SOME STATISTICAL PROPERTIES OF MULTIVARIATE PROPER DISPERSION

MODELS, WITH SPECIAL REFERENCE TO A MULTIVARIATE GAMMA MODEL

by

Jeevanantham Rajeswaran

B.Sc., University of Jaffna, Jaffna, Sri Lanka, 1992.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Statistics

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 31, 1998

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

Department of _Statistics_

The University of British Columbia
Vancouver, Canada

Date _28th Aug, 1998_

Abstract

A broad class of error distributions for generalized linear models is provided by the class of dispersion models which was introduced by Jørgensen (1987a, 1997a) and a detailed study on dispersion models was made by Jørgensen (1997b). In this thesis we study multivariate proper dispersion models. Our aim is to do multivariate analysis for non-normal data, particularly data from the multivariate gamma distribution which is an example of a multivariate proper dispersion model, introduced by Jørgensen and Lauritzen (1998). This class provides a multivariate extension of the dispersion model density, following the spirit of the multivariate normal density.

We consider the saddlepoint approximation for small dispersion matrices, which, in turn, implies that the multivariate proper dispersion model is approximately multivariate normal for small dispersion matrices.

We want to mimic the basic technique of testing in multivariate normal, Hotelling's T^2. Our version of the T^2 test applies asymptotically, for either small dispersion or large samples.

We also consider estimating the normalizing constant of the bivariate gamma by Monte Carlo simulation and we investigate the marginal density by using numerical integration. We also investigate the distribution of the T^2-statistic by Monte Carlo simulation.

Contents

List of Tables

List of Figures

Acknowledgments

Chapter 1

Introduction

The multivariate normal distribution has been studied for more extensively than any other multivariate distribution. Indeed, its position of preeminence among multivariate continuous distribution is more marked than that of the normal univariate continuous distribution. However, in recent years there have been signs that the need for more usable alternatives to the multivariate normal distribution is becoming recognized.

Although the bivariate normal distribution had been studied at the beginning of the 19th century, interest in multivariate distributions remained at a low level until it was stimulated by the work of Galton in the last quarter of the century. He did not, himself, introduce any new form of distribution, but he developed the idea of correlation and regression and focussed attention on the need for greater knowledge of possible forms of multivariate distribution. Much multivariate theory has become practically useful only with the advent of the electronic computer. This stimulus has resulted in a rapid development of multivariate theory over the past two decades.

Investigation of non-normal distributions, or "skew frequency surfaces" (as non-symmetrical forms have been denoted) has generally followed lines suggested by previ-

ous work on univariate distributions. Early work in this fields followed rather different lines, but was not very successful. It is noteworthy that, among non-normal distributions, methods of estimation appear to have been developed less fully than one might expect. Karl Pearson (1923), whose first investigations appear to have been prompted by noting distinctly non-normal properties of some observed joint distributions, initially tried to proceed by an analogy with the bivariate normal surface. Later, Neyman (1926) also considered methods of construction of joint distributions, starting from certain requirements on the regression and scedastic functions. This was an extension of the work initiated by Jule (1897), who showed that assuming multiple linear regression, the multiple regression function obtained by the method of least squares is identical with that of a multinormal distribution.

The multivariate distributions were derived from other distributions through transformations, projections, convolutions, extreme values, mixing, compounding, truncating, and censoring. Then the marginal distributions of various statistics were derived from them. The study of multivariate distributions is concerned mainly with distributions of the continuous and discrete types. Some distributions arise from multidimensional central limit theorems, many serve as models for random experiments, and others are of interest primarily as derived distributions. Prominent among limit distributions and those from which others derive is the multinormal distribution. Some distributions derived from it are known to be the same for all parent distributions in a class containing symmetric multivariate stable distributions. These facts bear on robustness and the validity of normal-theory procedures for use with non-normal data, including multivariate analysis, multivariate analysis of variance and the multivariate Bartlett test.

Often multivariate data are scattered more heavily away from the center of location than are multinormal data. The introduction of a class of generalized linear models by

Nelder and Wedderburn (1972) paved the way to analyze a large variety of non-normal data by a simple general technique. Since there was no suitable class of multivariate non-normal distributions, this analysis is restricted to univariate data. Analysis of univariate non-normal data became simpler and accommodate more non-normal data types after the introduction of dispersion models, a broad class of error distributions for generalized linear models by Jørgensen (1987b,1997a).

Having in mind to find a class of multivariate distributions suitable to analyze the multivariate non-normal responses in a generalized linear model setup and to extend the existing multivariate techniques to multivariate non-normal data, Jørgensen and Lauritzen (1998) proposed a multivariate extension to the dispersion models, the so called *multivariate dispersion models*. Since the form of the distribution is similar the multivariate normal distribution this distribution can handle a regression setup. It would be noteworthy that, unlike the other non-normal multivariate distributions proposed earlier, the definition of multivariate dispersion models is geared towards good statistical properties rather than requiring specific moments or marginals.

As a first step towards these goals, we, in this thesis, consider a sub-class of multivariate dispersion models, the *multivariate proper dispersion models* and investigate some properties of the sub-class and propose a generalization of Hotelling's T^2-statistic.

In Chapter 2, we give a brief introduction to the univariate dispersion models of Jørgensen (1997b). In Chapter 3, we introduce the multivariate dispersion models and multivariate proper dispersion models. Specifically, we consider a multivariate gamma distribution which is an example of the multivariate proper dispersion models. Then we investigate the asymptotic behaviour of the multivariate proper dispersion models. In fact, we will show that multivariate proper dispersion models can be approximated by a multivariate normal distribution.

In chapter 4, we investigate the behaviour of the bivariate gamma distribution by observing the shape of the density plots and estimating the moments and marginal densities. In chapter 5, we generalize the Hotelling's T^2-statistic so that it can accommodate non-normal data.

Chapter 2

Univariate dispersion models

In this chapter, we give an introduction to the theory of the univariate dispersion models of Jørgensen (1997b), which is the basis of the entire theory that we develop in this thesis. First we give the definition of dispersion models, then we look at the two types of dispersion models: exponential dispersion models and proper dispersion models, and finally we discuss the saddlepoint approximations for the dispersion models.

2.1   Definition

Nelder and Wedderburn (1972) were the first to show, by introducing the class of generalized linear models, that a large variety of non-normal data may be analyzed by a simple general technique. Generalized linear models were originally developed for exponential families of distributions, but the main ideas were extended to a wider class of models called dispersion models. Dispersion models were originally introduced by Jørgensen (1987b) and the models were studied in detail by Jørgensen (1997b).

The class of dispersion models covers a comprehensive range of non-normal distributions, including distributions for the following seven basic data types, where S

6

denotes the support of the distribution:

- Data on the real line, $S = \mathcal{R}$.

- Positive data, $S = \mathcal{R}_+ = (0, 1)$.

- Positive data with zeros, $S = [0, \infty)$.

- Proportions, $S = (0, 1)$.

- Directions, $S = [0, 2\pi)$.

- Count data, $S = \{0, 1, 2, \ldots\}$.

- Count data with finite support, $S = \{0, \ldots, m\}$.

The main idea behind the dispersion model is that the notions of location and scale may be generalized to position and dispersion, respectively, for all the above seven data types.

Definition 2.1 *Let $\Omega \subseteq C \subseteq \mathcal{R}$ be intervals with Ω open. A function $d : C \times \Omega \to \mathcal{R}$ is called a* **unit deviance** *if it satisfies*

$$d(y; y) = d(\mu; \mu) = 0 \ \ \forall y, \ \mu \in \Omega \tag{2.1}$$

and

$$d(y; \mu) > 0 \ \ \forall y \neq \mu. \tag{2.2}$$

A unit deviance d is called **regular** *if $d(y; \mu)$ is twice continuously differentiable with respect to (y, μ) on $\Omega \times \Omega$ and satisfies*

$$\frac{\partial^2 d}{\partial \mu^2}(\mu; \mu) > 0 \ \ \forall \mu \in \Omega. \tag{2.3}$$

7

The **unit variance function** $V : \Omega \to \mathcal{R}_+$ *of a regular unit deviance is defined by*

$$V(\mu) = \frac{2}{\frac{\partial^2 d}{\partial \mu^2}(\mu; \mu)}. \qquad (2.4)$$

A univariate (regular) **dispersion model** $DM(\mu, \sigma^2)$ *with position parameter* $\mu \in \Omega$ *and dispersion parameter* $\sigma^2 > 0$ *is defined by a probability density function of the form*

$$f\left(y; \mu, \sigma^2\right) = a\left(y; \sigma^2\right) \exp\left\{-\frac{1}{2\sigma^2} d\left(y; \mu\right)\right\}, y \in C \qquad (2.5)$$

where $a \geq 0$ *is a suitable function and d is a regular unit deviance.*

The important characteristics of the density are:

1. The factor a in (2.5) is independent of the position parameter μ.

2. The densities are often unimodal; in general, the deviance d tends to make the density peak near μ, and the peak will become higher and sharper the smaller the value of σ^2.

As an example, consider the gamma distribution with shape parameter λ and scale parameter θ. The corresponding distribution is given by

$$f\left(y; \lambda, \theta\right) = \frac{\theta^\lambda y^{\lambda-1} e^{-\theta y}}{\Gamma(\lambda)}; y > 0, \ \lambda > 0, \ \theta > 0. \qquad (2.6)$$

The mean and the variance of Y are given by $E(Y) = \lambda/\theta$, and $Var(Y) = \lambda/\theta^2$ respectively. Now we write the above density function (2.6) as a dispersion model. Let $E(Y) = \mu$ and $\lambda = 1/\sigma^2$. Then we have, $Var(Y) = \mu^2\sigma^2$ and $\theta = 1/(\mu\sigma^2)$, where μ and σ^2 are position and dispersion parameters, respectively. Then the density is given by

$$\begin{aligned} f\left(y; \mu, \sigma^2\right) &= \frac{\lambda^\lambda}{\Gamma(\lambda)} y^{-1} e^{-\lambda} \exp\left\{-\frac{1}{2\sigma^2} d\left(y; \mu\right)\right\} \\ &= a\left(y; \sigma^2\right) \exp\left\{-\frac{1}{2\sigma^2} d\left(y; \mu\right)\right\}, \end{aligned} \qquad (2.7)$$
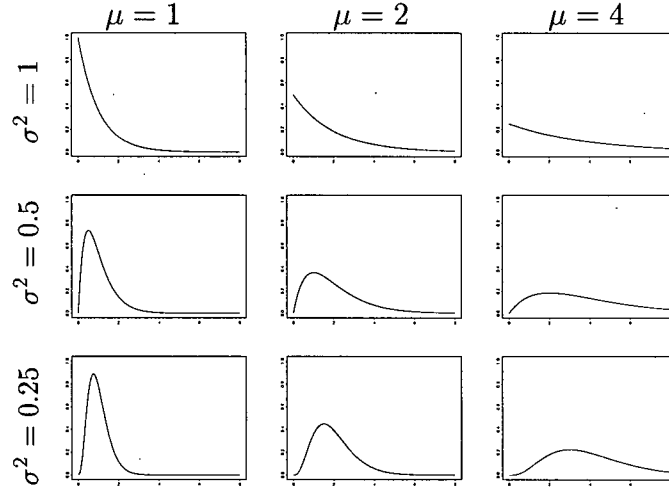
Figure 2.1: Some gamma densities.

where the regular unit deviance $d(y; \mu)$ is given by

$$d(y; \mu) = 2 \left\{ \frac{y}{\mu} - \log\left(\frac{y}{\mu}\right) - 1 \right\},$$ (2.8)

and the corresponding unit variance function is given by $V(\mu) = \mu^2$. Note that in the following we denote the a function as $c(y; \lambda)$ when we use λ instead of σ^2

For any regular unit deviance $d(y; \mu)$

$$\frac{\partial^2 d}{\partial y^2}(y; y) = \frac{\partial^2 d}{\partial \mu^2}(y; y) = -\frac{\partial^2 d}{\partial y \partial \mu}(y; y)$$ (2.9)

for details see Jørgensen (1997b, p. 24).

2.2   Exponential dispersion models

Exponential dispersion models were originally proposed by Tweedie (1947), and later reintroduced by Nelder and Wedderburn (1972) as a class of error distribution for generalized linear models.

9

A *reproductive exponential dispersion model* is defined by the following density for a random variable Y, with respect to a suitable measure,

$$f(y; \theta, \lambda) = c(y; \lambda) \exp[\lambda \{\theta y - \kappa(\theta)\}], y \in \mathcal{R}. \tag{2.10}$$

An *additive exponential dispersion model* for a random variable Z is defined by the probability density function of the following form, with respect to a suitable measure,

$$f^\star(z; \theta, \lambda) = c^\star(z; \lambda) \exp\{\theta z - \lambda \kappa(\theta)\}, z \in \mathcal{R}. \tag{2.11}$$

The two types are connected by the duality transformation $Z = \lambda Y$, so in principle, any exponential dispersion model has both an additive and a reproductive forms. The parameter θ is called the *canonical parameter* and λ is called the *index parameter*. The maximum possible domain for θ, denoted $\Theta \subseteq \mathcal{R}$ is called the *canonical parameter domain*. The domain for λ, called the *index set*, denoted by Λ, is a subset of $\mathcal{R}_+$. The function κ is called the *cumulant function*.

The reproductive form (2.10) is denoted by $Y \sim ED(\mu, \sigma^2)$, where $\mu = \tau(\theta) = \kappa'(\theta)$ is the mean and $\sigma^2 = \lambda^{-1}$ is the dispersion parameter. The variance of Y is $\sigma^2 V(\mu) = \sigma^2 \tau'\{\tau^{-1}(\mu)\}$, where V is the variance function.

The additive form (2.11) is denoted by $Z \sim ED^\star(\theta, \lambda)$, and the mean and the variance of Z are

$$E(Z) = \zeta = \lambda \tau(\theta) \text{ and } Var(Z) = \lambda V(\zeta/\lambda) = \lambda V(\mu). \tag{2.12}$$

We define the unit deviance for the above exponential dispersion models (see Jørgensen (1997b) for details) as follows,

$$d(y; \mu) = 2 \left[\sup_{\theta \in \Theta} \{y\theta - \kappa(\theta)\} - y\tau^{-1}(\mu) + \kappa \left\{ \tau^{-1}(\mu) \right\} \right]. \tag{2.13}$$

Then the density (2.10) can be rewritten in the form (2.5) in terms of d of (2.13), and

the a function is given by

$$a(y; \sigma^2) = c(y; \sigma^{-2}) \exp\left[\sigma^{-2} \sup_{\theta \in \Theta}\{y\theta - \kappa(\theta)\}\right].$$

This proves that the reproductive model (2.10) is indeed a dispersion model (2.5). However, the additive model (2.11) is not, in general, of the form (2.5).

Both types of exponential dispersion models satisfy a convolution property (see Jørgensen (1997b) for details) given as follows. Let $Y_1, \ldots, Y_n$ be independent and $Y_i \sim ED(\mu, \sigma^2/w_i)$ with weights w_i. Then

$$\frac{1}{w_+} \sum_{i=1}^{n} w_i Y_i \sim ED\left(\mu, \frac{\sigma^2}{w_+}\right) \tag{2.14}$$

where $w_+ = \sum_1^n w_i$.

If $Z_1, \ldots, Z_n$ are independent and $Z_i \sim ED^\star(\theta, \lambda_i)$, then

$$Z_+ = \sum_1^n Z_i \sim ED^\star(\theta, \lambda_1 + \cdots + \lambda_n). \tag{2.15}$$

Furthermore, we have $\Lambda = \mathcal{R}_+$ if and only if (2.11) is infinitely divisible (Jørgensen, 1997b). Note that the gamma distribution (2.6) can be written as

$$f(y; \lambda, \theta) = \frac{\lambda^\lambda y^{\lambda-1}}{\Gamma(\lambda)} \exp[\lambda\{(-\theta)y - (-\log(-\theta))\}], \tag{2.16}$$

which is of the form (2.10). Hence the gamma distribution is a reproductive exponential dispersion model.

One of the shortcomings of the exponential dispersion models is that there are no adequate exponential dispersion models available for data with bounded support, such as angles or proportions.

2.3   Proper dispersion models

In this section, we discuss another class of dispersion models, *proper dispersion models*, introduced by Jørgensen (1997a). The introduction of proper dispersion models pave

the way to accommodate distributions suitable for angles and proportions, namely the *von Mises* and *simplex distributions*, respectively.

Definition 2.2 *If d is a regular unit deviance with unit variance function V, a regular proper dispersion model, denoted by $PD(\mu, \sigma^2)$, is defined by*

$$f(y; \mu, \lambda) = c(\lambda) V^{-\frac{1}{2}}(y) \exp \left\{ -\frac{\lambda}{2} d(y; \mu) \right\}, y \in \Omega. \tag{2.17}$$

where $\lambda = 1/\sigma^2$. That is, the term a in (2.5) is factorized as $c(\lambda) V^{-\frac{1}{2}}(y)$.

A crucial property of this model is that the integral

$$\frac{1}{c(\lambda)} = \int_\Omega V^{-\frac{1}{2}}(y) \exp \left\{ -\frac{\lambda}{2} d(y; \mu) \right\} dy \tag{2.18}$$

does not depend on μ.

Since the term a in the gamma density (2.7) factorizes as $c(\lambda) V^{-\frac{1}{2}}(y)$ it is a proper dispersion model. Exactly three exponential dispersion models are also proper dispersion models, namely the normal, gamma and inverse Gaussian families. Other examples of proper dispersion models are the von Mises distribution and the simplex distribution of Barndorff-Nielsen and Jørgensen (1991). The von Mises distribution, for example, is given by

$$f\left(y; \mu, \sigma^2\right) = \frac{e^{\sigma^{-2}}}{2\pi I_0(\sigma^{-2})} \exp \left[-\frac{1}{2\sigma^2} 2\{1 - \cos(y - \mu)\} \right], \tag{2.19}$$

for $0 \le y < 2\pi$, where position parameter $\mu \in [0, 2\pi)$, dispersion parameter $\sigma^2 = 1/\lambda > 0$ and I_0 denotes the modified Bessel function given by

$$I_0(\lambda) = \int_0^{2\pi} \exp\left(\lambda \cos y\right) dy.$$

2.4 Saddlepoint approximation

In this section, we briefly discuss three methods of approximations that give an approximation to $a(y; \sigma^2)$.

2.4.1 Barndorff-Nielsen's $p^\star$-formula

Barndorff-Nielsen's $p^\star$-formula provides an approximation to the conditional distribution of the maximum likelihood estimator for a given statistical model, given an ancillary statistic. Here we apply the formula in the case of a single observation from a proper dispersion model (2.17) with λ known, in which case an ancillary statistic is degenerate, and the formula provides an approximation to the marginal distribution of the maximum likelihood estimator.

Since the maximum likelihood estimate of μ for λ fixed is y, Barndorff-Nielsen's formula hence provides an approximation to the marginal distribution of Y. The approximation is defined by

$$f(y; \mu, \lambda) \sim f_0(y; \mu, \lambda), \tag{2.20}$$

where f_0 is the renormalized saddlepoint approximation corresponding to unit deviance d, which we will discuss in the next section.

Now let us define the multivariate version (Barndorff-Nielsen and Cox, 1994, p.172) of the $p^\star$-formula.

Definition 2.3 *For a given ancillary statistic a, the log-likelihood can be considered as a function of μ and data. Let there be a one-to-one correspondence between the data $\mathbf{y}$ and $(\widehat{\mu}, a)$. Then we can write the log-likelihood ℓ as $\ell(\mu; \widehat{\mu}, a)$ and the normed form $\bar{\ell}$ as*

$$\bar{\ell}(\mu; \widehat{\mu}, a) = \ell(\mu; \widehat{\mu}, a) - \ell(\widehat{\mu}; \widehat{\mu}, a). \tag{2.21}$$

The $p^\star$-formula (or saddlepoint approximation) is defined by

$$p^\star(\widehat{\mu}; \mu | a) = (2\pi)^{-p/2} |\widehat{j}|^{1/2} e^{\bar{\ell}} \tag{2.22}$$

where, $|\widehat{j}|$ is the determinant of the observed information matrix for μ evaluated at $\mu = \widehat{\mu}$.

13

This formula (renormalized) determines a probability density, with respect to Lebesgue measure, if the range space of $\hat{\boldsymbol{\mu}}$ is an open subset of $\mathcal{R}^p$. Note that we deal with continuous distribution. The question of Lebesgue measure will not arise in our problem. In fact, formula (2.22) is called the $p^{\dagger}$-formula by Barndorff-Nielsen and Cox (1994). As compared to $p^{\dagger}$, the $p^{\star}$-formula which is originally defined as $c|\hat{j}|^{1/2}e^{\bar{\ell}}$, gives at least one degree higher accuracy but at the cost of having to calculate the normalizing constant c. So it is feasible to use $p^{\dagger}$-formula and we still call this formula the $p^{\star}$-formula. We use the formula (2.22) in the Section 3.3.1 to develop an asymptotic distribution for multivariate proper dispersion models..

2.4.2   Saddlepoint and Laplace approximations

In this section, we briefly discuss three methods of approximations that give an approximation to $a(y; \sigma^2)$. We now briefly discuss the saddlepoint approximation for dispersion models; a detailed discussion can be found in Jørgensen (1997b) . The saddlepoint approximation is important for the asymptotic theory of generalized linear models.

The *saddlepoint approximation* for a dispersion model with regular unit deviance d is defined for $y \in \Omega$ by

$$f\left(y; \mu, \sigma^2\right) \sim \{2\pi\sigma^2 V(y)\}^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}d\left(y; \mu\right)\right\} \text{ as } \sigma^2 \to 0, \qquad (2.23)$$

meaning that the ratio of the two sides goes to 1 as $\sigma^2 \to 0$. The saddlepoint approximation is exact in a few very special cases, such as the normal and simplex distributions.

The saddlepoint approximation may be interpreted as being half way to a normal approximation for the density. Thus, if we replace $V(y)$ by $V(\mu)$ in (2.23) and introduce a quadratic approximation to the unit deviance in the exponent of the density,

we get the normal approximation.

Note that the saddlepoint approximation on the right hand side of (2.23), while positive, is not in general a density function on Ω. However, it may be rescaled to become a density function.

The corresponding renormalized saddlepoint approximation is the density function defined for $y \in \Omega$ by

$$f_0\left(y; \mu, \sigma^2\right) = a_0\left(\mu, \sigma^2\right) V^{-1/2}(y) \exp\left\{-\frac{1}{2\sigma^2} d\left(y; \mu\right)\right\}, \qquad (2.24)$$

where $a_0\left(\mu, \sigma^2\right)$ is a normalizing constant defined by

$$\frac{1}{a_0\left(\mu, \sigma^2\right)} = \int_\Omega V^{-1/2}(y) \exp\left\{-\frac{1}{2\sigma^2} d\left(y; \mu\right)\right\} dy. \qquad (2.25)$$

Note that, in the case of proper dispersion models, a_0 in (2.25) does not depend on μ.

We now present the important Laplace approximation, as described in, for example, Barndorff-Nielsen and Cox (1989, p.59).

Proposition 2.1 (Laplace approximation) *Define*

$$I(\lambda) = \int_\Omega b(y) e^{\lambda t(y)} dy,$$

where Ω is an open interval. Suppose that b is positive, and continuous at $y = \mu$, that t is twice differentiable, has global maximum for $y = \mu \in \Omega$ and that

$$K(\mu) = -t''(\mu) > 0.$$

Then

$$I(\lambda) \sim \sqrt{\frac{2\pi}{\lambda K(\mu)}} b(\mu) e^{\lambda t(\mu)} \text{ as } \lambda \to \infty. \qquad (2.26)$$

15

The ordinary and renormalized saddlepoint approximations agree asymptotically for σ^2 small; this can be shown by using the Laplace approximation. We now present this important saddlepoint approximation result, for a proof see Jørgensen (1997b, p. 28).

Proposition 2.2 *Let d be a given regular unit deviance defined on $C \times \Omega$ and define $a_0\left(\mu, \sigma^2\right)$ by (2.25). Then, as $\sigma^2 \to 0$,*

$$a_0\left(\mu, \sigma^2\right) \sim \left(2\pi\sigma^2\right)^{-1/2}.$$

The essence of the saddlepoint approximation is the approximation of the function $a(y; \sigma^2)$. Specifically, the saddlepoint approximation is equivalent to

$$\sigma a(y; \sigma^2) \to \{2\pi V(y)\}^{\frac{1}{2}} \text{ as } \sigma^2 \to 0. \tag{2.27}$$

Definition 2.4 *The saddlepoint approximation is said to be* **uniform on compacts** *if the convergence in (2.27) is uniform in y on compact subsets of Ω.*

Proposition 2.3 *In the case of a proper dispersion model, the saddlepoint approximation is uniform on compacts, and the renormalized saddlepoint approximation is exact.*

The proof of this proposition contains the following arguments: For proper dispersion model we have $a_0\left(\mu, \sigma^2\right) = a(\sigma^2)$ making the renormalized saddlepoint approximation exact. The continuity of the unit variance function V implies that the saddlepoint approximation is uniform on compacts. For details, see Jørgensen (1997b).

Chapter 3

Multivariate dispersion models

In this chapter, we discuss multivariate dispersion models. In the first section, we discuss the definition of multivariate dispersion models. Basically, in the first section, we discuss the multivariate dispersion models defined by Jørgensen and Lauritzen (1998) which opened a new field in statistics that has a lot of problems to investigate. It is the basis for this whole thesis. In subsequent sections we develop and discuss some statistical properties of the multivariate proper dispersion models and in particular the multivariate gamma distribution. Note that the order in which we discuss the multivariate dispersion models is related to its univariate counterpart in the preceding section.

3.1 Definition

A class of multivariate dispersion models suitable as error distributions for generalized linear models with multivariate non-normal responses was introduced by Jørgensen and Lauritzen (1998). The main concern, in their paper, is with suitable distributions for generalized linear models with a multivariate response vector. The definition of multivariate dispersion models is geared towards good statistical properties similar to

that are possessed by multivariate normal distribution, rather than requiring specific moments or marginals.

Let the sample space be an open subset of $\mathcal{R}^p$. A *multivariate dispersion model* is defined by the following probability density function on $\mathcal{R}^p$:

$$
\begin{aligned}
f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= a(\mathbf{y}; \boldsymbol{\Sigma}) \exp\left\{ -\frac{1}{2} t^\top (\mathbf{y}; \boldsymbol{\mu}) \, \boldsymbol{\Sigma}^{-1} t (\mathbf{y}; \boldsymbol{\mu}) \right\} \\
&= a(\mathbf{y}; \boldsymbol{\Sigma}) \exp\left[-\frac{1}{2} \mathrm{tr} \left\{ \boldsymbol{\Sigma}^{-1} t (\mathbf{y}; \boldsymbol{\mu}) \, t^\top (\mathbf{y}; \boldsymbol{\mu}) \right\} \right]
\end{aligned}
\tag{3.1}
$$

where $\boldsymbol{\mu} \in \Omega$ (an open region in $\mathcal{R}^p$), $\boldsymbol{\Sigma}$ is a symmetric positive-definite $p \times p$ matrix and $t(\mathbf{y}; \boldsymbol{\mu})$ is suitably defined vector of deviance residuals satisfying $t(\boldsymbol{\mu}; \boldsymbol{\mu}) = 0$ for $\boldsymbol{\mu} \in \Omega$. The parameter $\boldsymbol{\mu}$, called the *position vector*, and $\boldsymbol{\Sigma}$, called the *dispersion matrix*, may be interpreted as analogies of respectively the mean vector and variance-covariance matrix of the multivariate normal distribution.

As an example, Jørgensen and Lauritzen (1998) defined the *multivariate Laplace distribution* by the following density with respect to Lebesgue measure on $\mathcal{R}^p$:

$$
f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = a(\boldsymbol{\Sigma}) \exp\left\{ -\frac{1}{2} t^\top (\mathbf{y} - \boldsymbol{\mu}) \, \boldsymbol{\Sigma}^{-1} t (\mathbf{y} - \boldsymbol{\mu}) \right\},
$$

where $a(\boldsymbol{\Sigma})$ is a normalizing constant, t is defined by

$$
t(\mathbf{y} - \boldsymbol{\mu}) = \left(\pm\sqrt{|y_1 - \mu_1|}, \ldots, \pm\sqrt{|y_p - \mu_p|} \right)^\top,
\tag{3.2}
$$

and $\pm = \mathrm{sgn}(y_j - \mu_j)$ denotes the sign of $y_j - \mu_j$ for each j. Note that when the dispersion matrix $\boldsymbol{\Sigma}$ is diagonal, the components of $\mathbf{Y}$ follow independent univariate Laplace distributions.

For a detailed discussion on generalization of the definition (3.1) and geometric construction of multivariate dispersion models refer to Jørgensen and Lauritzen (1998).

3.2 Multivariate proper dispersion models

Following the spirit of the univariate proper dispersion models Jørgensen and Lauritzen (1998) called (3.1) a *multivariate proper dispersion models* if $a(\mathbf{y}; \boldsymbol{\Sigma})$ in (3.1) factorizes as $a(\boldsymbol{\Sigma})b(\mathbf{y})$ for suitable functions a and b. That is, we define the *multivariate proper dispersion models* as follows

$$f(\mathbf{y}; \boldsymbol{\mu}; \boldsymbol{\Sigma}) = a_p(\boldsymbol{\Sigma})|\mathbf{V}(\boldsymbol{\mu})|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}t^\top(\mathbf{y}; \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}t(\mathbf{y}; \boldsymbol{\mu})\right\}, \qquad (3.3)$$

where $\mathbf{V}(\boldsymbol{\mu}) = \mathrm{diag}\{V(\mu_1), \ldots, V(\mu_p)\}$. The distribution (3.3) is denoted by $PD_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In the following we often use $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ instead of $\boldsymbol{\Sigma}^{-1}$.

As an example, consider the von Mises distribution given by (2.19). Define the deviance residual by

$$r(y; \mu) = \pm\sqrt{2\{1 - \cos(y - \mu)\}},$$

where the sign $\pm$ denotes the sign of $\sin(y - \mu)$. Since $r(y; \mu)$ is a pivot, a function of Y and μ whose distribution under μ does not depend on the value of μ, the standard construction proposed by Jørgensen and Lauritzen (1998) leads to a multivariate von Mises distribution of the form

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = a(\boldsymbol{\Sigma})\exp\left[-\frac{1}{2}\{t^\top(\mathbf{y}; \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}t(\mathbf{y}; \boldsymbol{\mu})\}\right],$$

on $[0, 2\pi)^p$, where $t(\mathbf{y}; \boldsymbol{\mu})$ is the vector of deviance residuals. This is clearly a multivariate proper dispersion model.

Jørgensen and Lauritzen (1998) discussed a standard construction that takes almost any univariate proper dispersion model and gives a multivariate proper dispersion model as a result. They noted, however, that this standard construction does not work for the Inverse Gaussian distribution which is a univariate proper dispersion model.

3.2.1 Multivariate gamma distribution

In this section, we construct the multivariate gamma distribution which was described by Jørgensen and Lauritzen (1998). Consider the univariate gamma density (2.7)

$$f(y; \mu, \lambda) = c_1(\lambda) y^{-1} \exp\left\{-\frac{\lambda}{2} d(y; \mu)\right\}, y > 0 \qquad (3.4)$$

where the regular unit deviance d is given by (2.8), the unit variance function $V(\mu) = \mu^2$ and

$$c_1(\lambda) = \frac{\lambda^\lambda e^{-\lambda}}{\Gamma(\lambda)}.$$

Note that the geometric measure is $\nu(dy) = V^{-1/2}(y) dy = y^{-1} dy$, where the geometric measure is defined as follows, see Jørgensen and Lauritzen (1998) for more details. Suppose $\mathbf{G_\Lambda}(y) = \{g_\Lambda^{jk}(y)\}$ is a matrix with entries

$$g_\Lambda^{jk}(y) = \frac{1}{2} \frac{\partial^2}{\partial \mu_j \mu_k} \mathrm{tr}\{\Lambda T(\mathbf{y}; \boldsymbol{\mu})\}\Big|_{\mu = y}$$

then the geometric measure is defined by

$$\nu_\Lambda(dy) = |\mathbf{G_\Lambda}(\mathbf{y})|^{1/2} dy.$$

The deviance residual is

$$r(y; \mu) = \pm\sqrt{d(y; \mu)} = \pm\sqrt{2\{\frac{y}{\mu} - \log\frac{y}{\mu} - 1\}},$$

where $\pm$ denotes the sign of $(y - \mu)$. Clearly $r(Y; \mu)$ is a pivot. For more details about pivot refer to Jørgensen (1997b). Let

$$t(\mathbf{y}; \boldsymbol{\mu}) = \{r(y_1; \mu_1), \ldots, r(y_p; \mu_p)\}^\top$$

denote the vector of residuals for a p-vector of data. Then the yoke

$$T(\mathbf{Y}; \boldsymbol{\mu}) = t(\mathbf{Y}; \boldsymbol{\mu}) t^\top(\mathbf{Y}; \boldsymbol{\mu}) = \left\{\begin{array}{c} r(Y_1; \mu_1) \\ \vdots \\ r(Y_p; \mu_p) \end{array}\right\} \{r(Y_1; \mu_1), \ldots, r(Y_p; \mu_p)\} \qquad (3.5)$$

is also a pivot with respect to the product measure

$$
\begin{aligned}
\nu(dy_1, \ldots, dy_p) &= \nu(dy_1) \otimes \cdots \otimes \nu(dy_p) \\
&= \prod_{j=1}^{p} \left\{ V^{-1/2}(y_j) dy_j \right\} \\
&= \prod_{j=1}^{p} \left\{ y_j^{-1} dy_j \right\}.
\end{aligned}
\tag{3.6}
$$

A multivariate version of the univariate gamma therefore appears as

$$
f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = a_p(\boldsymbol{\Sigma})(y_1 \ldots y_p)^{-1} \exp\left[-\frac{1}{2} \text{tr}\{\boldsymbol{\Sigma}^{-1} T(\mathbf{y}; \boldsymbol{\mu})\} \right].
\tag{3.7}
$$

We denote this model $Ga_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The new model (3.7) is clearly a multivariate proper dispersion model and in the case of $\boldsymbol{\Lambda}$ being diagonal, it has gamma marginals distributed independently. So we call this model a *multivariate gamma distribution*. The main problem in this multivariate model is determining the normalizing constant $a_p(\boldsymbol{\Lambda})$. This normalizer can be estimated by numerical integration or by Monte Carlo simulation. We defer the discussion of this problem until Chapter 4.

This multivariate gamma distribution seems to be new and different from other multivariate gamma distributions proposed in the literature cf. Krishnaiah (1985). In particular, its marginals are not gamma except when $\boldsymbol{\Lambda}$ is diagonal. Several procedures exist for constructing a multivariate gamma distribution given that the marginal distributions are gamma. For example Johnson and Kotz (1972, p. 216) proceeds as follows: Suppose that $X_0, X_1, \ldots, X_m$ are independent random variables and that X_j has a standard gamma distribution. The joint distribution of $Y_j = X_0 + X_j$ $(j = 1, 2, \ldots, m)$ is called a multivariate gamma distribution and the marginal of Y_j is a standard gamma distribution.

Krishnamoorthy and Parthasarathy (1951) defined a multivariate gamma as follows: Let $\mathbf{X}_j^\top = (X_{1j}, \ldots, X_{pj})$, $(j = 1, 2, \ldots, n)$ be distributed independently and identically as a multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Also, let $Z_i = \frac{1}{2} \sum_{j=1}^n X_{ij}^2$ for $i = 1, 2, \ldots, p$. Then the joint distribution of $(Z_1, \ldots, Z_p)$ is a *central multivariate gamma distribution* with $n/2$ as shape parameter and $\boldsymbol{\Sigma}$ as the covariance parameter.

The main application of the proposed multivariate gamma distributions has so far been in the study of reliability. Since the forms of the densities are quite complicated, it is difficult to accommodate a multivariate regression set up. The form of multivariate gamma model, in general multivariate proper dispersion model, proposed by Jørgensen and Lauritzen (1998) is very similar to that of a multivariate normal model. This will enable us to accommodate a multivariate regression setup. The problems related to the inference on the position vector is very much similar to that of a multivariate normal model.

We now introduce some properties of the multivariate gamma distribution.

Theorem 3.1 *The Multivariate gamma model is closed with respect to scale transformations.*

Proof: Let $\mathbf{Y} \sim Ga_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{C}$ be a diagonal matrix with elements $c_1, \ldots, c_p$, $c_i \neq 0 \; \forall i$. Consider the transformation $\mathbf{X} = \mathbf{CY}$. The Jacobian matrix is given by

$$\mathbf{J} = \frac{\partial \mathbf{X}}{\partial \mathbf{Y}^\top} = \mathbf{C}.$$

Hence the density is given by

$$
\begin{aligned}
g(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= |J|^{-1} f(\mathbf{C}^{-1}\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= |\mathbf{C}|^{-1} a_p(\boldsymbol{\Sigma}) \left(\frac{x_1}{c_1} \ldots \frac{x_p}{c_p} \right)^{-1} \exp\left\{ -\frac{1}{2} t^\top (\mathbf{C}^{-1}\mathbf{x}; \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} t(\mathbf{C}^{-1}\mathbf{x}; \boldsymbol{\mu}) \right\}
\end{aligned}
$$

$$= a_p(\mathbf{\Sigma})(x_1 \dots x_p)^{-1} \exp\left\{-\frac{1}{2} t^\top(\mathbf{C}^{-1}\mathbf{x}; \boldsymbol{\mu}) \mathbf{\Sigma}^{-1} t(\mathbf{C}^{-1}\mathbf{x}; \boldsymbol{\mu})\right\}. \qquad (3.8)$$

Now let us consider the deviance

$$\begin{aligned}
d\left(\frac{x}{c}; \mu\right) &= 2\left(\frac{x/c}{\mu} - \log\frac{x/c}{\mu} - 1\right) \\
&= 2\left(\frac{x}{c\mu} - \log\frac{x}{c\mu} - 1\right) \\
&= d(x; c\mu) \qquad (3.9)
\end{aligned}$$

Hence, it follows from (3.9) that $t(\mathbf{C}^{-1}\mathbf{x}; \boldsymbol{\mu}) = t(\mathbf{x}; \mathbf{C}\boldsymbol{\mu})$. Hence we may write (3.8) as

$$g(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma}) = a_p(\mathbf{\Sigma})(x_1 \dots x_p)^{-1} \exp\left\{-\frac{1}{2} t^\top(\mathbf{x}; \mathbf{C}\boldsymbol{\mu}) \mathbf{\Sigma}^{-1} t(\mathbf{x}; \mathbf{C}\boldsymbol{\mu})\right\}.$$

Therefore, we have $\mathbf{X} \sim Ga_p(\mathbf{C}\boldsymbol{\mu}, \mathbf{\Sigma})$. Hence the theorem. $\square$

3.2.2 Maximum likelihood estimate of the position vector $\boldsymbol{\mu}$

One of the main advantages of the multivariate dispersion model defined by Jørgensen and Lauritzen (1998), as in the univariate dispersion models defined by Jørgensen (1997b), is that the normalizer $a_p(\mathbf{\Sigma})$ does not involve the position vector $\boldsymbol{\mu}$ and therefore it allows us to make inferences about $\boldsymbol{\mu}$ without knowing the exact formula for $a_p(\mathbf{\Sigma})$.

From now on, in this chapter, we use quite a number of techniques for matrix derivatives. We give an overview of these formulas in *Appendix A.4*. For more details cf. Dwyer (1962).

Theorem 3.2 *Let* $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ *be* n *random vectors, in* $\mathcal{R}^p$, *from a multivariate proper dispersion model* $PD_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. *Then, for arbitrary* $\mathbf{\Sigma}$, *the score function for*

μ_j *based on the full sample is*

$$\frac{\partial \ell\left(\boldsymbol{\mu}\right)}{\partial \mu_j} = -\frac{1}{2}\left\{\sum_{k=1}^{p}\lambda_{jk}\sum_{i=1}^{n}\frac{d'(Y_{ij};\mu_j)}{r(Y_{ij};\mu_j)}r(Y_{ik};\mu_k)\right\} \qquad (3.10)$$

where $\lambda_{jk} = (\boldsymbol{\Lambda})_{jk}$ and $d'(y;\mu)$ denotes the derivative of d with respect to μ.

Proof. The log-likelihood function of $\boldsymbol{\mu}$ is given by

$$\ell\left(\boldsymbol{\mu}\right) = n\log a_p(\boldsymbol{\Lambda}) + \prod_{i=1}^{n}\log b(\mathbf{y}_i) - \frac{1}{2}\sum_{i=1}^{n}t^{\top}\left(\mathbf{y}_i;\boldsymbol{\mu}\right)\boldsymbol{\Lambda}t\left(\mathbf{y}_i;\boldsymbol{\mu}\right). \qquad (3.11)$$

Differentiating (3.11) with respect to $\boldsymbol{\mu}$ gives

$$
\begin{aligned}
\frac{\partial \ell\left(\boldsymbol{\mu}\right)}{\partial \boldsymbol{\mu}} &= -\frac{1}{2}\sum_{i=1}^{n}\frac{\partial t^{\top}\left(\mathbf{Y}_i;\boldsymbol{\mu}\right)\boldsymbol{\Lambda}t\left(\mathbf{Y}_i;\boldsymbol{\mu}\right)}{\partial \boldsymbol{\mu}} \\
&= -\frac{1}{2}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\{\frac{\partial t^{\top}\left(\mathbf{Y}_i;\boldsymbol{\mu}\right)\boldsymbol{\Lambda}t\left(\mathbf{Y}_i;\boldsymbol{\mu}\right)}{\partial t\left(\mathbf{Y}_i;\boldsymbol{\mu}\right)}\right\}_m\frac{\partial\{t\left(\mathbf{Y}_i;\boldsymbol{\mu}\right)\}_m}{\partial \boldsymbol{\mu}} \\
&= -\sum_{i=1}^{n}\sum_{m=1}^{p}\left\{\boldsymbol{\Lambda}t\left(\mathbf{Y}_i;\boldsymbol{\mu}\right)\right\}_m\frac{\partial\{t\left(\mathbf{Y};\boldsymbol{\mu}\right)\}_m}{\partial \boldsymbol{\mu}} \\
&= -\frac{1}{2}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\{\boldsymbol{\Lambda}t\left(\mathbf{Y}_i;\boldsymbol{\mu}\right)\right\}_m\mathbf{D}'(\mathbf{Y}_i;\boldsymbol{\mu})\mathbf{e}_m, \qquad (3.12)
\end{aligned}
$$

where $\mathbf{e}_m$ is a $p \times 1$ vector with 1 at the mth position and zero elsewhere and we define

$$\mathbf{D}'(\mathbf{Y};\boldsymbol{\mu}) = \text{diag}\left\{\frac{d'(Y_1;\mu_1)}{r(Y_1;\mu_1)},\ldots,\frac{d'(Y_p;\mu_p)}{r(Y_p;\mu_p)}\right\}. \qquad (3.13)$$

Therefore, from (3.12) it follows that

$$\frac{\partial \ell(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = -\frac{1}{2}\sum_{i=1}^{n}\mathbf{D}'(\mathbf{Y}_i;\boldsymbol{\mu})\boldsymbol{\Lambda}t\left(\mathbf{Y}_i;\boldsymbol{\mu}\right). \qquad (3.14)$$

Matrix multiplication of diagonal matrix $\mathbf{D}'$ with $\boldsymbol{\Lambda}$ gives

$$\lambda_{jk}\frac{d'(Y_{ij};\mu_j)}{r(Y_{ij};\mu_j)}$$

24

as the $(j,k)^{th}$ element. Now the multiplication of the resulting matrix from the above with the vector $t(\mathbf{Y}_i; \boldsymbol{\mu})$ gives a vector with

$$\sum_{k=1}^{p} \lambda_{jk} \frac{d'(Y_{ij}; \mu_j)}{r(Y_{ij}; \mu_j)} r(Y_{ik}; \mu_k)$$

as j^{th} element. Now (3.10) follows immediately from (3.14). $\square$

Lemma 3.1 *For a multivariate gamma model, for arbitrary $\boldsymbol{\Sigma}$, the score function for μ_j based on a full sample given by*

$$\frac{\partial \ell(\boldsymbol{\mu})}{\partial \mu_j} = \sum_{k=1}^{p} \lambda_{jk} \sum_{i=1}^{n} \frac{Y_{ij} - \mu_j}{\mu_j^2 r(Y_{ij}; \mu_j)} r(Y_{ik}; \mu_k) \tag{3.15}$$

Proof. For a multivariate gamma model the $\mathbf{D}'(\mathbf{y}_i; \boldsymbol{\mu})$ in (3.12) can be written as

$$\mathbf{D}'(\mathbf{Y}_i; \boldsymbol{\mu}) = -2\boldsymbol{\Delta}(\mathbf{Y_i}, \boldsymbol{\mu}) \mathbf{V}^{-1}(\boldsymbol{\mu})$$

where

$$\boldsymbol{\Delta}(\mathbf{Y}, \boldsymbol{\mu}) = \text{diag}\left\{\frac{Y_1 - \mu_1}{r(Y_1; \mu_1)}, \ldots, \frac{Y_p - \mu_p}{r(Y_p; \mu_p)}\right\},$$

and

$$\mathbf{V}(\boldsymbol{\mu}) = \text{diag}\left\{\mu_1^2, \ldots, \mu_p^2\right\}$$

Now by using the fact that $d(y; \mu) = 2(\frac{y}{\mu} - \log\frac{y}{\mu} - 1)$ and by using Theorem 3.2, it is easy to show that (3.15) holds. $\square$

Note that, if we make a small-dispersion approximation to d' at $Y = \mu$, we have (Jørgensen, 1997 p. 25),

$$d'(Y; \mu) \approx 2r(Y; \mu)V^{-1/2}(\mu). \tag{3.16}$$

Thus,

$$\mathbf{D}'(\mathbf{Y}; \boldsymbol{\mu}) \approx 2\mathbf{V}^{-1/2}(\boldsymbol{\mu}). \tag{3.17}$$

We will use the approximations (3.16) and (3.17) frequently in the next section.

There is no closed form solution to $\frac{\partial \ell(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = 0$ except for the case where we have a single observation vector. But this can be solved by numerical techniques; for example, we may use optimization technique such as quasi-Newton to find numerical values.

Lemma 3.2 *Let* $\mathbf{Y}$ *be a random vector, in* $\mathcal{R}^p$, *from a multivariate proper dispersion model* $PD_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Then, for arbitrary* $\boldsymbol{\Sigma}$, *the maximum likelihood estimator of* $\boldsymbol{\mu}$ *is* $\mathbf{y}$.

Proof. We prove this by the following arguments. The log-likelihood function of $\boldsymbol{\mu}$ is given by

$$\ell(\boldsymbol{\mu}) = \log a_p(\boldsymbol{\Lambda}) + \log b(\mathbf{y}) - \frac{1}{2}t^{\top}(\mathbf{y};\boldsymbol{\mu})\,\boldsymbol{\Lambda}\,t(\mathbf{y};\boldsymbol{\mu}). \tag{3.18}$$

Let $t = t(\mathbf{y};\boldsymbol{\mu})$. Since $\boldsymbol{\Sigma}$ is positive-definite so is $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda}$. Hence $t^{\top}\boldsymbol{\Lambda}t \geq 0$ and clearly we have $t(\mathbf{y};\mathbf{y}) = 0$. Now if $\mathbf{y} \neq \boldsymbol{\mu}$ we have $t(\mathbf{y};\boldsymbol{\mu}) \neq 0$ and hence $t^{\top}\boldsymbol{\Lambda}t > 0$ again by $\boldsymbol{\Lambda}$ being positive-definitive. Hence, $t^{\top}\boldsymbol{\Lambda}t$ is minimum for $\boldsymbol{\mu} = \mathbf{y}$. Therefore it follows from (3.18) that the maximum likelihood estimator of $\boldsymbol{\mu}$ is $\mathbf{y}$. $\square$

We now give the second derivative of the log likelihood (3.18) with respect to $\boldsymbol{\mu}$, which will be used in the next section

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\mu})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^{\top}} &= -\frac{1}{2}\left[\frac{\partial \mathbf{D}'(\mathbf{y};\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^{\top}}\left\{\boldsymbol{\Sigma}^{-1}t(\mathbf{y};\boldsymbol{\mu}) \otimes \mathbf{I}_p\right\} + \left\{\mathbf{D}'(\mathbf{y};\boldsymbol{\mu}) \otimes \mathbf{I}_1\right\}\frac{\partial \boldsymbol{\Sigma}^{-1}t(\mathbf{y};\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^{\top}}\right] \\
&= -\frac{1}{2}\left[\frac{\partial \mathbf{D}'(\mathbf{y};\boldsymbol{\mu})}{\partial \boldsymbol{\mu}^{\top}}\left\{\boldsymbol{\Sigma}^{-1}t(\mathbf{y};\boldsymbol{\mu}) \otimes \mathbf{I}_{\mathbf{p}}\right\} + \frac{1}{2}\mathbf{D}'(\mathbf{y};\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{D}'(\mathbf{y};\boldsymbol{\mu})\right].
\end{aligned}
$$

$$\tag{3.19}$$

3.3 Saddlepoint approximations

We now introduce the saddlepoint approximations for the multivariate proper dispersion model. The term 'saddlepoint approximation' was used by Jørgensen (1997b) in

univariate dispersion models and we keep this term for the multivariate cases also. The saddlepoint approximation in multivariate dispersion models, as in the univariate case, fortuitously avoids approximation of the term in the exponent. This is useful because the position vector μ enters the density only via the exponent, so that the saddlepoint approximations preserves the statistical properties regarding the inference on μ.

In the following sections we investigate the limiting behaviour of the multivariate proper dispersion model when $||\Sigma|| \to 0$, that is, $||\Lambda|| \to \infty$ where $|| \cdot ||$ is the Euclidean norm of the matrix.

3.3.1 Barndorff-Nielsen's $p^{\star}$-formula

In this section, we develop a saddlepoint approximation for multivariate proper dispersion model by using the $p^{\star}$-formula (2.22):

Theorem 3.3 *Let the multivariate proper dispersion model be given by*

$$f(\mathbf{y}; \mu, \Sigma) = a(\Sigma)|\mathbf{V}(\mathbf{y})|^{-1/2} \exp\left\{ -\frac{1}{2} t^{\top}(\mathbf{y}; \mu)\,\Sigma^{-1} t(\mathbf{y}; \mu) \right\}. \tag{3.20}$$

Then the saddlepoint approximation for the multivariate proper dispersion model is given by

$$f(\mathbf{y}; \mu, \Sigma) \quad \sim \quad (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} |\mathbf{V}(\mathbf{y})|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} t^{\top}(\mathbf{y}; \mu)\,\Sigma^{-1} t(\mathbf{y}; \mu) \right\},$$
$$\mathbf{y} \in \Omega^p \tag{3.21}$$

Proof. Let $\mathbf{Y} \sim PD_p(\mu, \Sigma)$. For one observation, we have by Lemma 3.2 that the maximum likelihood estimator of μ is $\hat{\mu} = \mathbf{y}$. Note that for a single observation the ancillary statistic is degenerate. Using (3.17) and (3.19), the observation matrix of μ

at $\mathbf{y}$ is given by

$$\widehat{j} = -\left.\frac{\partial^2 \ell(\boldsymbol{\mu})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^\top}\right|_{\boldsymbol{\mu}=\mathbf{y}} = \mathbf{V}^{-1/2}(\mathbf{y})\,\boldsymbol{\Sigma}^{-1}\mathbf{V}^{-1/2}(\mathbf{y})\,. \tag{3.22}$$

Therefore, by using the $p^\star$-formula (2.22) the saddlepoint approximation for the multivariate proper dispersion model is given by

$$
\begin{aligned}
f(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) \;\sim\;& (2\pi)^{-\frac{p}{2}}\,|\mathbf{V}^{\frac{1}{2}}(\mathbf{y})\,\boldsymbol{\Sigma}\mathbf{V}^{\frac{1}{2}}(\mathbf{y})\,|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}t^\top(\mathbf{y};\boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}t(\mathbf{y};\boldsymbol{\mu})\right\} \\
=\;& (2\pi)^{-\frac{p}{2}}\,|\boldsymbol{\Sigma}|^{-\frac{1}{2}}|\mathbf{V}(\mathbf{y})\,|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}t^\top(\mathbf{y};\boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}t(\mathbf{y};\boldsymbol{\mu})\right\}, \quad (3.23)
\end{aligned}
$$

Hence the result. $\square$

We show below, by a Laplace approximation that the above approximation holds for $\|\boldsymbol{\Sigma}\| \to 0$. That is, the saddlepoint approximation is equivalent to

$$|\boldsymbol{\Sigma}|^{1/2}a_p(\mathbf{y};\boldsymbol{\Sigma}) \to (2\pi)^{-\frac{p}{2}}\,|\mathbf{V}(\mathbf{y})\,|^{-\frac{1}{2}} \text{ as } \|\boldsymbol{\Sigma}\| \to 0$$

or, equivalently, in multivariate proper dispersion models

$$|\boldsymbol{\Sigma}|^{1/2}a_p(\boldsymbol{\Sigma}) \to (2\pi)^{-\frac{p}{2}} \text{ as } \|\boldsymbol{\Sigma}\| \to 0$$

and this approximation is clearly uniform on compacts.

3.3.2 The Laplace method

The Laplace method provides a proof of the saddlepoint approximation. The Laplace method for multidimensional case is direct generalization of the corresponding univariate discussion in (2.1). The key argument in the univariate case was to produce in the integrand a factor close to a normal density of small variance. Essentially the same argument applies in p dimensions.

28

We now present the important Laplace approximation, as described in, e.g., Barndorff-Nielsen and Cox (1989, p.169). Before we give the Laplace approximation, note that for any matrix $\mathbf{\Lambda}$, we can write $\mathbf{\Lambda} = \lambda \mathbf{\Lambda}_0$ where $\mathbf{\Lambda}_0$ is a fixed constant matrix and λ is a constant.

Proposition 3.1 (Laplace approximation) *Define*

$$I(\lambda) = \int_\Omega q(\mathbf{y}) e^{-\lambda r(\mathbf{y})} d\mathbf{y}$$

where the functions $r(\mathbf{y})$ and $q(\mathbf{y})$ are defined over the region Ω of integration, $\Omega \subset \mathcal{R}^p$. Suppose that $r(\mathbf{y})$ has a unique minimum in the interior of $\mathbf{D}$ at $\mathbf{y} = \boldsymbol{\mu}$, then

$$I(\lambda) \sim \frac{e^{-\lambda r(\boldsymbol{\mu})} q(\boldsymbol{\mu}) (2\pi)^{p/2}}{|r''(\boldsymbol{\mu})|^{1/2}}, \quad as \ \lambda \to \infty, \tag{3.24}$$

where r'' is the second derivative of $r(\mathbf{y})$ evaluated at $\mathbf{y} = \boldsymbol{\mu}$ and $|r''(\boldsymbol{\mu})| > 0$.

Theorem 3.4 *Let the multivariate proper dispersion model be given by*

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = a(\boldsymbol{\Sigma}) |\mathbf{V}(\mathbf{y})|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} t^\top (\mathbf{y}; \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} t(\mathbf{y}; \boldsymbol{\mu}) \right\}, \mathbf{y} \in \Omega^p.$$

where

$$a(\boldsymbol{\Sigma})^{-1} = \int_{\Omega^p} |\mathbf{V}(\mathbf{y})|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} t^\top (\mathbf{y}; \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} t(\mathbf{y}; \boldsymbol{\mu}) \right\} d\mathbf{y}.$$

Then as $||\boldsymbol{\Sigma}|| \to 0$,

$$a(\boldsymbol{\Sigma}) \sim (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2}. \tag{3.25}$$

Proof. Define

$$r(\mathbf{y}) = \frac{1}{2} t^\top (\mathbf{y}; \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} t(\mathbf{y}; \boldsymbol{\mu}).$$

and

$$\boldsymbol{\Delta}'(\mathbf{y}; \boldsymbol{\mu}) = \text{diag}\left\{ \frac{d'(y_1; \mu_1)}{r(y_1; \mu_1)}, \ldots, \frac{d'(y_p; \mu_p)}{r(y_p; \mu_p)} \right\}$$

where d' is the differentiation of d with respect to y. Differentiating $r(\mathbf{y})$ twice with respect to $\mathbf{y}$ gives,

$$
\begin{aligned}
\frac{\partial r(\mathbf{y})}{\partial \mathbf{y}} &= \frac{1}{2}\boldsymbol{\Delta}'(\mathbf{y};\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}t\,(\mathbf{y};\boldsymbol{\mu}) \\
\frac{\partial^2 r(\mathbf{y})}{\partial \mathbf{y}\partial \mathbf{y}^{\top}} &= \frac{1}{2}\frac{\partial \boldsymbol{\Delta}'(\mathbf{y};\boldsymbol{\mu})}{\partial \mathbf{y}^{\top}}\left\{\boldsymbol{\Sigma}^{-1}t\,(\mathbf{y};\boldsymbol{\mu})\otimes \mathbf{I}_p\right\} + \frac{1}{2}\boldsymbol{\Delta}'(\mathbf{y};\boldsymbol{\mu})\frac{\partial \boldsymbol{\Sigma}^{-1}t\,(\mathbf{y};\boldsymbol{\mu})}{\partial \mathbf{y}^{\top}} \\
&= \frac{1}{2}\frac{\partial \boldsymbol{\Delta}'(\mathbf{y};\boldsymbol{\mu})}{\partial \mathbf{y}^{\top}}\left\{\boldsymbol{\Sigma}^{-1}t\,(\mathbf{y};\boldsymbol{\mu})\otimes \mathbf{I}_p\right\} + \frac{1}{4}\boldsymbol{\Delta}'(\mathbf{y};\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}'(\mathbf{y};\boldsymbol{\mu}). \quad (3.26)
\end{aligned}
$$

Note that by using (3.16), we have $\boldsymbol{\Delta}'(\boldsymbol{\mu};\boldsymbol{\mu}) = 2\mathbf{V}^{-1/2}(\boldsymbol{\mu})$. We have shown in Lemma 3.2 that $r(\mathbf{y})$ has unique minimum at $\mathbf{y} = \boldsymbol{\mu}$. At $\mathbf{y} = \boldsymbol{\mu}$ the first term on the right-hand side of (3.26) vanishes and we have

$$
\left.\frac{\partial^2 r(\mathbf{y})}{\partial \mathbf{y}\partial \mathbf{y}^{\top}}\right|_{\mathbf{y}=\boldsymbol{\mu}} = \mathbf{V}^{-1/2}(\boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}\mathbf{V}^{-1/2}(\boldsymbol{\mu})\,.
$$

Now by applying this to (3.24), with the fact that $r(\boldsymbol{\mu}) = 0$, we obtain

$$
a\,(\boldsymbol{\Sigma})^{-1} \sim \frac{\mathbf{V}^{-\frac{1}{2}}(\boldsymbol{\mu})(2\pi)^{\frac{p}{2}}}{|\mathbf{V}^{\frac{1}{2}}(\boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}\mathbf{V}^{\frac{1}{2}}(\boldsymbol{\mu})|} = (2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}.
$$

Now by taking the inverse of the above, we get (3.25). Hence the theorem. $\quad\square$

Suppose we think of the approximation obtained by using the p^*-formula as a saddlepoint approximation, then the approximation obtained by Laplace's method can be regarded as renormalized saddlepoint approximation. Therefore the ordinary and renormalized saddlepoint approximation agree asymptotically in the case of multivariate proper dispersion model. That is, in the case of a multivariate proper dispersion model the renormalized saddlepoint approximation is exact. In fact, this is one of the main results for the univariate proper dispersion models.

3.4 Multivariate normal approximation

We now consider a multivariate normal approximation to the multivariate proper dispersion model. This will pave the way to mimic some basic multivariate techniques in the multivariate proper dispersion model with small dispersion matrix.

Theorem 3.5 *Let the multivariate proper dispersion model be given by*

$$f(\mathbf{y};\boldsymbol{\mu};\boldsymbol{\Sigma}) = a(\boldsymbol{\Sigma})|\mathbf{V}(\mathbf{y})|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}t^\top(\mathbf{y};\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}t(\mathbf{y};\boldsymbol{\mu})\right\}. \tag{3.27}$$

Then

$$\mathbf{Z} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}_p) \ \ as \ ||\boldsymbol{\Sigma}|| \to \mathbf{0}, \tag{3.28}$$

where $\mathbf{Z}$ *is a linear transformation given by*

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}\mathbf{V}^{-1/2}(\boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu}). \tag{3.29}$$

Proof. For small dispersion $\mathbf{y}$ gets closer to $\boldsymbol{\mu}$. So we prove this by introducing a second-order Taylor series approximation to the term in the exponent of (3.27) at $\mathbf{y} = \boldsymbol{\mu}$ and saddlepoint approximation (uniformly convergent saddlepoint approximation) to the normalizing constant. Let

$$d_{\boldsymbol{\Sigma}}(\mathbf{y};\boldsymbol{\mu}) = t^\top(\mathbf{y};\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}t(\mathbf{y};\boldsymbol{\mu})$$

$$\left.\frac{\partial d_{\boldsymbol{\Sigma}}(\mathbf{y};\boldsymbol{\mu})}{\partial \mathbf{y}}\right|_{\mathbf{y}=\boldsymbol{\mu}} = \mathbf{0}$$

$$\left.\frac{\partial^2 d_{\boldsymbol{\Sigma}}(\mathbf{y};\boldsymbol{\mu})}{\partial \mathbf{y}\partial \mathbf{y}^\top}\right|_{\mathbf{y}=\boldsymbol{\mu}} = 2\mathbf{V}^{-1/2}(\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{V}^{-1/2}(\boldsymbol{\mu}). \tag{3.30}$$

Therefore the second-order Taylor series approximation to $d_{\boldsymbol{\Sigma}}(\mathbf{y};\boldsymbol{\mu})$ at $\mathbf{y} = \boldsymbol{\mu}$ is given by

$$d_{\boldsymbol{\Sigma}}(\mathbf{y};\boldsymbol{\mu}) \approx (\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{V}^{-1/2}(\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}\mathbf{V}^{-1/2}(\boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu}). \tag{3.31}$$

The Jacobian matrix is given by

$$J(\mu) = \frac{\partial \mathbf{Z}}{\partial \mathbf{Y}^\top} = \Sigma^{-1/2}\mathbf{V}^{-1/2}(\mu).$$

Hence the density of $\mathbf{Z}$ is

$$
\begin{aligned}
f(\mathbf{z}; \mu, \Sigma) &= |J(\mu)|^{-1} f\left(\mathbf{V}^{1/2}(\mu)\Sigma^{1/2}\mathbf{z} + \mu; \mu, \Sigma\right) \\
&\approx |\Sigma^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}(\mu)|^{-1} a(\Sigma)|\mathbf{V}\left\{\mathbf{V}^{\frac{1}{2}}(\mu)\Sigma^{\frac{1}{2}}\mathbf{z} + \mu\right\}|^{-\frac{1}{2}}\exp\left\{\mathbf{z}^\top\mathbf{z}\right\} \quad (3.32) \\
&= |\Sigma^{-1/2}\mathbf{V}^{-1/2}(\mu)|^{-1}(2\pi)^{-p/2}|\Sigma|^{-1/2}|\mathbf{V}(\mu)|^{-1/2}\exp\left\{\mathbf{z}^\top\mathbf{z}\right\} \quad (3.33) \\
&\sim (2\pi)^{-p/2}\exp\left\{\mathbf{z}^\top\mathbf{z}\right\}. \quad (3.34)
\end{aligned}
$$

where the term in the exponent of (3.32) was obtained by using the second-order Taylor approximation and the normalizing constant in (3.34) was obtained by the saddlepoint approximation for small-dispersion. Hence the results follows. $\square$

Note that we may write (3.28) as follows

$$\mathbf{Y} \dot{\sim} \mathcal{N}_p\left\{\mu, \mathbf{V}^{1/2}(\mu)\,\Sigma\,\mathbf{V}^{1/2}(\mu)\right\}.$$

for small dispersion matrix.

We note that

$$
\begin{aligned}
\frac{\partial^2 d_\Sigma(\mathbf{y}; \mu)}{\partial \mathbf{y} \partial \mu^\top} &= \frac{\partial\left\{2\Delta'(\mathbf{y}; \mu)\Sigma^{-1}t(\mathbf{y}; \mu)\right\}}{\partial \mu^\top} \\
\left.\frac{\partial^2 d_\Sigma(\mathbf{y}; \mu)}{\partial \mathbf{y} \partial \mu^\top}\right|_{\mu=\mathbf{y}} &\approx -2\mathbf{V}^{-1/2}(\mathbf{y})\,\Sigma^{-1}\mathbf{V}^{-1/2}(\mathbf{y}). \quad (3.35)
\end{aligned}
$$

Hence we have, generalizing (2.13),

$$\frac{\partial^2 d_\Sigma}{\partial \mathbf{y} \partial \mathbf{y}^\top}(\mathbf{y}; \mathbf{y}) = \frac{\partial^2 d_\Sigma}{\partial \mu \partial \mu^\top}(\mathbf{y}; \mathbf{y}) = -\frac{\partial^2 d_\Sigma}{\partial \mathbf{y} \partial \mu^\top}(\mathbf{y}; \mathbf{y}).$$

We now consider a normal approximation to another transformation of $\mathbf{Y}$ which is approximately equal to (3.29).

Theorem 3.6 *The normal approximation to the multivariate proper dispersion model (3.27) with the transformation*

$$\mathbf{Z} = \mathbf{\Sigma}^{-1/2} t\left(\mathbf{Y}; \boldsymbol{\mu}\right) \tag{3.36}$$

is given by

$$\mathbf{Z} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_p\right) \ \text{ as } \|\mathbf{\Sigma}\| \to \mathbf{0}. \tag{3.37}$$

Proof. For the transformation (3.36), the Jacobian matrix is given by

$$\mathbf{J}\left(\mathbf{Y}, \boldsymbol{\mu}\right) = \frac{\partial \mathbf{Z}}{\partial \mathbf{Y}^{\top}} = \frac{1}{2} \mathbf{\Sigma}^{-1/2} \mathbf{\Delta}'(\mathbf{Y}; \boldsymbol{\mu}).$$

Now by using a second-order Taylor series approximation for small dispersion at $\boldsymbol{\mu} = \mathbf{y}$, we get $\mathbf{\Delta}'\left(\mathbf{Y}; \boldsymbol{\mu}\right) \approx 2\mathbf{V}^{-1/2}(\mathbf{Y})$ and thus

$$\mathbf{J}\left(\mathbf{Y}\right) \approx \mathbf{\Sigma}^{-1/2} \mathbf{V}^{-1/2}\left(\mathbf{Y}\right).$$

Hence, the density of $\mathbf{Z}$ is given by

$$
\begin{aligned}
f\left(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Sigma}\right) &\approx \ \left|\mathbf{J}\left(\mathbf{y}\right)\right|^{-1} a(\mathbf{\Sigma}) |\mathbf{V}(\mathbf{y})|^{-1/2} \exp\left\{\mathbf{z}^{\top}\mathbf{z}\right\} \\
&= \ \left|\mathbf{\Sigma}^{-1/2}\mathbf{V}^{-1/2}\left(\mathbf{y}\right)\right|^{-1} a(\mathbf{\Sigma}) |\mathbf{V}(\mathbf{y})|^{-1/2} \exp\left\{\mathbf{z}^{\top}\mathbf{z}\right\} \\
&\sim \ \left|\mathbf{V}^{1/2}\left(\mathbf{y}\right) \mathbf{\Sigma} \mathbf{V}^{1/2}\left(\mathbf{y}\right)\right|^{-1/2} (2\pi)^{-p/2} |\mathbf{\Sigma}|^{-1/2} |\mathbf{V}\left(\mathbf{y}\right)|^{-1/2} \exp\left\{\mathbf{z}^{\top}\mathbf{z}\right\} \\
&\sim \ (2\pi)^{-p/2} \exp\left\{\mathbf{z}^{\top}\mathbf{z}\right\}. \tag{3.38}
\end{aligned}
$$

The normalizing constant in Equation (3.38) is obtained by the saddlepoint approximation. Therefore (3.37) follows immediately. $\square$

Note that we may write (3.37) as follows

$$t\left(\mathbf{Y}; \boldsymbol{\mu}\right) \stackrel{.}{\sim} \mathcal{N}_p\left(\mathbf{0}, \mathbf{\Sigma}\right).$$

It is noted that since $t\left(\mathbf{Y}; \boldsymbol{\mu}\right) \approx \mathbf{V}^{-1/2}\left(\boldsymbol{\mu}\right)\left(\mathbf{Y} - \boldsymbol{\mu}\right)$ we have the normal approximations (3.28), and (3.37) are approximately equal, as we expected.

Chapter 4

Bivariate gamma distribution

In this chapter, we discuss the bivariate gamma distribution. In Section 1, we discuss the normalizing constant in the case of the multivariate gamma model and study the normalizer numerically. In fact, we employ two methods to estimate the normalizer. Although in their construction of multivariate dispersion models, Jørgensen and Lauritzen (1998) did not require the definition of multivariate dispersion models to have specific moments or marginals, we, in Section 3, investigate the moments up to order two in a bivariate situation. And finally, in Section 4, we investigate the marginals numerically.

4.1 Numerical study of normalizing constant

Let the multivariate proper dispersion model be given by

$$f(\mathbf{y}; \boldsymbol{\mu}; \boldsymbol{\Lambda}) = a_p(\boldsymbol{\Lambda})|\mathbf{V}(\mathbf{y})|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\{t^\top(\mathbf{y}; \boldsymbol{\mu})\boldsymbol{\Lambda} t(\mathbf{y}; \boldsymbol{\mu})\}\right]. \qquad (4.1)$$

Jørgensen and Lauritzen (1998) have shown the following derivations to calculate the normalizing constant.

$$a_p^{-1}(\boldsymbol{\Lambda}) = \int \cdots \int \left\{V^{-\frac{1}{2}}(y_1) \cdots V^{-\frac{1}{2}}(y_p)\right\} \exp\left\{-\frac{1}{2}t^\top(\mathbf{y}; \boldsymbol{\mu})\boldsymbol{\Lambda} t(\mathbf{y}; \boldsymbol{\mu})\right\} \mathbf{dy} \quad (4.2)$$

$$
\begin{aligned}
&= \int \ldots \int \left\{ V^{-\frac{1}{2}}(y_1) \ldots V^{-\frac{1}{2}}(y_p) \right\} \exp\left\{ \frac{1}{2} \sum_{i=1}^{p} \lambda_{ii} d\left(y_i; \mu_i\right) \right\} \\
&\qquad \exp\left\{ -\frac{1}{2} \sum_{i \neq j} \lambda_{ij} r\left(y_i; \mu_i\right) r\left(y_j; \mu_j\right) \right\} \mathbf{dy} \\
&= \Pi_{j=1}^{p} c^{-1}(\lambda_{jj}) \int \ldots \int \exp\left\{ -\frac{1}{2} \sum_{i \neq j} \lambda_{ij} r\left(y_i; \mu_i\right) r\left(y_j; \mu_j\right) \right\} \\
&\qquad \Pi_{j=1}^{p} \left\{ c(\lambda_{jj}) V^{-\frac{1}{2}}(y_j) dy_j \right\} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} \lambda_{ii} d\left(y_i; \mu_i\right) \right\} \\
&= \Pi_{j=1}^{p} c^{-1}(\lambda_{jj}) \mathbf{E}\left[\exp\left\{ -\frac{1}{2} \sum_{i \neq j} \lambda_{ij} r\left(y_i; \mu_i\right) r\left(y_j; \mu_j\right) \right\} \right] \\
&= \Pi_{j=1}^{p} c^{-1}(\lambda_{jj}) \mathbf{E}\left(\exp\left[-\frac{1}{2} \{ t^{\top}(\mathbf{y}; \boldsymbol{\mu}) \boldsymbol{\Lambda}_0 t(\mathbf{y}; \boldsymbol{\mu}) \} \right] \right),
\end{aligned}
\tag{4.3}
$$

where $\boldsymbol{\Lambda}_0$ has zeroes on the diagonal, and off-diagonal elements equal to $\boldsymbol{\Lambda}$, c is the normalizer of the marginal univariate proper dispersion model, and the expectation is with respect to the distribution with independent marginals corresponding to the diagonal of $\boldsymbol{\Lambda}$.

In fact, Song (1996) used a derivation same as the above to estimate the normalizing constant for the multivariate gamma model. By using (4.3), the normalizing constant for multivariate gamma model is given by

$$
a_p^{-1}(\boldsymbol{\Sigma}) = \left\{ \frac{\left(\Pi_1^p \lambda_{ii}^{\lambda_{ii}} \right) \exp\left\{ -\sum_1^p \lambda_{ii} \right\}}{\Pi_1^p \Gamma(\lambda_{ii})} \right\}^{-1} \mathbf{E}\left[\exp\left\{ -\sum_{i>j} \lambda_{ij} r\left(X_i; 1\right) r\left(X_j; 1\right) \right\} \right].
\tag{4.4}
$$

The expectation is taken under the probability measure generated by p-independent univariate gamma random variables $X_i \sim Ga\left(1, \lambda_{ii}^{-1}\right)$, $i = 1, \ldots, p$. This expectation in (4.4) can be calculated by using Monte Carlo simulation. Note that we may estimate the normalizer by using numerical integration directly applied to (4.2). Note that this approach may not be a good one to estimate the normalizer in the case of higher dimensions.

In the case of Monte Carlo simulation, we have to estimate the following for the bivariate gamma:

$$\left\{ \frac{\left(\prod_1^2 \lambda_{ii}^{\lambda_{ii}} \right) \exp \left\{ -\sum_1^2 \lambda_{ii} \right\}}{\prod_1^2 \Gamma \left(\lambda_{ii} \right)} \right\}^{-1} \mathbf{E}\left[\exp \left\{ -\lambda_{12} r(X_1; 1) r(X_2; 1) \right\} \right],$$

where the expectation is taken over two independent gamma random variables with mean 1 and dispersion parameter λ_{ii}^{-1}. Before estimating the normalizer, let us define the following as a measure of association between two variables

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}},$$

where $(\mathbf{\Sigma})_{ij} = \sigma_{ij}$. We consider the following dispersion matrices $(\mathbf{\Sigma})$ and the measure of association (ρ) to study the normalizing constants .

Table 4.1: Dispersion Matrix.

ρ	$\mathbf{\Sigma}$
0.75	$\mathbf{\Sigma}_1 = \begin{pmatrix} 0.04 & 0.03 \\ 0.03 & 0.04 \end{pmatrix}$
0.3	$\mathbf{\Sigma}_2 = \begin{pmatrix} 0.2 & 0.12 \\ 0.12 & 0.8 \end{pmatrix}$
0.1	$\mathbf{\Sigma}_3 = \begin{pmatrix} 0.5 & 0.112 \\ 0.112 & 2.5 \end{pmatrix}$
−0.1	$\mathbf{\Sigma}_4 = \begin{pmatrix} 0.5 & -0.112 \\ -0.112 & 2.5 \end{pmatrix}$
−0.3	$\mathbf{\Sigma}_5 = \begin{pmatrix} 0.2 & -0.12 \\ -0.12 & 0.8 \end{pmatrix}$
−0.75	$\mathbf{\Sigma}_6 = \begin{pmatrix} 0.04 & -0.03 \\ -0.03 & 0.04 \end{pmatrix}$

An S-Plus program interfaced with C was written to estimate the normalizing constant by using Monte Carlo (MC) simulation. In fact, S-Plus was used to generate

Table 4.2: Estimation of normalizing constant by using MC method.

ρ	Σ	n			
		1000	10000	50000	100000
0.75	Σ_1	6.25 (0.21)	6.02 (0.14)	6.13 (0.05)	5.99 (0.01)
0.3	Σ_2	0.3814 (1.3×10^{-3})	0.37869 (5.3×10^{-4})	0.37874 (2.3×10^{-4})	0.37905 (1.3×10^{-4})
0.1	Σ_3	0.11263 (1.2×10^{-4})	0.112513 (4.4×10^{-5})	0.112507 (1.9×10^{-5})	0.112498 (1.09×10^{-5})

Table 4.3: A comparison of estimations of normalizing constant.

ρ	Σ	MC method	Saddlepoint approximation
0.75	Σ_1	5.99 (0.01)	6.02
0.3	Σ_2	0.37905 (1.3×10^{-4})	0.41709
0.1	Σ_3	0.112498 (1.09×10^{-5})	0.14307

the gamma random variables and C was used to do the calculation. Since C was used to do the calculation, the time taken to complete a simulation was very small. For example, for a simulation study on 100000 random numbers, the average time was 8 sec. using an HP C200 machine. In Tables 4.2 and 4.3, the estimates are given with their corresponding standard errors (given in brackets). Each simulation was repeated 10 times to get the standard errors.

It is observed from Table 4.2 that as we increase the number of simulations (n) from 1000 to 100000, we obtain almost the same values of the estimate but with higher precision.

In Table 4.3, we compared the estimate obtained by using MC methods with ns 100000 and the estimate obtained by using saddlepoint approximation. For small

dispersion ($\|\Sigma\|$ small) both estimates are very close to each other. Then as $\|\Sigma\|$ increases saddlepoint estimates tends to over estimate the normalizing constant.

Another interesting and potentially important empirical observation was that when a similar simulation study was done to estimate the normalizing constant in the cases of Σ_4, Σ_5 and Σ_6 the values obtained were almost equal to that of Σ_1, Σ_2 and Σ_3, respectively. This situation is similar to a bivariate normal where the values of normalizing constant in the cases of Σ_4, Σ_5 and Σ_6 are equal to that of Σ_1, Σ_2 and Σ_3, respectively.

4.2   Density plots

In this section, we give some density plots of the bivariate gamma. First we shall investigate the behaviour of contour plots and perspective plots when we have two independent gamma margins. That is, the dispersion matrix is diagonal. In this case, we consider the position vectors and dispersion matrices given in Tables 4.4 and 4.5.

Table 4.4: Position vectors.

$\boldsymbol{\mu}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$	$\boldsymbol{\mu}_3 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$

Table 4.5: Dispersion matrices.

$\Sigma_7 = \begin{pmatrix} 0.04 & 0 \\ 0 & 0.04 \end{pmatrix}$	$\Sigma_8 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.3 \end{pmatrix}$	$\Sigma_9 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

Figure 4.1 shows some contour plots of independent bivariate gamma densities and Figure 4.2 shows the corresponding perspective plots. The main reason for plotting the contour plots of independent bivariate gamma densities is to investigate the role
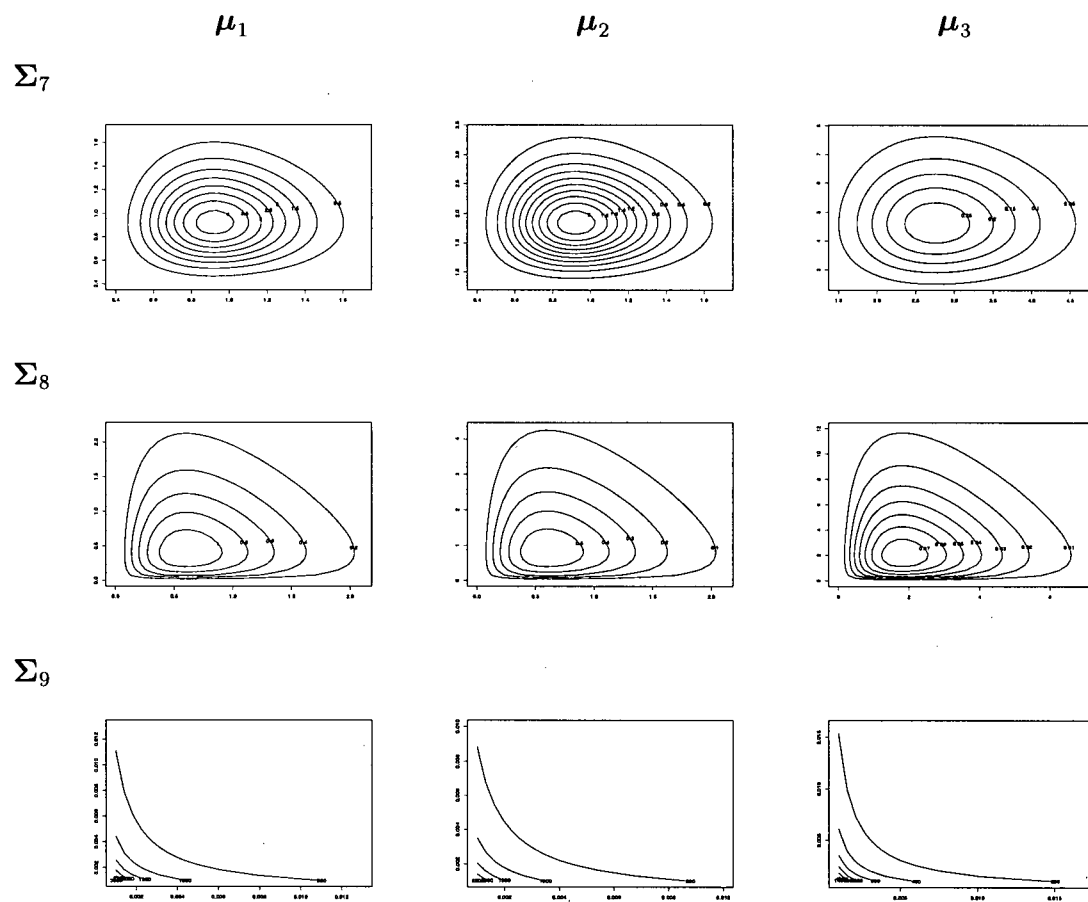
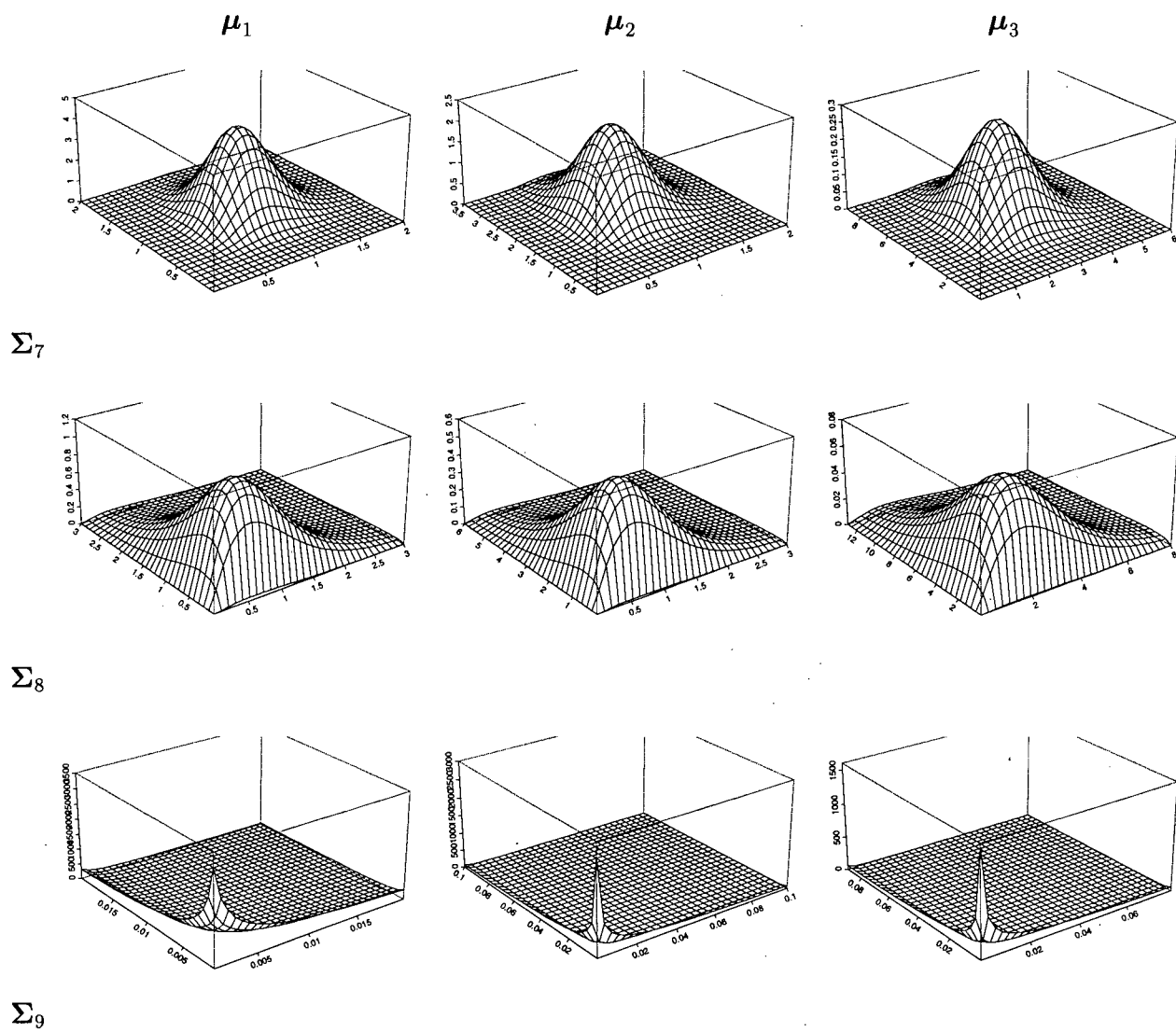Figure 4.1: Contour plots of independent bivariate gamma densities.

Figure 4.2: Perspective plots of independent bivariate gamma densities.

of the position vector $\boldsymbol{\mu}$ on the shape of the densities. In the first row of Figure 4.1 the dispersion matrix is small and thus we can observe clearly that as position varies the peak of the density varies accordingly. When we have large dispersion matrix (in row 3) the peak becomes blurred, essentially because the large value of the dispersion squeezes a lot of probability mass down towards the origin. Large values of the dispersion, as explained by Jørgensen (1997b), corresponds to a value of the Euclidean-norm of the dispersion matrix for which the non-normality of the distribution becomes evident.

Contour plots and perspective plots for bivariate gamma densities when the two variables have some dependence structure were discussed in three cases: 1) dispersion matrix with small dispersion; 2) dispersion matrix with moderately large dispersion; 3) dispersion matrix with large dispersion.

Small dispersion: The dispersion matrices used in this case are given in Table 4.6. And the contour plots and perspective plots corresponds to this case are given in Figures 4.3 and 4.4.

Table 4.6: Small dispersion matrices.

$$\Sigma_{11} = \begin{pmatrix} 0.04 & 0.03 \\ 0.03 & 0.04 \end{pmatrix} \quad \Sigma_{12} = \begin{pmatrix} 0.04 & -0.03 \\ -0.03 & 0.04 \end{pmatrix} \quad \Sigma_{13} = \begin{pmatrix} 0.03 & 0.015 \\ 0.015 & 0.09 \end{pmatrix} \quad \Sigma_{14} = \begin{pmatrix} 0.03 & -0.015 \\ -.015 & 0.09 \end{pmatrix}$$

Moderately large dispersion: The dispersion matrices used in this case are given in Table 4.7. And the contour plots and perspective plots corresponds to this case are given in Figures 4.5 and 4.6.

Table 4.7: Moderately large dispersion matrices.

$$\Sigma_{15} = \begin{pmatrix} 0.1 & 0.138 \\ 0.138 & 0.3 \end{pmatrix} \quad \Sigma_{16} = \begin{pmatrix} 0.1 & -0.138 \\ -0.138 & 0.3 \end{pmatrix} \quad \Sigma_{17} = \begin{pmatrix} 0.2 & 0.085 \\ 0.085 & 0.4 \end{pmatrix} \quad \Sigma_{18} = \begin{pmatrix} 0.2 & -0.085 \\ -.085 & 0.4 \end{pmatrix}$$

Large dispersion: The dispersion matrices used in this case are given in Table 4.8.

41

$$\mu_1 \qquad \mu_2 \qquad \mu_3$$

$\rho = 0.75 \; \Sigma_{11}$

$\rho = -0.75 \Sigma_{12}$

$\rho = 0.3 \;\; \Sigma_{13}$

$\rho = -0.3 \; \Sigma_{14}$
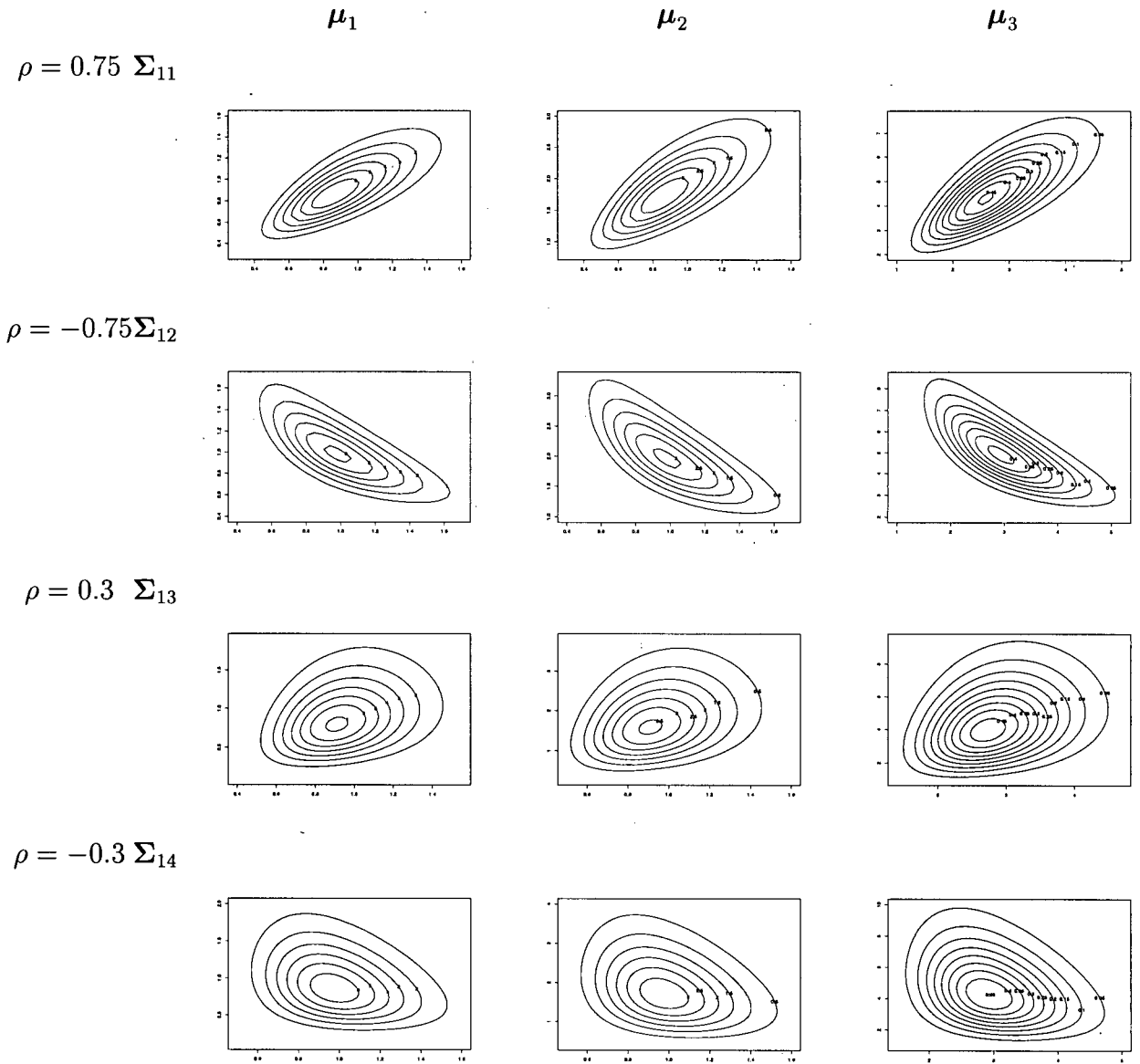
Figure 4.3: Contour plots of bivariate gamma densities with small dispersion.
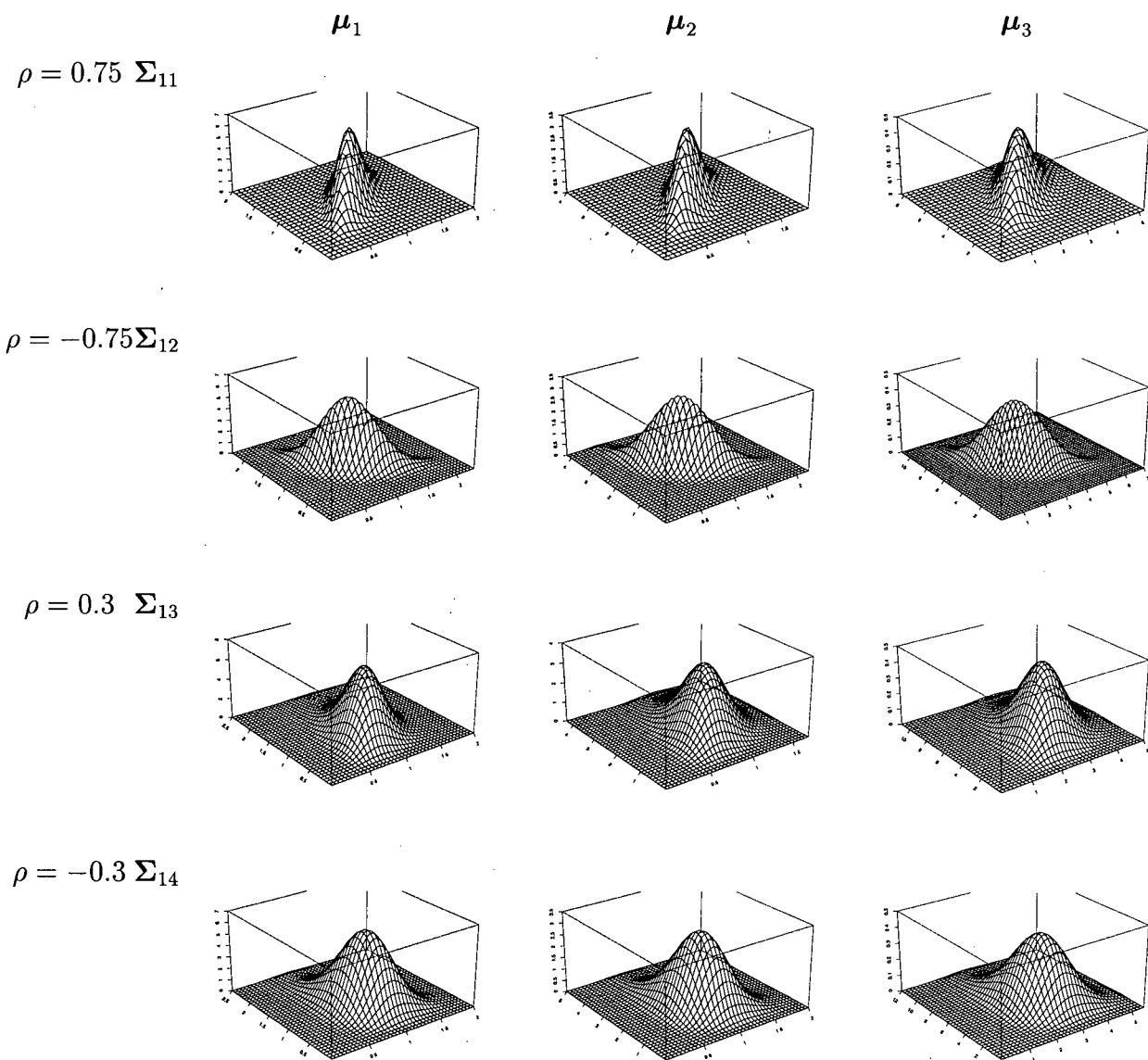
Figure 4.4: Perspective plots of bivariate gamma densities with small dispersion.
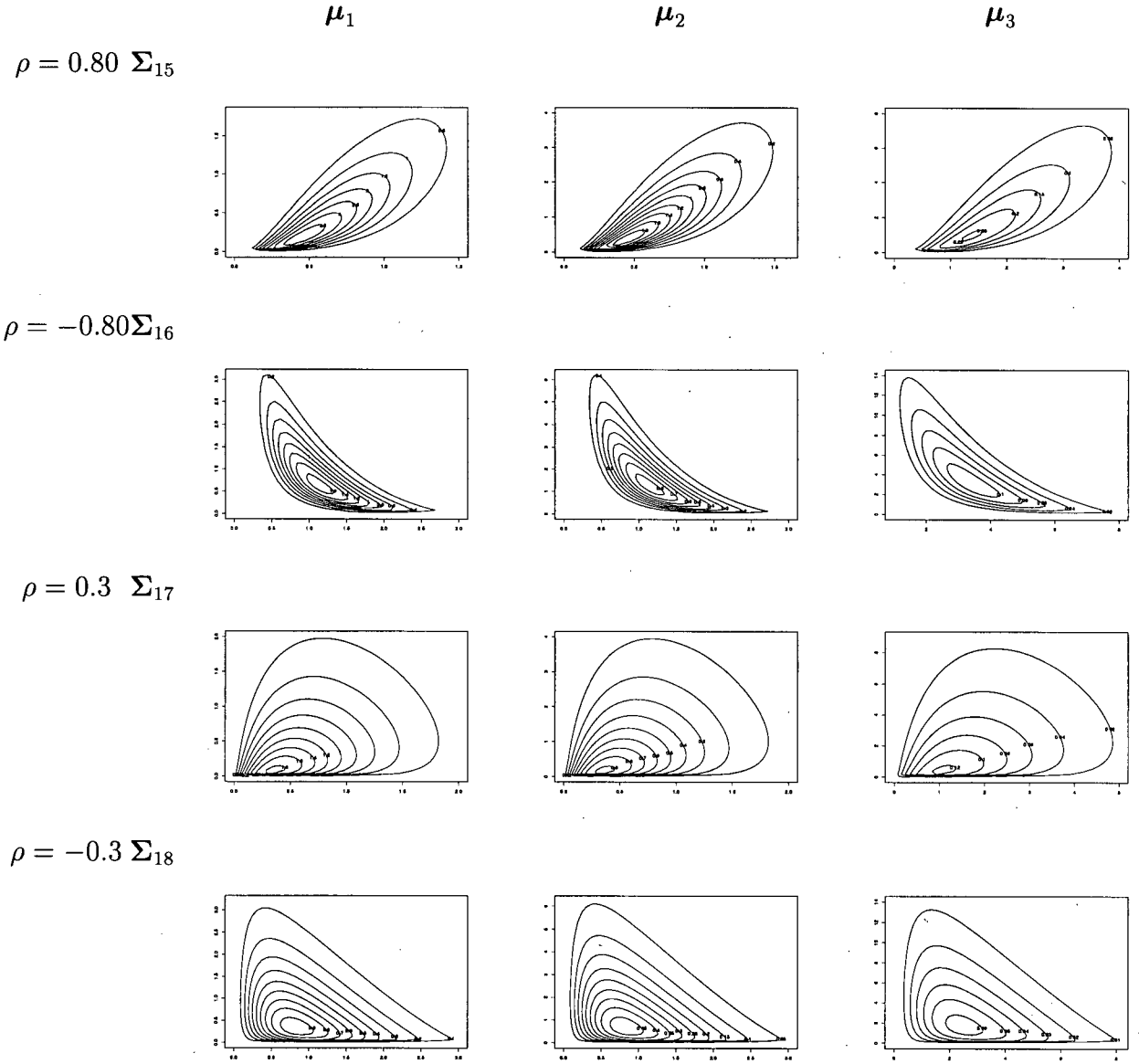
Figure 4.5: Contour plots of bivariate gamma densities with moderately large dispersion.
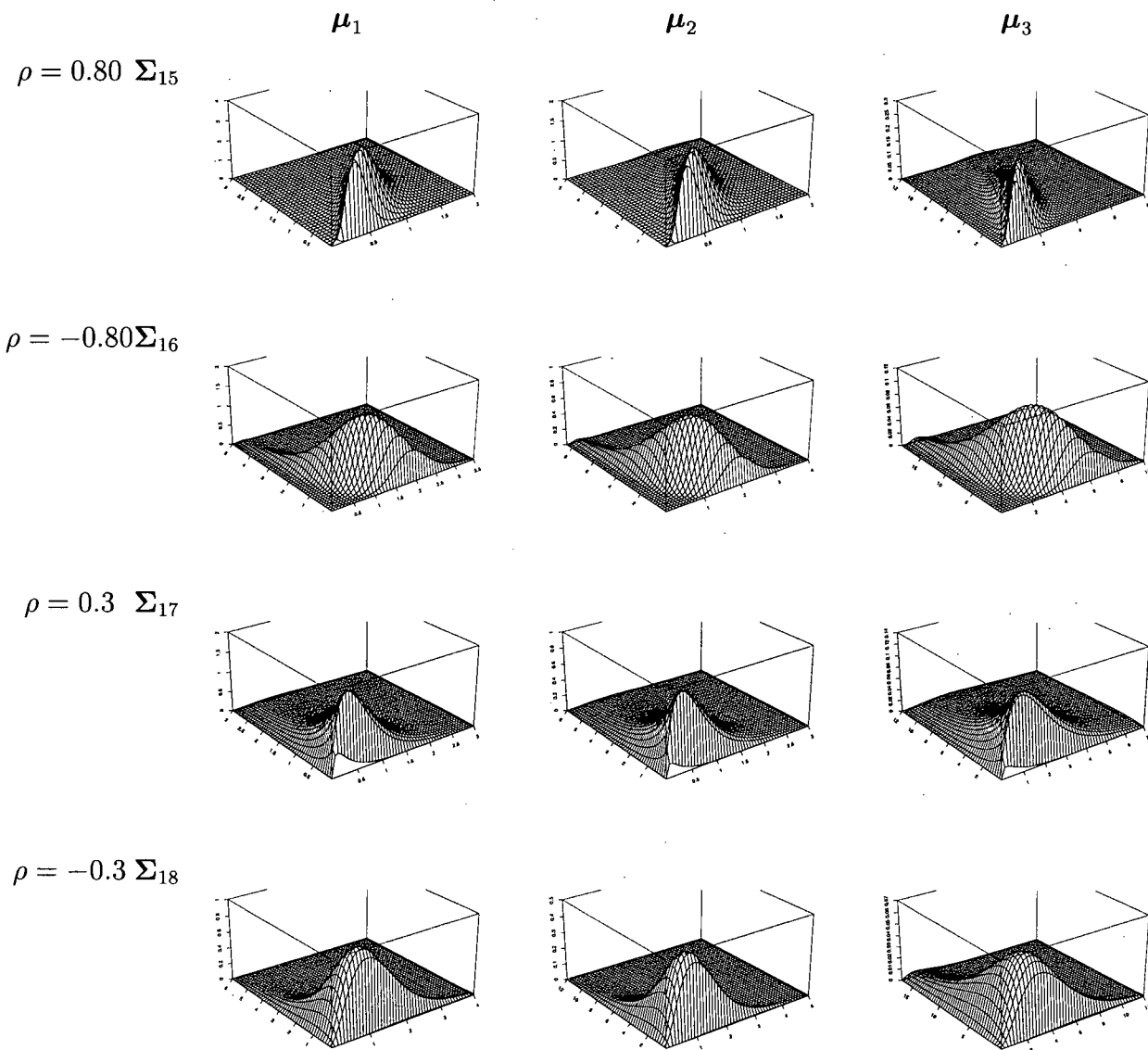
Figure 4.6: Perspective plots of bivariate gamma densities with moderately large dispersion.

μ_1 $\qquad\qquad$ μ_2 $\qquad\qquad$ μ_3

$\rho = 0.80 \ \Sigma_{19}$

$\rho = -0.80 \Sigma_{20}$
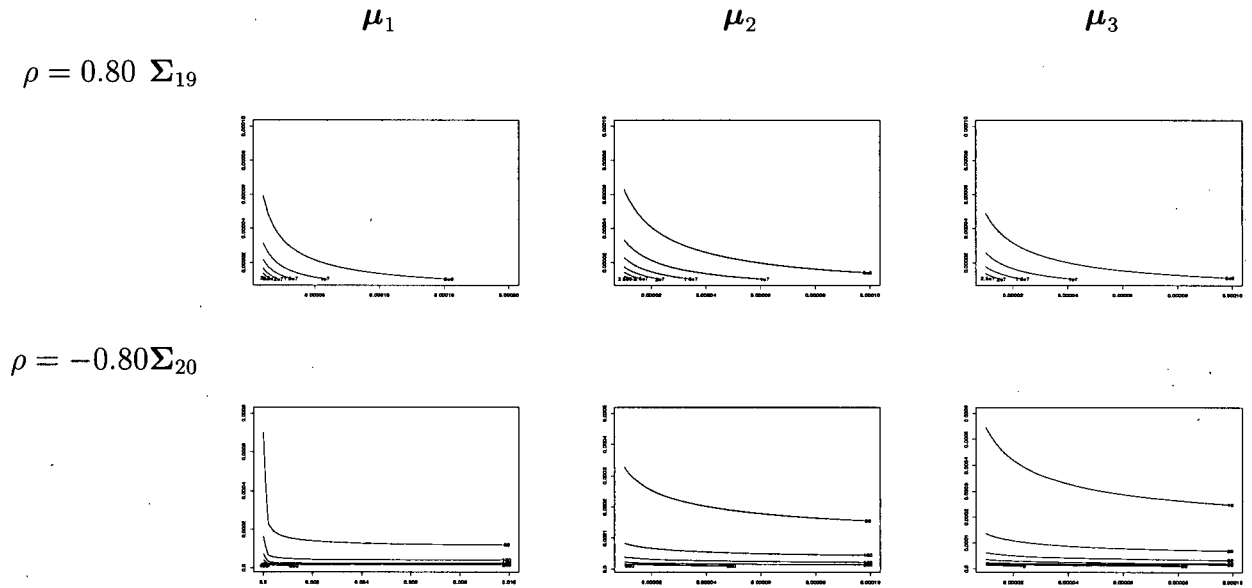
Figure 4.7: Contour plots of bivariate gamma densities with large dispersion.

And the contour plots and perspective plots corresponds to this case are given in Figures 4.7 and 4.8.

Table 4.8: Large Dispersion matrices.

$$\Sigma_{19} = \begin{pmatrix} 2.4 & 3.33 \\ 3.33 & 7.2 \end{pmatrix} \qquad \Sigma_{20} = \begin{pmatrix} 2.4 & -3.33 \\ -3.33 & 7.2 \end{pmatrix}$$

Discussion:

In the independence case, (Figure 4.1) the contour plots are almost ellipsoid like bivariate normal distribution where the axes (major and minor) of the contours are parallel to the xy-axes.

When the two variables have some dependence structure, for small dispersion (Figure 4.3), we considered two cases of dispersion matrices: one with diagonal elements equal: and the other with diagonal elements of ratio 3 and with large and small measure of

46
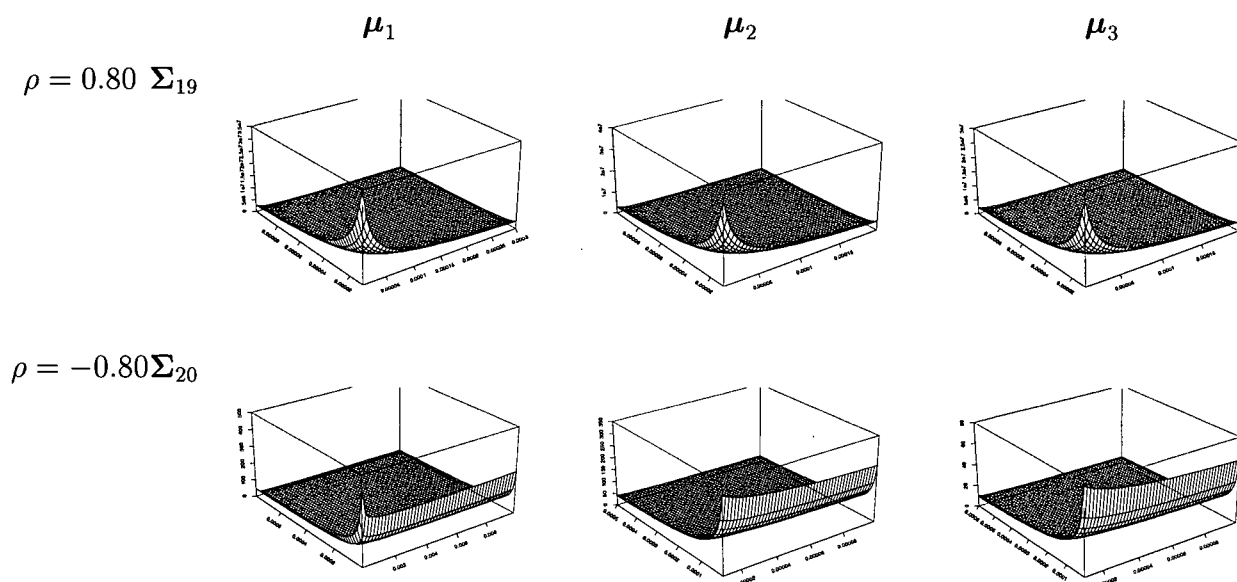
Figure 4.8: Perspective plots of bivariate gamma densities with large dispersion.

association. In the first case of dispersion matrix, for positive measure of association, the major axis of the contour plot (almost ellipsoid) is almost along the 45° line through μ. When the measure is negative the major axis of the contour plot lie along a line at right angle to the 45° line through μ. That is, for small dispersion, the bivariate gamma behaves like bivariate normal. We may also note how the larger measure of association causes the probability to concentrate along a line.

When we dealt with moderately large dispersion matrices (Figure 4.5), this large dispersion causes some deviation from normality but still preserve the fact that according to the sign of the dependency measure the major axis is either along the 45° line or perpendicular to it.

Since a large dispersion in gamma density squeezes lot of probability mass down towards the origin the corresponding plots (Figures 4.7 and 4.8) are no longer ellipsoids.

4.3 Estimation of moments

In this section, we estimate the moments up to order two. The estimation, for dimension p, can be done by using Monte Carlo simulation. The formula for this simulation can be obtained by employing the same kind of derivation that was used to obtain a formula for the normalizer (4.4). We give the equation in the case of two variables Y_1 and Y_2.

$$E\left(Y_i\right) = a_p\left(\Sigma\right) \left\{ \frac{\left(\prod_1^2 \lambda_{ii}^{\lambda_{ii}}\right) \exp\left\{-\sum_1^2 \lambda_{ii}\right\}}{\prod_1^2 \Gamma\left(\lambda_{ii}\right)} \right\}^{-1} E\left[Y_i \exp\left\{-\lambda_{12}r\left(Y_1;\mu_1\right)r\left(Y_2;\mu_2\right)\right\}\right].$$
(4.5)

$$E\left(Y_i^2\right) = a_p\left(\Sigma\right) \left\{ \frac{\left(\prod_1^2 \lambda_{ii}^{\lambda_{ii}}\right) \exp\left\{-\sum_1^2 \lambda_{ii}\right\}}{\prod_1^2 \Gamma\left(\lambda_{ii}\right)} \right\}^{-1} E\left[Y_i^2 \exp\left\{-\lambda_{12}r\left(Y_1;\mu_1\right)r\left(Y_2;\mu_2\right)\right\}\right].$$
(4.6)

The expectation is taken under the probability measure generated by two independent univariate gamma random variables $Y_i \sim Ga\left(\mu_i, \lambda_{ii}^{-1}\right)$, $i = 1, 2$. This expectation can be calculated by using Monte Carlo simulation. In the case of small dimensions we may also use numerical integration to estimate the moments. In Tables 4.9 and 4.10, we give estimates of moments, using Monte Carlo simulation, for two dispersion matrices. The first one with $\rho = 0.75$ and the second with $\rho = 0.3$. We compare the estimated values with reference values (Ref. Val.) obtained by using the marginals $Y_i \sim Ga(\mu_i, \sigma_{ii})$ $i = 1, 2$.

A C program interfaced with S-Plus was used to estimate the moments. A set of 100000 random numbers was generated from corresponding gamma distributions to estimate the moments and this was repeated ten times to estimate the standard error. It was noticed from Tables 4.9 and 4.10, that although the estimates are almost equal for large and small measures of association, the corresponding standard errors are small for small measures of association. This implies that the off-diagonal term

in the dispersion matrix plays an important role, along with the diagonal elements, in estimating the moments.

Table 4.9: Estimation of moments.

$\Sigma = \begin{pmatrix} 0.04 & 0.03 \\ 0.03 & 0.04 \end{pmatrix}$	$\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$		$\mu = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$	
	Ref. Val.	Estimated	Ref. Val.	Estimated
$E(Y_1)$	1	0.9893 (0.0019)	3	2.9879 (0.0044)
$E(Y_2)$	1	0.9894 (0.0017)	4	4.0083 (0.0426)
$Var(Y_1)$	0.04	0.0404 (0.0019)	0.36	0.3122 (0.0578)
$Var(Y_2)$	0.04	0.0341 (0.0115)	0.64	0.62 (0.26)

4.4 Numerical study of marginal density

In this section, we investigate marginals from the bi-variate gamma. For this study, we considered the position vector and dispersion matrices given in Table 4.4.

A numerical integration routine in C was used to integrate out the second variable at some specified points of the first variable. The reference density and the estimated density are given in Figure 4.9. The first plot is for small dispersion, the second and third are for moderately large dispersions and the last is for large dispersion. It is noted, from the first plot, that the estimated and the reference marginal densities are almost exactly the same. In the second and third plots the estimated and the reference marginal densities are very similar. But for the large dispersion, in the last

49

$Ga(3, 0.04)$

$Ga(3, 0.2)$

$Ga(3, 0.5)$

$Ga(3, 1)$
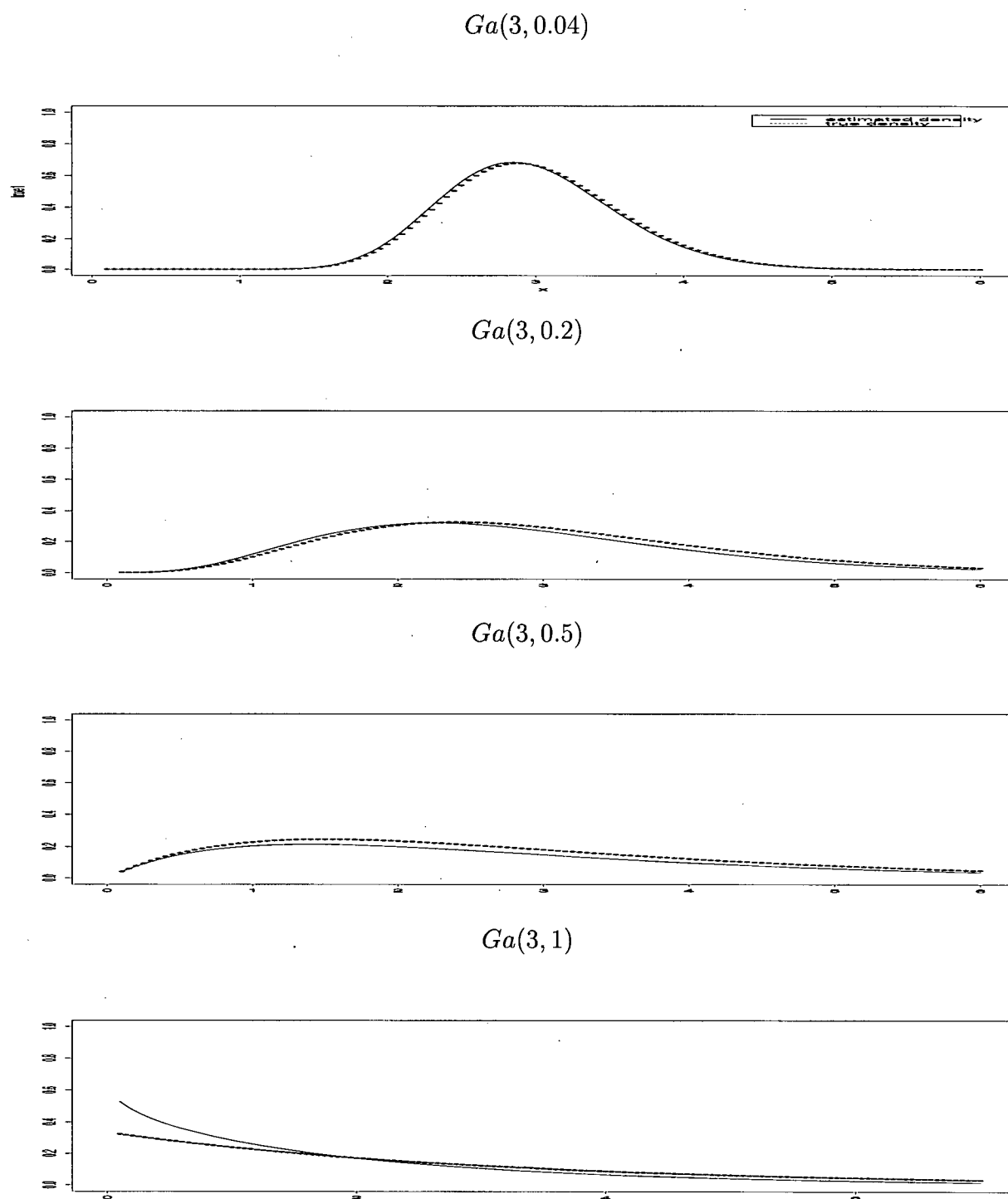
Figure 4.9: A comparison between true and estimated marginal densities (estimated density is given in solid lines).

$\Sigma = \begin{pmatrix} 0.03 & 0.015 \\ 0.015 & 0.09 \end{pmatrix}$	$\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$		$\mu = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$	
	Ref. Val.	Estimated	Ref. Val.	Estimated
$E(Y_1)$	1	0.9949 (0.0003)	3	2.9841 (0.0006)
$E(Y_2)$	1	0.9953 (0.0003)	4	3.9784 (0.0001)
$Var(Y_1)$	0.03	0.0298 (0.00001)	0.27	0.2717 (0.00001)
$Var(Y_2)$	0.09	0.0951 (0.0008)	1.44	1.4326 (0.0093)

Table 4.11: Dispersion matrices.

$\mu = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$	$\Sigma_{21} = \begin{pmatrix} 0.04 & 0.03 \\ 0.03 & 0.04 \end{pmatrix}$	$\Sigma_{22} = \begin{pmatrix} 0.2 & 0.12 \\ 0.12 & 0.8 \end{pmatrix}$	$\Sigma_{23} = \begin{pmatrix} 0.5 & 0.112 \\ 0.112 & 2.5 \end{pmatrix}$	$\Sigma_{24} = \begin{pmatrix} 1 & 0.735 \\ 0.735 & 1.5 \end{pmatrix}$

plot, there is a marked difference between the estimated and the reference marginal densities.

Chapter 5

A generalization of Hotelling's T^2-statistic

5.1 Introduction

One of the most important test problems in univariate statistics is the test of the mean of a given distribution when the variance of the distribution is unknown. The statistic usually used in univariate statistics is the difference between the sample mean, $\overline{x}$, and the hypothetical population mean, divided by the sample standard deviation, s. If the population sampled is $N(\mu, \sigma^2)$, then

$$t = \sqrt{n}\frac{\overline{x} - \mu}{s} \tag{5.1}$$

has the well-known t-distribution with $n-1$ degrees of freedom, where n is the sample size. On the basis of this fact, one can test the hypothesis $H_0 : \mu = \mu_0$, where μ_0 is specified.

The statistic (5.1) was generalized for non-normal distributions in the framework of generalized linear models, cf. Jørgensen (1997a). Jørgensen used the corresponding estimated dispersion parameter in the denominator of (5.1). He, in fact, showed

that the statistic follows asymptotically a t-distribution for large samples and small dispersions.

The multivariate analogue of the square of t given in (5.1) is

$$T^2 = n\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)^{\top} \mathbf{S}^{-1} \left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right) \tag{5.2}$$

where $\overline{\mathbf{X}}$ is the mean vector of a sample of size n and $\mathbf{S}$ is the sample covariance matrix. (5.2) is referred to as the *Hotelling T^2-statistic* in the multivariate normal model.

In the following sections, first we discuss the general likelihood ratio method in the multivariate normal model which will be used to derive (5.2). Then we generalize the T^2-statistic to the multivariate proper dispersion model.

5.2 General likelihood ratio method in multivariate normal model

Let $\boldsymbol{\theta}$ be a vector consisting of all the unknown population parameters, and let $L\left(\boldsymbol{\theta}\right)$ be the likelihood function function obtained by evaluating the joint density of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ at their observed values $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The parameter $\boldsymbol{\theta}$ takes its values in the parameter set $\boldsymbol{\Theta}$.

Let us consider the p-dimensional multivariate case, where the $\mathbf{X}_i$s are independent and identically distributed as $\mathcal{N}_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$. Then $\boldsymbol{\Theta}$ consists of the product of the p-dimensional space for means and the $p(p+1)/2$-dimensional space of variances and covariances such that $\boldsymbol{\Sigma}$ positive-definite. Therefore, $\boldsymbol{\Theta}$ has dimension $v = p + p(p+1)/2$. Under the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, $\boldsymbol{\mu}_0$ is restricted to lie in a subset $\boldsymbol{\Theta}_0$ of $\boldsymbol{\Theta}$. So $\boldsymbol{\Theta}_0$ has dimension $v_0 = p(p+1)/2$ (mean vector is known).

The likelihood ratio test of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ rejects H_0 in favour of $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ if

$$\Lambda = \frac{\max_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu}\neq\boldsymbol{\mu}_0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} < c$$

where c is a suitably chosen constant.

When the sample size is large and under certain regularity conditions, the sampling distribution of $-2 \log \Lambda$ is well approximated by the chi-squared distribution with degrees freedom $v - v_0$. That is, $-2 \log \Lambda \sim \chi^2_{v-v_0}$.

5.2.1 T^2-statistic in multivariate normal

Before we make a connection between likelihood ratio test for the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ and Hotelling's T^2-statistic, we first give the general definition of the Hotelling's T^2-statistic and its distribution.

Definition 5.1 *Let* $\overline{\mathbf{Z}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{W} \sim \mathcal{W}(\mathbf{I}_p, n-1)$ *(Wishart distribution), and let the two random variables be mutually independently distributed. Then the random variable*

$$T^2 = (n-1)\overline{\mathbf{Z}}^\top \mathbf{W}^{-1} \overline{\mathbf{Z}} \tag{5.3}$$

is called Hotelling's T^2-statistic. *The distribution of T^2 is called* Hotelling's T^2-distribution with n d.f.

The T^2-statistic and its distribution were originally developed by Hotelling (1931).

Now let us consider the likelihood ratio test for the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ on the basis of the random sample $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ drawn from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ positive-definite , where $n > p$, $\boldsymbol{\mu}_0$ is a given vector, and $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is unknown. Let

$$\boldsymbol{\Theta} = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : -\infty < \mu_i < \infty, i = 1, \dots, p, \boldsymbol{\Sigma} \text{ positive-definite}\},$$

$$\boldsymbol{\Theta}_0 = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \boldsymbol{\Sigma} \text{ positive-definite}\}.$$

The likelihood ratio criterion Λ is given by

$$
\begin{aligned}
\Lambda &= \frac{\max_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta})} \\
&= \left(1 + \frac{T^2}{n-1}\right)^{-n/2},
\end{aligned}
$$

where $T^2 = n\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)^\top \mathbf{S}^{-1}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)$. Note that this T^2 can be written in the form of (5.3) by letting $\overline{\mathbf{Z}} = \sqrt{n}\boldsymbol{\Sigma}^{-1/2}\left(\overline{\mathbf{X}} - \boldsymbol{\mu}_0\right)$ and $\mathbf{V} = n\mathbf{S}_n$ and $\mathbf{W} = \boldsymbol{\Sigma}^{-1/2}\mathbf{V}\boldsymbol{\Sigma}^{-1/2}$.

Since Λ is a monotonically decreasing function of T^2, the likelihood ratio test based on Λ is equivalent to the following:

$$
\text{Accept } H_0 \quad \text{if} \quad T^2 \le T^2(\alpha),
$$

$$
\text{Reject} H_0 \quad \text{if} \quad T^2 > T^2(\alpha),
$$

where $T^2(\alpha)$ is the upper α-point of the distribution of T^2 for a specified significance level α. Note that when $p = 1$, T^2 reduces to the square of Student's t-statistic.

Derivation of distribution of T^2-statistic

Consider the statistic $T^2 = (n-1)\overline{\mathbf{Z}}^\top \mathbf{W}^{-1}\overline{\mathbf{Z}}$. Since T^2 is invariant under any non-singular transformation, we make use of a random orthogonal transformation

$$
\overline{\mathbf{Z}}_\star = \mathbf{H}\overline{\mathbf{Z}}, \quad \mathbf{W}_\star = \mathbf{H}\mathbf{W}\mathbf{H}^\top,
$$

where $\mathbf{H}_{p \times p}$ is an orthogonal matrix such that its first row is $\dfrac{\overline{\mathbf{z}}^\top}{\left(\overline{\mathbf{z}}^\top \overline{\mathbf{z}}\right)^{1/2}}$. Noting that $\mathbf{H}\overline{\mathbf{Z}} = \left\{\left(\overline{\mathbf{Z}}^\top \overline{\mathbf{Z}}\right)^{1/2}, 0, \ldots, 0\right\}^\top$, T^2 reduces to the following simple form:

$$
T^2 = (n-1)\left(\mathbf{H}\overline{\mathbf{Z}}\right)^\top \left(\mathbf{H}\mathbf{W}\mathbf{H}^\top\right)^{-1}\left(\mathbf{H}\overline{\mathbf{Z}}\right) = (n-1)\frac{\overline{\mathbf{Z}}^\top \overline{\mathbf{Z}}}{w_{\star 11}};
$$

where $w_{\star 11}^{-1}$ is the $(1,1)^{th}$ element of $\mathbf{W}_\star^{-1}$.

Now we have $\left(\overline{\mathbf{Z}}^\top \overline{\mathbf{Z}}\right) \sim \chi_p^2$. By the property of the Wishart distribution, we have the distribution of $\mathbf{W}_\star$ condition on $\mathbf{H}$ is $\mathcal{W}\left(\mathbf{I}_p, n-1\right)$. Suppose we partition the $\mathbf{W}_\star$ as follows:

$$\begin{pmatrix} w_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix},$$

then we can write $w_{\star 11} = w_{11} - \mathbf{W}_{12}\mathbf{W}_{22}^{-1}\mathbf{W}_{21}$ and this has a chi-squared distribution of $n-p$ degrees of freedom independent of $\mathbf{H}$ and hence of $\overline{\mathbf{Z}}$. Hence the distribution of T^2 is given by

$$\frac{T^2}{(n-1)} \sim \frac{\chi_p^2}{\chi_{n-p}^2} = \frac{p}{(n-p)}F_{p,n-p}.$$

The above method of derivation of the T^2-distribution by making use of a random orthogonal transformation is due to Wijsman (1957).

5.3 Maximum likelihood estimators of μ and Σ

In this section, we discuss the small-dispersion asymptotic maximum likelihood estimators of μ and Σ.

Theorem 5.1 *Let* $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, *be a random sample from* $PD_p\left(\mu, \Sigma\right)$ *(multivariate proper dispersion model with p-dimension) where* μ *and* Σ *are the position vector and dispersion matrix, respectively. Then asymptotically, the maximum likelihood estimators of* μ *and* Σ *are given by*

$$\widehat{\mu} \approx \overline{\mathbf{Y}} \text{ as } ||\Sigma|| \to 0, \tag{5.4}$$

$$\widehat{\Sigma} \approx \frac{1}{n}\sum_{i=1}^{n}\left\{t\left(\mathbf{Y}_i; \overline{\mathbf{Y}}\right)t^\top\left(\mathbf{Y}_i; \overline{\mathbf{Y}}\right)\right\} \text{ as } ||\Sigma|| \to 0. \tag{5.5}$$

Proof. We have, from (3.14) in Theorem 3.2, that

$$\frac{\partial \ell(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = -\frac{1}{2}\sum_{i=1}^{n}\mathbf{D}'(\mathbf{Y}_i;\boldsymbol{\mu})\boldsymbol{\Lambda}t\,(\mathbf{Y}_i;\boldsymbol{\mu})\,. \tag{5.6}$$

For small-dispersion, we have

$$\mathbf{D}'(\mathbf{Y}_i;\boldsymbol{\mu}) \approx 2\mathbf{V}^{-1/2}(\boldsymbol{\mu}),$$

and

$$t\,(\mathbf{Y}_i;\boldsymbol{\mu}) \approx \mathbf{V}^{-1/2}(\boldsymbol{\mu})(\mathbf{Y}_i - \boldsymbol{\mu}).$$

Hence, we have

$$\frac{\partial \ell\,(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} \approx \sum_{i=1}^{n}\mathbf{V}^{-1/2}(\boldsymbol{\mu})\boldsymbol{\Lambda}\mathbf{V}^{-1/2}(\boldsymbol{\mu})\,(\mathbf{Y}_i - \boldsymbol{\mu})\,. \tag{5.7}$$

Hence (5.4) follows immediately from (5.7) by taking $\frac{\partial \ell(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = 0$.

For small-dispersion, by the saddlepoint approximation, we obtain

$$f\,(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) \approx (2\pi)^{-\frac{p}{2}}\,|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\,|\mathbf{V}\,(\mathbf{y})\,|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}t^{\top}\,(\mathbf{y};\boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}t\,(\mathbf{y};\boldsymbol{\mu})\right\}. \tag{5.8}$$

Hence the log-likelihood function is given by

$$\ell(\boldsymbol{\mu},\boldsymbol{\Sigma}) \approx \log\left\{(2\pi)^{-\frac{np}{2}}\prod_{1}^{n}|\mathbf{V}\,(\mathbf{y}_i)\,|^{-1/2}\right\} - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}t^{\top}(\mathbf{y}_i;\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}t(\mathbf{y}_i;\boldsymbol{\mu}). \tag{5.9}$$

We now differentiate (5.9) with respect to $\boldsymbol{\Sigma}$.

$$\frac{\partial \ell\,(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} = -\frac{n}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}\left\{t\,(\mathbf{y}_i;\boldsymbol{\mu})\,t^{\top}\,(\mathbf{y}_i;\boldsymbol{\mu})\right\}\boldsymbol{\Sigma}^{-1}. \tag{5.10}$$

Now taking $\frac{\partial \ell(\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} = 0$ and inserting the estimator $\hat{\boldsymbol{\mu}} = \overline{\mathbf{y}}$ gives a maximization problem similar to that for the multivariate normal and we obtain (5.5). $\square$

5.4 T^2 in multivariate proper dispersion model

In the general setup of the T^2-statistic, we consider the likelihood ratio test for the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is known and the dispersion matrix

Σ is unknown. For the discussion of T^2-statistic, we consider two types of asymptotic theory, namely standard large-sample theory $(n \rightarrow \infty)$ and small-dispersion asymptotics. But since we use the estimator of Σ in the likelihood ratio test, we cannot make $||\Sigma|| \rightarrow 0$. So we introduce a known weight matrix $\mathbf{A}$ for each observation vector such that $\mathbf{A}$ is symmetric and regular and consider $||\mathbf{A}_i|| \rightarrow \infty$ for each $i = 1, \ldots, n$. The limit $||\mathbf{A}_i|| \rightarrow \infty$ is the theoretically simplest way of expressing small-dispersion asymptotics; in practice the results may be applied for $||\Sigma|| \rightarrow 0$ or more generally $||\mathbf{A}_i^{-1/2}\Sigma\mathbf{A}_i^{-1/2}|| \rightarrow 0$ for all i.

Consider the likelihood ratio test for the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ on the basis of the random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, with distribution $\mathbf{Y}_i \sim PD_p\left(\boldsymbol{\mu}, \mathbf{A}_i^{-1/2}\Sigma\mathbf{A}_i^{-1/2}\right)$, $i = 1, \ldots, n$, where Σ is positive-definite, $n > p$ and $(\boldsymbol{\mu}, \Sigma)$ is unknown. Let

$$\Theta \quad = \quad \{(\boldsymbol{\mu}, \Sigma) : \mu_i \in \Omega, i = 1, \ldots, p, \Sigma \text{ positive-definite}\},$$

$$\Theta_0 \quad = \quad \{(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \Sigma \text{ positive-definite}\}.$$

We know from Theorem 3.6 that

$$\Sigma^{-1/2}\mathbf{A}_i^{1/2}t\left(\mathbf{Y}_i; \boldsymbol{\mu}\right) \quad \dot{\sim} \quad \mathcal{N}_p\left(\mathbf{0}, \mathbf{I}_p\right) \text{ as } ||\mathbf{A}_i|| \rightarrow \infty$$

$$\text{that is, } \mathbf{A}_i^{1/2}t\left(\mathbf{Y}_i; \boldsymbol{\mu}\right) \quad \dot{\sim} \quad \mathcal{N}_p\left(\mathbf{0}, \Sigma\right) \text{ as } ||\mathbf{A}_i|| \rightarrow \infty.$$

Lemma 5.1 *Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample from $PD_p\left(\boldsymbol{\mu}, \mathbf{A}_i^{-1/2}\Sigma\mathbf{A}_i^{-1/2}\right)$. Then, for $||\mathbf{A}_i|| \rightarrow \infty$ the maximum likelihood estimate of Σ is approximated by*

$$\widehat{\Sigma} \approx \frac{1}{n}\sum_{i=1}^n \mathbf{A}_i^{1/2}t\left(\mathbf{y}_i; \overline{\mathbf{y}}\right)t^\top\left(\mathbf{y}_i; \overline{\mathbf{y}}\right)\mathbf{A}_i^{1/2}. \tag{5.11}$$

Suppose $\boldsymbol{\mu}$ is known and equals to $\boldsymbol{\mu}_0$. Then, for $||\mathbf{A}_i|| \rightarrow \infty$ the maximum likelihood estimate of Σ is given by

$$\widehat{\Sigma}_0 \approx \frac{1}{n}\sum_{i=1}^n \mathbf{A}_i^{1/2}t\left(\mathbf{y}_i; \boldsymbol{\mu}_0\right)t^\top\left(\mathbf{y}_i; \boldsymbol{\mu}_0\right)\mathbf{A}_i^{1/2}. \tag{5.12}$$

Proof. The results (5.11) and (5.12) follow immediately from Theorem 5.1. □

Lemma 5.2 *Let* $\mathbf{Z}_i = \mathbf{A}_i^{1/2} t\left(\mathbf{Y}_i; \boldsymbol{\mu}_0\right)$ *and* $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} \mathbf{A}_i^{1/2} t\left(\mathbf{y}_i; \overline{\mathbf{y}}\right) t^{\top}\left(\mathbf{y}_i; \overline{\mathbf{y}}\right) \mathbf{A}_i^{1/2}$. *Then* $\mathbf{S}$ *and* $\overline{\mathbf{Z}}$ *are asymptotically independent for* $\|\mathbf{A}_i\| \to \infty$ *for all i, where* $\overline{\mathbf{Z}} = 1/n \sum \mathbf{Z}_i$, *and furthermore*

$$\overline{\mathbf{Z}} \dot{\sim} \mathcal{N}_p\left(0, \frac{\Sigma}{n}\right) \tag{5.13}$$

and

$$(n-1)\mathbf{S} \dot{\sim} \mathcal{W}_{n-1}\left(\Sigma\right). \tag{5.14}$$

where $\mathcal{W}_{n-1}(\Sigma)$ *is the Wishart distribution with* $n-1$ *degrees of freedom.*

Proof. Since $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent, so are $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. We have by Theorem 3.6

$$\mathbf{Z}_i \dot{\sim} \mathcal{N}_p\left(0, \Sigma\right) \text{ as } \|\mathbf{A}_i\| \to \infty \tag{5.15}$$

and hence

$$\overline{\mathbf{Z}} \dot{\sim} \mathcal{N}_p\left(0, \frac{\Sigma}{n}\right) \text{ as } \|\mathbf{A}_i\| \to \infty \text{ for all } i. \tag{5.16}$$

Now by applying a second-order Taylor series approximation to $\mathbf{Z}_i$ we obtain

$$
\begin{aligned}
\mathbf{A}_i^{1/2} t\left(\mathbf{Y}_i; \boldsymbol{\mu}_0\right) &\approx \mathbf{A}_i^{1/2} \mathbf{V}^{-1/2}\left(\boldsymbol{\mu}_0\right)\left(\mathbf{Y} - \boldsymbol{\mu}_0\right) \\
&= \mathbf{A}_i^{1/2} \left\{ \mathbf{V}^{-1/2}\left(\boldsymbol{\mu}_0\right)\left(\mathbf{Y} - \overline{\mathbf{Y}}\right) + \mathbf{V}^{-1/2}\left(\boldsymbol{\mu}_0\right)\left(\overline{\mathbf{Y}} - \boldsymbol{\mu}_0\right) \right\} \\
&\approx \mathbf{A}_i^{1/2} t\left(\mathbf{Y}_i; \overline{\mathbf{Y}}\right) + \frac{1}{n} \sum_{i=1}^{n} \mathbf{A}_i^{1/2} t\left(\mathbf{Y}_i; \boldsymbol{\mu}_0\right).
\end{aligned}
$$

Hence

$$\mathbf{A}_i^{1/2} t\left(\mathbf{Y}_i; \overline{\mathbf{Y}}\right) \simeq \mathbf{Z}_i - \overline{\mathbf{Z}}. \tag{5.17}$$

Let $\mathbf{Z}_{p \times n} = \left(\mathbf{Z}_1, \ldots, \mathbf{Z}_n\right)$, and $\mathbf{X} = \mathbf{Z}\mathbf{H}$, where $\mathbf{H}_{n \times n}$ is an orthogonal matrix and n^{th} column is $n^{1/2} \mathbf{1}_{n \times 1}$ so that $\mathbf{X}_n = n^{1/2} \overline{\mathbf{Z}}$. Now

$$\mathbf{X} \dot{\sim} \mathcal{N}_p\left(0, \Sigma \otimes \mathbf{I}_n\right), \tag{5.18}$$

59

so, the columns of $\mathbf{X}$ are independently normally distributed with the same covariance matrix $\boldsymbol{\Sigma}$. Now let $\mathbf{X}' = (\mathbf{X}_1, \ldots, \mathbf{X}_{n-1})$, then we have that $\mathbf{X}'$ and $\mathbf{X}_n$ are asymptotically independent. Now consider

$$
\begin{aligned}
(n-1)\mathbf{S} &= \sum_{i=1}^{n} \mathbf{Z}_i \mathbf{Z}_i^{\top} - n\bar{\mathbf{Z}}\bar{\mathbf{Z}}^{\top} \\
&= \mathbf{Z}\mathbf{Z}^{\top} - \mathbf{X}_n \mathbf{X}_n^{\top} \\
&= \mathbf{X}\mathbf{X}^{\top} - \mathbf{X}_n \mathbf{X}_n^{\top} \\
&= \sum_{i=1}^{n-1} \mathbf{X}_i \mathbf{X}_i^{\top}.
\end{aligned}
\tag{5.19}
$$

Hence we have that $\mathbf{S}$ and $\bar{\mathbf{Z}}$ are asymptotically independent for $\|\mathbf{A}_i\| \to \infty$ for all i. The result (5.15) now follows immediately from (5.18) and (5.19). $\square$

Note that $\mathbf{S}$ is an asymptotically unbiased estimator of $\boldsymbol{\Sigma}$ for $\|\mathbf{A}_i\| \to \infty$ for all i. We define the T^2-statistic for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ with $\boldsymbol{\Sigma}$ unknown by the following monotone transformation of the likelihood ratio test Λ,

$$
T^2 = (n-1)\left(\Lambda^{-2/n} - 1\right).
\tag{5.20}
$$

H_0 is rejected for small values of Λ or, equivalently, for large values of T^2. Note that in the case of multivariate normal the Equation (5.20) is exactly the Hotelling's T^2-statistic.

Theorem 5.2 *Consider the test of the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ on the basis of the random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ drawn from $PD_p\left(\boldsymbol{\mu}, \mathbf{A}_i^{-1/2}\boldsymbol{\Sigma}\mathbf{A}_i^{-1/2}\right)$. The T^2 in (5.20) then follows approximately the Hotelling T^2-distribution. Hence, furthermore, we can show that*

$$
T^2 \,\dot{\sim}\, \frac{p(n-1)}{n-p} F_{p,n-p} \;\; as \; \|\mathbf{A}_i\| \to \infty \; for \; all \; i.
\tag{5.21}
$$

60

Proof. It can easily be shown that the likelihood criterion for the above hypothesis, under multivariate normal approximation, is given by

$$\Lambda^{2/n} \approx \frac{\left|\widehat{\boldsymbol{\Sigma}}\right|}{\left|\widehat{\boldsymbol{\Sigma}}_0\right|}$$

$$= \frac{\left|\widehat{\boldsymbol{\Sigma}}\right|}{\left|\widehat{\boldsymbol{\Sigma}} + \left(\widehat{\boldsymbol{\Sigma}}_0 - \widehat{\boldsymbol{\Sigma}}\right)\right|}$$

$$= \frac{\left|\widehat{\boldsymbol{\Sigma}}\right|}{\left|\widehat{\boldsymbol{\Sigma}} + \overline{\mathbf{Z}}\overline{\mathbf{Z}}^{\top}\right|}.$$

Now by taking

$$\mathbf{B} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}} & \overline{\mathbf{Z}} \\ \overline{\mathbf{Z}}^{\top} & -1 \end{pmatrix},$$

we have

$$(-1)\left|\widehat{\boldsymbol{\Sigma}} + \overline{\mathbf{Z}}\overline{\mathbf{Z}}^{\top}\right| = \left|\widehat{\boldsymbol{\Sigma}}\right|\left|-1 - \overline{\mathbf{Z}}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{Z}}\right|$$

$$\frac{\left|\widehat{\boldsymbol{\Sigma}}\right|}{\left|\widehat{\boldsymbol{\Sigma}} + \overline{\mathbf{Z}}\overline{\mathbf{Z}}^{\top}\right|} = \left(1 + \overline{\mathbf{Z}}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{Z}}\right)^{-1}$$

$$= \left(1 + n\overline{\mathbf{Z}}^{\top}\frac{\mathbf{S}^{-1}}{n-1}\overline{\mathbf{Z}}\right)^{-1}.$$

Hence

$$\Lambda^{2/n} \approx \left(1 + \frac{T_0^2}{n-1}\right)^{-1}$$

$$T_0^2 \approx (n-1)\left(\Lambda^{-2/n} - 1\right)$$

and we have

$$T^2 \approx n\overline{\mathbf{Z}}^{\top}\mathbf{S}^{-1}\overline{\mathbf{Z}} \qquad (5.22)$$

Since

$$\sqrt{n}\overline{\mathbf{Z}} \;\dot{\sim}\; \mathcal{N}_p\left(\mathbf{0}, \boldsymbol{\Sigma}\right),$$

61

$$\mathbf{S} \quad \sim \quad \mathcal{W}_{n-1}\left(\boldsymbol{\Sigma}\right)$$

and $\overline{\mathbf{Z}}$ and $\mathbf{S}$ are asymptotically independent, by Definition 5.1, T^2 is approximately the Hotelling T^2-distribution. Hence

$$T^2 \overset{.}{\sim} \frac{p(n-1)}{n-p} F_{p,n-p} \text{ as } ||\mathbf{A}_i|| \to \infty \text{ for all } i.$$

Hence the theorem. $\square$

Theorem 5.3 *For large sample, Equation (5.21) still holds. That is,*

$$T^2 \overset{.}{\sim} \frac{p(n-1)}{n-p} F_{p,n-p} \text{ as } n \to \infty \text{ for all } i.$$

Proof. By large sample theory, we have

$$\nu = -2 \log \Lambda \to \chi_p^2 \text{ as } n \to \infty.$$

Using (5.20), we have

$$
\begin{aligned}
T^2 &= (n-1)\left(e^{\frac{\nu}{n}} - 1\right) \\
&\approx (n-1)\frac{\nu}{n} \sim \frac{p(n-1)}{n} \frac{\chi_p^2}{p} \\
&= \frac{p(n-1)}{n-p} F_{p,\infty} \text{ as } n \to \infty.
\end{aligned}
$$

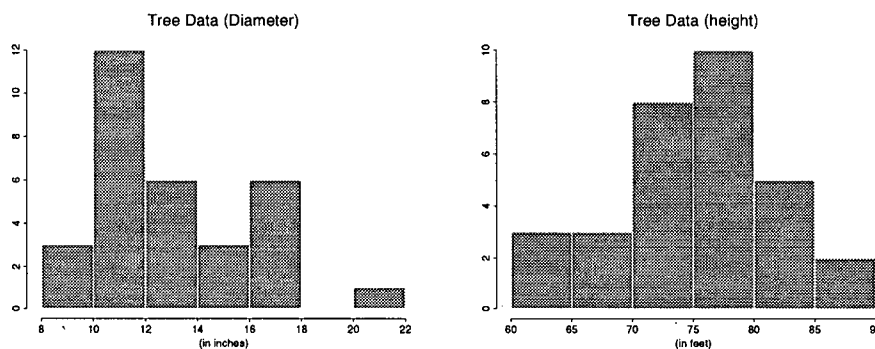Hence $T^2 \overset{.}{\sim} \frac{p(n-1)}{n-p} F_{p,n-p}$ as $n \to \infty$. $\square$

Figure 5.1: Histograms of diameter and height of the trees.

5.4.1 Numerical study on the T^2-statistic

In this section, we investigate the density of the generalized T^2 statistic via two ways;
1. by using data, 2. by simulating random numbers from the bivariate gamma model.

By using data: The data set used in this analysis was obtained from Ryan, Jointer
and Ryan (1985). The *Trees data* contains three variables. The variables are diameter
(d), height (h) and volume (v) for 31 black cherry trees in Allegheny National Forest,
Pennsylvania. The variables are clearly strong candidates for a multivariate gamma
model. We consider the first two variables in our density estimation problem. The
histograms of these two variables are given in Figure 5.1.

Asymptotic estimates for the position vector and dispersion matrix is given as
follows

$$\widehat{\boldsymbol{\mu}} = \begin{pmatrix} 13.25 \\ 76.00 \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.05 & 0.01 \\ 0.01 & 0.007 \end{pmatrix}$$

and $||\widehat{\boldsymbol{\Sigma}}|| = 0.05$.

To estimate the density of T^2, we bootstrap 1000 samples from the data and the
values were calculated for both the classical T^2-statistic (5.2) and the generalized T^2-

63

statistic (5.22). A non-parametric density estimation was obtained by using kernel smoother (with Silverman's band width). The estimated density is compared with the true density, $F_{p,n-p} = F_{2,29}$. The plots are given in Figure 5.2.

From the Figure 5.2, we observe that our generalized T^2-statistic behave almost exactly the same as the classical T^2-statistic for this data set. The main reason may be that the dispersion matrix is small. Even there is some discrepancy at the upper part of the distribution, the approximation is very accurate around the critical region. Note that $F_{2,29,5\%} = 3.33$ and $F_{2,29,1\%} = 5.02$ Although our generalized T^2-statistic obtained from asymptotic theory, we may use this when we have a non-normal data, data from multivariate gamma, in this case.

By Simulation study: One of the hardest part in this thesis was to find a suitable method or procedure to generate random numbers from the multivariate gamma distribution, bivariate gamma distribution in this case. The main problem in employing the well known Markov chain Monte Carlo methods such as Gibbs sampling and Metropolis algorithm, is that it is hard to find the conditional distributions in the multivariate gamma. Neal (1997) proposed a Markov chain Monte Carlo method based on 'Slicing' the density function. The method does not require the whole density function, only requires a function that is proportional to the density. Therefore, it is feasible to use this method in the multivariate gamma where the normalizing constant is unknown.

The idea is that suppose we need to sample from a distribution over some subset of $\mathcal{R}^n$, with density function proportional to some function $f(x)$. One simple idea is to sample uniformly from $n+1$ dimensional region that lies under the plot of $f(x)$. This idea can be formalized by introducing an auxiliary real variable, y, and defining a joint distribution over x and y that is uniform over region $U = \{(x,y) : 0 < y < f(x)\}$
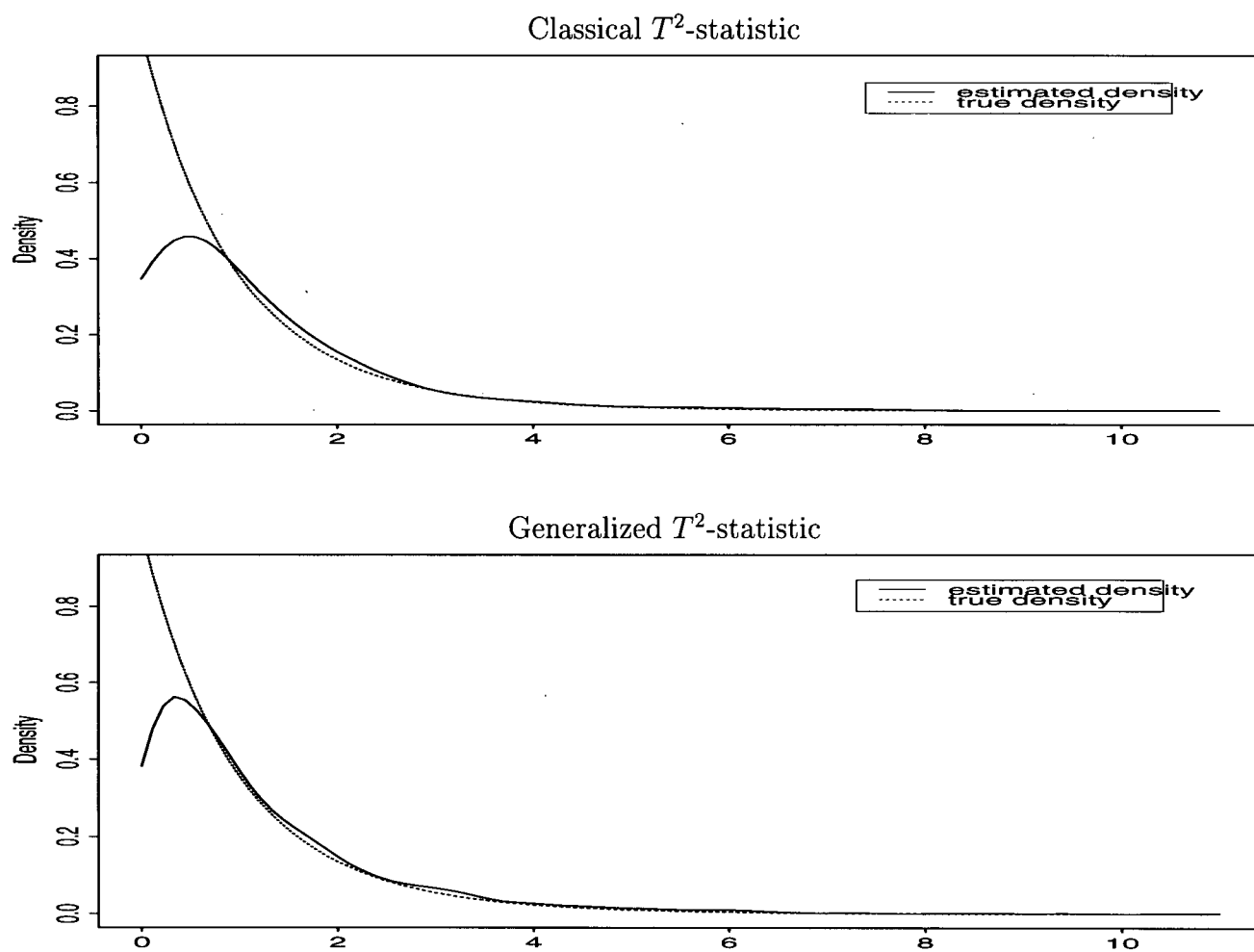
Classical T^2-statistic

Generalized T^2-statistic

Figure 5.2: Comparisons of estimated and true density of T^2 .

below the curve defined by $f(x)$.

The algorithm for a single variable (x) is briefly as follows:

We sample alternately from the conditional distribution for y given the current x – which is uniform over the interval $(0, f(x))$ – and from the conditional distribution for x given y – which is uniform over the region $S = \{x : y < f(x)\}$, which is called the 'slice' defined by y. For multivariables: each real-valued variable, x_i, can be updated by a single-variable slice sampling procedure.

A S-plus program interfaced with C routine was used to generate the random numbers from $Ga_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 0.04 & 0.015 \\ 0.015 & 0.09 \end{pmatrix}$$

and $||\boldsymbol{\Sigma}|| = 0.1$.

Five hundred samples each with 50 random numbers were generated and the classical T^2-statistic (5.2) and the generalized T^2-statistic (5.22) were calculated and the corresponding estimated densities were compared with true density, $F_{2,48}$ (see Figure 5.3).

From the Figure 5.3, we observe that the generalized T^2-statistic behave a little bit better than classical one. The discrepancy between the estimated and the true densities is more in the classical case than in generalized case. Although there is a discrepancy between the estimated and the true densities in the upper part of the plot, there is pretty close agreement between the estimated and the true densities in the tail area in both cases. That is, the approximation is very close around the critical region. This is a very important observation because we are interested in hypothesis testing. Note that $F_{2,48,5\%} = 3.19$ and $F_{2,48,1\%} = 5.08$.
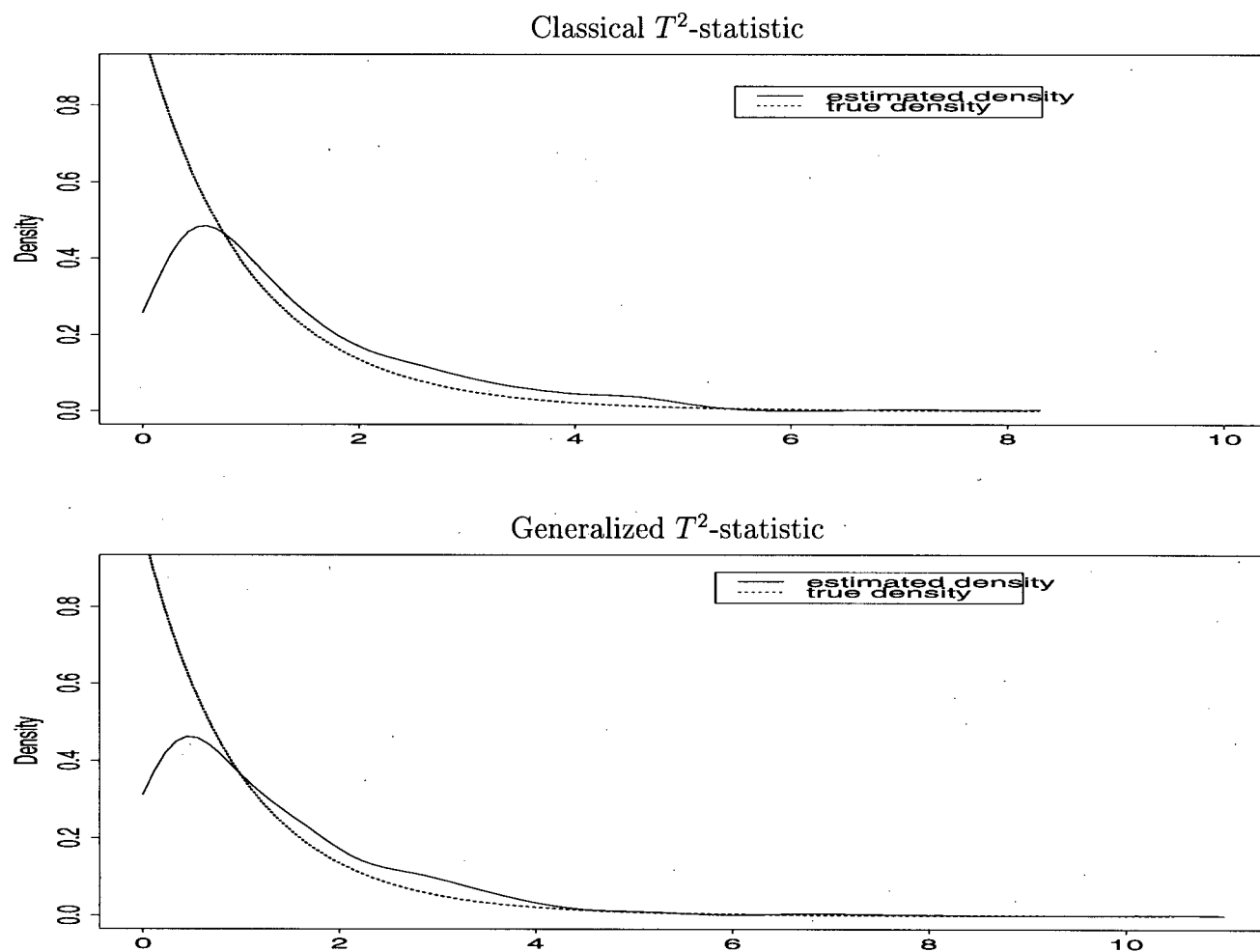
Classical T^2-statistic

Generalized T^2-statistic

Figure 5.3: Comparisons of estimated and true density of T^2 .

Discussion: Both in the simulation study and in the data analysis, we observed that the generalized T^2-statistic behave almost as the same as the classical T^2-statistic. Then the following question arises here: what is the usefulness of the generalized T^2-statistic? The one answer for this question is that in most situations gamma data behave almost like normal data. That is why we could not observe a marked difference between the generalized T^2-statistic (gamma case) and the classical T^2-statistic (normal case, of course). Another answer may be that we simulated data from gamma with small dispersion matrix, in the case of small dispersion, the multivariate proper dispersion model is very similar to the multivariate normal model.

Most of the results in non-normal data analysis have developed by using asymptotic theory. But this does not mean that we have to apply the results only when we have small-dispersion or large sample. As we do in generalized linear model, we can apply in any situation. The main point here is that we feel quite comfortable and satisfied when we do the data analysis of non-normal data with the relevant non-normal distribution rather than with the normal distribution.

As a final note to the non-normal data analysis, we quote the following from Jørgensen (1997b): *"While normal distribution is useful for many types of data the normal distribution is the exception, rather than the rule, except for data with small dispersion. There is an importance of describing data in their natural habitat. Analysis of non-normal data should hence take into account the actual form of the sample space for each type of data".*

Appendix A

Some results on matrices

Here we give basic results on matrices. For more details, see for eg. Dwyer (1962), Johnson and Wichern (1992) and Rao (1985).

A.1 Determinant

1. Let $\mathbf{A} : p \times p, \mathbf{B} : p \times q, \mathbf{C} : q \times p, \mathbf{D} : q \times q$. Then

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = \begin{cases} |\mathbf{A}| \ \ |\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}| & \text{if } \mathbf{A} \text{ is nonsingular,} \\ |\mathbf{D}| \ \ |\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}| & \text{if } \mathbf{D} \text{ is nonsingular.} \end{cases}$$

2. Let $\mathbf{A} : p \times q, \mathbf{B} : q \times p$. Then

$$|\mathbf{I}_p + \mathbf{A}\mathbf{B}| = |\mathbf{I} + \mathbf{B}\mathbf{A}|.$$

A.2 Inverse

1. Let $\mathbf{A}$ be an $n \times n$ nonsingular matrix that is partitioned as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

where $\mathbf{A}_{ij} : n_i \times n_j, i, j = 1, 2$ and $n_1 + n_2 = n$. Then

$$\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}$$

is the upper left-hand block of $\mathbf{A}^{-1}$.

A.3 Kronecker product (Direct product)

Definition A.1 *Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix and $\mathbf{B} = (b_{ij})$ be a $p \times q$ matrix. Then the $(mp) \times (nq)$ matrix*

$$\begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \ldots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \ldots & a_{2n}\mathbf{B} \\ & & \ldots & \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \ldots & a_{mn}\mathbf{B} \end{pmatrix}$$

is called the Kronecker product or the direct product of $\mathbf{A}$ and $\mathbf{B}$ and is denoted by $\mathbf{A} \otimes \mathbf{B}$.

Some elementary properties

1. $(c\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (c\mathbf{B})$,

2. $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$,

3. $(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top$,

4. $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ for nonsingular $\mathbf{A}$ and $\mathbf{B}$,

5. $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C}) \otimes (\mathbf{B}\mathbf{D})$,

6. $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{B} \otimes \mathbf{A}| = |\mathbf{A}|^p|\mathbf{B}|^m$ for $\mathbf{A} : m \times m$ and $\mathbf{B} : p \times p$.

Definition A.2 *For a given $\mathbf{A} : p \times q$, we denote by $Vec\mathbf{A}$ the vector obtained by writing the columns of $\mathbf{A}$ one below the other starting with the first.*

A.4 Matrix derivatives and Jacobians

Definition A.3 *For a scalar function f of a matrix variable $\mathbf{X} = (x_{ij}) : m \times n$ we define its matrix derivative by*

$$\frac{\partial f}{\partial \mathbf{X}} = \left(\frac{\partial f}{\partial x_{ij}} \right) : (m \times n).$$

When $n = 1$, $\mathbf{X}$ is a column vector $\mathbf{x}$ and the corresponding vector $\partial f / \partial \mathbf{x}$ is called the vector derivative.

Table A.1: Vector Derivatives $(\mathbf{x} : p \times 1)$

f	$\partial f / \partial x$
$\mathbf{a}^\top \mathbf{x}$ (a is constant)	$\mathbf{a}$
$\mathbf{x}^\top \mathbf{x}$	$2\mathbf{x}$
$\mathbf{x}^\top \mathbf{A} \mathbf{x}, \mathbf{A} : p \times p$	$(\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$
$\mathbf{x}^\top \mathbf{A} \mathbf{x}, (\mathbf{A} = \mathbf{A}^\top)$	$2\mathbf{A}\mathbf{x}$

Matrix Derivatives

- Let $w = w(\mathbf{Y})$ be a scalar valued function of a matrix $\mathbf{Y}$ which is a function of matrix $\mathbf{X}$ then

$$\frac{\partial w}{\partial \mathbf{X}} = \sum_{\gamma, \delta} \left(\frac{\partial w}{\partial \mathbf{Y}} \right)_{\gamma \delta} \frac{\partial (\mathbf{Y})_{\gamma \delta}}{\partial \mathbf{X}}$$

- The derivative of an element with respect to the matrix

$$\frac{\partial \left(\mathbf{C} \mathbf{X}^\top \mathbf{D} \right)_{ij}}{\partial \mathbf{X}} = \mathbf{D} \mathbf{K}^\top \mathbf{C},$$

where $\mathbf{C} \mathbf{X}^\top \mathbf{D}$ is a $p \times q$ matrix and $\mathbf{K}$ is $p \times q$ matrix with zero in every position except for $k_{ij} = 1$.

- Let $\mathbf{U}(\mathbf{X}) : p \times q$ and $\mathbf{V}(\mathbf{X}) : q \times r$ be matrix functions of a matrix $\mathbf{X}$. Then the Vec derivative of $\mathbf{U}(\mathbf{X})\mathbf{V}(\mathbf{X})$ respect to $\mathbf{X}$ is given by

$$(\mathbf{V} \otimes \mathbf{I}_p)^\top \frac{\partial \mathrm{Vec}(\mathbf{U})}{\partial \mathrm{Vec}^\top(\mathbf{X})} + (\mathbf{I}_r \otimes \mathbf{U}) \frac{\partial \mathrm{Vec}(\mathbf{V})}{\partial \mathrm{Vec}^\top(\mathbf{X})}$$

Bibliography

[1] Barndorff-Nielsen, O.E. (1988). *Parametric Statistical Models and Likelihood*, Lecture notes in Statistics, 50, New York, Springer-Verlag.

[2] Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for use in Statistics*, New York, Chapman and Hall.

[3] Barndorff-Nielsen, O.E., and Cox, D.R. (1994). *Inference and Asymptotics*, New York, Chapman and Hall.

[4] Barndorff-Nielsen, O.E. and Jørgensen, B. (1991). "Some parametric models on the simplex", *J. Multivariate Anal.*, 39, 106–116.

[5] Chambers, J.M. (1977). *Computational Methods for Data Analysis*, Wiley series in Prob. and Math. Statistics, New York, John Wiley and Sons.

[6] Dwyer, P.S. (1962). "Some applications of matrix derivatives in multivariate analysis", *Journal of American Statistical Association*, 62, 607–625.

[7] Everitt, B.S. (1987). *Introduction to optimization methods and their application in Statistics*, New York, Chapman and Hall

[8] Hotelling, H. (1931). The generalization of Student's ratio *Ann. Math. Statist.* 2 360–378.

[9] Hutchinson, T.P. and Lai, C.D. (1990). *Continuous Bivariate Distributions, Emphasising Applications*, South Australia, Rumsby Scientific Publishing.

[10] Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*, Third edition, New-Jersey, Prentice Hall Inc.

[11] Johnson, W.L and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley series in Prob. and Math. Statistics-Applies, New York, John Wiley and Sons.

[12] Jørgensen, B. (1987a). "Small Dispersion Asymptotics", *Brazilian Journal of Prob. and Statistics*, 1, 59–90.

[13] Jørgensen, B. (1987b). "Exponential Dispersion Models", *Journal of Royal Statistical Society*, B49, 127–162.

[14] Jørgensen, B. (1997a). Proper dispersion models (with discussion). *Brazilian Journal of Probability and Statistics* 11, 89–140.

[15] Jørgensen, B. (1997b). "The theory of dispersion models" , New York, Chapman & Hall.

[16] Jørgensen, B. and Lauritzen, S.L. (1998). "Constructing Multivariate Proper Dispersion Models", Research report 2: Department of Statistics and Demography, Odense University, Denmark.

[17] Krishnaiah, P.R (1985). Multivariate gamma distributions. In *Encyclopedia of Statistical Sciences* Vol. 6 (eds. S.Kotz, C.B.Read and D.L.Banks), 63-66. New York: Wiley.

[18] Krishnamoorthy, A.S. and Parthasarathy, M (1951). *Ann. Math. Statist.*, 22, 549; erratum 31, 229.

[19] Minoru Siotani, Takesi Hayakawa, and Yasunori Fujikoshi (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*, Ohio, American Sciences Press, Inc.

[20] McCullagh, P. and Nelder, J.A (1989). *Generalized linear models*, Second edition. London: Chapman & Hall.

[21] Neal, M.R. (1997). "Markov Chain Monte Carlo Methods based on 'Slicing' the density function", Technical report No: 9772, Department of Statistics, University of Toronto, Canada.

[22] Nelder, J.A and Wedderburn, R.W.M. (1972). "Generalized linear models", *J. Roy. Statist. Soc. Ser.*, A 135, 370–384.

[23] Neyman, J. (1926). "Further notes on non-linear regression", *Biometrika*, 18, 257–262.

[24] Pearson, K. (1923). "Notes on skew frequency surfaces", *Biometrika*, 15, 222–230.

[25] Radhakrishna Rao, C (1985). Matrix derivatives: Applications in statistics. In *Encyclopedia of Statistical Sciences* Vol. 6 (eds. S.Kotz, C.B.Read and D.L.Banks), 320–325. New York: Wiley.

[26] Rogers, G.S. (1980). *Matrix Derivatives*, Volume2, New York, Marcel Decker Inc.

[27] Ryan, B.F., Joiner, B.L. and Ryan, T.A. (1985). *Minitab*, Handbook, Second Edition, Duxbury Press, Boston.

[28] Song, P.X. (1996). "Some Statistical Models for the Multivariate Analysis of Longitudinal Data", Ph.D thesis, Dept. of Statistics, University of British Columbia.

[29] Tweedie, M.C.K. (1947). "Functions of a statistical variate with given mean, with special reference to Laplacian distributions", *Proc. Cambridge Phil. Soc.* 49, 41–49.

[30] Wijsman, R.A. (1957). "Random Orthogonal Transformations and their use in some classical distribution problems in Multivariate Analysis", *Ann. Math. Statistics*, 28, 415–423.

[31] Yule, G.U. (1897). " On the significance of Bravais' formulae for regression in the case of skew correlation", *Proceedings of the Royal Society of London*, 60, 477–489.