it follows that the optimal choice of \mathbf{W}_i is $\mathbf{W}_{opt,i} = \Omega_i^{-1} \Delta_i$. The corresponding covariance matrix is

$$\mathbf{V}_{Wopt} = \left(\sum_i \mathbf{X}_i^{*'} \Omega_i^{-1} \mathbf{X}_i^* \right)^{-1},$$

where $\mathbf{X}_i^* = \Delta_i \mathbf{X}_i$.

Since the optimal choice of weights, $\mathbf{W}_{opt,i}$, depends on the unknown parameters $\boldsymbol{\theta}$, we have to replace them with estimates. For this reason, the weighted estimates need to be implemented by an iterated method such as the one described below.

1. Estimate $\boldsymbol{\theta}$ without weights, and set $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$.
2. Calculate the weights $\mathbf{W}_{opt,i}(\tilde{\boldsymbol{\theta}})$.
3. Update $\hat{\boldsymbol{\theta}}$ using the weights $\mathbf{W}_{opt,i}(\tilde{\boldsymbol{\theta}})$.
4. Set $\tilde{\boldsymbol{\theta}}$ equal to the new $\hat{\boldsymbol{\theta}}$ and update $\mathbf{W}_{opt,i}(\tilde{\boldsymbol{\theta}})$.
5. Repeat 3-4 until convergence or for a few times.

Let $\tilde{\boldsymbol{\theta}}_{Wopt}$ be the estimate from the above iteration and $\hat{\boldsymbol{\theta}}_{Wopt}$ be the estimate when the $\mathbf{W}_{opt,i}$ are known. We conjecture that $\|\tilde{\boldsymbol{\theta}}_{Wopt} - \hat{\boldsymbol{\theta}}_{Wopt}\| \xrightarrow{P} 0$ and the asymptotic variance of $\tilde{\boldsymbol{\theta}}_{Wopt}$ converges to \mathbf{V}_{Wopt} in probability. However, we do not pursue a rigorous proof of the asymptotic properties of $\tilde{\boldsymbol{\theta}}_{Wopt}$ in this thesis for the reason given below. Theoretically, the weighted method is more efficient. However, the improvement is achieved at the price of increasing computation. Often, the evaluation of \mathbf{W}_i is too difficult or time consuming to carry out.

To reduce computation, we could consider weighting only \mathbf{g}_{UCL} . Let

$$\mathbf{W}_i = \begin{pmatrix} \mathbf{W}_{par,i} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$

where $\mathbf{W}_{par,i} = \Delta_i^{(1,1)} (\Omega_i^{(1,1)})^{-1}$, which is the optimal choice for \mathbf{g}_1 when $\boldsymbol{\theta}_2$ is known. Comparing with the optimal weights $\Omega_i^{-1} \Delta_i$, $\mathbf{W}_{par,i}$ is much easier to compute since it only involves the bivariate marginal distributions. The partial weights will mainly improve the efficiency of the estimate of $\boldsymbol{\theta}_1$ which, according to our investigation, suffers more efficiency loss (see details in the

next chapter). Under the partial weights $\mathbf{W}_{par,i}$,

$$\text{Var}(\hat{\boldsymbol{\theta}}_{1,W_{par}}) \approx \left(\sum_i \mathbf{X}_{i1}^* (\boldsymbol{\Omega}_i^{(1,1)})^{-1} \mathbf{X}_{i1}^* \right)^{-1}$$

where $\mathbf{X}_{i1}^* = \Delta_i^{(1,1)} \mathbf{X}_{i1}$. Since no weights modify \mathbf{g}_{2BCL} , $\text{Var}(\hat{\boldsymbol{\theta}}_2)$ remains the same as in (3.12).

Estimation with the partial weights can be implemented as follows:

1. Estimate $\boldsymbol{\theta}$ without weights, and set $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$.
2. Calculate the weights $W_{par,i}(\tilde{\boldsymbol{\theta}})$.
3. Update $\hat{\boldsymbol{\theta}}_1$ using the weights $\mathbf{W}_{par,i}(\tilde{\boldsymbol{\theta}})$.

3.2.3 Estimating Equations Based on BCL (CL2)

Inference Functions

Since both $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ can be identified by the bivariate marginal distributions, an alternative to the two-stage method is to estimate the two sets of parameters simultaneously by maximizing the BCL function Ψ_{BCL} defined in (3.5). However, this method could overemphasize large families since they contribute a quadratically increasing number of pairs to the BCL. It may be sensible to weight the contribution of each family according to its size. An individual from a family of size k appears in $k - 1$ pairs. Therefore we weight the contribution of the family by $1/(k - 1)$. With $w_i = 1/(k_i - 1)$ ($k_i > 1$), the weighted BCL function is:

$$\Psi_{BCL}^*(\boldsymbol{\theta}) = \sum_{i=1}^n w_i \sum_{j>j'} l_2(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}). \quad (3.16)$$

When Y_{i1}, \dots, Y_{ik_i} ($k_i > 1$) are independent,

$$w_i \sum_{j>j'} l_2(Y_{ij}, Y_{ij'}) = \frac{1}{k_i - 1} \sum_{j>j'} [l_1(Y_{ij}) + l_1(Y_{ij'})] = \sum_j l_1(Y_{ij}).$$

We can see that under independence Ψ_{BCL}^* is indeed the log likelihood function of \mathbf{Y}_i . Instead of using a family size-based weight, in theory we can derive the optimal weights. However, as mentioned before, such weights depend on the unknown $\boldsymbol{\theta}$ and are difficult to evaluate, so we do not pursue this.

The estimating function based on Ψ_{BCL}^* is

$$\psi_{CL2}(\boldsymbol{\theta}) = \frac{\partial \Psi_{BCL}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_i w_i [\mathbf{H}(\tilde{\mathbf{X}}_i)]' \mathbf{g}_{CL2}(\mathbf{Y}_i, \gamma_i), \quad (3.17)$$

where $\mathbf{H}(\tilde{\mathbf{X}}_i) = \partial \mathbf{h}(\tilde{\mathbf{X}}_i; \boldsymbol{\theta}, k_i) / \partial \boldsymbol{\theta}$ and

$$\mathbf{g}_{CL2}(\mathbf{Y}_i, \gamma_i) = \sum_{j>j'} \frac{\partial l_2(Y_{ij}, Y_{ij'})}{\partial \gamma_i}.$$

The function \mathbf{g}_{CL2} can be written as $(\mathbf{g}_{1BCL}, \mathbf{g}_{2BCL})'$, in terms of the partial derivatives with respect to γ_{i1} and γ_{i2} , respectively. For ψ_{CL2} ,

$$\mathbf{D}_\psi = \mathbf{D}_{CL2} = n^{-1} \sum_i w_i [\mathbf{H}(\tilde{\mathbf{X}}_i)]' \boldsymbol{\Delta}_i^* \mathbf{H}(\tilde{\mathbf{X}}_i)$$

with $\boldsymbol{\Delta}_i^* = -E\{\partial \mathbf{g}_{CL2}(\mathbf{Y}_i, \gamma_i) / \partial \gamma_i'\}$ and

$$\mathbf{M}_\psi = \mathbf{M}_{CL2} = n^{-1} \sum_i w_i^2 [\mathbf{H}(\tilde{\mathbf{X}}_i)]' \boldsymbol{\Omega}_i^* \mathbf{H}(\tilde{\mathbf{X}}_i)$$

with $\boldsymbol{\Omega}_i^* = \text{Var}(\mathbf{g}_{CL2}(\mathbf{Y}_i, \gamma_i))$.

Note that the weight $w_i = 1/(k_i - 1)$ can only be applied when $k_i > 1$. When $k_i = 1$, we set $w_i = 1$ and use the univariate log-likelihood, i.e.,

$$\Psi_{BCL}^*(\boldsymbol{\theta}) = \sum_{i=1}^n w_i l_{(i)}(\boldsymbol{\theta}), \quad (3.18)$$

where $l_{(i)}(\boldsymbol{\theta}) = l_1(Y_{i1}; \boldsymbol{\theta})$ when $k_i = 1$, and $l_{(i)}(\boldsymbol{\theta}) = \sum_{j>j'} l_2(Y_{ij}, Y_{ij'}; \boldsymbol{\theta})$ otherwise.

Performance of CL2 Estimates in Independence Case

Theorem 3.3 *Assume the same conditions in Theorem 3.1. Under independence, i.e., $f_k(\mathbf{Y}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) = \prod_i f_1(Y_i; \boldsymbol{\theta}_{0,1})$, we have*

$$\begin{aligned} \text{AVar}(\hat{\boldsymbol{\theta}}_{1,CL2}) &= \text{AVar}(\hat{\boldsymbol{\theta}}_{1,MLE}) \\ \text{AVar}(\hat{\boldsymbol{\theta}}_{2,CL2}) &\succeq_{pd} \text{AVar}(\hat{\boldsymbol{\theta}}_{2,MLE}), \end{aligned}$$

where AVar denotes asymptotic covariance matrix. The relations hold for the limits of the covariance matrices if $\boldsymbol{\theta}_{0,2}$ is on the boundary of the parameter space.

In our proof, we will need the following matrix version of the Cauchy-Schwarz inequality given by Chaganty (1997).

Lemma 3.1 *Let \mathbf{B}_i , \mathbf{D}_i , $1 \leq i \leq m$ be $t \times p$ matrices. Let Σ_i , $1 \leq i \leq k_i$ be $t \times t$ positive definite matrices. Then, assuming that the inverses exist,*

$$\left(\sum_i \mathbf{B}_i' \mathbf{D}_i \right)^{-1} \left(\sum_i \mathbf{B}_i' \Sigma_i \mathbf{B}_i \right) \left(\sum_i \mathbf{D}_i' \mathbf{B}_i \right)^{-1} \succeq_{pd} \left(\sum_i \mathbf{D}_i' \Sigma_i^{-1} \mathbf{D}_i \right)^{-1}.$$

The equality holds if $\mathbf{D}_i = \Sigma_i \mathbf{B}_i$.

Proof of Theorem 3.3: To avoid extra notation, we assume $k_i > 1$ for all i .

Based on the results in the proof of Theorem 3.2, $\mathbf{F}_{12}(\boldsymbol{\theta}_0) = \mathbf{0}$, so we have

$$\text{AVar}(\hat{\boldsymbol{\theta}}_{1,MLE}) = \frac{1}{n} \mathbf{F}_{11}^{-1}(\boldsymbol{\theta}_0), \quad (3.19)$$

$$\text{AVar}(\hat{\boldsymbol{\theta}}_{2,MLE}) = \frac{1}{n} \mathbf{F}_{22}^{-1}(\boldsymbol{\theta}_0). \quad (3.20)$$

We write

$$\mathbf{D}_{CL2} = \frac{1}{n} \begin{pmatrix} \mathbf{D}_{2,11} & \mathbf{D}_{2,12} \\ \mathbf{D}_{2,21} & \mathbf{D}_{2,22} \end{pmatrix},$$

where

$$\mathbf{D}_{2,ll'} = - \sum_i \frac{1}{k_i - 1} \sum_{j > j'} \mathbb{E} \ddot{l}_2^{(l,l')}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}), \quad l, l' = 1, 2.$$

Similarly, we write

$$\mathbf{M}_{CL2} = \frac{1}{n} \begin{pmatrix} \mathbf{M}_{2,11} & \mathbf{M}_{2,12} \\ \mathbf{M}_{2,21} & \mathbf{M}_{2,22} \end{pmatrix},$$

where

$$\mathbf{M}_{2,ll'} = \sum_i \frac{1}{(k_i - 1)^2} \mathbb{E} \left(\sum_{j > j'} \ddot{l}_2^{(l)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\sum_{j > j'} \ddot{l}_2^{(l')}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right)',$$

$l, l' = 1, 2$.

(1) Claim $\mathbf{D}_{2,11}(\boldsymbol{\theta}_0) = -\mathbf{F}_{11}(\boldsymbol{\theta}_0)$.

Since under independence, $\check{l}_2^{(1,1)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) = \check{l}_1^{(1,1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) + \check{l}_1^{(1,1)}(Y_{ij'}; \boldsymbol{\theta}_{0,1})$,

$$\begin{aligned} \mathbf{D}_{2,11}(\boldsymbol{\theta}_0) &= - \sum_i \frac{1}{k_i - 1} \sum_{j > j'} \left\{ \check{l}_1^{(1,1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) + \check{l}_1^{(1,1)}(Y_{ij'}; \boldsymbol{\theta}_{0,1}) \right\} \\ &= - \sum_i \sum_j \check{l}_1^{(1,1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \\ &= \mathbf{D}_{1,11}(\boldsymbol{\theta}_0) = \mathbf{F}_{11}(\boldsymbol{\theta}_0). \end{aligned}$$

The last step is from step (5) in the proof of Theorem 3.2.

(2) Claim $\mathbf{D}_{2,21}(\boldsymbol{\theta}_0) = \mathbf{0}$. The proof is similar to step (2) in the proof of Theorem 2. Then $\mathbf{D}_{2,12}(\boldsymbol{\theta}_0) = \mathbf{0}$ as well.

(3) Claim $\mathbf{M}_{2,11}(\boldsymbol{\theta}_0) = \mathbf{F}_{11}(\boldsymbol{\theta}_0)$.

Under independence,

$$\frac{1}{k_i - 1} \sum_{j > j'} \check{l}_2^{(1)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) = \sum_j \check{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}).$$

Therefore,

$$\mathbf{M}_{2,11}(\boldsymbol{\theta}_0) = \sum_i \mathbf{E} \left(\sum_j \check{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right) \left(\sum_j \check{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right)' = \mathbf{F}_{11}(\boldsymbol{\theta}_0).$$

(4) $\mathbf{M}_{2,21}(\boldsymbol{\theta}_0) = \mathbf{0}$.

Under independence,

$$\mathbf{M}_{2,21}(\boldsymbol{\theta}_0) = \sum_i \frac{1}{k_i - 1} \mathbf{E} \left(\sum_{j > j'} \check{l}_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}) \right) \left(\sum_j \check{l}_1^{(1)}(Y_{ij}; \boldsymbol{\theta}_{0,1}) \right)'.$$

Then $\mathbf{M}_{2,21}(\boldsymbol{\theta}_0) = \mathbf{0}$, by a similar proof as we gave for $\mathbf{M}_{1,21}(\boldsymbol{\theta}_0) = \mathbf{0}$ in Step (3) of Theorem 3.2.

Then $\mathbf{M}_{2,12}(\boldsymbol{\theta}_0) = \mathbf{0}$ also.

By (2) and (4), we have

$$\mathbf{D}_{CL2}^{-1}(\boldsymbol{\theta}_0) = \frac{1}{n} \begin{pmatrix} \mathbf{D}_{2,11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{2,22}^{-1} \end{pmatrix}$$

and

$$\mathbf{M}_{CL2}(\boldsymbol{\theta}_0) = \frac{1}{n} \begin{pmatrix} \mathbf{M}_{2,11} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{2,22} \end{pmatrix}.$$

Hence

$$\begin{aligned}\text{AVar}(\hat{\boldsymbol{\theta}}_{1,CL2}) &= \frac{1}{n} \mathbf{D}_{2,11}^{-1} \mathbf{M}_{2,11} \mathbf{D}_{2,11}^{-1}, \\ \text{AVar}(\hat{\boldsymbol{\theta}}_{2,CL2}) &= \frac{1}{n} \mathbf{D}_{2,22}^{-1} \mathbf{M}_{2,22} \mathbf{D}_{2,22}^{-1}.\end{aligned}$$

By (1) and (3),

$$\mathbf{D}_{2,11}^{-1} \mathbf{M}_{2,11} \mathbf{D}_{2,11}^{-1} = \mathbf{F}_{11}^{-1}(\boldsymbol{\theta}_0).$$

Therefore, under independence, $\text{AVar}(\hat{\boldsymbol{\theta}}_{1,CL2}) = \text{AVar}(\hat{\boldsymbol{\theta}}_{1,MLE})$.

By a similar proof as in Step (4) of Theorem 3.2, we can show that

$$\begin{aligned}\mathbf{D}_{2,22}(\boldsymbol{\theta}_0) &= \sum_i \frac{1}{k_i - 1} \sum_{j>j'} \mathbb{E}(i_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2})) (i_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}))', \\ \mathbf{M}_{2,22}(\boldsymbol{\theta}_0) &= \sum_i \frac{1}{(k_i - 1)^2} \sum_{j>j'} \mathbb{E}(i_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2})) (i_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}))' .\end{aligned}$$

Let

$$\mathbf{W}_i = \text{diag} \left(\frac{1}{k_i - 1}, \dots, \frac{1}{k_i - 1} \right)$$

and

$$\mathbf{B}_i = \mathbb{E}(i_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2})) (i_2^{(2)}(Y_{ij}, Y_{ij'}; \boldsymbol{\theta}_{0,1}, \boldsymbol{\theta}_{0,2}))'.$$

Then $\mathbf{D}_{2,22}(\boldsymbol{\theta}_0) = \sum_i \mathbf{W}_i \mathbf{B}_i$ and $\mathbf{M}_{2,22}(\boldsymbol{\theta}_0) = \sum_i \mathbf{W}_i \mathbf{B}_i \mathbf{W}_i$; hence

$$\text{AVar}(\hat{\boldsymbol{\theta}}_{2,BCL}) = \frac{1}{n} \left(\sum_i \mathbf{W}_i \mathbf{B}_i \right)^{-1} \left(\sum_i \mathbf{W}_i \mathbf{B}_i \mathbf{W}_i \right) \left(\sum_i \mathbf{W}_i \mathbf{B}_i \right)^{-1}.$$

By the matrix Cauchy-Schwarz inequality (Lemma 3.1,

$$\left(\sum_i \mathbf{W}_i \mathbf{B}_i \right)^{-1} \left(\sum_i \mathbf{W}_i \mathbf{B}_i \mathbf{W}_i \right) \left(\sum_i \mathbf{B}_i \mathbf{W}_i \right)^{-1} \succeq_{pd} \left(\sum_i \mathbf{B}_i \right)^{-1}.$$

Meanwhile, we have shown in step (7) of Theorem 3.2 that $\mathbf{F}_{22} = \sum_i \mathbf{B}_i$. Therefore

$$\text{AVar}(\hat{\boldsymbol{\theta}}_{2,BCL}) \succeq_{pd} \text{AVar}(\hat{\boldsymbol{\theta}}_{2,MLE}).$$

Equality holds when k_i is a constant. \square

The performance of the CL2 estimate in the dependent case is investigated in the next chapter.

3.3 Estimating the Asymptotic Covariance Matrix of $\hat{\theta}$

Different methods can be considered to estimate the asymptotic covariance matrix of $\hat{\theta}$. One approach is to evaluate the Godambe information matrix at $\hat{\theta}$ analytically. This can be time-consuming or even sometimes impossible. For example, with survival data, the matrix cannot be evaluated without specifying the censoring distribution. An alternative approach is to use resampling methods such as jackknife or bootstrapping. Naturally, the sampling units are the families. A third approach is to evaluate the Godambe information matrix empirically or using parametric resampling techniques based on the asymptotic distribution of the CL estimate.

We use the jackknife method in our examples in Chapter 6. Let $\tilde{\theta}_{-i}$ be the estimator of θ with \mathbf{Y}_i deleted, $i = 1, \dots, n$. The jackknife estimator of \mathbf{V} , as defined in (3.2), is

$$\mathbf{V}_J = \sum_i^n (\tilde{\theta}_{-i} - \hat{\theta})(\tilde{\theta}_{-i} - \hat{\theta})', \quad (3.21)$$

where $\hat{\theta}$ is the estimate based on all families. A proof of consistency of this estimator for the i.i.d. case and the case with covariates is given by Joe (1997) (Chapter 10). The theorem can be directly applied to familial data with fixed family structure. In general, family structures in the data are not identical, but we assume that the population under consideration is a finite mixture of types of family structures. Since the result holds for each type of family and the estimating function is a summation over all types of families, the consistency can be inferred from Joe's proof.

For large samples, the jackknife estimator can be obtained by deleting more than one family at a time in order to reduce the computation time.

3.4 Remarks

In this section, we provide some general remarks on the CL and related estimating equation methods.

Remark 1: For the multivariate probit model for binary data, a related estimating equation approach was proposed by Reboussin and Liang (1998). In their approach, a set of weighted estimating equations is constructed from the first and second moments. The estimating equations for θ_1 are the same as the weighted CL1 estimating equations, but the estimating equations for θ_2

are different. For multivariate probit models, under independence, the method of moments leads to the same estimating equations as the ML method for the regression parameters. However, when the estimating equations based on moments differ from the score equations, the method of moments is known to be less efficient than the ML method. We have shown that under independence, the CL1 method is equivalent to the ML method. So under independence, the CL1 method is better than the method of moments in general. Hence, Reboussin and Liang's method of moments idea cannot be expected to have high efficiency if it is extended to some other models. Also, this method cannot be directly applied to the Poisson-log normal mixture model, in which the regression parameters can not be estimated with only the first moment as the moment depends on σ^2 .

Remark 2: The CL1 and CL2 methods are not restricted to the models discussed in this thesis. In particular, the CL2 method works for multivariate models, whether or not the parameters are common to the different margins. It works even when the univariate margins do not belong to the same distribution family. The parameters can be estimated as long as the bivariate marginal likelihoods can be evaluated. Therefore, we can consider a more general form of the BCL function:

$$\Psi_{BCL}^*(\theta) = \sum_{i=1}^n w_i \sum_{j>j'} l_{i,jj'}(Y_{ij}, Y_{ij'}; \theta),$$

where $l_{i,jj'}$ is the log likelihood of the bivariate margin of Y_{ij} and $Y_{ij'}$. This estimation approach can be considered if the high-dimensional likelihood is too difficult or impossible to compute.

Remark 3: The idea can be extended to trivariate CL. However, as the dimension becomes higher, the computational demands increase, so the implementation of such an approach is more difficult. When all the parameters are either univariate or bivariate parameters, it is not clear that the efficiency will be improved by using the likelihoods of marginal distribution of higher order.

Remark 4: Typically, the CL methods can be implemented by numerically maximizing the CL functions. For example, the quasi-Newton method can be used to avoid solving the system derivative equations. Numerical optimization becomes more difficult as the total number of parameters increases since the search for the optimal point has to be carried out in a higher dimensional space. So the CL2 method is numerically more difficult to implement than the CL1 method. To improve the numerical optimization, we can use the CL1 estimate as the starting value for the CL2 method.

Remark 5: If the sample size is not large, the numerical optimization has to be done subject to

a constraint that the estimated dependence parameters must lie inside the parameter space of the multivariate joint distribution. Without this constraint, the solution of the estimating equations may be outside of the parameter space. For models based on the MVN distribution, the constraint on the dependence parameters is that the correlation matrix of each family is positive definite.

Remark 6: In the CL2 method, we weight the BCLs by the family size. Naturally, one would consider the same weights in the second stage of the CL1 method. However, those weights are inappropriate when the dependence is weak. As shown in Theorem 3.2, the unweighted CL1 method is equivalent to the ML method under independence. Hence, the weights will only worsen the efficiency when the dependence is weak. When the dependence is strong, the weights can be helpful. But as shown in the next chapter, the CL1 estimates of the dependence parameters perform well even under strong dependence. So weights are not essential. For these reasons, this weighting scheme is not used in the CL1 method.

Chapter 4

Efficiency Comparisons

The CL methods are not fully efficient. However, it is not known at what parameter values the CL estimates tend to lose more efficiency and how much efficiency is lost for familial data. In this chapter, we intend to partially answer these questions in the context of familial data analysis. We define the asymptotic relative efficiency (ARE) of the CL estimate of the parameter θ as $ARE(\hat{\theta}_{CL}) = AVar(\hat{\theta}_{MLE})/AVar(\hat{\theta}_{CL})$, where $AVar(\hat{\theta}_{MLE})$ and $AVar(\hat{\theta}_{CL})$ are derived from the inverted Fisher and Godambe information matrices respectively.

We plan to make comparisons for four types of data: continuous, binary, count and survival subject to right censoring. For each type of data we select one distribution as a representative: the MVN for continuous, the MVP for binary, the PLNM for count and the MVN with random right censoring for survival. For each model, we select some cases in which comparisons are possible.

For the MVN model, we are able to make theoretical comparisons of the ML method and the CL methods. For the MVP model, we randomly generate covariates and family size and compute the information matrices based on the generated families. For the other two models, the comparisons are done by simulations in which the likelihood is evaluated by numerical integration or approximation.

By these comparisons, we will gain some insight into how the performance is affected by different factors, such as degree of dependence, family size and data type.

4.1 Continuous Response: MVN model

First, we consider the multivariate normal model: $\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2\mathbf{R}_i(\boldsymbol{\alpha}))$, where $\boldsymbol{\beta}$ is an m -dimensional vector of unknown regression parameters; and $\mathbf{R}_i(\boldsymbol{\alpha}) = [\rho_{ijj'}(\boldsymbol{\alpha})]$ is a correlation matrix, with $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)'$, a vector of dependence parameters.

The likelihood of a MVN distribution has a closed form and therefore we are able to derive the theoretical asymptotic variances of both the ML estimates and the CL estimates, so that the efficiency can be compared analytically. For this reason, we conduct comparisons in various settings. The results from the MVN distribution can provide insight into the general behavior of the CL estimates, not limited to just the continuous response.

We first derive Fisher's information matrix and the asymptotic variances of the ML estimates and then investigate the AREs of the CL1 and CL2 estimates with conclusions of the comparisons at the end of each subsection.

4.1.1 MLE

Assuming \mathbf{R}_i^{-1} exists, the log-likelihood is:

$$l = \sum_i -\frac{k_i}{2} \log 2\pi - \frac{k_i}{2} \log \sigma^2 - \frac{1}{2} \log |\mathbf{R}_i| - \frac{1}{2\sigma^2} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})' \mathbf{R}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}),$$

The score functions are given by:

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2} \sum_i \mathbf{X}_i' \mathbf{R}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{\sum_i k_i}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})' \mathbf{R}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\ \frac{\partial l}{\partial \alpha_j} &= -\frac{1}{2} \text{tr}(\mathbf{R}_i^{-1} \boldsymbol{\Gamma}_i) + \frac{1}{2\sigma^2} \sum_i (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})' \mathbf{R}_i^{-1} \boldsymbol{\Gamma}_{ij} \mathbf{R}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}), \quad j = 1, \dots, q \end{aligned}$$

where $\boldsymbol{\Gamma}_{ij} = \frac{\partial}{\partial \alpha_j} \mathbf{R}_i$. These functions are derived based on the general results in Appendix B.1.

For $\theta' = (\beta', \sigma^2, \alpha')$, Fisher's information matrix is:

$$\mathbf{F} = \frac{1}{n} \begin{pmatrix} \mathbf{F}_{11} & 0 & 0 \\ 0 & \mathbf{F}_{22} & \mathbf{F}'_{32} \\ 0 & \mathbf{F}_{32} & \mathbf{F}_{33} \end{pmatrix},$$

where $\mathbf{F}_{11} = \sigma^{-2} \sum_i \mathbf{X}_i' \mathbf{R}_i^{-1} \mathbf{X}_i$; $\mathbf{F}_{22} = (2\sigma^2)^{-1} \sum_i k_i$; \mathbf{F}_{23} is a $q \times 1$ vector with the j th entry equal to $-(2\sigma^2)^{-1} \sum_i \text{tr} \Gamma_{ij} \mathbf{R}_i^{-1}$ and \mathbf{F}_{33} is a $q \times q$ matrix with the entry in the j th row and j' th column equal to $(2\sigma^2)^{-1} \sum_i \text{tr} \Gamma_{ij} \mathbf{R}_i^{-1} \Gamma_{ij'} \mathbf{R}_i^{-1}$.

As $\text{AVar}(\hat{\theta}_{MLE}) = n^{-1} \mathbf{F}^{-1}$, we obtain

$$\text{AVar}(\hat{\beta}_{MLE}) = \frac{1}{n} \mathbf{F}_{11}^{-1} = \sigma^2 (\sum_i \mathbf{X}_i' \mathbf{R}_i^{-1} \mathbf{X}_i)^{-1}, \quad (4.1)$$

$$\text{AVar} \begin{pmatrix} \hat{\sigma}_{MLE}^2 \\ \hat{\alpha}_{MLE} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{22} & \mathbf{F}'_{32} \\ \mathbf{F}_{32} & \mathbf{F}_{33} \end{pmatrix}^{-1}. \quad (4.2)$$

(With the MVN distribution, the variance of $\hat{\beta}$ is the exact variance. But to keep our notation simple, we do not distinguish Var and AVar in this section. Instead, we use AVar everywhere.)

4.1.2 Efficiency of the Two-Stage Estimating Approach (CL1)

In this section, we first derive the estimating equations and then compare the efficiency for different cases. Finally, we provide a summary of the comparison results.

Estimating Equations:

In Step 1, β and σ^2 are estimated by maximizing the sum of the univariate log likelihood of Y_{ij} :

$$\Psi_{UCL}(\mathbf{Y}; \beta, \sigma^2) = -\frac{1}{2} \sum_i \sum_j \left(\log 2\pi + \log \sigma^2 + \frac{(y_{ij} - \mathbf{X}_{ij}\beta)^2}{\sigma^2} \right).$$

The estimates of β and σ^2 are obtained by solving

$$\sum_i \mathbf{X}_i' (\mathbf{Y}_i - \mathbf{X}_i \beta)' = \mathbf{0} \quad (4.3)$$

$$\sigma^2 \sum_i k_i - \sum_i (\mathbf{Y}_i - \mathbf{X}_i \beta)' (\mathbf{Y}_i - \mathbf{X}_i \beta) = 0, \quad (4.4)$$

leading to the ordinary least squares estimators of β and σ^2 . If the weighted CL1 method is used, the estimating equations for β with the optimal weights are the same as in the full likelihood method:

$$\sum_i \mathbf{X}_i' \mathbf{R}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta)' = 0.$$

In Step 2, α is estimated by maximizing the sum of the log-likelihoods of the bivariate margins with β , σ^2 replaced by $\hat{\beta}_{CL1}$, and $\hat{\sigma}_{CL1}^2$ obtained from Step 1. The BCL function is

$$\begin{aligned} \Psi_{BCL}(\mathbf{Y}; \beta, \sigma^2, \alpha) &= - \sum_i \sum_{j \neq j'} \left\{ \log 2\pi + \log \sigma^2 + \frac{1}{2} \log(1 - \rho_{ijj'}^2) + \right. \\ &\quad \left. \frac{(Y_{ij} - \mathbf{X}_{ij}\beta)^2 - 2\rho_{ijj'}(Y_{ij} - \mathbf{X}_{ij}\beta)(Y_{ij'} - \mathbf{X}_{ij'}\beta) + (Y_{ij'} - \mathbf{X}_{ij'}\beta)^2}{2\sigma^2(1 - \rho_{ijj'}^2)} \right\} \\ &= c - \frac{1}{2} \log \sigma^2 \sum_i k_i(k_i - 1) - \frac{1}{2} \sum_i \sum_{j \neq j'} \log(1 - \rho_{ijj'}^2) - \\ &\quad \frac{1}{2\sigma^2} \sum_i (\mathbf{Y}_i - \mathbf{X}_i\beta)' \mathbf{A}_i (\mathbf{Y}_i - \mathbf{X}_i\beta) \end{aligned} \quad (4.5)$$

where c is a constant and $\mathbf{A}_i = (a_{ijj'})$ with

$$a_{ijj'} = \begin{cases} \sum_{l \neq j} (1 - \rho_{ijl}^2)^{-1} & \text{for } j = j', \\ -\rho_{ijj'} (1 - \rho_{ijj'}^2)^{-1} & \text{for } j \neq j'. \end{cases}$$

This leads to the following estimating equation for α_l :

$$2\hat{\sigma}_{CL1}^2 \sum_{i, j \neq j'} \frac{\rho_{ijj'}}{1 - \rho_{ijj'}^2} \frac{\partial \rho_{ijj'}}{\partial \alpha_l} - \sum_i (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{CL1})' \mathbf{B}_{il} (\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}_{CL1}) = 0, \quad l = 1, \dots, q, \quad (4.6)$$

where $\mathbf{B}_{il} = \partial \mathbf{A}_i / \partial \alpha_l$.

Efficiency

The investigation is divided into two parts: (1) efficiency of $\hat{\beta}_{CL1}$ and (2) efficiency of all estimates under different dependence structures.

1. Efficiency of $\hat{\beta}_{CL1}$

From (4.3) we have $\hat{\beta}_{CL1} = (\sum_i \mathbf{X}_i^T \mathbf{X}_i)^{-1} \sum_i \mathbf{X}_i^T \mathbf{Y}_i$. Therefore,

$$\text{AVar}(\hat{\beta}_{CL1}) = \sigma^2 \left(\sum_i \mathbf{X}_i^T \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i^T \mathbf{R}_i \mathbf{X}_i \right) \left(\sum_i \mathbf{X}_i^T \mathbf{X}_i \right)^{-1}. \quad (4.7)$$

The variance of $\hat{\beta}_{MLE}$ is given in (4.1).

We consider the case of only one covariate, denoted by \mathbf{x}_i . Then, $m = 2$, $\mathbf{X}_i = (\mathbf{1}, \mathbf{x}_i)$ and $\beta = (\beta_0, \beta_1)$. The results can be extended to more than one covariate. In the previous chapter, we treated the covariates as known constants. To simplify our analysis of the efficiency, we make the following assumptions about \mathbf{x}_i : The values of x_{i1}, \dots, x_{ik_i} are independent realizations of a random variable with mean 0 and variance 1. The assumptions about the mean and the variance can easily be satisfied by centering and standardization. Here we assume that the value of x_{ij} is not correlated within families. Cases where the x 's are related within family will be considered later. We will derive the asymptotic efficiency of $\hat{\beta}_{CL1}$ based on the following results: when the number of families $n \rightarrow \infty$,

Result 4.1 $\sum_{ij} x_{ij}/K \xrightarrow{P} 0$;

Result 4.2 $\sum_{ij} x_{ij}^2/K \xrightarrow{P} 1$;

Result 4.3 $K^{-1} \sum_i \mathbf{x}_i' \mathbf{R}_i^{-1} \mathbf{1} \xrightarrow{P} 0$;

Result 4.4 $K^{-1} \sum_i \mathbf{x}_i' \mathbf{R}_i \mathbf{x}_i - K^{-1} \sum_i \text{tr} \mathbf{R}_i \xrightarrow{P} 0$, and $K^{-1} \sum_i \mathbf{x}_i' \mathbf{R}_i^{-1} \mathbf{x}_i - K^{-1} \sum_i \text{tr} \mathbf{R}_i^{-1} \xrightarrow{P} 0$.

where $K = \sum_{i=1}^n k_i$, the total number of observations. These results are derived based on the assumption that M , the size of the largest family, is bounded. See Appendix B.2 for a proof of results 4.3 and 4.4.

It follows that when n is large

$$\frac{1}{K} \sum_i \mathbf{X}_i' \mathbf{X}_i \simeq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\frac{1}{K} \sum_i \mathbf{X}_i' \mathbf{R}_i \mathbf{X}_i \simeq \begin{pmatrix} K^{-1} \sum_i \mathbf{1}' \mathbf{R}_i \mathbf{1} & 0 \\ 0 & K^{-1} \sum_i \text{tr} \mathbf{R}_i \end{pmatrix} = \begin{pmatrix} K^{-1} \sum_i \mathbf{1}' \mathbf{R}_i \mathbf{1} & 0 \\ 0 & 1 \end{pmatrix},$$

since $\text{tr} \mathbf{R}_i = k_i$. Also,

$$\frac{1}{K} \sum_i \mathbf{X}_i' \mathbf{R}_i^{-1} \mathbf{X}_i \simeq \begin{pmatrix} K^{-1} \sum_i \mathbf{1}' \mathbf{R}_i^{-1} \mathbf{1} & 0 \\ 0 & K^{-1} \sum_i \text{tr} \mathbf{R}_i^{-1} \end{pmatrix}$$

follows from Result 4.4. Based on the above results, (4.7) and (4.1) become

$$\text{AVar}(\hat{\beta}_{CL1}) \simeq \sigma^2 \begin{pmatrix} K^{-2} \sum_i \mathbf{1}' \mathbf{R}_i \mathbf{1} & 0 \\ 0 & K^{-1} \end{pmatrix} \quad (4.8)$$

$$\text{AVar}(\hat{\beta}_{MLE}) \simeq \sigma^2 \begin{pmatrix} (\sum_i \mathbf{1}' \mathbf{R}_i^{-1} \mathbf{1})^{-1} & 0 \\ 0 & (\sum_i \text{tr} \mathbf{R}_i^{-1})^{-1} \end{pmatrix}. \quad (4.9)$$

The relative efficiency of $\hat{\beta}_{0,CL1}$ is approximately

$$\text{ARE}_{\hat{\beta}_{0,CL1}} = \frac{\text{AVar}(\hat{\beta}_{0,MLE})}{\text{AVar}(\hat{\beta}_{0,CL1})} = \frac{K^2}{(\sum_i \mathbf{1}' \mathbf{R}_i \mathbf{1})(\sum_i \mathbf{1}' \mathbf{R}_i^{-1} \mathbf{1})}, \quad (4.10)$$

whereas

$$\text{ARE}_{\hat{\beta}_{1,CL1}} = \frac{\text{AVar}(\hat{\beta}_{1,MLE})}{\text{AVar}(\hat{\beta}_{1,CL1})} = \frac{K}{\sum_i \text{tr} \mathbf{R}_i^{-1}}. \quad (4.11)$$

So with this MVN model, the ARE of $\hat{\beta}_{CL1}$ only depends on the correlation matrix \mathbf{R}_i and family sizes.

When $m > 2$, $\mathbf{X}_i = (\mathbf{1}, \mathbf{x}_{i1}, \dots, \mathbf{x}_{i(m-1)})$. If the covariates are realizations of $m - 1$ mutually independent random variables with mean 0 and variance 1, then $K^{-1} \sum_i \mathbf{x}'_{il} \mathbf{x}_{il'}$ and $K^{-1} \sum_i \mathbf{x}'_{il} \mathbf{R}_i \mathbf{x}_{il'}$ converge to 0 as $l \neq l'$ (see Appendix B.2 for a proof). Therefore, $\text{ARE}_{\hat{\beta}_{l,CL1}}$, $l = 1, \dots, m - 1$, is also given by (4.11).

Returning to the case of one covariate, we next investigate some cases in which the x values are correlated within families. The results 1 – 3 still hold when the x values are correlated within families. Therefore the asymptotic relative efficiency of $\hat{\beta}_{0,CL1}$ will not be affected. We only need to consider the asymptotic relative efficiency of $\hat{\beta}_{1,CL1}$.

The extreme case is that all family members share the same x value x_i . Suppose x_i is a realization of a random variable X_i with mean 0 and variance σ_x^2 . It can be shown that

$$\text{AVar}(\hat{\beta}_{1,CL1}) \simeq \frac{\sigma_x^2 \sum_i \mathbf{1}' \mathbf{R}_i \mathbf{1}}{\sigma_x^2 K^2}$$

and

$$\text{AVar}(\hat{\beta}_{1,MLE}) \simeq \frac{\sigma_x^2}{\sigma_x^2 \sum_i \mathbf{1}' \mathbf{R}_i^{-1} \mathbf{1}}.$$

As a result, $\text{ARE}_{\hat{\beta}_{1,CL1}}$ is the same as (4.10). For this case, since the information on β_1 is only from the difference between families, $\hat{\beta}_{1,CL1}$ is not affected much by the dependence between \mathbf{Y}_i .

More generally, suppose the components of \mathbf{x}_i are correlated with covariance matrix $\sigma_x^2 \mathbf{R}_{\mathbf{x}_i}$. Then we have

$$\begin{aligned} \frac{1}{K} \sum_i \mathbf{x}_i' \mathbf{R}_i \mathbf{x}_i - \sigma_x^2 \frac{\sum_i \text{tr}(\mathbf{R}_i \mathbf{R}_{\mathbf{x}_i})}{K} &\xrightarrow{P} 0 \\ \frac{1}{K} \sum_i \mathbf{x}_i' \mathbf{R}_i^{-1} \mathbf{x}_i - \sigma_x^2 \frac{\sum_i \text{tr}(\mathbf{R}_i^{-1} \mathbf{R}_{\mathbf{x}_i})}{K} &\xrightarrow{P} 0. \end{aligned}$$

This leads to

$$\begin{aligned} \text{AVar}(\hat{\beta}_{1,CL1}) &\simeq \frac{\sigma_x^2}{\sigma_x^2} \frac{\sum_i \text{tr}(\mathbf{R}_i \mathbf{R}_{\mathbf{x}_i})}{K^2} \\ \text{AVar}(\hat{\beta}_{1,MLE}) &\simeq \frac{\sigma_x^2}{\sigma_x^2} \frac{1}{\sum_i \text{tr}(\mathbf{R}_i^{-1} \mathbf{R}_{\mathbf{x}_i})}, \end{aligned}$$

so that the ratio $\text{ARE}_{\hat{\beta}_{1,CL1}}$ is given by

$$\frac{K^2}{\sum_i \text{tr}(\mathbf{R}_i \mathbf{R}_{\mathbf{x}_i}) \sum_i \text{tr}(\mathbf{R}_i^{-1} \mathbf{R}_{\mathbf{x}_i})}.$$

For the very special case that $\mathbf{R}_{\mathbf{x}_i} = \mathbf{R}_i$, we have

$$\text{ARE}_{\hat{\beta}_{1,CL1}} = \sum \frac{K}{\text{tr} \mathbf{R}_i^2} = \frac{1}{1 + \sum_i \sum_{j \neq j'} \rho_{ijj'}^2 / K},$$

and this efficiency decreases when $\rho_{ijj'}^2$ increases.

2. Efficiency of All Estimates under Different Dependence Structures

We consider the following dependence structures of \mathbf{Y}_i : exchangeable dependence structure, dependence structure of a Type-3 family; dependence structure of a Type-4 family. For each case, correlations are the same across families. We also consider families with either fixed or varying family size. These results can give us some idea of efficiency for data with more complicated family structure.

a. Exchangeable \mathbf{R}_i

If the largest family is of size M , then $-1/(M-1) \leq \rho \leq 1$. Since M can be much larger than 1, the lower limit of ρ is close to 0. For this reason, we consider $0 \leq \rho \leq 1$. The asymptotic relative efficiencies are:

$$\begin{aligned}
\text{ARE}_{\hat{\beta}_{0,CL1}} &= \frac{1}{(1+a\rho)(1-b\rho)} \\
\text{ARE}_{\hat{\beta}_{1,CL1}} &= \frac{1}{1+b\rho^2/(1-\rho)} \\
\text{ARE}_{\hat{\sigma}_{CL1}^2} &= \frac{1}{(1+a\rho^2)(1-d\rho^2)} \\
\text{ARE}_{\hat{\rho}_{CL1}} &= \frac{(1+\rho^2)^2}{(1-d\rho)ce},
\end{aligned}$$

where

$$\begin{aligned}
a &= \frac{1}{K} \sum k_i(k_i - 1) \\
b &= \frac{1}{K} \sum \frac{k_i(k_i - 1)}{1 + (k_i - 1)\rho} \\
c &= \frac{1}{K} \sum \frac{k_i(k_i - 1)(1 + (k_i - 1)\rho^2)}{(1 + (k_i - 1)\rho)^2} \\
d &= b^2/c \\
e &= \rho^2(1 + \rho)^2(1 + a\rho^2) + \frac{2\rho^3(1 - \rho^2) \sum k_i(k_i - 2)(k_i - 1)}{\sum k_i(k_i - 1)} + \frac{\sum f_i}{(\sum k_i(k_i - 1))^2} \\
f_i &= k_i(k_i - 1)((1 - (k_i - 3)\rho^2)(1 + (k_i - 1)\rho)^2 + k_i(k_i - 2)^2\rho^4).
\end{aligned}$$

$\text{ARE}_{\hat{\beta}_{0,CL1}}$ and $\text{ARE}_{\hat{\beta}_{1,CL1}}$ are obtained directly from (4.10) and (4.11). $\text{Var}(\hat{\sigma}_{CL1}^2)$ and $\text{Var}(\hat{\rho}_{CL1})$ are derived using Maple (see Appendix B.3 for the Maple code).

Some remarks concerning the dependence of these AREs on family size and the dependence parameter ρ are given below:

1. Family size:

(a) When $k_i = k$ for all i , $\text{ARE}_{\hat{\beta}_{0,CL1}} = \text{ARE}_{\hat{\sigma}_{CL1}^2} = \text{ARE}_{\hat{\rho}_{CL1}} = 1$, while

$$\text{ARE}_{\hat{\beta}_{1,CL1}} = 1 - \frac{(k-1)\rho^2}{1 + (k-1)\rho}.$$

(b) The term $\frac{k(k-1)}{1 + (k-1)\rho}$ is a strictly convex function of k since

$$\frac{d^2}{dk^2} \left[\frac{k(k-1)}{1 + (k-1)\rho} \right] = \frac{2(1-\rho)}{[1 + (k-1)\rho]^3} \geq 0.$$

By Jensen's Inequality,

$$b \geq \frac{\bar{k} - 1}{1 + (\bar{k} - 1)\rho},$$

where $\bar{k} = K/n$. This implies that

$$\text{ARE}_{\hat{\beta}_{1,CL1}} \leq 1 - \frac{(\bar{k} - 1)\rho^2}{1 + (\bar{k} - 1)\rho}.$$

Equality holds if and only if $k_i = \bar{k}$ for all i . This implies that: for given K and n , the efficiency is always lower when family sizes vary, and further that, roughly speaking, $\text{ARE}_{\hat{\beta}_{1,CL1}}$ decreases when \bar{k} increases.

- (c) Roughly speaking, $\text{ARE}_{\sigma_{CL1}^2}$ and $\text{ARE}_{\hat{\beta}_{0,CL1}}$ increase with \bar{k} and decrease with the variation in k_i .

2. Dependence parameter ρ :

- (a) $\hat{\beta}_{1,CL1}$ loses efficiency when $|\rho|$ increases. As ρ approaches 1, $\text{ARE}_{\hat{\beta}_{1,CL1}}$ goes to 0. This is because when $\rho = 1$, $\hat{\beta}_{1,MLE}$ has no estimation error, whereas $\hat{\beta}_{1,CL1}$ has estimation error.
- (b) When $\rho = 1$, $\text{ARE}_{\hat{\beta}_{1,CL1}} = 0$ and

$$\text{ARE}_{\hat{\beta}_{0,CL1}} = \text{ARE}_{\hat{\sigma}_{CL1}^2} = \frac{K^2}{n \sum_i k_i^2}.$$

Clearly $\text{ARE}_{\hat{\beta}_{0,CL1}}$ and $\text{ARE}_{\hat{\sigma}_{CL1}^2}$ are less than one unless k_i is constant.

- (c) $\text{ARE}_{\hat{\sigma}_{CL1}^2}$ is a function of ρ^2 . Comparing to $\text{ARE}_{\hat{\beta}_{0,CL1}}$, it decreases slower when ρ increases.

Figure 4.1 shows the four ratios with the maximum family size M varying from 2 to 10 and mN_m/K fixed at $1/M$, where N_m is the number of families with m members, $m = 1, \dots, M$. (i.e. there are the same total number of individuals from each size of family.)

b. Type-3 family with varying number of offspring

Let ρ_1 be the parent-offspring correlation and ρ_2 be the sib-sib correlation. Assuming that the index of the parent is 1, then

$$\mathbf{R}_i = \begin{pmatrix} 1 & \rho_1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_2 \\ \rho_1 & \rho_2 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1 & \rho_2 & \rho_2 & \cdots & 1 \end{pmatrix}.$$

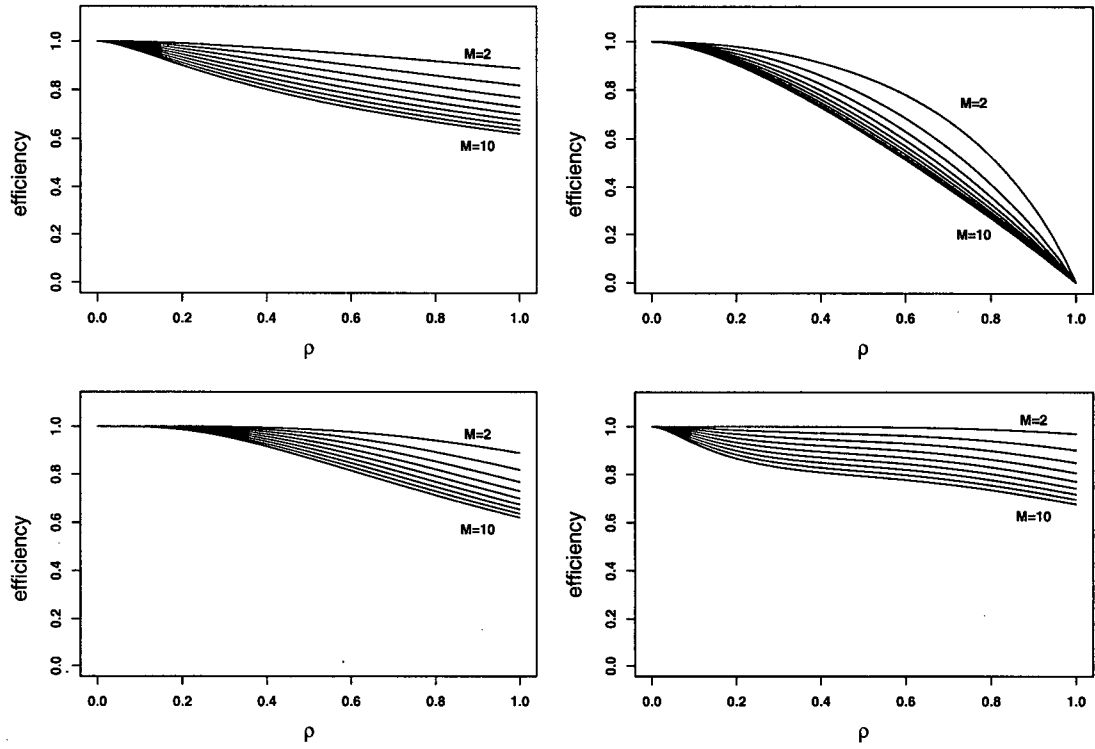


Figure 4.1: Efficiency of CL1 for exchangeable dependence structure. (MVN model)

The correlations ρ_1 and ρ_2 need to satisfy the following condition: $M\rho_1^2 - (M-1)\rho_2 < 1$, where M is the maximum number of offspring in a family (Srivastava and Katapa, 1986).

For a fixed number of offspring k^* in all families, the relative efficiency was calculated analytically using Maple. The Maple code is attached in Appendix B.3. Since there is no simple expression of the efficiency, we show the results graphically. Figure 4.2 shows the case when $k^* = 3$. When k^* increases, the efficiency generally decreases.

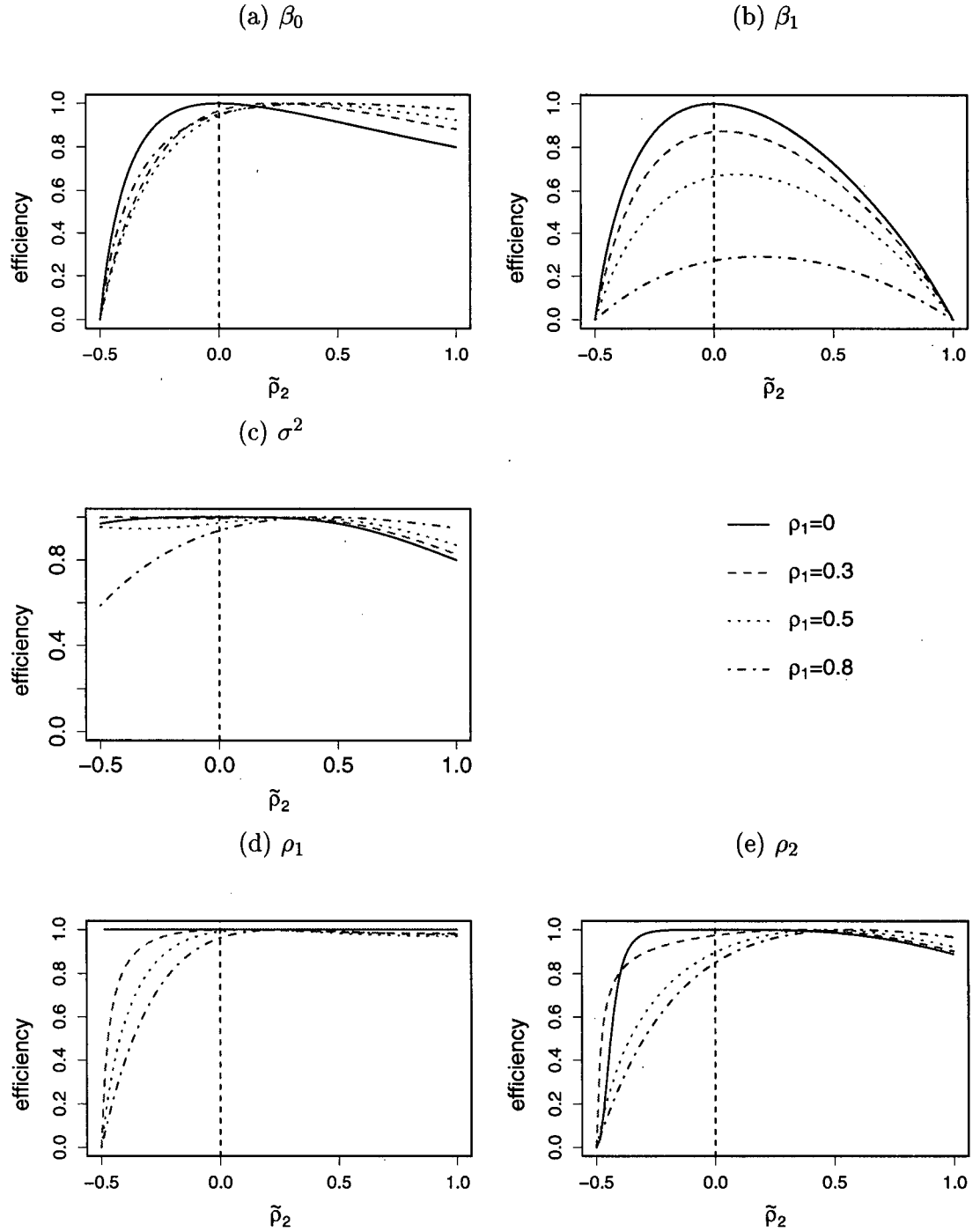


Figure 4.2: Efficiency of CL1 for Type-3 families with $k^* = 3$ offspring (MVN model); $\tilde{\rho}_2 = (\rho_2 - \rho_1^2)/(1 - \rho_1^2)$ is the sib-sib correlation conditional on their parent.

In Figure 4.2, the efficiency is plot at different levels of ρ_1 (0, 0.3, 0.5 and 0.8). To make the curves more comparable, the efficiency was plotted against the conditional sib-sib correlation,

$\tilde{\rho}_2 = (\rho_2 - \rho_1^2)/(1 - \rho_1^2)$. Some remarks follow:

1. The efficiency of $\hat{\beta}_{1,CL1}$ approaches zero when ρ_2 approaches to its lower and upper bounds. Given ρ_1 , $\hat{\beta}_{1,CL1}$ is most efficient when the conditional sib-sib correlation is close to 0.
2. The estimates of the other parameters perform well when the conditional dependence between siblings is non-negative. Except for $\hat{\sigma}_{CL1}^2$, the efficiency decreases rapidly as the conditional dependence approaches its lower bound. However, in analysis of familial data, we expect the conditional sib-sib correlation to be non-negative.
3. When the parent-offspring correlation ρ_1 is 0, the CL1 estimate of ρ_1 is as efficient as the MLE regardless of the strength of the dependence among the siblings.

When there is a varying number of offspring in each family, the efficiency is affected by both the mean and relative dispersion of family size. We randomly generated families with numbers of offspring varying from 1 to 7. We generated 500 sets of families. Each set has a different combination of family sizes. We calculated the efficiency of the parameter estimates for each combination at $\rho_1 = 0.4$ and $\rho_2 = 0.8$, then plotted the efficiency against the mean of k_i^* , denoted by \bar{k}^* , and the relative dispersion, measured by the variance-mean ratio of k_i^* 's, V/M . We deliberately chose ρ_2 far away from ρ_1 since the effect of the relative dispersion of family sizes is more obvious under strong conditional sib-sib dependence. The following patterns were observed:

1. Since $\text{tr } \mathbf{R}_i^{-1}$ is almost linear with \bar{k}^* , $\text{ARE}_{\hat{\beta}_{1,CL1}}$ mainly depends on \bar{k}^* . Figure 4.3 shows that the efficiency is monotonely decreasing in \bar{k}^* .
2. In Figure 4.4, we plot the efficiency of the other four parameters. The x-axis is \bar{k}^* and the y-axis is the variance-mean ratio of the k_i^* 's. For $\hat{\beta}_{0,CL1}$, $\hat{\sigma}_{CL1}^2$ and $\hat{\rho}_{2,CL1}$, the efficiency is almost linearly decreasing with both \bar{k}^* and V/M . On the other hand, $\text{ARE}_{\hat{\rho}_{1,CL1}}$ is mainly affected by V/M .

For different values of ρ_1 and ρ_2 , similar patterns were observed.

c. Type-4 family

In a Type-4 family, there are two parents with their common offspring. Assume that the correlation between father and child is the same as that between mother and child. Let ρ_1 and ρ_2

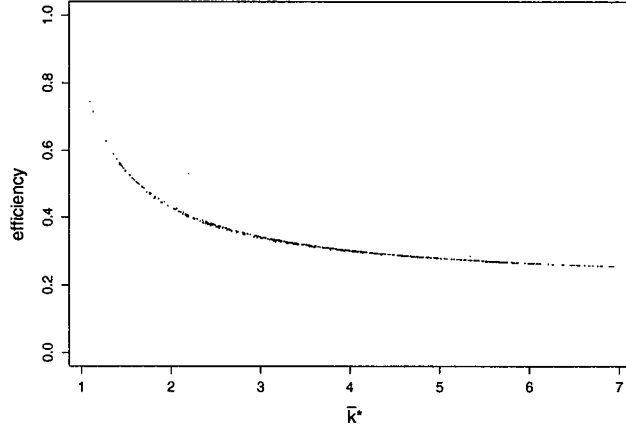


Figure 4.3: ARE of $\hat{\beta}_{1,CL1}$ against average number of offspring, \bar{k}^* (MVN model).

be the parent-offspring and sib-sib correlations, respectively, and ρ_3 be the correlation between the parents. The three dependence parameters must satisfy the following conditions:

$$\frac{(\rho_3 + 1)[(M - 1)\rho_2 + 1]}{2M} - \rho_1^2 > 0,$$

where M is the maximum number of offspring in a family. This can be proved with the same techniques used by Srivastava and Katapa (1986).

Case **c** is actually an extension of case **b**. The effect of ρ_3 is symmetric to that of ρ_2 . We first examine the boundary cases:

1. When the dependence parameters approach the boundary:

$$\frac{(\rho_3 + 1)[(M - 1)\rho_2 + 1]}{2M} = \rho_1^2,$$

$\text{ARE}_{\hat{\beta}_{0,CL1}}$ and $\text{ARE}_{\hat{\beta}_{1,CL1}}$ approach 0.

2. When either ρ_2 or ρ_3 approaches 1, $\text{ARE}_{\hat{\beta}_{1,CL1}}$ approaches 0.

Table 4.1 presents some results for constant family sizes. Except for β_1 , the efficiency of the estimates of other parameters is generally high.

Discussion

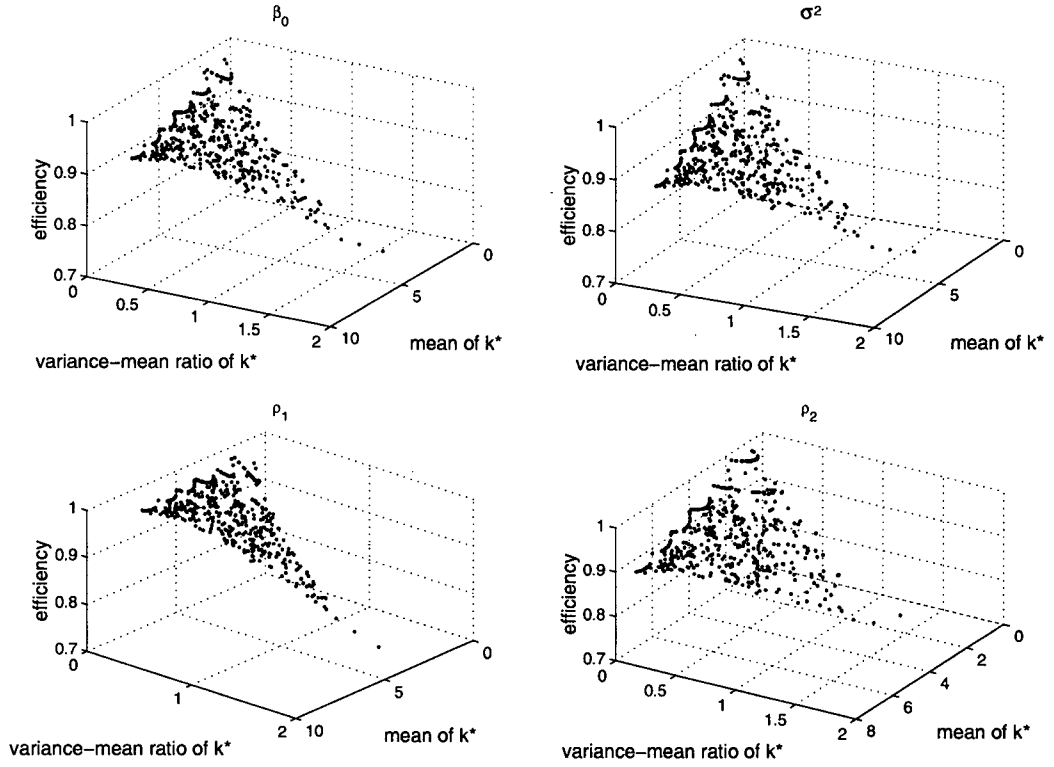


Figure 4.4: Efficiency of $\hat{\beta}_{0,CL1}$, $\hat{\sigma}_{CL1}^2$, $\hat{\rho}_{1,CL1}$ and $\hat{\rho}_{2,CL1}$ (MVN model).

The performance of the CL1 estimators is affected by the degree of dependence among the family members. When the correlation parameters approach the boundaries, \mathbf{R}_i is close to singular and the efficiency of $\hat{\beta}_{1,CL1}$ goes to 0. The CL1 estimators of the other parameters are less affected by the correlation parameters and perform fairly well. Their efficiency can reach 0 at some particular parameter values, such as in case b, when the conditional sib-sib correlation approaches its lower boundary.

The efficiency is also affected by family size. In general, the efficiency tends to decrease as the average family size increases and the relative dispersion of family size increases. When the relative dispersion of family sizes is large, the data are mainly formed by either small or large families. Individuals or pairs contributed by different families are equally weighted in the CL1 method. That is why large relative dispersion of family size combined with high dependence has a strong effect on their efficiency.

	$\rho_1 = 0.2$ $\rho_2 = 0.4$ $\rho_3 = 0.3$	$\rho_1 = 0.2$ $\rho_2 = 0.7$ $\rho_3 = 0.5$	$\rho_1 = 0.5$ $\rho_2 = 0.7$ $\rho_3 = 0.5$	number of offspring
β_0	.987	.921	.982	$k^* = 3$
β_1	.777	.475	.475	
σ^2	.993	.936	.974	
ρ_1	.989	.904	.972	
ρ_2	.999	.985	.995	
ρ_3	.991	.914	.987	
β_0	.978	.893	.961	$k^* = 5$
β_1	.727	.412	.412	
σ^2	.987	.898	.942	
ρ_1	.987	.888	.960	
ρ_2	.997	.947	.969	
ρ_3	.980	.865	.966	

Table 4.1: AREs of CL1 for Type-3 families with equal number of offspring (MVN model)

4.1.3 Estimates Based on Bivariate Composite Likelihood (CL2)

Estimating Equations

In the CL2 estimating approach, β , σ^2 and α are estimated simultaneously by maximizing the product of the likelihoods of the bivariate margins. First, we will show why it is appropriate to adjust the BCL by family size. Suppose $k_i > 1$ for each i . We estimate all the parameters by maximizing the BCL function defined in (4.5). Differentiating and setting to zero lead to the following set of equations:

$$\psi_1 = \frac{\partial l_2}{\partial \beta} \propto \sum_i \mathbf{X}_i' \mathbf{A}_i (\mathbf{Y}_i - \mathbf{X}_i \beta) = 0 \quad (4.12)$$

$$\psi_2 = \frac{\partial l_2}{\partial \sigma^2} \propto \sigma^2 \sum_i k_i (k_i - 1) - \sum_i (\mathbf{Y}_i - \mathbf{X}_i \beta)' \mathbf{A}_i (\mathbf{Y}_i - \mathbf{X}_i \beta) = 0 \quad (4.13)$$

$$\psi_3 = \frac{\partial l_2}{\partial \alpha} \propto 2\sigma^2 \sum_{i,j \neq j'} \frac{\rho_{ijj'}}{1 - \rho_{ijj'}^2} \frac{\partial \rho_{ijj'}}{\partial \alpha} - \sum_i (\mathbf{Y}_i - \mathbf{X}_i \beta)' \frac{\partial \mathbf{A}_i}{\partial \alpha} (\mathbf{Y}_i - \mathbf{X}_i \beta) = 0 \quad (4.14)$$

Let $\hat{\beta}_{BCL}$ be the estimator based on the above estimating equations. From (4.12), we have $\hat{\beta}_{BCL} = (\sum_i \mathbf{X}_i' \mathbf{A}_i \mathbf{X}_i)^{-1} \sum_i \mathbf{X}_i' \mathbf{A}_i \mathbf{Y}_i$. Therefore,

$$\text{AVar}(\hat{\beta}_{BCL}) = \sigma^2 \left(\sum_i \mathbf{X}_i' \mathbf{A}_i \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{A}_i \mathbf{R}_i \mathbf{A}_i \mathbf{X}_i \right) \left(\sum_i \mathbf{X}_i' \mathbf{A}_i \mathbf{X}_i \right)^{-1} \quad (4.15)$$

We consider the case of one covariate and make the same assumptions for the covariate as on page 72: the values of x_{i1}, \dots, x_{ik_i} are independent realizations of a random variable with mean 0 and variance 1. Asymptotically,

$$\text{AVar}(\hat{\beta}_{0,BCL}) = \frac{\sum_i \mathbf{1}' \mathbf{B}_i \mathbf{1}}{(\sum_i \mathbf{1}' \mathbf{A}_i \mathbf{1})^2}$$

$$\text{AVar}(\hat{\beta}_{1,BCL}) = \frac{\sum_i \text{tr} \mathbf{B}_i}{\sigma_x^2 (\text{tr} \mathbf{A}_i)^2},$$

where $\mathbf{B}_i = \sigma^2 \mathbf{A}_i \mathbf{R}_i \mathbf{A}_i$.

The efficiency relative to the MLE is:

$$\text{ARE}_{\hat{\beta}_{0,BCL}} = \frac{K(\sum_i \mathbf{1}' \mathbf{A}_i \mathbf{1})^2}{(\sum_i \mathbf{1}' \mathbf{B}_i \mathbf{1})(\sum_i \mathbf{1}' \mathbf{R}_i^{-1} \mathbf{1})}$$

$$\text{ARE}_{\hat{\beta}_{1,BCL}} = \frac{(\text{tr} \mathbf{A}_i)^2}{(\sum_i \text{tr} \mathbf{B}_i)(\sum_i \text{tr} \mathbf{R}_i^{-1})}$$

Using the exchangeable case as an example, in Figure 4.5, we plot the efficiency of both the CL1 and BCL estimators relative to the MLE for the special case of a mixture of two family sizes: size two and six with proportion 3:1. For β_0 , the efficiency of the BCL estimator is much lower than the CL1 estimator when the association is high. For β_1 , the BCL estimator is better when the correlation is strong, whereas the CL1 estimator is better when the correlation is weak. It is

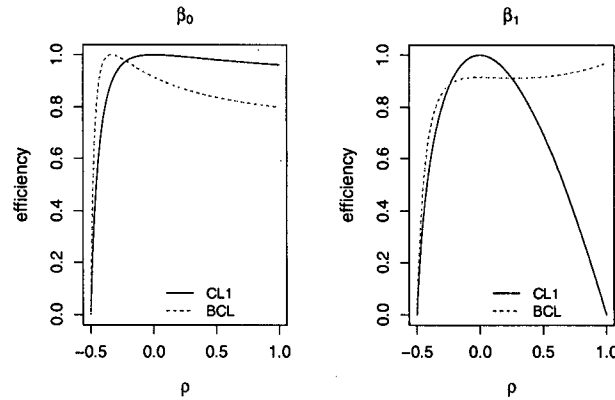


Figure 4.5: Efficiency of CL1 and BCL for the exchangeable case. Families are of size two and six with proportion 3:1 (MVN model).

not surprising that the estimates based on BCL are inefficient when the correlation ρ is close to

zero, since (4.12) can be seen as the CL1 estimating equation (4.3) weighted by \mathbf{A}_i . When $\rho = 0$, $\mathbf{A}_i = (k_i - 1)\mathbf{I}$. By Theorem 3.2, the CL1 is as efficient as the MLE when $\rho = 0$. Hence it is more efficient than a method with unequal weights. This justifies the previous consideration in Section 3.2.3 of weighting the BCL function by family size. In the CL2 method, the weight of family i is given by $w_i = 1$ for $k_i = 1$ and $w_i = 1/(k_i - 1)$ if $k_i > 1$. Let $l_{(i)}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha})$ be the contribution to the CL from the i th family before weighting. After weighting by w_i , the BCL function becomes:

$$\begin{aligned}\Psi_{BCL}^*(\mathbf{Y}; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}) &= \sum_i w_i l_{(i)}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}) \\ &= c^* - \frac{1}{2} \log \sigma^2 \sum_i k_i - \frac{1}{2} \sum_i w_i \sum_{j \neq j'} \log(1 - \rho_{ijj'}^2) - \\ &\quad \frac{1}{2\sigma^2} \sum_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{A}_i^* (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}),\end{aligned}$$

where c^* is a constant and

$$\mathbf{A}_i^* = \begin{cases} 1 & \text{when } k_i = 1 \\ \mathbf{A}_i / (k_i - 1) & \text{when } k_i > 1 \end{cases}.$$

After taking partial derivatives of l_2^* and rescaling, we obtain estimating equations of the CL2 method:

$$\psi_1^* = \sum_i \mathbf{X}_i' \mathbf{A}_i^* (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) = 0 \quad (4.16)$$

$$\psi_2^* = \sigma^2 \sum_i k_i - \sum_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{A}_i^* (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) = 0 \quad (4.17)$$

$$\psi_3^* = 2\sigma^2 \sum_i \frac{1}{k_i - 1} \sum_{j \neq j'} \frac{\rho_{ijj'}}{1 - \rho_{ijj'}^2} \frac{\partial \rho_{ijj'}}{\partial \boldsymbol{\alpha}} - \sum_i (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})' \frac{\partial \mathbf{A}_i^*}{\partial \boldsymbol{\alpha}} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) = 0 \quad (4.18)$$

The estimate of $\boldsymbol{\beta}$ generated from the CL2 method is

$$\hat{\boldsymbol{\beta}}_{CL2} = \left(\sum_i \mathbf{X}_i' \mathbf{A}_i^* \mathbf{X}_i \right)^{-1} \sum_i \mathbf{X}_i' \mathbf{A}_i^* \mathbf{Y}_i.$$

The variance of $\hat{\boldsymbol{\beta}}_{CL2}$ is

$$\text{Var}(\hat{\boldsymbol{\beta}}_{CL2}) = \sigma^2 \left(\sum_i \mathbf{X}_i' \mathbf{A}_i^* \mathbf{X}_i \right)^{-1} \left(\sum_i \mathbf{X}_i' \mathbf{A}_i^* \mathbf{R}_i \mathbf{A}_i^* \mathbf{X}_i \right) \left(\sum_i \mathbf{X}_i' \mathbf{A}_i^* \mathbf{X}_i \right)^{-1}.$$

For the example illustrated in Figure 4.5, $\hat{\boldsymbol{\beta}}_{0,CL2}$ is as efficient as $\hat{\boldsymbol{\beta}}_{0,CL1}$ and $\hat{\boldsymbol{\beta}}_{1,CL2}$ is generally more efficient than $\hat{\boldsymbol{\beta}}_{1,CL1}$. This will be shown in the next section.

Efficiency

First, we show that the CL2 method provides a better estimate of β_1 than the CL1 method.

Theorem 4.1 *Consider the case of one covariate. Under the assumption that the values of x_{i1}, \dots, x_{ik_i} are independent realizations of a random variable with mean 0 and variance 1, we have*

$$AVar(\hat{\beta}_{1,CL2}) \leq AVar(\hat{\beta}_{1,CL1}).$$

Proof: From (4.8), we have $AVar(\hat{\beta}_{1,CL1}) \simeq 1/K$. By a similar deduction, we also have

$$AVar(\hat{\beta}_{1,CL2}) \simeq \frac{\sum \text{tr } \mathbf{B}_i^*}{(\sum \text{tr } \mathbf{A}_i^*)^2},$$

where $\mathbf{B}_i^* = \mathbf{A}_i^* \mathbf{R}_i \mathbf{A}_i^*$.

(1) We first show that $\sum \text{tr } \mathbf{A}_i^* \geq K$. It is trivial for $k_i = 1$. For $k_i > 1$,

$$\text{tr } \mathbf{A}_i^* = \frac{1}{k_i - 1} \sum_{j=1}^{k_i} \sum_{l \neq j} \frac{1}{1 - \rho_{ijl}^2} \geq \frac{1}{k_i - 1} \sum_{j=1}^{k_i} (k_i - 1) = k_i.$$

(2) Next we claim that $\text{tr } \mathbf{B}_i^* \leq \text{tr } \mathbf{A}_i^*$. It is true when $k_i = 1$. For $k_i > 1$, we have

$$\text{tr } \mathbf{B}_i^* = \frac{1}{k_i - 1} \sum_{j=1}^{k_i} a_{ijj}.$$

and

$$\text{tr } \mathbf{B}_i^* = \frac{1}{(k_i - 1)^2} \sum_j \mathbf{a}_{ij}' \mathbf{R}_i \mathbf{a}_{ij},$$

where \mathbf{a}_{ij} is the j th column of \mathbf{A}_i . To prove the claim, it suffices to show

$$\mathbf{a}_{ij}' \mathbf{R}_i \mathbf{a}_{ij} \leq a_{ijj}(k_i - 1).$$

By symmetry, we only need to show that this is true when $j = 1$.

$$\mathbf{a}_{i1}' \mathbf{R}_i \mathbf{a}_{i1} = a_{i11}^2 + 2a_{i11} \sum_{j=2}^{k_i} a_{i1j} \rho_{i1j} + \sum_{j=2}^{k_i} a_{i1j}^2 + \sum_{j \neq j' \geq 2} a_{i1j} a_{i1j'} \rho_{ijj'}.$$

Since

$$a_{i11} + \sum_{j=2}^{k_i} a_{i1j} \rho_{i1j} = \sum_{j=2}^{k_i} \frac{1}{1 - \rho_{i1j}^2} - \sum_{j=2}^{k_i} \frac{\rho_{i1j}^2}{1 - \rho_{i1j}^2} = k_i - 1,$$

$$\begin{aligned}
\mathbf{a}'_{i1} \mathbf{R}_i \mathbf{a}_{i1} &= (k_i - 1)a_{i11} + a_{i11} \sum_{j=2}^{k_i} a_{i1j} \rho_{i1j} + \sum_{j=2}^{k_i} a_{i1j}^2 + \sum_{j \neq j' \geq 2} a_{i1j} a_{i1j'} \rho_{ijj'} \\
&= (k_i - 1)a_{i11} - \left(\sum_{j=2}^{k_i} \frac{1}{1 - \rho_{i1j}^2} \right) \left(\sum_{j=2}^{k_i} \frac{\rho_{i1j}^2}{1 - \rho_{i1j}^2} \right) + \sum_{j=2}^{k_i} \frac{\rho_{i1j}^2}{(1 - \rho_{i1j}^2)^2} \\
&\quad + \sum_{j \neq j' \geq 2} \frac{\rho_{i1j} \rho_{i1j'} \rho_{ijj'}}{(1 - \rho_{i1j}^2)(1 - \rho_{i1j'}^2)} \\
&= (k_i - 1)a_{i11} - \sum_{j \neq j' \geq 2} \frac{\rho_{i1j}^2}{(1 - \rho_{i1j}^2)(1 - \rho_{i1j'}^2)} + \sum_{j \neq j' \geq 2} \frac{\rho_{i1j} \rho_{i1j'} \rho_{ijj'}}{(1 - \rho_{i1j}^2)(1 - \rho_{i1j'}^2)} \\
&= (k_i - 1)a_{i11} - \sum_{j \geq j' \geq 2} \frac{\rho_{i1j}^2 + \rho_{i1j'}^2 - 2\rho_{i1j} \rho_{i1j'} \rho_{ijj'}}{(1 - \rho_{i1j}^2)(1 - \rho_{i1j'}^2)}.
\end{aligned}$$

Since $\rho_{i1j}^2 + \rho_{i1j'}^2 \geq 2|\rho_{i1j} \rho_{i1j'}| \geq 2|\rho_{i1j} \rho_{i1j'} \rho_{ijj'}|$, $\rho_{i1j}^2 + \rho_{i1j'}^2 - 2\rho_{i1j} \rho_{i1j'} \rho_{ijj'} \geq 0$. Therefore, $\mathbf{a}'_{i1} \mathbf{R}_i \mathbf{a}_{i1} \leq (k_i - 1)a_{i11}$.

From (1) and (2) we easily obtain that $\text{AVar}(\hat{\beta}_{1,CL2}) \leq \text{AVar}(\hat{\beta}_{1,CL1})$. The equality holds when $\mathbf{R}_i = \mathbf{I}$ for all i . \square

Next we give some results for the exchangeable case and for Type-3 families.

a. Exchangeable ($k_i \geq 2$)

1. $\text{ARE}_{\hat{\beta}_{0,CL2}} = \text{ARE}_{\hat{\beta}_{0,CL1}}$ and $\text{ARE}_{\hat{\sigma}_{CL2}^2} = \text{ARE}_{\hat{\sigma}_{CL1}^2}$.
2. $\text{ARE}_{\hat{\beta}_{1,CL2}} = \frac{1}{[1 + \rho - \rho^2(1 - a)](1 - b\rho)}$,

where $a = K^{-1} \sum k_i / (k_i - 1)$ and $b = K^{-1} \sum k_i / [1 + (k_i - 1)\rho]$.

When $\rho = 0$, $\text{ARE}_{\hat{\beta}_{1,CL2}} = 1$.

When $\rho > 0$, since $0 \leq a \leq 1$ and $b \geq 1/\bar{k}$,

$$\text{ARE}_{\hat{\beta}_{1,CL2}} \geq \frac{1}{(1 + \rho - \rho^2)(1 - \rho/\bar{k})} \geq \frac{1}{(1 + \rho - \rho^2)} \geq 0.8.$$

3. In Figure 4.6 we plot the efficiency of $\hat{\rho}_{CL1}$, $\hat{\rho}_{CL2}$ and $\hat{\rho}_{BCL}$ for the case of family sizes two and six mixed with proportion 3:1. When ρ is close to 0, the CL2 estimator is worse than both the CL1 and BCL estimators. However, as ρ increases, the CL2 estimator surpasses them at around $\rho = 0.2$.

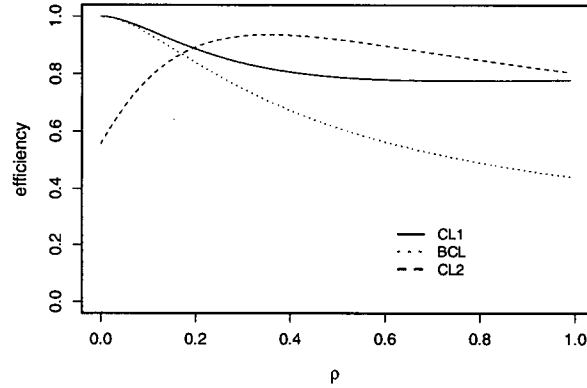


Figure 4.6: Efficiency of $\hat{\rho}_{CL1}$, $\hat{\rho}_{BCL}$ and $\hat{\rho}_{CL2}$ (MVN model); family sizes are 2 or 6 with proportion 3:1. The CL2 estimate is weighted by family size; whereas the BCL estimate is not.

b. Type-3 family

In Figure 4.7 we plot the efficiency of the CL2 estimators for all the parameters for the case of number of offspring varying from 1 to 5 with equal proportion. Similar to Figure 4.2, the efficiency was calculated at four different levels of $\rho_1 = (0, 0.3, 0.5, 0.8)$, and plotted against the conditional correlation $\tilde{\rho}_2$. For the purpose of comparison, we also plotted the efficiency of the CL1 estimators. The plots show that

1. β_0 : when $\tilde{\rho}_2$ is small, the two estimators are similar. For positive $\tilde{\rho}_2$, the CL2 estimator loses less efficiency when either ρ_1 or $\tilde{\rho}_2$ increases.
2. β_1 : unlike the CL1 estimator, the CL2 estimator performs very well when ρ_1 or $\tilde{\rho}_2$ is large.
3. σ^2 : the efficiency pattern is similar to β_0 .
4. ρ_1 and ρ_2 : The CL2 estimator loses efficiency when the dependence is small or negative. It performs better than the CL1 estimator when the dependence increases.

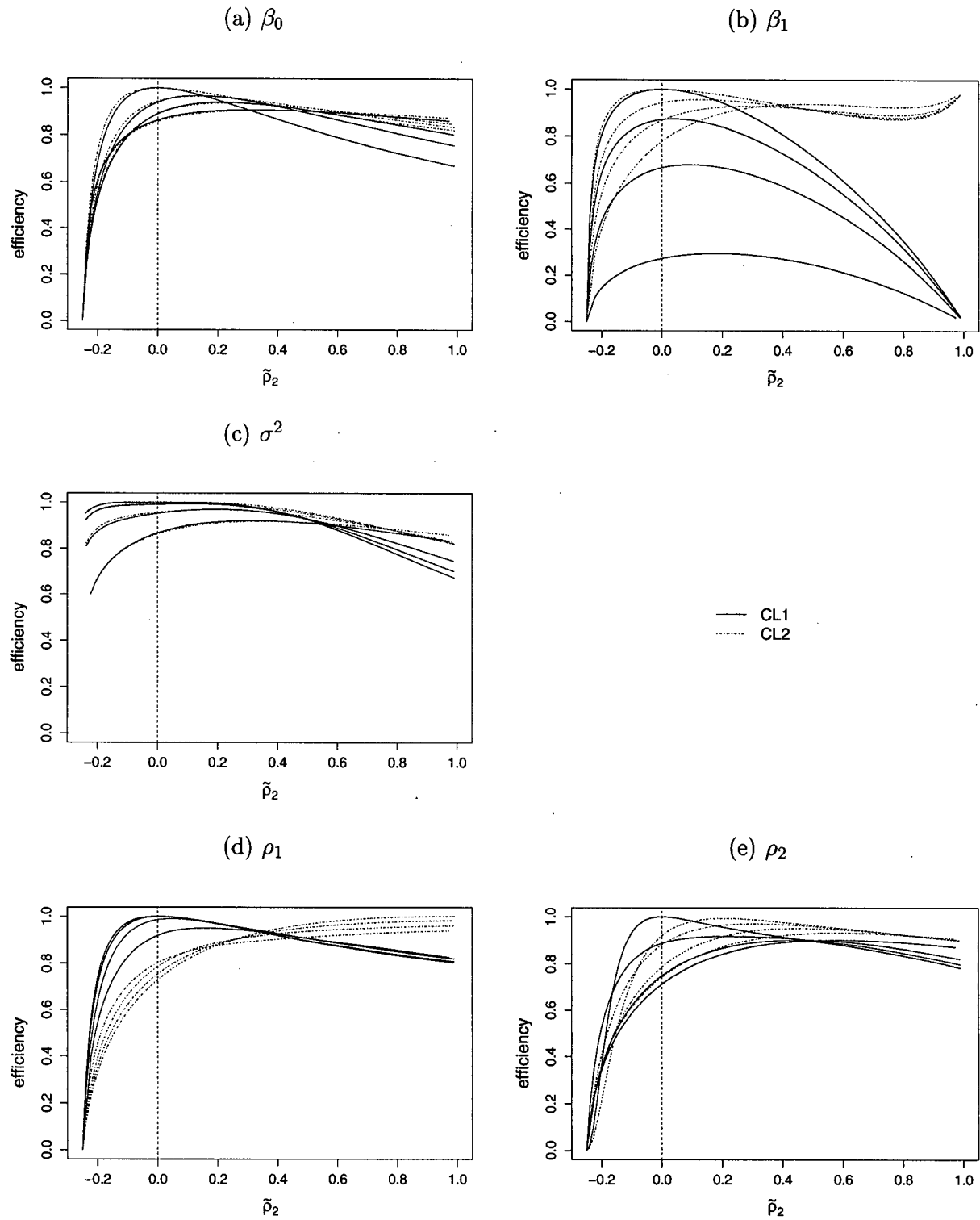


Figure 4.7: Efficiency CL1 and CL2 for Type-3 families. The family size varies; the number of offspring is from 1 to 5 with equal proportion (MVN model).

Discussion

In summary, compared with the CL1 approach, the advantages of the CL2 approach are:

1. the efficiency of the estimate of β_1 is much improved, while the efficiency of the estimates of β_0 and σ^2 is similar to the former approach; and
2. the efficiencies of the estimates of the dependence parameters are improved when the correlations are high.

The drawback of this approach is that when the dependency is weak, the estimates of α are less efficient compared with the CL1 approach.

4.2 Binary Response: MVP Model

The multivariate probit model for binary data is specified in Section 2.2.1 with the probability that $\mathbf{Y}_i = \mathbf{y}_i$ given in (2.5). Let $\rho_i = (\rho_{ijj'}, j > j')$, then $\gamma'_i = (\mu'_i, \rho'_i)$. Suppose $\mu_i = \mathbf{X}_i^{(1)}\beta$ and $\rho_i = \mathbf{X}_i^{(2)}\alpha$. Then $\theta' = (\beta', \alpha')$. We have

$$\tilde{\mathbf{X}}_i = \begin{pmatrix} \mathbf{X}_i^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i^{(2)} \end{pmatrix},$$

and $\gamma'_i = \tilde{\mathbf{X}}_i\theta$.

4.2.1 Likelihood and Composite Likelihood Estimating Functions

MLE

Suppose there are M_i different possible outcomes of \mathbf{Y}_i with non-zero probabilities. Let $\mathbf{y}_{i,l}$ denote the l th outcome and π_{il} denote the associated probability. For the MVP model,

$$\pi_{il} = \int_{\mathfrak{A}_{i,l}^{k_i}} \phi_{k_i}(\mathbf{z}; \mu_i, \mathbf{R}_i(\alpha)) d\mathbf{z},$$

where $\mathfrak{A}_{i,l}^{k_i}$ is the rectangular region in \mathfrak{R}^{k_i} in which \mathbf{Z}_i falls when $\mathbf{Y}_i = \mathbf{y}_{i,l}$. For each i , $\sum_l \pi_{il} = 1$.

The likelihood of \mathbf{Y}_i is

$$L(\mathbf{Y}_i; \theta) = \prod_{l=1}^{M_i} \pi_{il}^{I(\mathbf{Y}_i = \mathbf{y}_{i,l})},$$

where I is the indicator function. The estimating functions are

$$\sum_i \sum_l I(\mathbf{Y}_i = \mathbf{y}_{i,l}) \tilde{\mathbf{X}}_i' \frac{\partial \pi_{il} / \partial \boldsymbol{\gamma}_i}{\pi_{il}} = 0.$$

The Fisher Information matrix is

$$\mathbf{F} = \frac{1}{n} \sum_i \tilde{\mathbf{X}}_i' \left[\sum_l \frac{(\partial \pi_{il} / \partial \boldsymbol{\gamma}_i) (\partial \pi_{il} / \partial \boldsymbol{\gamma}_i)'}{\pi_{il}} \right] \tilde{\mathbf{X}}_i. \quad (4.19)$$

Let $\boldsymbol{\Lambda}_i = (\lambda_{ijj'})$ denote the term inside the square brackets in (4.19). Then

$$\mathbf{F} = \frac{1}{n} \sum_i \tilde{\mathbf{X}}_i' \boldsymbol{\Lambda}_i \tilde{\mathbf{X}}_i. \quad (4.20)$$

The Two-stage Estimating Approach (CL1)

Let $\nu_{ij} = EY_{ij} = 1 - \Phi(\mu_{ij})$, $\boldsymbol{\nu}_i = (\nu_{i1}, \dots, \nu_{ik_i})'$, $\sigma_{ij}^2 = \text{Var}(Y_{ij}) = \Phi(\mu_{ij})[1 - \Phi(\mu_{ij})]$, and $\phi_{ij} = \phi(\mu_{ij})$.

The univariate log likelihood of Y_{ij} is

$$l_1(Y_{ij}; \boldsymbol{\beta}) = (1 - Y_{ij}) \log \Phi(\mu_{ij}) + Y_{ij} \log[1 - \Phi(\mu_{ij})].$$

In the first step, $\boldsymbol{\beta}$ is estimated by the solution of the following equations

$$\sum_i (\mathbf{X}_i^{(1)})' \mathbf{g}_{UCL}(\mathbf{Y}_i, \boldsymbol{\mu}_i) = 0, \quad (4.21)$$

where

$$\mathbf{g}_{UCL}(\mathbf{Y}_i, \boldsymbol{\mu}_i) = \left(\frac{l_1(Y_{i1})}{\partial \mu_{i1}}, \dots, \frac{l_1(Y_{ik_i})}{\partial \mu_{ik_i}} \right).$$

Here

$$\frac{\partial l_1(Y_{ij})}{\partial \mu_{ij}} = \frac{(1 - Y_{ij})\phi_{ij}}{\Phi(\mu_{ij})} - \frac{Y_{ij}\phi_{ij}}{1 - \Phi(\mu_{ij})} = \frac{\phi_{ij}(Y_{ij} - [1 - \Phi(\mu_{ij})])}{\Phi(\mu_{ij})[1 - \Phi(\mu_{ij})]} = \frac{\phi_{ij}(Y_{ij} - \nu_{ij})}{\sigma_{ij}^2}.$$

Let $\mathbf{A}_i = \text{diag}(\phi_{ij}/\sigma_{ij}^2)$. Then we have $\mathbf{g}_{UCL}(\mathbf{Y}_i, \boldsymbol{\mu}_i) = \mathbf{A}_i(\mathbf{Y}_i - \boldsymbol{\nu}_i)$.

In the second step, the bivariate marginal log likelihood is given by:

$$\begin{aligned} l_2(Y_{ij}, Y_{ij'}; \mu_{ij}, \mu_{ij'}, \rho_{ijj'}) &= (1 - Y_{ij})(1 - Y_{ij'}) \log[\Phi_2(\mu_{ij}, \mu_{ij'}; \rho_{ijj'})] \\ &+ (1 - Y_{ij})Y_{ij'} \log[\Phi_1(\mu_{ij}) - \Phi_2(\mu_{ij}, \mu_{ij'}; \rho_{ijj'})] \\ &+ Y_{ij}(1 - Y_{ij'}) \log[\Phi_1(\mu_{ij'}) - \Phi_2(\mu_{ij}, \mu_{ij'}; \rho_{ijj'})] \\ &+ Y_{ij}Y_{ij'} \log[1 - \Phi_1(\mu_{ij}) - \Phi_1(\mu_{ij'}) + \Phi_2(\mu_{ij}, \mu_{ij'}; \rho_{ijj'})] \end{aligned}$$

The estimating equations for α are:

$$\sum_i (\mathbf{X}_i^{(2)})' \mathbf{g}_{2BCL}(\mathbf{Y}_i; \mu_i, \rho_i) = 0$$

where

$$\begin{aligned} \mathbf{g}_{2BCL}(\mathbf{Y}_i; \mu_i, \rho_i) &= \left(\frac{\partial l_2(Y_{ij}, Y_{ij'}; \mu_{ij}, \mu_{ij'}, \rho_{ijj'})}{\partial \rho_{ijj'}}, \quad j > j' \right) \\ &= \left(\frac{(-1)^{Y_{ij}+Y_{ij'}} \phi_2(\mu_{ij}, \mu_{ij'}; \rho_{ijj'})}{L_2(Y_{ij}, Y_{ij'}; \mu_{ij}, \mu_{ij'}, \rho_{ijj'})}, \quad j > j' \right). \end{aligned}$$

The Godambe information matrix for $\hat{\beta}$ is given by (3.12) where

$$\Delta_i^{1,1} = -E \frac{\partial \mathbf{g}_{UCL}(\mathbf{Y}_i, \mu_i)}{\partial \mu_i'} = -E \frac{\partial \mathbf{A}_i}{\partial \mu_i'} (\mathbf{Y}_i - \nu_i) + \mathbf{A}_i \frac{\partial \nu_i}{\partial \mu_i'} = \text{diag} \left(\frac{\phi_{ij}^2}{\sigma_{ij}^2} \right)$$

and $\Omega_i^{1,1} = \text{Cov}(\mathbf{g}_{UCL}(\mathbf{Y}_i, \mu_i)) = \mathbf{A}_i \text{Cov}(\mathbf{Y}_i) \mathbf{A}_i$.

3. CL2 method

To estimate both β and α based on the weighted BCL, the estimating functions are:

$$\begin{aligned} \psi_{1,CL2} &= \sum_{\{i:k_i=1\}} \frac{\partial l_1(Y_{i1})}{\partial \beta} + \sum_{\{i:k_i>1\}} \frac{1}{k_i-1} \sum_{j>j'} \frac{\partial l_2(Y_{ij}, Y_{ij'})}{\partial \beta} \\ &= \sum_{\{i:k_i=1\}} \frac{\partial \mu_{i1}}{\partial \beta} \frac{\phi_{i1}(Y_{i1} - \mu_{i1})}{\Phi_{i1}(1 - \Phi_{i1})} + \sum_{i \in \{i:k_i>1\}} \frac{1}{k_i-1} \sum_j \frac{\partial \mu_{ij}}{\partial \beta} \phi_{ij} \sum_{j \neq j'} \frac{(-1)^{1-Y_{ij}} \Phi_{ij'|j}^{Y_{ij'}} (1 - \Phi_{ij'|j}^{1-Y_{ij'}})}{L_2(Y_{ij}, Y_{ij'})}, \end{aligned}$$

$$\begin{aligned} \psi_{2,CL2} &= \sum_{\{i:k_i=1\}} \frac{1}{k_i-1} \sum_{j>j'} \frac{\partial l_2(Y_{ij}, Y_{ij'})}{\partial \alpha} \\ &= \sum_{\{i:k_i=1\}} \frac{1}{k_i-1} \sum_{j>j'} \frac{\partial \rho_{ijj'}}{\partial \alpha} \frac{(-1)^{Y_{ij}+Y_{ij'}} \phi_2(\mu_{ij}, \mu_{ij'}; \rho_{ijj'})}{L_2(Y_{ij}, Y_{ij'})}, \end{aligned}$$

where $\Phi_{i1} = \Phi(\mu_{i1})$ and $\Phi_{ij'|j} = \Phi \left(\frac{\mu_{ij'} - \rho_{ijj'} \mu_{ij}}{\sqrt{1 - \rho_{ijj'}^2}} \right)$.

4.2.2 Efficiency Comparison

In the first part of the comparison, we focus on the efficiency of $\hat{\beta}_{CL1}$, since in the previous section, we observed that the efficiency is strongly affected by the within-family dependence. In the second

part we focus on comparing the efficiencies of all parameter estimators from the CL1, CL1 with optimal weights (WCL1) and CL2 methods.

The AREs need to be evaluated from the information matrices which depend on the values of the covariates and their coefficients. In our comparison, we considered one covariate and simulated the values of the covariate from a uniform distribution on $(-1, 1)$. These values are independent both within a family and among families. The efficiency is compared at different levels of β_0 and β_1 which are listed in Table 4.2 along with the corresponding range of $\Pr(Y = 1)$. Only positive values of β were investigated; by symmetry, the results can be extended to negative values.

β_0	β_1	$\Pr(Y = 1)$
0	0	0.50
0	1	0.159 ~ 0.841
0	2	0.023 ~ 0.977
0.67	0	0.749
0.67	1	0.371 ~ 0.953
0.67	2	0.092 ~ 0.996
1.65	0	0.951
1.65	1	0.742 ~ 0.996
1.65	2	0.363 ~ 1.000

Table 4.2: Range of $\Pr(Y = 1)$ for each combination of β_0 and β_1 (MVP model).

Efficiency of $\hat{\beta}_{CL1}$

To gain some understanding of how the performance of the estimators is affected by the degree of dependence, we consider the bivariate case; that is, the family size $k_i = 2$ for all families.

We generated x values for 1000 families. The efficiency of $\hat{\beta}_{0,CL1}$ and $\hat{\beta}_{1,CL2}$ was evaluated at the nine combinations of β_0 and β_1 when ρ varied in the range $(-1, 1]$. The results are plotted in Figure 4.8. As expected, the efficiency of $\hat{\beta}_{0,CL1}$ is only slightly affected by ρ . For positive ρ , $ARE_{\hat{\beta}_{0,CL1}}$ is generally above 0.9. The efficiency of $\hat{\beta}_{1,CL1}$ is affected more by ρ . In all the cases, the stronger the dependence among \mathbf{Z}_i , the less efficient is $\hat{\beta}_{1,CL1}$. When $\beta_1 = 0$ and ρ_1 is close 1, $ARE_{\hat{\beta}_{1,CL1}}$ is close to 0.

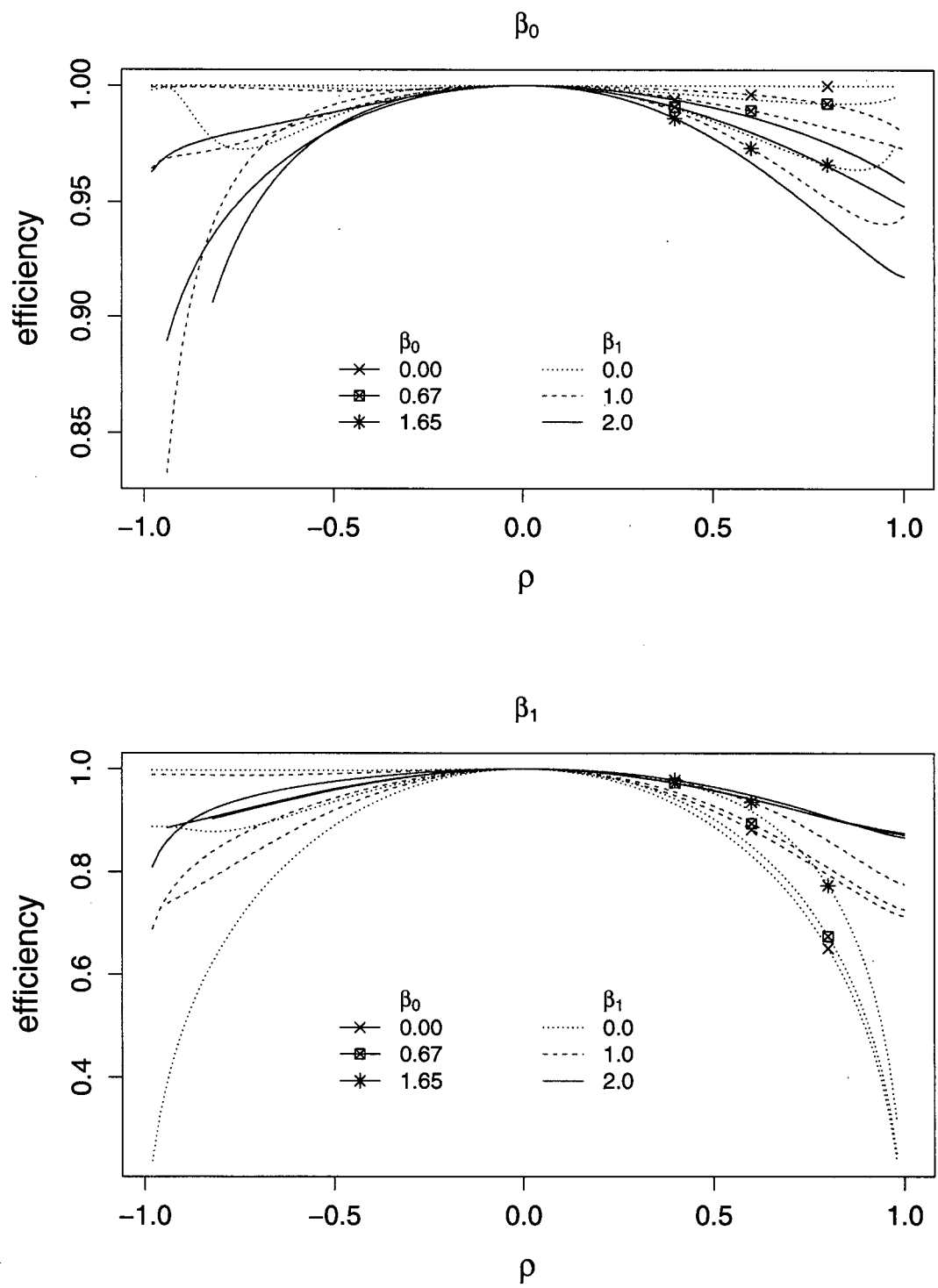


Figure 4.8: ARE of $\hat{\beta}_{0,CL1}$ and $\hat{\beta}_{1,CL1}$ when family size $k = 2$. (MVP model)

Next we will have a closer look at the worst scenario: Z 's are perfectly dependent within family and $\beta_1 = 0$. The family size is not restricted.

When $\beta_1 = 0$, then $\mu_{ij} = \beta_0$ for all j . Perfect dependence then implies $Z_{i1} = Z_{i2} = \dots = Z_{ik_i}$. As a result, $Y_{i1} = Y_{i2} = \dots = Y_{ik_i}$. By the ML method, β_1 can be estimated without error. The MLE of β_0 is $\hat{\beta}_{0,MLE} = \Phi^{-1}(\bar{Y})$, where $\bar{Y} = \sum_i Y_{i1}/n$. Let $\sigma^2 = \text{Var}(Y_{i1}) = \Phi(\beta_0)(1 - \Phi(\beta_0))$ and $\delta = \sigma^2/\phi^2(\beta_0)$. Then $\text{Var}(\hat{\beta}_{0,MLE}) = \delta/n$. For $\hat{\beta}_{CL1}$,

$$\frac{1}{n}\mathbf{D}_{1,11} = \frac{\delta}{n} \sum_i (\mathbf{X}_i^{(1)})' \mathbf{X}_i^{(1)} = \frac{\delta}{n} \begin{pmatrix} \sum_i k_i & \sum_{ij} x_{ij} \\ \sum_{ij} x_{ij} & \sum_{ij} x_{ij}^2 \end{pmatrix}$$

$$\frac{1}{n}\mathbf{M}_{1,11} = \frac{\delta}{n} \sum_i (\mathbf{X}_i^{(1)})' \mathbf{J}_i \mathbf{X}_i^{(1)} = \delta \begin{pmatrix} \sum_i k_i^2 & \sum_{ij} k_i x_{ij} \\ \sum_{ij} k_i x_{ij} & \sum_i (\sum_j x_{ij})^2 \end{pmatrix}$$

Suppose x is centred, i.e., $\sum_{ij} x_{ij} = 0$. Then $\text{Var}(\hat{\beta}_{0,CL1}) = \delta \sum_i k_i^2 / (\sum_i k_i)^2$. When k_i is constant across all families, $\text{Var}(\hat{\beta}_{0,CL1}) = \delta/n$ — the same as $\text{Var}(\hat{\beta}_{0,MLE})$. Meanwhile, $\text{Var}(\hat{\beta}_{1,CL1}) = \delta \sum_i (\sum_j x_{ij})^2 / (\sum x_{ij}^2)^2$. Since there is no estimation error in the MLE, the efficiency of $\hat{\beta}_1$ is zero.

As shown in Figure 4.8, when $\beta_1 \neq 0$, the efficiency is no longer zero since the MLE is also subject to estimating error.

Comparison of CL1, WCL1 and CL2 Estimates

a. Exchangeable

In the following comparisons, efficiency was computed under the first four combinations of β_0 and β_1 in Table 4.2. For each combination, ρ varies from 0 to 0.90. The other combinations of β_0 and β_1 are not included since in those cases the numerical results are not reliable when ρ exceeds a certain level.

The x values of 500 families were generated for each simulation with family size being randomly generated from 1 to 4 (the number of families of each size is 141, 139, 105 and 115, respectively).

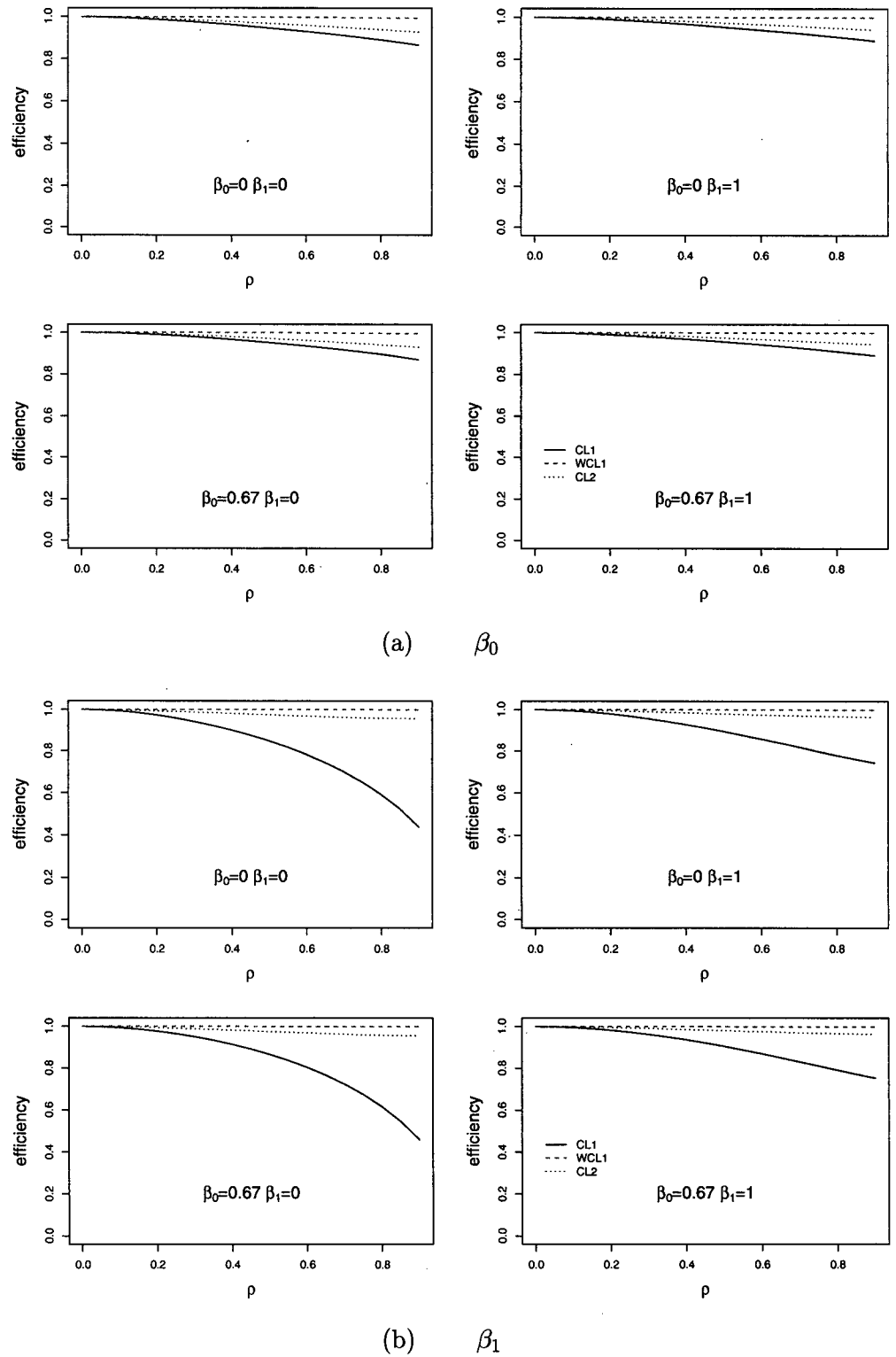


Figure 4.9: ARE of $\hat{\beta}_{CL1}$, $\hat{\beta}_{WCL1}$ and $\hat{\beta}_{CL2}$ for the exchangeable case with family size varying from 1 to 4 (MVP model).

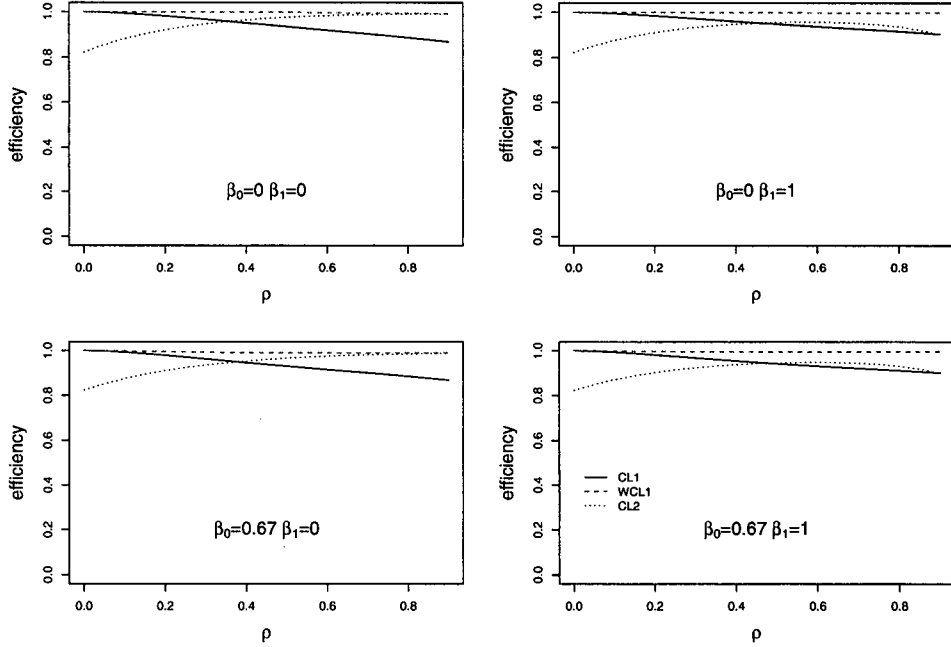
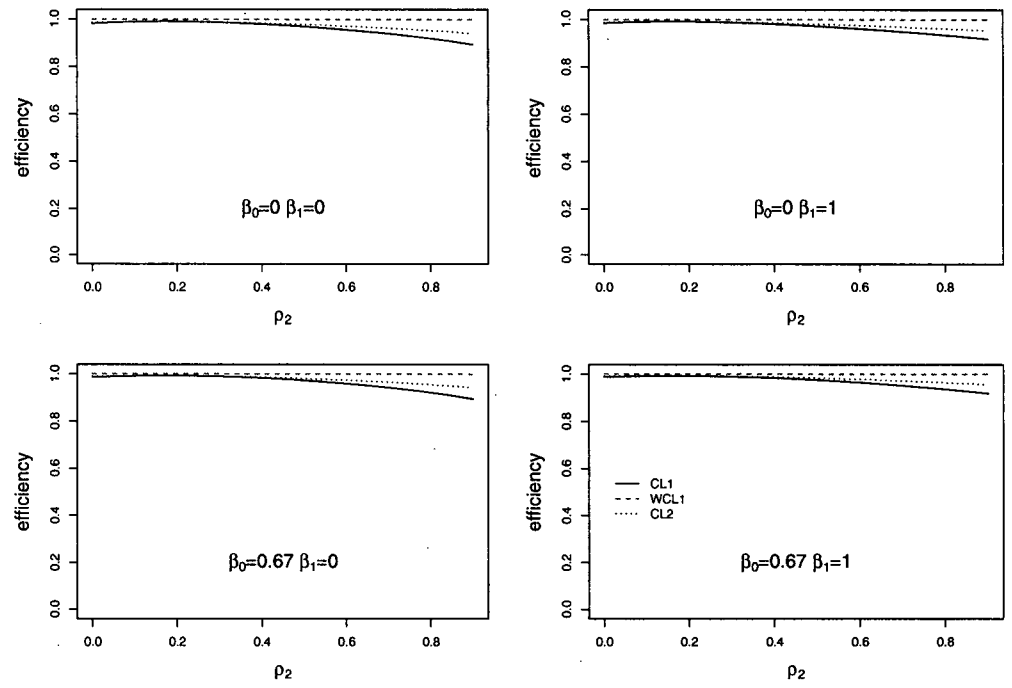


Figure 4.10: ARE of $\hat{\rho}_{CL1}$, $\hat{\rho}_{WCL1}$ and $\hat{\rho}_{CL2}$ for the exchangeable case with family size varying from 1 to 4. (MVP model)

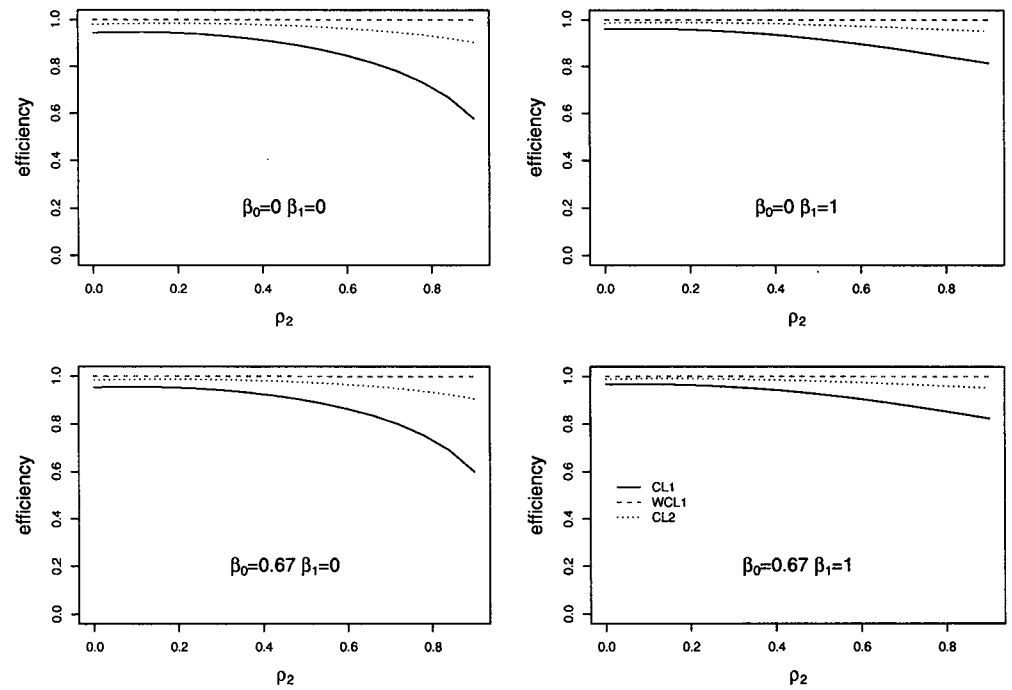
The results are plotted in Figure 4.9 and Figure 4.10. The following patterns are observed:

1. The efficiency of the weighted CL1 estimator is always close to 1 under all the conditions and for all three parameters.
2. The CL2 estimators of β_0 and β_1 are better than the CL1 estimators. The improvement of $\hat{\beta}_{1,CL2}$ is substantial when ρ is close to 1 and β_1 is close to 0.
3. $\hat{\rho}_{CL2}$ loses efficiency when ρ is close to 0 and β_1 is close to 1.

Recall that in the MVN model, $ARE(\hat{\beta}_{0,CL2}) = ARE(\hat{\beta}_{0,CL1})$ for the exchangeable case. However, in the MVP model, $ARE(\hat{\beta}_{0,CL2})$ is larger than $ARE(\hat{\beta}_{0,CL1})$. If we compare the efficiency loss of $\hat{\beta}_{0,CL1}$ in the MVP model to that in the MVN model, we find that there is more efficiency loss in the MVP model.



(a) β_0



(b) β_1

Figure 4.11: ARE of $\hat{\beta}_{CL1}$, $\hat{\beta}_{WCL1}$ and $\hat{\beta}_{CL2}$ for Type-3 families, at $\rho_1 = 0.3$ (MVP model). Solid line: CL1, dashed line: WCL1, dotted line: CL2.

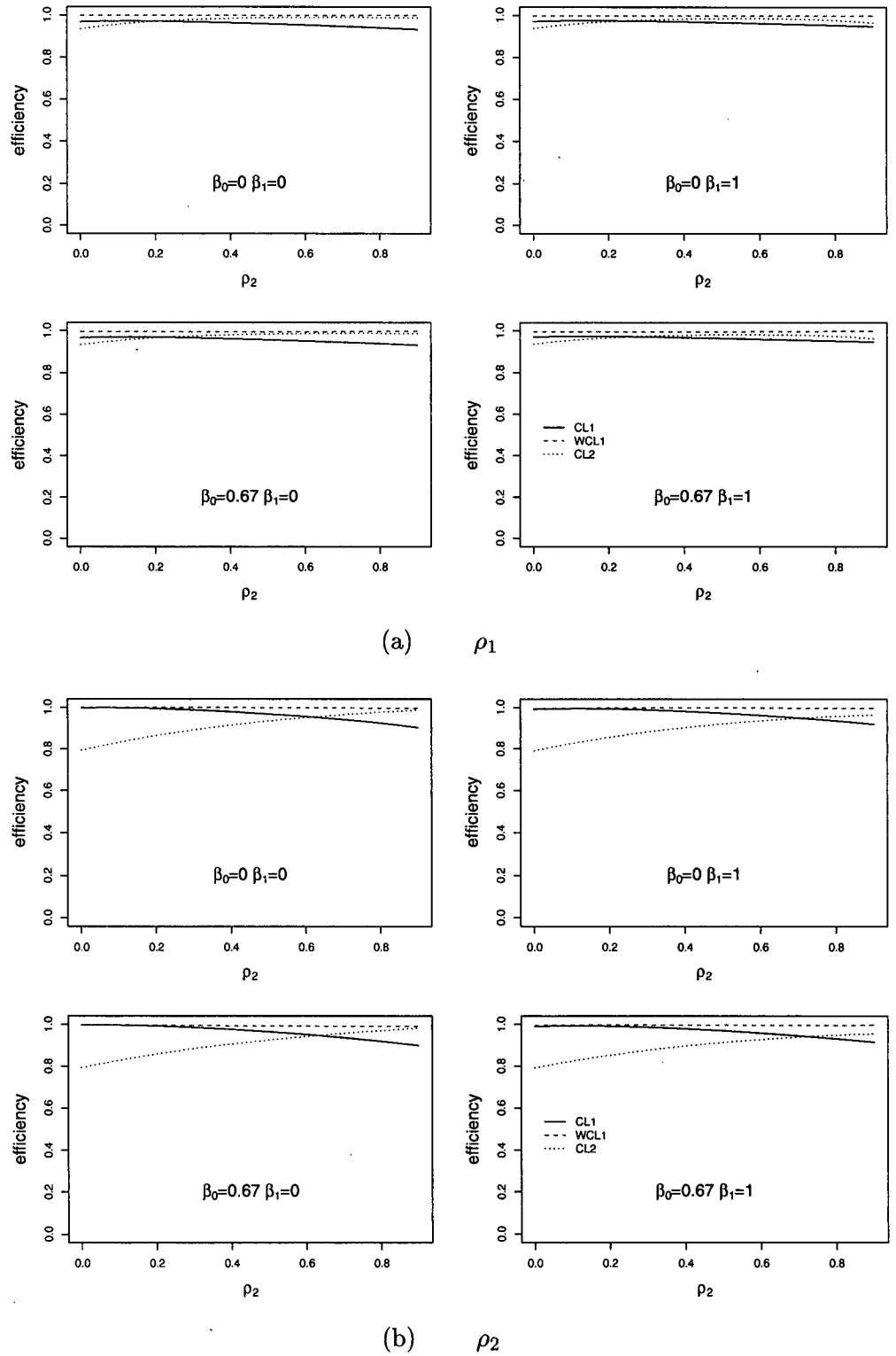
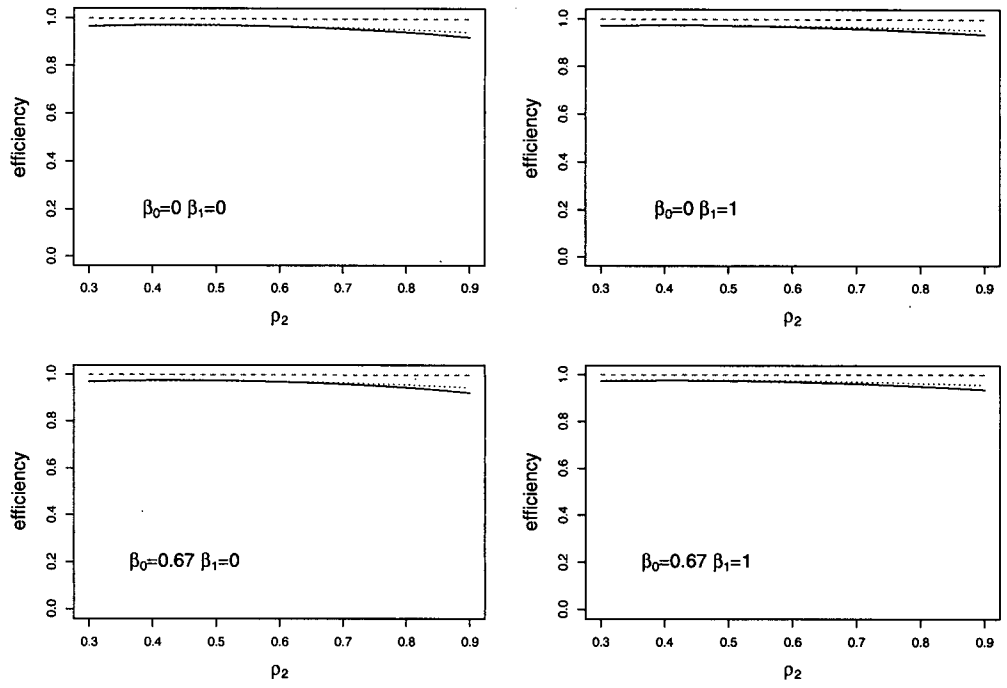
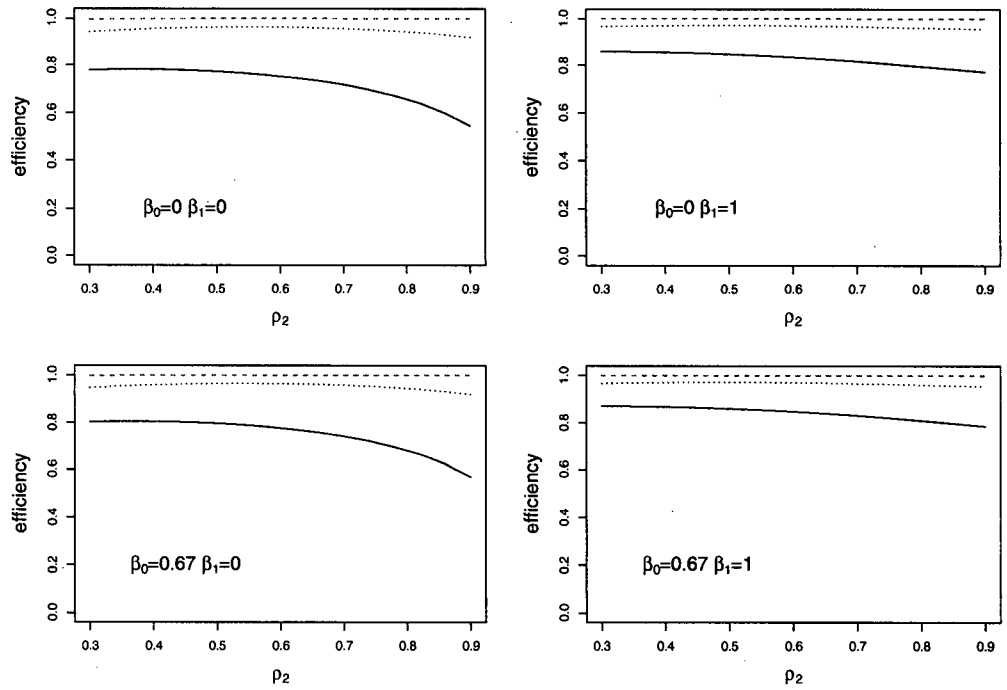


Figure 4.12: ARE of $\hat{\rho}_{CL1}$, $\hat{\rho}_{WCL1}$ and $\hat{\rho}_{CL2}$ at $\rho_1 = 0.3$ for Type-3 families (MVP model). Solid line: CL1, dashed line: WCL1, dotted line: CL2.



(a) β_0



(b) β_1

Figure 4.13: ARE of $\hat{\beta}_{CL1}$, $\hat{\beta}_{WCL1}$ and $\hat{\beta}_{CL2}$ for Type-3 families, at $\rho_1 = 0.6$ (MVP model). Solid line: CL1, dashed line: WCL1, dotted line: CL2.

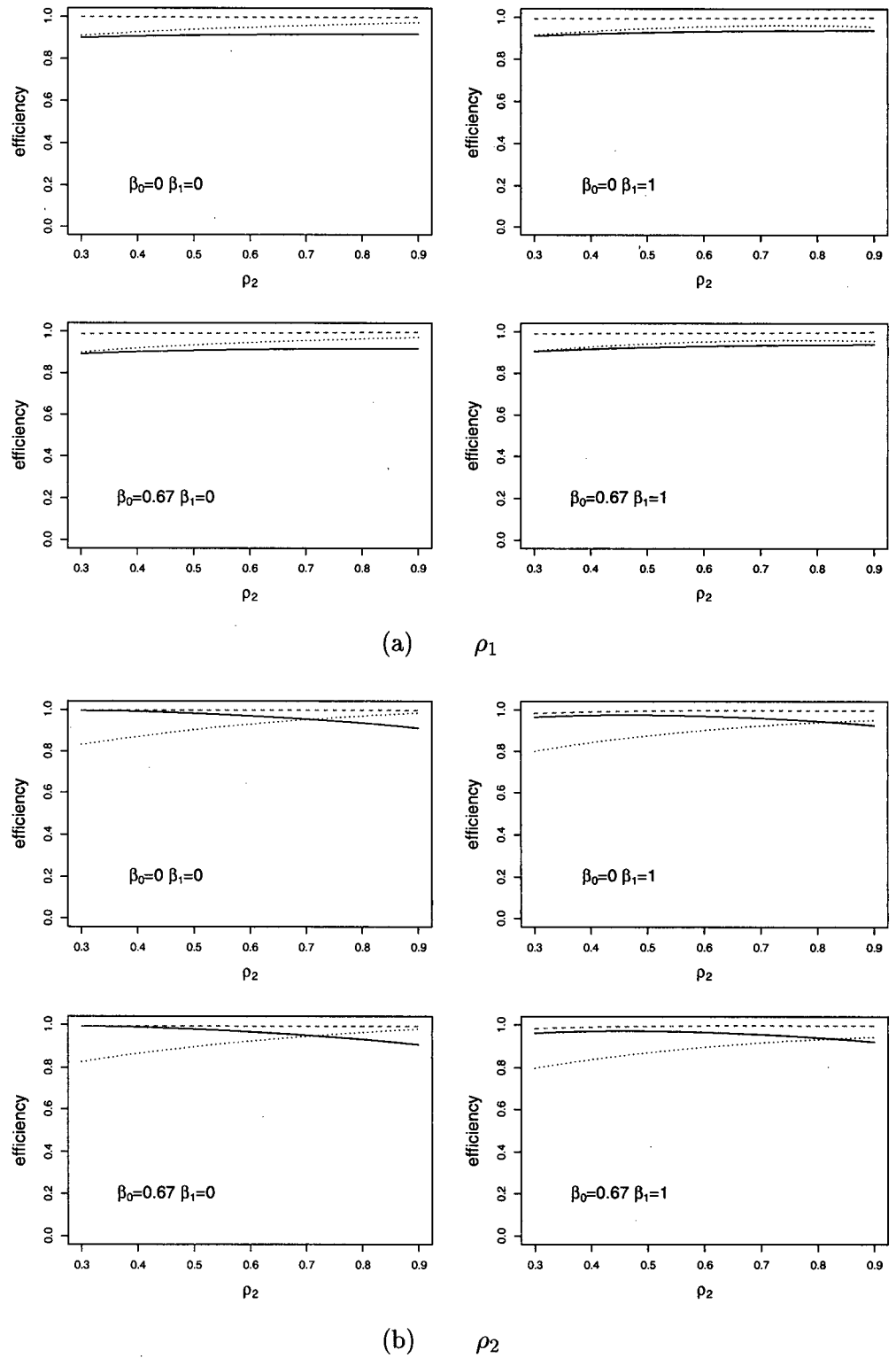


Figure 4.14: ARE of $\hat{\rho}_{CL1}$, $\hat{\rho}_{WCL1}$ and $\hat{\rho}_{CL2}$ for Type-3 families, at $\rho_1 = 0.6$ (MVP model). Solid line: CL1, dashed line: WCL1, dotted line: CL2.

b. Type-3 family

Again, 500 families were generated. In each family, the number of offspring was randomly generated from 1 to 3; there were 165 families with one offspring, 163 with two and 172 with three. The same four combinations for β_0 and β_1 were considered. In the first case, the parent-offspring correlation ρ_1 was fixed at 0.3 and the sib-sib correlation ρ_2 varied from 0 to 0.9. The efficiency of the four parameters was plotted in Figure 4.11 and Figure 4.12. In the second case, ρ_1 was set at 0.6 and ρ_2 was varied from 0.3 to 0.9. The results are in Figure 4.13 and Figure 4.14.

The patterns for the efficiency of estimation of β_0 and β_1 are the same as in the exchangeable case. $\hat{\beta}_{1,CL1}$ lost more efficiency when $\rho_1 = 0.6$ than when $\rho_1 = 0.3$. All three estimators of ρ_1 are highly efficient. The efficiency of $\hat{\rho}_{2,CL2}$ is relatively low when the dependence is weak.

4.2.3 Conclusion

For the MVP model, we observe patterns similar to those seen in the case of the MVN model. As familial dependence increases, $\hat{\beta}_{1,CL1}$ loses efficiency, while $\hat{\beta}_{1,CL2}$ is generally more efficient. The CL2 estimator of the dependence parameters loses efficiency when the dependence is weak.

Unlike the MVN, for the MVP model the efficiency is affected by the magnitude of β . The effect on the CL2 and weighted CL1 methods is rather small, but is greater for the CL1 estimators. In general, at a fixed dependence level, the efficiencies of the CL1 estimators are relatively higher when β_1 is far enough from 0 so that a change in x causes a notable change in the marginal probability.

4.3 Count Response: Multivariate Poisson-Lognormal Mixture Model

The model is defined in Section 2.3.1 with pmf given in (2.11). In this model, $\theta_1 = (\beta, \sigma^2)$, $\theta_2 = \alpha$, $\gamma'_{i1} = (\mu'_i, \sigma_i^2)$ and $\gamma_{i2} = (\rho_{ijj'}, j > j')$.

4.3.1 ML and CL estimating functions

ML

The score functions are

$$\begin{aligned} S(\theta) &= \sum_i \sum_{y_{ij}=0}^{\infty} \cdots \sum_{y_{ik_i}=0}^{\infty} \frac{1}{\Pr(\mathbf{Y}_i = \mathbf{y}_i)} \int \prod_i \Pr(Y_i = y_i | \Lambda_i = e^{z_i}) \left(\frac{\partial \gamma_i}{\partial \theta} \right)' \frac{\partial \phi_k(\mathbf{z}; \gamma_i)}{\partial \gamma_i} d\mathbf{z} \\ &= \sum_i \left(\frac{\partial \gamma_i}{\partial \theta} \right)' \sum_{y_{ij}=0}^{\infty} \cdots \sum_{y_{ik_i}=0}^{\infty} \frac{1}{\Pr(\mathbf{Y}_i = \mathbf{y}_i)} \int \prod_i \Pr(Y_i = y_i | \Lambda_i = e^{z_i}) \frac{\partial \phi_k(\mathbf{z}; \gamma_i)}{\partial \gamma_i} d\mathbf{z}. \end{aligned}$$

Let $\pi(\mathbf{y}_i; \gamma_i(\theta)) = \Pr(\mathbf{Y}_i = \mathbf{y}_i)$.

CL1 and CL2

The univariate marginal likelihood function of Y_{ij} is

$$l_1(Y_{ij}) = \log \int \Pr(Y_{ij} = y_{ij} | \Lambda_{ij} = e^{z_{ij}}) \phi_1(z_{ij}; \mu_{ij}, \sigma^2) dz_{ij}$$

and the bivariate marginal likelihood function of Y_{ij} and $Y_{ij'}$ is

$$\begin{aligned} l_2(Y_{ij}, Y_{ij'}) &= \\ \log \int \int \Pr(Y_{ij} = y_{ij} | \Lambda_{ij} = e^{z_{ij}}) \Pr(Y_{ij'} = y_{ij'} | \Lambda_{ij'} = e^{z_{ij'}}) \phi_2(z_{ij}, z_{ij'}; \mu_{ij}, \mu_{ij'}, \sigma^2, \rho_{ijj'}) dz_{ij} dz_{ij'} \end{aligned}$$

The CL1 and CL2 estimating functions are formed as described in Section 3.2.2 and Section 3.2.3.

4.3.2 Efficiency Comparison

Since the calculation of the information matrices is difficult to carry out, the efficiency comparison is conducted by simulations. We considered the exchangeable dependence structure.

In each simulation we generated 2,000 samples each containing 3,000 families. The family sizes were randomly generated from 1 to 4. To reduce the computing time, we use a discrete covariate which was randomly generated with value -1 , 0 or 1 .

We first fixed $\beta_1 = 0.5$, $\sigma^2 = 0.25$. We chose three different levels of $\beta_0 = -1, 0$ and 1 . For $\beta_0 = -1$, the count variable is close to a binary variable — around 90% of the counts are 0 or 1. As β_0 increases, the count variable becomes closer to a continuous random variable. At each level of β_0 , we conducted three simulations with the correlation between the latent variables, ρ , set at 0.2, 0.5 and 0.8, respectively. The AREs are reported in Table 4.3. The mean and variance reported in the table are calculated at $x_{ij} = 0$; ρ^* is $\text{Corr}(Y_{ij}, Y_{ij'})$ with $x_{ij} = x_{ij'} = 0$. In some cases, the ARE is slightly over 1 due to simulation noise.

	ρ	ρ^*	CL1				CL2			
			β_0	β_1	σ^2	ρ	β_0	β_1	σ^2	ρ
$\beta_0 = -1$	0.2	0.019	1.002	1.002	1.002	0.997	1.001	1.002	1.001	0.876
mean: 0.417	0.5	0.050	0.997	0.990	0.999	0.997	0.997	0.996	1.001	0.938
variance: 0.466	0.8	0.083	0.952	0.986	0.907	0.989	0.984	0.995	0.976	0.954
$\beta_0 = 0$	0.2	0.044	0.998	0.992	1.006	1.003	0.998	0.997	1.002	0.875
mean: 1.133	0.5	0.114	0.982	0.966	0.985	0.982	0.988	0.989	0.994	0.912
variance: 1.498	0.8	0.190	0.982	0.927	0.973	0.968	0.994	0.981	0.995	0.959
$\beta_0 = 1$	0.2	0.084	1.000	0.987	0.999	0.992	0.998	0.996	1.001	0.908
mean: 3.080	0.5	0.219	0.984	0.924	0.966	0.972	0.989	0.984	0.985	0.933
variance: 5.775	0.8	0.364	0.956	0.830	0.927	0.936	0.980	0.971	0.977	0.943

Table 4.3: ARE of the CL1 and CL2 estimators for the exchangeable case at different levels of β_1 and ρ (MPLN model).

At all levels of β_0 , the efficiency of $\hat{\beta}_{1,CL1}$ decreases as ρ increases and the efficiency loss increases with β_0 . The efficiency of $\hat{\sigma}_{CL1}^2$ is relatively low when $\rho = 0.8$. $\hat{\beta}_{1,CL2}$ is more efficient than $\hat{\beta}_{1,CL1}$ and less affected by ρ and β_0 . The efficiency of $\hat{\rho}_{CL2}$ is lower than $\hat{\rho}_{CL1}$ except for the last case, $\beta_0 = 1$ and $\rho = 0.8$ and that could be due to simulation noise. Also we see that ρ^* is small for all the other cases. This is consistent with the MVN and the MVP models, in which $\hat{\rho}_{CL2}$ also tends to be less efficient when the familial dependence is weak.

When σ^2 increases, the counts become more and more overdispersed relative to the Poisson distribution. We also did simulations to investigate how this affects the estimates. In these simulations, we fixed $\beta_0 = 0$, $\beta_1 = 0.5$, $\rho = 0.5$, and three different level of σ^2 were chosen: 0.1, 0.25 and 1. The AREs are reported in Table 4.4. Even though ρ is fixed, as σ^2 increases, ρ^* increases and $\hat{\beta}_{1,CL1}$ becomes less efficient. The estimator $\hat{\sigma}_{CL1}^2$ also loses efficiency as σ^2 increases. As the CL1 method tends to lose efficiency as dependence is high, we also did a simulation with $\sigma^2 = 1$ and $\rho = 0.8$ to get an idea of how much worse the estimate can be. The results are also reported in Table 4.4. In this case, the efficiencies of $\hat{\beta}_{1,CL1}$ and $\hat{\sigma}_{CL1}^2$ are much lower than the corresponding estimates from CL2. Moreover, the CL2 method provides a better estimate of ρ as well.

We also conducted simulations with $\beta_0 = 0$, $\sigma^2 = 0.25$, $\rho = 0.5$, and different levels of β_1 . There is very little change in the efficiencies of all the parameters, and therefore the results are not tabulated here.

σ^2	Mean	Variance	ρ^*	CL1				CL2			
				β_0	β_1	σ^2	ρ	β_0	β_1	σ^2	ρ
$\rho = 0.5$											
0.10	1.051	1.167	0.049	1.0021	0.992	1.001	0.998	1.009	0.996	1.001	0.913
0.25	1.133	1.498	0.114	0.997	0.990	0.999	0.997	0.997	0.996	1.001	0.938
1.00	1.649	6.319	0.279	0.955	0.852	0.896	0.940	0.977	0.972	0.989	0.949
$\rho = 0.8$											
1.00	0.527	1.649	.632	0.895	0.673	0.809	0.932	0.955	0.954	0.953	0.963

Table 4.4: ARE of the CL1 and CL2 estimators for the exchangeable case at different levels of σ^2 and ρ (MVPL model).

In conclusion, both β_0 and σ^2 affect the efficiency of the CL1 estimates, especially $\hat{\beta}_{1,CL1}$. Once again, we observed (1) efficiency loss of $\hat{\beta}_{1,CL1}$ when the dependence is strong; (2) the CL2 method is less efficient than the CL1 method for the dependence parameters when the dependence is weak.

4.4 Survival Data Subject to Right Censoring: Multivariate Log-normal Model

The goal of this section is to gain some knowledge on how the CL1 or CL2 estimates are affected by the presence of censoring. Let us consider failure times with a multivariate lognormal distribution subject to right censoring. The model is described in Section 2.4.1. The weighted CL1 method is not considered in this section since the optimal weights are difficult to compute in the presence of censoring.

When censoring is present, the asymptotic variances of the estimates generally depend on the censoring mechanism. For the case of random censoring, these depend on the distribution of the censoring time C . Moreover, the information matrices are difficult to derive even when a censoring distribution is specified. For this reason, a simulation study was conducted to assess the efficiency.

In the simulation study, we consider an exchangeable dependence structure among failure times within a family. The marginal distribution of $\log(T_{ij})$ is normal with mean $\mu_{ij} = \beta_0 + \beta_1 x_{ij}$ and variance σ^2 . The value of x_{ij} was generated from uniform $(-1, 1)$. For convenience, the distribution of the log censoring time, $\log(C_{ij})$, is also chosen as normal with parameters μ_c and

ρ	r	CL1				CL2			
		β_0	β_1	σ^2	ρ	β_0	β_1	σ^2	ρ
0.1	0	0.996	0.985	1.000	0.991	0.998	0.997	1.000	0.913
	0.2	0.995	0.984	0.999	0.997	0.996	0.995	1.000	0.889
	0.5	0.994	0.988	0.996	0.992	0.995	0.996	0.999	0.876
	0.8	0.998	0.994	0.999	0.996	1.000	0.999	1.000	0.884
0.4	0	0.937	0.788	0.979	0.957	0.960	0.968	0.987	0.967
	0.2	0.945	0.807	0.980	0.946	0.974	0.970	0.993	0.964
	0.5	0.945	0.842	0.973	0.936	0.978	0.969	0.992	0.928
	0.8	0.962	0.906	0.976	0.943	0.986	0.977	0.993	0.903
0.7	0	0.885	0.447	0.924	0.916	0.942	0.958	0.961	0.957
	0.2	0.882	0.480	0.909	0.908	0.944	0.949	0.961	0.952
	0.5	0.874	0.547	0.902	0.891	0.949	0.945	0.966	0.922
	0.8	0.894	0.677	0.905	0.887	0.964	0.939	0.966	0.898
0.9	0	0.846	0.160	0.863	0.885	0.924	0.966	0.934	0.943
	0.2	0.841	0.183	0.845	0.876	0.926	0.947	0.937	0.940
	0.5	0.840	0.239	0.839	0.869	0.943	0.932	0.949	0.925
	0.8	0.871	0.353	0.867	0.889	0.982	0.913	0.976	0.919

Table 4.5: ARE of the CL1 and CL2 estimators for the exchangeable case with family size varying from 1 to 4 (MVN model with right censoring).

σ^2 . C_{ij} is independent from T_{ij} and independent both within and between families. By varying μ_c , we can achieve the desired censoring rate. Suppose r is the target censoring rate. We choose μ_c such that $\Pr(C_{ij} < T_0) = r$, where $\log(T_0) \sim N(\beta_0, \sigma^2)$. In each generated sample, the censoring rate fluctuates around r , but the average is close enough to r .

In each simulation we generated 5,000 samples each containing 2,000 families. The family size was randomly generated from 1 to 4. We fixed $\beta_0 = 1$, $\beta_1 = 1$ and $\sigma^2 = 4$ since they have little impact on the efficiency. The censoring rate r was chosen at four different levels: 0, 0.2, 0.4 and 0.8. The case of $r = 0$ was also considered for the purpose of comparison. Four levels of dependence were considered: $\rho = 0.1, 0.4, 0.7$ and 0.9 . The AREs are shown in Table 4.5.

The following is a summary of the simulation results:

1. The CL1 method:

- Regardless of the degree of censoring, $\hat{\beta}_{1,CL1}$ loses efficiency as ρ increases. However, the loss of efficiency becomes less when r increases.

- (b) The efficiency of other parameters is generally high and not much affected by the censoring rate, especially when ρ is small. When ρ is relatively large, there is a decreasing trend in $ARE_{\hat{\sigma}_{CL1}^2}$ and $ARE_{\hat{\rho}_{CL1}}$.

2. The CL2 method:

- (a) The estimator $\hat{\rho}_{CL2}$ loses efficiency when ρ is close to 0 and, among the four parameters, it is also the most affected by the censoring rate. Its efficiency decreases with r , especially when $\rho = 0.4$ or 0.7 .
- (b) The other parameters are not much affected by the censoring rate. The performance of $\hat{\beta}_{1,CL2}$ is satisfactory. We observed an increasing trend in $ARE_{\hat{\beta}_{0,CL2}}$ and $ARE_{\hat{\sigma}_{CL2}^2}$, and a decreasing trend in $ARE_{\hat{\beta}_{1,CL2}}$ when $\rho = 0.7$ or 0.9 .

4.5 Conclusion

From the different models and dependence structures investigated in this chapter, we obtained similar results. The following is a summary of the main patterns for relative efficiency.

1. Regression parameters: the CL1 method is easily affected by the following factors. (1) dependence: it tends to lose more efficiency when dependence becomes stronger. The efficiency can be 0 in certain extreme cases. (2) data type: there is more efficiency loss for continuous responses than discrete responses. (3) censoring rate: when right censoring occurs, there is less efficiency loss when the censoring rate increases. (4) family size: the efficiency decreases with the average family size. The CL2 method is less affected by these factors. The efficiency of the CL2 estimate is close to 1 most of the time and generally better than that of the CL1 estimate.
2. Other univariate parameters: both methods are reasonable.
3. Dependence parameters: the CL2 method is generally better for stronger dependence, even though exceptions can occur; the CL1 method is better for weaker dependence.
4. Effect of family size: the efficiency of all the parameters is negatively associated with the mean and relative dispersion of family size (measured by the variance-mean ratio).

We make a few final remarks about these two approaches. They are not limited to familial data. We can apply them to other correlated data. We recommend the CL2 method when the initial data analysis suggests a strong dependence. It generally provides better estimation for all the parameters, especially the regression parameters. However, it is numerically more difficult to implement since computation increases when the total number of parameters increases. Therefore we recommend the CL1 method when the initial data analysis suggests a weak dependence.

Chapter 5

Inferences for Log Odds Ratio with Dependent Pairs

This chapter is motivated by an association study of clinical features on NF1 patients (see Szudek et al. (2000)). The objective of the study is to quantify the association between the occurrence of the same clinical feature in relatives affected with NF1. Two associations are of special interest in the study: interclass association, i.e., association between parent and offspring. (2) intraclass association, i.e., association between siblings. As mentioned in an earlier chapter, comparing the strength of these two types of associations may help to reveal characteristics that separate familial and non-familial factors.

For continuous variables with a multivariate normal distribution, the intraclass and interclass correlations are used to measure the sib-sib and parent-offspring associations. For a binary response, we study the intraclass and interclass (log) odds ratios as analogies of the intraclass and interclass correlations.

Let X and Y be the indicators of the occurrence of a feature for a parent and an offspring, respectively. If the feature is present, i.e., the individual is affected, the indicator equals 1. Then there are four possible outcomes: (1, 1), (1, 0), (0, 1), and (0, 0). Suppose the probabilities associated with the four outcomes are $\mathbf{P} = (P_1, P_2, P_3, P_4)'$. The interclass odds ratio is

$$r_{po} = \frac{P_1 P_4}{P_2 P_3}.$$

The intraclass odds ratio, r_{ss} , is defined in a similar way.

If we have a random sample of n independent parent and offspring pairs, (X_i, Y_i) , $i = 1, \dots, n$, and these pairs are identically distributed Bernoulli random variables. We can form a 2×2 table with counts, O_1, O_2, O_3 and O_4 , of the four outcomes (1, 1), (1, 0), (0, 1), and (0, 0) (Table 5.1). The odds ratio can be estimated by

$$\hat{r} = \frac{O_1 O_4}{O_2 O_3}, \quad (5.1)$$

and the log odds ratio, $\gamma = \log[(P_1 P_4)/(P_2 P_3)]$, can be estimated by $\hat{\gamma} = \log(\hat{r})$. Based on the i.i.d. assumption, the counts in the 2×2 table can be considered as a random sample from the multinomial $(n; \mathbf{P})$ distribution. As a result, the asymptotic variance of $\hat{\gamma}$ is $n^{-1} \sum_l 1/P_l$.

		Offspring	
		1	0
Parent	1	O_1	O_2
	0	O_3	O_4

Table 5.1: Contingency table formed by independent parent-offspring pairs

In NF1 association study, the sample units are not independent relative pairs. For the interclass odds ratio, the data contain families with one NF1 parent and at least one NF1 offspring. The odds ratio was estimated based on a contingency table formed by all possible parent-offspring pairs. For the intraclass odds ratio, the data contain families with at least two NF1 siblings. The odds ratio was estimated based on a contingency table formed by all possible sibling pairs.

In this case, pairs in the contingency table are not independent. Firstly, family members are dependent if there exists a familial association; secondly, the same individual appears in more than one pair when the family size is larger than two. Assuming exchangeability within a family and closure of multivariate binary distribution under margins, O_l/N , where N is the total number of pairs, is an unbiased estimate of P_l , $l = 1$ to 4. Therefore, the cross-product ratio of the table is still a valid point estimate of the odds ratio. However, the dependence among pairs will affect the variance of the estimators and this has to be taken into consideration.

The idea of forming a 2×2 contingency table by all possible relative pairs has been used by Hunt et al. (1988) to test familial aggregation for families of a fixed size. (Familial aggregation is a terminology commonly used in the literature referring to the resemblance among family members.) The test is based on the standard χ^2 statistic generated from the contingency table. He showed that

the standard χ^2 statistic is appropriate for intraclass aggregation despite the dependence among the pairs. For interclass aggregation, an adjustment based on the affected rate and the intraclass aggregation is needed. However, Hunt did not consider to estimate of the intraclass/interclass odds ratio and provide an appropriate standard error for the estimate.

In this chapter, we derive the asymptotic variance of the estimators based on such a contingency table and then compare their efficiency with the MLE from a parametric model. The rest of this chapter is organized as follows. In Section 5.1, we first derive the asymptotic variance of the estimator of the interclass log odds ratio and write it in a form that allows an interpretation of its magnitude as a function of the degree of association among siblings conditional on their parents. The method is derived for both Type-4 families, i.e., families containing two parents and multiple offspring, and Type-3 families, i.e., families containing one parent and multiple offspring. The latter is discussed as a special case. Then we propose estimation of the asymptotic variance of the estimator followed by an efficiency comparison of the proposed estimator with the MLE generated from a parametric model. We find that except for very strong dependence, the efficiency of our estimator is generally above 0.85. In Section 5.2, we consider estimating the intraclass log odds ratio based on all sibling pairs. Since the simple estimator tends to emphasize large families disproportionately, weights are considered to improve the efficiency. We then propose a method to evaluate the asymptotic variance and compare the efficiencies of the unweighted and weighted estimators with the MLE based on a parametric model. Two applications of these methods will be given later in Section 6.1.

5.1 Interclass Odds Ratio

Assume a random sample of n families. Let k_{i1} ($= 1$ or 2) be the number of parents and k_{i2} be the number of offspring in the i th family. Let $\mathbf{X}_i = (X_{ij_1}, j_1 = 1, \dots, k_{i1})$ be the vector of indicators for the parents in family i and $\mathbf{Y}_i = (Y_{ij_2}, j_2 = 1, \dots, k_{i2})$ be the vector corresponding to the offspring.

Within a family, the parents and the offspring are considered as two classes. We assume that the individuals belonging to the same class are exchangeable. We also assume closure under margins. That is, there is a probability distribution, $p(\mathbf{x}, \mathbf{y}; k_1, k_2)$ closed under margins, such that

$$\Pr(\mathbf{X}_i = \mathbf{x}, \mathbf{Y}_i = \mathbf{y}) = \Pr(\mathbf{X}_i = \mathbf{x}^*, \mathbf{Y}_i = \mathbf{y}^*) = p(\mathbf{x}, \mathbf{y}; k_{i1}, k_{i2})$$

where \mathbf{x}^* and \mathbf{y}^* are arbitrary permutations of \mathbf{x} and \mathbf{y} , respectively. This guarantees a common intraclass odds ratio for different families:

$$\gamma_{po} = \frac{P_1 P_4}{P_2 P_3} = \frac{p(1, 1; 1, 1)p(0, 0; 1, 1)}{p(1, 0; 1, 1)p(0, 1; 1, 1)}.$$

5.1.1 Estimator of γ_{po}

We pair each parent with each of his/her offspring and then form a contingency table based on all the pairs as in Table 5.1. We estimate the interclass odds ratio using the cross-product ratio of the counts O_1 , O_2 , O_3 and O_4 . The estimator is denoted by \hat{r}_{po} . The log odds ratio is estimated by $\hat{\gamma}_{po} = \log \hat{r}_{po}$.

Since the probability that a pair falls in cell l , $l = 1, \dots, 4$, is P_l for all the pairs, the expected value of O_l is $E(O_l) = NP_l$, where N is the total number of parent-offspring pairs. Therefore O_l/N is an unbiased estimate of P_l . As a result, \hat{r}_{po} is asymptotically unbiased.

5.1.2 Asymptotic Variance of $\hat{\gamma}_{po}$

General

Since $\hat{\gamma}_{po}$ is a function of $\mathbf{O} = (O_1, O_2, O_3, O_4)'$, we first obtain $\text{Var}(\mathbf{O})$, then use the Delta method to obtain the asymptotic variance of $\hat{\gamma}_{po}$. Clearly the distribution of \mathbf{O} is no longer multinomial since the pairs contributed by the same family are not independent due to the intraclass dependence and repeated use of the same individual in multiple pairs.

Let $\mathbf{O}_i = (O_{il}, l = 1, \dots, 4)$, be the counts in the contingency table contributed by the i th family. The number of parent-offspring pairs formed by the i th family is $N_i = k_{i1}k_{i2}$. Let $I_{ij}^{(l)}$ be the indicator that the j th pair contributed by the i th family falls in cell l . Then $O_{il} = \sum_{j=1}^{N_i} I_{ij}^{(l)}$. The variance of O_{il} is

$$\begin{aligned} \text{Var}(O_{il}) &= \sum_{j=1}^{N_i} \text{Var}(I_{ij}^{(l)}) + \sum_{j' \neq j} \text{Cov}(I_{ij}^{(l)}, I_{ij'}^{(l)}) \\ &= N_i P_l (1 - P_l) + b_i (P_{ll}^{(i)} - P_l^2), \end{aligned}$$

where $b_i = N_i(N_i - 1)$ and $P_{ll}^{(i)}$ is the probability that pairs j and j' both fall in cell l . The

covariance between O_{il} and O_{im} , $l \neq m$, is

$$\begin{aligned}
\text{Cov}(O_{il}, O_{im}) &= \text{Cov}\left(\sum_{j=1}^{N_i} I_{ij}^{(l)}, \sum_{j'=1}^{N_i} I_{ij'}^{(m)}\right) \\
&= \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} \text{Cov}(I_{ij}^{(l)}, I_{ij'}^{(m)}) \\
&= \sum_{j=1}^{N_i} \text{Cov}(I_{ij}^{(l)}, I_{ij}^{(m)}) + \sum_{j' \neq j} \text{Cov}(I_{ij}^{(l)}, I_{ij'}^{(m)}) \\
&= -N_i P_l P_m + b_i (P_{lm}^{(i)} - P_l P_m),
\end{aligned}$$

where $P_{lm}^{(i)}$ is the probability that pairs j and j' fall in cells l and m respectively. Since the two pairs are symmetric, $P_{lm}^{(i)} = P_{ml}^{(i)}$.

Next we derive $P_{lm}^{(i)}$, $l, m = 1, \dots, 4$. If we randomly choose two distinct pairs from all the possible pairs within a family, there are three types of combinations:

type I: two pairs containing the same parent and different offspring. This occurs with probability

$$a_I^{(i)} = \frac{k_{i1} k_{i2} (k_{i2} - 1)}{N_i (N_i - 1)} = \frac{k_{i2} - 1}{N_i - 1};$$

type II: two pairs containing the same offspring and different parents. This occurs with probability

$$a_{II}^{(i)} = \frac{k_{i2} k_{i1} (k_{i1} - 1)}{N_i (N_i - 1)} = \frac{k_{i1} - 1}{N_i - 1} = \begin{cases} 0 & k_{i1} = 1, \\ 1/(2k_{i2} - 1) & k_{i1} = 2; \end{cases}$$

type III: two pairs containing no common individuals. This occurs with probability

$$a_{III}^{(i)} = 1 - a_I^{(i)} - a_{II}^{(i)}.$$

For families with one parent, $a_I^{(i)} = 1$, $a_{II}^{(i)} = a_{III}^{(i)} = 0$; for families with two parents, $a_I^{(i)} = a_{III}^{(i)} = (k_{i2} - 1)/(2k_{i2} - 1)$, $a_{II}^{(i)} = 1/(2k_{i2} - 1)$.

Let $P_{lm|d}$ be the probability that two distinct pairs fall in cells l and m given that the pair set is of type d , $d = \text{I, II or III}$. Then

$$P_{lm}^{(i)} = \sum_d a_d^{(i)} P_{lm|d}.$$

The variance and covariance of \mathbf{O} can be easily obtained based on the above results for \mathbf{O}_i . Since $O_l = \sum_i O_{il}$ and families are independent, we have

$$\text{Var}(O_l) = NP_l(1 - P_l) + \sum_{i=1}^n b_i(P_{il}^{(i)} - P_l^2) \quad (5.2)$$

$$\text{Cov}(O_l, O_m) = -NP_lP_m + \sum_{i=1}^n b_i(P_{lm}^{(i)} - P_lP_m), \quad \text{for } l \neq m. \quad (5.3)$$

As

$$\hat{\gamma}_{po} = \log \frac{\hat{P}_1 \hat{P}_4}{\hat{P}_2 \hat{P}_3}, \quad (5.4)$$

where $\hat{P}_l = O_l/N$, by the Delta method, the asymptotic variance of $\hat{\gamma}$ is given by:

$$\sigma^2 = \frac{\partial \gamma_{po}}{\partial \mathbf{P}'} \text{Cov}(\hat{\mathbf{P}}) \frac{\partial \gamma_{po}}{\partial \mathbf{P}} = \frac{1}{N^2} \frac{\partial \gamma_{po}}{\partial \mathbf{P}'} \text{Cov}(\mathbf{O}) \frac{\partial \gamma_{po}}{\partial \mathbf{P}},$$

since $\text{Cov}(\hat{\mathbf{P}}) = \text{Cov}(\mathbf{O})/N^2$. As

$$\frac{\partial \gamma_{po}}{\partial \mathbf{P}'} = \left(\frac{1}{P_1}, -\frac{1}{P_2}, -\frac{1}{P_3}, \frac{1}{P_4} \right)',$$

algebraic simplification leads to the following result:

Result 5.1 *The asymptotic variance of the interclass log odds ratio estimator given in (5.4) is*

$$\begin{aligned} \sigma^2 &= \frac{1}{N} \sum_{l=1}^4 \frac{1}{P_l} + \frac{1}{N^2} \sum_i b_i \left(\left[\sum_{l=1}^4 \frac{P_{il}^{(i)}}{P_l^2} \right] + \frac{2P_{14}^{(i)}}{P_1P_4} + \frac{2P_{23}^{(i)}}{P_2P_3} - \frac{2P_{12}^{(i)}}{P_1P_2} - \frac{2P_{13}^{(i)}}{P_1P_3} - \frac{2P_{24}^{(i)}}{P_2P_4} - \frac{2P_{34}^{(i)}}{P_3P_4} \right) \\ &= \frac{1}{N} \sigma_0^2 + \frac{1}{N} \sum_d B_d v_d, \end{aligned} \quad (5.5)$$

where $\sigma_0^2 = \sum_{l=1}^4 P_l^{-1}$, $B_d = \sum_i b_i a_d^{(i)}/N$ and

$$v_d = \left[\sum_{l=1}^4 \frac{P_{ll|d}}{P_l^2} \right] + \frac{2P_{14|d}}{P_1P_4} + \frac{2P_{23|d}}{P_2P_3} - \frac{2P_{12|d}}{P_1P_2} - \frac{2P_{13|d}}{P_1P_3} - \frac{2P_{24|d}}{P_2P_4} - \frac{2P_{34|d}}{P_3P_4}, \quad (5.6)$$

$d = I, II$ and III .

In the above result, the term σ_0^2/N is the asymptotic variance given by N independent pairs.

Next, we show that there is an easy interpretation of v_I and v_{II} . When $d = I$, the same parent is in both pairs. Therefore, $P_{lm|I} = 0$ when $l = 1, 2$ and $m = 3, 4$ or when $l = 3, 4$ and $m = 1, 2$. Hence

$$v_I = \sum_{l=1}^4 \frac{P_{ll|I}}{P_l^2} - 2 \frac{P_{12|I}}{P_1P_2} - 2 \frac{P_{34|I}}{P_3P_4}. \quad (5.7)$$

Let

$$c_l = \frac{P_{ll|I}}{P_l} - \frac{P_{ll'|I}}{P_{l'}}, \quad l' = \begin{cases} 1 & \text{if } l = 2, \\ 2 & \text{if } l = 1, \\ 3 & \text{if } l = 4, \\ 4 & \text{if } l = 3. \end{cases} \quad (5.8)$$

Then v_I can be written as $\sum_{l=1}^4 c_l / P_l$.

Let X be the indicator of a parent, and Y_1, Y_2 be those of two offspring. We write $p_{y_i|x} = \Pr(Y_i = y_i | X = x)$ and $p_{y_1 y_2 | x} = \Pr(Y_1 = y_1, Y_2 = y_2 | X = x)$, for $x, y_i = 0$ or 1 . Then

$$\begin{aligned} \frac{P_{11|I}}{P_1} &= \frac{\Pr(X = 1, Y_1 = 1, Y_2 = 1)}{\Pr(X = 1, Y_2 = 1)} = \frac{p_{11|1}}{p_{1|1}}, \\ \frac{P_{12|I}}{P_2} &= \frac{\Pr(X = 1, Y_1 = 1, Y_2 = 0)}{\Pr(X = 1, Y_2 = 0)} = \frac{p_{10|1}}{p_{0|1}} = \frac{p_{1|1} - p_{11|1}}{1 - p_{1|1}}. \end{aligned}$$

It follows that

$$c_1 = \frac{p_{11|1} - p_{1|1}^2}{p_{1|1}(1 - p_{1|1})},$$

which is the conditional correlation coefficient of Y_1 and Y_2 given $X = 1$. It can be shown that $c_1 = c_2$. By symmetry, $c_3 = c_4$ is the conditional correlation coefficient of Y_1 and Y_2 given $X = 0$.

This indicates that the magnitude of v_I depends on the degree of conditional association between the siblings given the status of one of the parents. In general

$$v_I \begin{cases} < 0 & \text{if the siblings are conditionally negatively dependent,} \\ = 0 & \text{if the siblings are conditionally independent,} \\ > 0 & \text{if the siblings are conditionally positively dependent.} \end{cases}$$

Since $|c_l| \leq 1$, $|v_I| \leq \sigma_0^2$. Moreover, v_I can be simplified as

$$v_I = c_1 \left(\frac{1}{P_1} + \frac{1}{P_2} \right) + c_4 \left(\frac{1}{P_3} + \frac{1}{P_4} \right).$$

Likewise, v_{II} depends on the degree of conditional correlation between the two parents given the value of one of their offspring and can be rewritten as

$$v_{II} = c'_1 \left(\frac{1}{P_1} + \frac{1}{P_3} \right) + c'_4 \left(\frac{1}{P_2} + \frac{1}{P_4} \right)$$

where $c'_1 = P_{11|II}/P_1 - P_{13|II}/P_3$ and $c'_4 = P_{44|II}/P_4 - P_{24|II}/P_2$.

Unfortunately, there is no such intuitive interpretation for v_{III} .

For a special case of no interclass or intraclass dependence, $v_I = 0$ and $v_{II} = 0$ since the conditional correlations, c_1 , c_4 , c'_1 and c'_4 , are equal to 0. Under independence, $P_{lm|d} = P_l P_m$, and therefore $v_{III} = 0$ as well. This leads to $\sigma^2 = \sigma_0^2/N$.

Type-3 Family

If the sample contains only Type-3 families, that is $k_{i1} = 1$ for all i , $a_{II}^{(i)} = a_{III}^{(i)} = 0$. Then we have the following result:

Result 5.2 When $k_{i1} = 1$ for all i , the asymptotic variance of $\hat{\gamma}_{po}$ is

$$\sigma^2 = (\sigma_0^2 + B_I v_I)/N,$$

with $N = \sum_{i=1}^n k_{i2}$ and $B_I = \sum_{i=1}^n k_{i2}(k_{i2} - 1)/N$.

Particularly, when the family size is constant, i.e., $k_{i2} = k$, $\sigma^2 = (\sigma_0^2 + (k - 1)v_I)/N$. When family size varies, B_I can be written in terms of the sample mean (M) and variance ($V = \sum(k_{i2} - M)^2/n$) of k_{i2} : $B_I = V/M + (M - 1)$.

In general, the variance of $\hat{\gamma}_{po}$ increases when either c_1 or c_4 becomes larger. Now we consider two cases: (1) the offspring are independent conditional on the status of the parent, and (2) the offspring are perfectly dependent conditional on the status of the parent. In the first case, $c_1 = c_4 = 0$, and $v_I = 0$. Hence $\sigma^2 = \sigma_0^2/N$. Therefore, when conditional independence occurs, the resulting asymptotic variance is the same as having N independent pairs. In the second case, $c_1 = c_4 = 1$ and $\sigma^2 = (B_I + 1)\sigma_0^2/N$ since $v_I = \sigma_0^2$ in this case. So when perfect dependence occurs, if the family size is constant,

$$\sigma^2 = k\sigma_0^2/N = \sigma_0^2/n.$$

As would be expected in this case, it is the variance achieved considering only one offspring from each family. If the family size is not a constant,

$$\frac{B_I + 1}{N} > \frac{1}{n}.$$

The estimate is less efficient than considering only one offspring in each family.

5.1.3 Methods to Estimate the Asymptotic Variance of $\hat{\gamma}_{po}$ in (5.5)

First consider data with Type-4 families. To calculate the asymptotic variances of $\hat{\gamma}_{po}$, we need to know $P_{lm|d}$, which can be derived from the joint probabilities of two parents and two offspring. A natural way to estimate these joint probabilities is to use all the possible parent-offspring sets containing two parents and two offsprings. The sample proportion of counts corresponding to one outcome can be used as the estimate of the probability of that outcome.

If the dataset is formed by Type-3 families, instead of estimating the $P_{lm|d}$, we estimate the conditional correlations c_1 and c_4 in (5.8). In this case, the ANOVA estimator for unbalanced data can be borrowed for our use. To estimate c_1 , we construct an ANOVA table by all siblings whose parent is affected. Suppose there are n_1 families with affected parent and a total number of N_1 siblings from these families. Let $\bar{Y}_i = \sum_j Y_{ij}/k_i$ for $i \in I_1$ and $\bar{\bar{Y}}_1 = N_1^{-1} \sum_{i \in I_1} \sum_j Y_{ij}$, where I_1 is the set of indices of families with affected parent. The sum of squares within and between families are:

$$SSE = \sum_{i \in I_1} \sum_j (Y_{ij} - \bar{Y}_i)^2$$

$$SSB = \sum_{i \in I_1} k_i (\bar{Y}_i - \bar{\bar{Y}}_1)^2.$$

The corresponding mean squares are $MSE = SSE/(N_1 - n_1)$ and $MSB = SSB/(n_1 - 1)$. From the expected sum of squares (Searle et al., 1992),

$$\hat{\sigma}_B^2 = \frac{(n_1 - 1)(MSB - MSE)}{\sum_{i \in I_1} k_{i2}(1 - k_{i2}/N_1)}$$

is an unbiased estimate of $\text{Cov}(Y_{ij}, Y_{ij'})$, for $i \in I_1$, and $\hat{\sigma}_B^2 + MSE$ is an unbiased estimate of $\text{Var}(Y_{ij})$ for $i \in I_1$. The above results only depend on the moments of the distribution. Then an estimate of c_1 is given by

$$\hat{c}_1 = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_B^2 + MSE}.$$

In a similar way, c_4 can be estimated based the ANOVA table formed by all siblings whose parent is unaffected. The ANOVA approach can be inefficient when c_1 and c_4 is close to 1. The combination estimator proposed by Srivastava (1993) may be considered.

An alternative approach to estimate the variance of $\hat{\gamma}_{po}$ is to use the jackknife method with one family or one block of families removed at a time.

5.1.4 Asymptotic Relative Efficiency of $\hat{\gamma}_{po}$

Asymptotic Variance of the MLE

To compare the efficiency based on a parametric model, we need the asymptotic variance of the ML estimate of γ_{po} . Let $p(\mathbf{x}, \mathbf{y}; k_1, k_2; \boldsymbol{\theta})$ be a multivariate binary distribution with k_1, k_2 being the dimension of the two classes. We assume that it is exchangeable in \mathbf{x} and in \mathbf{y} , and is closed under margins. Then

$$\gamma_{po}(\boldsymbol{\theta}) = \log p(1, 1; 1, 1; \boldsymbol{\theta}) + \log p(0, 0; 1, 1; \boldsymbol{\theta}) - \log p(1, 0; 1, 1; \boldsymbol{\theta}) - \log p(0, 1; 1, 1; \boldsymbol{\theta})$$

Let $\hat{\boldsymbol{\theta}}$ be the MLE and $V_{\hat{\boldsymbol{\theta}}}$ be the asymptotic covariance matrix of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, i.e., the inverse of Fisher's information \mathbf{F} . Then the asymptotic variance of $n^{1/2}[\gamma(\hat{\boldsymbol{\theta}}) - \gamma(\boldsymbol{\theta})]$ is

$$\frac{\partial \gamma}{\partial \boldsymbol{\theta}'} V_{\hat{\boldsymbol{\theta}}} \frac{\partial \gamma}{\partial \boldsymbol{\theta}}.$$

If the proportion of families with sizes k_1, k_2 is (asymptotically) q_{k_1, k_2} for $k_1 = 1, 2, k_2 = 1, \dots, M$, then Fisher's information matrix is given by

$$\mathbf{F} = \sum_{k_1=1}^2 \sum_{k_2=1}^M q_{k_1, k_2} \sum_{\mathbf{x} \in \{0,1\}^{k_1}} \sum_{\mathbf{y} \in \{0,1\}^{k_2}} \frac{\partial p(\mathbf{x}, \mathbf{y}; k_1, k_2; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial p(\mathbf{x}, \mathbf{y}; k_1, k_2; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \bigg/ p(\mathbf{x}, \mathbf{y}; k_1, k_2; \boldsymbol{\theta}).$$

Results

We compare the efficiency with specific computations from the multivariate binary beta-binomial and multivariate probit models for Type-3 families. In our comparisons, we set the number of siblings in a family varying from 1 to 4 with the same frequency for each size. This is reasonable because most families contains no more than 4 offspring. If we change the maximum family size, it will not change the general patterns addressed below.

- (a) The relative efficiency is typically above 0.85.
- (b) The relative efficiency decreases as the dependence becomes strong.
- (c) The relative efficiency tends to decrease as the variance-mean ratio of family size increases.

We first show the comparison based on the multivariate binary beta-binomial distribution. Suppose the probability of the presence of the binary trait among parents is π , and that given the

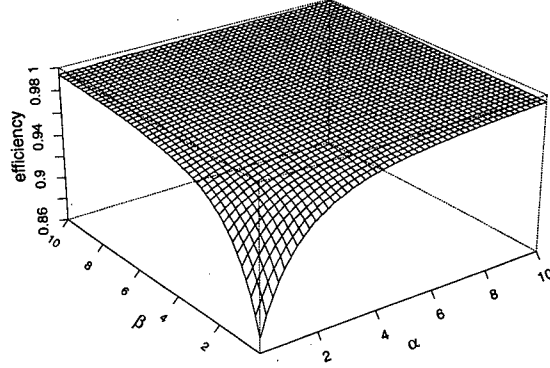


Figure 5.1: ARE of $\hat{\gamma}_{po}$ based on a multivariate binary beta-binomial model when $\alpha_1 = \beta_0 = \alpha, \alpha_0 = \beta_1 = \beta$

parent's state, $X_i = I$, $I = 0$ or 1 , (the subscript j_1 is omitted since there is only one parent per family), the joint conditional distribution of \mathbf{Y}_i is a multivariate binary beta-binomial distribution with parameters (α_I, β_I) . The conditional pmf of \mathbf{Y} is

$$\Pr(\mathbf{Y} = \mathbf{y} | X_i = I) = \frac{B(\alpha_I + y_{i+}, \beta_I + k - y_{i+})}{B(\alpha_I, \beta_I)},$$

where $B(\cdot, \cdot)$ is the Beta function and $y_{i+} = \sum y_j$. The interclass odds ratio is $(\alpha_1 \beta_0) / (\alpha_0 \beta_1)$.

In general, the efficiency depends on π , and the conditional correlation between offspring, which is given by $1/(\alpha_I + \beta_I + 1)$. As a simple case, we set $\alpha_1 = \beta_0 = \alpha$ and $\alpha_0 = \beta_1 = \beta$ so that the conditional correlation among siblings given $X_i = 1$ is the same as that given $X_i = 0$, and is equal to $1/(\alpha + \beta + 1)$. In this case, the efficiency is not affected by the value of π by the symmetry of the conditional distribution. For this reason we set $\pi = 0.5$ for the calculation.

The efficiency of $\hat{\gamma}_{po}$ is plotted in Figure 5.1. When both α and β are small, i.e. the siblings are strongly correlated, the efficiency is relatively low. But even when the correlation is almost 1, it is still above 0.85.

We next show the comparison based on the multivariate Probit model. Let Z_{i0} be the latent variable of the parent and Z_{ij} , $j = 1, \dots, k_i$, be those of the offspring. Let $\mu_1 = EZ_{i0}$ and $\mu_2 = EZ_{ij}$, $j > 0$, and ρ_1, ρ_2 be the parent-offspring and sib-sib correlations between the latent variables.

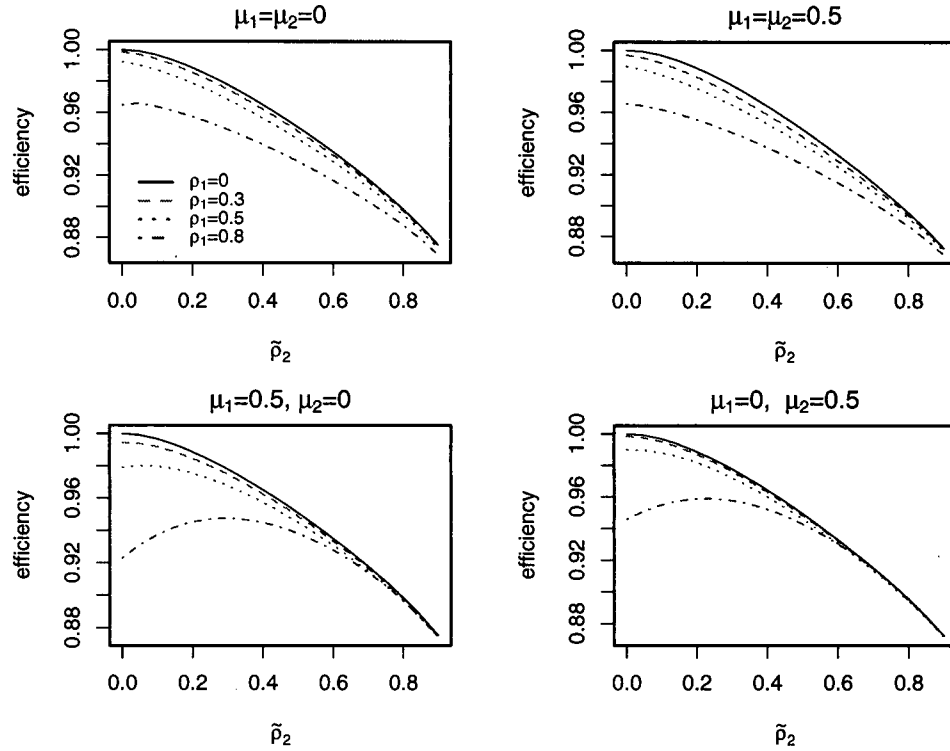


Figure 5.2: ARE of $\hat{\gamma}_{po}$ based on the multivariate Probit model.

We calculate the efficiency at different combinations of the means and different levels of ρ_1 and ρ_2 . The results are displayed in Figure 5.2. In order to make the results comparable at different levels of ρ_1 , the efficiency is plotted against the conditional sib-sib correlation $\tilde{\rho}_2$. The efficiencies based on this multivariate Probit model are similar to those based on the multivariate binary beta-binomial model.

If we change the distribution of family sizes, the efficiency of the estimator exhibits the same general patterns but with different magnitudes. We find that the efficiency tends to be lower when the variation of the family size is large compared with the mean family size. We illustrate this pattern by a comparison based on the multivariate binary beta-binomial model. The results are shown in Figure 5.3, in which each point corresponds to the efficiency associated with a particular distribution of family size. In all cases, the number of offspring in a family is set to be 1, 2, 3 or 4 but with a randomly generated frequency. The efficiencies are calculated based on a multivariate binary beta-binomial model with $\alpha_1 = 2$, $\beta_1 = 1$, $\alpha_0 = 1$, $\beta_0 = 2$, $\pi = 0.5$.

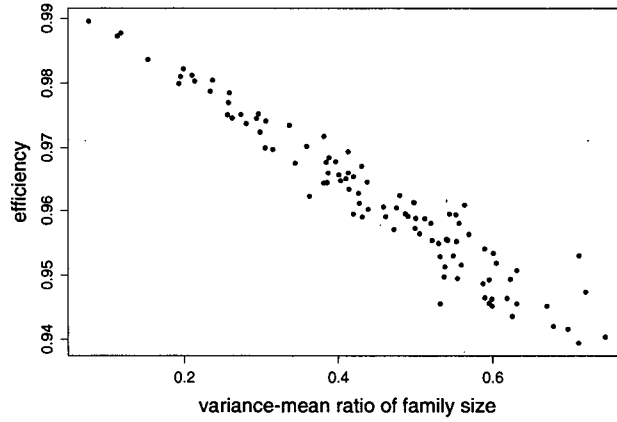


Figure 5.3: Efficiency of $\hat{\gamma}_{po}$ versus the variance-mean ratio (V/M) of family size

5.2 Intraclass Odds Ratio

Consider a random sample of n families, each containing a varying number of siblings. Let k_i be the number of siblings in the i th family and $\mathbf{Y}_i = (Y_{ij}, j = 1, \dots, k_i)$ be the indicator vector for the siblings.

We assume exchangeability and closure under margins, that is, there is a distribution $p(\mathbf{y}; k)$ closed under margins, such that for \mathbf{y} of dimension k_i ,

$$\Pr(\mathbf{Y}_i = \mathbf{y}) = \Pr(\mathbf{Y}_i = \mathbf{y}^*) = p(\mathbf{y}; k_i)$$

for \mathbf{y}^* being any arbitrary permutation of \mathbf{y} . From the closure under margins, the odds ratio is

$$r_{ss} = \frac{P_1 P_4}{P_2 P_3} = \frac{p(1, 1; 2)p(0, 0; 2)}{p(1, 0; 2)p(0, 1; 2)}$$

Due to the exchangeability, $P_2 = P_3 = (1 - P_1 - P_4)/2$. Therefore, the intraclass odds ratio is completely determined by P_1 and P_4 :

$$r_{ss} = \frac{4P_1 P_4}{(1 - P_1 - P_4)^2}.$$

5.2.1 The Estimator of γ_{ss}

For family i , there are $\binom{k_i}{2}$ possible sib-sib pairs. The indicators of a pair fall into three categories: both equal to 1, one equal to 1 and the other equal to 0, and both equal to 0. To form the 2×2

contingency table of pairs from all the families, let O_1 be the count of pairs both equal to 1 and O_4 be the count of those both equal to 0. As O_2 and O_3 are not distinguished, we set each of them equal to half of the counts for heterogeneous pairs; that is, $O_2 = O_3 = (N - O_1 - O_4)/2$, given the total number of pairs, N .

The estimate of the intraclass odds ratio from such a contingency table is

$$\hat{r}_{ss} = \frac{O_1 O_4}{O_2 O_3} = \frac{4O_1 O_4}{(N - O_1 - O_4)^2}. \quad (5.9)$$

The corresponding log odds ratio estimate is

$$\hat{\gamma}_{ss} = \log O_1 - 2 \log(N - O_1 - O_4) + \log O_4 + \log 4. \quad (5.10)$$

5.2.2 Asymptotic Variance of $\hat{\gamma}_{ss}$

With similar derivations as in the preceding section, the variance and covariance of (O_1, O_4) is given by

$$\text{Var}(O_l) = NP_l(1 - P_l) + \sum_{i=1}^n b_i(P_{li}^{(i)} - P_l^2), \quad \text{for } l = 1, 4, \quad (5.11)$$

$$\text{Cov}(O_1, O_4) = -NP_1 P_4 + \sum_{i=1}^n b_i(P_{14}^{(i)} - P_1 P_4), \quad (5.12)$$

with $N = \sum_{i=1}^n \binom{k_i}{2}$, $b_i = \binom{k_i}{2}(\binom{k_i}{2} - 1)$ and

$$P_{lm}^{(i)} = (1 - a_i)P_{lm|4} + a_i P_{lm|3}, \quad l, m = 1, 4,$$

where a_i is the probability that two random pairs from the $\binom{k_i}{2}$ possible pairs contain only three distinct individuals and $P_{ij|3}$ is the probability that two such pairs fall in the cell l and m respectively, while $P_{ij|4}$ is the same probability for two random pairs containing four individuals. It can be shown that

$$a_i = \frac{k_i(k_i - 1)(k_i - 2)/2}{\binom{k_i}{2}(\binom{k_i}{2} - 1)/2} = \frac{4}{k_i + 1}, \quad k_i > 2$$

(Hunt et al., 1988). As $P_{14|3} = 0$, we obtain $P_{14}^{(i)} = (1 - a_i)P_{14|4}$.

With substitution for $P_{lm}^{(i)}$, (5.11) and (5.12) become

$$\text{Var}(O_l) = NP_l(1 - P_l) + \sum_{i=1}^n (b_i a_i P_{l|3} + b_i(1 - a_i)P_{l|4} - b_i P_l^2), \quad \text{for } l = 1, 4, \quad (5.13)$$

$$\text{Cov}(O_1, O_4) = -NP_1 P_4 + \sum_{i=1}^n (b_i(1 - a_i)P_{14|4} - b_i P_1 P_4). \quad (5.14)$$

Let

$$B_1 = \sum_i b_i a_i / N, \quad B_2 = \sum_i b_i(1 - a_i) / N \quad \text{and} \quad B_3 = \sum_i b_i / N; \quad (5.15)$$

$$\mathbf{V}_0 = \begin{pmatrix} P_1 & 0 \\ 0 & P_4 \end{pmatrix} - \begin{pmatrix} P_1^2 & P_1 P_4 \\ P_1 P_4 & P_4^2 \end{pmatrix}, \quad \mathbf{V}_1 = \begin{pmatrix} P_{11|3} & 0 \\ 0 & P_{44|3} \end{pmatrix},$$

$$\mathbf{V}_2 = \begin{pmatrix} P_{11|4} & P_{14|4} \\ P_{14|4} & P_{44|4} \end{pmatrix}, \quad \mathbf{V}_3 = - \begin{pmatrix} P_1^2 & P_1 P_4 \\ P_1 P_4 & P_4^2 \end{pmatrix}.$$

The covariance matrix of (O_1, O_4) , denoted by \mathbf{V} , can be written as:

$$\mathbf{V} = N\mathbf{V}_0 + N \sum_{d=1}^3 B_d \mathbf{V}_d.$$

Let

$$\phi = \frac{1}{N} \left(\frac{\partial \gamma_{ss}}{\partial P_1}, \frac{\partial \gamma_{ss}}{\partial P_4} \right)',$$

where

$$\frac{\partial \gamma_{ss}}{\partial P_1} = \frac{1}{P_1} + \frac{1}{P_2} \quad \text{and} \quad \frac{\partial \gamma_{ss}}{\partial P_4} = \frac{1}{P_4} + \frac{1}{P_2}.$$

The asymptotic variance of $\hat{\gamma}_{ss}$ given by the Delta method is:

$$\sigma^2 = N\phi' \mathbf{V}_0 \phi + N \sum_{d=1}^3 B_d \phi' \mathbf{V}_d \phi.$$

If we write

$$\begin{aligned}
\sigma_0^2 &= N^2 \phi' \mathbf{V}_0 \phi = \frac{1}{P_1} + \frac{1}{P_4} + \frac{2}{P_2}, \\
v_1 &= N^2 \phi' \mathbf{V}_1 \phi = P_{11|3} \left(\frac{1}{P_1} + \frac{1}{P_2} \right)^2 + P_{44|3} \left(\frac{1}{P_2} + \frac{1}{P_4} \right)^2, \\
v_2 &= N^2 \phi' \mathbf{V}_2 \phi = \sum_{l,m \in \{1,4\}} P_{lm|4} \left(\frac{1}{P_l} + \frac{1}{P_2} \right) \left(\frac{1}{P_m} + \frac{1}{P_2} \right), \\
v_3 &= N^2 \phi' \mathbf{V}_3 \phi = -\frac{1}{P_2^2},
\end{aligned} \tag{5.16}$$

then we have the following result.

Result 5.3 *The asymptotic variance of the intraclass log odds ratio estimator given in (5.10) is*

$$\sigma^2 = \frac{1}{N} \sigma_0^2 + \frac{1}{N} \sum_{d=1}^3 B_d v_d, \tag{5.17}$$

with B_d defined in (5.15) and v_d defined in (5.16).

If each family contains the same number of siblings, say k , then $N = n \binom{k}{2}$ and

$$B_1 = 2(k-2), \quad B_2 = \frac{(k-2)(k-3)}{2}, \quad B_3 = \frac{(k+1)(k-2)}{2}, \tag{5.18}$$

from which it follows that

$$\sigma^2 = \frac{2}{nk(k-1)} \left[\sigma_0^2 + \frac{k-2}{2} (4v_1 + (k-3)v_2 + (k+1)v_3) \right].$$

For the special case where the siblings are mutually independent, let $p = \Pr(Y_{ij} = 1)$ and $q = 1 - p$. For this case, $P_1 = p^2$, $P_4 = q^2$, $P_2 = pq$, $p_{14|4} = p^2 q^2$ and $P_{11|L} = p^L$, $P_{44|L} = q^L$, for $L = 3$ or 4 . Consequently,

$$\begin{aligned}
P_{11}^{(i)} &= (1 - a_i) p^4 + a_i p^3 \\
P_{14}^{(i)} &= (1 - a_i) p^2 q^2 \\
P_{44}^{(i)} &= (1 - a_i) q^4 + a_i q^3.
\end{aligned}$$

Substituting these probabilities into (5.17), we find $\sum_{d=1}^3 B_d v_d = 0$. Hence $\sigma^2 = \sigma_0^2/N$, which is the same variance when the sample contains N independent pairs. Surprisingly, although the pairs are not independent due to the inclusion of the same individuals in multiple pairs, the asymptotic variance of $\hat{\gamma}_{ss}$ is not inflated.

Generally speaking, when siblings are positively dependent, $\sigma^2 > \sigma_0^2$. When there are large families in the sample, σ^2 may be substantially larger than σ_0^2 .

5.2.3 Weighted Estimator of γ_{ss}

In the contingency table, a large family contributes many more pairs than a small family. For example, the contribution from a family of six is equivalent to that from 5 families of three or 15 families of two. Since the number of pairs is disproportional to the family size, the large families might be over-emphasized. With this concern, we consider assigning weights to each family based on the size of the family.

Let n_k be the number of families of size k and $\mathbf{O}^{(k)} = (O_1^{(k)}, O_4^{(k)})$ be the counts from these families. Consider a weight w_k for $\mathbf{O}^{(k)}$. Then the weighted total counts are $\mathbf{O}_w = \sum_k w_k \mathbf{O}^{(k)}$. Let $\hat{\gamma}_{w,ss}$ be the weighted estimate of the log odds ratio calculated from \mathbf{O}_w , that is

$$\hat{\gamma}_{w,ss} = \log \frac{4O_{w1}O_{w4}}{(N' - O_{w1} - P_{w4})^2}, \quad (5.19)$$

where $N' = \sum_k w_k n_k \binom{k}{2}$. Next we derive its asymptotic variance.

Let $\mathbf{V}^{(k)} = \text{Var}(\mathbf{O}^{(k)})/n_k \binom{k}{2}$. Then the covariance matrix of \mathbf{O}_w is

$$\mathbf{V}_w = \sum_k w_k^2 n_k \binom{k}{2} \mathbf{V}^{(k)}.$$

Let

$$\phi_w = \frac{1}{N'} \left(\frac{1}{P_1} + \frac{1}{P_2}, \frac{1}{P_4} + \frac{1}{P_2} \right).$$

If we write $\tau_k = (N')^2 \phi_w' \mathbf{V}^{(k)} \phi_w$, then the asymptotic variance of $\hat{\gamma}_{w,ss}$ is

$$\sigma_w^2 = \phi_w' \mathbf{V}_w \phi_w = \frac{1}{(N')^2} \sum_k w_k^2 n_k \binom{k}{2} \tau_k.$$

It can be shown that $\tau_k = \sigma_0^2 + \sum_{d=1}^3 B_d^{(k)} v_d$, where the $B_d^{(k)}$ are given by (5.18). In particular, $\tau_2 = \sigma_0^2$ since $B_d^{(2)} = 0$ for all d . Without loss of generality, let $w_2 = 1$, then

$$\sigma_w^2 = \frac{1}{(N')^2} \left(n_2 \tau_2 + \sum_{k \geq 3} w_k^2 n_k \binom{k}{2} \tau_k \right).$$

with $N' = n_2 + \sum_{k \geq 3} w_k n_k \binom{k}{2}$.

To find the optimal weights, we solve the following equations:

$$w_k \tau_k - N' \sigma_w^2 = 0, \quad \text{for } k \geq 3,$$

obtained by differentiating σ_w^2 with respect to the w_k s and then setting the derivatives equal to 0.

Result 5.4 *The optimal weights which minimize the asymptotic variance of $\hat{\gamma}_{w,ss}$ (5.19) are*

$$w_k = \tau_2 / \tau_k, \quad \text{for } k \geq 2.$$

The corresponding asymptotic variance is $\sigma_{w_{opt}}^2 = \sigma_0^2 / N'$.

The $\sigma_{w_{opt}}^2$ is derived as the following:

$$\begin{aligned} \sigma_{w_{opt}}^2 &= \frac{1}{(N')^2} \left[n_2 \tau_2 + \sum_{k \geq 3} \frac{\tau_2^2}{\tau_k} n_k \binom{k}{2} \right] \\ &= \frac{\tau_2}{N'} \left[\frac{n_2}{N'} + \frac{1}{N'} \sum_{k \geq 3} w_k n_k \binom{k}{2} \right] \\ &= \frac{\tau_2}{N'} = \frac{\sigma_0^2}{N'} \end{aligned}$$

by the definition of N' . Since the weight w_k is a function of τ_2 and τ_k , it depends on P_1 , P_4 and $P_{lm|L}$, for $l, m = 1, 4$ and $L = 3, 4$. In practice, these probabilities are unknown and need to be estimated. Therefore $\hat{\gamma}_w$ has to be obtained iteratively.

Instead of using the optimal weights, a simple approach to down-weight the large families is to set the weights such that $w_k N_k$, where N_k is the number of pairs contributed by a family of size k , is proportional to k , that is, $w_k = 2/(k-1)$.

Result 5.5 *When set $w_k = 2/(k-1)$, the asymptotic variance of $\hat{\gamma}_w$ is*

$$\sigma_{w_{simple}}^2 = \frac{1}{(N')^2} \sum_k \frac{n_k k}{k-1} [2\sigma_0^2 + (k-2)(4v_1 + (k-3)v_2 + (k+1)v_3)]. \quad (5.20)$$

with $N' = \sum_k n_k k$.

5.2.4 Methods to Estimate the Asymptotic Variance of $\hat{\gamma}_{ss}$

To estimate the asymptotic variance of $\hat{\gamma}_{ss}$, we need to estimate $P_{ij|3}$ and $P_{ij|4}$. A simple way to obtain the estimates is to use the sample proportions based on all possible sets of four siblings to estimate the $P_{ij|4}$ and use all the possible sets of three siblings to estimate the $P_{ij|3}$. However, if the dataset has limited number of families containing three or more siblings, this method will not be able to provide reliable estimates of the probabilities. In this case, the jackknife method can be used to estimate the variance of $\hat{\gamma}_{ss}$.

5.2.5 Efficiency Comparison

In this subsection, we investigate the efficiency of the weighted and unweighted estimators by comparing them with the MLE based on a parametric model.

Asymptotic Variance of the MLE

Consider a parametric exchangeable model $p(\mathbf{y}; k, \boldsymbol{\theta})$ which is closed under margins and defined for all $k \geq 2$. This means that the bivariate margin of Y_i, Y_j is $p(y_i, y_j; 2, \boldsymbol{\theta})$ for all $i \neq j$. The log odds ratio based on the parametric model is

$$\gamma(\boldsymbol{\theta}) = \log p(1, 1; 2, \boldsymbol{\theta}) + \log p(0, 0; 2, \boldsymbol{\theta}) - 2 \log p(1, 0; 2, \boldsymbol{\theta}).$$

Let $\hat{\boldsymbol{\theta}}$ be the MLE and $V_{\hat{\boldsymbol{\theta}}}$ be the asymptotic covariance matrix of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. Then the asymptotic variance of $n^{1/2}[\gamma(\hat{\boldsymbol{\theta}}) - \gamma(\boldsymbol{\theta})]$ is

$$\frac{\partial \gamma}{\partial \boldsymbol{\theta}'} V_{\hat{\boldsymbol{\theta}}} \frac{\partial \gamma}{\partial \boldsymbol{\theta}}.$$

The matrix $V_{\hat{\boldsymbol{\theta}}}$ is the inverse of Fisher's information \mathbf{F} . If the proportion of families with size k is (asymptotically) q_k for $k = 2, \dots, M$, then

$$\mathbf{F} = \sum_{k=2}^M q_k \sum_{\mathbf{y}_k \in \{0,1\}^k} \frac{\partial p(\mathbf{y}_k; k, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial p(\mathbf{y}_k; k, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} / p(\mathbf{y}_k; k, \boldsymbol{\theta}).$$

Results

We compare the efficiency with specific computations from the multivariate binary beta-binomial and multivariate probit models. In our comparisons, we set the number of siblings in a family varying from 2 to 5 with the same frequency for each size. The conclusions are the following.

- (a) The relative efficiency with the optimal weights is close to 1 (> 0.95).
- (b) The relative efficiency of the unweighted estimator decreases when the odds ratio increases and is not much affected by p ; the relative efficiency can get below 0.6.
- (c) The relative efficiency for the estimator with simple weights is lower when the odds ratio is low, but it is typically above 0.8.

For a multivariate binary beta-binomial model with parameters (α, β) . In this model the affected rate, p , is $\alpha/(\alpha + \beta)$ and the intraclass odds ratio is $(\alpha + 1)(\beta + 1)/(\alpha\beta)$.

For a multivariate Probit model, let Z_{ij} be the latent variable corresponding to Y_{ij} . Let μ be the mean of Z_{ij} , and $\rho = \text{Corr}(Z_{ij}, Z_{ij'})$ for $j \neq j'$. Then, the affected rate is $p = \Phi(\mu)$.

In our comparisons, we vary the degree of the intraclass dependence at four different affected rates (p): 0.05, 0.20, 0.35 and 0.50. When p is above 0.5, the results are same as that for $1 - p$. For each case, we calculate the efficiency under three situations: unweighted, weighted by optimal weights and weighted according to family size (referred to as simple weights).

In Figures 5.4 and 5.5, we plot the results for the multivariate binary beta-binomial model and the multivariate Probit model, respectively. The plots show that the unweighted estimator (solid line) loses efficiency when the dependence increases, whereas its efficiency does not seem much affected by p . The optimal weights (dotted line) greatly enhance the efficiency, particularly when the dependence is strong. When $p = 0.05$, the efficiency of the optimally weighted estimator is relatively low. In other cases, it is always above 0.95. Using the simple weights is worse than no weights when the odds ratio is low, however, the performance of this estimator improves when the dependence increases and is almost as good as the optimally weighted estimator for strong dependence. The plots also show that its efficiency climbs faster when p is closer to 0.5.

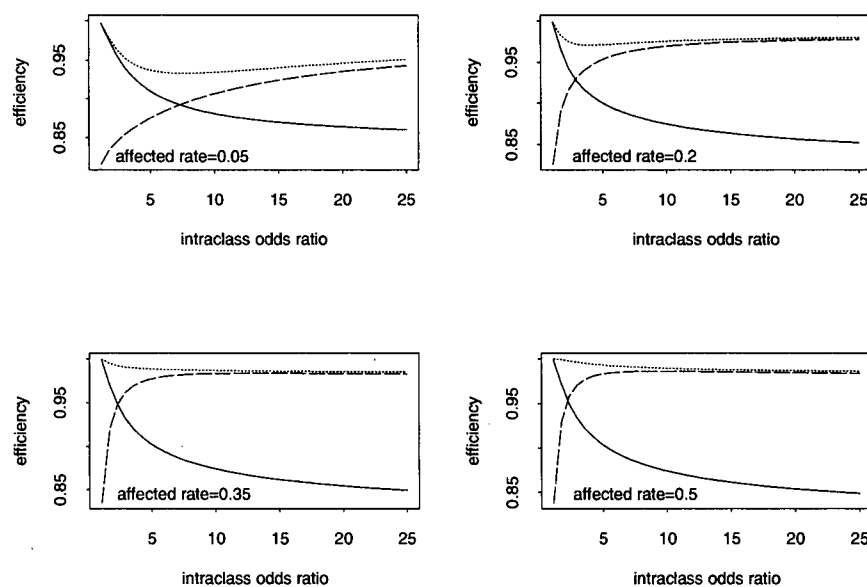


Figure 5.4: Efficiencies of γ_{ss} estimators (multivariate binary beta-binomial model). Solid line: no weights, dotted line: optimal weights, dashed line: simple weights.

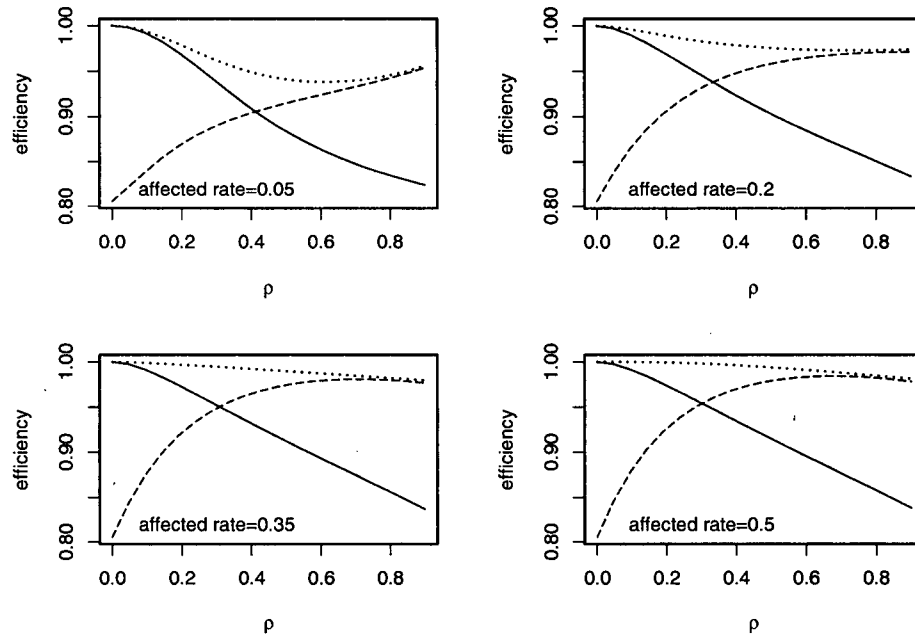


Figure 5.5: Efficiencies of γ_{ss} estimators (multivariate Probit model). Solid line: no weights, dotted line: optimal weights, dashed line: simple weights.

Next, we investigate how the distribution of family sizes affects the efficiency of the three estimators based on the multivariate binary beta-binomial model. In this case, we fix p at 0.5 and the intraclass odds ratio at a low level (1.44) and a high level (9). We only consider two different family sizes: 2 and 6. In Figure 5.6 the efficiencies are plotted against n_2/n , the proportion of families of size 2. At both odds ratio levels, the unweighted estimator is least efficient when $n_2/n = 15/16$, that is, when each type of family contributes equal number of pairs to the contingency table. The efficiency of the estimator with optimal weights increases monotonically with the proportion n_2/n . When the odds ratio is small, the estimator with simple weights is least efficient when $n_2/n = 6/7$, that is, each type of family contributes an equal number of individuals to the sample, whereas the efficiency of this estimator is almost coincident with that of the optimal weights when the odds ratio is high. Moreover, analogous to the interclass odds ratio, the efficiency of both the estimators, unweighted or weighted by family size, decreases almost linearly with the variance-mean ratio of the number of pairs contributed by each family.

Lastly we give some considerations on using weights. When the dependence among siblings

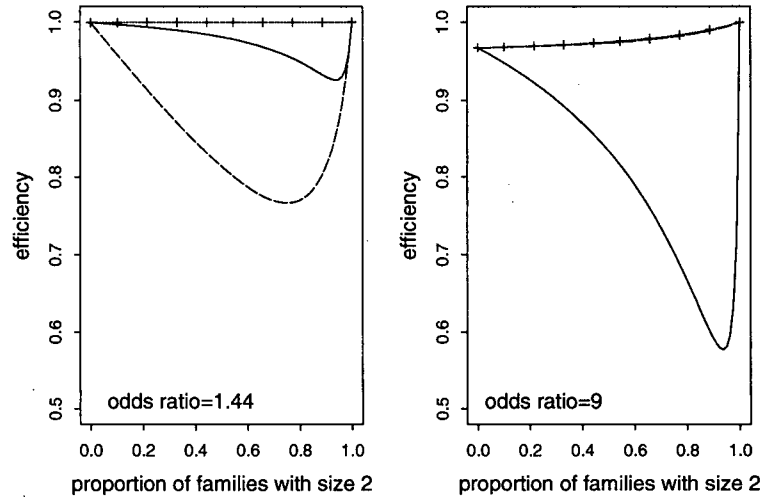


Figure 5.6: Efficiency comparison of γ_{ss} estimators with varied proportion of families with size 2 and 6. Solid line: no weights, dotted line with "+": optimal weights, dashed line: simple weights. In the figure on the right hand side, the dotted line with "+" and the dashed line are coincident.

is weak, the loss of efficiency is modest if the unweighted estimator is used and, therefore, there is no compelling reason to use any weights. On the contrary, when the dependence is very strong (odds ratio > 5 when the affected rate p is not too small) and there is a great deal of variation in family size, the simple weights are recommended since they work almost as well as the optimal weights and are straightforward to use.

5.3 Discussion

In this chapter, we proposed a simple method to estimate common odds ratios across families with mixed sizes. The method is directly extended from the odds ratio estimator based on a standard contingency table and is easy to carry out. The method can also be used in many other studies. For example, the interclass odds ratio estimator is useful for a family based case-control study. The methods are not limited to family studies either. They can be applied to clustered binary data.

Chapter 6

Applications to NF1 and NF2 Data

In this chapter, we apply the models and estimation methods studied in the previous chapters to some measurements on NF1 and NF2 patients. On the one hand, these measurements have a remarkable variation among patients. On the other hand, they exhibit a certain degree of similarity among relatives. In this chapter, we analyze the strength of association between different types of relatives. The measurements are treated as response variables in our analyses. These variables belong to different data types, and therefore require different models and methods of analysis. For some variables, adjustment for covariates, such as age and mutation type, is required.

Among the four response variables analyzed in this chapter, the first two are from NF1 patients and the data are from two large NF1 databases: the National Neurofibromatosis Foundation International Database and the Neurofibromatosis Institute Database; the second two are from NF2 patients and the data are from various sources. The data from these sources can not be considered as random samples from certain population. Our inferences are valid based on the assumption that these data are representative of the NF1/2 population.

The first variable is studied using Type-2 and Type-3 families. The other three variables are studied using Type-I pedigrees. The first variable is the presence of intertriginous freckling (freckling in intertriginous area, such as armpits, where skin comes into contact with itself) in NF1 patients. We estimated the interclass and intraclass odds ratio using the method presented in Chapter 5. The second variable is the presence of subcutaneous neurofibromas (tumors or growths located along a nerve or nervous tissue under the skin) in NF1 patients. A multivariate probit model and a two-component series model are compared in this example. The third variable is

age at first symptom of NF2, where right censoring is present. We consider a MVN model for this variable. The CL1, CL2 methods are compared with the MLE method in this example. The last variable is number of intracranial meningiomas (a type of brain tumor) in NF2 patients. A multinormal copula model with negative-binomial margin and a Poisson log-normal mixture model are compared in this example.

6.1 Intertriginous Freckling in NF1 Patients

In this section we use the method presented in Chapter 5 to evaluate the parent-offspring (interclass) and sib-sib (intraclass) odds ratios in intertriginous freckling.

6.1.1 Interclass Odds Ratio

There are 261 families containing one affected parent and his/her affected offspring. Among these families there are 187 with one child, 52 with two children, 22 with three to five. In total there are 364 parent-offspring pairs. The prevalence of intertriginous freckling among the parents is 0.88 and 0.80 among the offspring. The estimates of P_1 to P_4 from these pairs are

$$\begin{aligned}\hat{P}_1 &= 0.714 && \text{(both affected),} \\ \hat{P}_2 &= 0.165 && \text{(parent affected, offspring unaffected),} \\ \hat{P}_3 &= 0.091 && \text{(parent unaffected, offspring affected),} \\ \hat{P}_4 &= 0.030 && \text{(both unaffected).}\end{aligned}$$

To estimate the standard error of the log odds ratio (5.17), we need to estimate the conditional correlations among siblings given the state of their parent, c_1 and c_4 defined in (5.8). The estimation method is presented in Section 5.1.3. $\hat{c}_1 = 0.004$, obtained from an ANOVA table of all sibling groups with affected parent and $\hat{c}_4 = 0.200$, obtained from all sibling groups with unaffected parent.

The jackknife method, in which one family was removed at a time, is also used to estimate the standard error of the log odds ratio. The estimates are given in Table 6.1. The naive standard error in the table is calculated ignoring the dependence among the pairs. The other two standard errors are not very different from the naive standard error, since the conditional associations among siblings are mild. The estimate of the standard error based on (5.5) is slightly larger than the

odds ratio	log odds ratio	s.e.*	naive s.e.*	Jackknife s.e.*
1.51	0.368	0.402	0.377	0.387

* standard error

Table 6.1: Interclass odds ratio in intertriginous freckling

jackknife estimate.

6.1.2 Intraclass Odds Ratio

There are 193 families with at least two siblings. Among them there are 151 families with two siblings, 30 with three and 11 with four. There is one large family with eight siblings, all affected. In total, there are 335 sibling pairs.

To obtain good estimates of $P_{ij|3}$ and $P_{ij|4}$, the probabilities of two pairs falling in respective categories, usually a large number of families with size over two is needed. In this data set, there are not enough families to produce the estimates. However, the exchangeability assumption can help narrow down the region of these values. For k exchangeable binary variables, the joint probabilities are fully determined by $\Pr(Y_i = 1, i = 1, \dots, j)$, $j = 1, \dots, k$. (George and Bowman, 1995). In fact,

$$P_{44|3} = 1 - 3P_2 + P_{11|3}$$

$$P_{44|4} = 1 - 4P_2 + 2P_1 - 4P_{11|3} + P_{11|4}$$

$$P_{14|4} = P_1 - 2P_{11|3} + P_{11|4}.$$

Therefore, we only need to know $P_{11|3}$ and $P_{11|4}$. Given P_1 to P_4 , $P_{11|3}$ and $P_{11|4}$ are restricted to a relatively small region (Figure 6.1). Since an estimate is not available, we choose the point at the center of the region as the values of $P_{11|3}$ and $P_{11|4}$ and carry out the standard error calculation at these values. We calculated both the unweighted and optimal weighted estimates and the corresponding standard errors based on (5.17) and (5.20). The results are reported in Table 6.2. For the optimal weighted estimates, the estimated weights are 0.684, 0.526, 0.292 for families of sizes 3, 4 and 8, respectively. After weighting, there is a slight decrease in the estimate of P_1 and an increase in the log odds ratio estimate. We also calculate the naive standard error (0.314) and the jackknife standard error (0.357) for the unweighted estimate. The standard error reported in Table 6.2 is larger than the jackknife estimate. This could be due to the choice of $P_{11|3}$ and $P_{11|4}$.

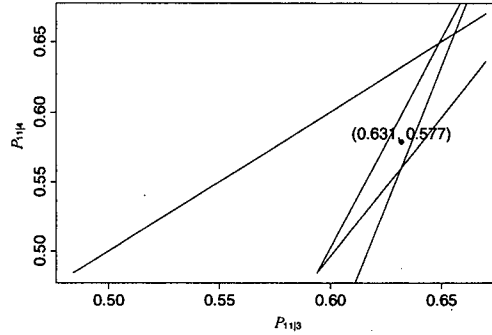


Figure 6.1: Region of $P_{11|3}$ and $P_{11|4}$ (intraclass odds ratio).

	equal weights	optimal weights
P1 (both affected)	0.704	0.680
P2 (one affected, one unaffected)	0.110	0.116
P4 (both unaffected)	0.075	0.088
log odds ratio	1.461	1.490
standard error	*0.385	*0.359

* The standard errors are calculated at $P_{ij|3} = 0.631$, and $P_{ij|4} = 0.577$.

Table 6.2: Intraclass odds ratio of intertriginous freckling

The naive standard error is about 10% smaller than the jackknife standard error.

6.2 Presence of Subcutaneous Neurofibromas in NF1 Patients

Subcutaneous neurofibromas are a common manifestation among NF1 patients. Significant parent-offspring and sib-sib associations have been observed in a previous study (Szudek et al., 2002).

In this section we first apply a multivariate probit model on the dataset used by Szudek. As an alternative, we then fit a two-component series model. The purposes of this section are to (1) demonstrate the CL1 and CL2 methods in multivariate probit regression, and (2) compare the two-component model with the probit model as a standard method in binary data analysis. In addition, in the probit analysis, we also compare a model of conditional independence structure with that of a more general dependence structure. The model comparison is based on the Akaike information criterion (AIC) which is defined as $-2 \log(L) + 2m$, where L is the likelihood and m is the number of parameters in the model.

6.2.1 Summary of Patients

There are 867 NF1 patients included in our analysis. These patients are from 371 independent families. Among all the families there are 12 of size one, 264 of size two, 68 with three, 18 with four, 5 with five, 2 with six and 2 with seven. The numbers of affected pairs included are summarized in Table 6.3.

Type of relative pair	Number of Pairs
siblings	353
parent-offspring	264
second degree	57
third degree	23
fourth degree	1

Table 6.3: Number of affected pairs included in the analysis of presence of subcutaneous neurofibromas in NF1 patients

Subcutaneous neurofibromas are known to be increasing in frequency with age, therefore age is included as a covariate in both models. The average age of the patients is 20.5 (range: 0.2 - 78.3) at the time of examination. Around 95% of the patients are under 50 years old.

6.2.2 Probit Model

We first consider a dependence structure with three parameters: ρ_{ss} (sib-sib correlation), ρ_{po} (parent-offspring correlation) and ρ_{2+} (correlation between any relative pair higher than second degree). The correlation between a third or fourth degree relative pair is set to be equal to the correlation between a second degree relative pair because there are only 24 pairs of third or fourth degree. The parameters are estimated using the CL1, CL2 and MLE methods. The maximum family size is seven in this dataset, so we are able to compute the MLEs using the second order approximation (Joe 1995) for MVN rectangle probabilities. The results are reported in Table 6.4. The covariate age is incorporated on its natural scale, log scale and log-log scale. The log-log transformation provides the best fit of these three and is used in the final model.

The point estimates from all the three methods basically agree, especially the regression parameter estimates. The standard errors of the CL1 and CL2 estimates are close to each other, but around 20% larger than those of the MLEs.

	CL1		CL2		MLE	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
Covariate coefficients:						
β_0	-3.109	0.291	-3.121	0.288	-3.199	0.240
age*	2.517	0.254	2.536	0.250	2.606	0.207
Correlations:						
ρ_{ss}	0.660	0.099	0.594	0.121	0.652	0.093
ρ_{po}	0.493	0.116	0.443	0.109	0.463	0.096
ρ_{2+}	0.710	0.311	0.698	0.322	0.670	0.275
age*=log(log(age+2))						

Table 6.4: Parameter estimates of the Probit model for the presence of subcutaneous neurofibromas in NF1 patients. The three dependence parameters are: ρ_{ss} (sib-sib correlation), ρ_{po} (parent-offspring correlation) and ρ_{2+} (correlation between any relative pair higher than 2nd degree).

Incidentally, the unconstrained estimates from the CL2 method lead to an inappropriate correlation matrix (not non-negative definite) for a multi-generation family. This shows the drawback of the CL methods if the positive definiteness constraint is not used in the estimating procedure.

It is no surprise that the estimates of ρ_{2+} have large standard errors because there are only a small number of 2+ relative pairs and half of them are from four families. Since there is not enough information on the dependence between those relatives, we then consider a dependence structure based on the conditional independence (CI) assumption. In this case, the correlation between any relative pair higher than first degree is a function of ρ_{po} and ρ_{ss} as shown in Example 2.1. Again the results from the three methods agree.

	CL1		CL2		MLE	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
Covariate coefficients:						
β_0	-3.109	0.291	-3.130	0.284	-3.224	0.288
age*	2.517	0.254	2.544	0.247	2.626	0.250
Correlations:						
ρ_{ss}	0.676	0.096	0.604	0.118	0.658	0.085
ρ_{po}	0.536	0.111	0.470	0.107	0.489	0.092
age*=log(log(age+2))						

Table 6.5: Parameter estimates of the Probit model with CI structure for the presence of subcutaneous neurofibromas in NF1 patients.

The negative log-likelihoods of the two models are 425.0 and 426.0, respectively (based on

the MLEs). It indicates that the model with the CI structure provides almost as good a fit as the first model. The AICs of the two models are 860.7 and 860.0, respectively.

Based on the estimates of ρ_{ss} and ρ_{po} and their standard errors, we find that the sib-sib dependence is slightly stronger than the parent-offspring dependence. The estimate of ρ_{2+} is high but has a large standard error. Therefore, we do not have sufficient information to draw any conclusion on the dependence between more distant relatives.

6.2.3 Two-Component Series Model

Now we consider a two-component series model for the presence of subcutaneous neurofibromas. In this model, the occurrence of subcutaneous neurofibromas is assumed to be caused by the heritable and non-heritable components jointly. The non-heritable component Y_N is modelled by a logistic regression. Age is included as a covariate on its natural scale. The multivariate binary beta-binomial model is used to specify distribution of the heritable components of a family unit. Let $Y_{I,il}$ be the heritable component of the parent l and $\mathbf{Y}_{I,-il}$ the corresponding components of l 's offspring. Let k_{il} be the length of $\mathbf{Y}_{I,-il}$. Then,

$$\begin{aligned}\Pr(Y_{I,ij}) &= p_I, \text{ for all } i \text{ and } j \\ \mathbf{Y}_{I,-il} | Y_{I,il} = 0 &\sim \text{beta-binomial}(\alpha_0, \beta_0, k_{il}) \\ \mathbf{Y}_{I,-il} | Y_{I,il} = 1 &\sim \text{beta-binomial}(\alpha_1, \beta_1, k_{il})\end{aligned}$$

Meanwhile,

$$p_I \frac{\alpha_1}{\alpha_1 + \beta_1} + (1 - p_I) \frac{\alpha_0}{\alpha_0 + \beta_0} = p_I.$$

Therefore, there are actually four independent parameters for the heritable component. The parameterization used in the ML estimation procedure is $\boldsymbol{\theta} = (\text{logit } p_I, \log \alpha_0, \log(\beta_0/\alpha_0), \log \alpha_1)$. The transformations improve numerical stability. The ML estimates of the model are reported in Table 6.6.

The estimates are transformed into probabilities and correlations and reported in Table 6.7. To avoid the tedious calculation of the Delta method, the 95% confidence intervals are obtained by resampling $\hat{\boldsymbol{\theta}}$ from its estimated asymptotic MVN distribution.

The estimates of α_0 and α_1 have relatively large standard errors and are sensitive to the starting point. These two parameters are difficult to estimate because they involve the joint dis-

	estimate	SE
Non-heritable:		
intercept	-3.399	0.353
age	0.232	0.027
Heritable:		
logit(p_I)	0.409	0.128
log(α_0)	-3.504	4.475
log(β_0/α_0)	0.922	0.291
log(α_1)	0.186	1.459

Table 6.6: The ML estimates of the two-component series model for presence of subcutaneous neurofibromas in NF1 patients

	estimate	95% confidence interval
Univariate probability:		
p_I	0.601	(0.539, 0.658)
Conditional probabilities:		
$\Pr(Y_{I,ij} = 1 Y_{I,il} = 0)$	0.285	(0.183, 0.414)
$\Pr(Y_{I,ij} = 1 Y_{I,il} = 1)$	0.811	(0.708, 0.884)
Parent-offspring correlation:		
$\text{Corr}(Y_{I,il}, Y_{I,ij})$	0.526	(0.309, 0.695)
Sib-sib correlation:		
$\text{Corr}(Y_{I,il}, Y_{I,ij'})$	0.677	(0.340, 0.750)

Table 6.7: Probabilities and correlations of the heritable component in the two-component series model for the presence of subcutaneous neurofibromas in NF1 patients. Index l is the index of a parent and j, j' are the indices of l 's two children.

tribution of three variables: indicators of the parent and his/her two offspring. To obtain a more informative inference, we need more families with at least two offspring.

6.2.4 Comparison of the Probit Model and the Two-Component Series Model

We first compare the AICs of the two models. Recall that the AIC of the probit model is 860.0. The AIC of the two-component model is 849.4. This indicates that the two-component series model provides a better overall fit than the probit model.

Next, we use the two models to predict the prevalence of subcutaneous neurofibromas in NF1 patients at different ages and compare the results with that observed in the data. The results are plotted in Figure 6.2. The dashed line is the probit model prediction and the solid line is the

two-component model prediction. The circles on the broken line are the proportions of affected patients in different age groups. Patients under 50 years old are grouped in 5-year intervals. There are only 40 patients over 50 years old, so those patients are grouped in 10-year intervals. From the plot we observe a decrease in the estimated prevalence after age 50. This decrease is very likely due to the death of patients with more serious forms of the disease. Subcutaneous neurofibroma itself is a benign complication of NF1, but it has been reported to have an association with other more severe tumors, such as plexiform neurofibroma (Szudek, 2000). Therefore there could be a bias in older patients. For this reason, we limit our comparison to before age 50. The estimate from the probit model has a major departure from the data around age 20; whereas the estimate based on the two-component model follows the data well. This indicates that the two-component is better in modelling the age covariate while adequately approximating the dependence.

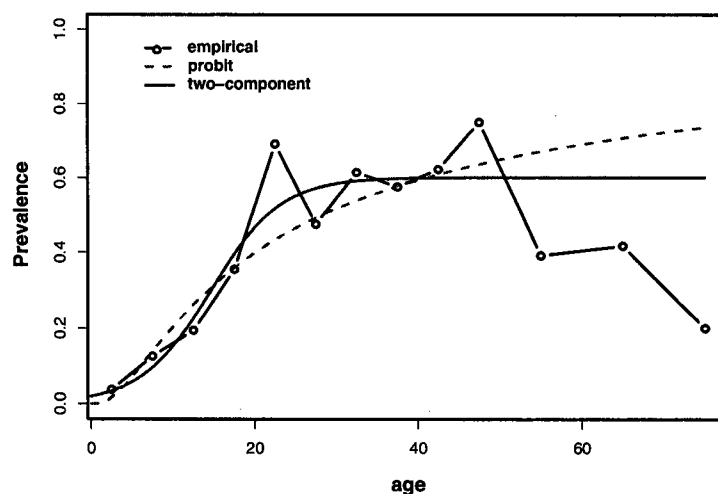


Figure 6.2: Predictions of the prevalence of subcutaneous neurofibromatosis in NF1 patients by the probit model and the two-component model.

The above comparison is based on the fit of the models. In terms of implementation, the two-component model is easier to implement than the probit model as it takes much less computing time.

Type of relative pair	Number of Pairs
siblings	206
parent-offspring	152
second degree	196
third degree	96
other	280

Table 6.8: Number of affected pairs included in the analysis of age at first symptom of NF2

6.3 Age at First Symptom of NF2

In this section we apply the CL1 and CL2 methods to a MVN model for survival data. The response is the age at first symptom of NF2. Censoring can occur because some patients were diagnosed by gene testing before any manifestation was present. We assume that the censoring time is independent of the age at first symptom.

6.3.1 Summary of Patients

There are 721 individuals from 526 unrelated families. The family sizes ranged from 1 to 23 affected members. There are 378 with one member, 35 with two, 18 with three, 16 with four, 13 with 5-8 and 6 with at least ten. Among the 721 individuals included, 70 were right censored, i.e., asymptomatic. There are 30 families with one censored case, 9 with two, 3 with three, and 2 families, each of size ten, with five and eight censored cases, respectively. The number of affected pairs included in the analysis is summarized in Table 6.8.

6.3.2 Model

In previous work (Zhao et al., 2002), we used a simple random effects model to assess the intrafamilial correlation without distinguishing the correlations between different types of relatives. In the study, we only considered inherited cases. An intrafamilial correlation of 0.51 was observed with a 95% confidence interval (0.35, 0.64). In this section, we further examine the correlations between various types of relative pairs, such as parent-offspring, siblings and more distant relatives. Individuals with both new and inherited mutation are included. The previous analysis indicated that the univariate marginal distribution of the data is close to a normal distribution. Therefore, we use a MVN model in our analysis.

Mutations have been found throughout the *NF2* gene, and different mutations tend to occur in different families (Friedman, 1999). These mutations are of various types, but most can be classified as nonsense, frameshift, splice-site, missense, or large deletions. In general, patients with a truncating mutation (i.e. nonsense or frameshift) are associated with an earlier onset. Therefore we include the indicator of truncating mutation as a covariate. Another covariate, the indicator of new mutation (1 = new mutation and 0 = inherited mutation), is included as well.

We consider four dependence parameters: the sib-sib correlation (ρ_{ss}), the parent-offspring correlation (ρ_{po}), the correlation between second degree relatives (ρ_2), the correlation between relatives more distant than second degree relatives (ρ_{3+}).

6.3.3 Results

We implement the CL1 and the CL2 methods by minimizing the negative CL functions using the quasi-Newton algorithm (Nash, 1990). The standard errors are estimated by the jackknife method with one family removed each time. We also compute the ML estimates. The likelihood for censored family members is computed using the second order approximation proposed by Joe (1995) for MVN rectangle probabilities. The results are reported in Table 6.9.

	CL1		CL2		MLE	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
Marginal:						
β_0	27.776	1.353	27.228	1.395	22.814	1.031
New mutation	-3.496	1.415	-2.966	1.494	1.028	1.149
Truncating	-7.295	1.159	-7.377	1.165	-5.997	1.245
σ^2	12.823	0.575	12.751	0.556	12.253	0.368
Correlations:						
ρ_{ss}	0.425	0.093	0.454	0.079	0.616	0.051
ρ_{po}	0.192	0.060	0.138	0.086	0.303	0.068
ρ_2	0.170	0.093	0.100	0.085	0.220	0.069
ρ_{3+}	0.105	0.115	0.081	0.109	0.115	0.067

Table 6.9: Parameter estimates of the MVN model for age at first symptom of NF2

The point estimates and the standard errors from the CL1 and CL2 methods are similar. The ML estimates of the regression coefficients and the two correlations, ρ_{ss} and ρ_{po} , are quite different from those CL estimates. This could be due to the normality assumption being violated.

We observed a certain degree of right skewness in the sample distribution. The ML method could be more sensitive to the departure from the model assumptions than the CL methods. Another possibility is that there are a few large families which are highly influential. Further investigation is needed to explain the differences.

The results indicate that the sib-sib correlation is considerably higher than the correlation between other type of relatives. It also suggests a decreasing pattern when two relatives become more distantly related.

In terms of computation, the computing times are 20, 42 and 342 seconds for the CL1, CL2 and ML methods, respectively. Moreover, the ML method is relatively sensitive to the starting point.

6.4 Number of Intracranial Meningiomas in NF2 Patients

The variable considered in this section is a count trait: number of intracranial meningiomas in NF2 patients. This variable was analyzed in the previous study (Zhao et al., 2002). Since the information on relative types was not available, we assumed an exchangeable dependence structure within a family and modelled the counts using a gamma mixture of negative binomial distribution. An significant familial correlation was observed. In the current study, the information on relatives has become available and we would like to assess the dependence between different types of relatives. The gamma mixture of negative binomial model is not suitable for this analysis since it can only provide one level of dependence. Therefore, in this section, we consider a multinormal copula model (Section 2.3.2) with negative binomial margin and a Poisson-lognormal mixture model (Section 2.3.1) to analyze the data.

6.4.1 Summary of Patients

In the dataset, there are 628 individuals from 431 families with information on meningiomas. Among the families there are 35 of size two, 16 with three, 15 with four, 11 with 5-7, 2 with nine and 2 with ten. The number of affected pairs included is summarized in Table 6.10.

Among the 628 patients, there are 350 new mutation cases and 278 inherited cases. The counts of meningiomas are summarized in Table 6.11 by new mutation cases and inherited cases.

Type of relative pair	Number of Pairs
siblings	104
parent-offspring	157
second degree	109
third degree	44
other	67

Table 6.10: Number of affected pairs included in the analysis of number of meningiomas in NF2 patients.

There are some patients recorded with meningiomas, where the count is unknown. These individuals are summarized in the last column. The sample mean and variance calculated based on the known counts are 1.75 and 6.12 for new mutation cases and 0.87 and 3.79 for the inherited cases. This indicates overdispersion with respect to the Poisson distribution.

Type of mutation	Number of Meningioma													
	0	1	2	3	4	5	6	7	8	9	10	19	20	> 0, but no count
New	130	71	48	19	17	12	10	1	5	3	0	6	1	27
Inherited	160	48	19	7	5	5	2	0	1	2	1	1	0	27

Table 6.11: Summary of meningioma counts in NF2 patients

6.4.2 Multinormal Copula Model with Negative Binomial Margin for Number of Meningiomas

We first considered a multinormal copula model. Since overdispersion relative to the Poisson distribution is observed, we chose the negative-binomial distribution to model the univariate margin of the count variable Y_{ij} . The univariate pmf of Y_{ij} is

$$\Pr(Y_{ij} = y_{ij}) = \frac{\Gamma(\theta_{ij} + y_{ij})}{\Gamma(\theta_{ij})y_{ij}!} \frac{\mu_{ij}^{y_{ij}} \theta_{ij}^{\theta_{ij}}}{(\mu_{ij} + \theta_{ij})^{\theta_{ij} + y_{ij}}}.$$

The parameter μ_{ij} is the mean of Y_{ij} and we specify $\log \mu_{ij} = \beta_0 + \beta_1 x_{ij}$, where $x_{ij} = 1$ for a new mutation case and 0 for an inherited case. We also assume different θ_{ij} values for the new and inherited mutation cases: $\theta_{ij} = \theta_1$ if it is a new mutation, otherwise, $\theta_{ij} = \theta_0$.

We consider 3 dependence parameters: the sib-sib correlation (ρ_{ss}), the parent-offspring correlation (ρ_{po}), and the correlation between relatives of second or higher degree (ρ_{2+}).

We use the CL1 and CL2 methods to estimate the parameters. The jackknife method is used to estimate the standard errors. Individuals with positive but unknown count are included in the analysis by calculating the probability of $Y_{ij} > 0$. The results are reported in Table 6.12. The CL1 estimates of the marginal parameters are very close to the CL2 estimates. As for the dependence parameters, the standard error of the CL2 estimate for ρ_{ss} is considerably larger than that of the CL1 estimate. This is due to two special influential cases. The first case is a family containing two siblings with counts 5 and 9. The removal of this family results a decrease of 0.102 in the CL2 estimate of ρ_{ss} . The second case is a family containing two siblings with counts 1 and 10. The removal of this family results an increase of 0.156 in the CL2 estimate of ρ_{ss} . Similar changes occur in the CL1 estimate, but with smaller magnitudes (a decrease of 0.04 in the first case and an increase of 0.05 in the second case). The CL2 estimate is more sensitive to unusual pairs from small families since pairs from a small family are weighted more in the CL2 method than in the CL1 method. When there are influential observations in the data, the jackknife standard error can be inflated. To show the effect, we calculate the jackknife standard errors without using the two sets of estimates obtained with the two influential families deleted. The results are reported as s.e.* in Table 6.12.

	CL1			CL2		
	estimate	s.e.	s.e.*	estimate	s.e.	s.e.*
Marginal:						
β_0	0.024	0.158	0.144	0.030	0.155	0.143
β_1	0.590	0.175	0.162	0.587	0.171	0.160
θ_0	0.454	0.110	0.105	0.442	0.104	0.099
θ_1	0.829	0.115	0.115	0.827	0.114	0.114
Correlations:						
ρ_{ss}	0.481	0.119	0.100	0.408	0.202	0.080
ρ_{po}	0.436	0.081	0.080	0.400	0.092	0.097
ρ_{2+}	0.454	0.229	0.228	0.382	0.218	0.217
–log likelihood	911.4			910.7		

s.e.*: jackknife standard error with two outliers removed.

Table 6.12: Parameter estimates of the multinormal copula model with negative binomial margin for number of meningiomas in NF2 patients

6.4.3 Multivariate Poisson-Lognormal Mixture Model for Number of Meningiomas

We also use a multivariate Poisson-lognormal mixture model to analyze the same data. In this model, $Y_{ij}|\Lambda_{ij} = \lambda_{ij} \sim \text{Poisson}(\lambda_{ij})$ and $\log \Lambda_{ij} \sim \text{Normal}(\mu_{ij}, \sigma_{ij}^2)$, where $\mu_{ij} = \beta_0 + \beta_1 x_{ij}$ with $x_{ij} = 1$ for a new mutation case and 0 for an inherited case, and $\sigma_{ij}^2 = \sigma_1^2$ if it is a new mutation, otherwise, $\sigma_{ij}^2 = \sigma_0^2$. We consider the same dependence parameters for $(\log \lambda_{i1}, \dots, \log \lambda_{ik_i})$ as in the multinormal copula model. The CL2 and CL2 estimates and jackknife standard errors are reported in Table 6.13. The results show that the two influential families have a even stronger effect on the estimates of ρ_{ss} in this model.

	CL1			CL2		
	estimate	s.e.	s.e.*	estimate	s.e.	s.e.*
Marginal:						
β_0	-0.823	0.207	0.207	-0.867	0.205	0.203
β_1	0.914	0.219	0.218	0.955	0.215	0.214
σ_0^2	1.718	0.329	0.307	1.838	0.356	0.335
σ_1^2	1.083	0.146	0.146	1.085	0.147	0.147
Correlations:						
ρ_{ss}	0.677	0.167	0.137	0.584	0.288	0.102
ρ_{po}	0.599	0.123	0.123	0.578	0.137	0.137
ρ_{2+}	0.652	0.285	0.285	0.574	0.297	0.296
-log likelihood	911.8			910.9		

s.e.*: jackknife standard error with two outliers removed.

Table 6.13: Parameter estimates of the multivariate Poisson-lognormal mixture model for number of meningiomas in NF2 patients

6.4.4 Comparison of the Two Models

Since both models contain the same number of parameters, we use the negative log likelihood evaluated at the CL estimates as a measure of the overall fit. The negative log likelihood of the copula model is computed using Joe's second order approximation for MVN rectangle probabilities. As for the Poisson-lognormal mixture model, the likelihood of a family is computed using numerical integration if the family size is less than 3; otherwise, the likelihood is computed using the Monte Carlo method. The results are reported at the bottom of Tables 6.12 and 6.13. Since the negative

new mutation:	0	1	2	3	4	5	6	7-9	>= 10	Pearson χ^2
observed	130	71	48	19	17	12	10	9	7	
expected (MCNB)	132	66	42	27	18	12	8	12	5	5.4
expected (MPLN)	126	77	45	26	10	7	8	10	8	5.2
inherited mutation:	0	1	2	3,4	5,6	>= 7				
observed	160	48	19	12	7	5				
expected (MCNB)	163	39	19	18	7	5				4.2
expected (MPLN)	162	45	19	14	5	6				1.4

Table 6.14: Expected number of meningiomas based on the multinormal copula model with negative binomial margin (NBMC) and the multivariate Poisson-lognormal mixture (MPLN) model

log likelihoods of the two models are very close to each other (910.7 and 910.9 based on the CL2 estimates), there is no indication that one model fits better than the other.

To check the marginal fit, we calculate the expected counts based on the CL2 estimates and compare them with the observed counts (Table 6.14). The expected counts for those individuals with positive but unknown counts are calculated conditional on $Y_{ij} > 0$ then subtracted from the total expected counts for all the individuals. The Pearson χ^2 statistic is also reported in the table. Since the data are correlated, the statistic is not chi-square distributed. However, a small value of the statistic still indicates a reasonable fit. In all the cases, the Pearson χ^2 statistics are less than the number of the categories. For the new mutation case, the fits of the two models are similar. For the inherited case, the Poisson-lognormal mixture model fits the data better.

The estimated means based on the two models are almost the same. For the first model, the estimated mean is 1.85 for the new mutation cases and 1.03 for the inherited mutation cases. For the second model, the estimated means are 1.89 and 1.05, respectively. However, there is a difference in the standard deviations. For the first model, the estimated standard deviation is 2.45 for the new mutation cases and 1.85 for the inherited mutation cases. For the second model, the estimated standard deviations are 2.99 and 2.63, respectively. With the estimated parameters, the Poisson-lognormal mixture distribution has a longer right tail than the negative binomial distribution. For example, in the new mutation case, the fitted negative binomial distribution has a probability of 0.0003 to produce a count over 20 (the largest count in our data set), whereas the fitted Poisson-lognormal mixture distribution has a probability 10 times higher to produce such a count.

The estimates of the dependence parameters from both models suggest an exchangeable

dependence structure within a family. However, the estimate of ρ_{2+} has a large standard error.

6.5 Discussion

The examples demonstrate that the CL1 and CL2 methods work well in practice. They also revealed some aspects that need to be investigated in future research. For example, it will be useful to gain some understanding of the consequences of departures from model assumptions, the influence of outliers or influential observations on the parameter estimates and the influence of large families on the estimates of the dependence parameters. In Chapter 4, we studied the asymptotic properties of the CL methods. The properties of the CL estimates may not be the same when the sample size is close to the data size in our examples. It would be interesting to further investigate the behavior of the CL estimators when the sample size is not very large. We will discuss these issues further in Chapter 7.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The major contributions of our work to familial data analysis are in modelling and estimation approaches for non-normal traits. The CI model proposed in Chapter 2 provides a new approach to construct models for familial data. We gave two examples: the two-component model for binary trait and the CI model for count trait using a distribution of convolution-closed infinitely divisible class. Potentially, many other models can be constructed based on the CI assumption. The example in Section 6.2 demonstrated that the CI model works well with real data and can be easily implemented. We also considered two existing classes of models: random effects models and multi-normal copula models, which are suitable for modelling familial data, but hindered in practice by their computational difficulties. Equipped with the CL estimation approaches discussed in Chapter 3, these models become feasible. Our investigations in the asymptotic efficiency comparisons showed that the relative asymptotic efficiencies of these two methods are satisfactory except for some extreme cases. Those investigations help us better understand under what situations these methods work well and provide hints on how to improve them.

7.2 Future Work

Due to the complexity in the nature of familial data, there are still many variations and extensions that remain to be explored. Listed below are several research problems that look interesting for

future exploration.

1. Extend the multinormal copula models to semiparametric margins. There is of special interest in survival models. A Cox-type regression model has been considered to analyze clustered failure time data (Wei et al., 1989; Lee et al., 1992; Spiekerman and Lin, 1998). The dependence of the failure times within a cluster is assumed to be exchangeable, but not specified by a parametric model. The marginal hazard function for the failure time T_{ij} is

$$\lambda_{ij}(t; \mathbf{X}_{ij}) = \lambda_0(t)e^{\mathbf{X}_{ij}\beta},$$

where λ_0 is the common baseline hazard function. β is estimated from “quasi-partial likelihood”, which takes the form of the partial likelihood of independent observations. The cumulative baseline hazard function is estimated by an Aalen-Breslow type estimator. Spiekerman and Lin (1998) established the asymptotic results for the estimators and developed procedures to approximate the covariance matrix. To combine the Cox-type of regression with the multinormal copula, we allow different dependence structure among the variables. It is natural to use the two-stage method to estimate the marginal and dependence parameters. We can adapt Spiekerman and Lin’s procedure to estimate the regression parameters and the cumulative base-line hazard, then estimate the dependence parameters based on BCLs. King et al. (1996) and Bandeen-Roche and Liang (1996) used a similar procedure for a gamma frailty model with unspecified common base-line hazard, but they only provided asymptotic results for their first stage.

The remaining work is to establish the asymptotic results for the estimating procedure and develop a method to estimate the standard errors of the dependence parameters. Even though we use the existing procedure in the first step, the asymptotic results need to be modified, since they are based on an exchangeable dependence structure. The major challenge is to develop the asymptotic properties for the dependence parameter estimators. To estimate those parameters, we need to plug-in the estimated marginal cdf in the bivariate marginal likelihood function. Unlike a full parametric model, the estimated marginal cdf is a step function.

2. Develop models for multiple responses and longitudinal familial data. In this thesis, we only consider one measurement on each individual. Often, values of more than one trait are recorded on each individual. It is of interest to see how the traits are related to each other. Sometimes patients are followed over time and the value of a trait is recorded longitudinally. There is a need to model the change of the trait over time. In both cases, there are more associations

that need to be taken into account: the association of the two measurements on the same family member, the association of two measurements across family members. Moreover, for the first case, the data types of two traits might differ.

3. Investigate other properties of the CL methods. As mentioned in the data examples, in practice we might encounter problems such as departures from model assumptions or outliers. It would be useful to investigate the robustness of these methods compared to the ML method. Intuitively, the CL methods might be more robust to violations of the distribution assumptions. The estimates of the marginal parameters might be closer to the true marginal characteristics since only marginal information is used. On the other hand, the influence of outliers on the CL methods could be stronger. It is also important to see how these methods are affected by large families. Another direction is to examine their small sample properties. For many studies, the sample size is around a few hundred families. For the dependence parameters, the convergence is relatively slower than for the marginal parameters. The investigations can be done by conducting simulations with different sample sizes.

4. Further develop the weighted CL1 method. In Chapter 3, we derived the theoretical optimal weights for the CL1 estimating functions. Since the optimal weights depend on unknown parameters, the method needs to be implemented iteratively. In the future, we will develop the asymptotic theory and consider implementation of the method.

Appendix A

Dependence Structure of Quantitative Traits

In this appendix, we introduce the dependence structure of quantitative traits. Analysis of quantitative traits is fundamental to familial data analysis. It will help us to understand the patterns of association in non-normal traits. Moreover, the dependence structure of quantitative traits can be reasonably linked to non-quantitative traits. A good example is the multivariate probit model for binary traits. In such a model, the presence or absence of a binary trait is assumed to be determined by an unobservable quantitative trait which is under the influence of multiple factors. This idea can be used to generate other types of non-normal traits, such as the log-normal Poisson model for count data, or the log-normal frailty model for survival data. These models are discussed in detail in Chapter 2.

As mentioned in Section 1.2, the value of a quantitative trait can be affected by a single gene, several genes, a large number of genes with comparable effects, or a gene with a major effect plus a large number of genes with small effects. The trait is normally also affected by environmental factors. How and how much the genetic and environmental factors affect a trait is often reflected in the association pattern within family members. The idea of modelling the dependence structure of quantitative traits is to decompose the total phenotypic variance into different components attributable to various factors. Some variance can be attributed to known and measurable factors — these factors are treated as fixed effects. Genetic factors are normally

modelled as random effects since modelling is at the phenotypic level without genotype information at loci and therefore genetic effects are unobservable. The major gene model is an exception, in which information on the major gene is available, but genetic effects additional to the major gene are still unobservable. Environment effects due to factors such as a shared household are unobservable as well. These factors are modelled as random effects. Environmental factors and genetic factors are usually considered additive.

In this appendix, different models for quantitative traits will be introduced. Behind each model there is a different mechanism and therefore each model has a different dependence structure. But they share a common feature: the degree of association due to genetic factors decreases when two relatives are more distantly related. This is because of the fact that the more distantly related two individuals are, the less likely they are to share a gene by descent.

Section A.1 discusses the simplest case of Mendelian models. Section A.2 discusses polygenic models. The major gene models and models for autosomal-dominant diseases are discussed in Sections A.3 and A.4. The appendix ends with some general comments.

A.1 Mendelian Models

Mendelian models involve only one locus or, in a more general case, a finite number of loci. In order to proceed from the simple to the complex, we first consider a trait controlled by a single locus with two alleles A and A' . If the frequencies of A and A' are p and q , then according to the **Hardy-Weinberg** equilibrium there are three possible genotypes, AA , AA' and $A'A'$ with frequencies p^2 , $2pq$ and q^2 . Suppose the value of the phenotype (Y) is completely determined by the genotype (G) the variance of Y is purely phenotypic. Furthermore, suppose $Y(AA) = \mu + a$, $Y(AA') = \mu + d$ and $Y(A'A') = \mu - a$, where $-a < d < a$. When $d \neq 0$, we say one allele is dominant over the other. The variance of Y can be expressed as $\sigma_g^2 = 2pq(a + (q - p)d)^2 + (pqad)^2$. The first term is called the additive genetic variance σ_a^2 , and the second term is called the dominance genetic variance σ_d^2 . If we do a regression of Y on the number of A alleles, σ_a^2 is the variation explained by the regression and σ_d^2 is the unexplained variation (Li, 1976).

The additive genetic variance is closely related to the concept of heritability (h^2), which is defined as the ratio of additive genetic variance to the total variance. It represents the degree of

influence of the phenotype value on the next generation. The dominance genetic variance arises from the phenomenon of dominance among alleles. When there is no dominant allele, $\sigma_d^2 = 0$. To compare these two components, let us look at the ratio of σ_d^2 to σ_a^2

$$\frac{pq}{2} \left[\frac{ra}{1 + \delta(q-p)r} \right]^2,$$

where $r = |d/a|$ is the degree of dominance and δ is the sign of d . This ratio is mainly affected by a . When a is small and p is not too close to 1 or 0, the contribution of σ_d^2 is negligible comparing to σ_a^2 . (For example, when $p = q = 0.5$ and $a = 1$, the ratio is less than 0.125.) Based on random mating and Mendel's laws, the phenotypic covariance between any pair of relatives can be derived after some calculations. In Table A.1 we list the covariance for several common types of relative pairs.

Relationship	Covariance
Offspring and one parent	$\sigma_a^2/2$
Offspring and midparent*	$\sigma_a^2/2$
Identical twins	$\sigma_a^2 + \sigma_d^2$
Full siblings	$\sigma_a^2/2 + \sigma_d^2/4$
Half siblings	$\sigma_a^2/4$
Nephew and uncle	$\sigma_a^2/4$
Grandparent and grandchild	$\sigma_a^2/4$
First cousins	$\sigma_a^2/8$

*midparent is the average of the two parents

Table A.1: Covariances between relatives in a simple Mendelian model

A full explanation about how to derive the above covariances can be found in standard references, such as Falconer (1989). In general, the covariance between the trait values of two relatives, Y_i and Y_j , can be written as

$$\text{Cov}(Y_i, Y_j) = 2\Phi\sigma_a^2 + \Delta\sigma_d^2,$$

where Φ and Δ are known as the kinship coefficient and the identity coefficient in genetics literature (Falconer, 1989).

Remarks:

1. The table shows that closer biological relationship generally leads to greater phenotypic similarity.

2. Parent-offspring correlation is always less 0.5. This makes sense, since an offspring and parent only share 50% of their genetic materials.
3. The sib-sib correlation is generally higher than the parent-offspring correlation due to the dominance effect. When the locus has a strong effect on the trait (i.e., a is large) and one allele is obviously dominant over the other, the sib-sib correlation could be much stronger than the parent-offspring correlation. Otherwise, the difference between the sib-sib correlation and parent-offspring correlation is rather small.

More often the value of Y is not completely determined by the genotype. Instead, it may be subject to some random variation. In this case, we assume that given the genotype G , Y has mean $\mu + g$ and variance σ_e^2 , where g is the loci effect with value a , d and $-a$ when $G = AA$, AA' and $A'A'$ respectively; σ_e^2 is also called environmental variance, representing the influence of non-genetic circumstances. Then the total phenotypic variation in the population is $\sigma_g^2 + \sigma_e^2$. If the environmental contribution is uncorrelated within families, it will not affect the covariance between relatives, but will weaken the correlation since it increases the total variance.

When more than one locus is involved, if there is no interaction between the loci, the variance components can be derived by summing over each individual locus. When interactions among the loci are included, the genetic variance will be partitioned into three components:

$$\sigma_g^2 = \sigma_a^2 + \sigma_d^2 + \sigma_I^2,$$

where the new term σ_I^2 is associated with the interactions, and thus called the interaction genetic variance. The covariances between different types of relatives can be found in Kempthorne (1955a,b).

A.2 Polygenic Models

A polygenic (PG) model involves many loci. It is assumed that these loci act independently and additively, each with relatively small effects. Under this assumption, the vector of trait values \mathbf{Y} for a family is multivariate normally distributed by the central limit theorem (Lange, 1978). Moreover, the total genetic contribution to the phenotypic variation is the sum of the contribution

of the individual loci. Therefore the genotypic variability still can be partitioned into the additive and dominance components; each is summed over all participating loci.

The variation of \mathbf{Y} comes from different types of factors: measurable environmental factors which are typically modelled as fixed effects, unobservable factors such as a shared household effect which are modelled as random effects, and the polygenic factors which are modelled as random effects. The model is specified as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{G} + \boldsymbol{\Xi} \quad (\text{A.1})$$

where \mathbf{G} denotes the polygenic factor and is multivariate normal with mean $\mathbf{0}$ and covariance matrix

$$2\sigma_a^2\boldsymbol{\Phi} + \sigma_d^2\boldsymbol{\Delta} \quad (\text{A.2})$$

where σ_a^2 , σ_d^2 are defined as before, and $\boldsymbol{\Phi}$ and $\boldsymbol{\Delta}$ are matrices of the kinship and identity coefficients. $\boldsymbol{\Xi}$ is the environmental factor which can be further decomposed into components. For instance, if there is a shared household effect that needs to be considered, then $\boldsymbol{\Xi}$ can be specified as normal with mean $\mathbf{0}$ and covariance matrix $\sigma_h^2\mathbf{H} + \sigma_e^2\mathbf{I}$, where σ_h^2 is the common household variance, \mathbf{H} is a matrix of household indicators, σ_e^2 is the random variance, and \mathbf{I} is the identity matrix. As a result, the covariance matrix of \mathbf{Y} is given by

$$2\sigma_a^2\boldsymbol{\Phi} + \sigma_d^2\boldsymbol{\Delta} + \sigma_h^2\mathbf{H} + \sigma_e^2\mathbf{I}. \quad (\text{A.3})$$

In many polygenic models, the dominance variance is omitted. As we have mentioned before, when the effect of a locus is small, the additive component outweighs the dominance component. This was also justified by Morton (1974) and Amos (1988). On the contrary, when a notable dominance variance component is detected in a polygenic model, it could indicate the existence of gene or genes which have a large effect on the trait compared with the other genes. Such a gene is called a major gene.

A.3 Major Gene Models

In a major gene model, a quantitative trait is determined by the contribution of a major locus plus a polygenic contribution (Elston and Stewart, 1971; Morton and Maclean, 1974). In a linkage study (a study to identify the major gene using markers), a marker locus is linked with this major gene,

and the distribution is a mixture of normals. It is not surprising that computational problems arise when a large number of alleles are observed at the locus (Ott, 1979). Thus a random effects model may be a good approximation for such cases (Andrade et al., 1999). The model is represented as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{G} + \mathbf{M} + \boldsymbol{\Xi}, \quad (\text{A.4})$$

where \mathbf{G} is the polygenic effect with only the additive component, \mathbf{M} is the major gene effect whose covariance matrix depends on the recombination fraction and the IBD (identical by descent) value (Amos, 1994), and $\boldsymbol{\Xi}$ is the environmental effect. We will not discuss this model in detail as linkage study is not the focus of this thesis.

A.4 Models for Autosomal-Dominant Diseases

A genetic disorder is autosomal-dominant if it is caused by a single, abnormal gene on one of the autosomal chromosomes (one of the first 22 non-sex chromosomes) and a mutation at one copy of the allele is sufficient for expression of such a disease. One of the parents will usually have the disease in this mode of inheritance and only one parent must have an abnormal gene in order for the child to inherit the disease. Examples are NF1 and NF2. Since a new mutation is rare, we can assume that in a family the mutant alleles are inherited from a common ancestor and therefore identical. However, wide expressive variability can occur among the affected relatives. Differences in the other genes may be important reason for the variation. When we study a population of autosomal-dominant disease, the trait of interest is normally a clinical feature of the disease, such as the onset time. If the variation in the trait is due to a polygenic effect, then the polygenic models can be used to model those effects. If the variation is due to a polygenic effect plus a modifying gene effect, then the major gene models can be used. Differences in the normal allele at the responsible locus could be another reason for the phenotypic differences. For siblings, the mutant allele is from their affected parent, whereas the normal allele is from the unaffected parent. With no inbreeding, affected siblings have 50% chance to inherit the same normal allele from their unaffected parent. Two affected family members, except for siblings, do not share an identical normal allele by descent, since their unaffected parents are not related, i.e., they do not share any gene by descent.

Suppose we model the normal allele effect as a random effect Λ with distribution $N(0, \sigma_\Lambda^2)$. Let Λ_i and Λ_j denote the normal allele effects of two individuals i and j . If i and j are monozygotic

twins, they share an identical normal allele by descent, then $\Lambda_i = \Lambda_j$. If i and j are full siblings, there is a 50% chance that $\Lambda_i = \Lambda_j$ and 50% chance that Λ_i and Λ_j are independent. For any other types of relatives, Λ_i and Λ_j are independent. Now let p be the chance of two individuals to share an identical normal allele by descent. The covariance due to the normal allele effects is then

$$\begin{aligned}\text{Cov}(\Lambda_i, \Lambda_j) &= E(\Lambda_i \Lambda_j) \\ &= pE(\Lambda_i \Lambda_j | \Lambda_i = \Lambda_j) + (1 - p)E(\Lambda_i \Lambda_j | \Lambda_i \perp \Lambda_j) \\ &= p\sigma_\Lambda^2.\end{aligned}$$

The covariance $\text{Cov}(\Lambda_i, \Lambda_j)$ is σ_Λ^2 between monozygotic twins, $\sigma_\Lambda^2/2$ between full siblings and 0 between any other relative pairs.

If we also consider a polygenic effect and environmental effect, the model can be expressed as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{G} + \boldsymbol{\Lambda} + \boldsymbol{\Xi} \quad (\text{A.5})$$

where $\boldsymbol{\Lambda}$ is the vector of the normal allele effects of a family with covariance matrix $\sigma_\Lambda^2 \boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a matrix with γ_{ij} equal to the probability of individuals i and j sharing an identical normal allele by descent.

A.5 Comments

Up to now, we have discussed how to specify the dependence structure based on how genetic and environmental factors are involved. In practice, often it is the other way around — the correlation structure of a trait is estimable and the question to be answered is what are the genetic and environmental influences on the trait. This can be assessed by comparing the correlation between different type of relatives. For example, if the parent-offspring correlation is comparable with the sib-sib correlation and the correlation declines by half when the two individuals are one degree less related, this suggests a polygenic effect. To identify a modifying gene effect or a normal allele effect, it is important to compare the correlation among monozygotic twins, parent-offspring, full siblings and half siblings.

Appendix B

Appendix for Chapter 4

B.1 General Results

The first part are some general results on matrices from algebra and calculus. The second part are some results on expectations and covariances of quadratic forms of normal random variables. These results were used to derive the information matrices and the proof of Theorem B.1 in Appendix B.2.

Let $\mathbf{A} = (a_{ij})$ be a symmetric and positive definite matrix and A_{ij} be the cofactor corresponding to a_{ij} . Then (Lang, 1997)

$$|\mathbf{A}| = \sum_j a_{ij} A_{ij}, \quad \text{for any } i$$

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} (A_{ij})$$

$$\frac{\partial}{\partial \theta} \mathbf{A}^{-1} = -\mathbf{A}^{-1} \left(\frac{\partial}{\partial \theta} \mathbf{A} \right) \mathbf{A}^{-1}$$

$$\frac{\partial}{\partial \theta} \log |\mathbf{A}| = \text{tr} \left(\mathbf{A}^{-1} \frac{\partial}{\partial \theta} \mathbf{A} \right).$$

Let \mathbf{Z} and \mathbf{Y} be n dimensional random vectors, $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$ and $\mathbf{Y} \sim N(\mathbf{0}, \Sigma)$. \mathbf{A} and \mathbf{B} are $n \times n$ symmetric matrices.

1. $E(Z_i Z_j Z_k) = 0$

$$E(Y_i Y_j Y_k) = 0, \quad i, j, k = 1, \dots, n$$

More generally, $E(\mathbf{C}_1 \mathbf{Y}, \mathbf{Y}' \mathbf{C}_2 \mathbf{Y}) = 0$, where \mathbf{C}_1 and \mathbf{C}_2 are arbitrary matrices.

$$\begin{aligned}
2. \quad & E(\mathbf{Z}'\mathbf{A}\mathbf{Z}) = \text{tr}(\mathbf{A}) \\
& E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = \text{tr}(\mathbf{A}\mathbf{\Sigma}) \\
& \text{Var}(\mathbf{Z}'\mathbf{A}\mathbf{Z}) = 2\text{tr}(\mathbf{A}^2) \\
& \text{Var}(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = 2\text{tr}((\mathbf{A}\mathbf{\Sigma})^2) \\
& \text{Cov}(\mathbf{Z}'\mathbf{A}\mathbf{Z}, \mathbf{Z}'\mathbf{B}\mathbf{Z}) = 2\text{tr}(\mathbf{A}\mathbf{B}) \\
& \text{Cov}(\mathbf{Y}'\mathbf{A}\mathbf{Y}, \mathbf{Y}'\mathbf{B}\mathbf{Y}) = 2\text{tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma})
\end{aligned}$$

(Seber, 1977).

B.2 Asymptotic Properties of the Covariates

We prove the asymptotic properties of matrices in the information matrices under the assumptions in Section 4.1.2. Results (1) and (2) in Theorem B.1 are Results 4.3 and 4.4 in the case of one covariate; result (3) is used for the case of more than one covariate.

Lemma B.1 *Let S_n , $n = 1, 2, \dots$, be a sequence of random variables with finite variance. A sufficient condition for $S_n - ES_n \xrightarrow{P} 0$ is that*

$$\text{Var}(S_n) \rightarrow 0.$$

The proof follows from Chebyshev's inequality.

Theorem B.1 *Let $\mathbf{X}_i = (X_{i1}, \dots, X_{ik_i})'$ be a k_i -dimensional random vector, $i = 1, 2, \dots, n$. Suppose $1 \leq k_i \leq M$ and M is bounded. Assume $\{\mathbf{X}_i\}$ has the following properties: (a) $E\mathbf{X}_i = \mathbf{0}$, (b) $\text{Cov}(\mathbf{X}_i) = \mathbf{I}_{k_i \times k_i}$ and (c) $\mathbf{X}_1, \dots, \mathbf{X}_n$ are mutually independent. Suppose $\mathbf{D}_i = (d_{ijj'})$ is a symmetric $k_i \times k_i$ matrix with $|d_{ijj'}|$ bounded by T , a finite real number. As $n \rightarrow \infty$,*

$$(1) \quad K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{1} \xrightarrow{P} 0, \text{ where } K = \sum_i k_i.$$

$$(2) \quad K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i - K^{-1} \sum_i \text{tr} \mathbf{D}_i \xrightarrow{P} 0.$$

(3) *If $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ik_i})'$ is another sequence of random vectors with properties (a) – (c) and \mathbf{X}_i and $\mathbf{Y}_{i'}$ are independent for all i and i' , then as $n \rightarrow \infty$, $K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{Y}_i \xrightarrow{P} 0$.*

Proof: The proofs of (1)-(3) all follow from Lemma B.1.

(1) Let $\mathbf{a}_i = \mathbf{D}_i \mathbf{1}$. Firstly,

$$\mathbb{E} \left(\frac{1}{K} \sum_i \mathbf{X}_i' \mathbf{a}_i \right) = \frac{1}{K} \sum_i (\mathbb{E} \mathbf{X}_i)' \mathbf{a}_i = 0.$$

Secondly, we need to show that

$$\text{Var}(K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{1}) \rightarrow 0.$$

as $n \rightarrow \infty$. This holds because

$$\begin{aligned} \text{Var}(K^{-1} \sum_i \mathbf{X}_i' \mathbf{a}_i) &= K^{-2} \sum_i \mathbf{a}_i' \text{Var}(\mathbf{X}_i) \mathbf{a}_i \\ &= K^{-2} \sum_{i,j} a_{ij}^2 \\ &\leq T^2 \sum_i k_i^3 / K^2. \end{aligned}$$

The inequality holds because, for the j th element of \mathbf{a}_i , $|a_{ij}| = |\sum_{j'} d_{ijj'}| \leq k_i M$. Moreover, $\sum_i k_i^3 \leq n M^3$ and $K^2 = (\sum_i k_i)^2 \geq n^2$ since $1 \leq k_i \leq M$. This implies that

$$\frac{\sum_i k_i^3}{K^2} \leq \frac{M^3}{n}.$$

As $n \rightarrow \infty$, $K \rightarrow \infty$ and $T^2 M^3 / K \rightarrow 0$ since T and M are finite.

(2) We have $\mathbb{E}(K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i) = K^{-1} \sum_i \text{tr} \mathbf{D}_i$ (see Appendix B.1). Next we will show that $\text{Var}(K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i) \rightarrow 0$ as $n \rightarrow \infty$. Since $\text{Var}(\mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i) = 2 \text{tr} \mathbf{D}_i^2$ (general results of expectations) and $\text{tr} \mathbf{D}_i^2 = \sum d_{ijj'}^2 \leq k_i^2 T^2$, we have

$$\text{Var}(K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i) = \frac{2}{K^2} \sum_i \text{tr} \mathbf{D}_i^2 \leq 2T^2 \frac{\sum_i k_i^2}{K^2}.$$

Similarly as in (1), we have

$$\frac{\sum_i k_i^2}{K^2} \leq \frac{M^2}{n}.$$

Therefore, as $n \rightarrow \infty$, $\text{Var}(K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i) \rightarrow 0$.

(3) Firstly, $\mathbb{E}(K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{Y}_i) = K^{-1} \sum_i \sum_{j,j'} d_{ijj'} \mathbb{E}(X_{ij} Y_{ij'}) = 0$. Also,

$$\begin{aligned} \text{Var}(\mathbf{X}_i' \mathbf{D}_i \mathbf{Y}_i) &= \sum_{j,j'} \sum_{l,l'} \text{Cov}(d_{ijj'} X_{ij} Y_{ij'}, d_{ill'} X_{il} Y_{il'}) \\ &= \sum_{j,j'} \sum_{l,l'} d_{ijj'} d_{ill'} \mathbb{E}(X_{ij} Y_{ij'} X_{il} Y_{il'}). \end{aligned}$$

Due to independence, $E(X_{ij}Y_{ij'}X_{il}Y_{il'}) = 0$ as $j \neq j'$ or $l \neq l'$ and $EX_{ij}^2Y_{ij'}^2 = EX_{ij}^2EY_{ij'}^2$. Therefore,

$$\begin{aligned}\text{Var}(\mathbf{X}_i' \mathbf{D}_i \mathbf{Y}_i) &= 2 \sum_{j,j'} d_{ijj'}^2 EX_{ij}^2 EY_{ij'}^2 \\ &= 2 \sum_{j,j'} d_{ijj'}^2 \text{Var}X_{ij} \text{Var}Y_{ij'} \\ &= 2 \sum_{j,j'} d_{ijj'}^2 \\ &= \text{Var}(\mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i).\end{aligned}$$

Since we have shown that $K^{-1} \sum_i \text{Var}(\mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i) \rightarrow 0$ as $n \rightarrow \infty$ in (2), it follows that

$$K^{-1} \sum_i \text{Var}(\mathbf{X}_i' \mathbf{D}_i \mathbf{Y}_i) \rightarrow 0.$$

The cases (1) and (2) can be extended to the case that \mathbf{X}_i has non-identity correlation matrix \mathbf{R}_i . Then (1) still holds and (2) becomes: (2') $K^{-1} \sum_i \mathbf{X}_i' \mathbf{D}_i \mathbf{X}_i - K^{-1} \sum_i \text{tr} \mathbf{D}_i \mathbf{R}_i \xrightarrow{P} 0$

Proof: Let $\mathbf{Z}_i = \mathbf{R}_i^{-1/2} \mathbf{X}_i$, then $\mathbf{X}_i = \mathbf{R}_i^{1/2} \mathbf{Z}_i$. The left hand side of (1) and (2) can be written as $K^{-1} \sum_i \mathbf{Z}_i' \mathbf{R}_i^{1/2} \mathbf{D}_i \mathbf{1}$ and $K^{-1} \sum_i \mathbf{Z}_i' \mathbf{R}_i^{1/2} \mathbf{D}_i \mathbf{R}_i^{1/2} \mathbf{Z}_i$, respectively. Since \mathbf{Z}_i has correlation matrix $\mathbf{I}_{k_i \times k_i}$, we can apply (1) and (2) and use the identity $\text{tr} \mathbf{R}_i^{1/2} \mathbf{D}_i \mathbf{R}_i^{1/2} = \text{tr} \mathbf{R}_i \mathbf{D}_i$ (see Appendix B.1).

B.3 Maple Code

1. $\text{Var}(\hat{\sigma}_{CL1}^2)$ and $\text{Var}(\hat{\rho}_{CL1})$ under exchangeable dependence and with family size fixed at k .

```
readlib(linalg):
with(linalg):
###MLE
ss:=2*s^4*(1+(k-1)*a^2)/k:
rr:=2*(1-a)^2*(1+(k-1)*a)^2/k/(k-1):
###CL1
#M: covariance matrix of the estimating functions
a2:=rh^4*(1-kk+kk^2)+rh^3*(4*kk-2*kk^2)+rh^2*(kk^2-kk+2)+2*kk*rh+1:
m:=matrix(2,2):
m[1,1]:=2*ss*ss*(kk+1)*(1+kk*rh*rh):
m[1,2]:=-(kk+1)*kk*(kk-1)*ss*ss*rh^2*(1-rh)^2:
m[2,1]:=m[1,2]:
m[2,2]:=(kk+1)*kk*(1-rh)^2*a2*ss*ss/2:
#D: expectations of derivatives of the estimating functions
# with respect to theta
d:=matrix(2,2):
d[1,1]:=kk+1:
```

```

d[1,2]:=0:
d[2,1]:=(kk+1)*kk*rh*(1-rh^2)/2:
d[2,2]:=-(kk+1)*kk*(1+rh*rh)*ss/2:
#Godambe information matrix
d1:=inverse(d):
d1:=map(simplify,d1):
tem:=multiply(d1,m):
tem:=map(simplify,tem):
gm:=multiply(tem,transpose(d1)):
gm:=map(simplify,gm):
gm:=map(factor,gm):
#var(ss)
gm[1,1]:
#var(rr)
gm[2,2]:
quit;

```

2. $ARE_{\hat{\sigma}_{CL1}^2}$, $ARE_{\hat{\rho}_{1,CL1}}$ and $ARE_{\hat{\rho}_{2,CL1}}$ for Type-3 family with fixed number of offspring k .

```

#efficiency: 1 parent k offspring
readlib(linalg):
with(linalg):
with(codegen,C):

###constants
e:=a/(k*a^2-1-(k-1)*b):
c:=1-k*a*e:
f:=(a*(1-a*e)+e)/(a*(k-1)*(b-1)):
f:=simplify(f):
d:=1-a*e-(k-1)*b*f:
d:=simplify(d):
bb:=d+(k-1)*f:
bb:=simplify(bb):

###Fisher's information matrix (mle)
mle:=matrix(3,3):
mle[1,1]:=(k+1)/(2*s^4):

mle[1,2]:=k*e/s^2:
mle[2,1]:=mle[1,2]:

mle[1,3]:=k*(k-1)*f/(2*s^2):
mle[3,1]:=mle[1,3]:

tr1:=k^2*e^2+k*c*bb:
tr2:=k*(c*bb+e^2*k):
mle[2,2]:=(tr1+tr2)/2:

mle[2,3]:=e*(k-1)*k*bb:
mle[3,2]:=mle[2,3]:

```

```

c1:=(k-1)*f:
c2:=d+(k-2)*f:
mle[3,3]:=k*(c1^2+(k-1)*c2^2)/2:

#asy covariance matrix of the MLE
map(simplify, mle);
mv:=inverse(mle):
mv:=map(simplify,mv):

###asy covariance matrix of the CL1 estimate
###D^{-1}%%M%%D^{-1}
##matrix D and D^{-1}
tem1:=1/(1-a^2):
tem2:=-a/(1-a^2):
da1:=diff(tem1, a):
da2:=diff(tem2, a):
tem1:=1/(1-b^2):
tem2:=-b/(1-b^2):
db1:=diff(tem1, b):
db2:=diff(tem2, b):

D1:=matrix(3,3):
D1[1,1]:=k+1:
D1[1,2]:=s^2*k*e*2:
D1[1,3]:=s^2*k*(k-1)*f:

D1[2,1]:=2*k*a/(1-a^2):
dda1:=diff(da1,a):
dda2:=diff(da2,a):
tem:=dda1+dda2*a:
D1[2,2]:=2*s^2*k*(-da2-tem):
D1[2,3]:=0:

D1[3,1]:=k*(k-1)*b/(1-b^2):
D1[3,2]:=0:
ddb1:=diff(db1,b):
ddb2:=diff(db2,b):
tem:=ddb1+ddb2*b:
D1[3,3]:=s^2*k*(k-1)*(-db2-tem):

map(simplify, D1);
DV:=inverse(D1):
DV:=map(simplify, DV);

##matrix M
M:=matrix(3,3):
M[1,1]:=k+1:
tem1:=k*(da1+a*da2):
tem2:=k*(da1+a*da2):
M[1,2]:=tem1+tem2:

```



```

c1:=1:
c2:=b:
c3:=db1*(k-1):
c4:=db2:
M[1,3]:=k*(c1*c3+(k-1)*c2*c4):

M[2,1]:=M[1,2]:
tem1:=k*(da1+da2*a):
tem2:=da2+a*da1:
tem3:=k*da1*a+da2*(1+(k-1)*b):
tem4:=a*da2:
tr1:=tem1^2+k*tem2*tem3:
c1:=tem4+da1:
c2:=tem4+da1*b:
tr2:=k*tem2*tem3+k*(c1^2+(k-1)*c2^2):
M[2,2]:=tr1+tr2:
tr1:=k*(k-1)*tem3*a*(db1+db2):
c3:=(k-1)*(db1+b*db2):
c4:=db2+b*(k-1)*db1+(k-2)*db2*b:
tr2:=k*(c1*c3+(k-1)*c2*c4):
M[2,3]:=tr1+tr2:

M[3,1]:=M[1,3]:
M[3,2]:=M[2,3]:
M[3,3]:=k*(c3^2+(k-1)*c4^2):
MV:=map(simplify, M);

##D^{-1}%%M%%D^{-1}
tem:=multiply(DV, MV):
tem:=map(simplify, tem):
dvt:=transpose(DV):
gm:=multiply(tem, dvt):

###AREs
sigma:=simplify(mv[1,1]/(gm[1,1]*2*s^4)):
b:=(k*a^2-1)/(k-1):
rh1:=simplify(mv[2,2]/(gm[2,2]*2*s^4));
rh2:=simplify(mv[3,3]/(gm[3,3]*2*s^4));
quit;

```

Bibliography

- Abel, L., Golmard, J., and Mallet, A. (1993). An autologistic model for the genetic analysis of familial binary data. *American Journal of Human Genetics*, 53: 894–907.
- Amos, C.I. (1988). *Robust Methods for Detection of Genetic Linkage for Data from Extended Families and Pedigree*. PhD thesis, Louisiana State University, New Orleans.
- Amos, C.I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *American Journal of Human Genetics*, 54: 535–543.
- Andrade, M., Amos, C., and Thiel, T. (1999). Methods to estimate genetic components of variance for quantitative traits in family studies. *Genetic Epidemiology*, 17: 64–76.
- Bandeem-Roche, K. and Liang, K.-Y. (1996). Modelling failure-time associations in data with multiple levels of clustering. *Biometrika*, 83: 29–39.
- Bonney, G.E. (1986). Regressive logistic models for familial disease and other binary traits. *Biometrics*, 42: 611–625.
- Cameron, A. and Trivedi, P. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.
- Chaganty, N.R. (1997). Loss in efficiency due to misspecification of the correlation structure in gee. In *Proceedings of the 51st session of the International Statistical Institute, Istanbul, Turkey*, 2: 127–128.
- Connolly, M. and Liang, K.-Y. (1988). Conditional logistic regression models for correlated binary data. *Biometrika*, 75: 501–506.

- Curriero, F. and Lele, S. (1999). A composite likelihood approach to semivariogram estimation. *Journal of Agricultural, Biological and Environmental Statistics*, 4: 9–28.
- Donner, A. and Koval, J.J. (1980). The estimation of intraclass correlation in the analysis of family data. *Biometrics*, 36: 19–25.
- Donner, A., Eliasziw, M. and Shoukri, M. (1998). Review of inference procedures for the interclass correlation coefficient with emphasis on applications to family studies. *Genetic Epidemiology*, 15: 627–646.
- Elston, R.C. and Stewart, J. (1971). A general model for the genetic analysis of pedigree data. *Human Heredity*, 21: 523–542.
- Falconer, D. (1965). The inheritance of liability to certain diseases, estimated from the incidences among relatives. *Annals of Human Genetics*, 29: 51–79.
- Falconer, D. (1967). The inheritance of liability to diseases with variable age of onset, with particular reference of diabetes mellitus. *Annal of Human Genetics*, 31: 1–20.
- Falconer, D. (1989). *Introduction to Quantitative Genetics*. Longman Scientific and Technical, Essex.
- Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*. 52: 399–433.
- Friedman, J., editor (1999). *Neurofibromatosis, Phenotype, Natural History, and Pathogenesis*. Johns Hopkins University Press, Baltimore, 3rd edition.
- George, E.O. and Bowman, D. (1995). Full likelihood procedure for analyzing exchangeable binary data. *Biometrics*, 51: 512–523.
- Godambe, V.P. (1991). *Estimating Functions*. Oxford University Press, New York.
- Haldane, J. (1932). *The Causes of Evolution*. Longmans, Green, London.
- Heagerty, P. and Lele, S. (1998). A composite likelihood approach to binary data in space. *Journal of the American Statistical Association*, 93: 1099–1111.

- Hunt, S., Hasstedt, S., and Williams, R. (1988). Testing for familial aggregation of a dichotomous trait. *Genetic Epidemiology*, 3: 299–312.
- Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, 46: 262–282.
- Joe, H. (1995). Approximation to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association*, 90: 957–964.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.
- Jöreskog, K.G. and Moustaki, I. (2001). Factor analysis of ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research*, 36: 347–387.
- Kempthorne, O. (1955a). The correlations between relatives in random mating populations. *Cold Spring Harbor Symposia Quantitative Biology*, 20: 60–75.
- Kempthorne, O. (1955b). The theoretical values of correlations between relatives in random mating populations. *Genetics*, 40: 153–167.
- King, T.M., Beaty, T.H., and Liang, K.-Y. (1996). Comparison of methods for survival analysis of dependent data. *Genetic Epidemiology*, 13: 139–158.
- Kleffe, J. (1993). Parent-offspring and sibling correlation estimation based on MINQUE theory. *Biometrika*, 80: 393–404.
- Korsgaard, I. and Andersen, A. (1998). The additive genetic gamma frailty model. *Scandinavian Journal of Statistics Theory and Applications*, 25: 255–269.
- Lange, K. (1978). Central limit theorems for pedigrees. *Journal of Mathematical Biology*, 6: 59–66.
- Lange, K. (1997). *Mathematical and Statistical Methods for Genetic Analysis*. Springer, New York.
- Lee, E., Wei, L.J., and Amato, D.A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, Klein, J.P., Liang, K.-Y., and Goel, P.K., editors, Kluwer Academic Dordrecht, 237–247.

- Lele, S. and Taper, M.L. (2002). A composite likelihood approach to (co)variance components estimation. *Journal of Statistical Planning and Inference*, 103: 117–135.
- Li, C.-C. (1976). *First Course in Population Genetics*. The Boxwood Press.
- Lindsay, B.G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80: 221–239.
- Manatunga, A.K. and Williamson, J.M. (2001). Assessing familial aggregation with an ordinal response. *Communications in Statistics: Theory and Methods*, 30: 627–641.
- Morton, N. (1974). Analysis of family resemblance. I. introduction. *American Journal of Human Genetics*, 26: 318–330.
- Morton, N. and Maclean, C. (1974). Analysis of family resemblance. III. complex segregation analysis of quantitative traits. *American Journal of Human Genetics*, 26: 489–503.
- Nash, J. (1990). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*. Hilger, New York.
- Ott, J. (1979). Maximum likelihood estimation by counting methods under polygenic and mixed models in human pedigrees. *American Journal of Human Genetics*, 31: 161–175.
- Parner, E.T. (2001). A composite likelihood approach to multivariate survival data. *Scandinavian Journal of Statistics*, 28: 295–302.
- Pickles, A., Crouchley, R., Simonoff, E., Eaves, L., Meyer, J., Rutter, M., Hewitt, J., and Silberg, J. (1994). Survival models for developmental genetic data: age of onset of puberty and antisocial behavior in twins. *Genetic Epidemiology*, 11: 155–170.
- Reboussin, B.A. and Liang, K.-Y. (1998). An estimating equations approach for the LISCOMP model. *Psychometrika*, 63(1): 165–182.
- Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components*. John Wiley & Sons, Inc., New York.
- Seber, G.A.F. (1977). *Linear Regression Analysis*. Griffin, London.

- Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.
- Song, P. (2000). Multivariate dispersion models generated from Gaussian copula *Scandinavian Journal of Statistics*, 27: 305-320.
- Spiekerman, C.F. and Lin, D.Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, 93: 1164-1175.
- Srivastava, M.S. and Ng, F. (1990). Estimation of intraclass correlation in regression models. *Gujarat Statistics Review, Professor Khatri Memorial Volume*, 229-236.
- Srivastava, M.S. and Katapa, R. (1986). Comparison of estimators of interclass and intraclass correlations from familial data. *The Canadian Journal of Statistics*, 14: 29-46.
- Srivastava, M.S. (1993). Estimation of the intraclass correlation coefficient. *Annals of Human Genetics*, 57: 159-65.
- Szudek, J., Birch, P., Riccardi, V.M., Evans, D., and Friedman, J. (2000). Association of clinical features in neurofibromatosis 1 (nfl). *Genetic Epidemiology*, 19: 429-439.
- Szudek, J., Joe, H. and Friedman, J. (2002). Analysis of intrafamilial phenotypic variation in neurofibromatosis 1 (nfl). *Genetic Epidemiology*, 23: 150-164.
- Taylor, H.M. (1994). *An Introduction to Stochastic Modeling*. Academic Press, In, San Diego.
- Tempelman, R.J. (1996). A mixed effects model for overdispersed count data in animal breeding. *Biometrics*, 52: 265-279.
- Tempelman, R.J. and Gianola, D. (1999). Genetic analysis of fertility in dairy cattle using negative binomial mixed models. *Journal of Dairy Science*, 82: 1834-1847.
- Tosteson, T., Rosner, B., and Redline, S. (1991). Logistic regression for clustered binary data in proband studies with application to familial aggregation of sleep disorders. *Biometrics*, 47: 1257-1265.

- Trégouët, D.-A., Ducimetière, P., Bocquet, V., Visvikis, S., Soubrier, F., and Tiret, L. (1999). A parametric copula model for analysis of familial binary data. *American Journal of Human Genetics*, 64: 886–893.
- Vaupel, J., Manton, K., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16: 439–454.
- Wei, L.J., Lin, D.Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association*, 84: 1065–1073.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20: 557–585.
- Xu, J.J. (1996). *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*. PhD thesis, Department of Statistics, University of British Columbia.
- Yashin, A. and Iachine, I. (1995). Survival of related individuals: an extension of some fundamental results of heterogeneity analysis. *Mathematical Population Studies*, 5: 321–340.
- Yau, K.K.W. and Lee, A.H. (2001). Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. *Statistics in Medicine*, 20: 2907–2920.
- Zhao, Y., Kumar, R., Baser, M., Evans, D., Wallace, A., Kluwe, L., Mautner, V., Parry, D., Rouleau, G., Joe, H., and Friedman, J. (2002). Intrafamilial correlation of clinical manifestations in neurofibromatosis. *Genetic Epidemiology*, 23: 245–259.