THE ENERGY GOODNESS-OF-FIT TEST AND E-M TYPE ESTIMATOR FOR ASYMMETRIC LAPLACE DISTRIBUTIONS

John Haman

A Dissertation

Submitted to the Graduate College of Bowling Green
State University in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2018

Committee:

Maria Rizzo, Advisor

Joseph Chao,
Graduate Faculty Representative

Wei Ning

Craig Zirbel

ABSTRACT

Maria Rizzo, Advisor

Recently the asymmetric Laplace distribution and its extensions have gained attention in the statistical literature. This may be due to its relatively simple form and its ability to model skewness and outliers. For these reasons, the asymmetric Laplace distribution is a reasonable candidate model for certain data that arise in finance, biology, engineering, and other disciplines. For a practitioner that wishes to use this distribution, it is very important to check the validity of the model before making inferences that depend on the model. These types of questions are traditionally addressed by goodness-of-fit tests in the statistical literature.

In this dissertation, a new goodness-of-fit test is proposed based on energy statistics, a widely applicable class of statistics for which one application is goodness-of-fit testing. The energy goodness-of-fit test has a number of desirable properties. It is consistent against general alternatives. If the null hypothesis is true, the distribution of the test statistic converges in distribution to an infinite, weighted sum of Chi-square random variables. In addition, we find through simulation that the energy test is among the most powerful tests for the asymmetric Laplace distribution in the scenarios considered.

In studying this statistic, we found that the current methods for parameter estimation of this distribution were lacking, and proposed a new method to calculate the maximum likelihood estimates of the multivariate asymmetric Laplace distribution through the expectation-maximization (E-M) algorithm. Our proposed E-M algorithm has a fast computational formula and often yields parameter estimates with a smaller mean squared error than other estimators.

This dissertation is dedicated to my grandparents.

ACKNOWLEDGMENTS

First, I must recognize my advisor, Maria Rizzo, for all the time, energy, support, encouragement, and guidance you provided. I'm not sure you will ever really know how much I look up to you and have learned from you.

My wife, Jing-Yi Wu, deserves special recognition. This dissertation would simply not exist without you. You constantly remind me of what is important.

My parents and sister have always supported and encouraged my lengthy educational journey. Thank you, Mom, Dad, and Molly.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Introduction

The asymmetric Laplace distribution has proven to be useful for modeling a variety of data, and has so far enjoyed applications in fields such as engineering (Rossi and Spazzini, 2010; Lindsey and Lindsey, 2006), finance (Kozubowski and Podgórski, 2001; Levin and Tchernitser, 2003), quality control (Peterson and Silver, 1979), astronomy (Norris, Nemiroff, Bonnell, Scargle, Kouveliotou, Paciesas, Meegan, and Fishman, 1996), biology (Uppuluri, 1981), and environmental sciences (Fieller, 1993). The breadth of these applications shows that there is a need to apply simple, parametric models in situations where the Normal distribution may fail to be adequate.

One question that is central to modeling with a parametric assumption is "how well does my data actually match the distribution?" The solutions to questions of this type are answered in the field of goodness-of-fit testing, which helps practitioners select a candidate model from a set of plausible models. Selection of a reasonable statistical model is important to ensuring the validity of all statistical procedures that depend on that model. Goodness-of-fit tests have also been developed in the context of regression analysis, but in this dissertation we will focus on the goodness-of-fit of a distribution to a random sample. The purpose of a random sample goodness-of-fit test is to measure the plausibility that a hypothesized distribution generated the data of interest. If the plausibility is relatively low, then one may consider a different distribution as a model for the random sample.

Historically, much work on goodness-of-fit testing has been pointed at the Normal distribution, however we will develop a set of goodness-of-fit tests for the hypothesis that data are generated from the Laplace distribution and its extensions. In this chapter, the Laplace distribution and an extension called the asymmetric Laplace distribution are detailed. We present

properties and characterizations of the asymmetric Laplace distribution, and detail two common estimation methods.

1.2 Laplace Distribution

The symmetric Laplace distribution may be characterized by a density function with two parameters. The parameter θ regulates location, and σ regulates scale.

Definition 1.2.1. *(symmetric Laplace density)*

Let $x \in \mathbb{R}$, $\theta \in \mathbb{R}$, and $\sigma \geq 0$, then

$$f(x;\theta,\sigma) = \frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x-\theta|}$$

is the density of the symmetric Laplace distribution.

The distribution is historically important as the first probability distribution on the unbounded real line making it a precursor to the Normal distribution, appearing in Laplace (1774), four years prior to Laplace's discovery of the Normal distribution in 1778 (Wilson, 1923). For this reason, the Laplace distribution is sometimes called the First Laplace Law and the Normal distribution is called the Second Laplace Law. For analytical and practical reasons, the Normal distribution has received far greater attention than the Laplace distribution.

However the Laplace distribution did receive attention from economist John Maynard Keynes in (Keynes, 1911) wherein Keynes argued the Laplace distribution minimizes the absolute deviation from the median. This is in contrast to the Normal distribution which minimizes the squared deviation from the mean. Keynes' analysis that the median is the "most probable value" for the symmetric Laplace distribution's location parameter precedes Ronald Fisher's maximum likelihood technique introduced in 1912. Keynes did not include details about the support set or parameter space of the class of distributions he considered.

An alternative means of characterizing the Laplace distribution is through its characteristic function (CF).

Definition 1.2.2. *The CF of a Laplace random variable Y is given by*

$$\phi_Y(t) = \mathbb{E}[\exp(itY)] = \frac{e^{it\theta}}{1+\frac{1}{2}\sigma^2 t^2}.$$

Interestingly, this expression for the CF of a Laplace random variable is similar in form to the density of a Cauchy random variable.

1.3 Characterizations and Parameterizations of Asymmetric Laplace Distributions

A simple way to extend the Laplace distribution to a three parameter family is to include a skewness parameter. The presence of a skewness parameter controls the amount of probability that is assigned to either tail of the distribution. If one tail of the distribution is "enlarged", the other tail must be "shrunk" to ensure that the distribution remains proper.

Definition 1.3.1. *(asymmetric Laplace distribution)*

A random variable Y is said to have an asymmetric Laplace distribution denoted $\mathscr{AL}(\theta,\mu,\sigma)$ if there exist parameters $\theta \in \mathbb{R}$, $\mu \in \mathbb{R}$, and $\sigma \in \mathbb{R}^+$ such that the CF of Y has the form

$$\phi_Y(t) = \mathbb{E}[\exp(itY)] = \frac{e^{i\theta t}}{1+\frac{1}{2}\sigma^2 t^2 - i\mu t}.$$

The CF of the asymmetric Laplace distribution is similar to that of the classical Laplace distribution, with the addition of the term "$-i\mu t$" in the denominator. While the form of the CF is relatively simple, we will frequently make use of a change of variables to reparameterize the distribution when the data are univariate.

The following observations about the asymmetric Laplace distribution were made in Kotz, Kozubowski, and Podgorski (2001).

- If $\mu = \sigma = 0$ the distribution is degenerate, as the CF $\phi(t) = e^{i\theta t}$ is degenerate at θ. The distribution places a probability mass of 1 at θ and 0 elsewhere.
- If $\theta = \sigma = 0$ the distribution is exponential with mean μ.

- If $\mu = 0$ and $\sigma \neq 0$ the distribution is exactly the symmetric Laplace distribution with mean θ and variance σ^2.

1.3.1 Asymmetric Laplace Reparameterization

Frequently, it is useful to reparameterize the distribution, trading μ for a new parameter, κ. The substitution follows from factoring the denominator of the CF 1.3.1,

$$\phi_Y(t) = e^{i\theta t}\left(\frac{1}{1+i\frac{\sigma\kappa}{\sqrt{2}}t}\right)\left(\frac{1}{1-i\frac{\sigma}{\sqrt{2}\kappa}t}\right) = \frac{e^{i\theta t}}{1+\frac{1}{2}\sigma^2t^2 - i\frac{\sigma}{\sqrt{2}}(\frac{1}{\kappa}-\kappa)t}. \tag{1.3.2}$$

We denote the asymmetric Laplace distribution under the (θ, κ, σ) parameterization as $\mathscr{AL}^*(\theta, \kappa, \sigma)$.

The parameter κ is related to μ and σ in the following manner:

$$\kappa = \frac{\sqrt{2}\sigma}{\mu + \sqrt{2\sigma^2+\mu^2}} \quad \text{and} \quad \mu = \frac{\sigma}{\sqrt{2}}\left(\frac{1}{\kappa} - \kappa\right). \tag{1.3.3}$$

For analytical purposes, we will require the density and distribution functions of the asymmetric Laplace distribution.

Definition 1.3.4. *((θ, κ, σ) parameterization) Let $f(\cdot;\theta,\kappa,\sigma)$ denote the density function of an $\mathscr{AL}^*(\theta,\kappa,\sigma)$ random variable. Then*

$$f(x;\theta,\kappa,\sigma) = \frac{\sqrt{2}}{\sigma}\frac{\kappa}{1+\kappa^2}\begin{cases}\exp\{-\frac{\sqrt{2}\kappa}{\sigma}|x-\theta|\}, & \text{if } x \geq \theta;\\ \exp\{-\frac{\sqrt{2}}{\sigma\kappa}|x-\theta|\}, & \text{if } x < \theta.\end{cases}$$

Remark 1.3.5. *The derivative of f does not exist at $x = \theta$.*

Definition 1.3.6. *Let $F(\cdot;\theta,\kappa,\sigma)$ denote the distribution function of an $\mathscr{AL}^*(\theta,\kappa,\sigma)$ random variable. Then*

$$F(x;\theta,\kappa,\sigma) = \begin{cases}1 - \frac{1}{1+\kappa^2}\exp\{-\frac{\sqrt{2}\kappa}{\sigma}|x-\theta|\}, & \text{if } x \geq \theta;\\ \frac{\kappa^2}{1+\kappa^2}\exp\{-\frac{\sqrt{2}}{\sigma\kappa}|x-\theta|\}, & \text{if } x < \theta.\end{cases}$$

Definition 1.3.7. *Let $F^{-1}(\cdot;\theta,\kappa,\sigma)$ denote the quantile function of an $\mathscr{AL}^*(\theta,\kappa,\sigma)$ random variable respectively. Then*

$$F^{-1}(p;\theta,\kappa,\sigma)=\begin{cases}\theta+\frac{\sigma\kappa}{\sqrt{2}}\log\left(\frac{1+\kappa^2}{\kappa^2}p\right), & \text{if } p\in\left(0,\frac{\kappa^2}{1+\kappa^2}\right];\\ \theta-\frac{\sigma}{\sqrt{2}\kappa}\log\left((1+\kappa^2)(1-p)\right), & \text{if } p\in\left(\frac{\kappa^2}{1+\kappa^2},1\right).\end{cases} \tag{1.3.8}$$

Fernández and Steel (1998) proposed a general method for introducing skewness to a symmetric distribution by introducing inverse scale factors to the positive and negative support of the distribution. The transformation is given by

$$f(x)\to f(x;k)=\frac{2\kappa}{1+\kappa^2}\begin{cases}f(x\kappa), & x\geq 0;\\ f(x/\kappa), & x<0.\end{cases} \tag{1.3.9}$$

When f is the density of the symmetric Laplace distribution, we obtain the density

$$f(x;\theta,s,\kappa)=\frac{1}{s}\frac{\kappa}{1+\kappa^2}\begin{cases}\exp\left(-\frac{\kappa}{s}(x-\theta)\right), & x\geq 0;\\ \exp\left(\frac{1}{\kappa s}(x-\theta)\right), & x<0,\end{cases} \tag{1.3.10}$$

which agrees with the asymmetric Laplace density given in Definition 1.3.4 under $s=\frac{\sigma}{\sqrt{2}}$.

1.3.2 Quantile Regression Parameterization

In addition to the (θ,μ,σ) and (θ,κ,σ) parameterizations, there is also a common parameterization used in the quantile regression literature (Sánchez, Lachos, and Labra, 2013; Koenker, 2005).

Definition 1.3.11.

$$f(x;\theta,s,p)=\frac{p(1-p)}{s}\exp\left(-\rho_p\left(\frac{x-\theta}{s}\right)\right), \tag{1.3.12}$$

where

$$\rho_p(t)=t(p-1_{(-\infty,0)}(t)). \tag{1.3.13}$$

The function $1_C(x) = 1$ if $x \in C$ and 0 otherwise. This is a preferable parameterization in the quantile regression setting because the parameter $p \in (0,1)$ explicitly assigns p probability to the right tail.

1.3.3 Rizzo-Haman Parameterization

A convenient parameterization is used in Rizzo and Haman (2016). Let us identify

$$p_\kappa = \frac{1}{1+\kappa^2}, \qquad q_\kappa = \frac{\kappa^2}{1+\kappa^2}, \tag{1.3.14}$$

and

$$\lambda = \frac{\sqrt{2}\kappa}{\sigma}. \qquad \beta = \frac{\sqrt{2}}{\kappa\sigma}. \tag{1.3.15}$$

In this notation the asymmetric Laplace density of definition 1.3.4 becomes

$$f(y;\theta,\kappa,\sigma) = \lambda p_\kappa \begin{cases} \exp\left(-\lambda|y-\theta|\right), & y \geq \theta; \\ \exp\left(-\beta|y-\theta|\right). & y < \theta. \end{cases} \tag{1.3.16}$$

The distribution function is

$$F(y;\theta,\kappa,\sigma) = \begin{cases} 1 - p_\kappa \exp\left(-\lambda|y-\theta|\right), & y \geq \theta; \\ q_\kappa \exp\left(-\beta|y-\theta|\right), & y < \theta. \end{cases} \tag{1.3.17}$$

1.4 Properties of Asymmetric Laplace Distributions

Under parameterization given in Definition 1.3.4, if $\kappa = 1$, we have the symmetric Laplace distribution. Parameterization (1.3.16) allows one to write the density and distribution functions of the asymmetric Laplace distribution in a less cumbersome manner.

Parameter	Definition	Value
Mean	$\mathbb{E}[Y]$	$\theta+\mu$
Variance	$\mathbb{E}[Y-\mathbb{E}[Y]]^2$	$\sigma^2+\mu^2$
Median	$F^{-1}(0.5;\theta,\kappa,\sigma)$	$\begin{cases}\theta+\frac{\sigma}{\sqrt{2}\kappa}\log\left(\frac{2}{1+\kappa^2}\right), & \kappa\leq 1\\ \theta-\frac{\sigma\kappa}{\sqrt{2}}\log\left(\frac{2\kappa^2}{1+\kappa^2}\right), & \kappa>1\end{cases}$
Mean Absolute Deviation	$\mathbb{E\|}Y-\mathbb{E}[Y]\mathbb{|}$	$\frac{\sqrt{2}\sigma\exp(\kappa^2-1)}{1+\kappa^2}$

Table 1.1 Common parameter values for a $\mathscr{AL}(\theta,\mu,\sigma)$ or $\mathscr{AL}^*(\theta,\kappa,\sigma)$ distribution

Various density curves of the asymmetric Laplace distribution are shown in Figure 1.1. One can see that as the skewness parameter κ decreases from 1 to 0, the distribution places an increasing amount of probability on the right tail.

1.5 Representations of Asymmetric Laplace Distribution

In this section we present various characterizations of the $\mathscr{AL}^*(\theta,\kappa,\sigma)$ distribution and show how the characterizations may be used to generate asymmetric Laplace random variables. Because the distribution function given in Definition 1.3.6 is closed form, the probability inverse transform (Devroye, 1986) may be used. Otherwise, random variates may be drawn from other distributions mentioned in Proposition 1.5.1 below to generate asymmetric Laplace random variates.

The representation given in Proposition 1.5.1 shows that an asymmetric Laplace random variate is related to a Normal random variate with stochastic (Exponential) mean and variance.

Proposition 1.5.1. *Let Y be an $\mathscr{AL}(\theta,\mu,\sigma)$ random variable, let Z be a standard Normal random variable, and let W, W_1 and W_2 be independent exponential random variables with mean* 1. *Then*

$$Y \stackrel{d}{=} \theta+\mu W+\sigma\sqrt{W}Z \stackrel{d}{=} \theta+\frac{\sigma}{\sqrt{2}}\left(\frac{1}{\kappa}W_2-\kappa W_1\right). \tag{1.5.2}$$

The notation "$\stackrel{d}{=}$" denotes equality in distribution.

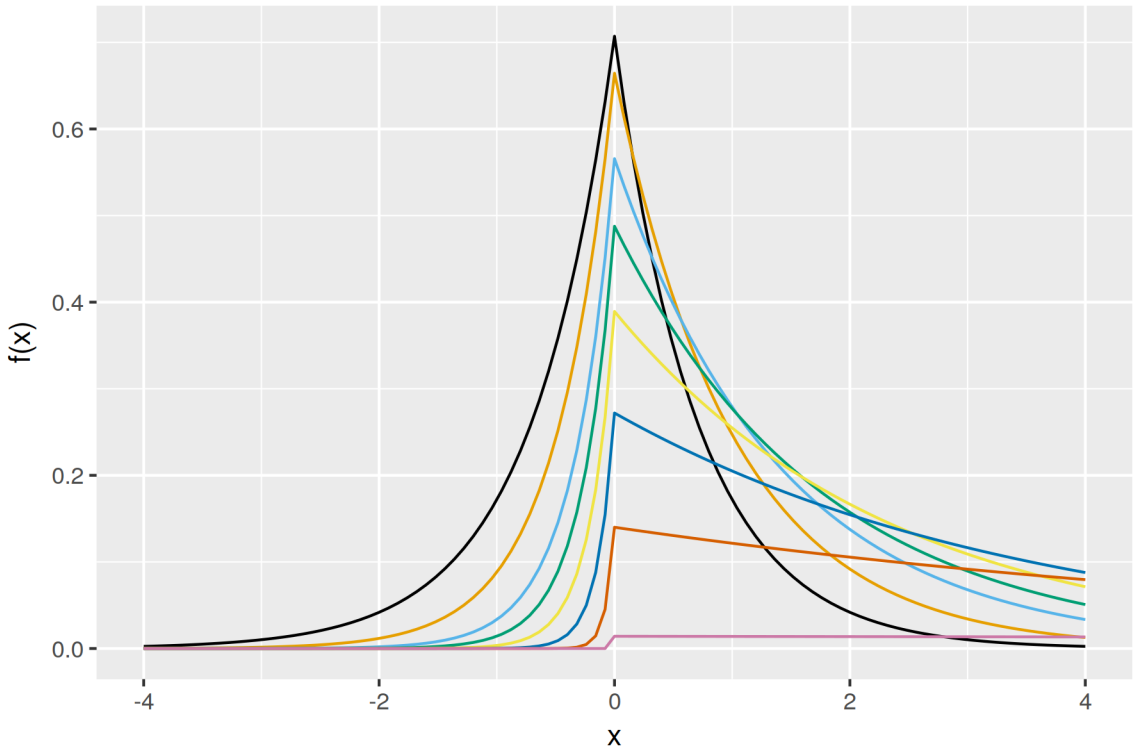

Figure 1.1 Asymmetric Laplace densities with $\sigma = 1$ and $\kappa = 1$ (black), 0.7, 0.5, 0.4, 0.3, 0.2, 0.1, 0.01 (pink).

Proof. The proof is due to Kotz et al. (2001). Begin by computing the CF of the right hand side. Note that the density of W is given by $f(w) = e^{-w}$. Then

$$\begin{aligned}\mathbb{E}(\exp\{it(\theta+\mu W+\sigma\sqrt{W}Z)\}) &= \mathbb{E}[\mathbb{E}(e^{it(\theta+\mu W+\sigma\sqrt{W}Z)})|W] \\ &= \int_0^\infty e^{it\theta+it\mu w}\mathbb{E}[e^{it\sigma\sqrt{w}Z}]e^{-w}dw.\end{aligned} \tag{1.5.3}$$

Now we note that $\mathbb{E}[e^{it\sigma\sqrt{w}Z}] = \phi_Z(\sigma\sqrt{w}t) = e^{\frac{-1}{2}t^2\sigma^2 w}$. So,

$$\begin{aligned}\mathbb{E}(\exp\{it(\theta+\mu W+\sigma\sqrt{W}Z)\} &= \int_0^\infty e^{it\theta+it\mu w-\frac{1}{2}t^2\sigma^2 w-w} \\ &= \frac{e^{it\theta}}{1+\frac{1}{2}t^2\sigma^2-it\mu}.\end{aligned} \tag{1.5.4}$$

□

Remark 1.5.5. *From proposition 1.5.1,* $Y \overset{d}{=} \theta + \frac{\sigma}{\sqrt{2}}\left(\frac{1}{\kappa}W_2 - \kappa W_1\right)$. *Because* $W \overset{d}{=} -\log(U)$, *where* $U \sim \text{Uniform}[0,1]$. *Letting* $U_1, U_2 \sim \text{Uniform}[0,1]$, *we also have that* $Y \overset{d}{=} \theta + \frac{\sigma}{\sqrt{2}}\log\left(\frac{U_1^\kappa}{U_2^{1/\kappa}}\right)$.

The symbol "$\sim$" is taken to mean "distributed as".

Proposition 1.5.6. *Let* $Y \sim \mathscr{ALS}^*(\theta,\kappa,\sigma)$ *with* $\theta = 0$ *and* $\sigma = 1$, *and let* (X_1,X_2) *and* (X_3,X_4) *be independent bivariate Normal random variables with mean zero and variance - covariance matrix*

$$\Sigma = \frac{1}{2\kappa}\begin{bmatrix}1+\kappa^2 & 1-\kappa^2 \\ 1-\kappa^2 & 1+\kappa^2\end{bmatrix}, \tag{1.5.7}$$

then $Y \overset{d}{=} X_1X_2 + X_3X_4$.

1.6 Parameter Estimation

We are aware of two methods of parameter estimation in the literature for the asymmetric Laplace distribution, Method of Moments (MM) and maximum likelihood (ML). We will discuss each of these methods in this section. The deficiencies of the ML method are demonstrated but an improvement is proposed in Chapter 4.

Representation	**Variables**
$\theta + \mu W + \sqrt{W}Z$	$Z \sim \mathcal{N}(0,1)$, $W \sim \text{Exp}(1)$, Z, W independent
$\theta + \frac{1}{\sqrt{2}}\left(\frac{1}{\kappa}W_1 - \kappa W_2\right)$	W_1, W_2 i.i.d. $\sim \text{Exp}(1)$
$\theta + \frac{1}{\sqrt{2}}\log\left(\frac{U_1^\kappa}{U_2^{1/\kappa}}\right)$	U_1, U_2 i.i.d. $\sim \text{Uniform}[0,1]$
$X_1X_2 + X_3X_4$	$(X_1, X_2), (X_3, X_4)$ i.i.d. $\mathcal{N}_2(0, \Sigma)$ where Σ is given in Proposition 1.5.7

Table 1.2 Common Representations of $Y \sim \mathcal{AL}^*(\theta, \kappa, \sigma)$.

1.6.1 Method of Moments

MM estimators (Casella and Berger, 2002) have closed form solutions for the parameters of the asymmetric Laplace distribution. For simplicity we assume the location parameter θ is known and without loss of generality equal to 0. In this case (Kotz et al., 2001) show that the MM estimates of μ and σ are given by

$$\widetilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$\widetilde{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\bar{X}^2}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

1.6.2 Maximum Likelihood Estimation

The ML estimation problem for asymmetric Laplace data has been given thorough treatment in Kotz et al. (2001). We will describe the two most common scenarios; when the location parameter θ is known, and when none of the parameters are known a priori.

1.6.3 Location Parameter Known

The density of a random variable $X \sim \mathscr{ALL}^*(\theta, \kappa, \sigma)$ is given in Definition 1.3.4. The likelihood function of the parameters θ, σ, and κ under this distribution is given by

$$L(\theta, \kappa, \sigma) = \frac{2^{n/2}}{\sigma^n} \frac{\kappa^n}{(1+\kappa^2)^n} \exp\left(\frac{-\sqrt{2}\kappa}{\sigma} \sum_{i=1}^{n} (x_i - \theta)^+ - \frac{\sqrt{2}}{\kappa\sigma} \sum_{i=1}^{n} (x_i - \theta)^- \right), \tag{1.6.1}$$

where

$$(x_i - \theta)^+ = \begin{cases} x_i - \theta & \text{if } x_i \geq \theta; \\ 0 & \text{if } x_i \leq \theta, \end{cases}$$

and

$$(x_i - \theta)^- = \begin{cases} \theta - x_i & \text{if } x_i \leq \theta; \\ 0 & \text{if } x_i \geq \theta. \end{cases}$$

The log likelihood is therefore

$$\ell(\theta, \kappa, \sigma) = \frac{n}{2}\log(2) - n\log(\sigma) + n\log\left(\frac{\kappa}{1+\kappa^2}\right) - \frac{\sqrt{2}}{\sigma}\left(\kappa \sum_{i=1}^{n} (x_i - \theta)^+ + \frac{1}{\kappa} \sum_{i=1}^{n} (x_i - \theta)^- \right). \tag{1.6.2}$$

A simplification of ℓ is given, which will be useful for the purpose of optimization. Let

$$Q(\kappa, \sigma) = \log(\kappa) - \log(1+\kappa^2) - \log(\sigma) - \frac{\sqrt{2}}{\sigma}\left[\kappa, \frac{1}{\kappa}\right] \bar{\mathbf{Z}}, \tag{1.6.3}$$

where

$$\bar{\mathbf{Z}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}^{(i)} = \left[\frac{1}{n}\sum_{i=1}^{n} Z_1^{(i)},\ \frac{1}{n}\sum_{i=1}^{n} Z_2^{(i)} \right]' = \left[\frac{1}{n}\sum_{i=1}^{n} (x_i - \theta)^+,\ \frac{1}{n}\sum_{i=1}^{n} (x_i - \theta)^- \right]'.$$

There are three cases to consider;

1. $x_{(1)} < \theta < x_{(n)}$,

2. $\theta < x_{(1)}$, and

3. $\theta > x_{(n)}$.

Cases 2 and 3 are dealt with in detail on p. 167 of Kotz et al. (2001). In case 2, the parameters that maximize the likelihood are $\kappa = 0$ and $\sigma = 0$. In case 3, $\kappa = \infty$ and $\sigma = 0$ jointly maximize the likelihood. Clearly, these solutions are not admissible because they do not lie in parameter space of (σ, κ). However it is plausible to conclude that the underlying distribution in these cases is a single tailed exponential distribution.

Case 3 admits a genuine solution. Taking derivatives,

$$\begin{aligned} \frac{\partial Q(\kappa,\sigma)}{\partial \sigma} &= -\frac{1}{\sigma} + \frac{\sqrt{2}}{\sigma^2}\left[\kappa, \frac{1}{\kappa}\right]\bar{\mathbf{Z}} = 0; \\ \frac{\partial Q(\kappa,\sigma)}{\partial \kappa} &= \frac{1}{\kappa} - \frac{2\kappa}{1+\kappa^2} - \frac{\sqrt{2}}{\sigma}\left[1, -\frac{1}{\kappa^2}\right]\bar{\mathbf{Z}} = 0. \end{aligned} \tag{1.6.4}$$

Equivalently, we can write

$$\begin{aligned} \left[-\kappa, \frac{1}{\kappa^2}\right]\bar{\mathbf{Z}} &= 0; \\ \sqrt{2}\left[\kappa, \frac{1}{\kappa}\right]\bar{\mathbf{Z}} &= \sigma. \end{aligned} \tag{1.6.5}$$

The system of equations (1.6.5) admits the following explicit solutions for $\hat{\kappa}$ and $\hat{\sigma}$ (Kotz, Kozubowski, and Podgórski, 2002)

$$\hat{\kappa} = \frac{\sqrt[4]{\beta(\theta)}}{\sqrt[4]{\alpha(\theta)}}, \tag{1.6.6}$$

and

$$\hat{\sigma} = \sqrt{2}\sqrt[4]{\alpha(\theta)}\sqrt[4]{\beta(\theta)}\left(\sqrt{\alpha(\theta)} + \sqrt{\beta(\theta)}\right). \tag{1.6.7}$$

1.6.4 Unknown Parameters

We consider the case where no parameters are known a priori. This scenario was first studied in Hartley and Revankar (1974) and Hinkley and Revankar (1977). To optimize the likelihood function, we may maximize the log-likelihood function $\ell(\theta, \kappa, \sigma)$, which is equivalent to

maximizing the function

$$\ell'(\theta,\kappa,\sigma) = -\log(\sigma) + \log\left(\frac{\kappa}{1+\kappa^2}\right) - \frac{\sqrt{2}}{\sigma}\left(\kappa\alpha(\theta) + \frac{1}{\kappa}\beta(\theta)\right), \tag{1.6.8}$$

where the functions α and β are given by

$$\alpha(\theta) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)1_{x_i>\theta}(\theta) \qquad \text{and} \qquad \beta(\theta) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)1_{x_i<\theta}(\theta). \tag{1.6.9}$$

As in Section 1.6.3, we consider three separate cases.

1. $x_{(1)} < \theta < x_{(n)}$,
2. $\theta < x_{(1)}$, and
3. $\theta > x_{(n)}$.

We may disregard situations (2) and (3) because, while θ technically has an unrestricted parameter space $\mathbb{R}$, if $\hat{\theta}$ does not lie within the range of the data, the asymmetric Laplace distribution may provide a poor fit to the data.

Now consider the parameter estimates $\hat{\sigma}$ and $\hat{\kappa}$ from equations (1.6.7) and (1.6.6) derived from the case in which the parameter θ is known. If $\hat{\theta} > x_{(1)}$ and $\hat{\theta} < x_{(n)}$ then $\alpha(\theta) > 0$ and $\beta(\theta) > 0$. Therefore,

$$\ell'(\theta,\kappa,\sigma) \leq \ell'(\theta,\hat{\sigma},\hat{\kappa}) \tag{1.6.10}$$

which is equivalent to

$$\ell'(\theta,\kappa,\sigma) \leq g(\theta), \tag{1.6.11}$$

where

$$g(\theta) = -\log(\sqrt{2}) - 2\log\left(\sqrt{\alpha(\theta)} + \sqrt{\beta(\theta)}\right) - \sqrt{\alpha(\theta)}\sqrt{\beta(\theta)}. \tag{1.6.12}$$

We will show that it is sufficient to examine the reduced function (1.6.12) rather than work with the original reduced log-likelihood function ℓ'. After determining the value of θ that maximizes

the function g, we can recalculate the parameter estimates $\hat{\sigma}$ and $\hat{\kappa}$ using $\hat{\theta} = \text{argmax} g(\theta)$. We restrict our attention to the function g when $\theta \in (x_{(1)}, x_{(n)})$. We may instead consider minimizing the function

$$h(\theta) = 2\log\left(\sqrt{\alpha(\theta)} + \sqrt{\beta(\theta)}\right) + \sqrt{\alpha(\theta)}\sqrt{\beta(\theta)}. \tag{1.6.13}$$

Lemma 1.6.14. *The function h (1.6.13) is continuous on the closed interval* $[x_{(1)}, x_{(n)}]$ *and concave down on each of the subintervals* $(x_{(i)}, x_{(i+1)})$ *for* $i = 1, 2, \ldots, n-1$.

In light of Lemma 1.6.14, the function h has a local minimum at each of the points $x_{(1)}, \ldots, x_{(n)}$. A global minimum must lie in this finite set.

We may check for the global minimum using exhaustion and set $\hat{\theta} = x_{(r)}$ where

$$x_{(r)} = \text{argmin}(h(x_{(i)})), \quad i = 1, \ldots, n. \tag{1.6.15}$$

Once the global minimum $x_{(r)}$ is determined, the maximum likelihood estimators of the remaining parameters may be computed by maximizing function (1.6.10). We find using techniques from calculus that

$$\hat{\kappa} = \frac{\sqrt[4]{\beta(\hat{\theta})}}{\sqrt[4]{\alpha(\hat{\theta})}} \tag{1.6.16}$$

and

$$\hat{\sigma} = \sqrt{2}\sqrt[4]{\alpha(\hat{\theta})}\sqrt[4]{\beta(\hat{\theta})}\left(\sqrt{\alpha(\hat{\theta})} + \sqrt{\beta(\hat{\theta})}\right) \tag{1.6.17}$$

1.6.5 Deficiencies in Maximum Likelihood Estimation

A simulation is performed to examine the applicability of the maximum likelihood estimators detailed in Section 1.6.4. In this simulation study, the sample size is allowed to vary within $\{30, 50, 100, 200\}$, and the skewness parameter within the set $\kappa \in \{1, 1.1, 1.2, 1.5, 2, 3, 5, 10\}$. The effect of increasing the parameter κ corresponds to a greater left skew.

For each κ, n deviates from a $\mathscr{ALL}^*(\theta = 0, \kappa, \sigma = \sqrt{2})$ distribution are drawn, and the ML estimators of Section 1.6.4 are computed. The process is repeated 500 times and the ML

parameter estimate of κ is recorded. If $\hat{\theta} = x_{(1)}$ or $\hat{\theta} = x_{(n)}$, an admissible ML estimate of κ does not exist, so $\hat{\kappa}$ is recorded as `NA`.

One of the striking findings of the simulation is that frequently the ML estimates fail to exist, even under moderate skewness and sample size. Table 1.3 shows the results of the simulation: each cell contains the estimated probability that bona fide ML estimates can be determined from genuine asymmetric Laplace data. For a fixed κ, it is clear that frequently the estimates of $\hat{\theta}$ fall on the first or last order statistic, leading in estimates of κ and σ that do not lie in their respective parameter spaces. Even under the assumption that there is no skewness present in the data ($\kappa = 1$), we still find that parameter estimation fails 25% of the time in samples of size 30. The problem is clearly exacerbated as κ increases. Larger sample sizes may be taken as a countermeasure against this estimation failure, but in practice, situations where additional data collection is possible are unlikely. The findings of this simulation were impetus for developing

	$\kappa = 1$	1.1	1.2	1.5	2	3	5	10
$n = 30$	0.252	0.290	0.250	0.158	0.050	0.002	0.000	0.000
50	0.514	0.504	0.466	0.382	0.100	0.002	0.000	0.000
100	0.838	0.836	0.816	0.630	0.428	0.040	0.000	0.000
200	0.982	0.976	0.974	0.906	0.706	0.274	0.000	0.000

Table 1.3 Estimated probability $\mathscr{AL}^*$ ML estimates exist

alternative parameter estimation techniques with attractive finite sample properties, which are discussed in Chapter 4.

Fragiadakis and Meintanis (2009) found that a modification to the function h in equation 1.6.13 may lead to more reliable maximum likelihood estimates. They consider minimizing the function

$$h'(\theta) = 2\log\left(\sqrt{\beta(\theta)} + \sqrt{\alpha(\theta)}\right) \tag{1.6.18}$$

over each of the sample order statistics $x_{(1)}, \ldots, x_{(n)}$ to arrive at the ML estimate of location, $\hat{\theta} = x_{(r)} = \text{argmin}(h'(x_{(i)}), i = 1, \ldots, n$. A graphical comparison of functions (1.6.13) and (1.6.18) is shown in Figure 1.2 for a sample data set of 10 observations from the

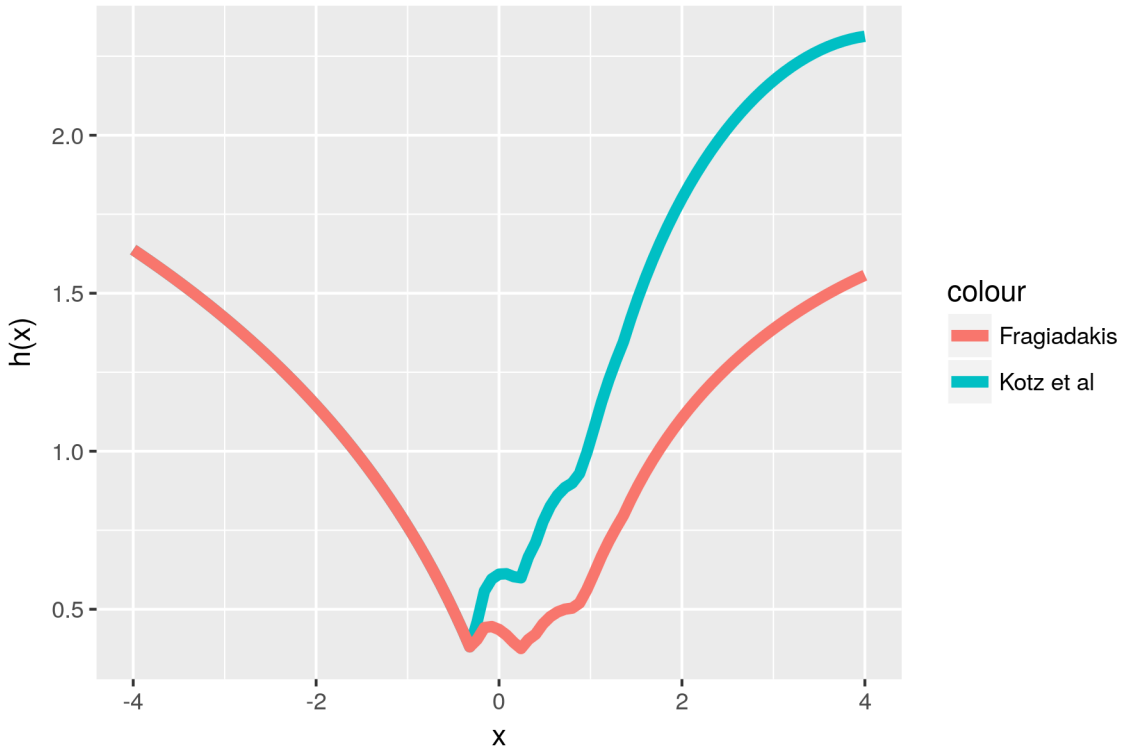

Figure 1.2 Plot of functions (1.6.13) and (1.6.18) for a random sample of size 10 from $\mathscr{ALL}^*(\theta = 0, \sigma = \sqrt{2}, \kappa = \frac{1}{2})$

$\mathscr{ALL}^*(\theta = 0, \kappa = \frac{1}{2}, \sigma = \sqrt{2})$ distribution. In this case, we can observe that the two objective functions choose the same estimate of $\hat{\theta}$.

We find that the estimates of the parameters chosen by optimizing (1.6.18) tend to exist more frequently than those chosen by the ML method of Kotz et al. (2001). Fragiadakis and Meintanis (2009) also claim that the estimators $(\hat{\theta}, \hat{\sigma}, \hat{\kappa})$ are more accurate and efficient, but it is difficult to validate this claim when the ML estimators detailed in (Kotz et al., 2001) frequently fail to exist.

A simulation study was performed to ascertain the frequency of existence under the same conditions that produced Table 1.3. The findings are shown in Table 1.4. Comparison of Tables 1.3 and 1.4 make it clear that using the new objective function (1.6.18) yields legitimate estimators more often.

	$\kappa = 1$	1.1	1.2	1.5	2	3	5	10
$n = 30$	0.978	0.960	0.964	0.882	0.710	0.378	0.088	0.050
50	1.000	1.000	1.000	0.982	0.926	0.622	0.248	0.046
100	1.000	1.000	1.000	0.998	0.994	0.902	0.446	0.058
200	1.000	1.000	1.000	1.000	1.000	0.994	0.804	0.194

Table 1.4 Estimated probability $\mathscr{A}\mathscr{L}^*$ maximum likelihood estimates exist (using 1.6.18)

CHAPTER 2 REVIEW OF EXISTING GOODNESS OF FIT TESTS

In this chapter, we review previous work developed to test the goodness-of-fit hypothesis for asymmetric Laplace distribution. Many goodness-of-fit tests are developed with some generality and applied specifically to this distribution.

The goodness-of-fit problem for asymmetric Laplace distributions may be formulated as follows. Let $X_1, \ldots, X_n$ be independent and identically distributed observations from a distribution F. Then two types of null hypotheses are typically considered:

$$H_0 : F = \mathscr{A}\mathscr{L}^*(\theta = \theta_0, \kappa = \kappa_0, \sigma = \sigma_0),$$

where the ordered triple of parameters $(\theta_0, \kappa_0, \sigma_0)$ is specified. or

$$H_0 : F \in \mathscr{A}\mathscr{L}^*(\theta, \kappa, \sigma).$$

This second hypothesis corresponds to the case where the parameters (θ, κ, σ) are all or partially unknown. In the literature, the first type of null hypothesis is known as a *simple* hypothesis and the second type of hypothesis is known as a *composite* hypothesis. Although these hypotheses look similar in notation, in practice they require different treatment. It is important to be mindful about the meaning of the decision of these tests. A rejection of the hypothesis means that data are not consistent with the distribution F: It is unlikely that F could have generated the data $X_1, \ldots, X_n$. However a failure to reject the hypothesis doesn't imply F generated the data, nor does it imply that F is the best fit for the data. Thus, a goodness-of-fit test may not provide an absolute indication of the usefulness of a parametric model, but may help practitioners select from a set of possible models.

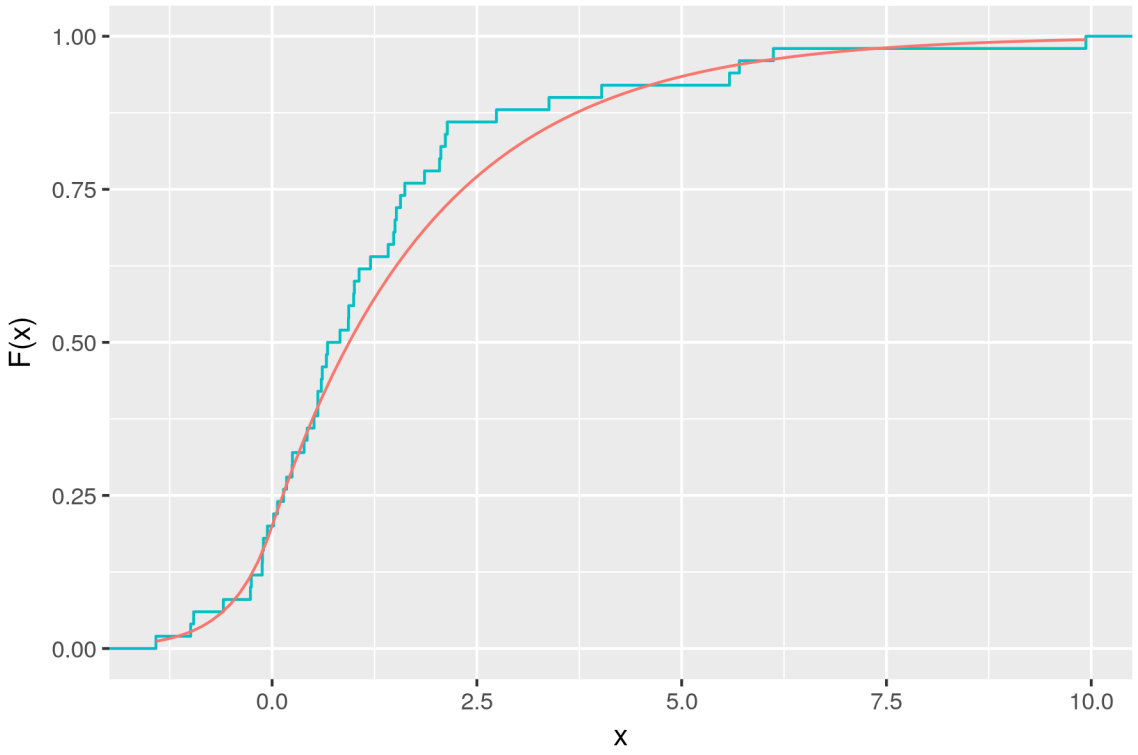

Figure 2.1 Comparison of the EDF (blue) of a random sample of 50 observations from $\mathscr{ALL}^*(\theta = 0, \sigma = \sqrt{2}, \kappa = 1/2)$ to the true distribution function (red)

2.1 Tests Based on the Empirical Distribution Function

The oldest and most widely used goodness-of-fit tests are based on empirical distribution functions. Many of these tests have a general form that makes them applicable to a wide range of distributions.

The empirical distribution function (EDF) is

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{x_i \leq x}. \tag{2.1.1}$$

These functions approximate the true distribution function of the data by a step-wise function. Use of EDFs is justified by the strong law of large numbers and the Glivenko-Cantelli theorem, which states that under the null hypothesis, the empirical distribution function converges uniformly to F, the true distribution function, as sample size n tends to infinity.

Therefore, to test the hypothesis of the data being generated by a specific distribution, one may check that the distance between the EDF and hypothesized distribution function is small.

Theorem 2.1.2. *Glivenko-Cantelli Theorem*

$$||\hat{F}_n(x) - F(x)||_\infty = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{a.s.} 0$$

as $n \to \infty$.

The left-hand side of Theorem 2.1.2 is the test statistic of the Kolmogorov-Smirnov test for the hypothesis $H_0 : X_1, \ldots, X_n \sim F$ (William, 1971; Durbin, 1973),

$$D_n = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)|.$$

When D_n is larger than the critical value $K_{\alpha,n}$, we reject the hypothesis that the data was generated by the distribution F at the α significance level. That is, there is evidence that the data $X_1, \ldots, X_n$ do not arise from the hypothesized distribution F.

Other measures of distance between the EDF and the hypothesized distribution function have also been proposed based on replacing the sup norm in Theorem 2.1.2 with the L_2 norm:

$$\int_{\mathbb{R}} n(\hat{F}_n(x) - F(x))^2 w(x)\, dF(x), \tag{2.1.3}$$

where $w(x)$ is a weight function. When $w(x) = 1$, the statistic is the Cramer-von Mises statistic. When $w(x) = (F(x)(1 - F(x)))^{-1}$, the statistic is the Anderson-Darling statistic (DasGupta, 2008). The weight function used in Anderson-Darling distance has the effect of placing more weight on outliers in the data. The statistical distances (2.1.3) are implemented by transforming the original order statistics $X_{(1)}, \ldots, X_{(n)}$ to uniform order statistics $u_{(i)} = F(x_{(i)})$ and computing the following goodness-of-fit statistics.

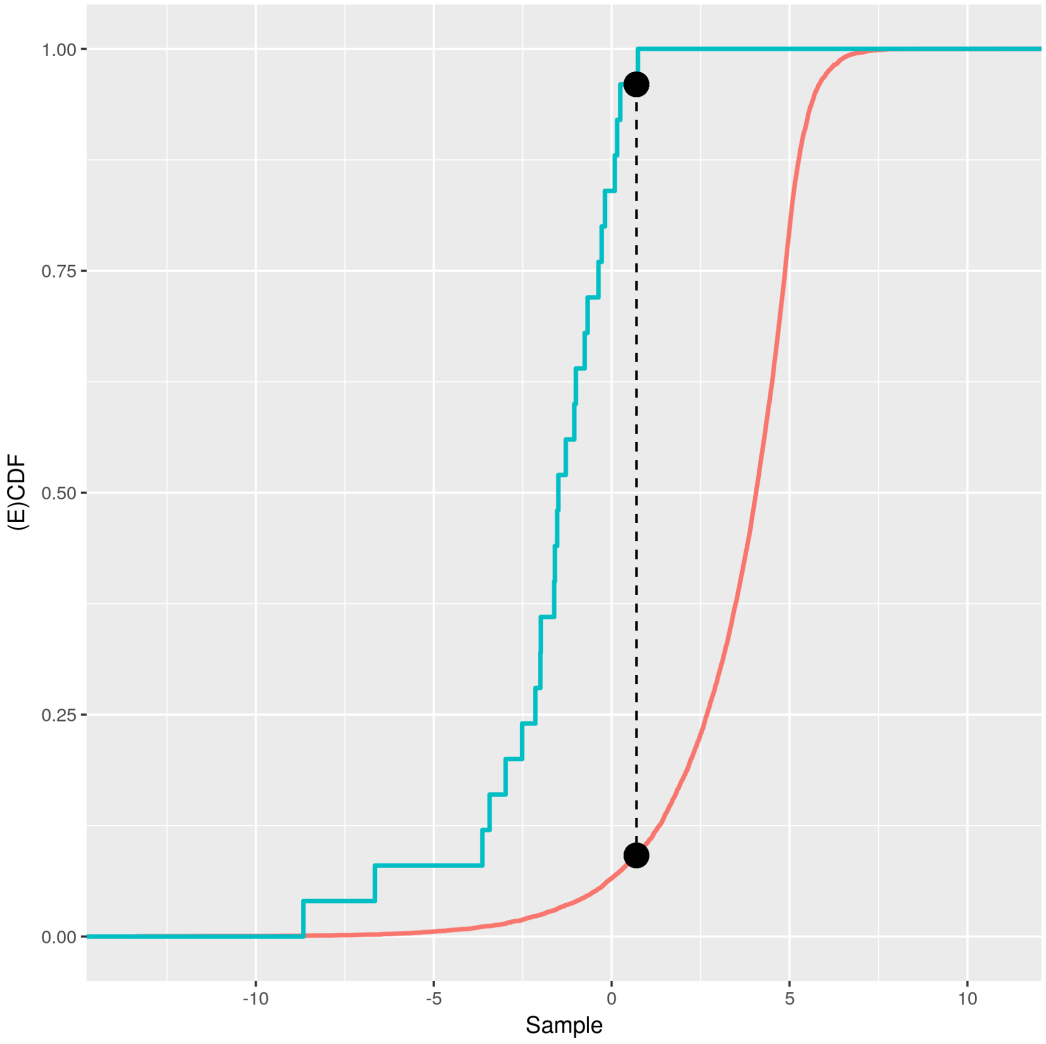

Figure 2.2 Illustration of Kolmogorov-Smirnov statistic D_n. The hypothesized distribution (red) is $\mathscr{ALS}^*(5,2,\sqrt{2})$

The Cramer-von Mises statistic (D'Agostino and Stephens, 2017):

$$C_n = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - F(x_{(i)}) \right)^2, \tag{2.1.4}$$

or the Anderson-Darling test statistic (D'Agostino and Stephens, 2017):

$$A_n = -n - \frac{1}{n} \left(\sum_{i=1}^{n} (2i-1)(\log(F(x_{(i)})) + \log(1 - F(x_{n-i+1}))) \right). \tag{2.1.5}$$

Performing goodness-of-fit tests with the statistics C_n and A_n is similar to using the statistic D_n; large values of the test statistics lead to a rejection of the null hypothesis at the prescribed level α. Puig and Stephens (2007) detail steps to test the composite hypothesis of asymmetric Laplacity.

A comparison of the Anderson-Darling, Cramer-von Mises, and Kolmogorov Smirnov test was studied by Chen (2002), concluding that the Anderson Darling test is the overall most powerful against the alternatives considered when the null hypothesis is the Laplace distribution. However, Chen did not include the asymmetric Laplace distribution in the study.

2.2 Tests Based on the Empirical Characteristic Function

Goodness-of-fit tests based on the empirical characteristic function present an alternative method to testing the goodness-of-fit hypothesis.

The empirical characteristic function (ECF) of a sample $x_1, \ldots, x_n$ is given by

$$\hat{\phi}(t) = \frac{1}{n} \sum_{j=1}^{n} \exp\left(itx_j\right), \tag{2.2.1}$$

where $i = \sqrt{-1}$. Under the null hypothesis, we expect the ECF to be close to the hypothesized CF, under some measure of distance between functions.

A goodness-of-fit test for the asymmetric Laplace based on the ECF was introduced in Fragiadakis and Meintanis (2009), although they consider a different parameterization of the

distribution, with the CF

$$\phi_Y(t) = \frac{\exp(i\theta t)}{1+\sigma^2 t^2 - i(\frac{1}{k}-k)\sigma t}, \tag{2.2.2}$$

which is only minimally different than the CF given in Definition 1.3.2. Under this parameterization, if the hypothesis of asymmetric Laplacity is true

$$D(t|\theta,\kappa,\sigma) \equiv \left(1+\sigma^2 t^2 - i\left(\frac{1}{\kappa}-\kappa\right)t\right)\phi_Y(t) - \exp(i\theta t) = 0. \tag{2.2.3}$$

So a reasonable goodness-of-fit test can rely on computing the deviation from 0 of the $D(t|\theta,\kappa,\sigma)$ from the standardized sample $Z_i = \frac{X_i-\hat{\theta}}{\hat{\sigma}}$. Use of this transformation is justified because the parameters θ and σ are location and scale parameters. Because the maximum likelihood estimates of these parameters are consistent (Kotz et al., 2001), the standardized sample tends to the $\mathscr{AL}^*(\theta=0,\sigma=1,\kappa)$ distribution. Therefore the function

$$\hat{D}(t|\hat{\kappa}) = \left(1+t^2 - i\left(\frac{1}{\hat{\kappa}}-\hat{\kappa}\right)t\right)\hat{\phi}_Z(t) - 1 \tag{2.2.4}$$

tends to 0 under the hypothesis $H_0 : F \in \mathscr{AL}$. The test statistic to measure the deviation from 0 is

$$\hat{T}_w = n\int_{\mathbb{R}} |\hat{D}(t|\hat{\kappa})|^2 w(t)\,dt, \tag{2.2.5}$$

where the function $w(t)$ is a weight function. The statistic 2.2.5 may be computed as

$$\begin{aligned}\hat{T}_w = \frac{1}{n}\sum_{j,k=1}^{n}&\left[I_c^{(4)}(z_j-z_k) + \left(2+\left(\frac{1}{\hat{\kappa}}-\hat{\kappa}\right)^2\right)I_c^{(2)}(z_j-z_k) + I_c^{(0)}(z_j-z_k)\right] \\ &+ nI_c^{(0)}(0) - 2\sum_{j=1}^{n}\left[I_c^{(2)}(z_j) + \left(\frac{1}{\hat{\kappa}}-\hat{\kappa}\right)I_s(z_j) + I_c^{(0)}(z_j)\right]\end{aligned} \tag{2.2.6}$$

where

$$I_c^{(m)}(b) = \int_{\mathbb{R}} t^m \cos(bt) w(t)\,dt, \qquad m = 0,2,4, \tag{2.2.7}$$

and

$$I_s(b) = \int_{\mathbb{R}} t \sin(bt) w(t)\, dt. \tag{2.2.8}$$

The statistic $\hat{T}_w$ is most convenient when w admits the integrals (2.2.7) and (2.2.8) with closed form expressions. Fragiadakis and Meintanis (2009) recommend using $w(t) = \exp(-a|t|)$.

Because the distribution of the test statistic (2.2.6) depends on the transformed data $Z_1, \ldots, Z_n$ and the unknown skewness parameter κ, the decision boundary of the test statistic cannot be tabulated. Rather, it needs to be investigated computationally for each sample depending on the skewness estimate $\hat{\kappa}$. Fragiadakis and Meintanis (2009) propose the following method, relying on a parametric bootstrap. The procedure is as follows:

1. Given the sample $X_1, \ldots, X_n$, compute the parameter estimates $(\hat{\theta}, \hat{\sigma}, \hat{\kappa})$ and the standardized data $Z_1, \ldots, Z_n$,

2. Calculate the test statistics T_w based on the Z_i's and parameter estimates.

3. (a) Generate a bootstrap sample $Z_1^*, \ldots, Z_n^* \sim \mathscr{ALS}^*(\theta = 0, \kappa = \hat{\kappa}, \sigma = 0)$.

   (b) Calculate the test statistic T_w^* based on the bootstrapped data and the parameter estimates $(\theta = \hat{\theta}^*, \kappa = \hat{\kappa}^*, \sigma = \hat{\sigma}^*)$.

   Repeat steps (a) and (b) M times, to calculate M bootstrap replicates of T_w^*.

4. Obtain the level α decision boundary, C_{T_w}, the $(1-\alpha)$-quantile of $T_{w,1}^*, \ldots, T_{w,M}^*$.

Simulations presented in Fragiadakis and Meintanis (2009) show their ECF test can be slightly more powerful than the AD test in certain situations. Unfortunately a tuning parameter needs to be set in the weight function $w(t) = \exp(-a|t|)$ which complicates the use of this test. The most appropriate value of the tuning parameter a depends on the distribution of the sample data, which generally won't be known in practice. Even if the distribution of the data is known (for example, in a simulation study) a search still needs to be performed to find an appropriate value for a. For these two reasons, we don't consider this ECF test as appropriate as the traditional, general tests such as Anderson-Darling and Kolmorgorov-Smirnov.

2.3 Other Goodness-of-Fit tests for Laplace Distribution

Visual inspection of Probability-Probability (P-P) plots or Quantile-Quantile (Q-Q) plots is a reasonable method for assessing the goodness-of-fit of a data set to a distribution. These visual tests serve in particular as excellent methods for diagnosing poor goodness-of-fit in the tails of the distribution. We recommend the use of Q-Q plots to practitioners that seek to answer questions of "Is my data close to being asymmetric Laplace?" rather than the binary goodness-of-fit question that is studied in this dissertation.

Choi and Kim (2006) suggest a goodness-of-fit test for the symmetric Laplace distribution based on its maximum entropy property. They show through simulation that their entropy based test is more powerful than EDF tests against several skewed and non-skew alternative hypotheses. The author is not aware of an attempt to extend this type of goodness-of-fit test to the asymmetric Laplace distribution. Other tests for the goodness-of-fit hypothesis for the symmetric Laplace distribution exist, for example Yen and Moore (1988) and Rublík (1997), but neither of these tests have been extended to the asymmetric case.

CHAPTER 3 UNIVARIATE ENERGY STATISTICS

In this chapter, a new goodness-of-fit test for the univariate asymmetric Laplace distribution is presented based on energy distance between the ECF and the asymmetric Laplace CF given in Definition 1.3.1. New results for the distance standard deviation of the asymmetric Laplace distribution are also shown. An energy goodness-of-fit test for the generalized asymmetric Laplace distribution (also called the Variance Gamma distribution) is derived under a parameter restriction.

3.1 Energy Distance Preliminaries

Energy distance, first described in (Székely, 2000), is a statistical distance between distributions. Functions of energy distances are called *energy statistics*. The name *energy* refers to the inherent "potential" energy of the data with respect to a hypothesized distribution. Justification for usage of the term "energy" can be found in Székely and Rizzo (2013).

Definition 3.1.1. *(Energy Distance, Székely and Rizzo (2013))*

The energy distance between two independent d-dimensional random variables X and Y is

$$\mathscr{E}(X,Y) = 2\mathbb{E}||X-Y||_d - \mathbb{E}||X-X'||_d - \mathbb{E}||Y-Y'||_d, \tag{3.1.2}$$

provided $\mathbb{E}||X|| < \infty$, $\mathbb{E}||Y|| < \infty$. Here X' is an i.i.d. copy of X and Y' is an i.i.d. copy of Y.

Remark 3.1.3. *When $d = 1$ the norms $||\cdot||$ in (3.1.2) are absolute values,*

$$\mathscr{E}(X,Y) = 2\mathbb{E}|X-Y| - \mathbb{E}|X-X'| - \mathbb{E}|Y-Y'|.$$

Remark 3.1.4. *The observations need not lie in space $\mathbb{R}^d$. Energy distance generalizes to separable Hilbert spaces as long as the underlying metric is conditionally negative definite (Székely and Rizzo, 2017).*

Energy distance measures the statistical potential energy between two distributions. We liken statistical distributions in a probability space to physical objects in the universe. As Newton's relative potential energy between objects in space is large when the bodies are far apart, intuitively, if two distributions are dissimilar, their statistical energy is large as well. When two statistical distributions have energy distance equal to 0, the distributions must agree. This matches the physical notion that potential energy of an object with respect to itself is 0. These ideas are stated in greater detail in Székely and Rizzo (2017).

Theorem 3.1.5. *(Energy distance properties)*

1. $\mathscr{E}(X,Y) \geq 0$
2. $\mathscr{E}(X,Y) = 0$ *if and only if* $X \overset{d}{=} Y$.

The square root of energy distance is a metric on the space of d-dimensional distribution functions (Székely and Rizzo, 2017) and therefore is suitable to handling hypotheses of the form $H_0 : F = G$, where F and G are distribution functions. Suppose X and Y are random variables with CFs $\phi_X(t)$ and $\phi_Y(t)$. The distance between two characteristic functions may be measured using a weighted integral,

$$\int_{\mathbb{R}} |\phi_X(t) - \phi_Y(t)|^2 w(t)\, dt. \tag{3.1.6}$$

According to Székely and Rizzo (2005), energy distance $\mathscr{E}(X,Y)$ is related to distance between CFs.

Proposition 3.1.7. *If d-dimensional random variables X and Y are independent with finite first absolute moments and CFs $\phi_X(t)$ and $\phi_Y(t)$, then their energy distance is*

$$\mathscr{E}(X,Y) = \frac{1}{c_d} \int_{\mathbb{R}} \frac{|\phi_X(t) - \phi_Y(t)|^2}{||t||_d^{d+1}}\, dt, \tag{3.1.8}$$

where

$$c_d = \frac{\pi^{(d+1)/2}}{\Gamma(\frac{d+1}{2})}. \tag{3.1.9}$$

3.1.1 One Sample Univariate Energy Goodness-of-Fit Test

Energy statistics may be formulated as V-statistics or U-statistics. It is traditional to write energy statistics as V-statistics to ensure the non-negativity of the statistic. Preventing the statistic from attaining negative values allows us to more closely identify energy distance as a statistical distance. However, the use of V-statistics does introduce bias in the estimation of energy distance.

Suppose a univariate random sample $X_1, \ldots, X_n$ is collected. Let $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a symmetric *kernel function*. Then energy statistics may be defined as a V-statistic

$$V_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(X_i, X_j) \tag{3.1.10}$$

or U-statistic

$$U_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} h(X_i, X_j) \tag{3.1.11}$$

where the kernel function h is given by

$$h(x,y) = \mathbb{E}|x-Y| + \mathbb{E}|y-Y| - \mathbb{E}|Y-Y'| - |x-y| \tag{3.1.12}$$

for the one sample goodness-of-fit test. Under the hypothesis $H_0 : F = G$, $\mathbb{E}[h(x,Y)] = 0$, so the kernel is said to be degenerate.

Use of this statistic for the energy goodness-of-fit test is bolstered by the theory of V-statistics. Because the kernel h of the statistic $\mathscr{E}_n$ is degenerate, $\mathbb{E}[h^2(Y,Y')] < \infty$, the limiting distribution of the test statistic $Q_n = n\mathscr{E}_n$ is an infinite quadratic form

$$Q_n \to \sum_{i=1}^{\infty} \lambda_i Z_i^2, \tag{3.1.13}$$

where the Z_i are i.i.d. standard Normal random variables and the coefficients λ_i are solutions to the integral equations

$$\int_{\mathbb{R}} h(x,y)\psi_i(y)dF(y) = \lambda_i \psi_i(y), \tag{3.1.14}$$

where h is the kernel function defined in equation (3.1.12) and $\psi_i \in L_2$ are called eigenfunctions. If the weight function $w(t)$ in equation (3.1.6) is $1/\pi t^2$, we arrive at the expression considered in Rizzo (2002):

$$Q_n = n \int_{\mathbb{R}} |\hat{\phi}(t) - \phi_F(t)| \frac{1}{\pi t^2}\, dt. \tag{3.1.15}$$

3.2 Energy Statistic for Asymmetric Laplace Distribution

The asymmetric Laplace energy statistic depends on the derivation of $\mathbb{E}|y-Y|$ and $\mathbb{E}|Y-Y'|$. These expectations were first derived in Rizzo and Haman (2016).

Proposition 3.2.1. *If $Y \sim \mathscr{AL}^*(\theta, \sigma, \kappa)$, then for any fixed $y \in \mathbb{R}$*

$$\mathbb{E}|y-Y| = \begin{cases} y - \theta - \mu + \frac{2p_\kappa}{\lambda} \exp(-\lambda|y-\theta|), & y \geq \theta; \\ -y + \theta + \mu + \frac{2q_\kappa}{\beta} \exp(-\beta|y-\theta|), & y < \theta. \end{cases} \tag{3.2.2}$$

Proof. (Rizzo and Haman (2016), Appendix)

Suppose that $Y \sim \mathscr{AL}^*(\theta, \kappa, \sigma)$ and $x \in \mathbb{R}$ is constant. Then

$$\begin{aligned} \mathbb{E}|x-Y| &= \int_{x \leq y} (y-x) f_Y(y) dy + \int_{x>y} (x-y) f_Y(y) dy \\ &= x(2F_Y(x) - 1) - E(Y) + 2\int_x^\infty y f_Y(y) dy \qquad (3.2.3) \\ &= x(2F_Y(x) - 1) + E(Y) - 2\int_{-\infty}^x y f_Y(y) dy. \qquad (3.2.4) \end{aligned}$$

Case 1: $x \geq \theta$. In this case $y \geq x \geq \theta$ in the integrand in (3.2.3), so that

$$\int_x^\infty y f_Y(y) dy = \int_x^\infty y p_\kappa \lambda e^{-\lambda|y-\theta|} dy = \frac{p_\kappa}{\lambda} e^{-\lambda|x-\theta|}(\lambda x + 1).$$

Thus (3.2.3) can be simplified to

$$\begin{aligned}\mathbb{E}|x-Y| &= x(2(1-p_\kappa e^{-\lambda|x-\theta|})-1)-\mathbb{E}(Y)-\frac{p_\kappa}{\lambda}e^{-\lambda|x-\theta|}(\lambda x+1)\\ &= x-\theta-\mu+\frac{2p_\kappa}{\lambda}e^{-\lambda|x-\theta|}, \qquad x\geq\theta.\end{aligned}$$

Case 2: $x<\theta$. In this case $y\leq x<\theta$ in the integrand in (3.2.4), so that

$$\int_{-\infty}^{x} y f_Y(y)dy = \int_{-\infty}^{x} y p_\kappa \lambda e^{-\beta|y-\theta|}dy = \frac{p_\kappa\lambda}{\beta^2}e^{-\beta|x-\theta|}(x\beta-1).$$

Hence using (3.2.4) we have

$$\begin{aligned}\mathbb{E}|x-Y| &= x(2F_Y(x)-1)+\mathbb{E}(Y)-2\int_{-\infty}^{x} y f_Y(y)dy\\ &= x\left(2q_\kappa e^{-\beta|x-\theta|}-1\right)+\theta+\mu-\frac{2p_\kappa\lambda}{\beta^2}e^{-\beta|x-\theta|}(x\beta-1)\\ &= -x+\theta+\mu-2xe^{-\beta|x-\theta|}\left(q_\kappa-\frac{p_\kappa\lambda}{\beta}\right)+\frac{2\lambda p_\kappa}{\beta^2}e^{-\beta|x-\theta|}\\ &= -x+\theta+\mu+\frac{2q_\kappa}{\beta}e^{-\beta|x-\theta|}, \qquad x<\theta.\end{aligned}$$

In the last step we used the identities $q_\kappa\beta = p_\kappa\lambda$ and $\frac{\lambda p_\kappa}{\beta^2}=\frac{q_\kappa}{\beta}$. □

Proposition 3.2.5. *If $Y\sim\mathscr{ALS}^*(\theta,\kappa,\sigma)$ or $Y\sim\mathscr{ALS}(\theta,\mu,\sigma)$ where $\mu=\frac{\sigma}{\sqrt{2}}(\frac{1}{\kappa}-\kappa)$, then*

$$\mathbb{E}|Y-Y'| = \frac{p_\kappa}{\beta}+\frac{q_\kappa}{\lambda}+\frac{p_\kappa^2}{\lambda}+\frac{q_\kappa^2}{\beta}. \tag{3.2.6}$$

$$= \frac{\sigma}{\sqrt{2}}\left(k+\frac{1}{k}-\frac{1}{k+\frac{1}{k}}\right) \tag{3.2.7}$$

$$= \frac{\sigma}{\sqrt{2}}\left(\sqrt{4+\frac{2\mu^2}{\sigma^2}}-\frac{1}{\sqrt{4+\frac{2\mu^2}{\sigma^2}}}\right). \tag{3.2.8}$$

Proof. (Rizzo and Haman (2016), Appendix)

Generally we can suppose that $\theta = 0$. Then if $Y \sim \mathscr{AL}^*(0,\kappa,\sigma)$ we have $\mathbb{E}|Y-Y'| = L+U$, where, by Proposition 3.2.2,

$$\begin{aligned} L &= \int_{-\infty}^{0} (-y+\mu+\frac{2q_\kappa}{\beta}e^{-\beta|y|})\cdot p_\kappa\lambda e^{-\beta|y|}dy \\ &= \frac{q_\kappa}{\beta}+\mu q_\kappa+\frac{q_\kappa\lambda}{\beta^2} = \frac{q_\kappa}{\beta}+\mu q_\kappa+\frac{q_\kappa^2}{\beta}, \end{aligned}$$

and

$$U = \int_0^{\infty} (y-\mu+\frac{2p_\kappa}{\lambda}e^{-\lambda|y|})\cdot p_\kappa\lambda e^{-\lambda|y|}dy = \frac{p_\kappa}{\lambda}-\mu p_\kappa+\frac{p_\kappa^2}{\lambda}.$$

Observe that $\mu = \frac{\sqrt{2}}{\sigma}(k-\frac{1}{k})$ implies $\mu = \frac{1}{\lambda}-\frac{1}{\beta}$. Substituting, we obtain

$$\begin{aligned} \mathbb{E}|Y-Y'| &= \frac{q_\kappa}{\beta}+\frac{p_\kappa}{\lambda}+\frac{q_\kappa^2}{\beta}+\frac{p_\kappa^2}{\lambda}+\mu(q_\kappa-p_\kappa) \\ &= \frac{q_\kappa}{\beta}+\frac{p_\kappa}{\lambda}+\frac{q_\kappa^2}{\beta}+\frac{p_\kappa^2}{\lambda}+\frac{q_\kappa}{\lambda}-\frac{q_\kappa}{\beta}-\frac{p_\kappa}{\lambda}+\frac{p_\kappa}{\beta} \\ &= \frac{p_\kappa}{\beta}+\frac{q_\kappa}{\lambda}+\frac{p_\kappa^2}{\lambda}+\frac{q_\kappa^2}{\beta}. \end{aligned}$$

To prove (3.2.6) we substitute $p_\kappa = \frac{\beta}{\lambda+\beta}$, $q_\kappa = \frac{\lambda}{\lambda+\beta}$, and apply several identities of the type $\frac{\lambda}{\beta} = \kappa^2$, $\lambda+\beta = \frac{\sqrt{2}}{\sigma}(k+\frac{1}{k})$, etc. After lengthy algebraic manipulation we obtain (3.2.7):

$$\frac{p_\kappa}{\beta}+\frac{q_\kappa}{\lambda}+\frac{p_\kappa^2}{\lambda}+\frac{q_\kappa^2}{\beta} = \frac{\sigma}{\sqrt{2}}\left(k+\frac{1}{k}-\frac{1}{k+\frac{1}{k}}\right)$$

and the right hand side equals (3.2.8) using another identity

$$\sqrt{4+\frac{2\mu^2}{\sigma^2}} = k+\frac{1}{k}.$$

□

The one-sample, univariate energy V-statistic for performing a goodness-of-fit test was introduced in Székely (2000):

$$n\hat{\mathscr{E}}(y_1,\ldots,y_n,X) = n\left(\frac{2}{n}\sum_{i=1}^{n}\mathbb{E}|X-y_i| - \mathbb{E}|X-X'| - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i-y_j|\right) \tag{3.2.9}$$

The statistic depends on the sample $y_1,\ldots,y_n$ and the hypothesized distribution F_x. A goodness-of-fit test statistic based on the energy statistic 3.2.9 and Propositions 3.2.5 and 3.2.1 is

$$\hat{Q}_n \equiv n\hat{\mathscr{E}}(y_1,\ldots,y_n,X) = 2\sum_{i=1}^{n}\begin{cases} y_i-\theta-\mu+\frac{2p_\kappa}{\lambda}\exp(-\lambda|y_i-\theta|), & y_i \geq \theta \\ -y_i+\theta+\mu+\frac{2q_\kappa}{\beta}\exp(-\beta|y_i-\theta|), & y_i < \theta \end{cases} \tag{3.2.10}$$

$$-n\left(\frac{p_\kappa}{\beta}+\frac{q_\kappa}{\lambda}+\frac{p_\kappa^2}{\lambda}+\frac{q_\kappa^2}{\beta}\right) - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i-y_j|.$$

where X represents an asymmetric Laplace random variable with parameters θ, μ, λ, β, p_κ, and q_κ. The energy statistic (3.2.10) has a simple computational form that is amenable to computer implementation. The double sum $\sum_{i,j=1}^{n}|y_i-y_j|$ can be linearized to improve the computational speed of the energy test. Following Rizzo (2002), we may use

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i-y_j| = 2\sum_{k=1}^{n}((2k-1)-n)y_{(k)} \tag{3.2.11}$$

to modify the univariate energy statistic. The notation $y_{(k)}$ denotes the k^{th} order statistic of the sample $y_1,\ldots,y_n$. This amounts to a reduction in computational complexity from $\mathscr{O}(n^2)$ to $\mathscr{O}(n\log(n))$ because of sorting time.

3.3 Univariate Energy Goodness-of-Fit Simulations

In this section, we consider the empirical power of the univariate energy test for the asymmetric Laplace distribution compared to other popular tests: Anderson-Darling (AD), Cramer von-Mises (CvM), and Kolmogorov Smirnov (KS). Multiple testing scenarios are

considered in this simulation study, but all scenarios have a similar procedure, which is presently described:

1. For a given sample size n and $\alpha = \mathbb{P}(\text{Type I error})$, generate data under an alternative distribution.

2. Calculate the energy test statistic $\hat{Q}_n$ for the asymmetric Laplace distribution, and any other applicable test statistics such as the AD statistic A_n (2.1.5), CvM statistic C_n (2.1.4), and the KS statistic D_n.

3. Calculate the p-value of each statistic, compare it to α, and record the result of the test.

4. Repeat steps (1) — (3) a large number of times.

The AD and CvM tests were conducted using the `R` library `goftest` (Faraway, Marsaglia, Marsaglia, and Baddeley, 2017), and the KS test was implemented with the function `ks.test()` from the `R` library `stats` (R Core Team, 2018). We take $\alpha = 0.1$ for all tests.

3.3.1 Simple Hypotheses

We investigate the power of the energy test for the asymmetric Laplace distribution under the simple hypothesis. Under the simple hypothesis, we consider the parameters of the distribution to be known. Therefore, the test may be written

$$H_0 : F_n \sim \mathscr{AL}(\theta_0, \sigma_0, \kappa_0) \tag{3.3.1}$$

In these tests, we take $\theta_0 = 0$ and $\sigma_0 = 1$, and the number of simulation replicates is 5000. p-values for the energy test are obtained via "on-the-fly" Monte Carlo with 200 replicates. The standard error of each empirical power estimate is $\sqrt{\frac{p(1-p)}{5000}} < \frac{0.5}{\sqrt{5000}} = 0.007$.

Figure 3.1 shows the power of the energy test against other tests when the skewness parameter κ is misspecified in samples of size 30. Under the null hypothesis, $y_1, \ldots, y_{30} \sim \mathscr{AL}^*(0,1,1)$. The alternative is that $y_1, \ldots, y_{30} \sim \mathscr{AL}^*(0,1,\kappa)$. Under the null hypothesis, the model is

symmetric Laplace. We observe that the energy test and the Anderson-Darling tests are essentially equally powerful. All tests empirically control the Type I error rate close to $\alpha = 0.1$.

Figure 3.2 shows a similar scenario in which we consider $\kappa = 2$ under the null hypothesis. Under this hypothesis, we find that the energy test performs better than alternative tests when the alternative distribution's κ exceeds 2. Otherwise the energy test is very competitive with the existing goodness-of-fit tests.

Figure 3.3 shows the empirical Type I errors rates of the four tests considered in this simulation. We find that in larger samples of size 250 or 500 the energy, KS, and AD tests control the type I error rate close to the nominal level. We observe that the CvM test empirically keeps the type I error rates below the significance level of the test in samples of size 100 and 250.

3.3.2 Alternative Distributions

Various alternatives to the (asymmetric) Laplace distribution are considered. The asymmetric Laplace distribution is unique in the family of univariate distributions in that its mode is peaked and its tails decay exponentially. In these simple hypothesis simulations, we consider distributions with location parameter 0 and scale parameter 1.

Normal

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Student's t

$$f(x|v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

Laplace-Normal Mixture $\mathscr{LN}(p) \overset{d}{=} p\mathscr{N}(0,1) + (1-p)\mathscr{L}(0,1)$

Random variates from the Normal distribution are generated with `rnorm`, and from the Student's t distribution with `rt`. Both functions are available in base `R`. Random variates from the Laplace-Normal mixture are generated from a custom function.

Figures 3.4 — 3.9 show the power of the tests against several alternative distributions detailed in Section 3.3.2. We observe that the energy test is generally the most powerful test one could employ in this limited set of testing situations. Testing against the Normal distribution (Figure

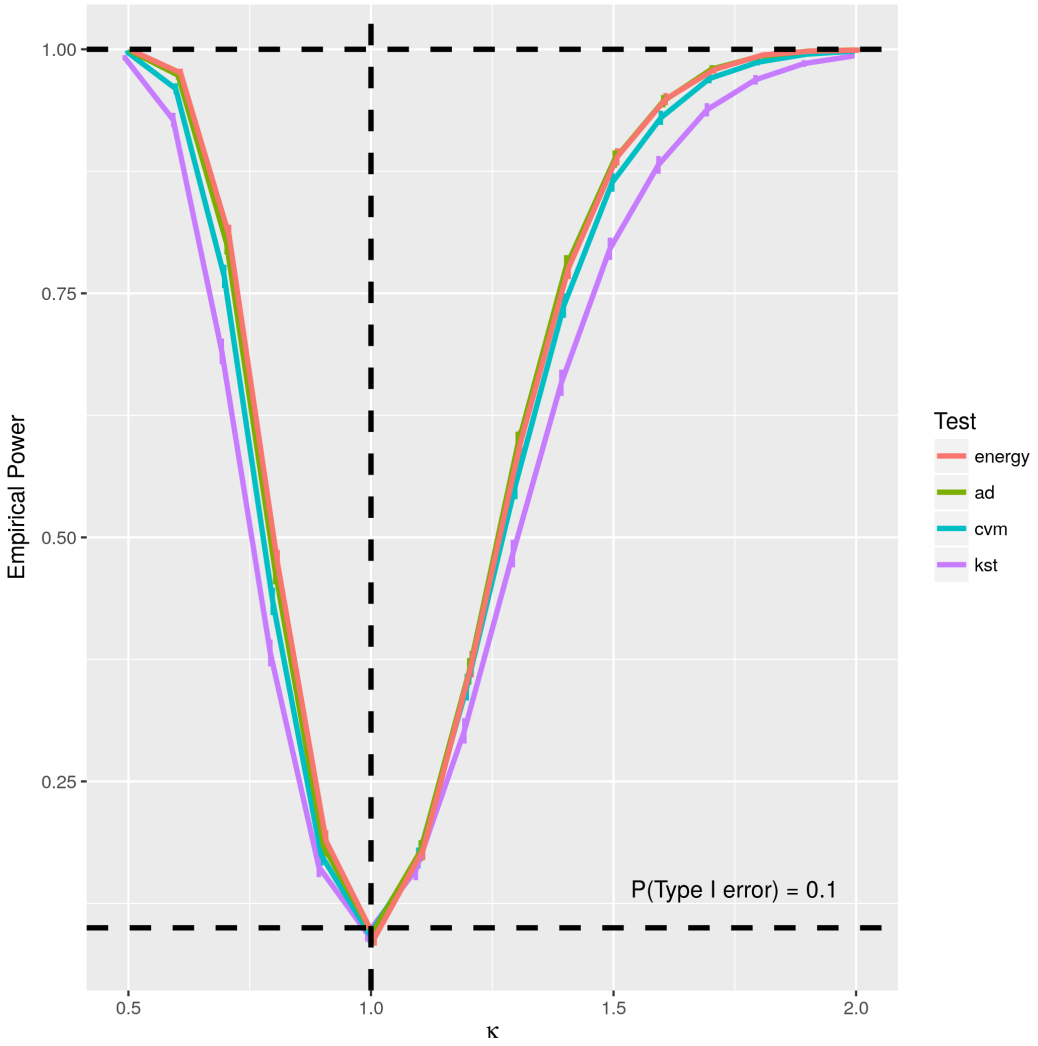

Figure 3.1 Power of testing $\mathscr{AL}(0,1,1)$ against $\mathscr{AL}(0,1,\kappa)$ with varying κ

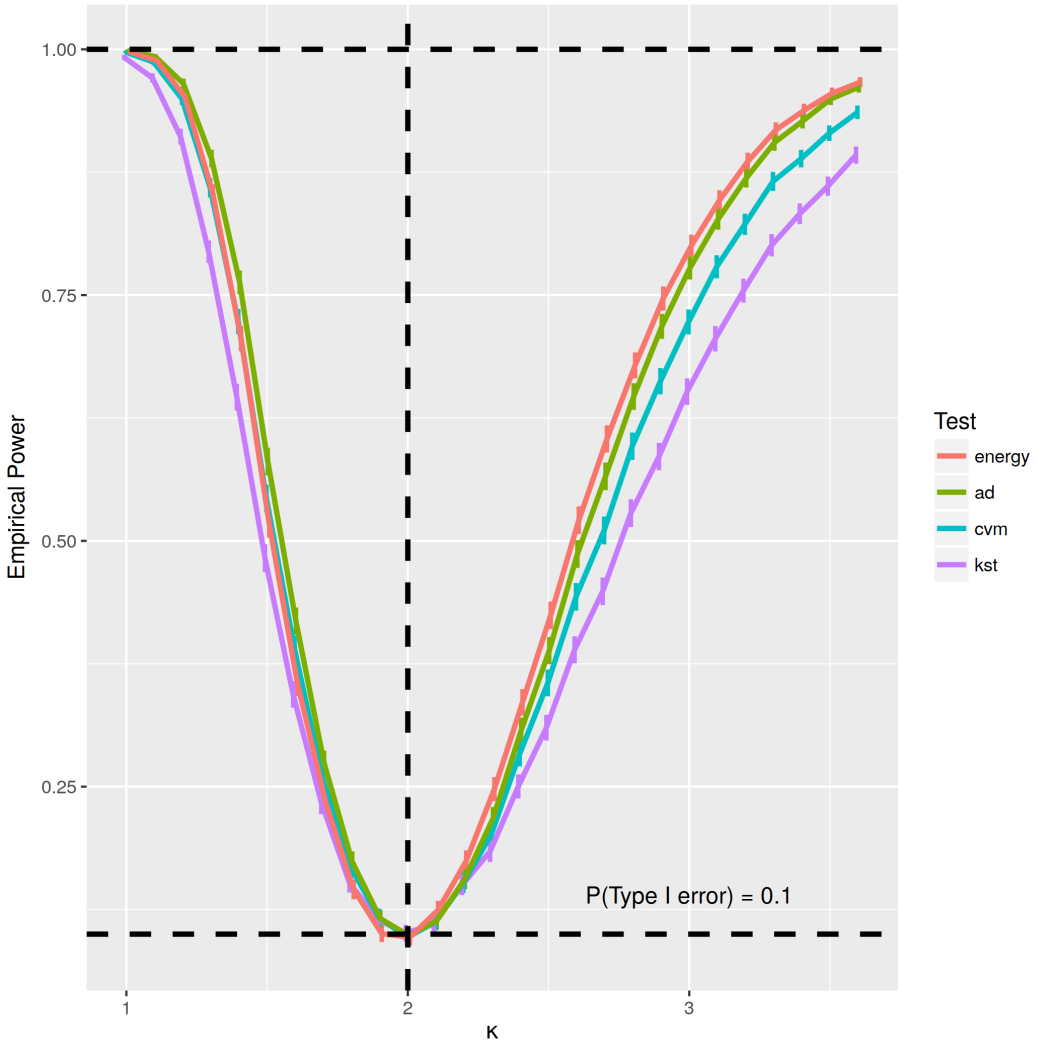

Figure 3.2 Power of testing $\mathscr{AL}(0,1,2)$ against $\mathscr{AL}(0,1,\kappa)$ with varying κ

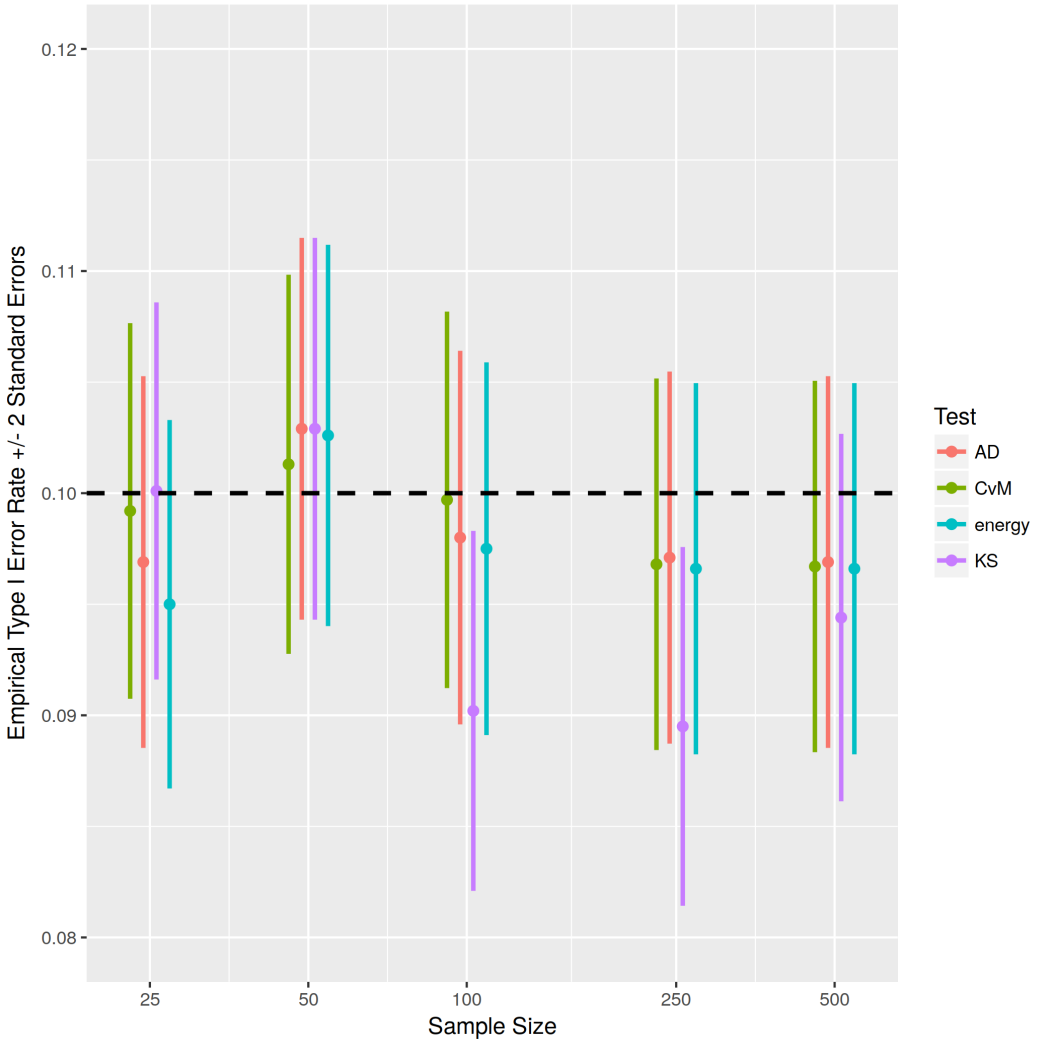

Figure 3.3 Type I error rates of four goodness-of-fit tests for the $\mathscr{AL}(0,1,1)$ hypothesis

3.4) or a mixture distribution (Figures 3.7 – 3.9) shows that the energy test and AD test are equally matched. Testing against Student's t distribution (Figures 3.5 and 3.6) show that the energy test is considerably more powerful than the AD test in samples of size less than 100.

3.3.3 Composite Hypotheses

In this section, we demonstrate the power of the energy test for the $\mathscr{AL}^*$ distribution using parameters estimated from the E-M algorithm. The details of this algorithm are considered separately in Section 4.4.1. We consider the distributions detailed in Section 3.3.2 and add to that list the skew-Normal distribution. Random variates from the skew-Normal distribution are

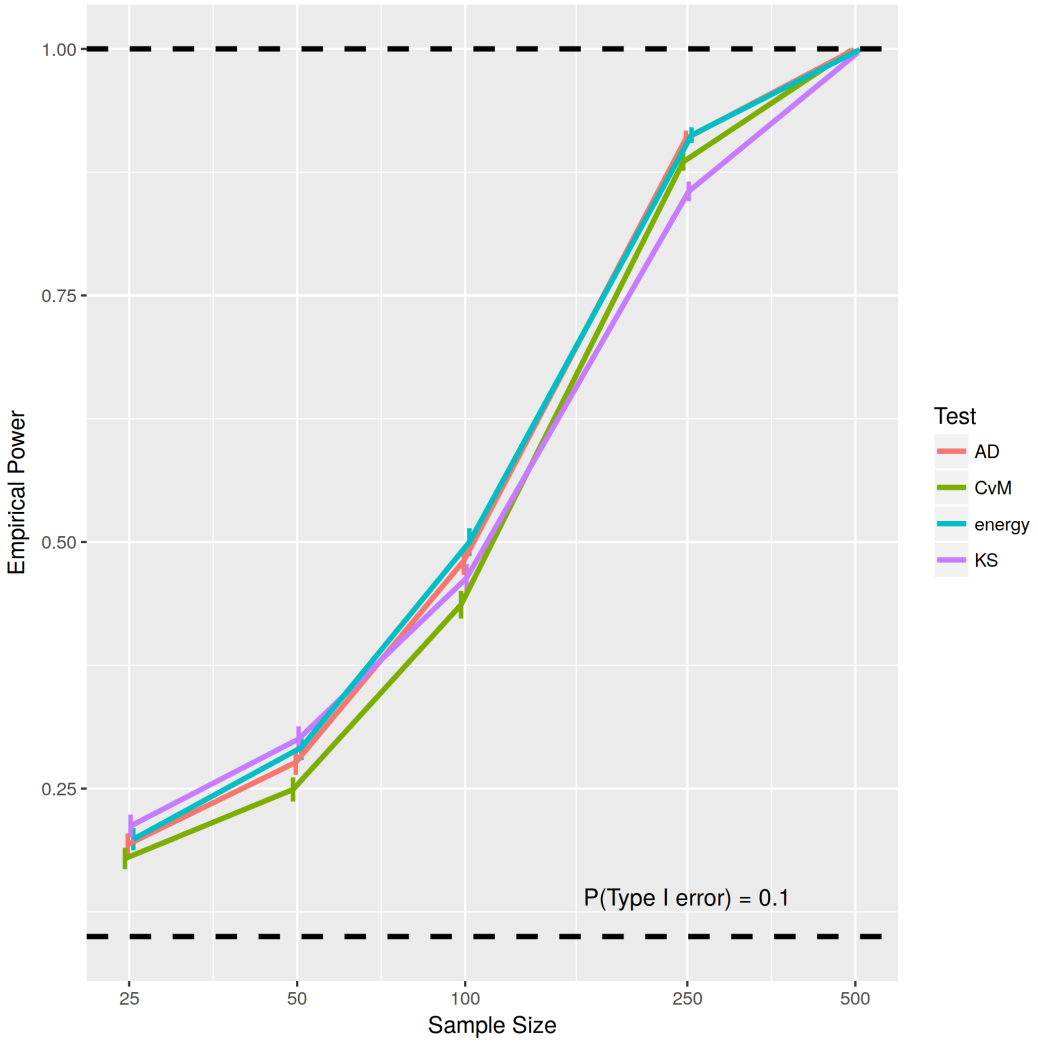

Figure 3.4 Power of testing $\mathscr{AL}(0,1,1)$ against the Normal distribution with varying sample size

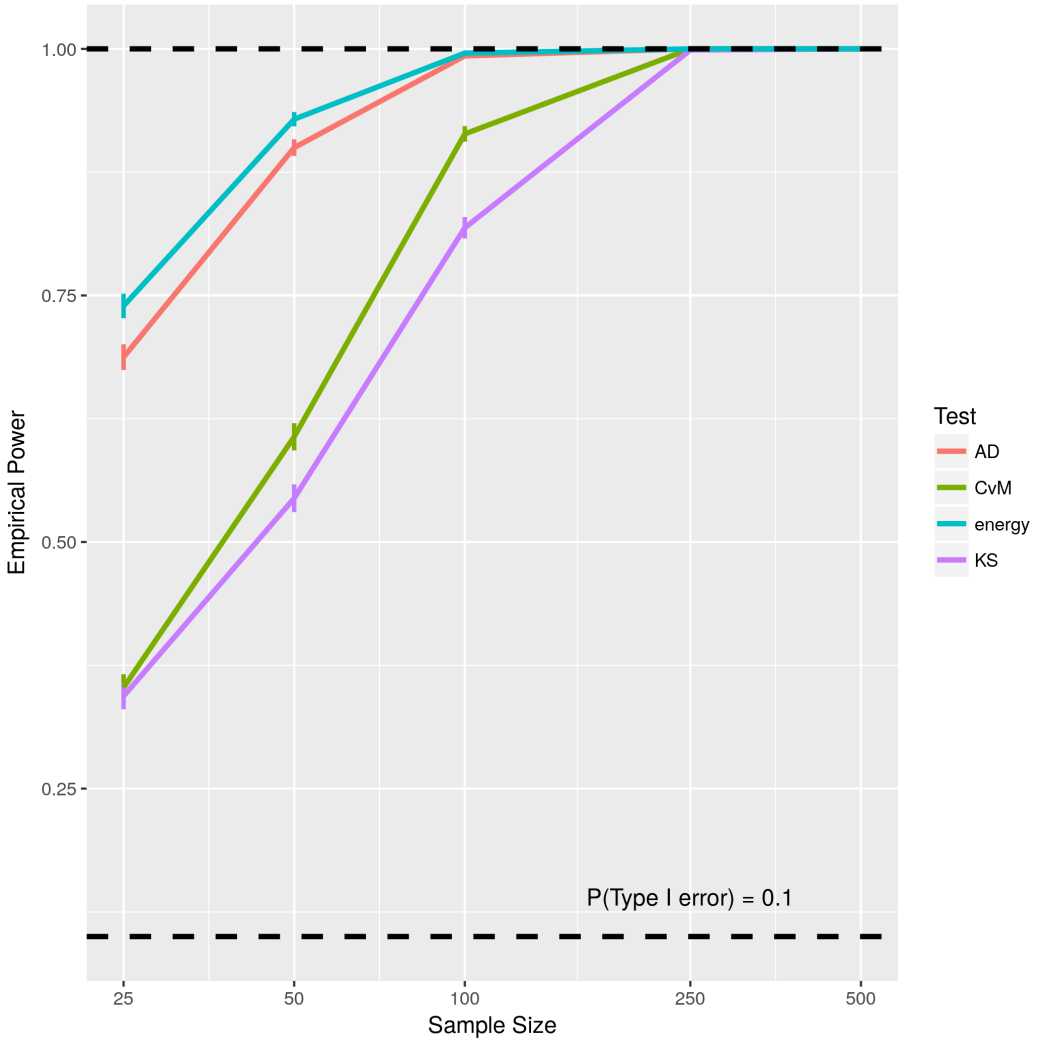

Figure 3.5 Power of testing $\mathscr{ALL}(0,1,1)$ against the Student's t distribution (df = 1) with varying sample size

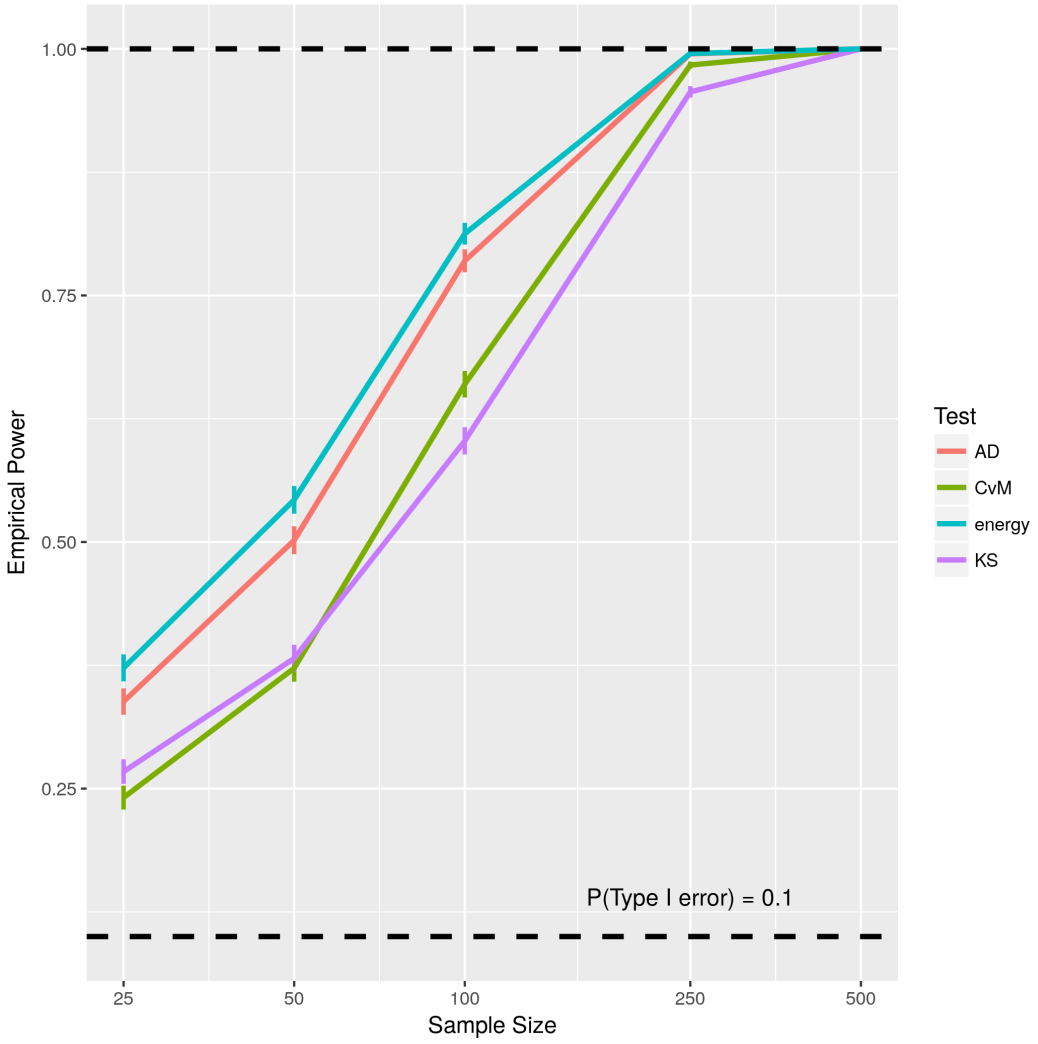

Figure 3.6 Power of testing $\mathscr{AL}(0,1,1)$ against the Student's t distribution (df = 5) with varying sample size

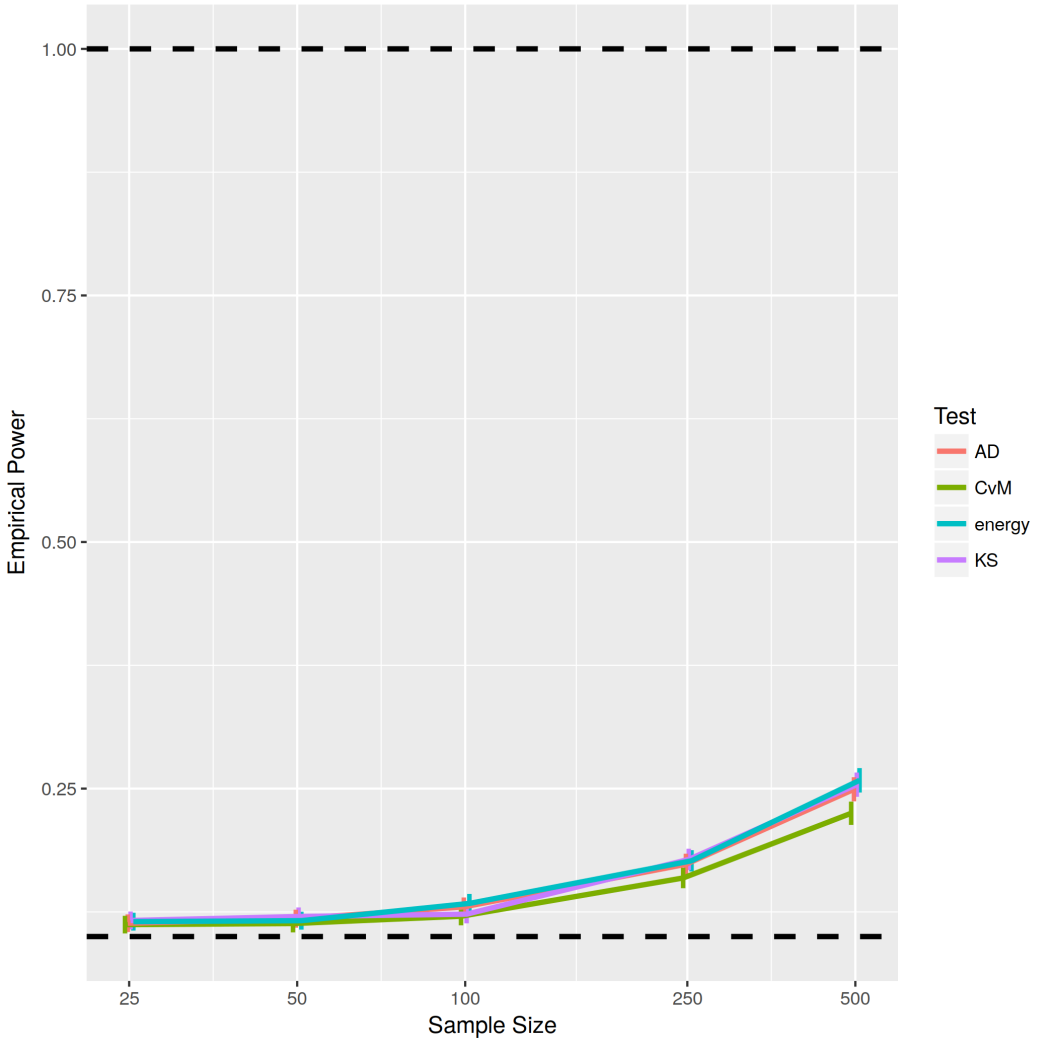

Figure 3.7 Power of testing $\mathscr{ALL}(0,1,1)$ against $\mathscr{LN}(0.3)$ with varying sample size

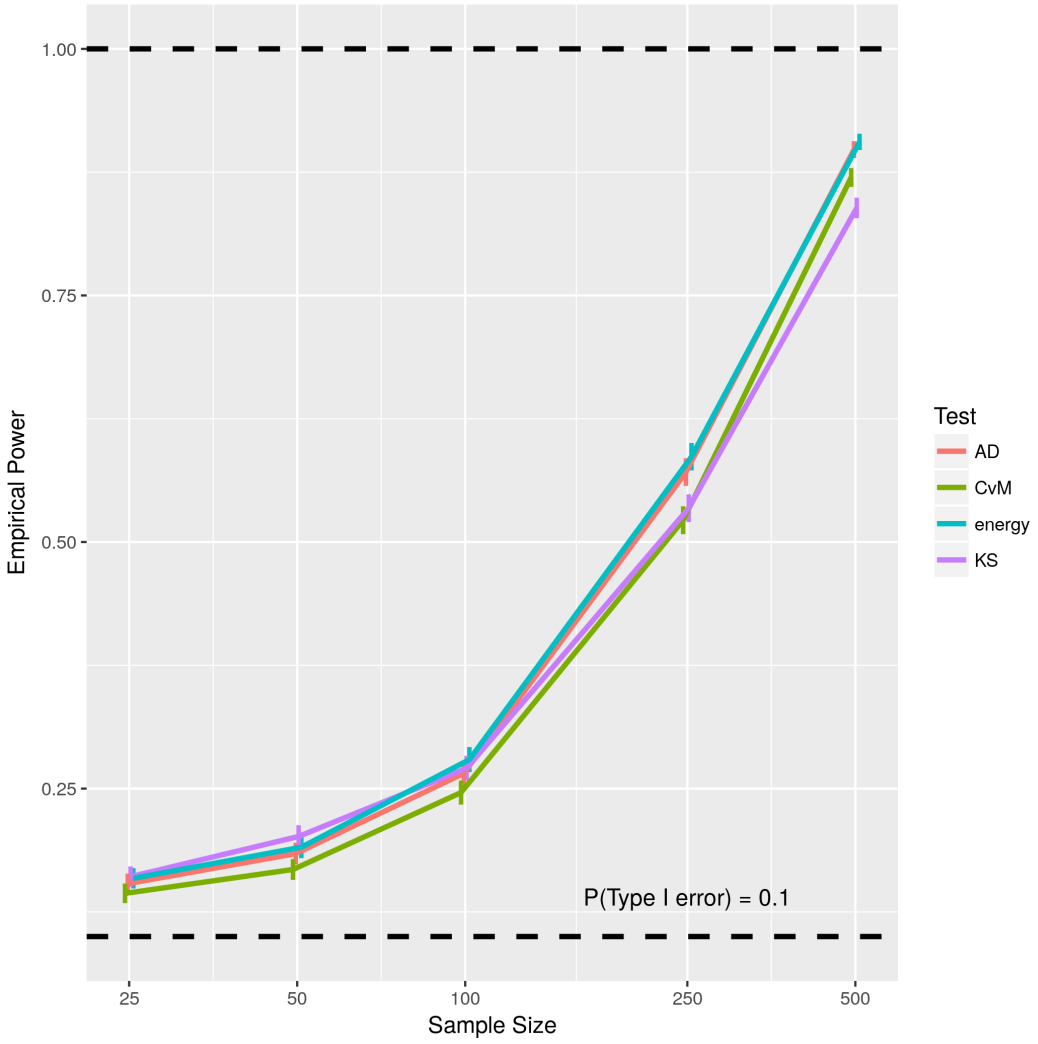

Figure 3.8 Power of testing $\mathscr{ALL}(0,1,1)$ against $\mathscr{LN}(0.7)$ with varying sample size

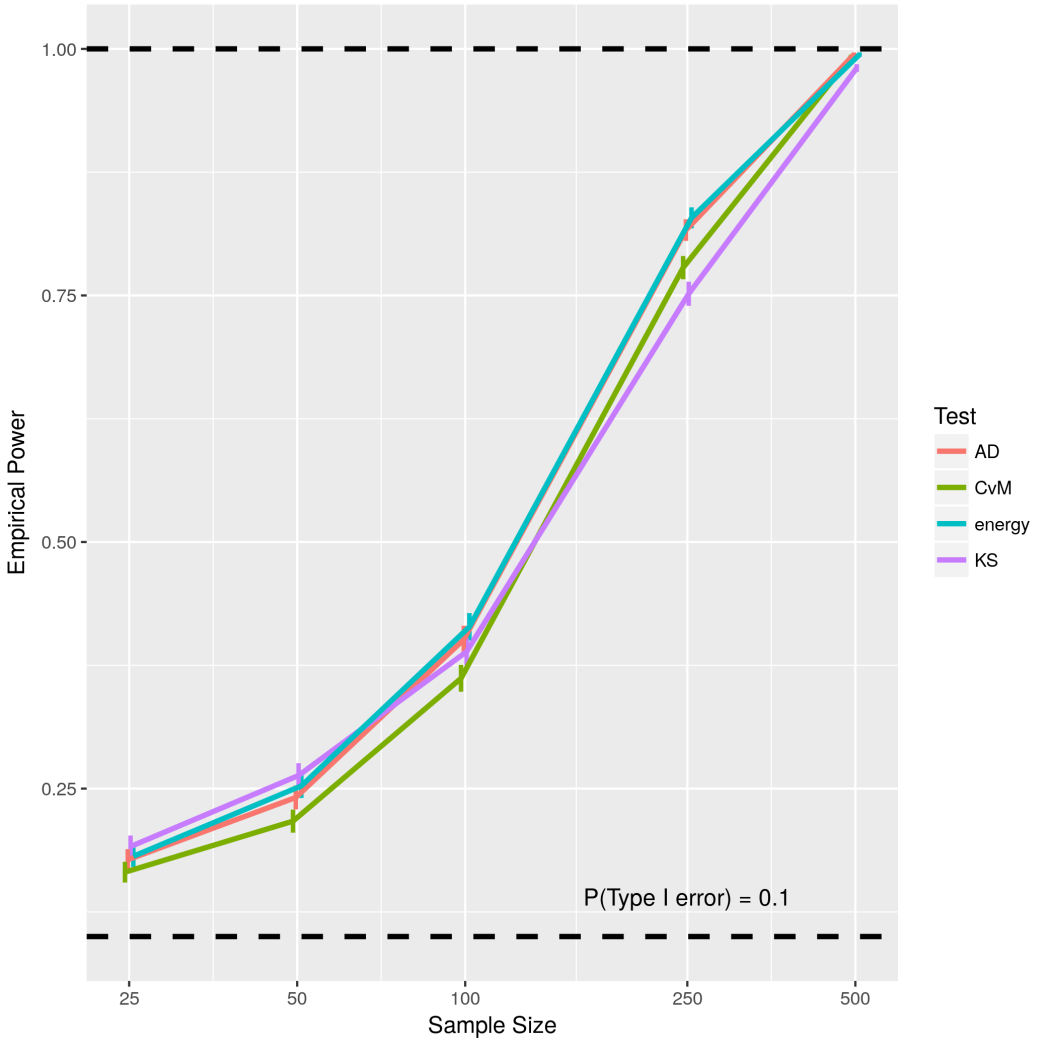

Figure 3.9 Power of testing $\mathscr{ALL}(0,1,1)$ against $\mathscr{LN}(0.9)$ with varying sample size

generated using the function `rsn` from the package `sn` (Azzalini, 2017). The density of the normalized skew-Normal distribution considered in this simulation study is $2\phi(x)\Phi(\alpha x)$ where ϕ denotes the density of the standard Normal distribution and Φ denotes the distribution function of the standard Normal distribution. The parameter α regulates the asymmetry of the distribution and $\alpha = 0$ recovers the standard Normal distribution.

Figure 3.10 shows the empirical Type I error rates of the energy, AD, and CvM tests for sample sizes $n = 50, 100, 250$. The KS test is omitted from the composite tests. Because these tests are conducted under estimated parameters, the test decision may not be calculated as if the parameters are known as in section 3.3.1. A solution to this difficulty is to use a parametric bootstrap to simulate the distribution of each test statistic, then choose the $1-\alpha$ quantile of the simulated distribution to make a test decision. For the asymmetric Laplace distribution, the situation is further complicated by the presence of the skewness parameter κ. Thus the distribution of test statistics depends on the sample size of the data and the skewness of the data. The parametric bootstrap procedure is detailed:

1. Calculate $\hat{\kappa}_{EM}$, the estimate of κ, from the centered and scaled data $y_1, \ldots, y_n$.
2. Calculate each test statistic $(\hat{Q}_n, A_n, C_n)$ of the transformed data using the parameters $\theta_0 = 0$, $\sigma_0 = 1$, and $\kappa_0 = \hat{\kappa}_{EM}$.
3. Generate B Monte Carlo samples from the $\mathscr{AL}(0, 1, \hat{\kappa}_{EM})$ distribution.
4. Calculate test statistics $\hat{Q}_n^{(b)}$, $A_n^{(b)}$ and $C_n^{(b)}$, $b = 1, \ldots, B$ conditioned on $\hat{\kappa}_{EM}^{(b)}$.
5. Calculate $c_{Q,1-\alpha}$, the $(1-\alpha)$ quantile of $\hat{Q}_n^{(1)}, \ldots, \hat{Q}_n^{(B)}$ (and similarly for other test statistics).
6. Reject $H_0 : y_1, \ldots, y_n \sim \mathscr{AL}$ if $\hat{Q}_n > c_{Q,1-\alpha}$ (and similarly for other test statistics).

In our simulations we use $B = 200$ bootstrap replicates and 2000 Monte Carlo simulations are performed in the composite hypothesis simulation study.

Figure 3.10 shows the empirical Type I error rates of the energy, AD, and CvM tests under estimated parameters. The significance level of the simulation is set at 10% for all tests. We observe that these tests generally regulate the Type I error rate below the 10% threshold in samples of size 50. The energy and AD tests control Type I error close to 10% in larger sample sizes, but the CvM test appears to empirically control the Type I error below the nominal level in each sample size studied.

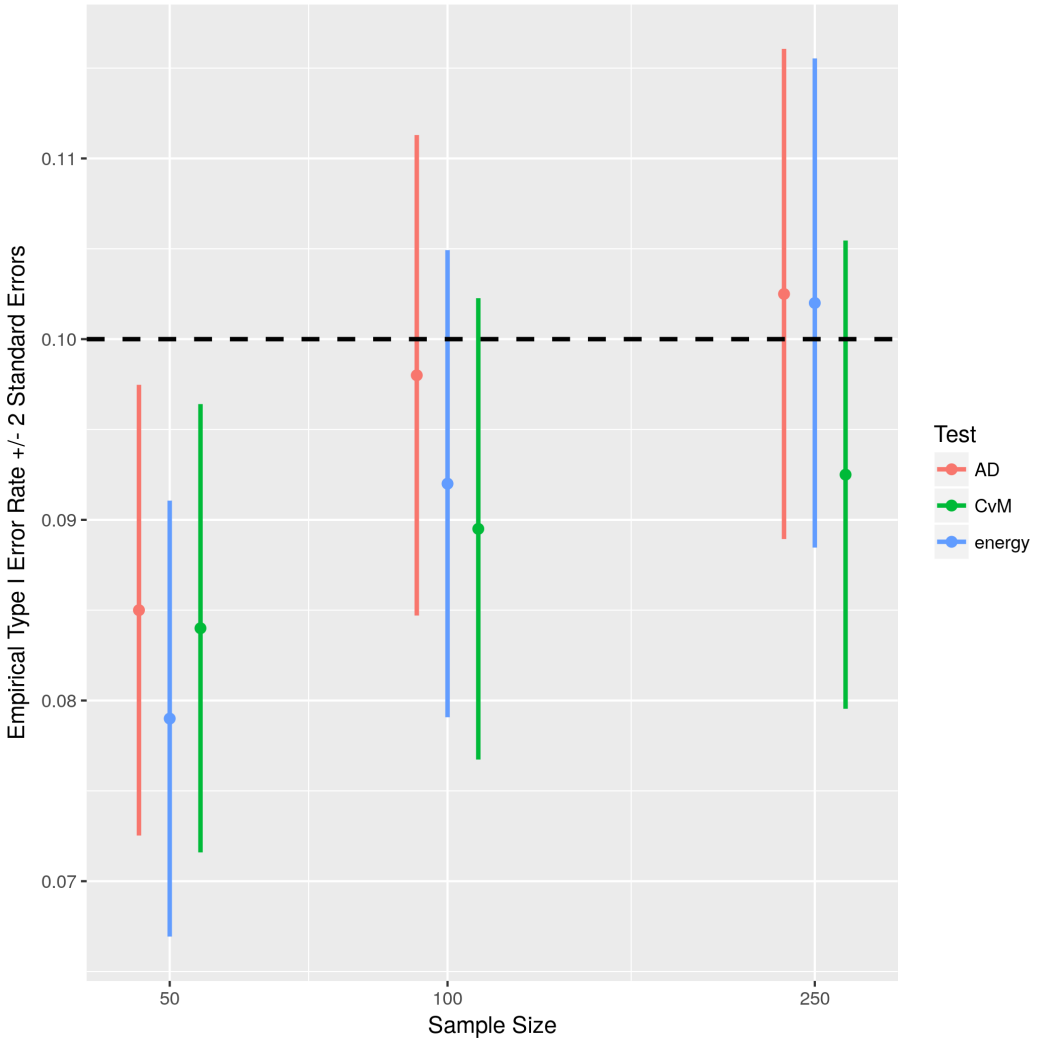

Figure 3.10 Type I error rates for three composite tests of the asymmetric Laplace distribution

Figure 3.11 shows the power of goodness-of-fit tests against the Normal distribution with unknown parameters. We observe that the AD and energy tests have virtually indistinguishable power at each of the sample sizes. The CvM test is slightly less powerful. As with each

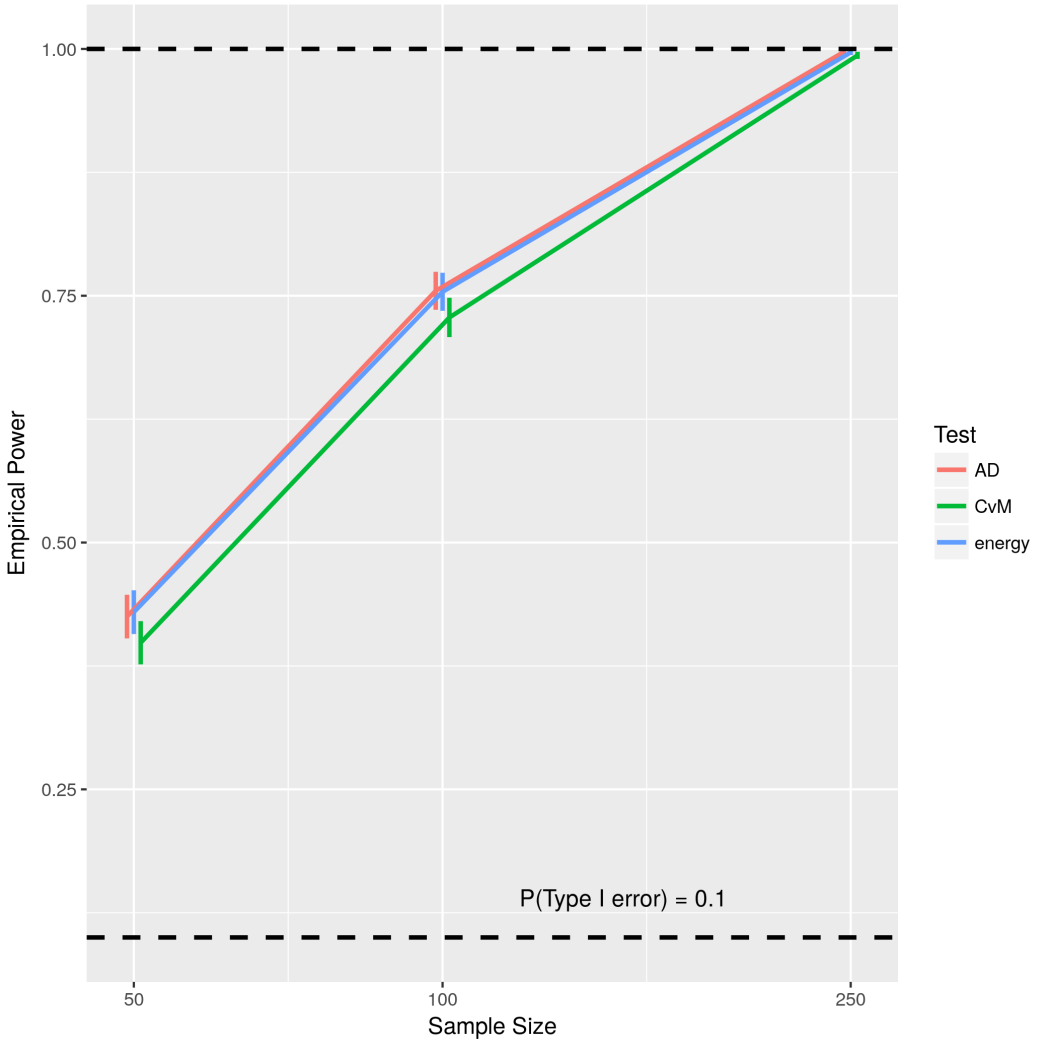

Figure 3.11 Power of testing the composite $\mathscr{AL}$ hypothesis against Normally distributed data (sample size varies)

alternative, the power of a consistent test is expected to approach 1, and in the case of the Normal distribution, we find that the power of the energy is nearly 1 in samples as small as 250.

Figures 3.12 and 3.13 show the power of each test statistic when the alternative distribution is from the Student's t family. We observe higher powers among all statistics when the alternative distribution is t_2. This is due to the heavier tails exhibited in the t_2 distribution. Testing against the t_2 distribution, we find that the energy test outperforms the AD and CvM tests. The difference between the tests' powers disappears as the degrees of freedom is increased to 5. Figure 3.13 shows that all three tests are equally powerful when testing against the t_5 distribution.

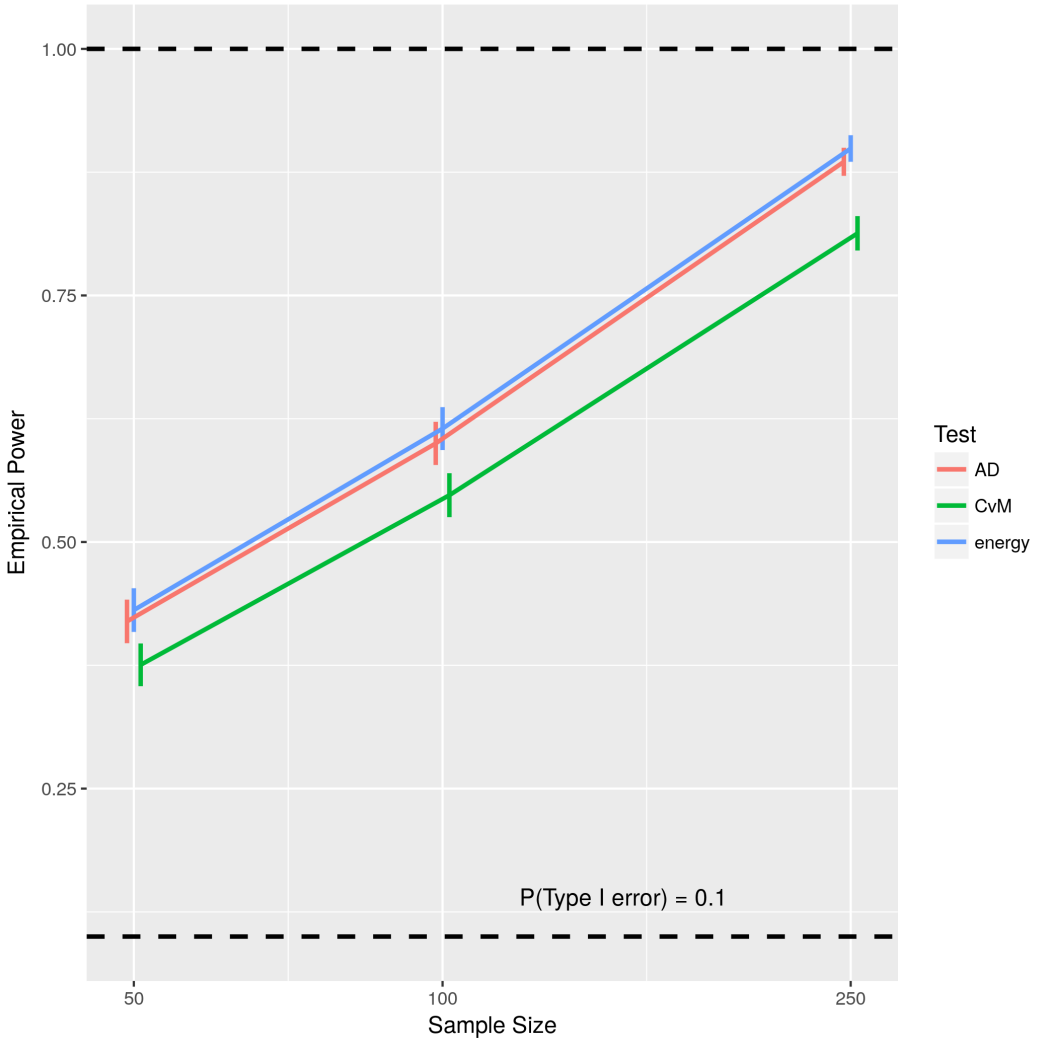

Figure 3.12 Power of testing the composite $\mathscr{A}\mathscr{L}$ hypothesis against t_2 distributed data (sample size varies)

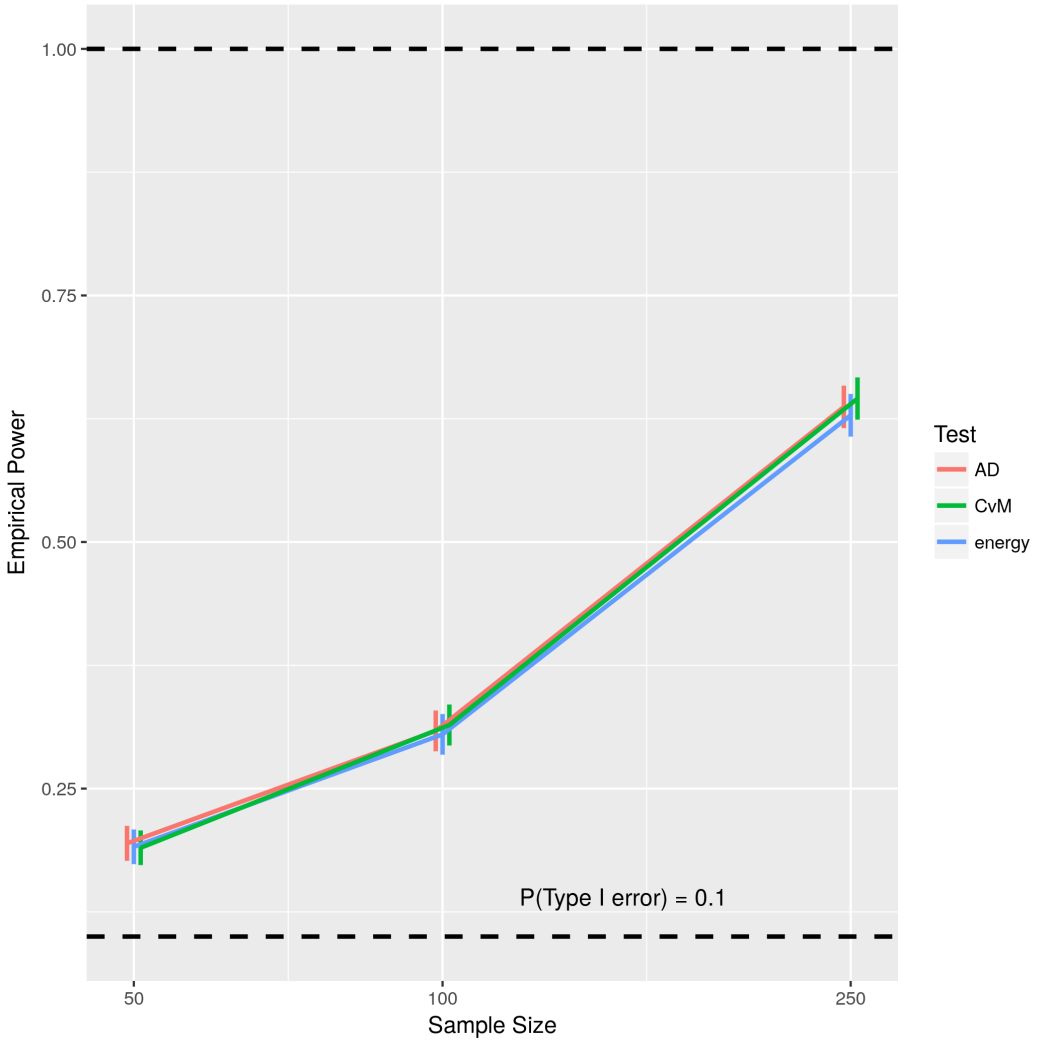

Figure 3.13 Power of testing the composite $\mathscr{AL}$ hypothesis against t_5 distributed data (sample size varies)

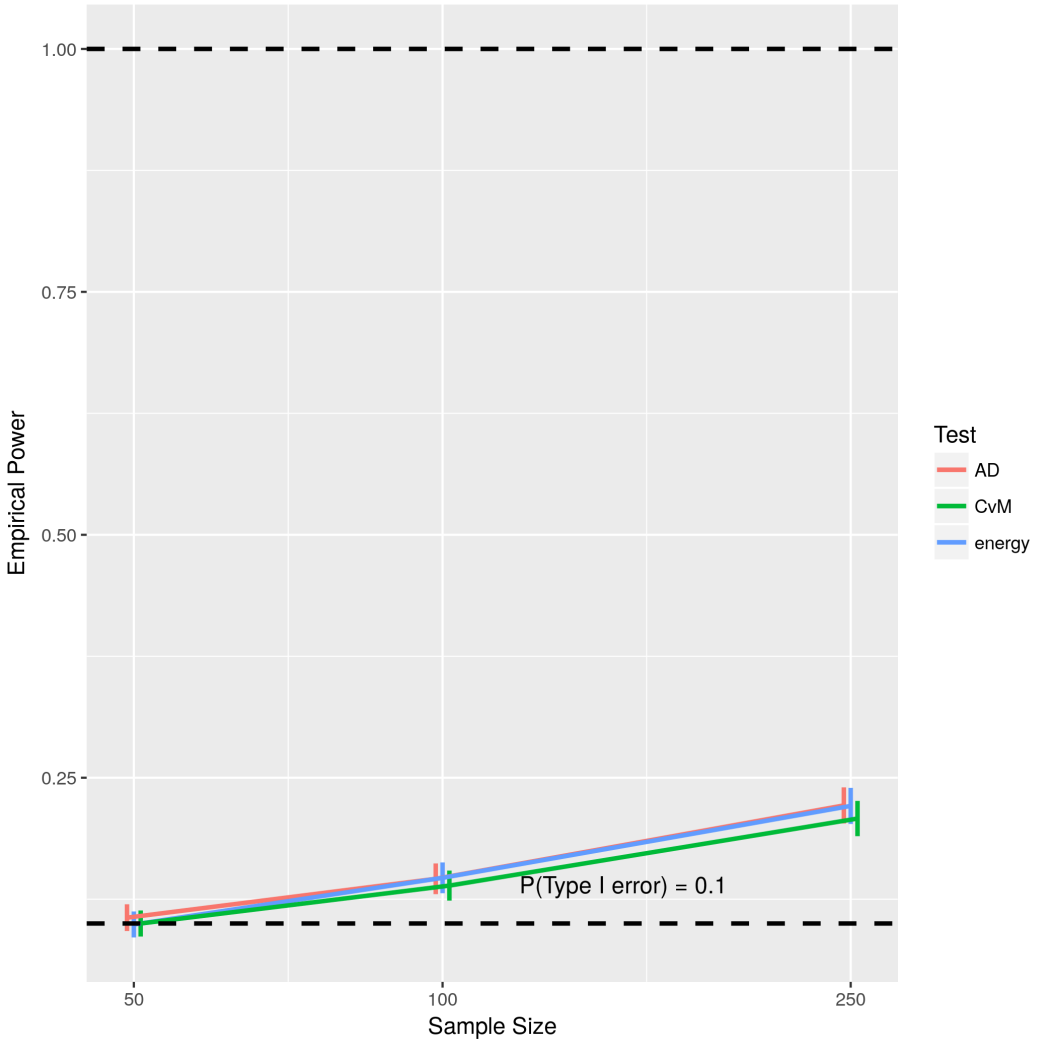

Figure 3.14 Power of testing the composite $\mathscr{AL}$ hypothesis against $\mathscr{LN}(0.25)$ distributed data (sample size varies)

Figures 3.14 — 3.16 show the power of the composite tests against a Laplace-Normal mixture alternative where the mixing parameter p is taken to be 0.25, 0.50, and 0.75. An increase in the parameter p results in a larger proportion of the data being generated from the Normal distribution. Accordingly, we observe greater powers among all goodness-of-fit tests under $p = 0.75$ (Figure 3.16) than $p = 0.25$ (Figure 3.14).

Figures 3.17 — 3.20 show the power of each composite asymmetric Laplace test against a range of different skew-Normal distributions with varying skewness parameter α. Each of these power curves exhibit roughly the same pattern: The asymmetric Laplace hypothesis is not likely

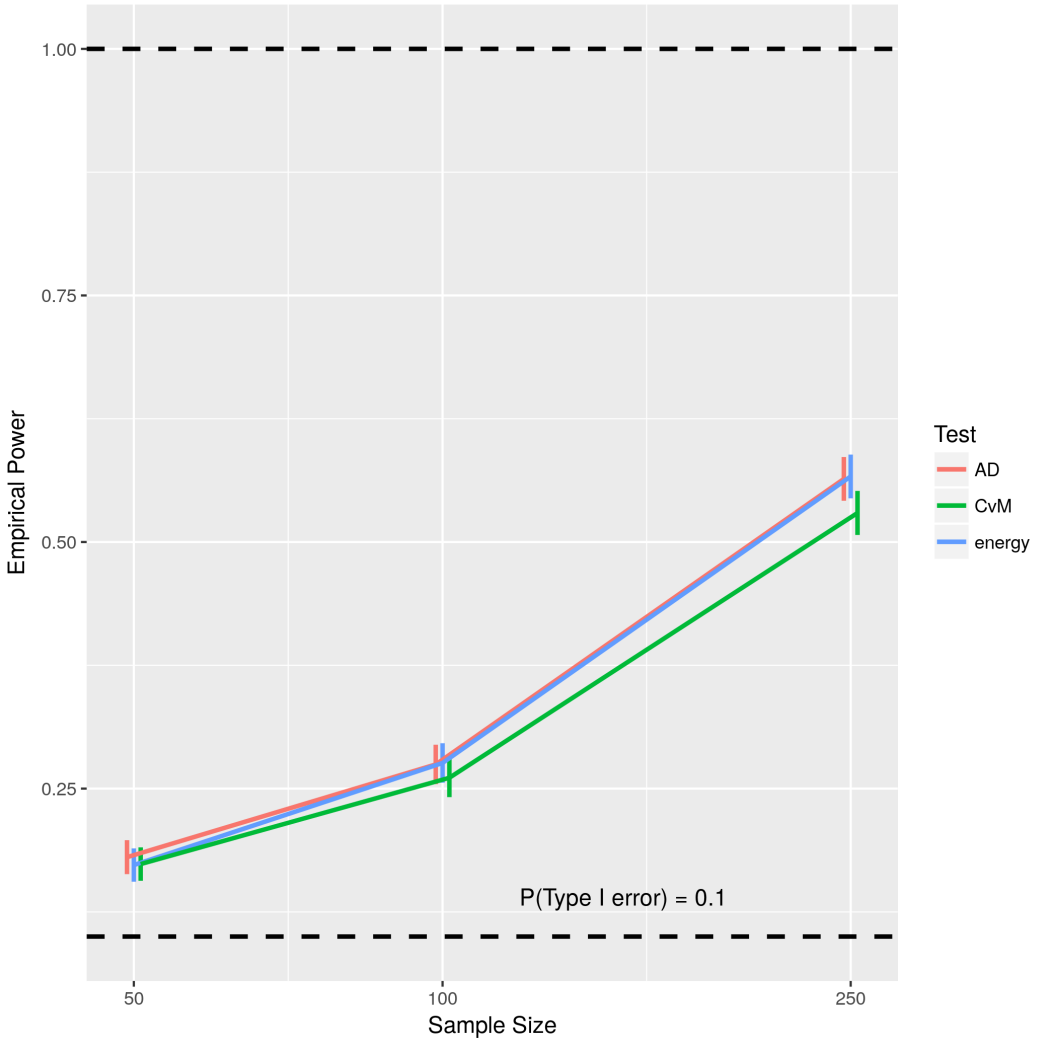

Figure 3.15 Power of testing the composite $\mathscr{AL}$ hypothesis against $\mathscr{LN}(0.50)$ distributed data (sample size varies)

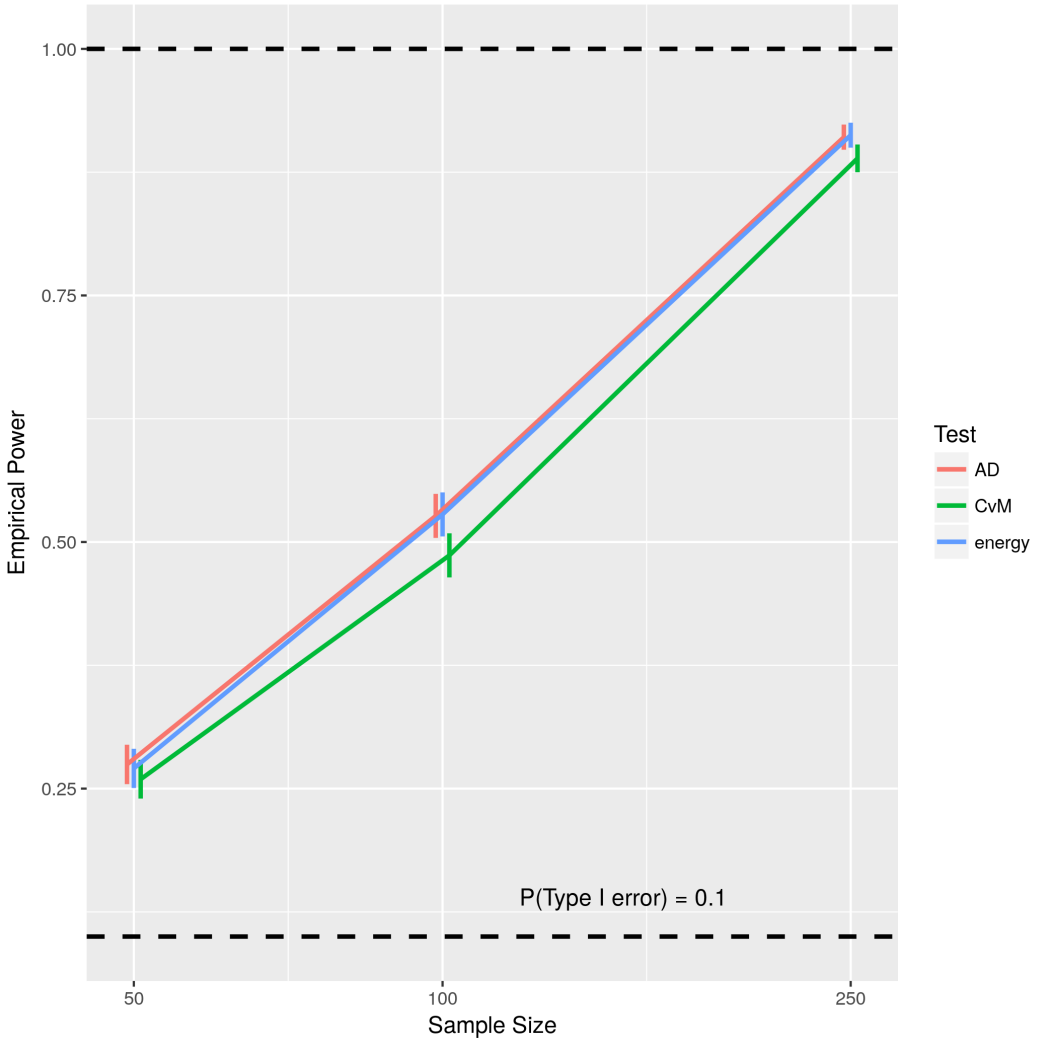

Figure 3.16 Power of testing the composite $\mathscr{AL}$ hypothesis against $\mathscr{LN}(0.75)$ distributed data (sample size varies)

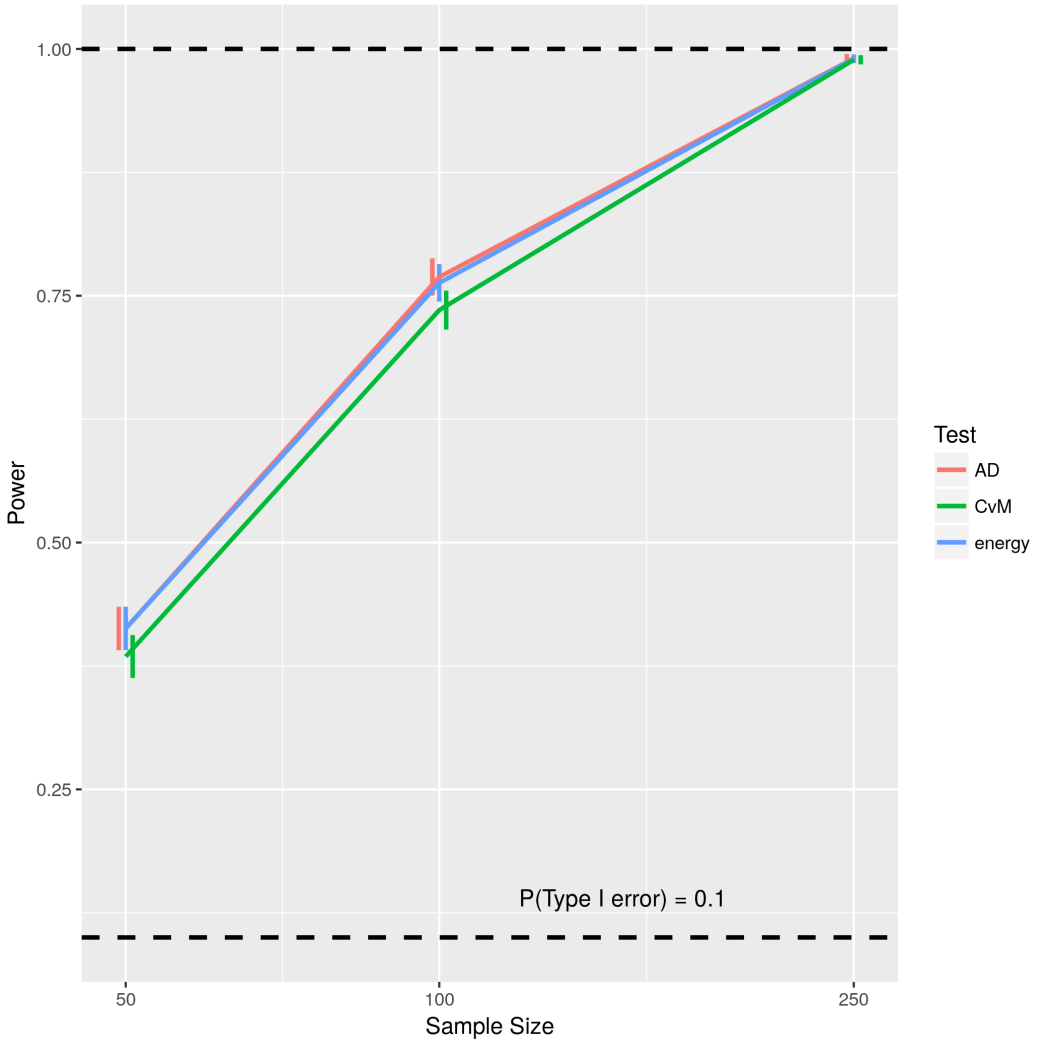

Figure 3.17 Power of testing the composite $\mathscr{AL}$ hypothesis against $\mathscr{SN}(\alpha = 1)$ distributed data (sample size varies)

to be rejected in small samples of size 50, but likely to be rejected in samples of size 100. Additionally, as we increase the skewness parameter, we find that the AD test performs slightly better than the energy test. Under each level of skewness, we find that power is nearly 1 in samples of size 250.

Our simulation study shows that the energy test for the asymmetric Laplace hypothesis is a powerful competitor to the AD test under a variety of alternative distributions. It's important to note that this simulation study is not comprehensive, as not every relevant sample size, alternative distribution, and completing goodness-of-fit test can be examined. Other the other hand, the KS

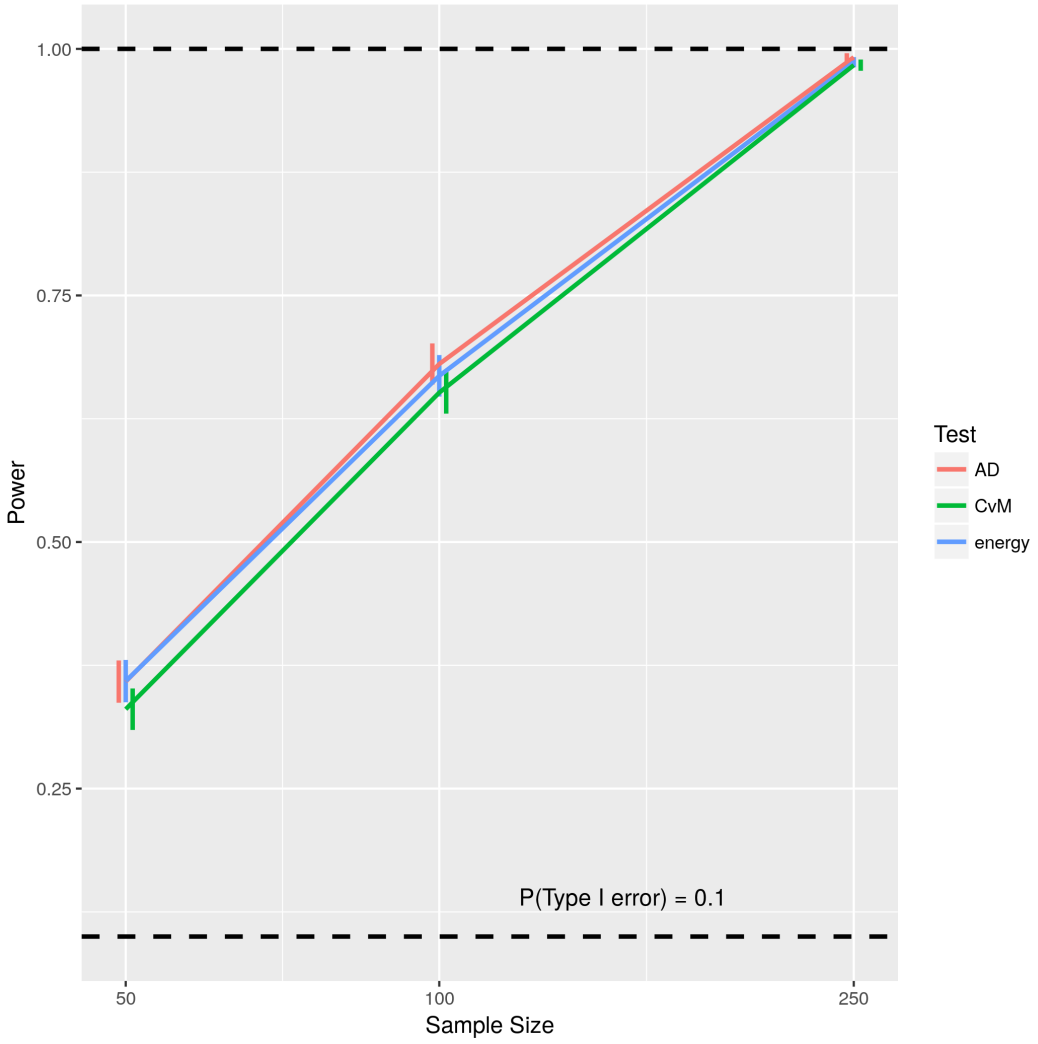

Figure 3.18 Power of testing the composite $\mathscr{AL}$ hypothesis against $\mathscr{SN}(\alpha = 2)$ distributed data (sample size varies)

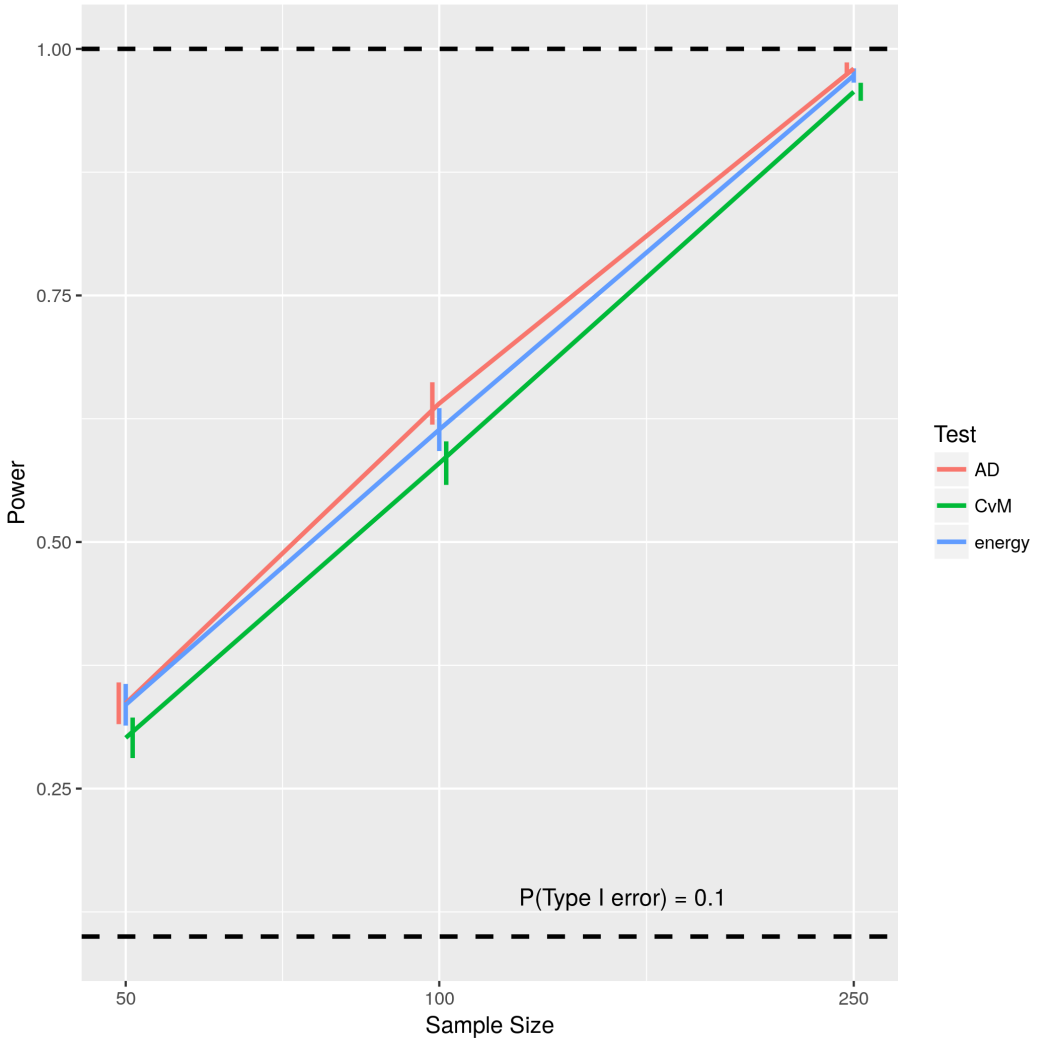

Figure 3.19 Power of testing the composite $\mathscr{AL}$ hypothesis against $\mathscr{SN}(\alpha = 3)$ distributed data (sample size varies)

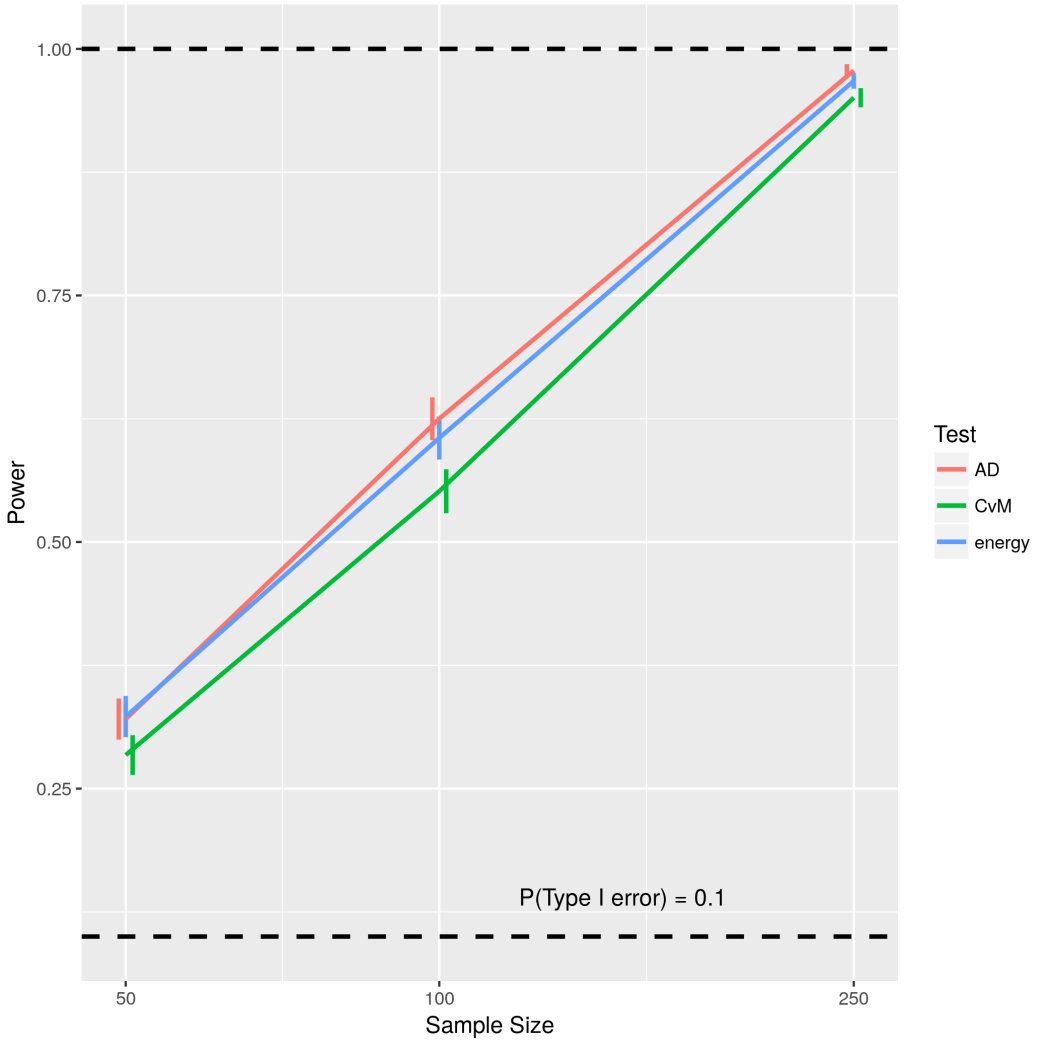

Figure 3.20 Power of testing the composite $\mathscr{AL}$ hypothesis against $\mathscr{SN}(\alpha = 4)$ distributed data (sample size varies)

test is a widely applied test for assessing a univariate goodness-of-fit hypothesis for simple hypotheses, and it is clear that under a range of alternative hypotheses, the energy test bests the KS test in terms of power.

3.4 Distance Variance and Distance Standard Deviation for Asymmetric Laplace Distribution

Distance correlation (Székely, Rizzo, and Bakirov, 2007) is a measure of dependence between random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$. Unlike classical correlation which measures the linear dependence, distance correlation is sensitive to linear and non-linear types of dependence. Additionally, sample distance correlation has a fast and simple computing formula for bivariate data (Huo and Székely, 2016). These properties have made this measure of dependence popular in both theoretical and applied settings.

Random variables X and Y are independent if the joint CF of (X,Y) factors. In the hypothesis testing setting, this is equivalent to testing

$$\begin{aligned} H_0 &: \phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t); \\ H_1 &: \phi_{X,Y}(s,t) \neq \phi_X(s)\phi_Y(t), \end{aligned} \tag{3.4.1}$$

where ϕ represents the Fourier transform (characteristic function). Distance covariance is defined to be

$$\begin{aligned} \mathcal{V}^2(X,Y) &= ||\phi_{X,Y}(s,t) - \phi_X(s)\phi_Y(t)||^2 \\ &= \int_{\mathbb{R}^{p+q}} \frac{1}{c_p c_q} \frac{|\phi_{X,Y}(s,t) - \phi_X(s)\phi_Y(t)|^2}{||s||_p^{p+1}||t||_q^{q+1}}\, ds\, dt, \end{aligned} \tag{3.4.2}$$

where $c_p = \frac{\pi^{(p+1)/2}}{\Gamma((p+1)/2)}$. Distance standard deviation (Székely et al., 2007) is defined to be the square root of

$$\mathcal{V}^2(X) \equiv \mathcal{V}^2(X,X) = ||\phi_{X,X}(s,t) - \phi_X(s)\phi_X(t)||^2, \tag{3.4.3}$$

which has a remarkable expression (Székely et al., 2007) in terms of distances between i.i.d. copies of X if we assume $\mathbb{E}[||X||^2] < \infty$,

$$\mathscr{V}^2(X) = \mathbb{E}[||X - X'||^2] + [\mathbb{E}||X - X'||]^2 - 2\mathbb{E}[||X - X'|| \cdot ||X - X''||]. \tag{3.4.4}$$

Recently Edelmann, Richards, and Vogel (2017) showed that a computing formula for distance variance can be given in terms of the moments of spacings of order statistics if X is real valued.

Theorem 3.4.5. *(Edelmann et al., 2017) Let X be a real valued random variable with $\mathbb{E}|X| < \infty$, and let X' and X'' be i.i.d. copies of X. If $X_{1:3} \leq X_{2:3} \leq X_{3:3}$ are the order statistics of (X, X', X''), then*

$$\mathscr{V}^2(X) = (\mathbb{E}|X - X'|)^2 - \frac{4}{3}\mathbb{E}[(X_{2:3} - X_{1:3})(X_{3:3} - X_{2:3})]. \tag{3.4.6}$$

The following theorem is particularly useful for computing distance variances for statistical distributions.

Theorem 3.4.7. *(Edelmann et al., 2017)*

$$\mathscr{V}^2(X) = 8 \int_{x=-\infty}^{x=\infty} \int_{y=x}^{y=\infty} F^2(x)[1 - F(y)]^2 \, dy \, dx. \tag{3.4.8}$$

A proof is given in Edelmann et al. (2017), along with explicit computations of the distance variance for a select set of univariate distributions. Several of these distance variances are computed using a combination of equation (3.4.6) with the results from Gerstenberger and Vogel (2015). In this section we will show that one can compute the distance variance directly using (3.4.8) for the $\mathscr{AL}^*$ family of distributions.

Proposition 3.4.9. *Suppose $Y \sim \mathscr{AL}^*(\theta, \kappa, \sigma)$. Then*

$$\mathscr{V}^2(X) = 8(I_1 + I_2 + I_3), \tag{3.4.10}$$

where

$$I_1 = \frac{\sigma^2}{2\kappa^2(1+\kappa^2)^2}\left(\frac{1}{4} - \frac{1}{3(1+\kappa^2)} + \frac{1}{8(1+\kappa^2)^2}\right), \tag{3.4.11}$$

$$I_2 = \frac{\kappa^4\sigma^2}{8(1+\kappa^2)^4}, \tag{3.4.12}$$

$$I_3 = \frac{\sigma^2\kappa^6}{2(1+\kappa^2)^2}\left(\frac{1}{4} - \frac{\kappa^2}{3(1+\kappa^2)} + \frac{\kappa^4}{8(1+\kappa^2)^2}\right). \tag{3.4.13}$$

Proof. Without loss of generality, we may set the location parameter $\theta = 0$. Then we calculate

$$\mathscr{V}^2(X) = 8\int_{x=-\infty}^{x=\infty}\int_{y=x}^{y=\infty} F^2(x)[1-F(y)]^2\,dy\,dx$$

by evaluating the integral in three parts. Write (3.4.8) as the sum:

$$\mathscr{V}^2(X) = 8(I_1 + I_2 + I_3), \tag{3.4.14}$$

where

$$\begin{aligned} I_1 &= \int_0^\infty\int_x^\infty F^2(x)[1-F(y)]^2\,dy\,dx \\ I_2 &= \int_{-\infty}^0\int_0^\infty F^2(x)[1-F(y)]^2\,dy\,dx \\ I_3 &= \int_{-\infty}^0\int_x^0 F^2(x)[1-F(y)]^2\,dy\,dx \end{aligned} \tag{3.4.15}$$

then substitute F with the parameterization given in equation 1.3.17. We find that

$$\begin{aligned} I_1 &= \int_0^\infty\int_x^\infty [1-p_\kappa\exp(-\lambda x)]^2[p_\kappa\exp(-\lambda y)]^2\,dy\,dx \\ &= \frac{p_\kappa^2}{4\lambda^2} - \frac{2p_\kappa^3}{6\lambda^2} + \frac{p_\kappa^4}{8\lambda^2}, \end{aligned} \tag{3.4.16}$$

$$
\begin{aligned}
I_2 &= \int_{-\infty}^{0}\int_{0}^{\infty} [q_\kappa \exp(\beta x)]^2 [p_\kappa \exp(-\lambda y)]^2 \, dy \, dx \\
&= \frac{(p_\kappa q_\kappa)^2}{4\lambda\beta}
\end{aligned}
\tag{3.4.17}
$$

$$
\begin{aligned}
I_3 &= \int_{-\infty}^{0}\int_{x}^{0} [q_\kappa \exp(\beta x)]^2 [1 - q_\kappa \exp(\beta y)]^2 \, dy \, dx \\
&= \frac{q_\kappa^2}{4\beta^2} - \frac{2q_\kappa^3}{6\beta^2} + \frac{q_\kappa^4}{8\beta^2}
\end{aligned}
\tag{3.4.18}
$$

Changing back to parameterization (1.3.4) using equations (1.3.14) and (1.3.15) gives equation (3.4.11). □

Table 3.1 shows the approximate standard deviation $\sigma(X)$ and distance standard deviation $\mathscr{V}(X)$ for the $\mathscr{AL}^*$ distribution for various parameters σ and κ. Note that $\mathscr{V}(X) < \sigma(X)$ for each distribution in Table 3.1, as was shown in Edelmann et al. (2017).

$\sigma = \sqrt{2}\times$		1	1.2	1.4	1.5	1.6	1.7	1.8	2	3	4	5
$\kappa = 1$	$\sigma(X)$	1.41	1.70	1.98	2.12	2.26	2.40	2.55	2.83	4.24	5.66	7.07
	$\mathscr{V}(X)$	0.76	0.92	1.07	1.15	1.22	1.30	1.37	1.53	2.29	3.06	3.81
$\kappa = 1.5$	$\sigma(X)$	1.64	1.97	2.30	2.46	2.62	2.79	2.95	3.28	4.92	6.57	8.21
	$\mathscr{V}(X)$	0.91	1.09	1.27	1.36	1.46	1.55	1.64	1.82	2.73	3.64	4.55

Table 3.1 Values of $\sigma(X) = (\mathrm{Var}(X))^{1/2}$ and $\mathscr{V}(X)$ for $X \sim \mathscr{AL}^*(0, \sigma, \kappa)$.

3.5 Generalized Asymmetric Laplace Energy Test

The generalized asymmetric Laplace (also called variance gamma) distribution (Pearson, Jeffery, and Elderton, 1929; Teichroew, 1957) is an extension of asymmetric Laplace, and was first studied in relation to the sample covariance of the bivariate Normal distribution. Recently, the distribution has seen widespread use in financial modeling due to its flexibility and relatively excellent fit to empirical data (Madan and Seneta, 1990; Madan, Carr, and Chang, 1998; Levin and Tchernitser, 2003).

Parameter	Definition	Value
Mean	$\mathbb{E}[Y]$	$\theta + \tau\mu$
Variance	$\mathbb{E}[Y - \mathbb{E}[Y]]^2$	$\tau(\sigma^2 + \mu^2)$

Table 3.2 Common parameter values for a $\mathscr{GALL}(\theta, \mu, \sigma, \tau)$ distribution

Definition 3.5.1. *A random variable Y is said to have a generalized asymmetric Laplace (variance gamma) distribution if its CF is given by*

$$\phi_Y(t) = \frac{\exp(i\theta t)}{\left(1 + \frac{1}{2}\sigma^2 t^2 - i\mu t\right)^{\tau}}, \tag{3.5.2}$$

where $\theta, \mu \in \mathbb{R}$*, and* $\sigma, \tau > 0$*. We denote the distribution* $\mathscr{GALL}(\theta, \sigma, \mu, \tau)$*.*

Remark 3.5.3. *When* $\tau = 1$ *we recover the* $\mathscr{ALL}(\theta, \sigma, \mu)$ *distribution.*

As in 1.3.2, we may factor the CF, leading to a new parameterization,

$$\phi_Y(t) = e^{i\theta t}\left(\frac{1}{1 + i\frac{\sigma\kappa}{\sqrt{2}}t}\right)^{\tau}\left(\frac{1}{1 - i\frac{\sigma}{\sqrt{2}\kappa}t}\right)^{\tau}. \tag{3.5.4}$$

We denote this distribution $\mathscr{GALL}^*(\theta, \sigma, \kappa, \tau)$. This parameterization for the CF allows us to express the density function of the variance gamma distribution. Additionally, this parameterization constitutes a location-scale family in the parameters θ and σ (Kotz et al., 2001).

The expression of the density of the variance gamma distribution incorporates a Bessel function.

Definition 3.5.5. *The density of the* $\mathscr{GALL}^*(\theta, \sigma, \kappa, \tau)$ *distribution has the form*

$$\frac{\sqrt{2}\exp\left(\frac{\sqrt{2}}{2\sigma}(1/\kappa - \kappa)(x - \theta)\right)}{\sqrt{\pi}\sigma^{\tau + 1/2}\Gamma(\tau)}\left(\frac{\sqrt{2}|x - \theta|}{\kappa + \frac{1}{\kappa}}\right)^{\tau - \frac{1}{2}} K_{\tau - \frac{1}{2}}\left(\frac{\sqrt{2}}{2\sigma}\left(\frac{1}{\kappa} + \kappa\right)|x - \theta|\right). \tag{3.5.6}$$

Definition 3.5.7. *Abramowitz and Stegun (1964) The modified Bessel function of the first kind I with index λ is given by the convergent power series*

$$I_\lambda(x) = \sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m+\lambda+1)} \left(\frac{x}{2}\right)^{2m+\lambda}. \tag{3.5.8}$$

The modified Bessel function of the second kind K with index λ is given by

$$K_\lambda(x) = \begin{cases} \frac{\pi}{2}\frac{I_{-\lambda}(x)-I_\lambda(x)}{\sin(\lambda x)}, & \lambda \notin \mathbb{Z} \\ \lim_{\alpha\to\lambda} \frac{\pi}{2}\frac{I_{-\alpha}(x)-I_\alpha(x)}{\sin(\alpha x)}, & \lambda \in \mathbb{Z}. \end{cases} \tag{3.5.9}$$

Remark 3.5.10. *The function $K_\lambda(\cdot)$ is known by many names including the modified Bessel function of the Second Kind, the modified Bessel function of the Third Kind, and the modified Hankel function. In this dissertation, we will only refer to $K_\lambda(\cdot)$ as the modified Bessel function of the second kind.*

The following integral representations of K_λ are available respectively in the Abramowitz and Stegun (1964), Olver (1974), and Watson (1995).

Proposition 3.5.11. *The modified Bessel function of the second kind may be expressed as*

$$K_\lambda(x) = \frac{1}{2}\left(\frac{x}{2}\right)^\lambda \int_0^\infty t^{-\lambda-1} \exp\left(-t - \frac{x^2}{4t}\right) dt, \quad x > 0 \tag{3.5.12}$$

$$K_\lambda(x) = \frac{(x/2)^\lambda \Gamma(1/2)}{\Gamma(\lambda+1/2)} \int_1^\infty \exp(-xt)(t^2-1)^{\lambda-1/2}\, dt, \quad \lambda \geq -1/2 \tag{3.5.13}$$

$$K_\lambda(x) = \int_0^\infty \exp(-x\cosh(t))\cosh(\lambda t)\, dt, \quad \lambda \in \mathbb{R} \tag{3.5.14}$$

Lemma 3.5.15. *The modified Bessel function of the second kind is an even function of the index λ: $K_\lambda(x) = K_{-\lambda}(x)$.*

Proposition 3.5.16. *For any $\lambda = r + \frac{1}{2}$, r a non-negative integer, the modified Bessel function of the second kind has a closed form expression:*

$$K_{r+1/2}(x) = \sqrt{\frac{\pi}{2x}} \exp(-x) \sum_{k=0}^{r} \frac{(r+k)!}{(r-k)!k!} (2x)^{-k}. \tag{3.5.17}$$

In particular, if $r = 0$, we have

$$K_{1/2}(x) = \sqrt{\frac{\pi}{2x}} \exp(-x). \tag{3.5.18}$$

Proposition 3.5.16 leads to a crucial simplification of the variance gamma density 3.5.6 when $\tau = n$ is integer valued.

$$f(x|\theta = 0, \sigma = 1, \kappa, n) = \frac{1}{(n-1)!} \sum_{i=0}^{n-1} \frac{(n-1+i)!}{(n-1-j)!j!} \frac{2^{\frac{n-i}{2}} |x|^{n-1-i}}{(\kappa + 1/\kappa)^{n+i}} \begin{cases} \exp(-\sqrt{2}\kappa|x|), & x \geq 0; \\ \exp(-\frac{\sqrt{2}}{\kappa}|x|), & x < 0. \end{cases} \tag{3.5.19}$$

We will write the following to simplify notation.

Proposition 3.5.20. *The density given in equation* (3.5.19) *can be written as,*

$$f(x) = \sum_{i=0}^{n-1} A_i |x|^{n-1-i} \begin{cases} \exp(-\beta x), & x \geq 0; \\ \exp(\lambda x), & x < 0, \end{cases} \tag{3.5.21}$$

where $\lambda = \frac{\sqrt{2}}{\kappa}$ and $\beta = \sqrt{2}\kappa$, and

$$A_i = \frac{(n-1+i)! 2^{\frac{n-j}{2}}}{(n-1)!(n-1-i)!i!(\kappa + \frac{1}{\kappa})^{n+i}}. \tag{3.5.22}$$

3.5.1 Generalized Asymmetric Laplace Representations

The generalized asymmetric Laplace distribution admits many representations that mirror those studied in Section 1.5.

Proposition 3.5.23. *If* $Y \sim \mathscr{GAL}(\theta, \sigma, \mu, \tau)$ *then*

$$Y \stackrel{d}{=} \theta + \mu W + \sigma\sqrt{W}Z \tag{3.5.24}$$

where $Z \sim N(0,1)$ *and* $W \sim Gamma(shape = \tau,\ scale = 1)$. *Explicitly, the density function of* W *is given by*

$$f_W(x) = \frac{1}{\Gamma(\tau)} x^{\tau-1} \exp(-x), \qquad x > 0. \tag{3.5.25}$$

Proposition 3.5.26. *If* $Y \sim \mathscr{GAL}^*(\theta, \sigma, \kappa, \tau)$ *then*

$$Y \stackrel{d}{=} \theta + \frac{\sigma}{\sqrt{2}} \left(\frac{1}{\kappa} G_1 - \kappa G_2 \right), \tag{3.5.27}$$

where G_1, G_2 *are independent random variables with Gamma distribution given by (3.5.25).*

If $\tau = n$, a positive integer, then Gamma distribution G_i, $i = 1,2$ is equivalent to the Erlang distribution, the distribution of the sum of i.i.d. Exponential random variables.

Remark 3.5.28. *We prefer to use (3.5.27) to generate random variance gamma variates.*

If $\tau = n$, a positive integer, then the corresponding $\mathscr{GAL}^*(\theta, \sigma, \kappa, n)$ density is the density of the sum of n independent $\mathscr{AL}^*(\theta, \sigma, \kappa)$ random variables. In this case, the modified Bessel function of the second kind, $K_{n-1/2}(\cdot)$ admits the closed form expression (3.5.19).

Proposition 3.5.29. *Let* $(X_i, Y_i)'$, $i = 1, \ldots, n$, *be i.i.d.* $\mathscr{N}_2(0, \Sigma)$ *where*

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{3.5.30}$$

and define

$$\begin{aligned} T_n &= n \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}), \\ &= n \sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i \right) \left(\sum_{i=1}^{n} Y_i \right). \end{aligned} \tag{3.5.31}$$

Then $T_n \sim \mathscr{GALL}^*$ with parameters

$$\begin{aligned} \theta = 0, \quad \sigma = \sqrt{2}n\sqrt{1-\rho^2}, \\ \kappa = \sqrt{\frac{1-\rho}{1+\rho}}, \quad \tau = \frac{n-1}{2}. \end{aligned} \tag{3.5.32}$$

Remark 3.5.33. *For bivariate Normal data with an odd sample size, the distribution of T_n reduces to the simplified $\mathscr{GALL}$ because the parameter τ is an integer.*

3.5.2 Expected Distances for Generalized Asymmetric Laplace

The energy statistic (3.2.9) has a simple form, but as in Section 3.2, the expectations need to be further evaluated to ease computational load. The expectations $\mathbb{E}|Y-y|$ and $\mathbb{E}|Y-Y'|$ may be computed using techniques familiar from calculus.

Proposition 3.5.34. *Let $Y \sim \mathscr{GALL}^*(\theta = 0, \sigma = 1, \kappa, n)$. Then,*

$$\mathbb{E}|Y-y| = \sum_{i=0}^{n-1} A_i \begin{cases} \lambda^{i-n-1}(i-\lambda y-n)\Gamma(n-i) + \beta^{i-n-1}(n-i-\beta y)\Gamma(n-i) + \\ 2\lambda^{i-n-1}\gamma^*(n-i+1,-\lambda y) + 2\lambda^{i-n}\gamma^*(n-i,-\lambda y), & y<0; \\ \lambda^{i-n-1}(n+\lambda y-i)\Gamma(n-i) - \beta^{i-n-1}(n-i-\beta y)\Gamma(n-i) + \\ 2\beta^{i-n-1}\gamma^*(n-i+1,\beta y) - 2\beta^{i-n}\gamma^*(n-i,\beta y), & y<0. \end{cases} \tag{3.5.35}$$

The function $\gamma^*(a,x)$ is the upper incomplete gamma function (Abramowitz and Stegun, 1964).

Definition 3.5.36. *The upper incomplete gamma function is*

$$\gamma^*(a,x) = \int_x^\infty t^{a-1}\exp(-t)\,dt. \tag{3.5.37}$$

The function generally does not have an analytic form, but is accurately approximated by a number of computer implementations. If the index a in (3.5.37) is restricted to the positive integers, then γ^* has an analytic expression which is amenable to further calculations.

Proposition 3.5.38. *If a is a positive integer, then*

$$\gamma^*(a,x) = (a-1)!\exp(-x)\sum_{k=0}^{a-1}\frac{x^k}{k!}, \tag{3.5.39}$$

and the expression for the expected distance between two i.i.d. $\mathscr{GALL}^*$ random variables is analytic.

Proposition 3.5.40. *If $Y, Y' \sim \mathscr{GALL}^*(\theta = 0, \sigma = 1, \kappa, n)$, the expected distance between Y and Y' is*

$$\begin{aligned}\mathbb{E}|Y-Y'| = \sum_{i=0}^{n-1} A_i \big[&\lambda^{i-n-1}\Gamma(n-i)(iI_1 - \lambda I_2 - nI_1) + \beta^{i-n-1}\Gamma(n-i)(nI_1 - iI_1 - \beta I_2) \\ &+ \lambda^{i-n-1}\Gamma(n-i)(nI_3 + \lambda I_4 - iI_3) - \beta^{i-n-1}\Gamma(n-i)(nI_3 - iI_3 - \beta I_4) \\ &+ 2\beta^{i-n-1}I_5 + 2\lambda^{i-n-1}I_6 - 2\beta^{i-n}I_7 + 2\lambda^{i-n}I_8\big],\end{aligned} \tag{3.5.41}$$

where each of the functions I_j, $j = 1,\ldots,8$ are defined in the appendix.

The integrals I_j, $j = 1,\ldots,8$ may be computed using integration by parts. Using (3.5.40) and (3.5.34), the univariate energy goodness-of-fit statistic (3.2.9) may be computed for the variance gamma distribution. When the parameter τ is not integer valued, it is possible to use numerical integration to determine the values of the expected distances $\mathbb{E}|Y-y|$ and $\mathbb{E}|Y-Y'|$.

3.5.3 Variance Gamma Energy Test

We present simulations showing the power of the energy goodness-of-fit test for the simple variance gamma null hypothesis. Under the simple hypothesis, we consider the parameters of the distribution to be known. Therefore, the test may be written,

$$H_0 : F_n \sim \mathscr{GALL}(\theta_0, \sigma_0, \kappa_0, \tau_0). \tag{3.5.42}$$

In this simulation study, we take $\theta_0 = 0$ and $\sigma_0 = 1$, and the number of simulation replicates to be 5000. The p-values for the energy test are obtained via "on-the-fly" Monte Carlo with 200 replicates. The standard error of each empirical power estimate $\hat{p}$ is $\sqrt{\frac{\hat{p}(1-\hat{p})}{5000}} < \frac{0.5}{\sqrt{5000}} \approx 0.007$.

Figure 3.21 shows the power of the energy test against other tests when the skewness parameter κ is misspecified in samples of size 30. Under the null hypothesis, $y_1, \ldots, y_{30} \sim \mathscr{GAL}(0,1,1,2)$. The alternative is that $y_1, \ldots, y_{30} \sim \mathscr{GAL}(0,1,\kappa,2)$. We observe that the energy test and the Anderson-Darling tests are essentially equally powerful, however there is little difference between all four tests in this case. All tests empirically control the Type I error rate close to $\alpha = 0.1$ in this small sample size, as indicated on the plot at $\kappa = 1$.

Figure 3.22 shows the scenario of testing misspecified shape parameter τ. The null hypothesis for this simulation is that $y_1, \ldots, y_{30} \sim \mathscr{AL}(0,1,1.5)$ We find that all tests are very competitive when $\tau < 1$ in the alternative. When $\tau > 1$ under the alternative hypothesis, the energy test is the most powerful among the four tests we consider.

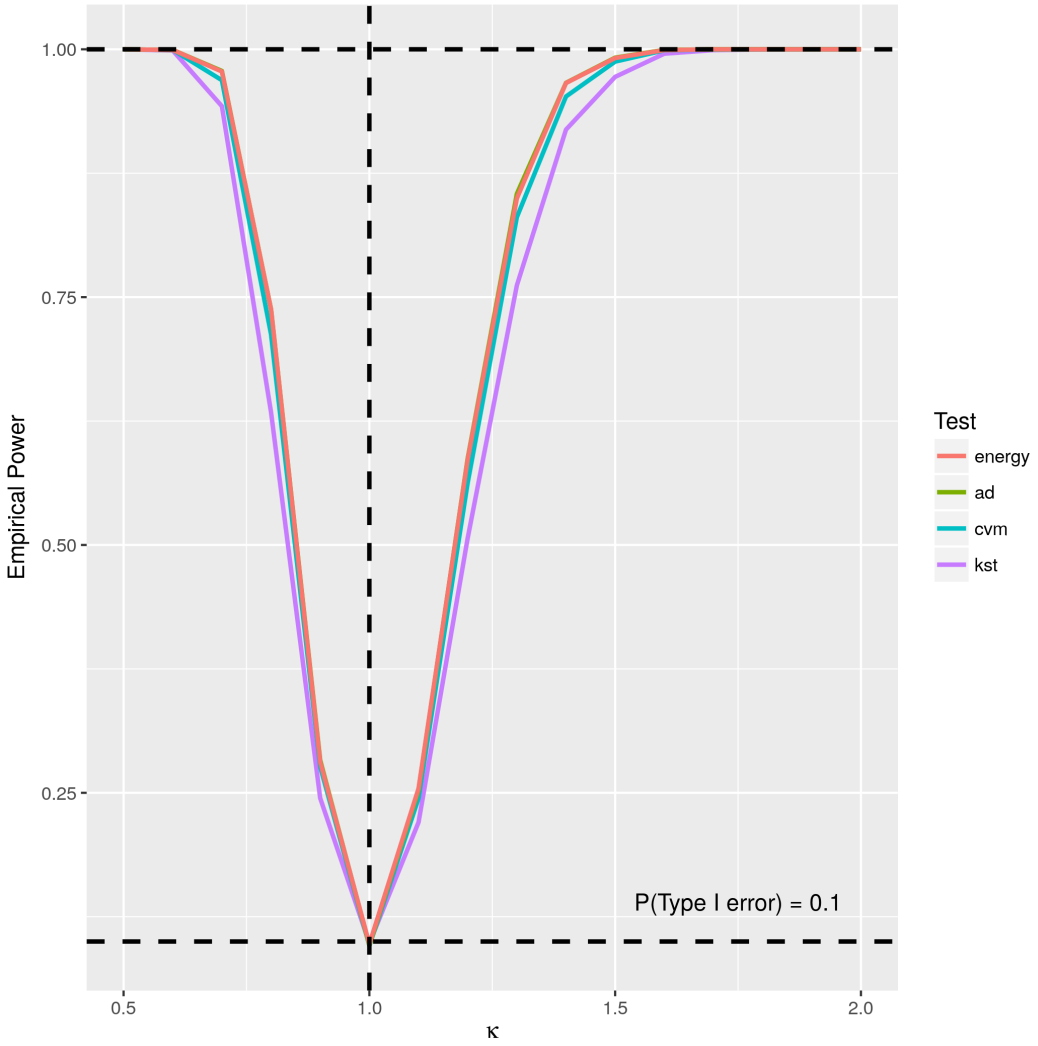

Figure 3.21 Power of testing $\mathscr{GAL}(0,1,1,2)$ against $\mathscr{GAL}(0,1,\kappa,2)$ with varying κ (sample size $= 30$)

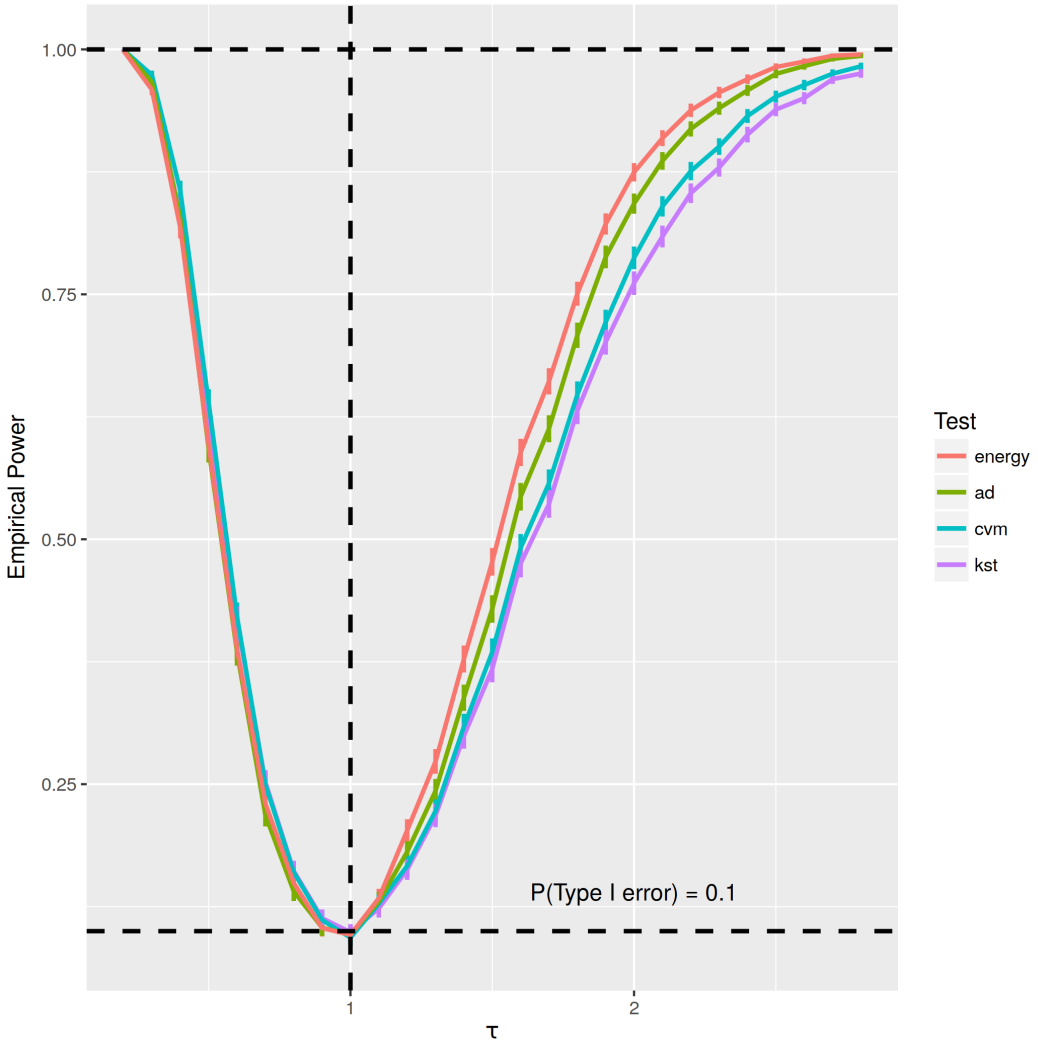

Figure 3.22 Power of testing $\mathscr{GALL}(0,1,1.5,1)$ against $\mathscr{GALL}(0,1,1.5,\tau)$ with varying τ (sample size $=30$)

CHAPTER 4 MULTIVARIATE LAPLACE PARAMETER ESTIMATION

Estimation of the parameters of the asymmetric Laplace distribution is in itself an interesting and important topic. In this dissertation, robust and reliable parameter estimation is crucial if a goodness of fit test is used for the composite goodness-of-fit hypothesis. Previous treatment of this problem has been performed in the univariate setting through maximum likelihood (ML) and method of moments (MM) estimation (Section 1.6). Furthermore, MM has been used to determine the parameters of the multivariate asymmetric Laplace distribution (Kollo and Srivastava, 2005; Visk, 2009; Hürlimann, 2013). In this chapter, the multivariate generalization of the asymmetric Laplace distribution is introduced, and an E-M type algorithm is detailed to address the parameter estimation problem from a ML perspective. The E-M estimator is applicable in both the univariate and multivariate setting and is shown to have better performance than the ML estimator of Kotz et al. (2001) in the univariate case, and the method of moments estimator of Visk (2009) in the multivariate case. Simulations that examine the bias and standard error of the E-M estimates are presented in Section 4.5.1. A proof-of-concept goodness-of-fit test for the multivariate asymmetric Laplace distribution is presented at the end of this chapter.

4.1 Multivariate Asymmetric Laplace Distribution

The multivariate Laplace distribution is the d-dimensional analog of the classical Laplace distribution. This distribution is useful for modeling multivariate data that exhibits leptokurtosis, asymmetry, and heavier than Normal tails. Let Σ be a $d \times d$ nonnegative definite symmetric matrix, and let $\theta = (\theta_1, \theta_2, \ldots, \theta_d)$ be a location parameter vector. The symbol $^\intercal$ denotes matrix transpose. Then a real, vector valued random variable $\mathbf{X}$ has multivariate Laplace distribution if it possesses the characteristic function

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}^\intercal \mathbf{X})] = \frac{e^{i\mathbf{t}^\intercal \theta}}{1 + \frac{1}{2}\mathbf{t}^\intercal \Sigma \mathbf{t}}. \tag{4.1.1}$$

We regard $\mathbf{t}$ as a fixed column vector and $\mathbf{t}^\intercal$ as a fixed row vector. A generalization of this distribution arises when a skewness parameter vector $\mathbf{m}$ is introduced. We follow the treatment of the multivariate asymmetric Laplace distribution given by Kotz, Kozubowkski and Podgórski (Kotz et al., 2001, Ch. 6), however we include $\theta = (\theta_1, \theta_2, \ldots, \theta_d)$ as a location parameter in this model. As in Chapter 1, we extend the model to account for asymmetry.

Definition 4.1.2. *A real, vector valued random variable* $\mathbf{Y}$ *has a multivariate asymmetric Laplace distribution* $\mathscr{AL}_d(\theta, \mathbf{m}, \Sigma)$ *if there exist parameters* $\theta \in \mathbb{R}^d$, $\mathbf{m} \in \mathbb{R}^d$, *and a* $d \times d$ *nonnegative definite symmetric matrix* Σ *such that the characteristic function of* $\mathbf{Y}$ *has the form*

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \frac{e^{i\mathbf{t}^\intercal \theta}}{1 + \frac{1}{2}\mathbf{t}^\intercal \Sigma \mathbf{t} - i\mathbf{m}^\intercal \mathbf{t}}, \quad \mathbf{t} \in \mathbb{R}^d. \tag{4.1.3}$$

When $\mathbf{m} = \mathbf{0}$, (4.1.3) is the CF of the symmetric multivariate Laplace distribution (4.1.1), which is an elliptically contoured distribution. This form of the CF has appeared earlier in articles on multivariate geometric stable distribution (Kozubowski, 1997; Kozubowski and Panorska, 1999).

The family $\mathscr{AL}_d(\theta, \mathbf{m}, \Sigma)$ is closed under translations and linear transformations.

Proposition 4.1.4. *(Visk (2009)) If* $\mathbf{y} \sim \mathscr{AL}_d(\theta, \mathbf{m}, \Sigma)$, *then for any* d*-dimensional vector* $\mathbf{a}$, $\mathbf{y} + \mathbf{a} \sim \mathscr{AL}_d(\theta + \mathbf{a}, \mathbf{m}, \Sigma)$.

Proposition 4.1.5. *If* $\mathbf{y} \sim \mathscr{AL}_d(\theta, \mathbf{m}, \Sigma)$, *then for any* $p \times d$ *full rank matrix* $\mathbf{A}$ *with* $p \leq d$, $\mathbf{Ay} \sim \mathscr{AL}_p(\mathbf{A}\theta, \mathbf{Am}, \mathbf{A}\Sigma\mathbf{A}^\intercal)$.

Proof. By properties of CFs,

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}[\exp(i(\mathbf{AY})^\intercal \mathbf{t})] = \mathbb{E}[\exp(i\mathbf{Y}^\intercal \mathbf{A}^\intercal \mathbf{t})] = \phi_{\mathbf{Y}}(\mathbf{A}^\intercal \mathbf{t}). \tag{4.1.6}$$

In terms of CF (4.1.3),

$$\frac{\exp(i(\mathbf{A}^\intercal \mathbf{t})^\intercal \theta)}{1 + \frac{1}{2}(\mathbf{A}^\intercal \mathbf{t})^\intercal \Sigma (\mathbf{A}^\intercal \mathbf{t}) - i\mathbf{m}^\intercal (\mathbf{A}^\intercal \mathbf{t})} = \frac{\exp(i\mathbf{t}^\intercal (\mathbf{A}\theta))}{1 + \frac{1}{2}\mathbf{t}^\intercal \mathbf{A}\Sigma\mathbf{A}^\intercal \mathbf{t} - i(\mathbf{Am})^\intercal \mathbf{t}}. \tag{4.1.7}$$

□

A consequence is that the multivariate asymmetric Laplace family may be centered and scaled to a "standard" distribution. If $\mathbf{Y} \sim \mathscr{AL}_d(\theta, \mathbf{m}, \Sigma)$, then $\Sigma^{-1/2}(\mathbf{Y} - \theta) \sim \mathscr{AL}_d(\mathbf{0}, \Sigma^{-1/2}\mathbf{m}, \mathbf{I})$.

Definition 4.1.8. *Let θ, $\mathbf{m} \in \mathbb{R}^d$ and Σ be a $d \times d$ positive definite matrix. The density of $\mathbf{Y} \sim \mathscr{AL}_d(\theta, \mathbf{m}, \Sigma)$ is*

$$f(\mathbf{y}; \theta, \mathbf{m}, \Sigma) = \frac{2e^{(\mathbf{y}-\theta)^\intercal \Sigma^{-1}\mathbf{m}}}{(2\pi)^{d/2}|\Sigma|^{1/2}} \left(\frac{(\mathbf{y}-\theta)^\intercal \Sigma^{-1}(\mathbf{y}-\theta)}{2 + \mathbf{m}^\intercal \Sigma^{-1}\mathbf{m}} \right)^{\nu/2} K_{\nu/2}\left(\sqrt{(2 + \mathbf{m}^\intercal \Sigma^{-1}\mathbf{m})((\mathbf{y}-\theta)^\intercal \Sigma^{-1}(\mathbf{y}-\theta)} \right), \tag{4.1.9}$$

where $\nu = \frac{2-d}{2}$ and $K_\nu(x)$ is the modified Bessel function of the second kind, defined in (3.5.9).

Remark 4.1.10. *We use the expression $q(\mathbf{y}) = (\mathbf{y}-\theta)^\intercal \Sigma^{-1}(\mathbf{y}-\theta)$ to simplify the expression of the density, allowing us to write*

$$f(\mathbf{y}; \theta, \mathbf{m}, \Sigma) = \frac{2e^{(\mathbf{y}-\theta)\Sigma^{-1}\mathbf{m}}}{(2\pi)^{d/2}|\Sigma|^{1/2}} \left(\frac{q(\mathbf{y})}{2 + \mathbf{m}^\intercal \Sigma^{-1}\mathbf{m}} \right)^{\nu/2} K_{\nu/2}\left(\sqrt{(2 + \mathbf{m}^\intercal \Sigma^{-1}\mathbf{m}) q(\mathbf{y})} \right). \tag{4.1.11}$$

The following representation for the multivariate asymmetric Laplace distribution is given. This representation is useful for generating $\mathscr{AL}_d$ variates, and is used in the derivation of the E-M algorithm.

Theorem 4.1.12. *Suppose $\mathbf{Y} \sim \mathscr{AL}_d(\theta, \mathbf{m}, \Sigma)$ and let $\mathbf{X} \sim \mathscr{N}_d(\mathbf{0}, \Sigma)$. Let Z be a standard, exponentially distributed random variable with mean 1, independent of $\mathbf{X}$. Then*

$$\mathbf{Y} \stackrel{d}{=} \theta + \mathbf{m}Z + \sqrt{Z}\mathbf{X}. \tag{4.1.13}$$

Proof. The proof is similar to that of Proposition 1.5.1. Let Z be an exponential random variable with mean 1. The density of Z is $f_Z(z) = \exp(-z)$. Let $\mathbf{X} \sim \mathscr{N}_d(0, \Sigma)$, independent of Z. Now we

consider the expression $\theta + \mathbf{m}Z + Z^{1/2}\mathbf{X}$. Conditioning on Z, we can express the CF of 4.1.13 as

$$\begin{aligned} &\mathbb{E}[\exp(i\mathbf{t}^\intercal(\theta + \mathbf{m}Z + Z^{1/2}\mathbf{X}))] \\ &= \int_0^\infty \exp(i\mathbf{t}^\intercal\theta + i\mathbf{t}^\intercal\mathbf{m}z)\mathbb{E}[\exp(i\mathbf{t}^\intercal z^{1/2}\mathbf{X})]\exp(-z)\,dz. \end{aligned} \tag{4.1.14}$$

The CF of $\mathbf{X}$ is $\phi_{\mathbf{X}}(\mathbf{t}) = \exp(-\frac{1}{2}\mathbf{t}^\intercal\Sigma\mathbf{t})$. It follows that $\phi_{\mathbf{X}}(z^{1/2}\mathbf{t}) = \exp(-\frac{1}{2}\mathbf{t}^\intercal\Sigma\mathbf{t}z)$. So,

$$\begin{aligned} &\int_0^\infty \exp(i\mathbf{t}^\intercal\theta + i\mathbf{t}^\intercal\mathbf{m}z)\mathbb{E}[\exp(i\mathbf{t}^\intercal z^{1/2}\mathbf{X})]\exp(-z)\,dz \\ &= \exp(i\mathbf{t}^\intercal\theta)\int_0^\infty \exp\left(-z\left(1 + \frac{1}{2}\mathbf{t}^\intercal\Sigma\mathbf{t} - t\mathbf{m}^\intercal\mathbf{t}\right)\right)dz \\ &= \frac{\exp(i\mathbf{t}^\intercal\theta)}{1 + \frac{1}{2}\mathbf{t}^\intercal\Sigma\mathbf{t} - i\mathbf{m}^\intercal\mathbf{t}}. \end{aligned} \tag{4.1.15}$$

This matches the CF given in (4.1.3). □

Parameter	Definition	Value
Mean	$\mathbb{E}[\mathbf{Y}]$	$\theta + \mathbf{m}$
Covariance	$\mathbb{E}[(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^\intercal]$	$\Sigma + \mathbf{m}\mathbf{m}^\intercal$

Table 4.1 Mean vector and covariance matrix for a $\mathscr{ALS}_d(\theta, \mathbf{m}, \Sigma)$ distribution.

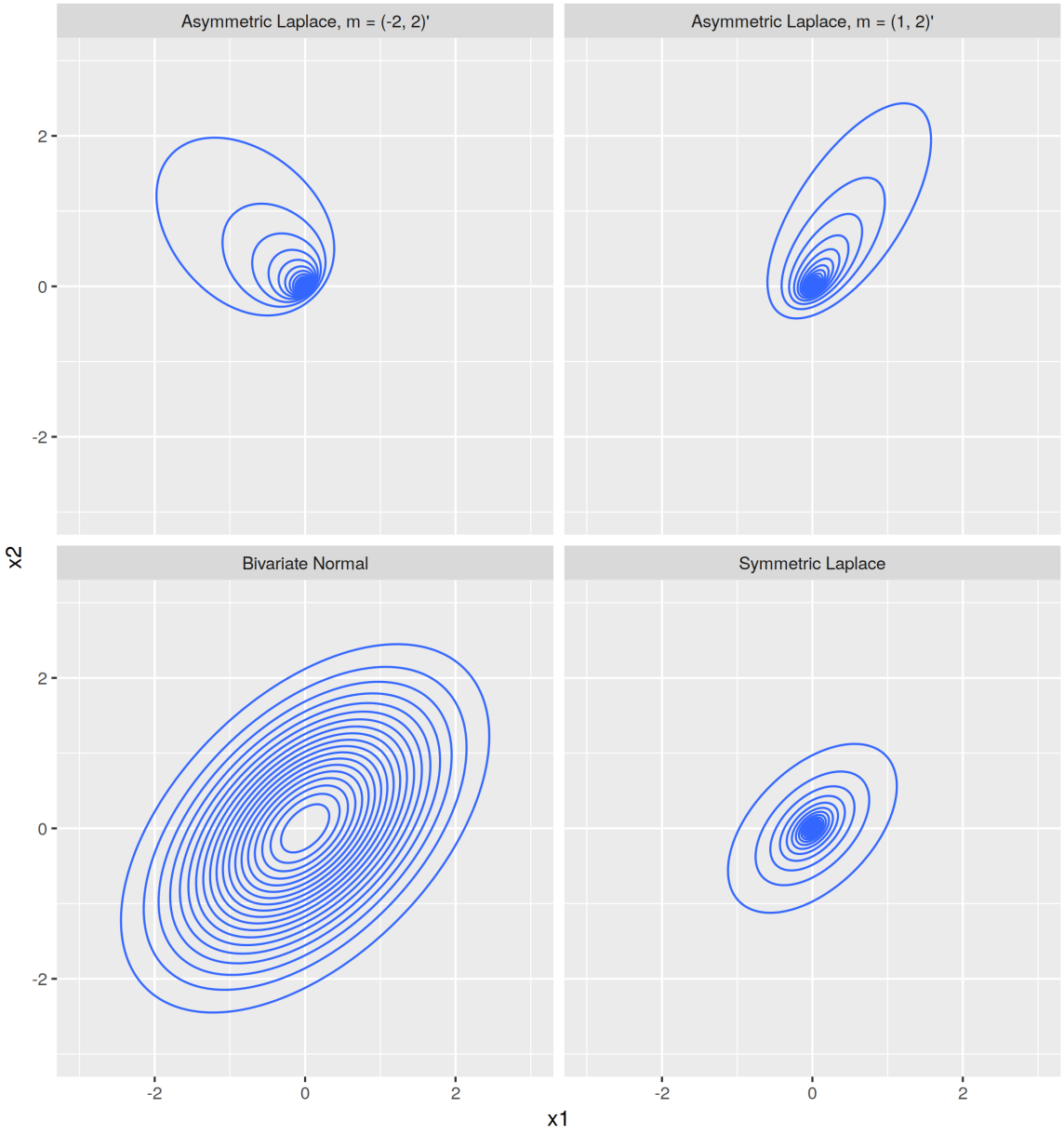

Figure 4.1 Contour plot of the densities of a bivariate Gaussian and three bivariate asymmetric Laplace distributions

4.2 Method of Moments Estimation

MM estimates when $\theta \neq 0$ have been studied in Visk (2009) and Hürlimann (2013), however these procedures can lead to negative estimates of the dispersion matrix Σ. These estimates are non-admissible. In the bivariate asymmetric Laplace distribution with location parameter $\theta = 0$, the distribution has CF

$$\phi(t_1,t_2) = \frac{1}{1+\frac{\sigma_1^2 t_1^2}{2}+\rho\sigma_1\sigma_2 t_1 t_2+\frac{\sigma_2^2 t_2^2}{2}-im_1 t_1-im_2 t_2}, \tag{4.2.1}$$

with parameters $\mathbf{m} = (m_1, m_2)^\intercal$ and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}. \tag{4.2.2}$$

If observations $\mathbf{Y} = (Y_{11}, Y_{12}), \ldots, (Y_{n1}, Y_{n2})$ are collected, then the corresponding MM estimators are:

$$\begin{aligned} \hat{m}_1 &= \frac{1}{n}\sum_{i=1}^{n} Y_{i1} = \bar{Y}_1; \\ \hat{m}_2 &= \frac{1}{n}\sum_{i=1}^{n} Y_{i2} = \bar{Y}_2; \\ \hat{\sigma}_1^2 &= \frac{1}{n}\sum_{i=1}^{n} Y_{i1}^2 - 2\bar{Y}_1^2; \\ \hat{\sigma}_2^2 &= \frac{1}{n}\sum_{i=1}^{n} Y_{i2}^2 - 2\bar{Y}_2^2; \\ \hat{\rho} &= \frac{\frac{1}{n}\sum_{i=1}^{n} Y_{i1}Y_{i2} - 2\bar{Y}_1\bar{Y}_2}{\hat{\sigma}_1\hat{\sigma}_2}. \end{aligned} \tag{4.2.3}$$

4.2.1 General Method of Moments

Visk's procedure applies in the general situation in which θ is unknown. The third moment of $\mathscr{ASL}_d(\theta,\Sigma,\mathbf{m})$ ("coskewness") is (Visk, 2009)

$$\bar{m}_3 \equiv \mathbb{E}[(\mathbf{Y}-\theta)\otimes(\mathbf{Y}-\theta)^\intercal\otimes(\mathbf{Y}-\theta)] = 2\mathbf{m}\otimes\mathbf{m}\mathbf{m}^\intercal + \mathrm{vec}(\Sigma)\mathbf{m}^\intercal + 2\mathbf{m}\otimes\Sigma + 2\Sigma\otimes\mathbf{m},$$

and is used when determining the moment estimators. The star product (MacRae, 1974) is defined for two matrices $\mathbf{A}_{p\times q}$ and $\mathbf{B}_{rp\times qs}$ where $\mathbf{B}$ consists of $r\times s$ size blocks $\mathbf{B}_{ij}$, $i=1,\ldots,p$, $j=1,\ldots,q$. The star product of $\mathbf{A}$ and $\mathbf{B}$ is defined to be the $r\times s$ matrix

$$\mathbf{A}\star\mathbf{B} = \sum_{i=1}^{p}\sum_{j=1}^{q} a_{ij}\mathbf{B}_{ij}.$$

The star product is a useful matrix algebra tool for "reducing" the size of a matrix. Visk derives the star product

$$\mathbf{1}_{p\times p}\star\bar{m}_3 = (S-M^2)\mathbf{m} + 2M\dot{(}\Sigma+\mathbf{m}\mathbf{m}^\intercal)\mathbf{1}_p \tag{4.2.4}$$

with $M=\sum_{i=1}^{d} m_i$ and $S=\sum_{i,j=1}^{d}(\Sigma+\mathbf{m}\mathbf{m}^\intercal)$. If M is known, the estimate for $\mathbf{m}$ can be computed. Summing (4.2.4) on both sides leads to a cubic equation, which yields an estimate for M:

$$M = \frac{1}{S-M^2}(H-2MS),$$

or

$$M - 3MS + H = 0, \tag{4.2.5}$$

with $H=\sum_{i=1}^{d^2}\sum_{j=1}^{d}(\bar{m}_3)_{ij}$. As in Kollo (2008), we estimate the $p^2\times p$ matrix of third central moments by

$$\hat{m}_3 = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})\otimes(Y_i-\bar{Y})^\intercal\otimes(Y_i-\bar{Y}) \tag{4.2.6}$$

so

$$\hat{H} = \sum_{i=1}^{d^2} \sum_{j=1}^{d} \hat{m}_3 \tag{4.2.7}$$

and

$$\hat{S} = \sum_{i=1}^{d} \sum_{j=1}^{d} \hat{\mathrm{Cov}}(Y). \tag{4.2.8}$$

There is one solution to (4.2.5) that preserves the positive definiteness of Σ:

$$M = \frac{-z}{4} - \frac{S}{z} - \frac{i\sqrt[3]{3}}{2}\left(\frac{z}{2} - \frac{2S}{z}\right), \qquad z = \sqrt[3]{-4H + 4\sqrt{H^2 - 4S^3}}. \tag{4.2.9}$$

The MM point estimate $\hat{\mathbf{m}}$ is given by (4.2.9) and (4.2.4). With $\hat{\mathbf{m}}$ computed, the remaining parameter estimates may be computed through

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^\intercal - \hat{\mathbf{m}}\hat{\mathbf{m}}^\intercal, \qquad \hat{\theta} = \bar{\mathbf{y}} - \hat{\mathbf{m}}. \tag{4.2.10}$$

The procedure is extended by Hürlimann (2013) to incorporate fourth central moments ("cokurtosis") of the asymmetric Laplace distribution.

4.3 Review of the E-M Algorithm

Let $\theta \in \Theta$ be a set of model parameters. The expectation-maximization (E-M) algorithm (Dempster, Laird, and Rubin, 1977) is an iterative procedure which may be used to produce ML parameter estimates, i.e., solutions to the equations

$$\frac{\partial}{\partial \theta} L(\theta | \mathbf{y}) = \mathbf{0}. \tag{4.3.1}$$

The procedure is useful when analytic estimates are difficult or impossible to derive due to the form of the assumed likelihood function underlying the data. It is applicable for estimation of parameters of a distribution from a data set that contains missing or incomplete data. The motivation of the procedure is based on the notion of missingness, and giving consideration to the distribution of what is missing given what is observed (Givens and Hoeting, 2012).

There are two main applications of this procedure. It may be that data are truly missing due to problems or limitations of the sampling process. The second (more common) application is to formulate an intractable likelihood function $L(\theta|\mathbf{y})$ (called the "incomplete data likelihood") in terms of a simpler likelihood $L_c(\theta|\mathbf{x})$ ("complete data likelihood") by positing the existence of latent variables X (McLachlan and Krishnan, 2007).

A brief description of the E-M algorithm is given based on details in McLachlan and Krishnan (2007). Suppose that $\mathscr{X}$ and $\mathscr{Y}$ are the sample spaces of $\mathbf{X}$ and $\mathbf{Y}$ respectively and let $f_c(\mathbf{x}|\theta)$ denote the probability density function of the complete data $\mathbf{X}$. Furthermore, let there be a many-to-one function $g : \mathscr{X} \to \mathscr{Y}$. Rather than observe the complete data $\mathbf{x}$ in $\mathscr{X}$, we observe the incomplete data $\mathbf{y}$ in $\mathscr{Y}$. Then the incomplete data probability density is the marginal density of the complete data density:

$$f(\mathbf{y}|\theta) = \int_{\mathscr{X}(y)} f_c(\mathbf{x}|\theta)\, d\mathbf{x}, \tag{4.3.2}$$

where $\mathscr{X}(y)$ is the subset of $\mathscr{X}$ determined by $\mathbf{y} = g(\mathbf{x})$. The E-M algorithm solves (4.3.1), the incomplete data likelihood equation, iteratively in terms of the complete data likelihood, L_c. Because L_c is unknown, it is replaced by its conditional expectation $\mathbb{E}[L_c]$ given the data $\mathbf{y}$ and the current estimate of θ.

First, initial parameters estimates $\theta^{(0)}$ are proposed. The "E-step" of the algorithm is to calculate the conditional expectation of L_c given $\theta^{(0)}$.

1. E-step:

$$Q(\theta|\theta^{(0)}) = \mathbb{E}_{\theta^{(0)}}[\log L_c(\theta|\mathbf{x})|\mathbf{y}] \tag{4.3.3}$$

2. M-step: Select $\theta^{(1)}$ such that

$$Q(\theta^{(1)}|\theta^{(0)}) \geq Q(\theta|\theta^{(0)}) \tag{4.3.4}$$

for all $\theta \in \Theta$. The E-step and the M-step are then iterated until some convergence criterion is met. An important result of Dempster et al. (1977) is that the E-M algorithm must converge to a local

or global maximum. This is guaranteed by the monotonicity of the incomplete likelihood, $L(\theta^{(t+1)}|\mathbf{y}) \geq L(\theta^{(t)}|\mathbf{y})$.

Stopping criteria for the E-M algorithm are usually built upon $(\theta^{(t+1)} - \theta^{(t)})^\intercal(\theta^{(t+1)} - \theta^{(t)})$ or $|Q(\theta^{(t+1)}|\theta^{(t)}) - Q(\theta^{(t)}|\theta^{(t-1)})|$.

4.4 Application of E-M to the Asymmetric Laplace Distribution

The E-M algorithm may be applied to determine the parameters of the multivariate asymmetric Laplace distribution. Given representation (4.1.13), we observe that the conditional distribution of $\mathbf{Y}|Z$ is multivariate Normal with mean $\theta + \mathbf{m}Z$ and variance-covariance matrix $Z\Sigma$. A Generalized Inverse Gaussian (GIG) random variable is defined in Barndorff-Nielsen (1997) as a random variable having probability density function

$$f_{GIG}(z;\gamma,\delta,\lambda) = \left(\frac{\gamma}{\delta}\right)^{\lambda} (2K_\lambda(\delta\gamma))^{-1} z^{\lambda-1} \exp\left(-\frac{1}{2}(\delta^2 z^{-1} + \gamma^2 z)\right). \tag{4.4.1}$$

Theorem 4.4.2. *The distribution of $Z|\mathbf{Y} = \mathbf{y}$ is GIG with parameters*

$$\gamma = \sqrt{2 + \mathbf{m}^\intercal \Sigma^{-1} \mathbf{m}}, \tag{4.4.3}$$

$$\delta = \sqrt{q(\mathbf{y})}, \tag{4.4.4}$$

$$\lambda = \frac{2-d}{2}. \tag{4.4.5}$$

Theorem 4.4.2 is due to Barndorff-Nielsen (1997), but we give a new proof.

Proof. Suppose $\mathbf{Y} \sim \mathscr{ALL}_d(\theta, \mathbf{m}, \Sigma)$. By Theorem 4.1.12, $\mathbf{Y} \stackrel{d}{=} \theta + \mathbf{m}Z + \sqrt{Z}\mathbf{X}$ with $Z \sim Exp(1)$ and $\mathbf{X} \sim \mathscr{N}_d(\mathbf{0}, \Sigma)$. It is clear that $[\mathbf{Y}|Z = z] \sim \mathscr{N}_d(\theta + \mathbf{m}z, z\Sigma)$. Because $\mathbf{Y}|Z$ is multivariate normal, the density of $[\mathbf{Y}|Z = z]$ is

$$f_{\mathbf{Y}|Z=z}(\mathbf{y}) = \frac{1}{(2\pi z)^{d/2}\det(\Sigma)^{1/2}} \exp\left[-\frac{1}{2z}(\mathbf{y} - (\theta + \mathbf{m}z))^\intercal \Sigma^{-1} (\mathbf{y} - (\theta + \mathbf{m}z))\right]. \tag{4.4.6}$$

The density of Z is $f_Z(z) = \exp(-z)$. The density of $[Z|\mathbf{Y}=\mathbf{y}]$ is $f_{Z|\mathbf{Y}=\mathbf{y}}(z) = \frac{f_{\mathbf{Y}|Z=z}(\mathbf{y})f_Z(z)}{f_{\mathbf{Y}}(\mathbf{y})}$ by Bayes' Theorem. Therefore, the density of $[Z|\mathbf{Y}=\mathbf{y}]$ is given by

$$\frac{\frac{\exp(-z)}{(2\pi z)^{d/2}\det(\Sigma)^{1/2}}\exp\left[-\frac{1}{2z}(\mathbf{y}-(\theta+\mathbf{m}z))^\intercal\Sigma^{-1}(\mathbf{y}-(\theta+\mathbf{m}z))\right]}{\frac{2\exp((\mathbf{y}-\theta)^\intercal\Sigma^{-1}\mathbf{m})}{(2\pi)^{d/2}\det(\Sigma)^{1/2}}\left(\frac{q(\mathbf{y})}{2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m}}\right)^{\frac{2-d}{4}}K_{\frac{2-d}{2}}\left(\sqrt{q(\mathbf{y})(2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m})}\right)}. \tag{4.4.7}$$

Grouping terms together,

$$f_{Z|\mathbf{Y}=\mathbf{y}}(z) = \left(\frac{\sqrt{q(\mathbf{y})}}{\sqrt{2+\mathbf{m}\Sigma^{-1}\mathbf{m}}}\right)^{\frac{-d}{2}-1}\frac{z^{-d/2}\exp\left[\frac{-1}{2z}(\mathbf{y}-\theta-\mathbf{m}z)^\intercal\Sigma^{-1}(\mathbf{y}-\theta-\mathbf{m}z)-z\right]}{2\exp((\mathbf{y}-\theta)^\intercal\Sigma^{-1}\mathbf{m})K_{1-\frac{d}{2}}(\sqrt{q(\mathbf{y})(2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m})})}. \tag{4.4.8}$$

Let $\lambda = \frac{2-d}{2}$, $\delta = \sqrt{q(\mathbf{y})}$, and $\gamma = \sqrt{2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m}}$. Then we may write the density as

$$f_{Z|\mathbf{Y}=\mathbf{y}}(z) = \frac{\left(\frac{\gamma}{\delta}\right)^\lambda z^{\lambda-1}\exp\left[\frac{-1}{2z}(\mathbf{y}-\theta-\mathbf{m}z)^\intercal\Sigma^{-1}(\mathbf{y}-\theta-\mathbf{m}z)-z-(\mathbf{y}-\theta)^\intercal\Sigma^{-1}\mathbf{m}\right]}{2K_\lambda(\delta\gamma)}. \tag{4.4.9}$$

The final part is to show that

$$-\frac{1}{2z}(\mathbf{y}-\theta-\mathbf{m}z)^\intercal\Sigma^{-1}(\mathbf{y}-\theta-\mathbf{m}z)-z-(\mathbf{y}-\theta)^\intercal\Sigma^{-1}\mathbf{m} = -\frac{1}{2}\left(\frac{q(\mathbf{y})}{z}+(2+\mathbf{m}\Sigma^{-1}\mathbf{m})z\right). \tag{4.4.10}$$

Denote the left hand side ★.

$$\begin{aligned}
\bigstar &= -\frac{1}{2z}\left[(\mathbf{y}-\theta)^\intercal\Sigma^{-1}(\mathbf{y}-\theta)-2z^2(\mathbf{y}-\theta)^\intercal\Sigma^{-1}\mathbf{m}+z^2\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m}\right]-z-(\mathbf{y}-\theta)^\intercal\Sigma^{-1}\mathbf{m},\\
&= -\frac{1}{2}\frac{q(\mathbf{y})}{z}+(\mathbf{y}-\theta)^\intercal\Sigma^{-1}\mathbf{m}z-\frac{1}{2}\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m}-z-(\mathbf{y}-\theta)^\intercal\Sigma^{-1}\mathbf{m}z,\\
&= -\frac{1}{2}\frac{q(\mathbf{y})}{z}-\frac{1}{2}z\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m}-z,\\
&= -\frac{1}{2}\left[\frac{q(\mathbf{y})}{z}+(2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m})z\right],\\
&= -\frac{1}{2}\left[\frac{\delta^2}{z}+\gamma^2 z\right].
\end{aligned}$$

Therefore density (4.4.7) may be written in the form of the density given in equation (4.4.1) and $Z|\mathbf{Y}=\mathbf{y}$ is GIG. □

Theorem 4.4.11. *The raw moments of a random variable Z that possesses a GIG distribution are (Barndorff-Nielsen, 1997)*

$$\mathbb{E}[Z^k] = \left(\frac{\delta}{\gamma}\right)^k \frac{K_{\lambda+k}(\gamma\delta)}{K_\lambda(\gamma\delta)}. \tag{4.4.12}$$

Two of these raw moments, $\mathbb{E}[Z|\mathbf{Y}=\mathbf{y}_i]$ and $\mathbb{E}[\frac{1}{Z}|\mathbf{Y}=\mathbf{y}_i]$ are useful for deriving the E-M algorithm. We define

$$\eta_i \equiv \mathbb{E}[Z|\mathbf{Y}=\mathbf{y}_i] = \frac{\sqrt{q(\mathbf{y}_i)}}{\sqrt{2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m}}} \frac{K_{2-d/2}\left(\sqrt{q(\mathbf{y}_i)(2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m})}\right)}{K_{1-d/2}\left(\sqrt{q(\mathbf{y}_i)(2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m})}\right)}, \tag{4.4.13}$$

and

$$\xi_i \equiv \mathbb{E}\left[\frac{1}{Z}|\mathbf{Y}=\mathbf{y}_i\right] = \frac{\sqrt{2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m}}}{\sqrt{q(\mathbf{y}_i)}} \frac{K_{-d/2}\left(\sqrt{q(\mathbf{y}_i)(2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m})}\right)}{K_{1-d/2}\left(\sqrt{q(\mathbf{y}_i)(2+\mathbf{m}^\intercal\Sigma^{-1}\mathbf{m})}\right)}. \tag{4.4.14}$$

Because the conditional distribution of $[\mathbf{Y}|Z=z_i]$ is multivariate Gaussian, one may write the log-likelihood function ℓ of $[\mathbf{Y}|Z=z_i]$ as

$$\ell(\theta,\mathbf{m},\Sigma) \propto -\frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{z_i}(\mathbf{y}_i-\mathbf{m}z_i-\theta)^\intercal\Sigma^{-1}(\mathbf{y}_i-\mathbf{m}z_i-\theta). \tag{4.4.15}$$

Theorem 4.4.16 (E-M Estimates for $\mathscr{ALL}_d$ — M-step)**.** *Taking partial derivatives of* (4.4.15) *and equating to zero yields*

$$\hat{\theta} = \frac{\sum_{i=1}^{n} \frac{1}{z_i}\mathbf{y}_i - n\hat{\mathbf{m}}}{\sum_{i=1}^{n} \frac{1}{z_i}}, \tag{4.4.17}$$

$$\hat{\mathbf{m}} = \frac{\sum_{i=1}^{n} \mathbf{y}_i - n\hat{\theta}}{\sum_{i=1}^{n} z_i}, \tag{4.4.18}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{z_i}\left(\mathbf{y}_i - \hat{\theta}\right)\left(\mathbf{y}_i - \hat{\theta}\right)^{\intercal} - z_i\hat{\mathbf{m}}\hat{\mathbf{m}}^{\intercal}. \tag{4.4.19}$$

Proof. Suppose $\mathbf{Y} \sim \mathscr{ALL}_d(\theta, \mathbf{m}, \Sigma)$. By Theorem 4.1.12, $[\mathbf{Y}|Z = z_i] \sim \mathscr{N}_d(\theta + \mathbf{m}z_i, z_i\Sigma)$. Thus, the log-likelihood $\ell(\theta, \mathbf{m}, \Sigma)$ of the data is given by equation (4.4.15). We calculate the partial derivatives of ℓ with respect to each parameter.

$$\frac{\partial \ell(\theta, \mathbf{m}, \Sigma)}{\partial \theta} = \frac{1}{2}\sum_{i=1}^{n} \frac{1}{z_i}\left[2\Sigma^{-1}(\mathbf{y}_i - \mathbf{m}z_i - \theta)\right], \tag{4.4.20}$$

$$\frac{\partial \ell(\theta, \mathbf{m}, \Sigma)}{\partial \mathbf{m}} = \frac{1}{2}\sum_{i=1}^{n} \frac{1}{z_i}\left[2\Sigma^{-1}(\mathbf{y}_i - \mathbf{m}z_i - \theta)\right] z_i, \tag{4.4.21}$$

$$\frac{\partial \ell(\theta, \mathbf{m}, \Sigma)}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{n} \frac{1}{z_i}\Sigma^{-1}(\mathbf{y}_i - \mathbf{m}z_i - \theta)(\mathbf{y}_i - \mathbf{m}z_i - \theta)^{\intercal}\Sigma^{-1}. \tag{4.4.22}$$

Solving for θ in (4.4.20),

$$\begin{aligned} \sum_{i=1}^{n} \frac{1}{z_i}\Sigma^{-1}(\mathbf{y}_i - \mathbf{m}z_i - \theta) &\stackrel{set}{=} 0, \\ \sum_{i=1}^{n} \frac{1}{z_i}(\mathbf{y}_i - \mathbf{m}z_i - \theta) &= 0, \\ \sum_{i=1}^{n} \frac{1}{z_i}(\mathbf{y}_i - \mathbf{m}z_i) &= \sum_{i=1}^{n} \frac{1}{z_i}\theta. \end{aligned} \tag{4.4.23}$$

Therefore

$$\hat{\theta} = \frac{\sum_{i=1}^{n} \frac{1}{z_i}(\mathbf{y}_i - \hat{\mathbf{m}}z_i)}{\sum_{i=1}^{n} \frac{1}{z_i}} = \frac{\sum_{i=1}^{n} \frac{1}{z_i}\mathbf{y}_i - n\hat{\mathbf{m}}}{\sum_{i=1}^{n} \frac{1}{z_i}}. \tag{4.4.24}$$

Solving for $\mathbf{m}$ in (4.4.21),

$$\begin{aligned}
\sum_{i=1}^{n} \frac{1}{z_i} \left[\Sigma^{-1}(\mathbf{y}_i - \mathbf{m}z_i - \theta)\right] &\stackrel{set}{=} 0, \\
\sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{m}z_i - \theta) &= 0, \\
\sum_{i=1}^{n} (\mathbf{y}_i - \theta) - \mathbf{m}\sum_{i=1}^{n} z_i &= 0, \\
\sum_{i=1}^{n} \mathbf{y}_i - n\theta &= \mathbf{m}\sum_{i=1}^{n} z_i.
\end{aligned} \tag{4.4.25}$$

Therefore

$$\hat{\mathbf{m}} = \frac{\sum_{i=1}^{n} \mathbf{y}_i - n\hat{\theta}}{\sum_{i=1}^{n} z_i}. \tag{4.4.26}$$

Solving for Σ in (4.4.22),

$$\begin{aligned}
-\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{n} \frac{1}{z_i}\Sigma^{-1}(\mathbf{y}_i - \mathbf{m}z_i - \theta)(\mathbf{y}_i - \mathbf{m}z_i - \theta)^{\intercal}\Sigma^{-1} &\stackrel{set}{=} \mathbf{0}, \\
-\frac{n}{2}I + \frac{1}{2}\sum_{i=1}^{n} \frac{1}{z_i}\Sigma^{-1}(\mathbf{y}_i - \mathbf{m}z_i - \theta)(\mathbf{y}_i - \mathbf{m}z_i - \theta)^{\intercal} &= \mathbf{0}, \\
\sum_{i=1}^{n} \frac{1}{z_i}\Sigma^{-1}(\mathbf{y}_i - \mathbf{m}z_i - \theta)(\mathbf{y}_i - \mathbf{m}z_i - \theta)^{\intercal} &= nI.
\end{aligned} \tag{4.4.27}$$

Therefore

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{z_i}(\mathbf{y}_i - \hat{\theta})(\mathbf{y}_i - \hat{\theta})^{\intercal} - z_i\hat{\mathbf{m}}\hat{\mathbf{m}}^{\intercal}. \tag{4.4.28}$$

□

We propose the following E-M algorithm, which corrects the algorithm given by Eltoft, Kim, and Lee (2006). The difference between the algorithm proposed in this text, and the one of Eltoft et al. (2006) is in the M-step of the E-M algorithm. Where the previous algorithm relied on a regression to estimate parameters, we propose an algorithm that relies the explicit maximization of $(\theta, \mathbf{m}, \Sigma)$.

4.4.1 E-M Algorithm for Multivariate Asymmetric Laplace Distribution

1. Initialize the parameter estimates

$$\hat{\theta}_j^{(0)} = \frac{1}{n}\sum_{i=1}^{n} y_{ij}, \quad j = 1,\ldots,d, \tag{4.4.29}$$

$$\hat{\mathbf{m}}^{(0)} = \mathbf{0}, \tag{4.4.30}$$

$$\hat{\Sigma}^{(0)} = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{y}_i - \hat{\theta})(\mathbf{y}_i - \hat{\theta})^{\intercal}. \tag{4.4.31}$$

2. Calculate η_i and ξ_i, $i = 1,\ldots,n$, using (4.4.13) and (4.4.14) and the most recent parameter estimates of θ, Σ and $\mathbf{m}$.

3. Update the values of the parameter estimates in the following order:

$$\begin{aligned}
\hat{\theta}^{(t)} &= \frac{\sum_{i=1}^{n} \xi_i \mathbf{y}_i - n\hat{\mathbf{m}}^{(t-1)}}{\sum_{i=1}^{n} \xi_i}, \\
\hat{\mathbf{m}}^{(t)} &= \frac{\sum_{i=1}^{n} \mathbf{y}_i - n\hat{\theta}^{(t)}}{\sum_{i=1}^{n} \eta_i}, \\
\hat{\Sigma}^{(t)} &= \frac{1}{n}\sum_{i=1}^{n} \xi_i \left(\mathbf{y}_i - \hat{\theta}^{(t)}\right)\left(\mathbf{y}_i - \hat{\theta}^{(t)}\right)^{\intercal} - \eta_i \mathbf{m}^{(t)}\mathbf{m}^{(t)\intercal}.
\end{aligned}$$

4. Repeat steps 2 and 3 until a convergence criterion is achieved. Because the density of the multivariate asymmetric Laplace distribution is unbounded at θ, we recommend stopping the procedure when $(\theta^{(t)} - \theta^{(t-1)})^{\intercal}(\theta^{(t)} - \theta^{(t-1)}) < \varepsilon$, where ε is a small threshold supplied by the user.

If any of the parameters are known a priori, we can skip the update step for the respective parameter. The code for this procedure was written in C++ and is given in Appendix B.

4.4.2 Bias Corrected E-M Estimation

Because the parameter estimates yielded by the E-M algorithm are generally biased (investigated empirically in Section 4.5.3), we propose a bootstrap method that ameliorates a

portion of the bias. The method is based on bootstrap resampling and detailed in Rizzo (2007). Let $\hat{\zeta}$ be an estimate of the parameter ζ of a statistical distribution. The bias of $\hat{\zeta}$ for ζ is defined to be $\text{bias}_{\zeta}(\hat{\zeta}) = \mathbb{E}[\hat{\zeta}] - \zeta$.

To estimate the bias of a parameter estimate $\hat{\zeta}$, we take B samples of size n with replacement from the original data and calculate $\hat{\zeta}$ on each of the B synthetic data sets. Thus we have a collection $\hat{\zeta}^{(1)}, \ldots, \hat{\zeta}^{(B)}$ for each of the B resampled data sets. The bootstrap estimate of bias is $\widehat{\text{bias}}_{\zeta}(\hat{\zeta}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\zeta}^{(b)} - \hat{\zeta}) = \overline{\hat{\zeta}^{(b)}} - \hat{\zeta}$. Therefore, to correct the bias inherent to an estimator $\hat{\zeta}$, we may use the estimator $\hat{\zeta}_{BC} \equiv \hat{\zeta} - \widehat{\text{bias}}_{\zeta}(\hat{\zeta}) = \hat{\zeta} - (\overline{\hat{\zeta}^{(b)}} - \hat{\zeta}) = 2\hat{\zeta} - \overline{\hat{\zeta}^{(b)}}$.

Empirically we find that the E-M algorithm is quick to converge to the ML estimates, usually taking less than 20 iterations to find optimal estimates. Due to this computational efficiency, bootstrapping the E-M method does not present a great computational burden, and we have found that compiling the algorithm in C++ and creating 500 bootstrap replicates takes a few seconds on an Intel dual core i7-4600U processor. However the decrease in bias of the estimators comes at a cost — larger standard errors.

The jack-knife estimator of bias cannot be used with our E-M algorithm as the estimator of θ is not a "smooth" function of the data (Rizzo, 2007). Simulations involving the bias corrected E-M estimator are shown in Section 4.6.

4.5 E-M Algorithm Simulation

In this section, results of an extensive simulation study are reported to investigate the empirical properties of the proposed E-M estimation methods. All simulations were implemented in the R statistical computing environment (R Core Team, 2018) and executed on a Dell E7240 laptop with an Intel i7-4600U CPU or at the Ohio Supercomputer Center (Ohio Supercomputer Center, 1987) Owens cluster.

A bivariate asymmetric Laplace distribution is used. We considered a distribution with the parameters given by

$$\mathbf{Y} \sim \mathscr{AL}_2 \left(\theta = \begin{bmatrix} -5 \\ 3 \end{bmatrix}, \mathbf{m} = \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & 1.5 \\ 1.5 & 1 \end{bmatrix} \right). \tag{4.5.1}$$

The simulation was conducted in two scenarios:

1. Location parameter θ is *known*.

2. Location parameter θ is *unknown*.

For sample sizes $n = 25, 50, 100, 250, 500, 1000, 2000$, we selected $10,000$ replicates from distribution (4.5.1). At each replication, the E-M algorithm and the MM algorithm were applied and the estimates were collected. An algorithm that describes our random variable generation procedure is detailed below.

4.5.1 Simulating Multivariate Asymmetric Laplace Variates

We use the simulation method of Kotz et al. (2001), which takes advantage of representation 4.1.12. The algorithm is as follows:

Generator for $\mathbf{Y} \sim \mathscr{AL}_d(\theta, \mathbf{m}, \Sigma)$:

1. Generate an exponentially distributed random variable W with mean 1.

2. Generate a multivariate Normal random variable $X \sim \mathscr{N}_d(\mathbf{0}, \Sigma)$.

3. Return $\mathbf{Y} = \theta + \mathbf{m}W + \sqrt{W}\mathbf{X}$.

4.5.2 Scenario 1: Fixed Location

With fixed location $\theta = (-5, 3)^\intercal$, we perform a simulation to investigate the bias and standard errors of the skewness parameter $\mathbf{m}$ and dispersion parameter Σ. Figures 4.2 and 4.3 show a large difference between the E-M estimator and the MM estimator when sample sizes are small. This

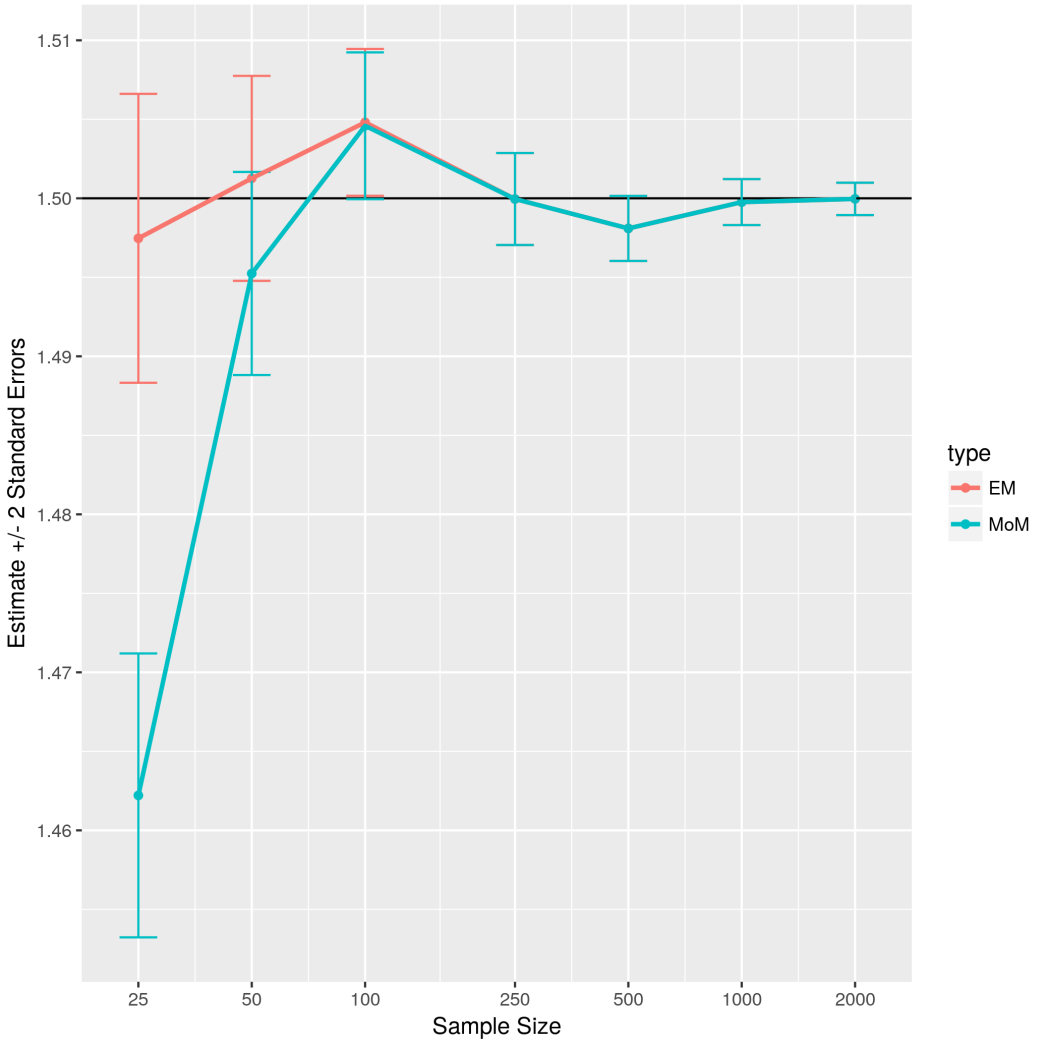

Figure 4.2 Bias of E-M and MM Estimates of m_1. The true value of m_1 is 1.5. The location parameter θ is known. Sample size varies from 25 to 2000.

disagreement is due to the removal of many method of moments samples that resulted in non positive definite covariance matrix estimates. As sample size increases, the proportion of admissible covariance estimates increases. We find that the E-M and MM estimators agree in these larger samples, and may have a negligible amount of bias for the parameter **m**.

Figures 4.4—4.6 show that parameter estimates of Σ are clearly biased. However the E-M estimates show significantly less bias and more precision than method of moments estimates. Both estimators tend to underestimate elements of Σ.

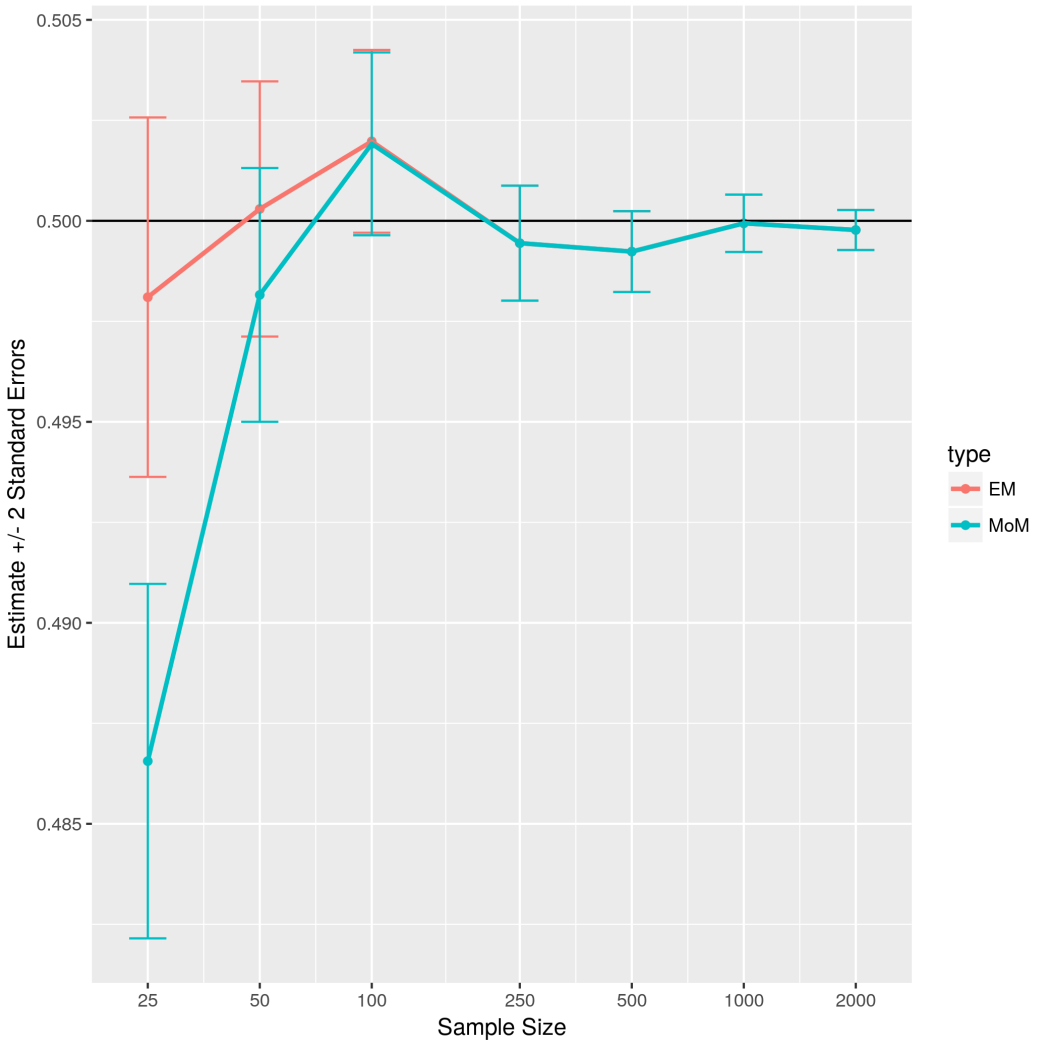

Figure 4.3 Bias of E-M and MM Estimates of m_2. The true value of m_2 is 0.5. The location parameter θ is known. Sample size varies from 25 to 2000.

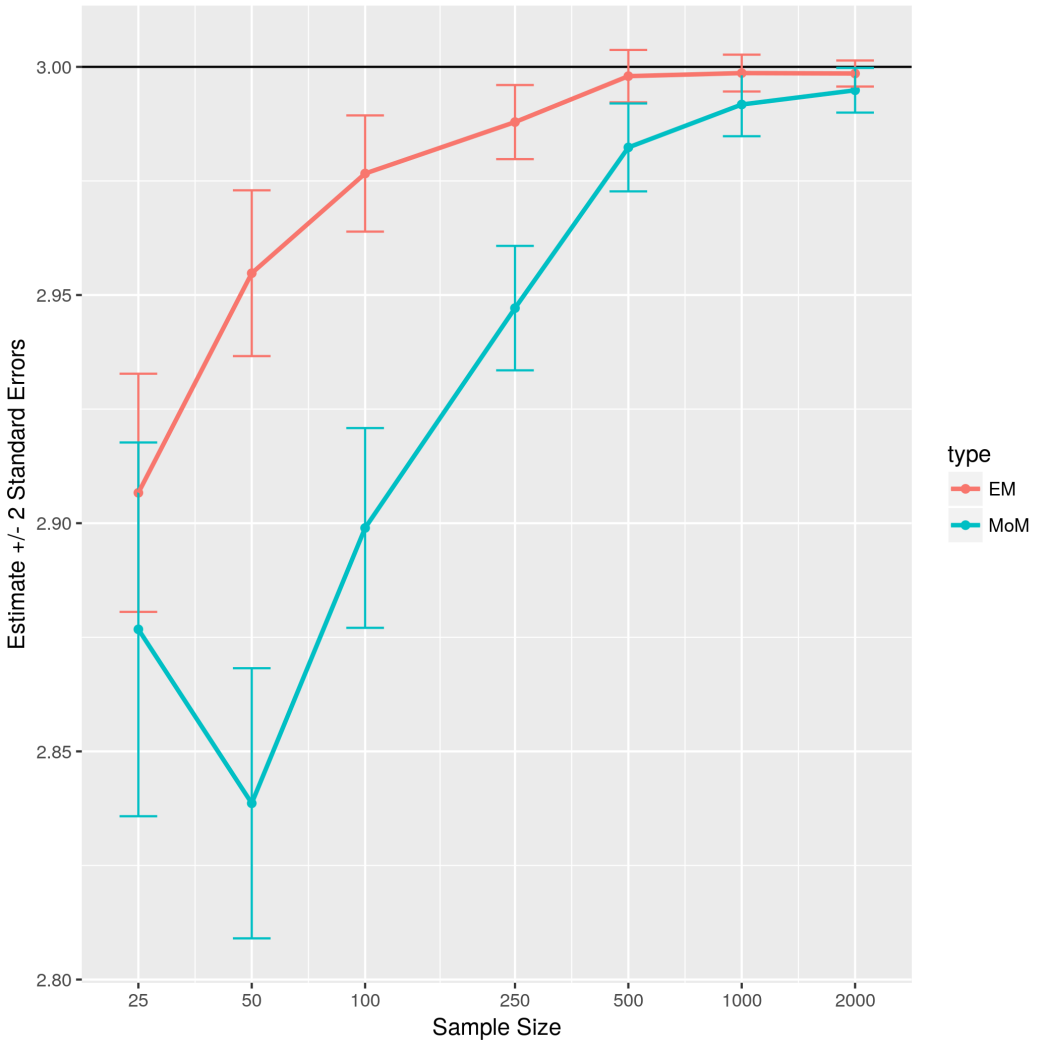

Figure 4.4 Bias of E-M and MM Estimates of Σ_{11}. The true value of Σ_{11} is 3. The location parameter θ is known. Sample size varies from 25 to 2000.

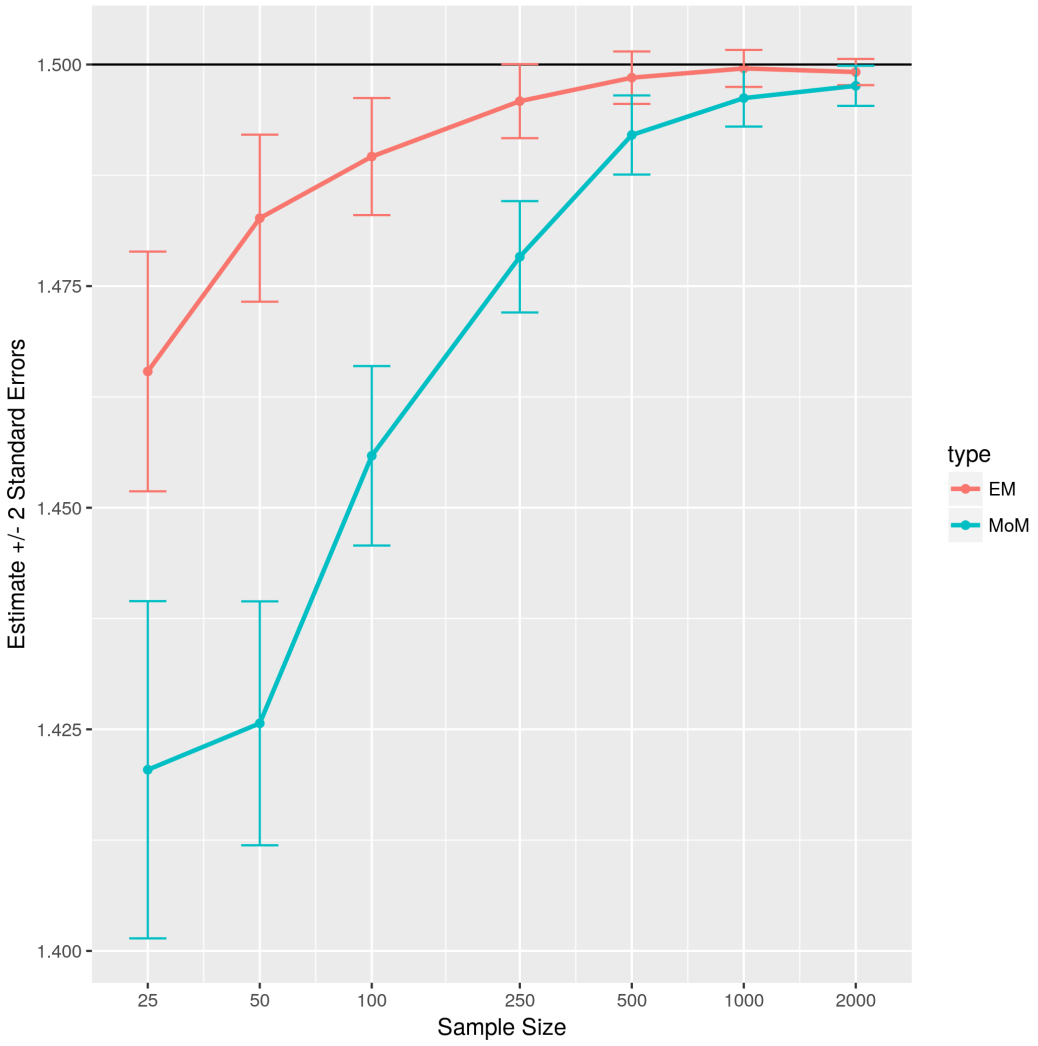

Figure 4.5 Bias of E-M and MM Estimates of Σ_{12}. The true value of Σ_{12} is 1.5. The location parameter θ is known. Sample size varies from 25 to 2000.

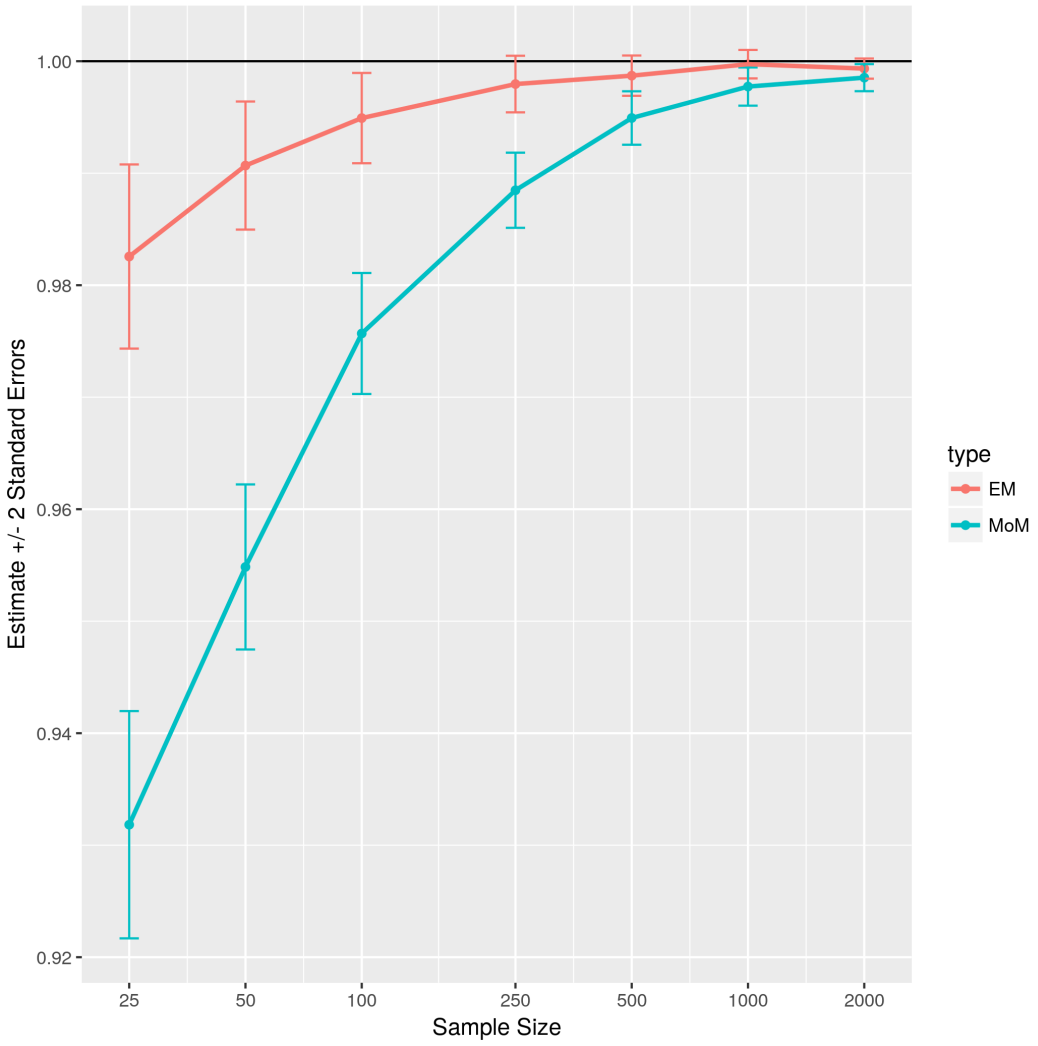

Figure 4.6 Bias of E-M and MM Estimates of Σ_{22}. The true value of Σ_{22} is 1. The location parameter θ is known. Sample size varies from 25 to 2000.

Scatter plots are created to show the MM estimates (vertical axis) and E-M estimates (horizontal axis) at a more granular level. The scatter plots only show the estimates derived when sample size is set at 2000 observations in the simulation. A red line is drawn to show where moment and E-M estimators agree, and a red dot is added at the true value of the population parameter. Only cases with admissible E-M and moment estimators are shown are plotted.

In Figures 4.7 and 4.8, we find that point estimates of the skewness parameter lie on the line indicating perfect agreement. This observation shows that E-M estimates of skewness converge to the sample average. However, in our E-M algorithm, we initialize skewness estimates at the sample average.

Contour lines from a two-dimensional kernel density estimator are added to the scatter plots of Σ estimates to show the concentration of estimates. From this one can see that many of the estimates are concentrated near the true parameter value. In this fixed parameters case, the contour lines form approximate ellipses with major axes aligned with the 45° line.

Figures 4.9 - 4.11 show scatter plots of estimates of elements of the dispersion matrix Σ. Generally there is more variability in the estimates that overestimate the parameter value, and much of that variability is spread in the vertical direction rather than the horizontal direction. This indicates that moment estimates may overestimate the dispersion parameters compared in the E-M estimates in some extreme cases.

4.5.3 Scenario 2: Unknown Location

The second part of the simulation study examines the behavior of our parameter estimation method under unknown parameters. As in Section 4.5.2, we generate $10,000$ replicates from samples of size, 25, 50, 100, 250, 500, 1000, and 2000. Because the location parameter θ is assumed unknown, both estimation procedures change to accommodate this lack of information. The MM procedure employed is as in Visk (2009) and detailed in Section 4.2.1. The E-M algorithm applied was detailed in Section 4.4.1.

Figures 4.12 - 4.18 show the bias of each estimation method with the true parameter denoted by a solid black line. Samples resulting with inadmissible estimates were discarded. We find that

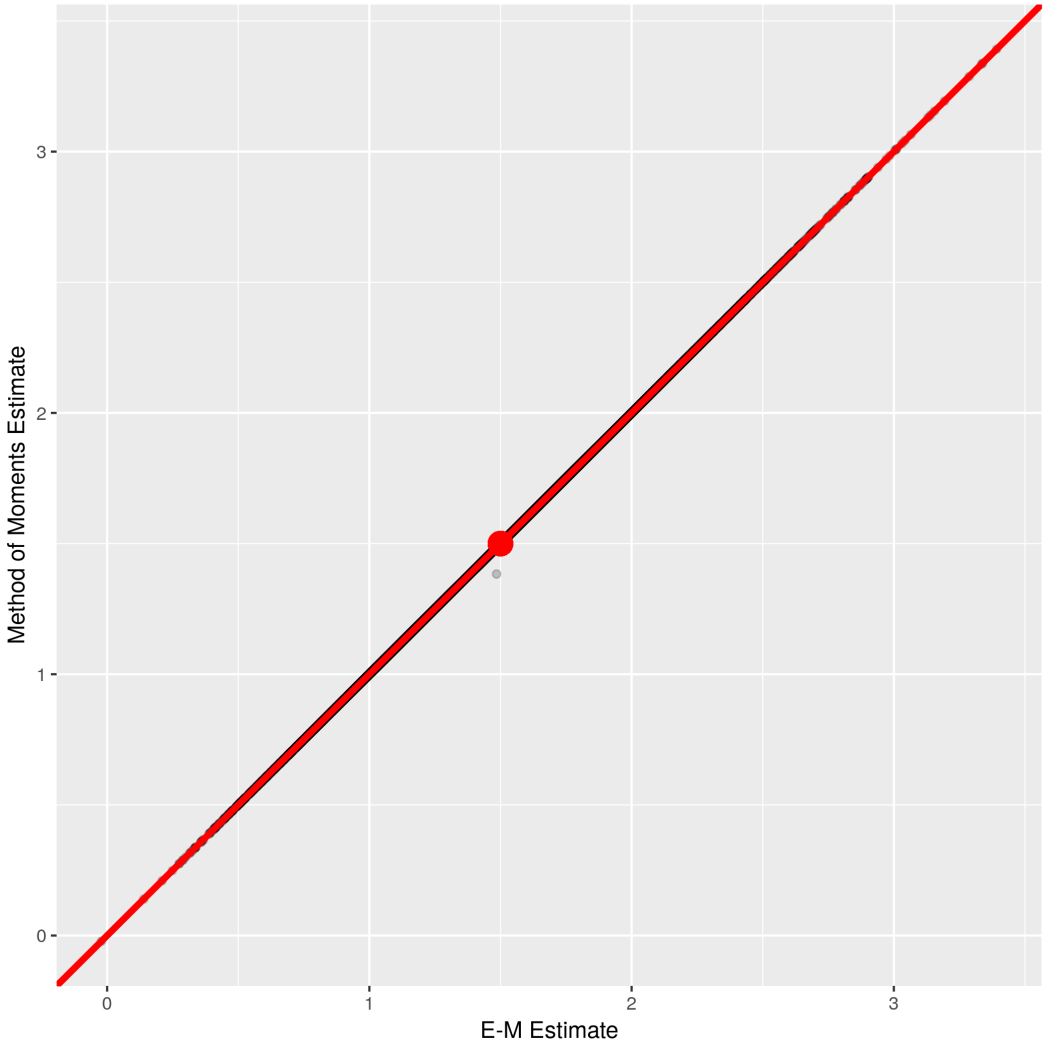

Figure 4.7 Scatter plot of Estimates of m_1. Sample size is 2000. The true value of m_1 is 1.5.

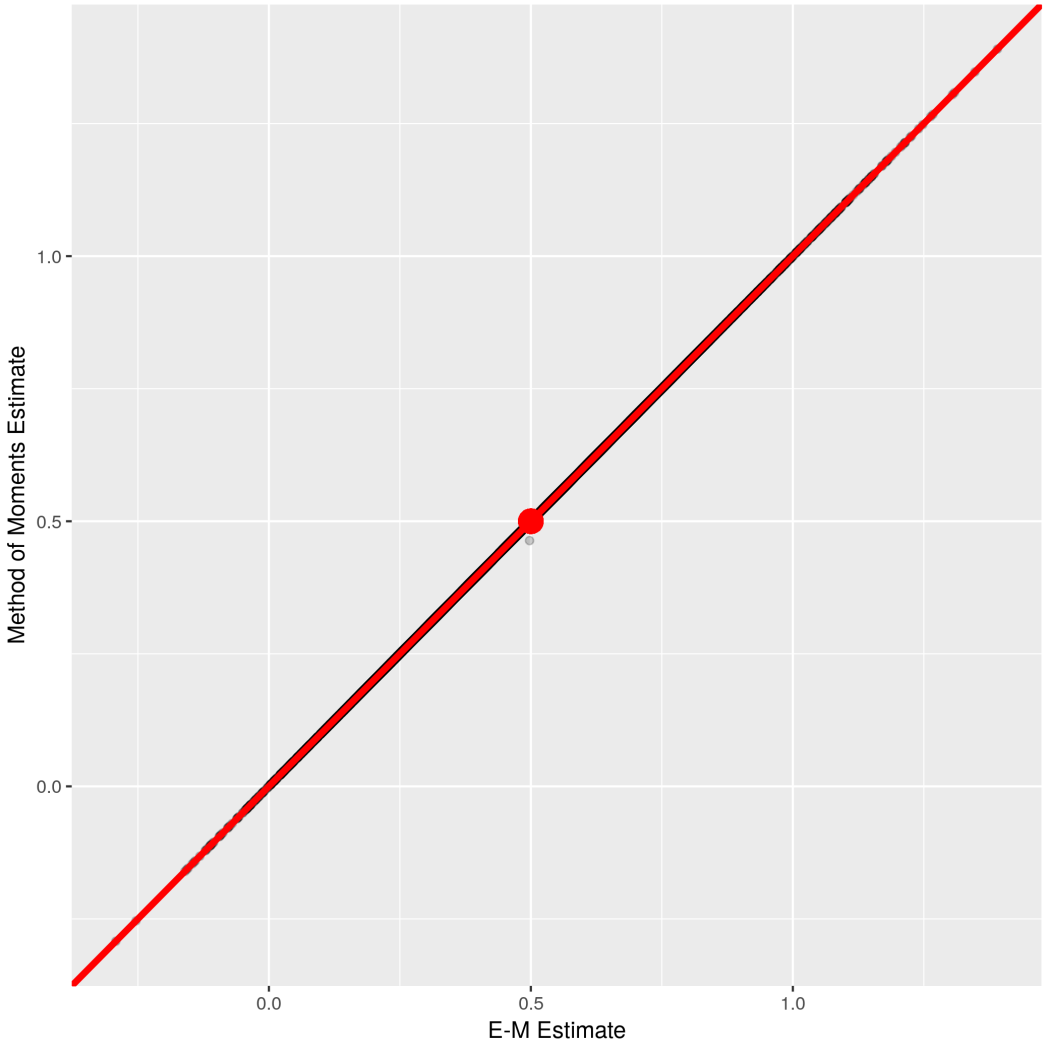

Figure 4.8 Scatter plot of Estimates of m_2. Sample size is 2000. The true value of m_2 is 0.5.

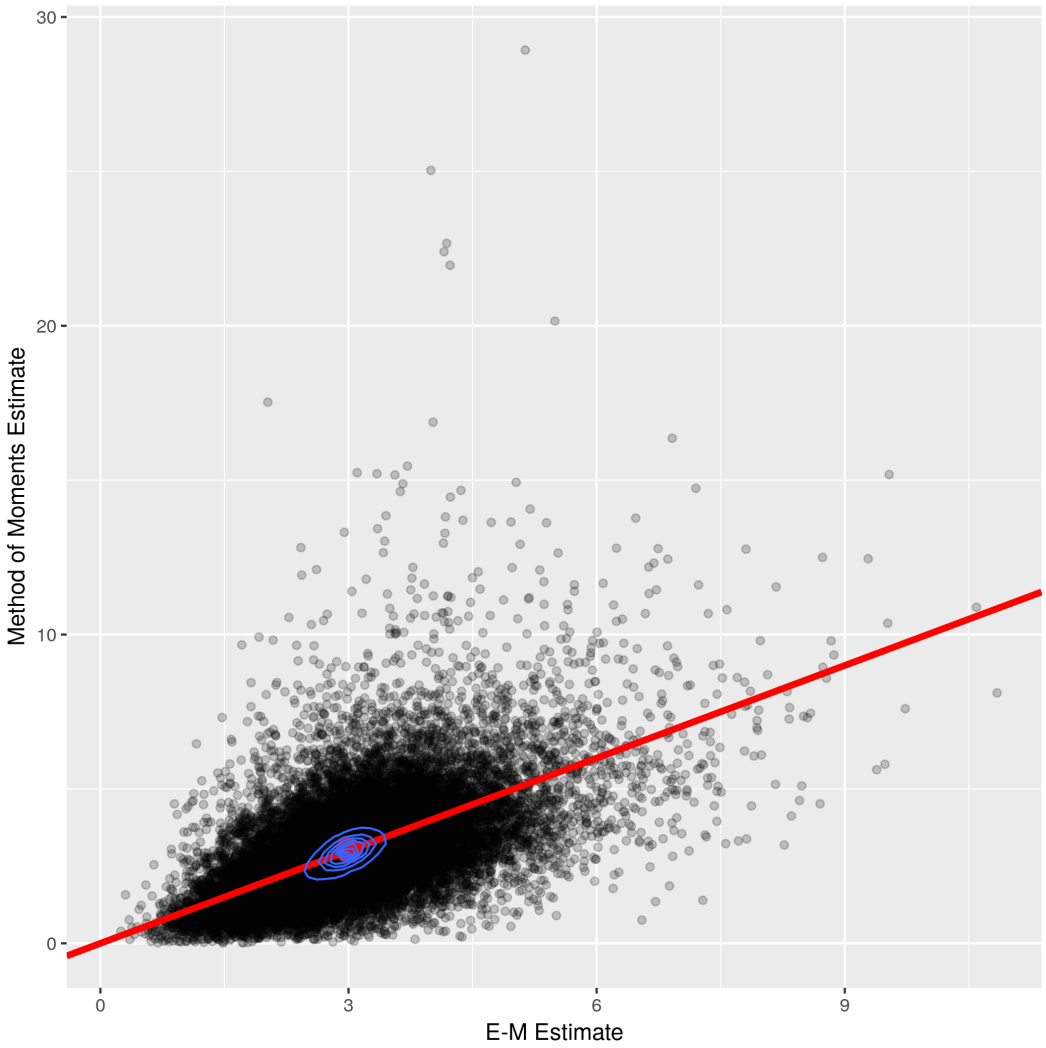

Figure 4.9 Scatter plot of Estimates of Σ_{11} with known location parameter. Sample size is 2000. The true value of Σ_{11} is 3.

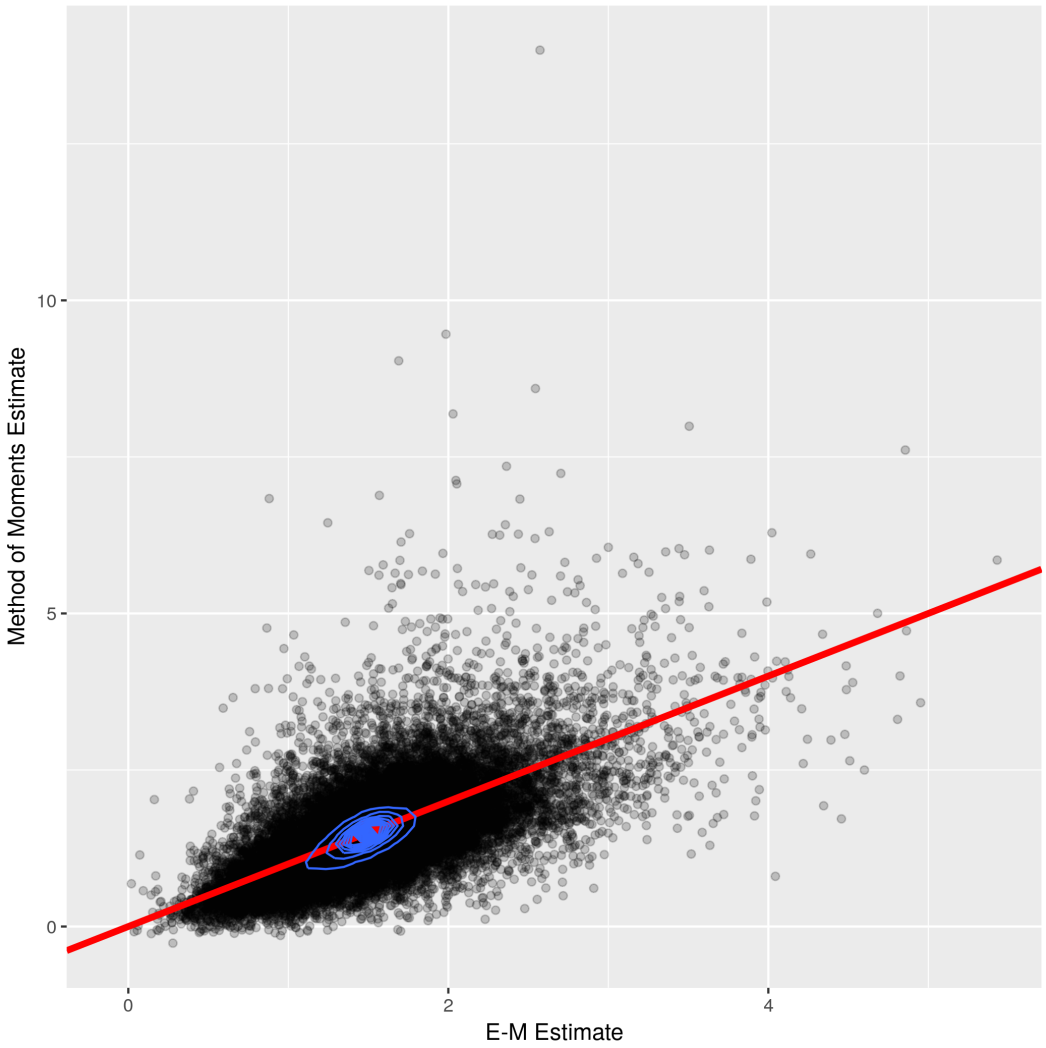

Figure 4.10 Scatter plot of Estimates of Σ_{12} with known location parameter. Sample size is 2000. The true value of Σ_{12} is 1.5.

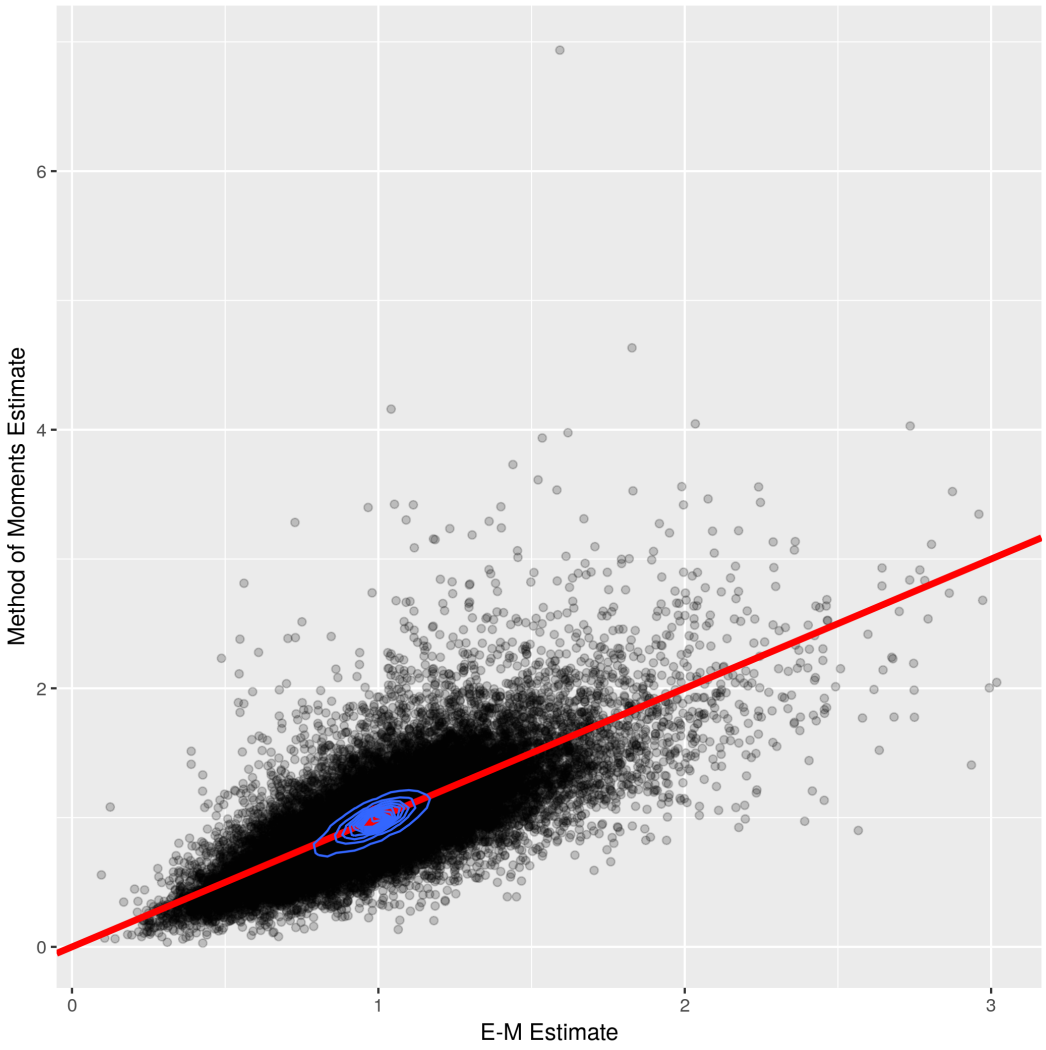

Figure 4.11 Scatter plot of Estimates of Σ_{22} with known location parameter. Sample size is 2000. The true value of Σ_{22} is 1.

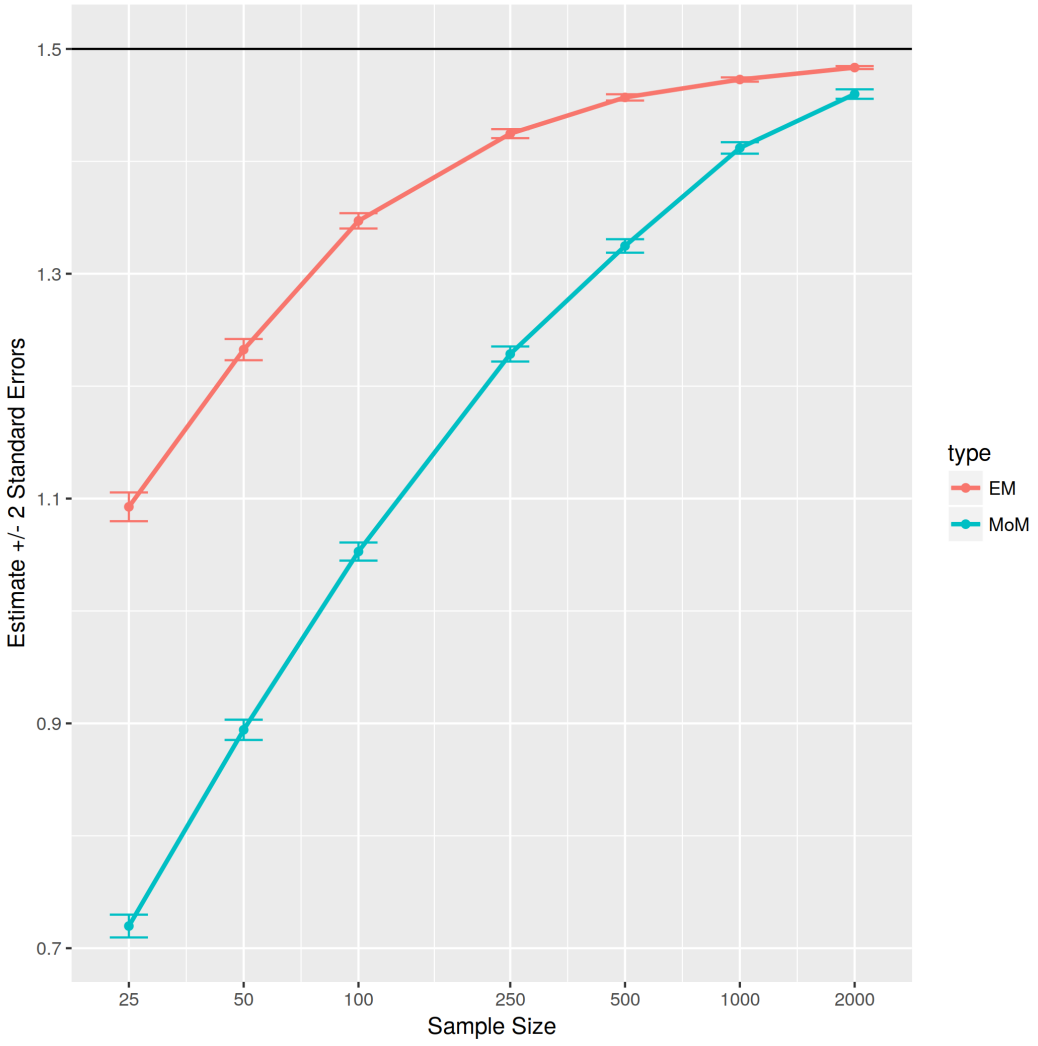

Figure 4.12 Bias of E-M and MM Estimates of m_1 with unknown location parameter. Sample size varies from 25 to 2000.

the standard errors of E-M estimates are generally smaller for each parameter, and that bias is decreased by a significant amount. An exception is found in the estimates of Σ_{22}, shown in Figure 4.18.

In samples with less than 100 observations, the E-M method outperforms the moment method by a wide margin. When sample size exceeds 500 the moment method shows less bias. For moderate sized samples it is unclear which method yields a better estimate of this parameter.

Scatter plots are drawn to show a detailed view of point estimates in samples of size 2000. Contour lines are drawn atop the scatter plots to show the concentration of the point estimates

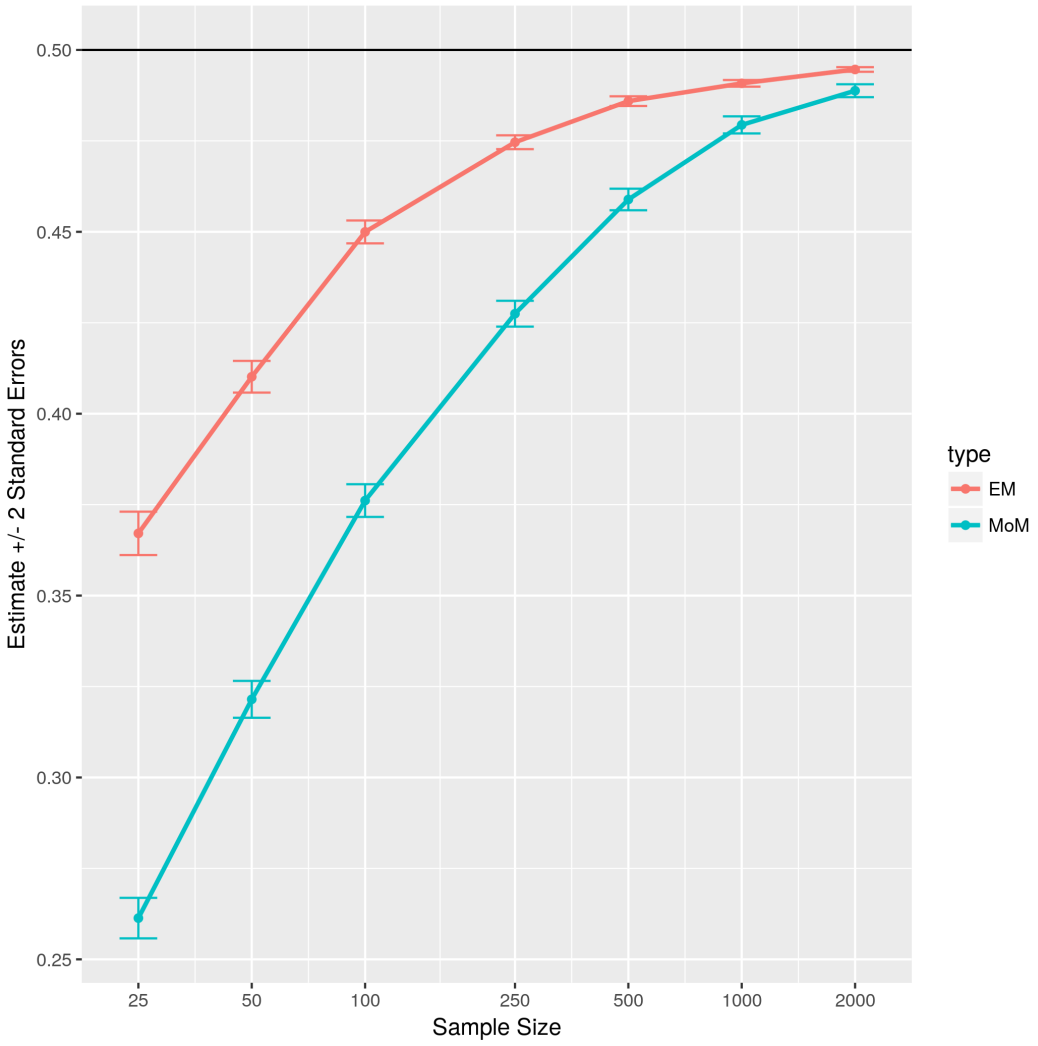

Figure 4.13 Bias of E-M and MM Estimates of m_2 with unknown location parameter. Sample size varies from 25 to 2000.

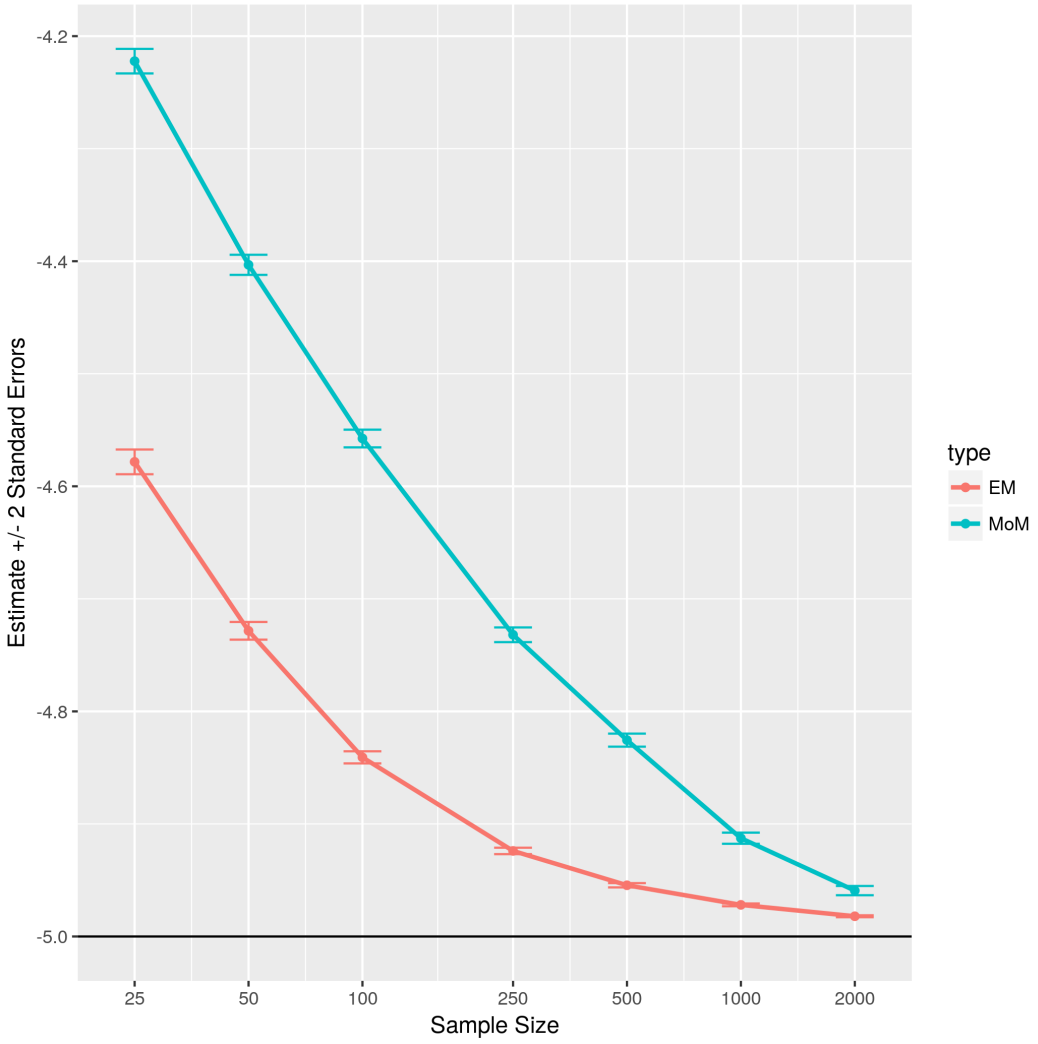

Figure 4.14 Bias of E-M and MM estimates of θ_1. Sample size varies from 25 to 2000.

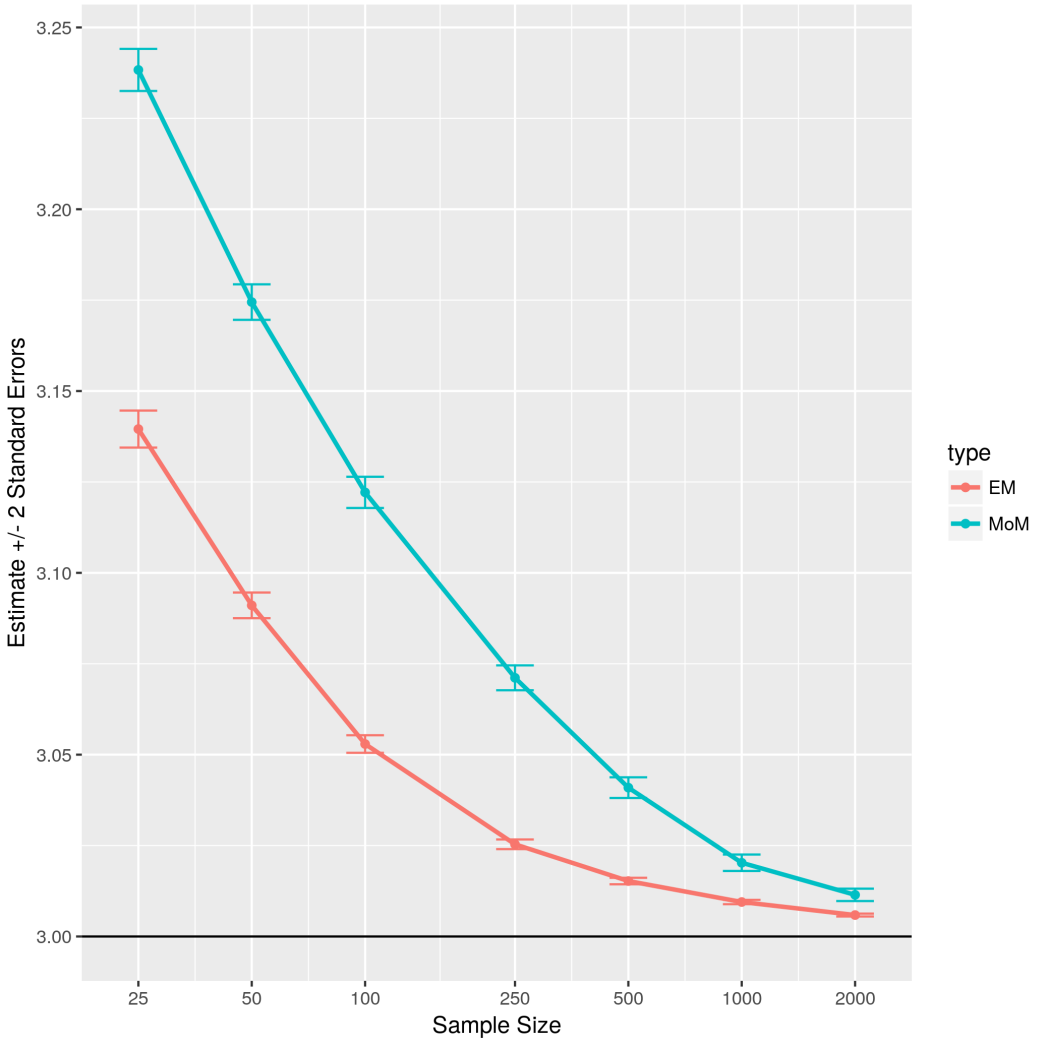

Figure 4.15 Bias of E-M and MM estimates of θ_2. Sample size varies from 25 to 2000.

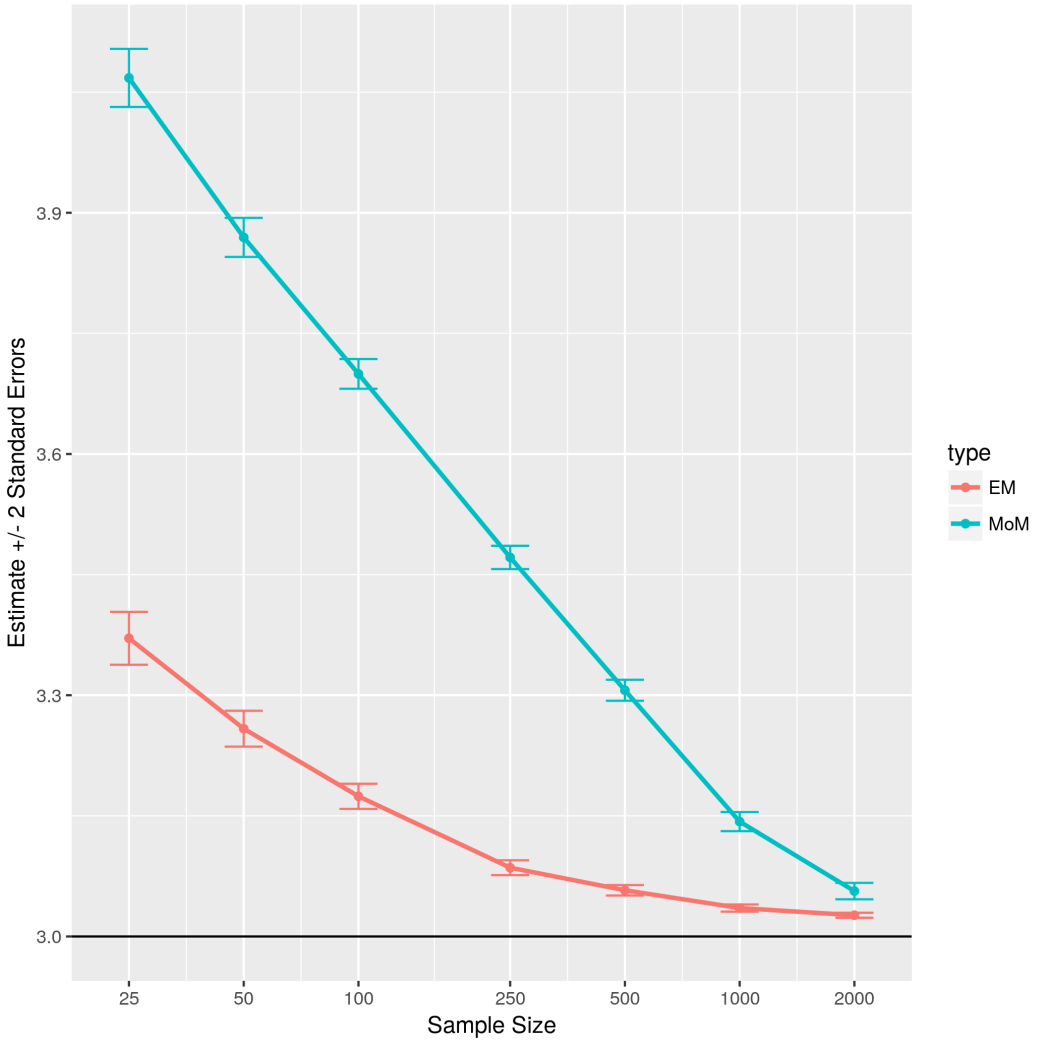

Figure 4.16 Bias of E-M and MM estimates of Σ_{11}. Sample size varies from 25 to 2000.

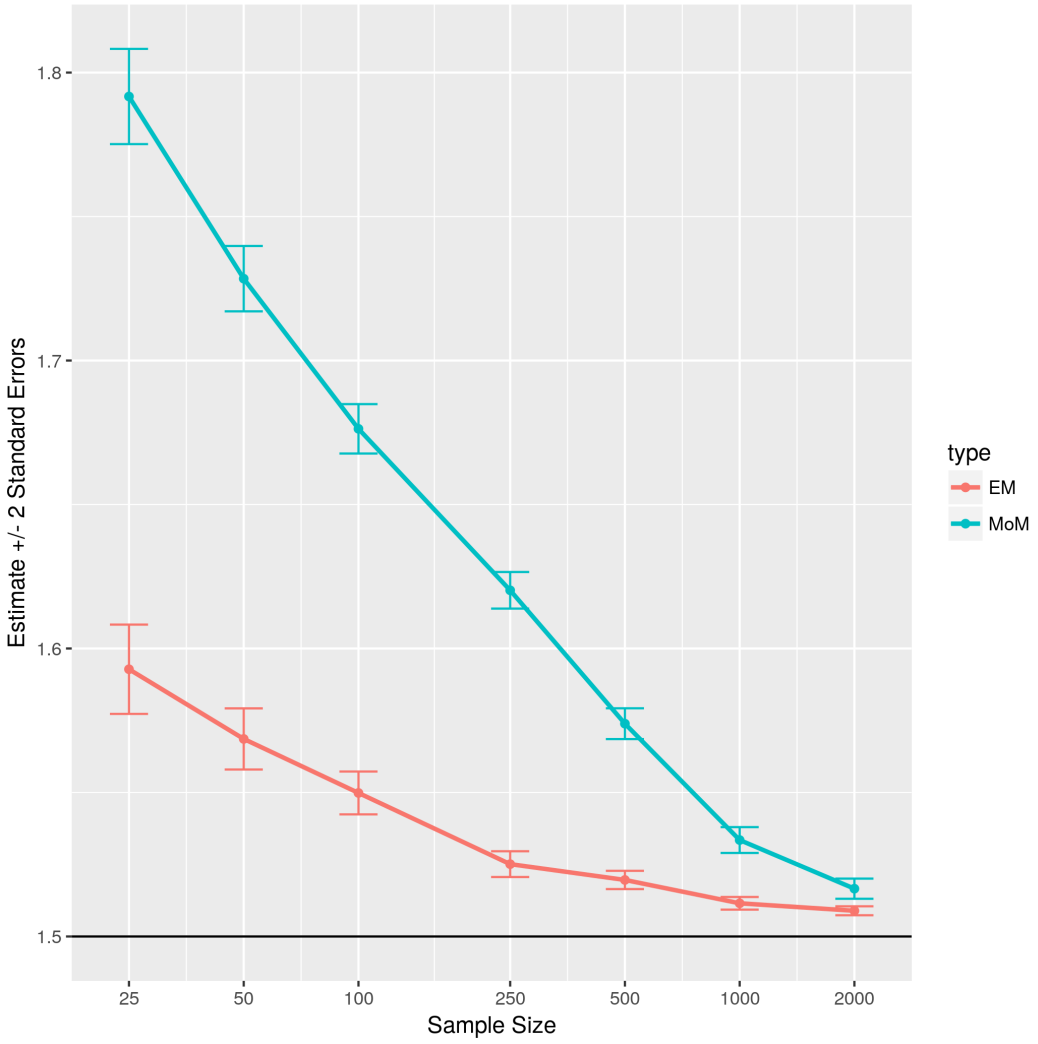

Figure 4.17 Bias of E-M and MM estimates of Σ_{12}. Sample size varies from 25 to 2000.

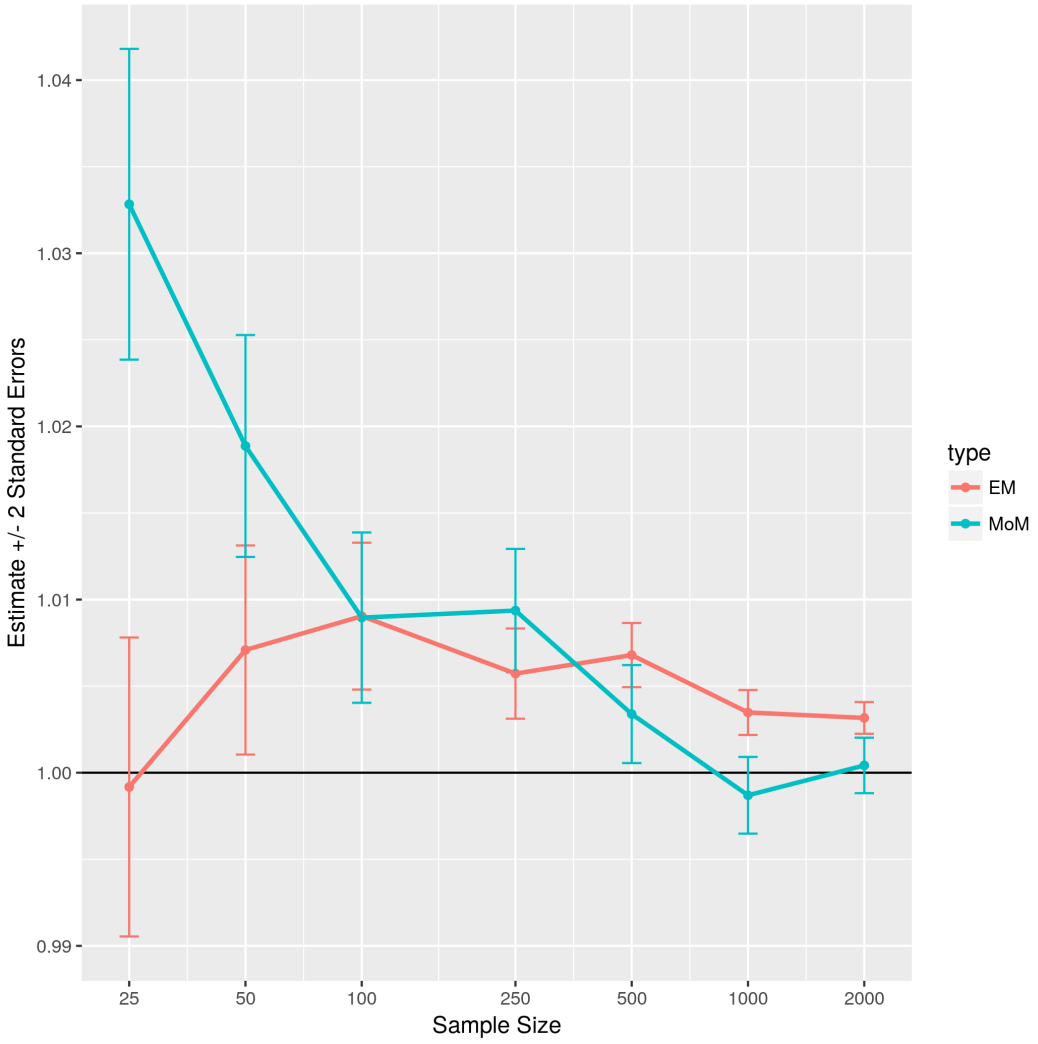

Figure 4.18 Bias of E-M and MM estimates of Σ_{22}. Sample size varies from 25 to 2000.

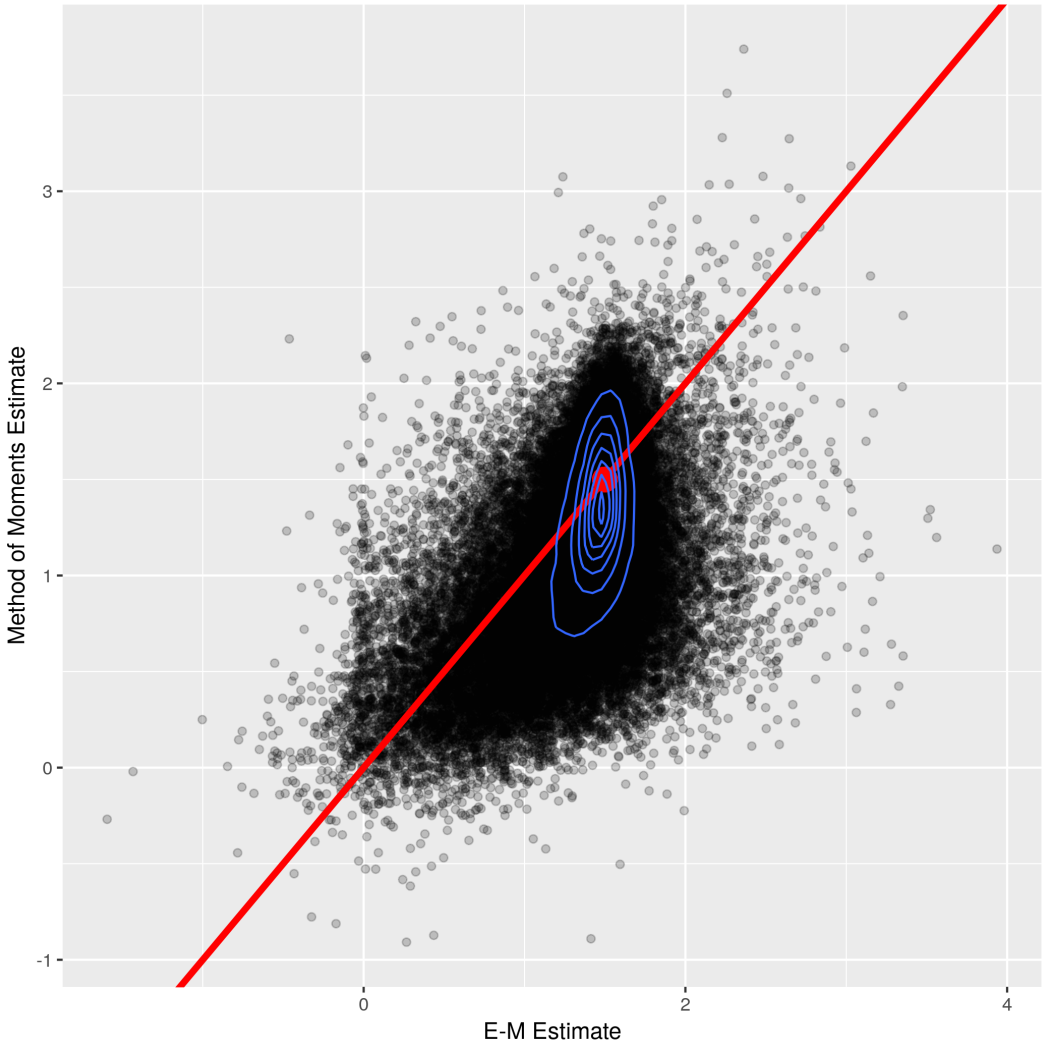

Figure 4.19 Scatter plot with non-parametric contours of estimates of m_1. Sample size is 2000 and location parameter θ is unknown.

from the $10,000$ replicates. Figures 4.19 - 4.22 show scatter plots for estimates of θ_1, θ_2, m_1, and m_2. The contour lines on these scatter plots show that most of the variation is concentrated in the vertical direction.

Contours drawn on the estimates of dispersion parameters show a different trend – a group of moment estimates tends to fall below the parameter estimate. This shape can be seen in the scatter plots of each dispersion parameter. This group of estimates that breaks from the elliptical trend that we would expect in the contour lines is due to moment estimates that trend toward inadmissibility; that is, non-positive definite Σ.

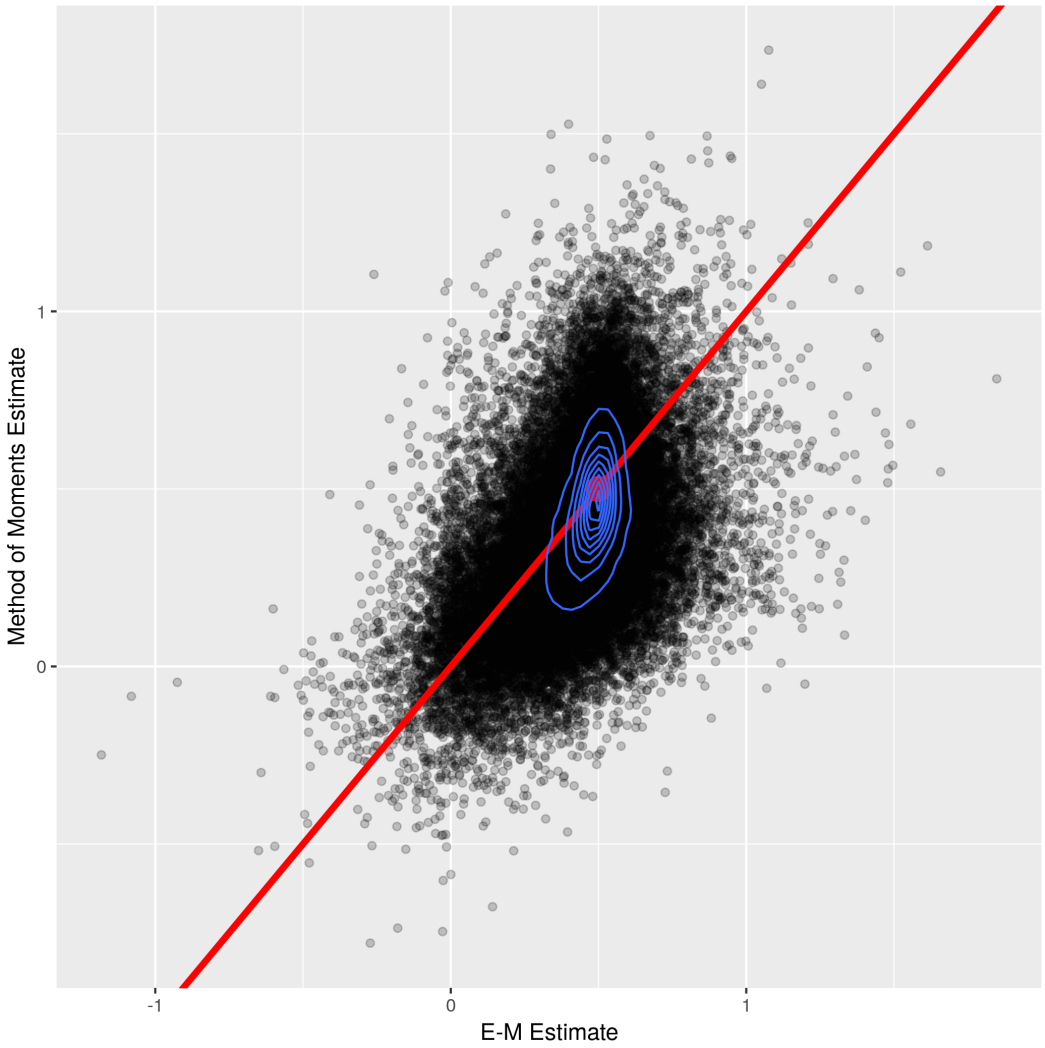

Figure 4.20 Scatter plot with non-parametric contours of estimates of m_2. Sample size is 2000 and location parameter θ is unknown.

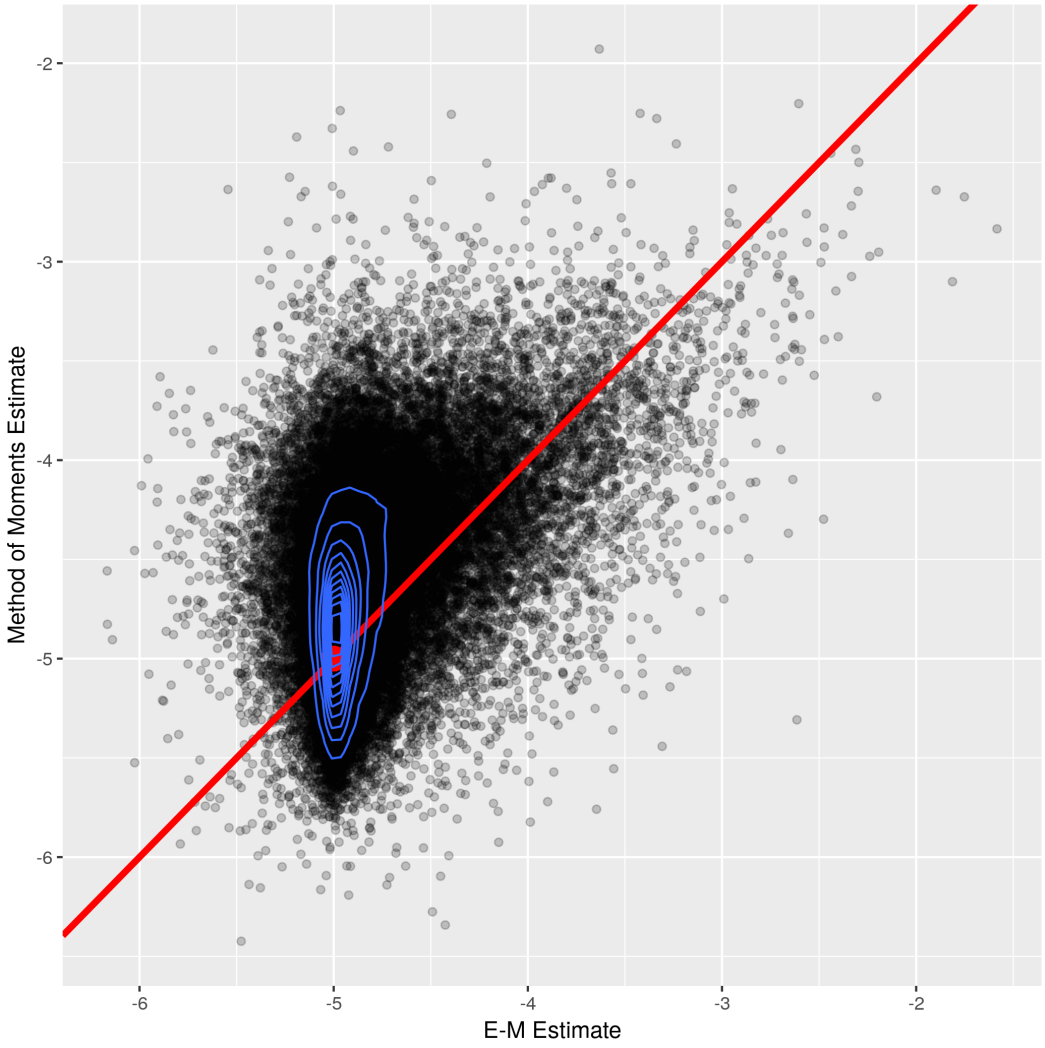

Figure 4.21 Scatter plot with non-parametric contours of estimates of θ_1. Sample size = 2000.

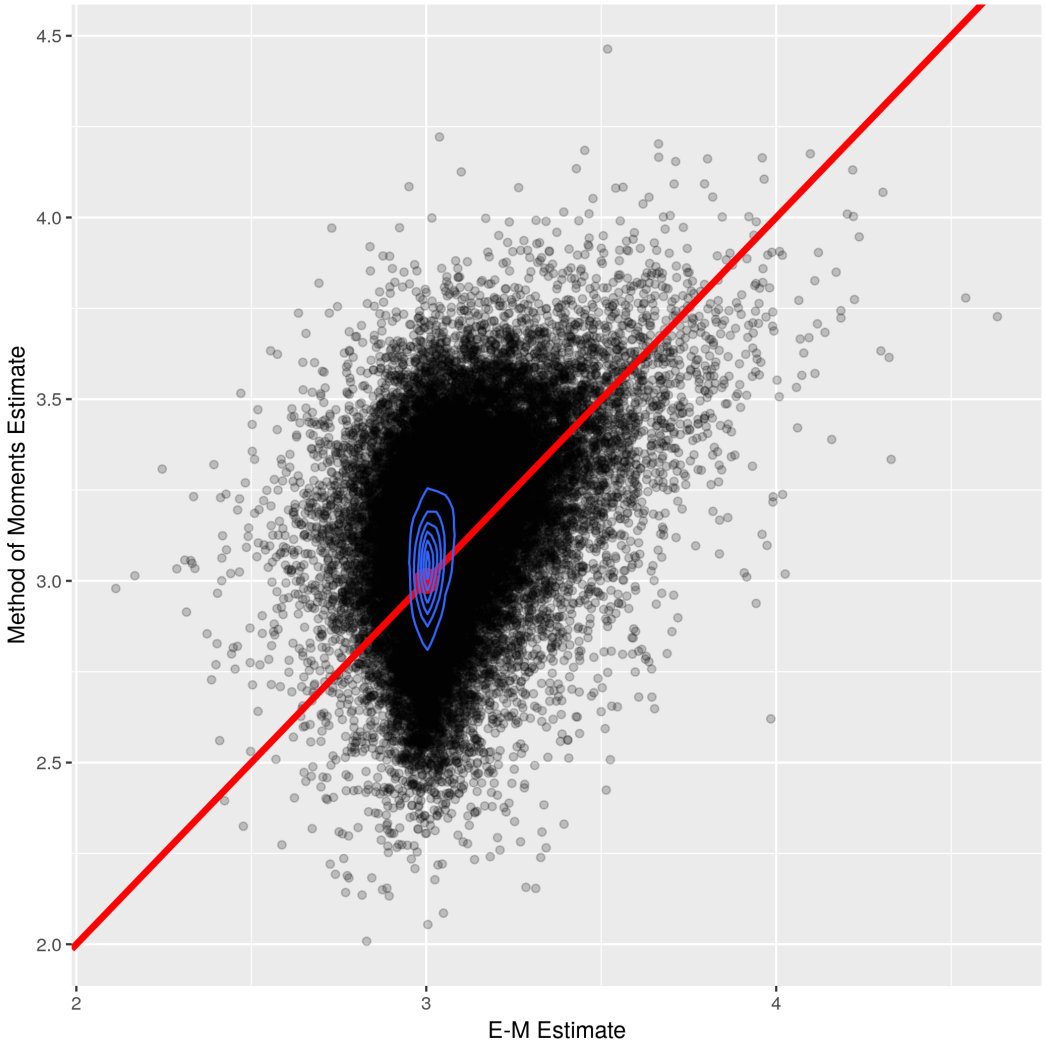

Figure 4.22 Scatter plot with non-parametric contours of estimates of θ_2. Sample size is 2000.

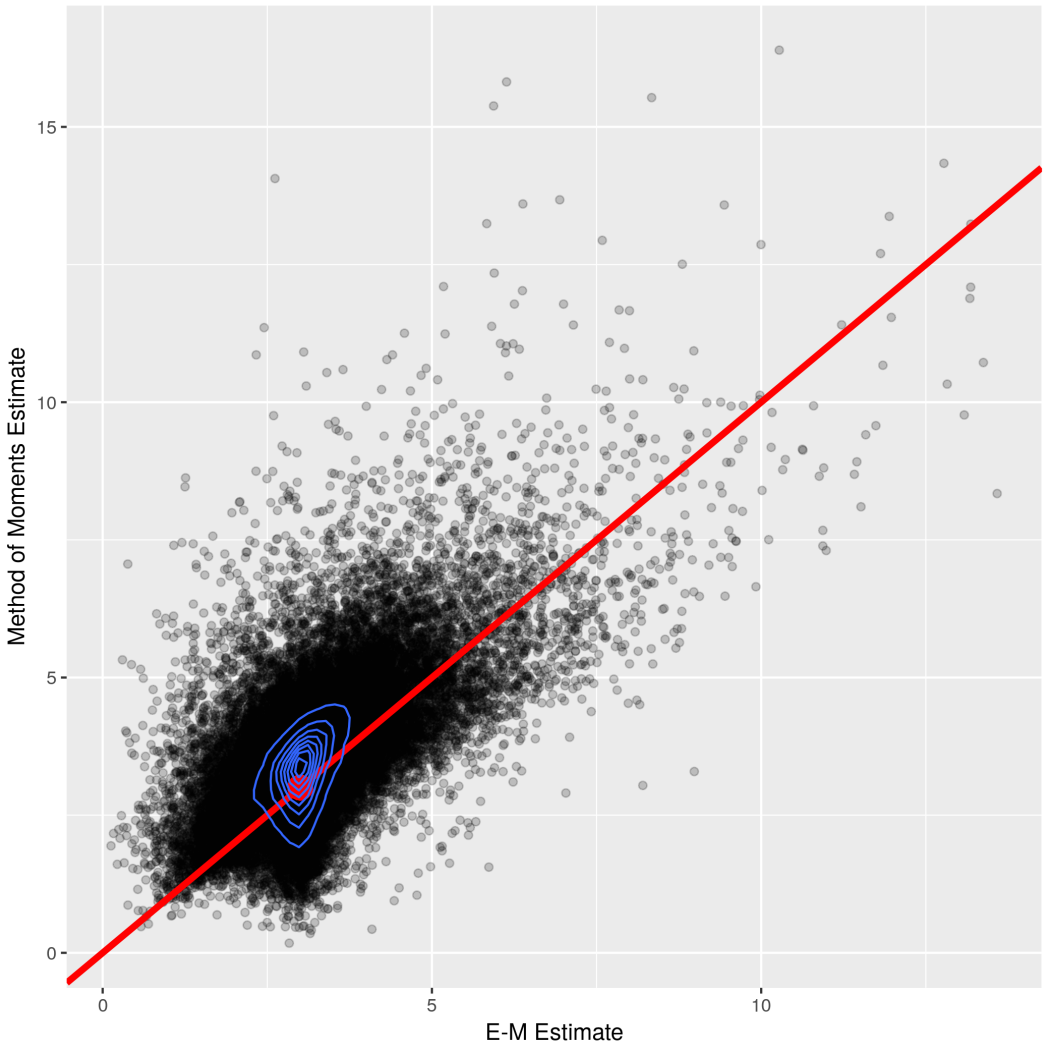

Figure 4.23 Scatter plot with non-parametric contours of estimates of Σ_{11}. Sample size is 2000 and location parameter θ is unknown.

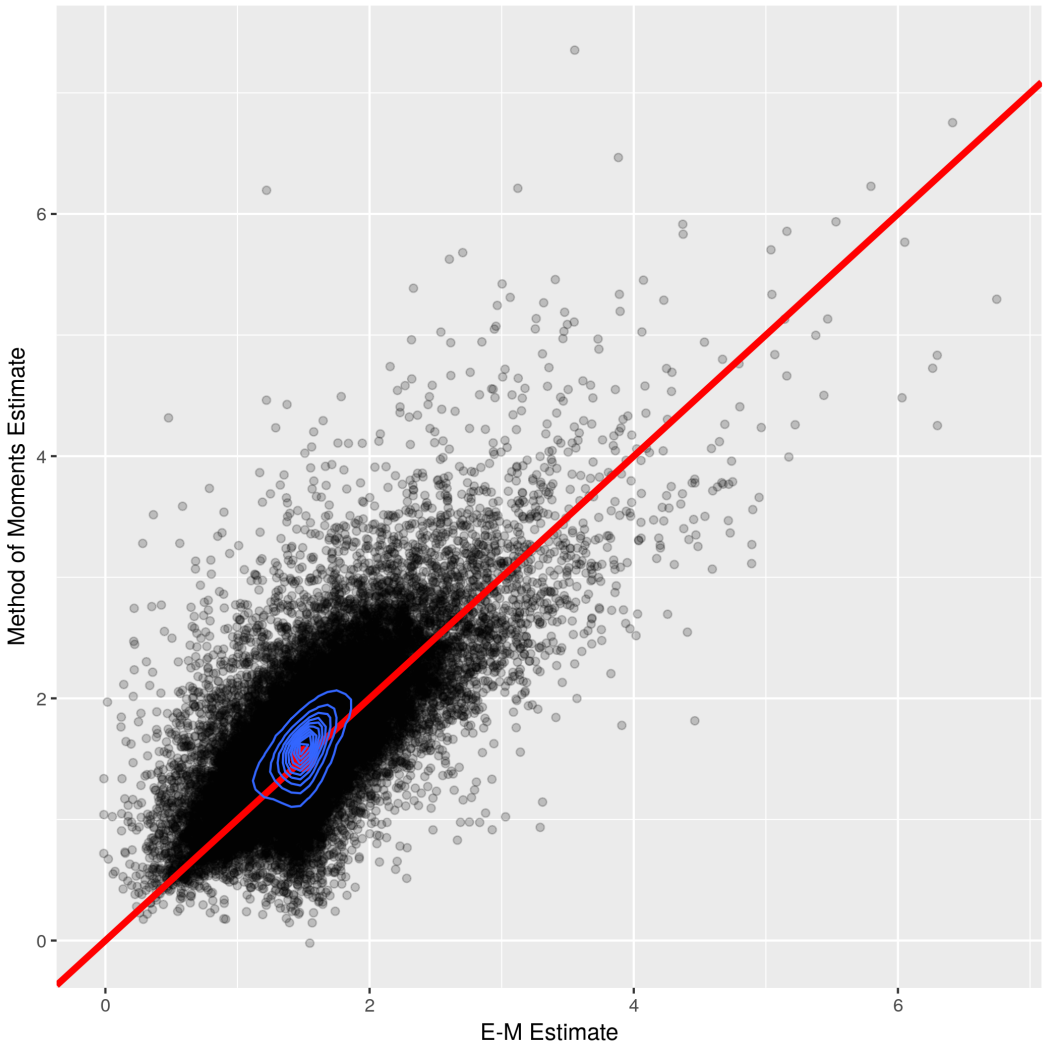

Figure 4.24 Scatter plot with non-parametric contours of estimates of Σ_{12}. Sample size is 2000 and location parameter θ is unknown.

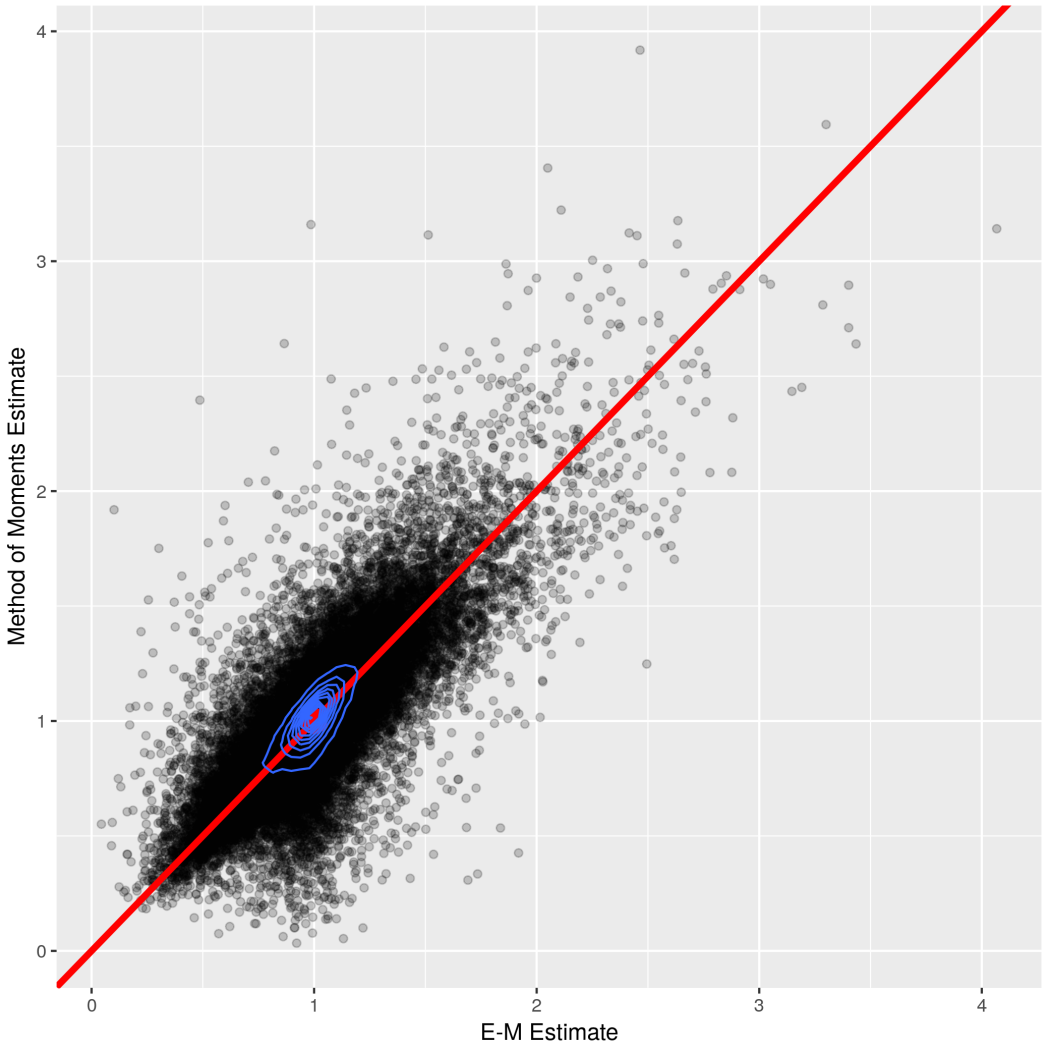

Figure 4.25 Scatter plot with non-parametric contours of estimates of Σ_{22}. Sample size is 2000 and location parameter θ is unknown.

4.6 Bias Corrected E-M Estimator

Following Section 4.4.2, a simulation study was performed to investigate the properties of E-M estimates after a bootstrap bias correction was performed. As in previous simulations, $10,000$ replications were performed of data generated from distribution (4.5.1). The sample sizes were taken to be $50, 100, 250, 500$ or $1,000$. The sample size 2000 taken in previous simulations was omitted due to time constraints at the Ohio Supercomputer Center. At each replication, E-M parameter estimates and bias corrected E-M parameter estimates were calculated. The number of bootstrap samples was taken to be 500 for all replications, though in practice a higher number e.g. 2000 is recommended.

The results of the procedure are shown in Figures 4.26 - 4.29. In this simulation, the location parameter θ and asymmetry parameter $\mathbf{m}$ were specifically investigated.

Generally we find a reduction in bias across the estimates of all parameters, however the bias correction method underestimates the magnitude of bias for each parameter. The result is that bias corrected estimates of parameters still admit a degree of bias, however we observe an over 50% decrease in bias for each parameter estimate at each sample size. Standard errors of the estimates on the other hand are larger due to the increased number of a parameters estimated.

4.7 Goodness-of-Fit Tests

We consider the hypothesis

$$H : \mathbf{Y_1}, \ldots, \mathbf{Y_n} \sim \mathscr{AL}_d(\theta, \Sigma, \mathbf{m}) \tag{4.7.1}$$

for some $\theta \in \mathbb{R}^d$, a positive definite $\Sigma \in \mathbb{R}^{d \times d}$, and $\mathbf{m} \in \mathbb{R}^d$. A test of this hypothesis has been considered in Fragiadakis and Meintanis (2011) as an extension of their test of symmetric Laplacity. Similar to the test of univariate asymmetric Laplacity (2.2), the test of the multivariate hypothesis can be constructed by observing that under the null hypothesis, the CF (4.1.1), when multiplied by its denominator, should be close (in some sense) to 1. We reduce the problem by

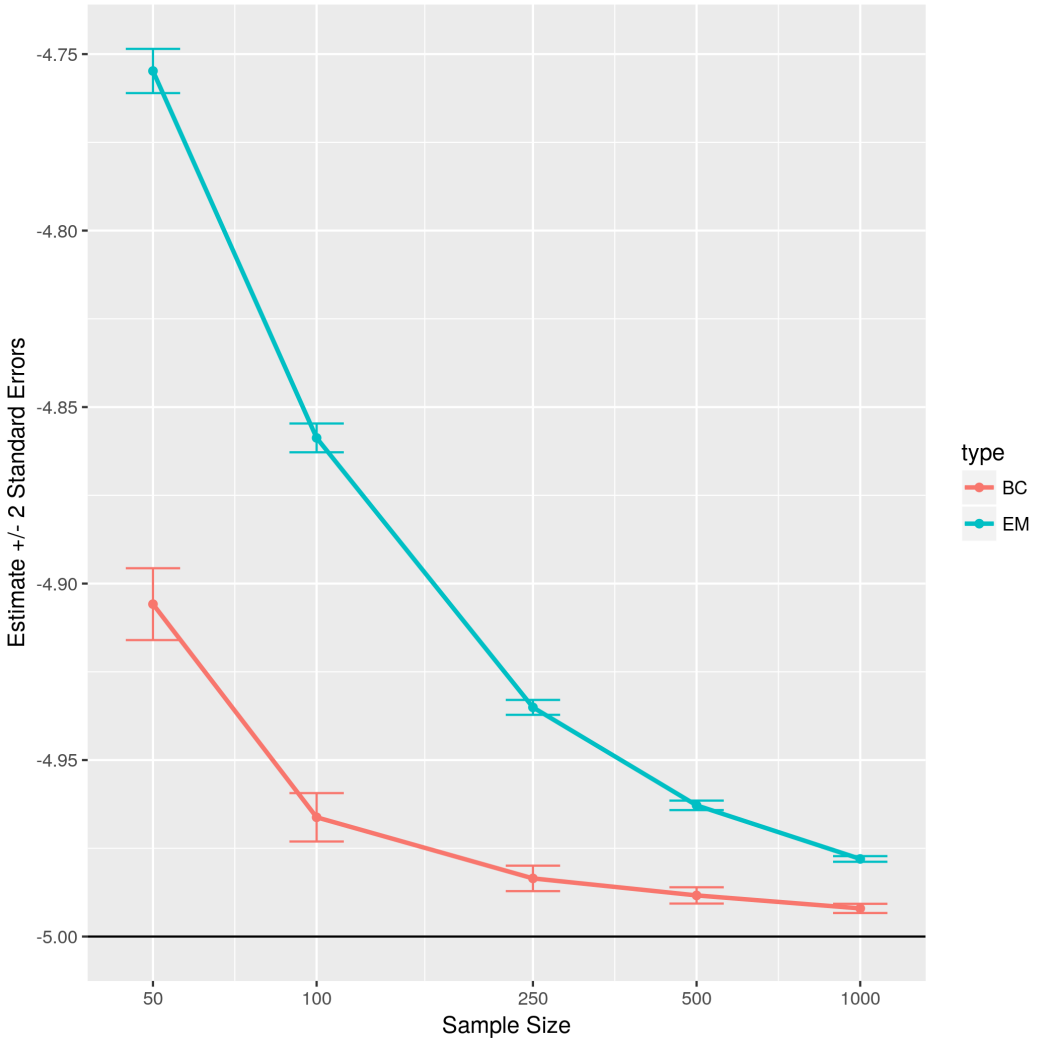

Figure 4.26 Bias of bias corrected E-M and E-M estimates of θ_1. Sample size varies from 50 to 1000.

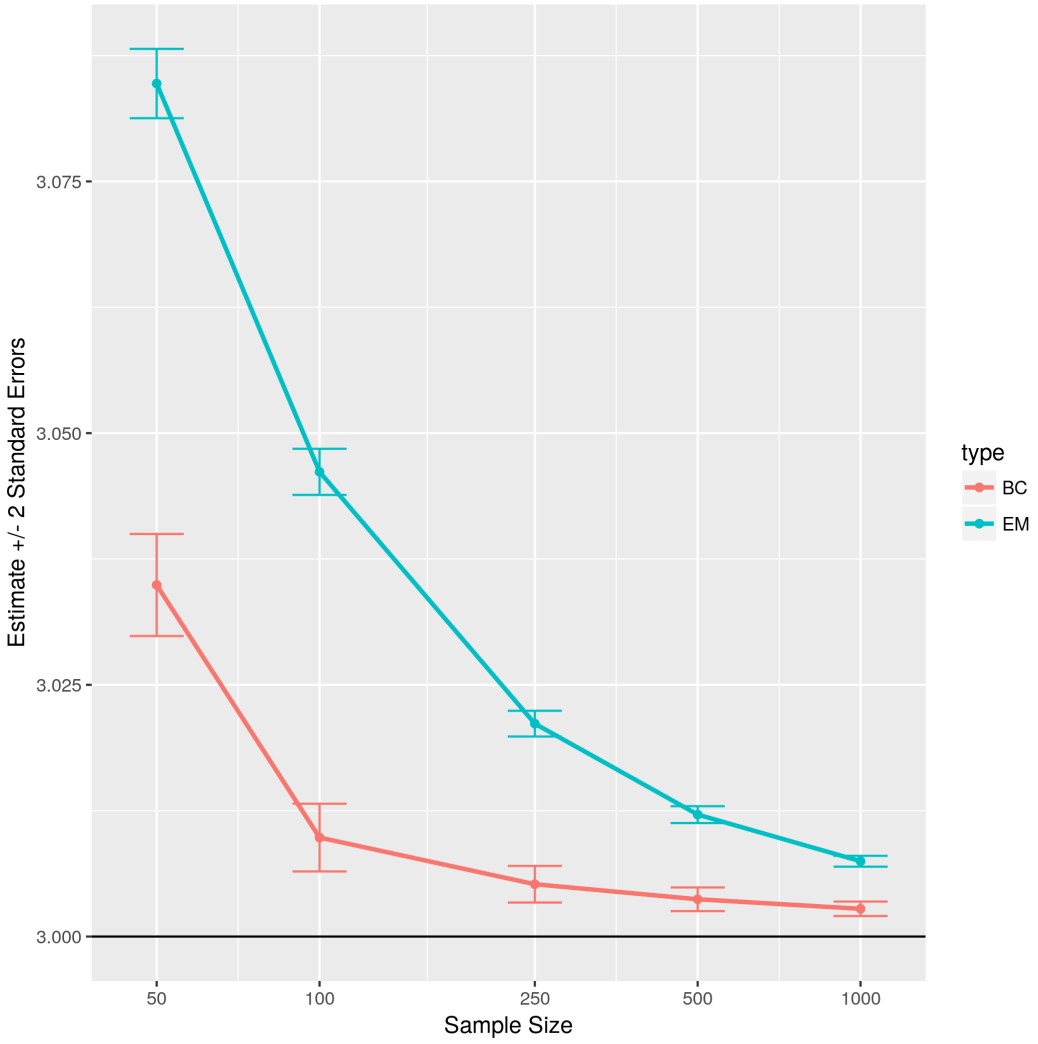

Figure 4.27 Bias of bias corrected E-M and E-M estimates of θ_2. Sample size varies from 50 to 1000.

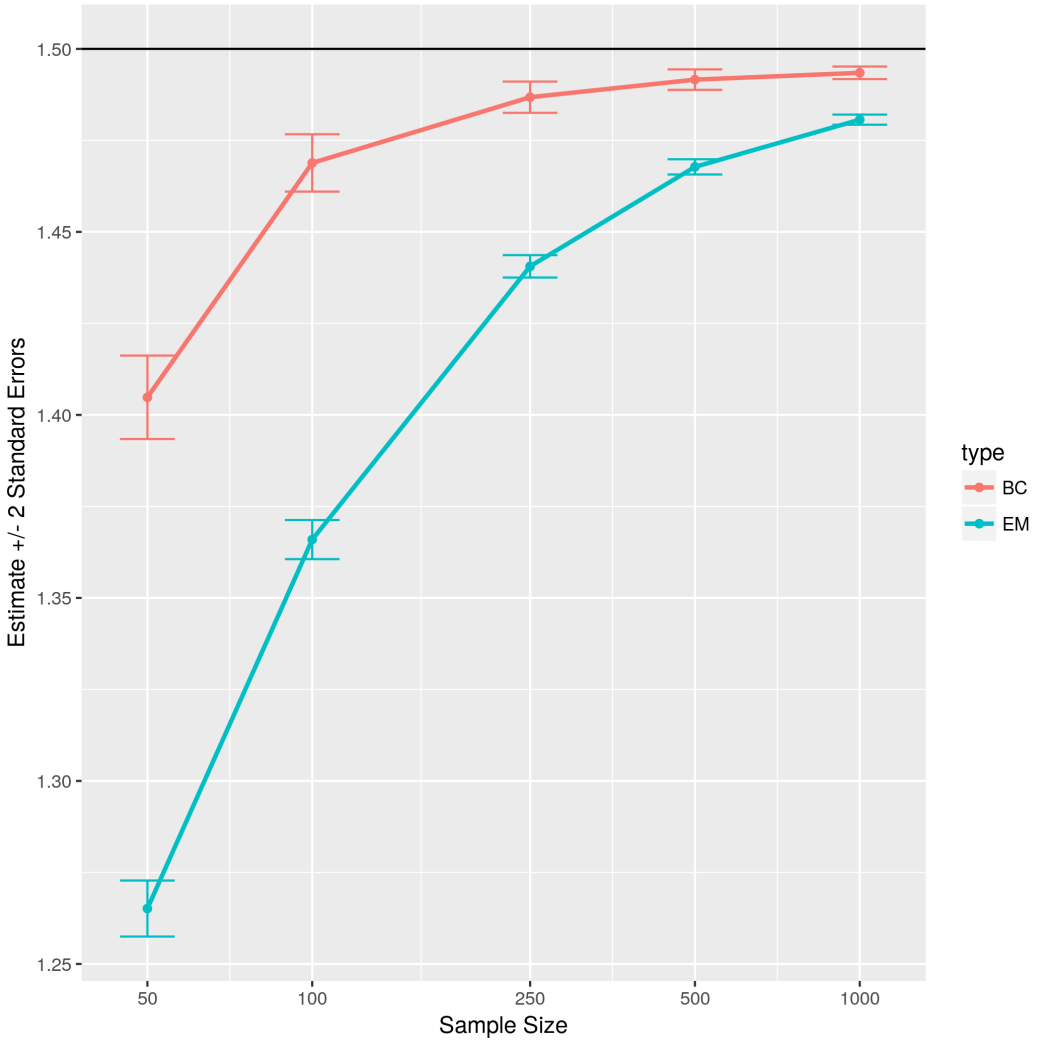

Figure 4.28 Bias of bias corrected and E-M estimates of m_1. Sample size varies from 50 to 1000.

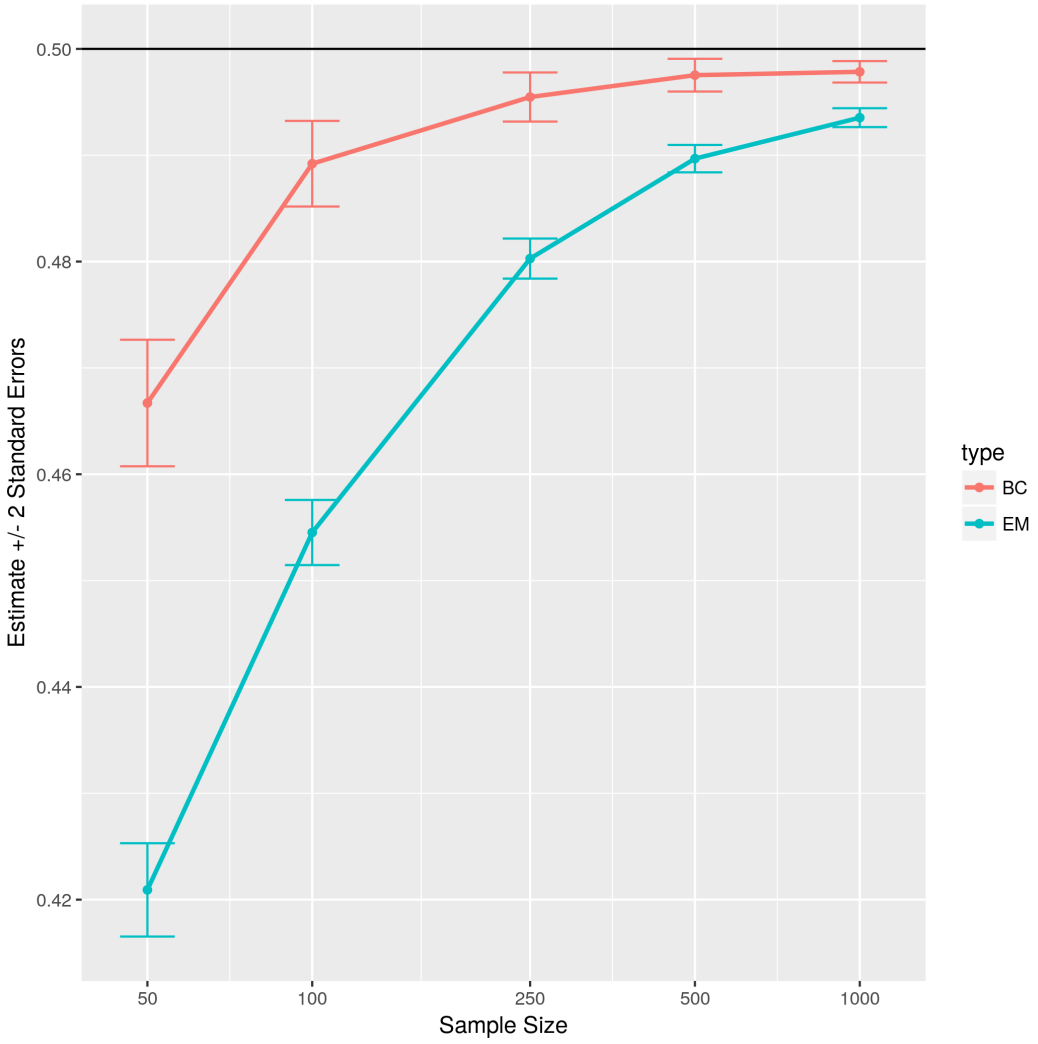

Figure 4.29 Bias of bias-corrected E-M and E-M estimates of m_2. Sample size varies from 50 to 1000.

considering the case where $\theta = \mathbf{0}$ and $\Sigma = \mathbf{I}_d$. Under the null,

$$\phi_{\mathbf{Y}}(\mathbf{t})\left(1 - i\hat{\mathbf{m}}^{\intercal}\mathbf{t} + \frac{1}{2}\mathbf{t}^{\intercal}\mathbf{t}\right) - 1 = 0. \tag{4.7.2}$$

Fragiadakis and Meintanis (2011) suggest the use of the test statistic

$$T_{n,W} = n\int_{\mathbb{R}^d}\left|\hat{\phi}_{\mathbf{Y}}(\mathbf{t})\left(1 - i\hat{\mathbf{m}}^{\intercal}\mathbf{t} + \frac{1}{2}\mathbf{t}^{\intercal}\mathbf{t}\right) - 1\right|^2 W(\mathbf{t})\, d\mathbf{t} \tag{4.7.3}$$

where $\hat{\phi}_{\mathbf{Y}}(\mathbf{t})$ is the multivariate empirical CF $\hat{\phi}_{\mathbf{Y}}(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n}\exp(i\mathbf{t}^{\intercal}\mathbf{y}_i)$ and W is a suitable weight function. For computational purposes, one may take $W(\mathbf{t}) = \exp(-a||\mathbf{t}||^2)$. In this composite test, parameters are estimated through MM.

4.8 Multivariate Asymmetric Laplace Energy Test

The energy test of the hypothesis (4.7.1) makes use of the statistic (3.2.9) with the absolute values replace by Euclidean norms. Therefore to conduct an energy goodness-of-fit test in the multivariate setting, we apply the test statistic

$$Q_{n,d} = n\left(\frac{2}{n}\sum_{i=1}^{n}\mathbb{E}||\mathbf{x}_i - \mathbf{X}_i|| - \mathbb{E}||\mathbf{X} - \mathbf{X}'|| - \frac{1}{n^2}\sum_{i,j=1}^{n}||\mathbf{x_i} - \mathbf{x_j}||\right). \tag{4.8.1}$$

The expressions

$$\begin{aligned}&\mathbb{E}||\mathbf{x}_i - \mathbf{X}|| = \\ &\int_{\mathbb{R}^d}||\mathbf{x}_i - \mathbf{x}||\frac{2e^{(\mathbf{x}-\theta)\Sigma^{-1}\mathbf{m}}}{(2\pi)^{d/2}|\Sigma|^{1/2}}\left(\frac{q(\mathbf{x})}{2 + \mathbf{m}^{\intercal}\Sigma^{-1}\mathbf{m}}\right)^{\nu/2}K_{\nu/2}\left(\sqrt{(2 + \mathbf{m}^{\intercal}\Sigma^{-1}\mathbf{m})q(\mathbf{x})}\right)d\mathbf{x}\end{aligned} \tag{4.8.2}$$

and

$$\begin{aligned}&\mathbb{E}||\mathbf{X} - \mathbf{X}'|| = \\ &\int_{\mathbb{R}^d}\mathbb{E}||\mathbf{x}_i - \mathbf{X}||\frac{2e^{(\mathbf{x}-\theta)\Sigma^{-1}\mathbf{m}}}{(2\pi)^{d/2}|\Sigma|^{1/2}}\left(\frac{q(\mathbf{x})}{2 + \mathbf{m}^{\intercal}\Sigma^{-1}\mathbf{m}}\right)^{\nu/2}K_{\nu/2}\left(\sqrt{(2 + \mathbf{m}^{\intercal}\Sigma^{-1}\mathbf{m})q(\mathbf{x})}\right)d\mathbf{x}\end{aligned} \tag{4.8.3}$$

are key to evaluating the one-sample energy goodness-of-fit statistic. Thus far, we do not know of an analytic expression for these expected values. As a proof-of-concept, one may conduct the energy goodness-of-fit test for the multivariate asymmetric Laplace distribution by first computing these expectations using numerical integration. Multivariate numerical integration in this scenario is complicated by two obstacles. The multivariate asymmetric Laplace model has an unbounded density at θ due to the presence of the Bessel function of the second kind and in order to compute $\mathbb{E}||\mathbf{X}-\mathbf{x}_i||$ and $\mathbb{E}||\mathbf{X}-\mathbf{X}'||$, we must integrate over the support of the $\mathbf{X}$ which is unbounded.

To deal with the unbounded domain of integration, a transformation is applied to change the bounds of integration to the hyper-cube $[\theta_1-1,\theta_1+1]\times\ldots\times[\theta_d-1,\theta_d+1]$. Specifically, we use the transformations

$$s_i = \frac{x_i-\theta_i}{1-(x_i-\theta_i)^2} \quad i=1,\ldots,d. \tag{4.8.4}$$

The resulting Jacobian determinant for the numerical integration is

$$\prod_{i=1}^{d}\frac{1+(x_i-\theta_i)^2}{(1-(x_i-\theta_i)^2)^2}. \tag{4.8.5}$$

Numerical integration will be unstable in a neighbor of θ. In response, we create a grid of quadrants that meet at θ so that any subdivision of the planes will not include θ. For example, if $d=2$, we divide $[\theta_1-1,\theta_1+1]\times[\theta_2-1,\theta_2+1]$ into four equal quadrants that each share a vertex with θ. If $d=3$, we divide the cube into 8 cubes of unit volume. The process may be continued for larger d. Unfortunately, the integration of (4.8.3) requires integration over $2d$ dimensions which results in a slower evaluation time.

We used the `Cuba` library (Hahn, 2005) to perform the numerical integration. The Cuhre algorithm (Berntsen, Espelid, and Genz, 1991) is used to approximate multivariate integrals. Cuhre is a deterministic algorithm which uses one of several quadrature rules of polynomial degree in a globally adaptive subdivision scheme. Cuhre is relatively fast and accurate for low dimensional multivariate integration.

$n = 50$	$n = 100$	$n = 250$
2.711	2.704	2.707

Table 4.2 Simulation estimates of critical values for testing $\mathscr{AL}_2(\mathbf{0},\mathbf{0},\mathbf{I})$ at the $\alpha = 0.10$ level for samples of size 50, 100, and 250.

Numerical integration was performed in `C++` and exported to the `R` environment using the `Rcpp` library for `R` (Eddelbuettel and François, 2011). Source code for the numerical integration of (4.8.2) and (4.8.3) is included in Appendix B.

Following the computation of expected values (4.8.2) and (4.8.3), the computation of the energy goodness-of-fit statistic may proceed by inserting these estimates into equation (4.8.1).

4.9 Multivariate Asymmetric Laplace Energy Test Simulations

A simulation was performed to assess the adequacy of the multivariate asymmetric Laplace test. In this section, we consider the simple hypothesis that the underlying distribution is bivariate, symmetric Laplace. Due to this simple hypothesis, we may calculate critical values of the test through Monte Carlo simulation. The critical values of the energy test are given in Table (4.2). Critical values are determined by Monte Carlo simulation with 5000 replicates. An example of the composite test is shown in Section 5.2, but the power analysis of the composite, multivariate Laplace distribution is omitted from this dissertation.

Random variates were generated from each alternative distribution in samples of size 50, 100, and 250. The simulation size is set to 2000 replicates to determine to empirical power of the energy test. Specific alternative distributions are listed in Section 4.9.1. None of the alternative distributions exhibit the excess kurtosis that is found in the bivariate Laplace distribution so the test is quite powerful for this simple hypothesis. Type I error rates are given in the first line of Table 4.3 and found to be controlled close to the to 10% level.

We find generally that the energy test is quite powerful against many of the alternative distributions featured in Table 4.3. The energy test can detect multivariate Normality with 72% power in samples of size 100, and is extremely powerful when testing against the multivariate Student's t distribution, and the skew Normal distribution.

4.9.1 Alternative Multivariate Distributions

Four alternative distributions are considered in this simulation study.

Bivariate Normal

$$f_{\mathcal{N}}(\mathbf{x};\theta,\Sigma) = \frac{1}{\sqrt{(2\pi)^2 \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x}-\theta)^{\intercal}\Sigma^{-1}(\mathbf{x}-\theta)\right)$$

Bivariate Student's t

$$f_{\mathcal{T}}(\mathbf{x};v) = \frac{\Gamma\left(\frac{v}{2}+1\right)}{\Gamma\left(\frac{v}{2}\right)\pi v \det(\Sigma)^{1/2}}\left[1+\frac{1}{v}\mathbf{x}^{\intercal}\Sigma^{-1}\mathbf{x}\right]^{-\frac{v}{2}-1}$$

$$\mathbf{Y}_{\mathcal{T}} \overset{d}{=} \left(\frac{\sqrt{\chi^2_v}}{v}\right)^{-1}\mathbf{Z}_{\mathcal{N}}$$

Bivariate Normal-Laplace Mixture

$$\mathbf{Y}_{\mathcal{LN}}(p) \overset{d}{=} p\mathbf{Y}_{\mathcal{L}} + (1-p)\mathbf{Z}_{\mathcal{N}}$$

Bivariate Skew-Normal

$$\mathbf{Y}_{\mathcal{SN}} \overset{d}{=} \lambda|Z_{\mathcal{N}}|\mathbf{1} + \sqrt{1-\lambda^2}\mathbf{Z}_{\mathcal{N}}$$

Distribution	$n = 50$	$n = 100$	$n = 250$
$\mathscr{L}_2(\mathbf{0}, \mathbf{0}, \mathbf{I})$	10	10	10
$\mathscr{LN}(0.75)$	12	13	19
$\mathscr{LN}(0.50)$	18	24	45
$\mathscr{LN}(0.25)$	26	44	87
$\mathscr{N}_2(\mathbf{0}, \mathbf{I})$	39	72	99
$\mathscr{T}_2(2)$	100	100	100
$\mathscr{T}_2(5)$	100	100	100
$\mathscr{SN}(0.25)$	77	98	100
$\mathscr{SN}(0.50)$	99	100	100
$\mathscr{SN}(0.75)$	100	100	100

Table 4.3 Power and type I error of the bivariate energy test for $H : \mathbf{x}_1, \ldots, \mathbf{x}_n \sim \mathscr{AL}(\mathbf{0}, \mathbf{0}, \mathbf{I})$ for $n = 50, 100, 250$.

CHAPTER 5 APPLICATION OF ESTIMATION AND TESTING TECHNIQUES

Characterizing the distribution of financial returns is often of interest in the field of quantitative finance. Financial returns may be the change in value of a stock, bond, currency, mutual fund, option or other financial instrument. Complex time series methods have been developed to model the value of various financial assets. However, in this chapter, we will focus on the simpler problem of testing the goodness-of-fit of the asymmetric Laplace distribution to the daily logarithmic returns of a mutual fund and a stock.

Often the distribution of financial returns is found to be leptokurtic, that is, their histogram is found to be fat-tailed and sharply peaked. Depending on the financial instrument, there may also be a degree of asymmetry to the returns. Various distributions have been proposed to capture these features such as Stable, Pareto, mixture models, and double Weibull. Kozubowski and Podgorski (Kozubowski and Podgórski, 2001) showed empirically that the asymmetric Laplace distribution is a strong competitor against these models for describing the distribution of currency exchange rates.

5.1 Univariate Example

The historical price of Vanguard LifeStrategy Growth fund (VASGX) is used as an example. VASGX is a mutual fund composed of 80% stocks and 20% bonds. A portion of stocks and bonds in the fund are international. VASGX is an example of a mutual fund that provides a great degree of diversification for an investor, but is subject to currency and international risks, in addition to the nominal risk inherent to the US stock market.

The daily return of the fund was downloaded from Yahoo! Finance via the `quantmod` package in `R`. A period of 2000 days was collected from 2007-01-03 to 2018-01-29. Let P_t be the daily closing price of VASGX on day t. The logarithmic daily return is calculated as $r_t \equiv \log\left(\frac{P_t}{P_{t-1}}\right)$. After this transformation, there are 1999 observations. A histogram of the data with fitted asymmetric Laplace density is presented in Figure 5.1.

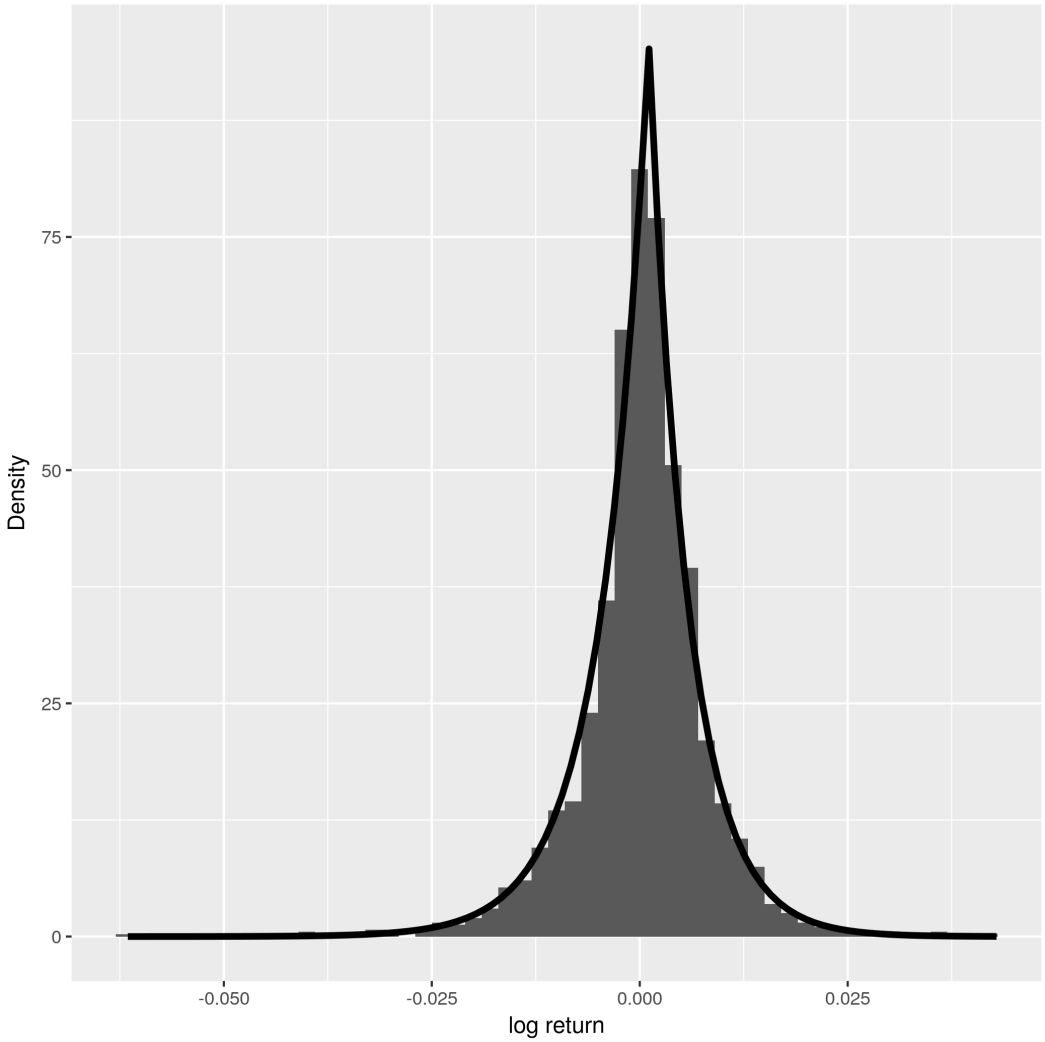

Figure 5.1 Histogram of daily log returns of VASGX with estimated asymmetric Laplace density

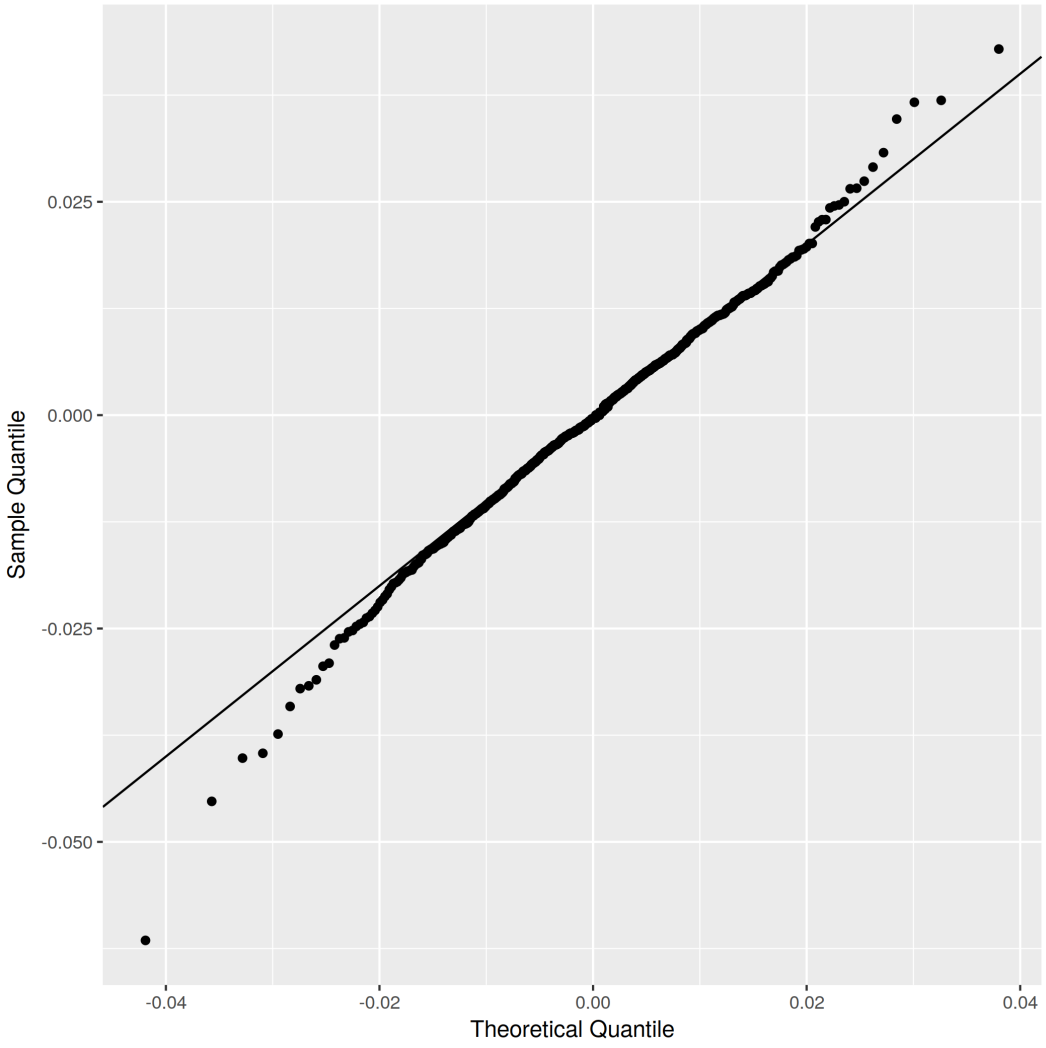

Figure 5.2 Q-Q plot of sample quantiles against $\mathscr{AL}^*$ quantiles with estimated parameters

Fitting the asymmetric Laplace model to the data results in the parameter estimates given in Table 5.1. We find the location parameter θ is nearly 0. The skewness parameter κ is slightly larger than 1 indicating that a symmetric Laplace model may be appropriate.

We wish to test the composite hypothesis $r_1, \ldots r_m \sim \mathscr{AL}$ with unknown parameters. An argument could be made that the location parameter θ is 0 a priori, but we do not make this assumption. The value of the energy test statistic is computed after centering and scaling the data. The p-value of the composite energy goodness-of-fit test is computed through a parametric bootstrap. This p-value only depends on the skewness estimate $\hat{\kappa}$ and sample size because we perform the composite test on standardized data. All parameters are estimated through the E-M algorithm (Section 4.4.1).

The p-value of the test (Table 5.1) is relatively large. This does not imply that the asymmetric Laplace hypothesis is true, rather, it shows that we lack sufficient evidence to conclude otherwise. However, we are comforted by the fact that this p-value was gathered from a sample of size 1999. The combination of the energy test and the "peakedness" of the histogram (Figure 5.1) lead us to believe that the asymmetric Laplace model is a good fit for this data (or conservatively, as least not a bad fit). A Q-Q plot is given in Figure 5.2. The Q-Q plot shows minimal cause for concern over our hypothesized distribution.

Ticker Symbol	Sample Size	$\hat{\theta}_{EM}$	$\hat{\sigma}_{EM}$	$\hat{\kappa}_{EM}$	Energy Statistic	p-value
VASGX	1999	0.001	0.007	1.072	1.062	0.335

Table 5.1 VASGX: Parameter estimates and goodness-of-fit

5.2 Multivariate Example

We now consider an example of testing the multivariate goodness-of-fit hypothesis. Previously we found that $\mathscr{AL}$ may be a sufficient model for the logarithmic daily returns on VASGX. We now wish to determine if the bivariate $\mathscr{AL}_2$ model is sufficient for modeling the joint distribution of the logarithmic daily returns of VASGX and BA, Boeing's stock price. Original stock prices are shown in Figure 5.3. The price of Boeing is larger and more volatile than the price of the Vanguard mutual fund, which is smaller but more stable.

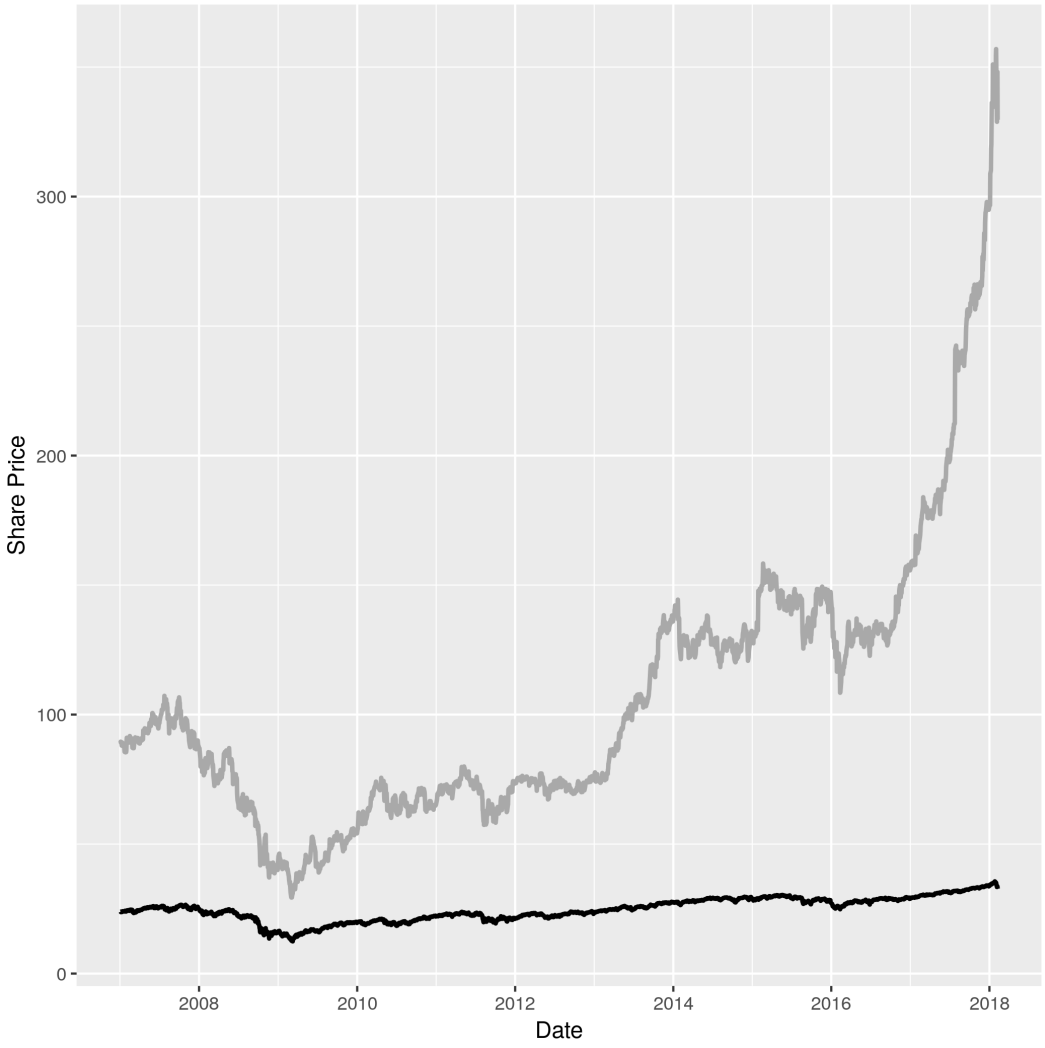

Figure 5.3 Share price of BA (gray) and VASGX (black) over the period of the study.

Ticker Symbol	Sample Size	$\hat{\theta}$	$\hat{\Sigma}$	$\hat{m}$
$\begin{pmatrix} VASGX \\ BA \end{pmatrix}$	1999	$\begin{pmatrix} 0.0009 \\ 0.0016 \end{pmatrix}$	$\begin{pmatrix} 0.00006 & 0.00007 \\ 0.00007 & 0.0002 \end{pmatrix}$	$\begin{pmatrix} -0.0007 \\ -0.0008 \end{pmatrix}$

Table 5.2 VASGX and BA $\mathscr{AL}_2$ parameter estimates

Before standardizing the data, we compute E-M estimates of bivariate asymmetric Laplace parameters. One may compute bias corrected parameter estimates, but in this example we calculate E-M estimates. The results of the procedure to shown in Table 5.2. As in the univariate example, we observe a location parameter close to the origin, and a small degree of asymmetry.

A multivariate energy test is conducted to assess to reasonableness of an asymmetric Laplace model. The results of the procedure are shown in Table 5.3. We find that the energy statistic is too large to justify the use of an $\mathscr{AL}_2$ model for this data. Figures 5.4 and 5.5 compare the contours of a bivariate kernel density estimator and the bivariate $\mathscr{AL}_2$ parametric fit, respectively. Comparing the non-parametric density estimator with the parametric one, we find the parameter estimate for the dispersion matrix Σ may be under-estimated — a large portion of observations fall outside of the outermost $\mathscr{AL}_2$ contour line. The contour lines of the non-parametric density estimator reveal a slight "S" shape in the scatter-plot, a pattern that the asymmetric Laplace distribution will fail to model with any combination of parameter estimates.

Ticker Symbol	Sample Size	Energy GoF Statistic	p-value
VASGX and BA	1999	2.86	< 0.01

Table 5.3 VASGX and BA Goodness-of-fit statistics

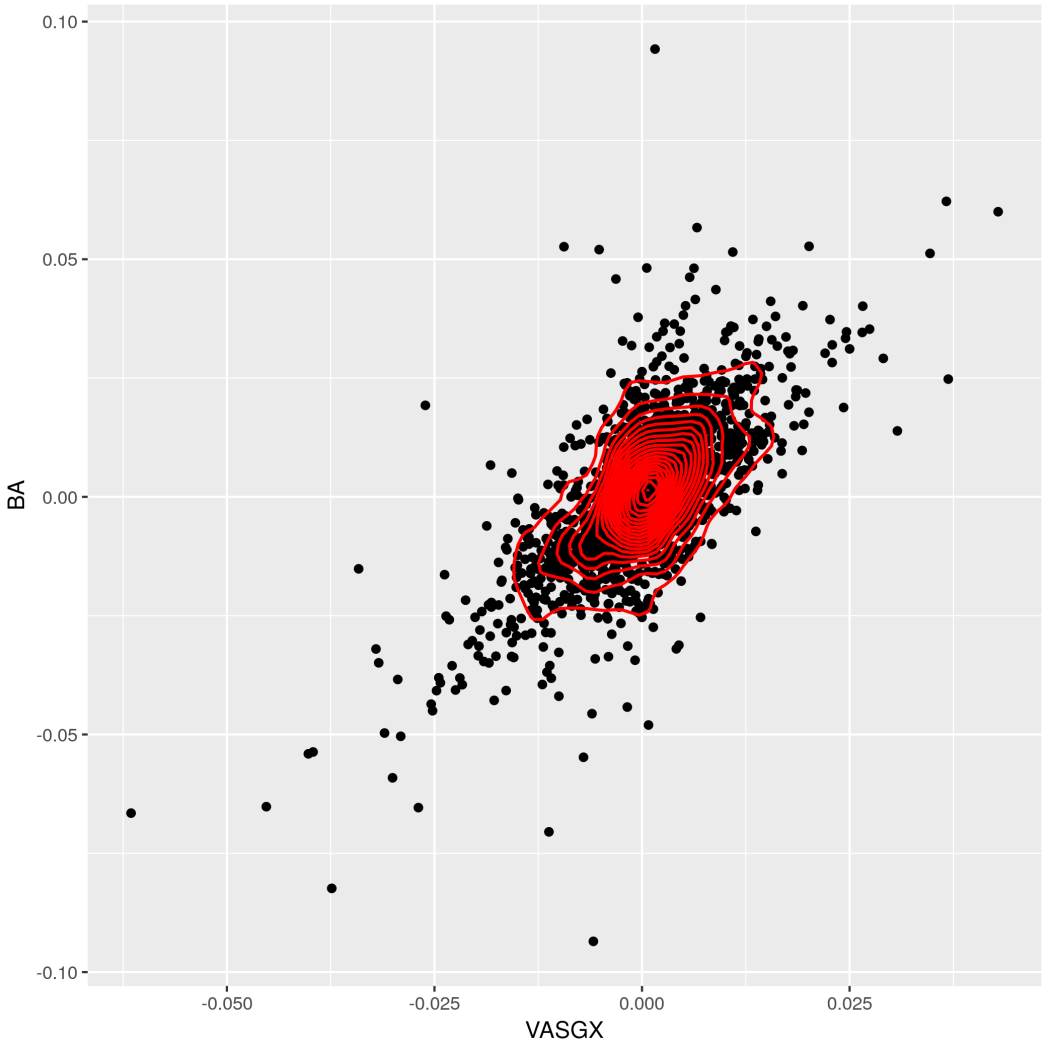

Figure 5.4 Scatter plot of VASGX and BA with kernel density estimator contours (20 contour lines).

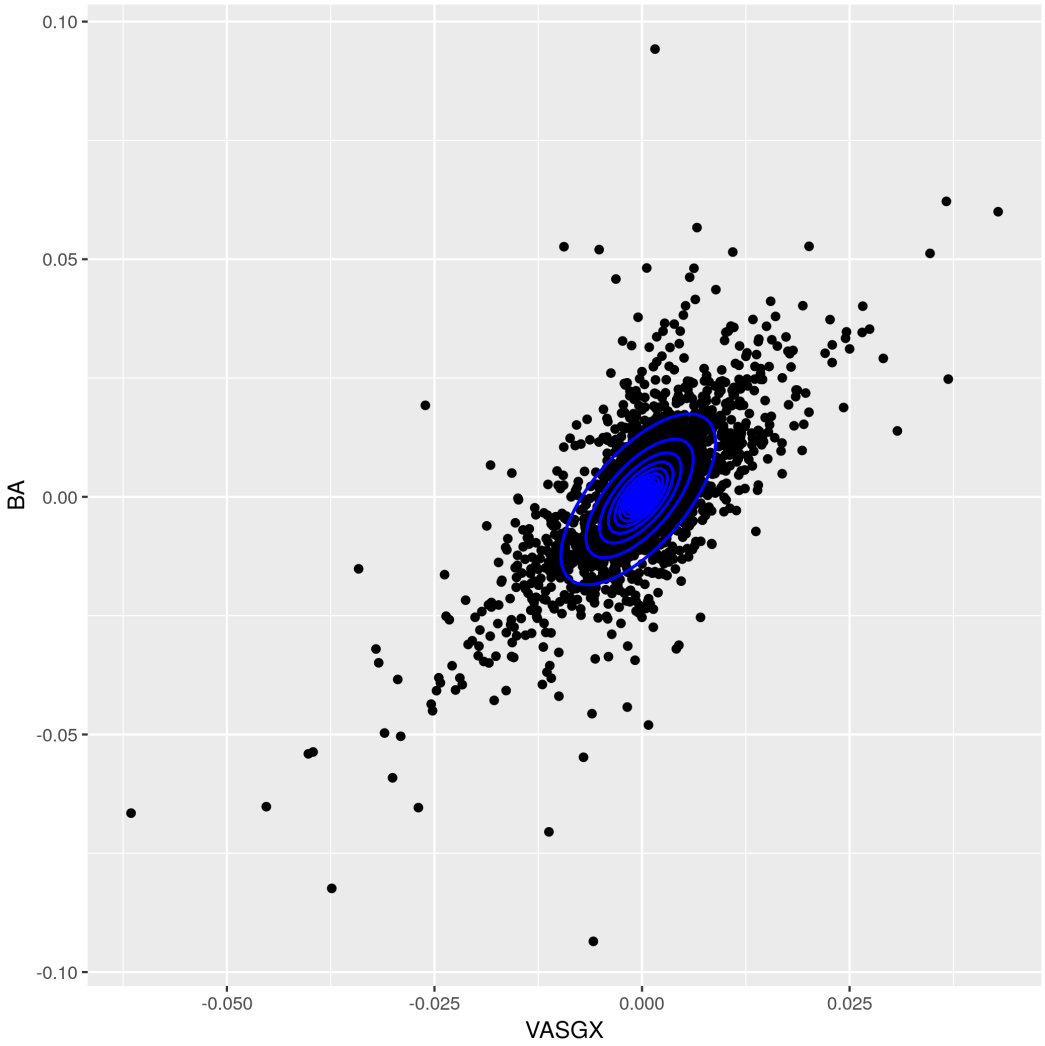

Figure 5.5 Scatter plot of VASGX and BA with $\mathscr{AL}_2$ contour plot overlay (estimated parameters, 20 contour lines)

CHAPTER 6 SUMMARY

In this dissertation, a variety of new results are shown in connection to the asymmetric Laplace distribution and its extensions — the multivariate asymmetric Laplace distribution and the generalized asymmetric Laplace distribution. A test for the univariate asymmetric Laplace distribution and its extension the variance gamma distribution are detailed in Chapter 3. These tests are among the most powerful tests that one may use to investigate the goodness-of-fit of these distributions. Another energy statistic, distance standard deviation, is computed for the $\mathscr{AL}$ distribution in Chapter 3 for use by researchers that are inclined to measure dispersion in this manner.

We found that the classical ML parameter estimates often fail to be admissible, and a new application of the E-M algorithm is derived in Chapter 4 to address the short-comings of ML estimation. The advantage of our E-M algorithm is that it provides admissible maximum likelihood estimates in the univariate case, and bests the MM estimates in the multivariate case in terms of bias and standard error. Using this multivariate E-M estimator, we approximate the energy statistics required to assess the multivariate $\mathscr{AL}_d$ hypothesis through numerical integration.

An application of our testing and estimation procedures is given in Chapter 5 to show how we can effectively characterize the distribution of logarithmic returns of financial instruments. We show in this application that one can use the E-M algorithm to estimate the parameters of the joint distribution of logarithmic returns of multiple stocks. Furthermore, the energy goodness-of-fit test can quickly assess the univariate asymmetric Laplace hypothesis for the logarithmic returns of individual stocks.

Further research is required to complete the work on the multivariate energy test for the asymmetric Laplace distribution. In this dissertation, we have found that one can use numerical integration to compute the test statistics in the multivariate case, but the use of numerical integration is slow and potentially fraught with numerical demons. There is potential for a fast

multivariate energy test if analytically useful approximations for the Bessel function of the second kind are found. If such an approximation is presented, progress could also be made on the energy test for generalized asymmetric Laplace distribution with real valued shape parameter τ.

BIBLIOGRAPHY

Abramowitz, M. and I. A. Stegun (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Volume 55. Courier Corporation.

Azzalini, A. (2017). *The R package sn: The Skew-Normal and Related Distributions such as the Skew-t (version 1.5-1).* Università di Padova, Italia.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics 24*(1), 1–13.

Berntsen, J., T. O. Espelid, and A. Genz (1991). An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Transactions on Mathematical Software (TOMS) 17*(4), 437–451.

Casella, G. and R. L. Berger (2002). *Statistical Inference*, Volume 2. Duxbury Pacific Grove, CA.

Chen, C. (2002). Tests for the goodness-of-fit of the Laplace distribution. *Communications in Statistics-Simulation and Computation 31*(1), 159–174.

Choi, B. and K. Kim (2006). Testing goodness-of-fit for Laplace distribution based on maximum entropy. *Statistics 40*(6), 517–531.

D'Agostino, R. B. and M. A. Stephens (2017). *Goodness-of-fit Techniques*. Routledge.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer Science & Business Media.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

Devroye, L. (1986). Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pp. 260–265. ACM.

Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM.

Eddelbuettel, D. and R. François (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software 40*(8), 1–18.

Edelmann, D., D. Richards, and D. Vogel (2017, May). The distance standard deviation. *ArXiv e-prints*.

Eltoft, T., T. Kim, and T. Lee (2006). On the multivariate Laplace distribution. *IEEE Signal Processing Letters 13*(5), 300–303.

Faraway, J., G. Marsaglia, J. Marsaglia, and A. Baddeley (2017). *goftest: Classical Goodness-of-Fit Tests for Univariate Distributions*. R package version 1.1-1.

Fernández, C. and M. F. Steel (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association 93*(441), 359–371.

Fieller, N. (1993). Archaeostatistics: old statistics in ancient contexts. *The Statistician 42*, 279–295.

Fragiadakis, K. and S. Meintanis (2009). Test of fit for asymmetric Laplace distributions with applications. *Journal of Statistics: Advances in Theory and Applications 1*(1), 49–63.

Fragiadakis, K. and S. G. Meintanis (2011). Goodness-of-fit tests for multivariate Laplace distributions. *Mathematical and Computer Modelling 53*(5), 769–779.

Gerstenberger, C. and D. Vogel (2015). On the efficiency of Gini's mean difference. *Statistical Methods & Applications 24*(4), 569–596.

Givens, G. H. and J. A. Hoeting (2012). *Computational Statistics*, Volume 710. John Wiley & Sons.

Hahn, T. (2005). Cuba — a library for multidimensional numerical integration. *Computer Physics Communications 168*(2), 78–95.

Hartley, M. J. and N. S. Revankar (1974). On the estimation of the Pareto law from under-reported data. *Journal of Econometrics 2*(4), 327–341.

Hinkley, D. V. and N. S. Revankar (1977). Estimation of the Pareto law from underreported data: A further analysis. *Journal of Econometrics 5*(1), 1–11.

Huo, X. and G. J. Székely (2016). Fast computing for distance covariance. *Technometrics 58*(4), 435–447.

Hürlimann, W. (2013). A moment method for the multivariate asymmetric Laplace distribution. *Statistics & Probability Letters 83*(4), 1247–1253.

Keynes, J. (1911). The principal averages and the laws of error which lead to them. *Journal of the Royal Statistical Society 74*(3), 322–331.

Koenker, R. (2005). *Quantile Regression*. Number 38. Econometric Society Monograph Series, Cambridge University Press.

Kollo, T. (2008). Multivariate skewness and kurtosis measures with an application in ICA. *Journal of Multivariate Analysis 99*(10), 2328–2338.

Kollo, T. and M. S. Srivastava (2005). Estimation and testing of parameters in multivariate Laplace distribution. *Communications in Statistics-Theory and Methods 33*(10), 2363–2387.

Kotz, S., T. Kozubowski, and K. Podgorski (2001). *The Laplace Distribution and Generalizations: a Revisit with Applications to Communications, Economics, Engineering, and Finance*. Springer Science & Business Media.

Kotz, S., T. J. Kozubowski, and K. Podgórski (2002). Ml estimation of asymmetric Laplace parameters. *Annals of the Institute of Statistical Mathematics 54*(4), 816–826.

Kozubowski, T. and A. Panorska (1999). Multivariate geometric stable distributions in financial applications. *Mathematical and Computer Modelling 29*(10-12), 83–92.

Kozubowski, T. J. (1997). Characterization of multivariate geometric stable distributions. *Statistics & Risk Modeling 15*(4), 397–416.

Kozubowski, T. J. and K. Podgórski (2001). Asymmetric Laplace laws and modeling financial data. *Mathematical and Computer Modelling 34*(9-11), 1003–1021.

Laplace, P. S. (1774). Memoir on the probability of the causes of events. *Statistical Science 1*(3), 364–378.

Levin, A. and A. Tchernitser (2003). Multifactor stochastic variance models in risk management: maximum entropy approach and Lévy processes. In S. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance*, pp. 443–480.

Lindsey, J. and P. Lindsey (2006). Multivariate distributions with correlation matrices for nonlinear repeated measurements. *Computational statistics & Data Analysis 50*(3), 720–732.

MacRae, E. C. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *The Annals of Statistics 2*(2), 337–346.

Madan, D. B., P. P. Carr, and E. C. Chang (1998). The variance gamma process and option pricing. *Review of Finance 2*(1), 79–105.

Madan, D. B. and E. Seneta (1990). The variance gamma (vg) model for share market returns. *Journal of Business 63*(4), 511–524.

McLachlan, G. and T. Krishnan (2007). *The EM Algorithm and Extensions*, Volume 382. John Wiley & Sons.

Norris, J., R. Nemiroff, J. Bonnell, J. Scargle, C. Kouveliotou, W. Paciesas, C. Meegan, and G. Fishman (1996). Attributes of pulses in long bright gamma-ray bursts. *The Astrophysical Journal 459*, 393.

Ohio Supercomputer Center (1987). `http://osc.edu/ark:/19495/f5s1ph73`.

Olver, F. W. J. (1974). *Asymptotics and Special Functions*. Academic Press.

Pearson, K., G. Jeffery, and E. M. Elderton (1929). On the distribution of the first product moment-coefficient, in samples drawn from an indefinitely large normal population. *Biometrika*, 164–201.

Peterson, R. and E. A. Silver (1979). *Decision Systems for Inventory Management and Production Planning*. Wiley New York.

Puig, P. and M. A. Stephens (2007). Goodness of fit tests for the skew-Laplace distribution. *SORT 31*(1).

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rizzo, M. L. (2002). *A new rotation invariant goodness-of-fit test*. Ph. D. thesis, Bowling Green State University.

Rizzo, M. L. (2007). *Statistical Computing with R*. CRC Press.

Rizzo, M. L. and J. T. Haman (2016). Expected distances and goodness-of-fit for the asymmetric Laplace distribution. *Statistics & Probability Letters 117*, 158–164.

Rossi, E. and F. Spazzini (2010). Model and distribution uncertainty in multivariate garch estimation: A monte carlo analysis. *Computational Statistics & Data Analysis 54*(11), 2786–2800.

Rublík, F. (1997). A quantile goodness-of-fit test applicable to distributions with non-differentiable densities. *Kybernetika 33*(5), 505–524.

Sánchez, B., H. Lachos, and V. Labra (2013). Likelihood based inference for quantile regression using the asymmetric Laplace distribution. *Journal of Statistical Computation and Simulation 81*, 1565–1578.

Székely, G. J. (2000). Technical report 03-05: E-statistics: energy of statistical samples. *Department of Mathematics and Statistics, Bowling Green State University*.

Székely, G. J. and M. L. Rizzo (2005). A new test for multivariate normality. *Journal of Multivariate Analysis 93*(1), 58–80.

Székely, G. J. and M. L. Rizzo (2013). Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference 143*(8), 1249–1272.

Székely, G. J. and M. L. Rizzo (2017). The energy of data. *Annual Review of Statistics and Its Application 4*(1), 447–479.

Székely, G. J., M. L. Rizzo, and N. Bakirov (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics 35*(6), 2769–2794.

Teichroew, D. (1957). The mixture of normal distributions with different variances. *The Annals of Mathematical Statistics 28*(2), 510–512.

Uppuluri, V. (1981). Some properties of log-Laplace distribution. *Statistical Distributions in Scientific Work 4*, 105–110.

Visk, H. (2009). On the parameter estimation of the asymmetric multivariate Laplace distribution. *Communications in Statistics–Theory and Methods 38*(4), 461–470.

Watson, G. N. (1995). *A Treatise on the Theory of Bessel Functions*. Cambridge University Press.

William, J. C. (1971). *Practical Nonparametric Statistics*. John Wiley & Sons, New York.

Wilson, E. B. (1923). First and second laws of error. *Journal of the American Statistical Association 18*(143), 841–851.

Yen, V. C. and A. H. Moore (1988). Modified goodness-of-fit test for the Laplace distribution. *Communications in Statistics-Simulation and Computation 17*(1), 275–281.

APPENDIX A EXPECTED DISTANCE FUNCTIONS

These functions are used in the expression of $\mathbb{E}|Y-Y'|$ for a $\mathscr{GAL}^*$ random variable. The expressions are referenced in Proposition 3.5.40. We make use of the variables λ, β, and A_j as defined in equation (3.5.22) and equation (3.5.20).

$$I_1 = \sum_{j=0}^{n-1} A_j \lambda^{j-n} \Gamma(n-j) \tag{A.0.1}$$

$$I_2 = -\sum_{j=0}^{n-1} A_j \lambda^{j-n-1} \Gamma(n-j+1) \tag{A.0.2}$$

$$I_3 = \sum_{j=0}^{n-1} A_j \beta^{j-n} \Gamma(n-j) \tag{A.0.3}$$

$$I_2 = \sum_{j=0}^{n-1} A_j \beta^{j-n-1} \Gamma(n-j+1) \tag{A.0.4}$$

$$I_5 = \sum_{j=0}^{n-1} \sum_{k=0}^{n-i} A_j \frac{(n-i)!\beta^k (2\beta)^{j-k-n} \Gamma(n-j+k)}{k!} \tag{A.0.5}$$

$$I_6 = \sum_{j=0}^{n-1} \sum_{k=0}^{n-i} A_j \frac{(n-i)!\lambda^k (2\lambda)^{j-k-n} \Gamma(k-j+n)}{k!} \tag{A.0.6}$$

$$I_7 = \sum_{j=0}^{n-1} \sum_{k=0}^{n-i-1} A_j \frac{(n-i-1)!\beta^k (2\beta)^{j-k-n-1} \Gamma(k+n-j+1)}{k!} \tag{A.0.7}$$

$$I_8 = -\sum_{j=0}^{n-1} \sum_{k=0}^{n-i-1} A_j \frac{(n-i-1)!\lambda^k (2\lambda)^{j-k-n-1} \Gamma(k+n-j+1)}{k!} \tag{A.0.8}$$

APPENDIX B SOURCE CODE FOR PROGRAMS

Source code for the numerical computation of equation 4.8.2 and equation 4.8.2 for the bivariate asymmetric Laplace distribution. Source is written in C++ for the Rcpp interface to R.

```
// [[Rcpp::depends(RcppEigen)]]
// [[Rcpp::depends(RcppNumerical)]]
// [[Rcpp::depends(RcppGSL)]]

#include <RcppNumerical.h>
#include <RcppGSL.h>
#include <math.h>
#include <gsl/gsl_sf_bessel.h>

/*
The following need to be loaded in the R session:
library(Rcpp)
library(RcppEigen)
library(RcppNumerical)
library(RcppGSL)
*/

using namespace Numer;

class BiLaplace: public MFunc
{
private:
  const double theta1;
  const double theta2;
  const double rho;
  const double m1;
  const double m2;
  const double sigma1;
  const double sigma2;
  double const1;
  double b;

public:
  BiLaplace(const double& theta1_,
       const double& theta2_,
       const double& m1_,
       const double& m2_,
       const double& sigma1_,
       const double& sigma2_,
       const double& rho_) : rho(rho_), sigma1(sigma1_),
```

```
            sigma2(sigma2_), m1(m1_), m2(m2_),
            theta1(theta1_), theta2(theta2_)
  {
    const1 = std::sqrt(2*sigma1*sigma2*(1-rho*rho)
           + m1*m1*sigma2/sigma1 - 2*m1*m2*rho
           + m2*m2*sigma1/sigma2)/(sigma1*sigma2*(1-rho*rho));
    b = -log(M_PI*sigma1*sigma2*std::sqrt(1-rho*rho));
  }

  // PDF of bivariate AL
  double operator()(Constvec& x)
  {
    double s = (x[0] - theta1) / (1 - std::pow(x[0] - theta1, 2));
    double t = (x[1] - theta2) / (1 - std::pow(x[1] - theta2, 2));
    double c =((m1*sigma2/sigma1 -m2*rho)*s +
         (m2*sigma1/sigma2 - m1*rho)*t) /
             (sigma1*sigma2*(1-rho*rho));
    double d = log(gsl_sf_bessel_K0(const1*std::sqrt(std::pow(s, 2)*sigma2/sigma1
       -
                    2 * rho * s * t +
                    std::pow(t, 2)*sigma1/sigma2)));
    double e = log((1 + std::pow(x[0] - theta1, 2)) / std::pow(1 - std::pow(x[0]
       - theta1, 2), 2));
    double f = log((1 + std::pow(x[1] - theta2, 2)) / std::pow(1 - std::pow(x[1]
       - theta2, 2), 2));
    return std::exp(b + c + d + e + f);
  }
};


// [[Rcpp::export]]
Rcpp::List pBiLaplace(double theta1 = 0,
           double theta2 =0,
           double m1 =0,
           double m2=0,
           double sigma1 =1,
           double sigma2=1,
           double rho=0,
           int evaluations = 500000)
{
  BiLaplace f(theta1, theta2, m1, m2, sigma1, sigma2,rho);
  Eigen::VectorXd lowerI(2);
  lowerI << theta1, theta2;
  Eigen::VectorXd upperI(2);
  upperI << theta1 + 1, theta2 + 1;
  Eigen::VectorXd lowerII(2);
  lowerII << theta1 - 1, theta2;
```

```
  Eigen::VectorXd upperII(2);
  upperII << theta1, theta2 + 1;
  Eigen::VectorXd lowerIII(2);
  lowerIII << theta1 - 1, theta2 - 1;
  Eigen::VectorXd upperIII(2);
  upperIII << theta1, theta2;
  Eigen::VectorXd lowerIV(2);
  lowerIV << theta1, theta2 - 1;
  Eigen::VectorXd upperIV(2);
  upperIV << theta1 + 1, theta2;
  double err_est;
  int err_code;
  int maxeval = evaluations;
  const double resI = integrate(f, lowerI, upperI, err_est, err_code, maxeval);
  const double resII = integrate(f, lowerII, upperII, err_est, err_code,
     maxeval);
  const double resIII = integrate(f, lowerIII, upperIII, err_est, err_code,
     maxeval);
  const double resIV = integrate(f, lowerIV, upperIV, err_est, err_code,
     maxeval);
  return Rcpp::List::create(
            Rcpp::Named("approximate") = resI + resII + resIII + resIV,
            Rcpp::Named("error_estimate") = err_est,
            Rcpp::Named("error_code") = err_code
            );
}
// Distance to a fixed point integrand
class BiLaplaceDistIntegrand: public MFunc
{
private:
  const double theta1;
  const double theta2;
  const double rho;
  const double m1;
  const double m2;
  const double sigma1;
  const double sigma2;
  const double a1;
  const double a2;
  double const1;
  double b;

public:
  BiLaplaceDistIntegrand(const double& theta1_,
          const double& theta2_,
          const double& m1_,
          const double& m2_,
```

```
        const double& a1_,
        const double& a2_,
        const double& sigma1_,
        const double& sigma2_,
        const double& rho_) : theta1(theta1_), theta2(theta2_),
                  rho(rho_), sigma1(sigma1_), sigma2(sigma2_),
                  m1(m1_), m2(m2_), a1(a1_), a2(a2_)
  {
    const1 = std::sqrt(2*sigma1*sigma2*(1-rho*rho)
           + m1*m1*sigma2/sigma1 - 2*m1*m2*rho
           + m2*m2*sigma1/sigma2)/(sigma1*sigma2*(1-rho*rho));
    b = -log(M_PI*sigma1*sigma2*std::sqrt(1-rho*rho));
  }

  // PDF of bivariate AL
  double operator()(Constvec& x)
  {
    double s = (x[0] - theta1) / (1 - std::pow(x[0] - theta1, 2));
    double t = (x[1] - theta2) / (1 - std::pow(x[1] - theta2, 2));
    double p1 = (a1) / (1 - std::pow(a1, 2));
    double p2 = (a2) / (1 - std::pow(a2, 2));
    double c =((m1*sigma2/sigma1 -m2*rho)*s +
         (m2*sigma1/sigma2 - m1*rho)*t) /
             (sigma1*sigma2*(1-rho*rho));
    double d = log(gsl_sf_bessel_K0(const1*std::sqrt(std::pow(s, 2)*sigma2/sigma1
       -
                    2*rho* s * t +
                    std::pow(t, 2)*sigma1/sigma2)));
    double e = log((1 + std::pow(x[0] - theta1, 2)) / std::pow(1 - std::pow(x[0]
       - theta1, 2), 2));
    double f = log((1 + std::pow(x[1] - theta2, 2)) / std::pow(1 - std::pow(x[1]
       - theta2, 2), 2));
    double g = log(std::sqrt(std::pow(s - a1 + theta1, 2) + std::pow(t - a2 +
       theta2, 2)));
    return std::exp(b + c + d + e + f + g);
  }
};

// [[Rcpp::export]]
Rcpp::List BiLaplaceExY(double theta1 = 0,
        double theta2 =0,
        double m1 = 0,
        double m2 = 0,
        double a1 = 0,
        double a2 = 0,
        double sigma1 = 1,
        double sigma2 = 1,
```

```
        double rho = 0,
        int evaluations = 500000)
{
  BiLaplaceDistIntegrand f(theta1, theta2, m1, m2, a1, a2, sigma1, sigma2, rho);
  Eigen::VectorXd lowerI(2);
  lowerI << theta1, theta2;
  Eigen::VectorXd upperI(2);
  upperI << theta1 + 1, theta2 + 1;
  Eigen::VectorXd lowerII(2);
  lowerII << theta1 - 1, theta2;
  Eigen::VectorXd upperII(2);
  upperII << theta1, theta2 + 1;
  Eigen::VectorXd lowerIII(2);
  lowerIII << theta1 - 1, theta2 - 1;
  Eigen::VectorXd upperIII(2);
  upperIII << theta1, theta2;
  Eigen::VectorXd lowerIV(2);
  lowerIV << theta1, theta2 - 1;
  Eigen::VectorXd upperIV(2);
  upperIV << theta1 + 1, theta2;
  double err_est;
  int err_code;
  int maxeval = evaluations;
  const double resI = integrate(f, lowerI, upperI, err_est, err_code, maxeval);
  const double resII = integrate(f, lowerII, upperII, err_est, err_code,
     maxeval);
  const double resIII = integrate(f, lowerIII, upperIII, err_est, err_code,
     maxeval);
  const double resIV = integrate(f, lowerIV, upperIV, err_est, err_code,
     maxeval);
  return Rcpp::List::create(
            Rcpp::Named("approximate") = resI + resII + resIII + resIV,
            Rcpp::Named("error_estimate") = err_est,
            Rcpp::Named("error_code") = err_code
            );
}



class BiLaplaceDistIntegrandYY: public MFunc
{
private:
  const double theta1;
  const double theta2;
  const double rho;
  const double m1;
  const double m2;
```

```
  const double sigma1;
  const double sigma2;
  double const1;
  double b1;
  double b2;

public:
  BiLaplaceDistIntegrandYY(const double& theta1_,
           const double& theta2_,
           const double& m1_,
           const double& m2_,
           const double& sigma1_,
           const double& sigma2_,
           const double& rho_) : theta1(theta1_), theta2(theta2_),
                 m1(m1_), m2(m2_), sigma1(sigma1_),
                 sigma2(sigma2_), rho(rho_)
  {
    const1 = std::sqrt(2*sigma1*sigma2*(1-rho*rho) +
             m1*m1*sigma2/sigma1 - 2*m1*m2*rho +
             m2*m2*sigma1/sigma2) /
                      (sigma1*sigma2*(1-rho*rho));
    b1 = -log(M_PI*sigma1*sigma2*std::sqrt(1-rho*rho));
    b2 = b1;
  }

  double operator()(Constvec& x)
  {
    double s = (x[0] - theta1) / (1 - std::pow(x[0] - theta1, 2));
    double t = (x[1] - theta2) / (1 - std::pow(x[1] - theta2, 2));
    double u = (x[2] - theta1) / (1 - std::pow(x[2] - theta1, 2));
    double v = (x[3] - theta2) / (1 - std::pow(x[3] - theta2, 2));

    double c1 =((m1*sigma2/sigma1 -m2*rho)*(s) +
      (m2*sigma1/sigma2 - m1*rho)*(t)) /
               (sigma1*sigma2*(1-rho*rho));
    double d1 = log(gsl_sf_bessel_K0(const1*std::sqrt(std::pow(s,
       2)*sigma2/sigma1 -
                      2*rho*s * t +
                      std::pow(t, 2)*sigma1/sigma2)));
    double c2 =((m1*sigma2/sigma1 -m2*rho)*(u) +
         (m2*sigma1/sigma2 - m1*rho)*(v)) /
              (sigma1*sigma2*(1-rho*rho));
    double d2 = log(gsl_sf_bessel_K0(const1*std::sqrt(std::pow(u,
       2)*sigma2/sigma1 -
                      2*rho*u*v +
                      std::pow(v, 2)*sigma1/sigma2)));
```

```
    double e1 = log((1 + std::pow(x[0] - theta1, 2)) / std::pow(1 - std::pow(x[0]
        - theta1, 2), 2));
    double f1 = log((1 + std::pow(x[1] - theta2, 2)) / std::pow(1 - std::pow(x[1]
        - theta2, 2), 2));
    double e2 = log((1 + std::pow(x[2] - theta1, 2)) / std::pow(1 - std::pow(x[2]
        - theta1, 2), 2));
    double f2 = log((1 + std::pow(x[3] - theta2, 2)) / std::pow(1 - std::pow(x[3]
        - theta2, 2), 2));
    double g = log(std::sqrt((s - u)*(s - u) + (t - v)*(t - v)));
    return std::exp(b1 + c1 + d1 + b2 + c2 + d2 + e1 + e2 + f1 + f2 + g);
  }
};

// Compute the Expected Distance between two bivariate ALs
// [[Rcpp::export]]
Rcpp::List BiLaplaceEYY(double theta1 = 0,
        double theta2 =0,
        double m1 = 0,
        double m2 = 0,
        double sigma1 = 1,
        double sigma2 = 1,
        double rho = 0,
        int evaluations = 500000)
{
  BiLaplaceDistIntegrandYY f(theta1, theta2, m1, m2, sigma1, sigma2, rho);
  Eigen::VectorXd lower1(4);
  lower1 << theta1, theta2, theta1, theta2;
  Eigen::VectorXd upper1(4);
  upper1 << theta1 + 1, theta2 + 1, theta1 + 1, theta2 + 1;
  Eigen::VectorXd lower2(4);
  lower2 << theta1, theta2, theta1, theta2-1;
  Eigen::VectorXd upper2(4);
  upper2 << theta1+1, theta2+1, theta1+1, theta2;
  Eigen::VectorXd lower3(4);
  lower3 << theta1, theta2, theta1-1, theta2;
  Eigen::VectorXd upper3(4);
  upper3 << theta1+1, theta2+1, theta1, theta2+1;
  Eigen::VectorXd lower4(4);
  lower4 << theta1, theta2, theta1-1, theta2-1;
  Eigen::VectorXd upper4(4);
  upper4 << theta1+1, theta2+1, theta1, theta2;
  Eigen::VectorXd lower5(4);
  lower5 << theta1, theta2-1, theta1, theta2;
  Eigen::VectorXd upper5(4);
  upper5 << theta1+1, theta2, theta1+1, theta2+1;
  Eigen::VectorXd lower6(4);
  lower6 << theta1, theta2-1, theta1, theta2-1;
```

```
Eigen::VectorXd upper6(4);
upper6 << theta1+1, theta2, theta1+1, theta2;
Eigen::VectorXd lower7(4);
lower7 << theta1, theta2-1, theta1-1, theta2;
Eigen::VectorXd upper7(4);
upper7 << theta1+1, theta2, theta1, theta2+1;
Eigen::VectorXd lower8(4);
lower8 << theta1, theta2-1, theta1-1, theta2-1;
Eigen::VectorXd upper8(4);
upper8 << theta1+1, theta2, theta1, theta2;
Eigen::VectorXd lower9(4);
lower9 << theta1-1, theta2, theta1, theta2;
Eigen::VectorXd upper9(4);
upper9 << theta1, theta2+1, theta1+1, theta2+1;
Eigen::VectorXd lower10(4);
lower10 << theta1-1, theta2, theta1, theta2-1;
Eigen::VectorXd upper10(4);
upper10 << theta1, theta2+1, theta1+1, theta2;
Eigen::VectorXd lower11(4);
lower11 << theta1-1, theta2, theta1-1, theta2;
Eigen::VectorXd upper11(4);
upper11 << theta1, theta2+1, theta1, theta2+1;
Eigen::VectorXd lower12(4);
lower12 << theta1-1, theta2, theta1-1, theta2-1;
Eigen::VectorXd upper12(4);
upper12 << theta1, theta2+1, theta1, theta2;
Eigen::VectorXd lower13(4);
lower13 << theta1-1, theta2-1, theta1, theta2;
Eigen::VectorXd upper13(4);
upper13 << theta1, theta2, theta1+1, theta2+1;
Eigen::VectorXd lower14(4);
lower14 << theta1-1, theta2-1, theta1, theta2-1;
Eigen::VectorXd upper14(4);
upper14 << theta1, theta2, theta1+1, theta2;
Eigen::VectorXd lower15(4);
lower15 << theta1-1, theta2-1, theta1-1, theta2;
Eigen::VectorXd upper15(4);
upper15 << theta1, theta2, theta1, theta2+1;
Eigen::VectorXd lower16(4);
lower16 << theta1-1, theta2-1, theta1-1, theta2-1;
Eigen::VectorXd upper16(4);
upper16 << theta1, theta2, theta1, theta2;
double err_est;
int err_code;
int maxeval = evaluations;
const double res1 = integrate(f, lower1, upper1, err_est, err_code, maxeval);
const double res2 = integrate(f, lower2, upper2, err_est, err_code, maxeval);
```

```
  const double res3 = integrate(f, lower3, upper3, err_est, err_code, maxeval);
  const double res4 = integrate(f, lower4, upper4, err_est, err_code, maxeval);
  const double res5 = integrate(f, lower5, upper5, err_est, err_code, maxeval);
  const double res6 = integrate(f, lower6, upper6, err_est, err_code, maxeval);
  const double res7 = integrate(f, lower7, upper7, err_est, err_code, maxeval);
  const double res8 = integrate(f, lower8, upper8, err_est, err_code, maxeval);
  const double res9 = integrate(f, lower9, upper9, err_est, err_code, maxeval);
  const double res10 = integrate(f, lower10, upper10, err_est, err_code,
     maxeval);
  const double res11 = integrate(f, lower11, upper11, err_est, err_code,
     maxeval);
  const double res12 = integrate(f, lower12, upper12, err_est, err_code,
     maxeval);
  const double res13 = integrate(f, lower13, upper13, err_est, err_code,
     maxeval);
  const double res14 = integrate(f, lower14, upper14, err_est, err_code,
     maxeval);
  const double res15 = integrate(f, lower15, upper15, err_est, err_code,
     maxeval);
  const double res16 = integrate(f, lower16, upper16, err_est, err_code,
     maxeval);
  return Rcpp::List::create(
            Rcpp::Named("approximate") = res1 + res2 + res3 + res4
            + res5 + res6 + res7 + res8 + res9 + res10
            + res11 + res12 + res13 + res14 + res15 + res16,
            Rcpp::Named("error_estimate") = err_est,
            Rcpp::Named("error_code") = err_code
            );
}
// Test for convergence
class QuadLaplaceIntegrand: public MFunc
{
private:
  const double theta1;
  const double theta2;
  const double rho;
  const double m1;
  const double m2;
  const double sigma1;
  const double sigma2;
  double const1;
  double b1;
  double b2;

public:
  QuadLaplaceIntegrand(const double& theta1_,
            const double& theta2_,
```

```
          const double& m1_,
          const double& m2_,
          const double& sigma1_,
          const double& sigma2_,
          const double& rho_) : theta1(theta1_), theta2(theta2_),
                m1(m1_), m2(m2_), sigma1(sigma1_),
                sigma2(sigma2_), rho(rho_)
{
  const1 = std::sqrt(2*sigma1*sigma2*(1-rho*rho)
          + m1*m1*sigma2/sigma1 - 2*m1*m2*rho
          + m2*m2*sigma1/sigma2)
                   /(sigma1*sigma2*(1-rho*rho));
  b1 = -log(M_PI*sigma1*sigma2*std::sqrt(1-rho*rho));
  b2 = b1;
}

double operator()(Constvec& x)
{
  double s = (x[0] - theta1) / (1 - std::pow(x[0] - theta1, 2));
  double t = (x[1] - theta2) / (1 - std::pow(x[1] - theta2, 2));
  double u = (x[2] - theta1) / (1 - std::pow(x[2] - theta1, 2));
  double v = (x[3] - theta2) / (1 - std::pow(x[3] - theta2, 2));

  double c1 =((m1*sigma2/sigma1 -m2*rho)*(s) +
    (m2*sigma1/sigma2 - m1*rho)*(t)) /
             (sigma1*sigma2*(1-rho*rho));
  double d1 = log(gsl_sf_bessel_K0(const1*std::sqrt(std::pow(s,
     2)*sigma2/sigma1 -
                    2*rho*s * t +
                    std::pow(t, 2)*sigma1/sigma2)));
  double c2 =((m1*sigma2/sigma1 -m2*rho)*(u) +
       (m2*sigma1/sigma2 - m1*rho)*(v)) /
            (sigma1*sigma2*(1-rho*rho));
  double d2 = log(gsl_sf_bessel_K0(const1*std::sqrt(std::pow(u,
     2)*sigma2/sigma1 -
                    2*rho*u*v +
                    std::pow(v, 2)*sigma1/sigma2)));

  double e1 = log((1 + std::pow(x[0] - theta1, 2)) / std::pow(1 - std::pow(x[0]
     - theta1, 2), 2));
  double f1 = log((1 + std::pow(x[1] - theta2, 2)) / std::pow(1 - std::pow(x[1]
     - theta2, 2), 2));
  double e2 = log((1 + std::pow(x[2] - theta1, 2)) / std::pow(1 - std::pow(x[2]
     - theta1, 2), 2));
  double f2 = log((1 + std::pow(x[3] - theta2, 2)) / std::pow(1 - std::pow(x[3]
     - theta2, 2), 2));
  return std::exp(b1 + c1 + d1 + b2 + c2 + d2 + e1 + e2 + f1 + f2);
```

```
  }
};
```

Source code for multivariate asymmetric Laplace E-M Algorithm (Section 4.4.1). Source is written in C++ for the Rcpp interface to R.

```
// [[Rcpp::depends(RcppArmadillo)]]
// [[Rcpp::depends(BH)]]

#include <RcppArmadillo.h>
#include <math.h>
#include <boost/math/special_functions/bessel.hpp>
#include <RcppArmadilloExtensions/sample.h>


/*
The following need to be loaded in the R session:
library(Rcpp)
library(BH)
library(RcppArmadillo)
*/

// do not load the arma namespace. Some users report namespace
// conflicts when using Rcpp.
using namespace Rcpp;

// helper function for computing the modified Bessel Fn of the third
// kind. Also available in GSL. But Boost documentation is much
// easier to read so I'm using the implementation from the Boost
// Library for C++

double compute_bessel(double v, double x) {
    return boost::math::cyl_bessel_k(v, x);
}

// Provide the Initial estimate of dispersion (Sigma) matrix in the EM
// algorithm for Skew laplace Data.

arma::mat Sigma1(arma::mat dat,
      arma::vec theta,
      arma::vec m){
  int n = dat.n_rows;
  int d = dat.n_cols;
  arma::mat Sigma(d,d);
  Sigma.zeros();
  for(int i = 0; i < n; i++){
```

```
    arma::mat update = (dat.row(i) - m.t() - theta.t()).t() * (dat.row(i) - m.t()
        - theta.t()) / n;
    Sigma = Sigma + update;
  }
  return Sigma;
}

// Estimate the Generalize inverse Gaussians in the EM algorithm for
// Skew Laplace data A tolerance parameter is used to prevent
// overflow. Overflow occurs due to the unbound likelihood when
// approaching theta. Overflow is guaranteed to occur without a
// stopping rule.
arma::mat get_xi_eta (arma::mat dat,
           arma::vec theta,
           arma::vec m,
           arma::mat Sigma,
           double tol){
  int d = dat.n_cols;
  double d1 = double(d);
  int n = dat.n_rows;
  arma::mat gamma = m.t() * arma::inv_sympd(Sigma) * m;
  double gamma1 = sqrt(gamma(0,0) + 2);
  arma::vec qy(n);
  arma::vec e(n);
  arma::vec f(n);
  arma::vec g(n);
  arma::vec h(n);
  arma::vec xi(n);
  arma::vec eta(n);
  arma::mat result(n,2);
  for(int i = 0; i < n; i++){
    arma::mat update = (dat.row(i).t() - theta).t() * inv_sympd(Sigma) *
        (dat.row(i).t() - theta);
    double updated = update(0,0);
    qy(i) = sqrt(updated);
  }
  arma::vec delta = qy;
  arma::vec check = gamma1 * delta;
  if(any(check < tol)){
    return result.fill(arma::datum::inf);
  }
  double lambda = 1 - (d1/2);
  for (int i = 0; i < n; i++){
    e(i) = log(delta(i) / gamma1);
    f(i) = log(compute_bessel(lambda - 1, gamma1 * delta(i)));
    g(i) = log(compute_bessel(lambda + 1, gamma1 * delta(i)));
    h(i) = log(compute_bessel(lambda, gamma1 * delta(i)));
```

```
    xi(i) = exp(-e(i) + f(i) - h(i));
    eta(i) = exp(e(i) + g(i) - h(i));
  }
  result.col(0) = xi;
  result.col(1) = eta;
  return result;
}

// function to update the estimate of dispersion in the EM algorithm
// for Skew Laplace data
arma::mat Sigma_update(arma::mat dat,
            arma::vec theta,
            arma::vec m,
            arma::mat xi_eta){
  arma::vec xi = xi_eta.col(0);
  arma::vec eta = xi_eta.col(1);
  int d = dat.n_cols;
  int n = dat.n_rows;
  arma::mat Sigma(d,d);
  Sigma.fill(0);
  for(int i = 0; i < n; i++){
    arma::mat A = ((dat.row(i).t() - theta - eta(i) * m) * (dat.row(i).t() -
        theta - eta(i) * m).t()) / n;
    arma::mat A1 = xi(i) * A;
    Sigma = Sigma + A1;
  }
  return(Sigma);
}

// function to update the estimate of location in the EM algorithm
arma::vec theta_update(arma::mat dat,
            arma::mat xi_eta,
                     arma::vec m){
  int n = dat.n_rows;
  int d = dat.n_cols;
  arma::vec xi = xi_eta.col(0);
  arma::vec eta = xi_eta.col(1);
  arma::mat xi_dat(n,d);
  for (int i = 0; i < n; i++){
    xi_dat.row(i) = xi(i) * dat.row(i);
  }
  arma::vec numerator = (sum(xi_dat, 0)).t() - n * m;
  double denominator = sum(xi);
  arma::vec result(d);
  for (int i = 0; i < d; i++){
    result(i) = numerator(i) / denominator;
  }
```

```
  return result;
}

// function to update the estimate of skewness parameter in EM
// algorithm
arma::vec m_update(arma::mat dat,
        arma::vec theta,
        arma::mat xi_eta){
  int n = dat.n_rows;
  int d = dat.n_cols;
  arma::vec eta = xi_eta.col(1);
  double denominator = sum(eta);
  arma::vec result(d);
  arma::vec numerator = sum(dat, 0).t() - (n * theta);
  for(int i = 0; i < d; i++){
    result(i) = numerator(i) / denominator;
  }
  return result;
}

// function to obtain skew laplace parameter estimates
// This function is exposed to the R session.
// [[Rcpp::export]]
List aml_em(arma::mat dat,
       double tol1=0.000000000000001,
       double tol2=0.000000000000001){
  int n = dat.n_rows;
  int d = dat.n_cols;
  int N = 1000;
  arma::mat store_theta(N,d);
  arma::mat store_m(N,d);
  arma::vec theta_temp(d);
  arma::vec m_temp(d);
  arma::mat xi_eta_cur(n,2);
  arma::mat Sigma_temp(d,d);

  double test2_d = 0;
  double test1_d= 0;
  NumericVector xx(2);
  int iter = 0;
  for(int i = 0; i < N; i++){
    if(i == 0){
      store_theta.row(i) = mean(dat);
      store_m.row(i).fill(0);
      Sigma_temp = Sigma1(dat, store_theta.row(i).t(), store_m.row(i).t());
    }else{
      xi_eta_cur = get_xi_eta(dat,
```

```
            store_theta.row(i-1).t(),
            store_m.row(i-1).t(),
            Sigma_temp,
            tol2);
    if(xi_eta_cur(0,0) == arma::datum::inf){
  break;
    }
    theta_temp = theta_update(dat, xi_eta_cur, m_temp);
    store_theta.row(i) = theta_temp.t();
    m_temp = m_update(dat, theta_temp, xi_eta_cur);
    store_m.row(i) = m_temp.t();
    arma::mat test1 = (store_theta.row(i) - store_theta.row(i-1)) *
        (store_theta.row(i) - store_theta.row(i-1)).t();
    double test1_d = test1(0,0);
    arma::mat test2 = (store_m.row(i) - store_m.row(i-1)) * (store_m.row(i) -
        store_m.row(i-1)).t();
    double test2_d = test2(0,0);
    if ((test1_d < tol1) && (test2_d < tol1)){
  break;
    }
    iter = i;
    Sigma_temp = Sigma_update(dat, theta_temp, m_temp, xi_eta_cur);
   }
 }
 return Rcpp::List::create(Rcpp::Named("theta") = theta_temp,
           Rcpp::Named("m") = m_temp,
           Rcpp::Named("Sigma") = Sigma_temp,
           Rcpp::Named("iterations") = iter);
           //Rcpp::Named("alltheta") = store_theta.rows(1,iter), // sometimes
               generates an indices error
           //Rcpp::Named("allm") = store_m.rows(1,iter));
}
```