Laplace Transform Based Estimation

Mathematical Sciences School
Shinji Ogaki
Student ID: 4061253

November 16, 2008

Contents

Abstract

The empirical transform based estimator is an attractive alternative of the maximum likelihood estimator especially when the likelihood is unavailable. The most difficult part of using the empirical transform based estimator is the choice of the transform parameter since the performance is greatly influenced by the parameter. This paper addresses this issue. Also, the performance comparison with other estimators and a performance improvement method are given. Past studies on the empirical transform based estimator were done with the characteristic function. However, this paper focuses on the Laplace transform based estimator because of its simplicity.

1 Introduction

The maximum likelihood estimator(MLE) is used in a various fields because of its asymptotic efficiency. However, it is not available in some instances. For example, the parameters the stable law can not be estimated by the MLE. In the case of that the likelihood does not exist or the involved computational cost is considerably high, the empirical transform based estimator is an attractive alternative.

The idea of estimating parameter by empirical transforms was brought by Parzen[19] in 1962. Since then, many methods which exploit empirical transforms have been proposed (a good overview is found in [23] and [26]). The estimation of the parameters of the stable law is one of the important issues in many fields such as finance. In 1972, Press[21] introduced a simple approach which uses the e.c.f. A similar but more efficient procedure was introduced by Koutrouvelis[17]. His approach uses many transform parameters and estimates the parameters in a regression style. This is a standard method for the parameter estimation of the stable law. In 1975, Paulson[20] proposed the estimation method minimizing the weighted integral $|\phi(t) - \hat{\phi}(t)|^2 e^{-t^2}$, where $\phi(t)$ and $\hat{\phi}(t)$ are the theoretical and empirical characteristic function(e.c.f) and t is a transform parameter. Later, Heatcote[15] noted that the weight function e^{-t^2} is not an optimal in terms of performance and suggested a better weight function. Paulson's method is classified to the minimum distance method and this can be used for many applications since its usage is not limited to the stable law. A more flexible method is the generalized method of moments(GMM) estimator. The GMM was introduced by Hansen[13] to deal with over-identified models. Feuerverger and McDunnough[11] derived the estimator in the GMM framework and claimed the GMM estimator is asymptotically efficient if the transform variables are chosen carefully. However, Carrasco[6] found it is not true because of a singularity. It arises when the the transform parameters are too close or too many. She introduced the regularization to avoid the singularity problem and showed the GMM estimation can be obtain a higher performance by considering a continuum of moment conditions (see detail [6] and [7]).

The use of the empirical transform based estimator is not easy. The efficiency of the estimator is significantly affected by the choice of the transform parameter, therefore, the parameter must be chosen so that the asymptotic variance of the estimator is minimized. This is the most difficult part of using the empirical transform based estimator since the asymptotic distribution involves the parameter itself in the general case. Furthermore, the parameter has to be determined before estimating. To address this issue Ball and Milne[3] introduced the adaptive estimator. Another similar approach is the empirical minimum variance estimator proposed by Yao and Morgan(see detail [3]). Both of the regression-type estimator and discrete GMM estimator use finite number of transform variables. For these methods, a proper choice of the number of transform parameters and their location is a key. Besbeas[5] considered the optimal choice of the transform variables to improve the performance of the Koutrouvelis method. For the minimum distance method, the problem is the choice of a weight function.

The purpose of this paper is to introduce the Laplace transform based estimator and its asymptotic property and to present a method which improves the performance of the estimator. We also describe valuable methods such as the adaptive estimator and briefly survey the methods for the stable law. For every method, the implementation detail and illustrative example are shown and a performance comparison is also given.

Numerous researchers have proposed the use of the e.c.f. whilst there are a few studies of the Laplace transform based estimation. Our study is limited to the Laplace transform based estimator because it is simpler than the e.c.f. based approach. Additionally, the results from the study on the e.c.f based estimator can be directly applied to the Laplace transform based estimator.

This paper is organized as follows. Section 2 introduces the Laplace transform based estimator and its asymptotic distribution. The adaptive estimator and the empirical minimum variance estimator are also introduced in this section. Section 3 compares its performance with that of MLE and also briefly surveys methods used for the stable law. Section 4 presents the method to improve the performance by exploiting the GMM. Section 5 concludes.

2 Laplace Transform Based Estimator and its Asymptotic Distribution

In this section the Laplace transform based estimator and its asymptotic distribution are introduced. The estimator is derived by equating the theoretical and empirical Laplace transform. We start by dealing with a statistical model with a single parameter. Then this idea can be extended easily to a multiple-parameter model. The efficiency of the estimator depends on the Laplace transform parameter t. The optimal value of t gives the most efficient estimator and can be obtained by minimizing the variance of the asymptotic distribution of the estimator. This is the main topic of this chapter. As its application, the scale and location problem will be discussed. In most cases the asymptotic distribution involves the parameter itself. Since we do not know the true value of the parameter in a practical situation, we cannot obtain the value of t which gives the most efficient estimator directory. At the end of this chapter, two methods of overcoming this problem are given.

2.1 Definitions of Laplace Transform and Empirical Laplace Transform

Here, the definitions of Laplace transformation and empirical Laplace transformation are given. There are a couple of small differences with the standard definition of Laplace transformation, which enable us to handle every kind of statistical models. Let t be the parameter of Laplace transformation. The form of the Laplace transformation used in this paper is given as

$$M_\theta(t) = \int_{-\infty}^{\infty} e^{-tx} f(x|\theta)dx, \tag{2.1}$$

.

where $f(x|\theta)$ is the probability distribution function of the random variable x and θ is a parameter. By the standard definition of Laplace transformation, integration is done from 0 to ∞ and t only takes a positive value. However, we use the above formula to handle all kinds of distribution. Moreover, we also allow negative values for t. Let X be a random variable and $x_1, x_2, \cdots, x_n$ be n observations. The empirical Laplace transformation is determined as

$$M_n(t) = \frac{1}{n}\sum_{i=1}^{n} e^{-tx_i}. \tag{2.2}$$

Throughout the paper, the above notations, $M_\theta(t)$ and $M_n(t)$ are used.

2.2 Single Parameter Case

2.2.1 Estimator

The derivation of the Laplace transform based estimator for a single parameter model is as follows. Let θ be a parameter of a statistical model and $\hat{\theta}$ be the estimator of θ. By letting $M_\theta(t) = M_n(t)$ and solving this formula for the parameter θ, $\hat{\theta}$ can be obtained. Assuming $M_\theta(t) = g_t(\theta)$ has an unique solution, the estimator is given as

$$\hat{\theta} = g_t^{-1}(M_n(t)). \tag{2.3}$$

An illustrative example is given below.

Example - Exponential Distribution

We show how to derive the estimator by taking the Exponential distribution $NE(\theta)$ as an example. The pdf of the Exponential distribution is

$$NE(x|\theta) = \theta \exp(-\theta x).$$

The theoretical Laplace transformation is given as

$$M_\theta(t) = \int_{-\infty}^{0} e^{-tx} \theta e^{-\theta x} dx = \frac{\theta}{\theta + t}.$$

Letting $M_\theta(t) = M_n(t)$ yields

$$\frac{\theta}{\theta + t} = M_n(t). \tag{2.4}$$

Solving this for θ yields

$$\hat{\theta} = \frac{tM_n(t)}{1 - M_n(t)}. \tag{2.5}$$

2.2.2 Asymptotic Distribution

In order to evaluate the performance of the estimator, it is crucial to consider its asymptotic distribution. This is easily derived by the Delta method. The derivation is as follows. Let $X_1, X_2, X_3, \cdots$ be independent and identically distributed random variables. The central limit theorem states as $n \to \infty$,

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2), \tag{2.6}$$

where $\mu = E[X_1]$ and $\sigma^2 = \mathrm{Var}[X_1]$. The Taylor expansion of the function $f(x)$ around μ is

$$f(x) = f(\mu) + f^{'}(\mu)(x - \mu) + \frac{1}{2} f^{''}(\mu)(x - \mu)^2 + \cdots . \tag{2.7}$$

Ignoring higher order terms, we can approximate $f(x) - f(\mu)$ as

$$f(x) - f(\mu) \simeq f^{'}(\mu)(x-\mu).$$

Replacing x with $\bar{X}$ and multiplying $\frac{\sqrt{n}}{\sigma}$ both sides yields

$$\frac{\sqrt{n}\left(f(\bar{X}) - f(\mu)\right)}{\sigma} \simeq \frac{\sqrt{n}\left(\bar{X}-\mu\right)}{\sigma} f^{'}(\mu).$$

Due to the central limit theorem the right hand side of the above equation converges to $N(0, f^{'}(\mu)^2)$ as $n \to \infty$. Hence

$$f(\bar{X}) \xrightarrow{D} N\left(f(\mu), \frac{1}{n}\sigma^2 f^{'}(\mu)^2\right). \tag{2.8}$$

This method is known as the univariate delta method (see [1]).

Let $X_i = e^{-x_i t}$ (Recall Equation (2.2)). Then $\bar{X} = M_n(t)$, $\mu = E[X] = M_\theta(t)$ and $\sigma^2 = \mathrm{Var}[X] = M(2t) - M(t)^2$. Furthermore, let $\mu = g_t(\theta)$ and f_t be the inverse function of g_t. From Equation (2.3) $\hat{\theta} = f_t(\bar{X})$ and clearly $\theta = f_t(\mu)$. By the inverse function theorem, $f_t^{'} = (g_t^{'})^{-1}$. Hence, the asymptotic distribution of $\hat{\theta}$ can be expressed as

$$\hat{\theta} \xrightarrow{D} N\left(\theta, \frac{1}{n}\frac{(M(2t) - M(t)^2)}{g_t^{'}(\theta)^2}\right). \tag{2.9}$$

We also show the multivariate delta method since it is required in later subsections. The Taylor expansion of the multivariate function $f(\boldsymbol{x})$ around $\boldsymbol{\mu}$ is

$$f(\boldsymbol{x}) = f(\boldsymbol{\mu}) + \nabla f(\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \nabla^2 f(\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu}) + \cdots.$$

Ignoring the higher order terms, $f(\boldsymbol{x}) - f(\boldsymbol{\mu})$ can be approximated as

$$f(\boldsymbol{x}) - f(\boldsymbol{\mu}) \simeq \nabla f(\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu}).$$

The variance of $f(\boldsymbol{x})$ is given as

$$\begin{aligned}
\mathrm{Var}[f(\boldsymbol{x})] &\simeq \mathrm{Var}[f(\boldsymbol{\mu}) + \nabla f(\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})] \\
&= \mathrm{Var}[\nabla f(\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})] \\
&= \nabla f(\boldsymbol{\mu})^T \mathrm{Var}[\boldsymbol{x}-\boldsymbol{\mu}] \nabla f(\boldsymbol{\mu}) \\
&= \nabla f(\boldsymbol{\mu})^T \mathrm{Var}[\boldsymbol{x}] \nabla f(\boldsymbol{\mu}).
\end{aligned}$$

Therefore,

$$\sqrt{n}\left(f(\boldsymbol{x}) - f(\boldsymbol{\mu})\right) \simeq \sqrt{n}\left(\nabla f(\boldsymbol{\mu})^T(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

Due to the central limit theorem, the right hand side of the above equation converges to

$N(0, n\nabla f(\boldsymbol{\mu})^T \mathrm{Var}[\boldsymbol{x}]\nabla f(\boldsymbol{\mu}))$ as $n \to \infty$. Replacing $\boldsymbol{x}$ with $\bar{\boldsymbol{X}}$ yields

$$f(\bar{\boldsymbol{X}}) \xrightarrow{D} N\left(f(\boldsymbol{\mu}), \frac{1}{n}\nabla f(\boldsymbol{\mu})^T \mathrm{Var}[\boldsymbol{x}]\nabla f(\boldsymbol{\mu})\right). \tag{2.10}$$

Example - Gamma Distribution

Here, we show how to derive the estimator and its asymptotic distribution of the shape parameter k of the Gamma distribution $Gamma(k, \theta)$ by the Laplace transform based estimation. Later, we will see the MLE is slightly complicated to calculate. The pdf of the Gamma distribution is

$$Gamma(x|k, \theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}\theta^{-k}}{\Gamma(k)}.$$

Let $X_0 \sim Gamma(k, \theta)$. Assuming the scale parameter θ is known, the estimator of the shape parameter $\hat{k}$ is derived as follows. We have

$$\begin{aligned} M(t) &= \int_{-\infty}^{\infty} \frac{x^{k-1}e^{-\frac{x}{\theta}}\theta^{-k}}{\Gamma(k)} e^{-tx} dx \\ &= (1+\theta t)^{-k}. \end{aligned}$$

Letting $M(t) = M_n(t)$ and taking logarithm of both sides gives

$$\hat{k} = -\frac{\log(M_n(t))}{\log(1+\theta t)}.$$

Let define $h(x)$ as

$$h(x) = -\frac{\log(x)}{\log(1+\theta t)}.$$

Hence, the asymptotic distribution is given as

$$\hat{k} \xrightarrow{D} N\left(k, \frac{1}{n}(\log(1+\theta t)(1+\theta t)^{-k})^{-2}((1+2\theta t)^{-k} - (1+\theta t)^{-2k})\right).$$

On the other hand, the MLE does not give a simple form. The estimator of the shape parameter is given as

$$\hat{k} = \mathrm{Digamma}^{-1}\left(\frac{\sum_{i=1}^{N}\log(x_i)}{N} - \log(\theta)\right). \tag{2.11}$$

The derivation of Equation (2.11) is as follows. Likelihood is given as

$$L = \prod_{i=1}^{n} \frac{x_i^{k-1} e^{-\frac{x_i}{\theta}} \theta^{-k}}{\Gamma(k)}.$$

Taking logarithm

$$l = \log(L) = k \sum_{i=1}^{n} \log(x_i) - nk\log(\theta) - n\log(\Gamma(k)) + C.$$

Differentiating the above equation with respect to k,

$$\frac{\partial l}{\partial k} = nk\log(\theta) - n\text{Digamma}(k),$$

where C is a constant and $\text{Digamma}(k) = \frac{\Gamma'(k)}{\Gamma(k)}$. Letting $\frac{\partial l}{\partial k} = 0$ yields

$$\text{Digamma}(\hat{k}) = \frac{\sum_{i=1}^{n} \log(x_i)}{n} - \log(\theta).$$

Therefore,

$$\hat{k} = \text{Digamma}^{-1}\left(\frac{\sum_{i=1}^{N} \log(x_i)}{N} - \log(\theta)\right).$$

The MLE is not easy to calculate.

2.3 Location and Scale Problem

In this subsection we consider the location and scale problem. Let X_0 be a random variable with pdf $f(x)$ and $X = \phi X_0 + \omega$. ϕ is a scale parameter and ω is a location parameter. The location and scale Problem is to estimate the value of location and scale parameters. Three cases are shown. The first is the location is unknown. The second is the scale is unknown. The last is both are unknown.

2.3.1 Location is Unknown

If only the location parameter is unknown, the problem is to estimate ω of the following formula.

$$X = X_0 + \omega.$$

Note that

$$M(t) = e^{-t\omega} M_0(t).$$

where $M_0(t)$ is the theoretical Laplace transformation of X_0. Letting $M(t) = M_n(t)$ yields

$$\hat{\omega} = \frac{1}{t}\left(\log(M_0(t) - \log(M_n(t)))\right).$$

The asymptotic distribution of ω is given as

$$\hat{\omega} \xrightarrow{D} N\left(\omega, \frac{M_0(2t) - M_0^2(t)}{t^2 M_0^2(t)}\right). \tag{2.12}$$

By solving the following formula for t, the most effective estimator which minimizes the asymptotic variance $\frac{M_0(2t)-M_0^2(t)}{t^2 M_0^2(t)}$ can be obtained.

$$M_0(t)\left(M_0(2t) - M_0^2(t)\right) = t\left(M_0^{'}(2t)M_0(t) - M_0(2t)M_0^{'}(t)\right).$$

More detail can be found in [3].

2.3.2 Scale is Unknown

Similarly, if only the scale parameter is unknown, the problem is to estimate ϕ of the following formula

$$X = \phi X_0.$$

Note that

$$M(t) = M_0(\phi t).$$

Letting $M(t) = M_n(t)$ yields

$$\hat{\phi} = \frac{1}{t} M_0^{-1}(M_n(t)). \tag{2.13}$$

The asymptotic distribution of ω is given as

$$\hat{\phi} \xrightarrow{D} N\left(\phi, \frac{M_0(2\phi t) - M_0^2(\phi t)}{t^2 M_0^{'2}(\phi t)}\right). \tag{2.14}$$

The value of t which minimizes the variance is given as the solution of the following formula

$$\phi t M_0^{'}(\phi t)\left(M_0^{'}(2\phi t) - M_0(\phi t)M_0^{'}(\phi t)\right) = \left(M_0(2\phi t) - M_0^2(\phi t)\right)\left(M_0^{'}(\phi t) + \phi t M_0^{''}(\phi t)\right). \tag{2.15}$$

More detail can be found in [3].

Example - Normal Scale Problem

Let X_0 be a random variable which follows the Normal distribution with mean 0 and variance θ^2 (i.e. $X_0 \sim N(0, \theta^2)$). $M_0(t)$ is given as

$$M_0(t) = \exp\left(\frac{t^2\theta^2}{2}\right). \tag{2.16}$$

Letting $M_0(t) = M_n(t)$ yields

$$\hat{\theta} = \frac{\sqrt{2\log M_n(t)}}{|t|}.$$

The variance V is

$$V = \frac{M_0(2t) - M_0^2(t)}{t^2 M_0'^2(t)}.$$

The value of t which minimizes V is given as

$$t = \pm\frac{\sqrt{2 + \mathrm{W}(-2e^{-}2)}}{\theta}. \tag{2.17}$$

The function W is the Lambert W function. It is defined as the inverse function of $f(W) = We^W$. It is also known as the Omega function. The derivation of Equation (2.17) is as follows. Substituting Equation (2.16) into Equation (2.15),

$$2\theta^2 t^2 \exp(\theta^2 t^2) - \theta^2 t^2 = \left(\exp(\theta^2 t^2) - 1)(2 + \theta^2 t^2\right).$$

Let $x = \theta^2 t^2$, then we have

$$(x-2)\exp(x) = -2.$$

Multiplying $\exp(-2)$ both sides,

$$(x-2)\exp(x-2) = -2\exp(-2).$$

It can be rewritten by using the Lambert W function,

$$(x-2) = \mathrm{W}(-2\exp(-2)),$$

which is equivalent to Equation (2.17).

Example - Uniform Scale Problem

Let X_0 be a random variable which follow the Uniform distribution with the parameter θ (i.e. $X_0 \sim U(0,\theta)$). $M_0(t)$ is given as

$$M_0(t) = \frac{1}{\theta t}\left(1 - e^{\theta t}\right).$$

Letting $M_0(t) = M_n(t)$ yields

$$\hat{\theta} = \frac{1}{t}\left(\mathrm{W}\left(\frac{-1}{M_n(t)}\exp\left(\frac{-1}{M_n(t)}\right)\right) + \frac{1}{M_n(t)}\right).$$

The variance V of the asymptotic distribution is given as

$$V = \frac{\theta t(1-e^{-2\theta t}) - 2(1-e^{-\theta t})^2}{2(e^{-\theta t}(\theta t+1)-1)}.$$

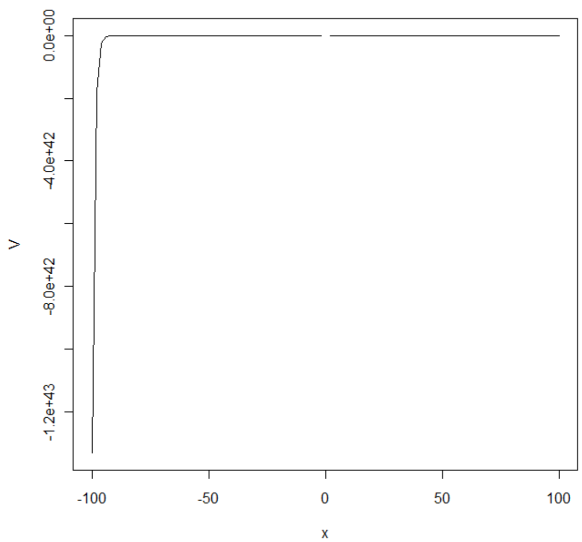

Figure 2.1: Plot of V

Figure 2.1 is a plot of V against $x = \theta t$. As this graph suggests, $t \to -\infty$ minimizes V.

2.3.3 Joint Estimation of Location and Scale

Let X_0 be a random variable with the probability distribution function $f(\theta)$ and X be a random variable defined as $X = \phi X_0 + \omega$. Now we show how to estimate ϕ and ω simultaneously. To obtain two unknowns, we need two Laplace transform parameters, say t_1 and t_2.

$$\begin{array}{l} \exp(-\omega t_1)M_0(\phi t_1) = M_n(t_1) \\ \exp(-\omega t_2)M_0(\phi t_2) = M_n(t_2) \end{array}. \tag{2.18}$$

The solution can be obtained as follows. Taking logarithm of both sides of Equation (2.18) gives

$$-\omega t_1 = \log \frac{M_n(t_1)}{M_0(\phi t_1)}$$
$$-\omega t_2 = \log \frac{M_n(t_2)}{M_0(\phi t_2)}.$$

Removing θ of the above two equations, we have

$$t_1 \log M_0(\phi t_2) - t_2 \log M_0(\phi t_1) = t_1 \log M_n(t_2) - t_2 \log M_n(t_1).$$

Let $f(\phi) = \frac{M_0(\phi t_1)^{t_2}}{M_0(\phi t_2)^{t_1}}$. Then

$$\hat{\phi} = f^{-1}\left(\frac{M_n(t_1)^{t_2}}{M_n(t_2)^{t_1}}\right).$$

Substituting $\hat{\phi}$ into Equation (2.18) yields

$$\hat{\omega} = \frac{1}{t_1} \log \frac{M_0(\hat{\phi} t_1)}{M_n(t_1)}.$$

By applying the multivariate delta method given in Equation (2.10), the asymptotic distribution of the estimates of these two variables are given as

$$\begin{pmatrix} \hat{\omega} \\ \hat{\phi} \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} \omega \\ \phi \end{pmatrix}, \frac{1}{n} H \Sigma H^T\right),$$

where

$$H = J^{-1}.$$

J is a Jacobian matrix which is given as

$$J = \begin{pmatrix} \frac{\partial M(t_1)}{\partial \omega} & \frac{\partial M(t_1)}{\partial \phi} \\ \frac{\partial M(t_2)}{\partial \omega} & \frac{\partial M(t_2)}{\partial \phi} \end{pmatrix}. \tag{2.19}$$

The covariance matrix is given as

$$\Sigma = \begin{pmatrix} E_{11} - E_1 E_1 & E_{12} - E_1 E_2 \\ E_{12} - E_1 E_2 & E_{22} - E_2 E_2 \end{pmatrix}, \tag{2.20}$$

where

$$E_i = e^{-t_i \omega} M_0(\phi t_i)$$
$$E_{ij} = e^{-(t_i + t_j)\omega} M_0(\phi(t_i + t_j)),$$

for $i = 1, 2$ and $j = 1, 2$.

There are a couple of methods to obtain good estimator. The determinant of the matrix $H\Sigma H^T$ has to be minimized. The determinant is given as

$$|H\Sigma H^T| = |\Sigma||J|^{-2}.$$

The right hand side of the above equation can be expressed as

$$\frac{\left(M_0(2\phi t_1) - M_0^2(\phi t_1)\right)\left(M_0(2\phi t_2) - M_0^2(\phi t_2)\right) - \left(M_0(\phi(t_1+t_2)) - M_0(\phi t_1)M_0(\phi t_2)\right)^2}{t_1^2 t_2^2\left(M_0(\phi t_1)M_0'(\phi t_2) - M_0(\phi t_2)M_0'(\phi t_1)\right)^2}. \tag{2.21}$$

which is independent of location. We need to minimize the absolute value of the determinant to derive the most effective estimator.

Example - Exponential Distribution

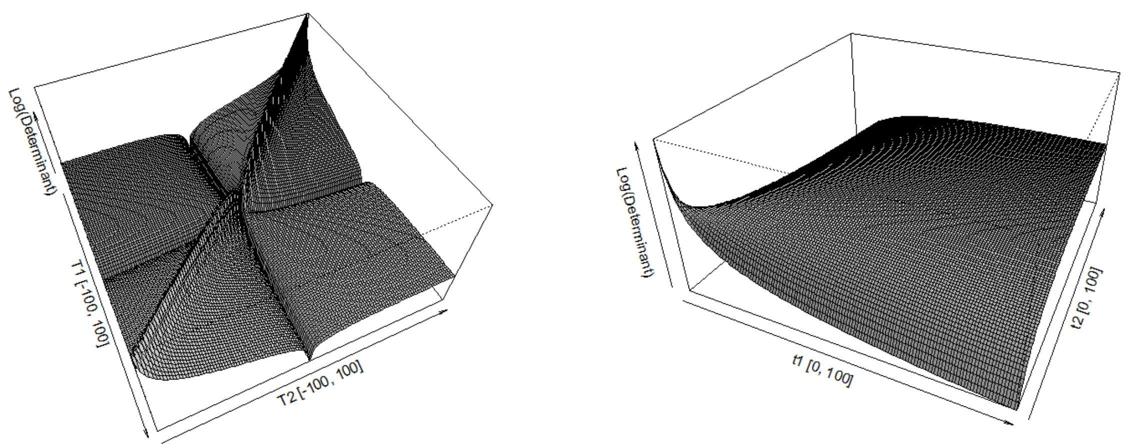

Figure 2.2: Logarithm of the absolute value of the Determinant against t1 and t2. Exponential distribution. Four quadrants (Left) and the first quadrants(Right). $\theta = 1$.

Let X_0 be a random variable which follows the Exponential distribution with the parameter θ (i.e. $X_0 \sim NE(\theta)$). Let $X = X_0 + \omega$. The estimates of θ and ω are obtained by solving

$$
\begin{aligned}
M_n(t_1) &= e^{-\omega t_1}\frac{\theta}{\theta+t_1}\\
M_n(t_2) &= e^{-\omega t_2}\frac{\theta}{\theta+t_2}.
\end{aligned}
$$

The solution can be obtained by a numerical method such as the Newton-Raphson method. From Equation (2.21), the determinant of the covariance matrix of the asymptotic distribution is given as

$$
\frac{(\theta+t_1)^2(\theta+t_2)^2}{\theta^2(\theta+2t_1)(\theta+2t_2)(\theta+t_1+t_2)^2}.
$$

It is obvious that $t_1 = -\theta$ or $t_2 = -\theta$ make this quantity 0. However, if $t_i = -\theta$, $M_0(t_i)$ cannot be determined (for $i = 1, 2$). Now we show $t_1 \to 0, t_2 \to \infty$ minimizes this quantity. By dividing the numerator and denominator by t_2^3, we have

$$
\frac{\frac{(\theta+t_1)^2}{t_2}\left(\frac{\theta}{t_2}+1\right)^2}{\theta^2(\theta+2t_1)\left(\frac{\theta}{t_2}+2\right)\left(\frac{\theta+t_1}{t_2}+1\right)^2}.
$$

It clearly goes 0 as $t_1 \to 0, t_2 \to \infty$. The same argument holds for $t_2 \to 0, t_1 \to \infty$. Figure 2.2 shows the 3D plot of the determinant. This also suggests $t_1 \to 0, t_2 \to \infty$ or $t_2 \to 0, t_1 \to \infty$ minimizes the determinant.

2.4 Diagonal Optimization

Here we introduce the idea of the diagonal optimization by considering the joint estimation of location and scale parameters. For some distributions the most efficient estimator is given when t_1 and t_2 are approximately the same. Suppose $t2 \simeq t_1$. Let $t = t2 \simeq t1$. The determinant of the covariance matrix of the asymptotic distribution can be expressed as a function of t and the problem of choosing the transform variables becomes a simple line search problem. This is called the diagonal optimization (see detail [25]). In this case, the estimates can be obtained in the following way.

By differentiating both sides of $\exp(-\omega t)M_0(\phi t) = M_n(t)$, we have

$$
\frac{\partial}{\partial t}\{\exp(-\omega t)M_0(\phi t)\} = \frac{\partial}{\partial t}M_n(t). \tag{2.22}
$$

Clearly

$$
\omega = \frac{1}{t}\log\frac{M_0(\phi t)}{M_n(t)}.
$$

By substituting this into Equation (2.22), we have

$$\phi \frac{M_0'(\phi t)}{M_0(\phi t)} + \frac{1}{t} \log \frac{M_n(t)}{M_0(\phi t)} = \frac{M_n'(t)}{M_n(t)}.$$

Let

$$f(\phi) = \phi \frac{M_0'(\phi t)}{M_0(\phi t)} + \frac{1}{t} \log \frac{M_n(t)}{M_0(\phi t)}.$$

Then

$$\hat{\phi} = f^{-1} \left(\frac{M_n'(t)}{M_n(t)} \right)$$

and

$$\hat{\omega} = \frac{1}{t} \log \frac{M_0(\hat{\phi} t)}{M_n(t)}.$$

Now one example is examined in detail.

Example - Normal Distribution

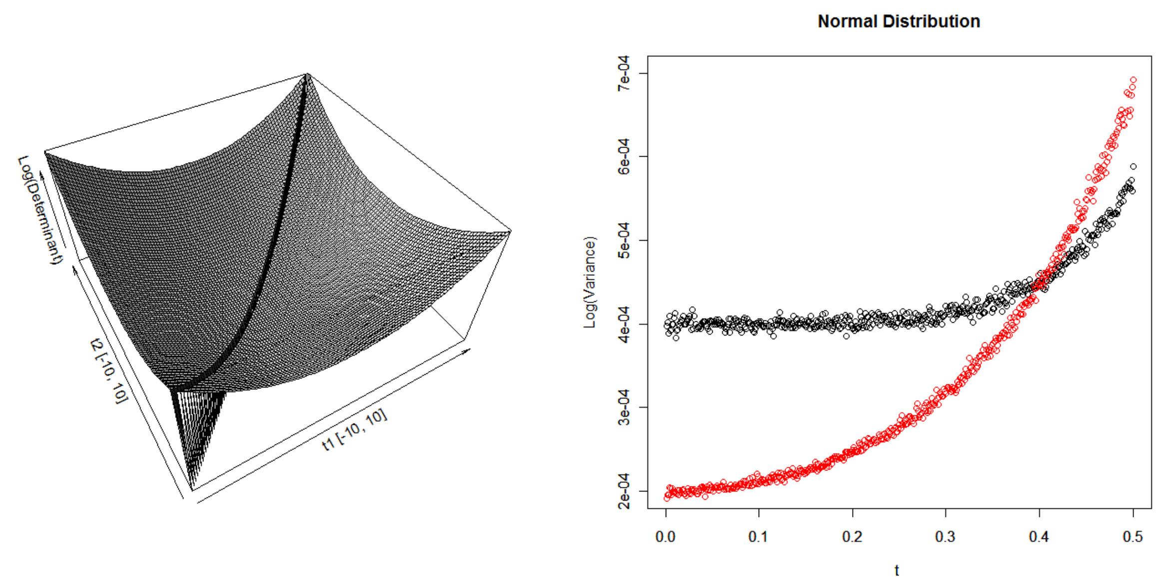

Figure 2.3: Left: Logarithm of the Determinant against t1 and t2. Normal distribution. Note that determinant is not determined on the line $t_1 = t_2$. Right: The sample variance of $\hat{\mu}$ and $\hat{\theta}$ against t. $N = 10000$ and 10000 samples are used. $\mu = 1$ and $\theta = 2$. t ranges from 0 to 0.5.

Let X_0 be a random variable which follows the Normal distribution with mean μ and variance σ^2 (i.e. $X_0 \sim N(\mu, \sigma^2)$). As $t_1 \to t_2$, we can obtain the estimators by solving the following formula

$$\begin{aligned} M_n(t) &= \exp\left(-\mu t + \frac{\sigma^2 t^2}{2}\right) \\ M_n^{'}(t) &= (-\mu + \sigma^2 t) \exp\left(-\mu t + \frac{\sigma^2 t^2}{2}\right). \end{aligned}$$

The solution is given as

$$\hat{\sigma} = \sqrt{\frac{2}{t}(\frac{M_n'(t)}{M_n(t)} - \frac{1}{t}\log M_n(t))}$$

and

$$\hat{\mu} = \frac{1}{t}\left(\frac{\hat{\sigma}^2 t^2}{2} - \log M_n(t)\right).$$

From Equation (2.21), the determinant of the covariance matrix of the asymptotic distribution is given as

$$|H\Sigma H^T| = \frac{(exp(\sigma^2 t_1^2) - 1)(\exp(\sigma^2 t_2^2) - 1) - (\exp(\sigma^2 t_1 t_2) - 1)^2}{\sigma^2 t_1^2 t_2^2 (t_2 - t_1)^2}.$$

When $t_1 \simeq t_2$, by applying l'Hospital's rule twice, we obtain

$$\begin{aligned} \lim_{t_1 \to t_2} |H\Sigma H^T| &= \lim_{t_1 \to t_2} \frac{(exp(\sigma^2 t_1^2) - 1)(\exp(\sigma^2 t_2^2) - 1) - (\exp(\sigma^2 t_1 t_2) - 1)^2}{\sigma^2 t_1^2 t_2^2 (t_2 - t_1)^2} \\ &= \lim_{t_1 \to t_2} \frac{t_1 \exp(\sigma^2 t_1^2)(\exp(\sigma^2 t_2^2) - 1) - t_2 \exp(\sigma^2 t_1 t_2)(\exp(\sigma^2 t_1 t_2) - 1)}{t_1 t_2^2 (t_2 - t_1)(t_2 - 2t_1)} \\ &= \lim_{t_1 \to t_2} \frac{(\exp(\sigma^2 t_2^2) - 1)\exp(\sigma^2 t_1^2)(1 + 2t_1^2 \sigma^2) - \sigma^2 t_2^2 \exp(\sigma^2 t_1 t_2)(2\exp(\sigma^2 t_1 t_2) - 1)}{t_2^2 (t_2^2 - 6t_1 t_2 + 6t_1^2)} \\ &= \frac{\exp(\sigma^2 t_2^2)\left(\exp(\sigma^2 t_2^2) - 1 - \sigma^2 t_2^2\right)}{t_2^4}. \end{aligned}$$

Let $t = t_2 \simeq t_1$. The problem of choosing the transform parameters t_1 and t_2 becomes a simple one dimensional search. Now we show $t \to 0$ minimizes $|H\Sigma H^T|$. By letting $x = \sigma^2 t^2$, $|H\Sigma H^T|$ can be expressed as

$$|H\Sigma H^T| = \frac{1}{\sigma^4} \times \frac{\exp(2x) - \exp(x) - x\exp(x)}{x^2}$$

and

$$\frac{d}{dx}\left(\frac{\exp(2x)-\exp(x)-x\exp(x)}{x^2}\right)=\frac{x\exp(x)\left(2x\exp(x)-2\exp(x)+2-x^2\right)}{x^4}.$$

Clearly $\frac{x\exp(x)}{x^4}>0$ for $x>0$. Denote $f(x)=2x\exp(x)-2\exp(x)+2-x^2$. $f(0)=0$ and $f^{'}(x)=2x(\exp(x)-1)>0$ for $x>0$, which means $f(x)$ is a monotonic increasing function. Therefore, $t_2\rightarrow 0$ minimizes $|H\Sigma H^T|$. Thus, the estimator is most effective as $t_1\rightarrow 0$ and $t_2\rightarrow 0$.

Figure 2.3 indicates that the values of $t\rightarrow +0$ minimizes the variance of μ and σ. Figure 2.3 shows the Mean Square Error of the estimates obtained with this method against t. This also suggests $t\rightarrow 0$ minimizes $|H\Sigma H^T|$.

2.5 Multiple-Parameter Case

In this subsection we deal with a statistical model with multiple parameters. The method of the estimation and derivation of the asymptotic distribution is almost the same with that of the joint estimation of location and scale parameters. Let $\boldsymbol{\theta}$ be a vector which represents all unknown parameters and p be the number of parameters. Let

$$\begin{array}{rcl} M_{\boldsymbol{\theta}}(t_1) & = & g_{t_1}(\boldsymbol{\theta}) \\ & \vdots & \\ M_{\boldsymbol{\theta}}(t_p) & = & g_{t_p}(\boldsymbol{\theta}) \end{array}.$$

The estimator of $\boldsymbol{\theta}$ is given by solving the following formula.

$$\begin{array}{rcl} g_{t_1}(\boldsymbol{\theta}) & = & M_n(t_1) \\ & \vdots & \\ g_{t_p}(\boldsymbol{\theta}) & = & M_n(t_p) \end{array}.$$

The asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{\theta}}\xrightarrow{D} N\left(\boldsymbol{\theta},\frac{1}{n}H\Sigma H^T\right).$$

H is the inverse matrix of a Jacobian J which is given as

$$J=\begin{pmatrix} \frac{\partial M(t_1)}{\partial\theta_1} & \cdots & \frac{\partial M(t_1)}{\partial\theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial M(t_p)}{\partial\theta_1} & \cdots & \frac{\partial M(t_p)}{\partial\theta_p} \end{pmatrix}. \tag{2.23}$$

Σ is a $p\times p$ covariance matrix. The ij-element of Σ is given as $\Sigma_{ij}=E_{11}-E_iE_j$. $E_i=M(t_i)$ and $E_{ij}=M(t_i+t_j)$. Similarly, we the most effective estimator is obtained by minimizing the determinant of the covariance of the asymptotic distribution. However, it is rather hard to show for the general case.

Example - Gamma Distribution

Here the estimator of the location, scale and shape parameters of the Gamma distribution and their asymptotic distribution are shown. Let X_0 be a random variable which follows the Gamma distribution with the shape parameter k and scale parameter θ (i.e. $X_0 \sim Gamma(k, \theta)$) and $X = X_0 + \omega$. Then we have

$$M(t_i) = e^{-t_i \omega}(1 + t_i \theta)^{-k}.$$

for $i = 1, 2, 3$.
The estimates are given by solving the following equations

$$M_n(t_i) = M(t_i).$$

This can be solved by a numerical method such as Newton-Raphson Method. The asymptotic distribution of k, ω and ϕ is given as

$$J = \begin{pmatrix} \frac{\partial M(t_1)}{\partial k} & \frac{\partial M(t_1)}{\partial \omega} & \frac{\partial M(t_1)}{\partial \phi} \\ \frac{\partial M(t_2)}{\partial k} & \frac{\partial M(t_2)}{\partial \omega} & \frac{\partial M(t_2)}{\partial \phi} \\ \frac{\partial M(t_3)}{\partial k} & \frac{\partial M(t_3)}{\partial \omega} & \frac{\partial M(t_3)}{\partial \phi} \end{pmatrix}. \tag{2.24}$$

$$\Sigma = \begin{pmatrix} E_{11} - E_1 E_1 & E_{12} - E_1 E_2 & E_{31} - E_1 E_3 \\ E_{12} - E_2 E_1 & E_{22} - E_2 E_2 & E_{23} - E_2 E_3 \\ E_{31} - E_3 E_1 & E_{23} - E_3 E_2 & E_{33} - E_3 E_3 \end{pmatrix}, \tag{2.25}$$

where

$$\begin{array}{ll} E_i & = e^{-t_1 \omega}(1 + t_i \theta)^{-k} \\ E_{ij} & = e^{-(t_i + t_j)\omega}(1 + (t_i + t_j)\theta)^{-k} \end{array}.$$

2.6 Adaptive Estimator

In the previous subsection, the Laplace transform based estimator and and its asymptotic distribution were derived. If the true value of the parameter is known, the optimal value of the Laplace transformation parameter t can be obtained easily. It is, however, realistically impossible. Hence, we need a method to estimate the optimal value of t. One solution of doing this is the adaptive estimator of Ball and Milne[3].

Example - Normal Scale Problem

Let X_0 be a random variable which follows the Normal distribution with mean 0 and variance σ^2 (i.e. $X_0 \sim N(0, \sigma^2)$).

$$M(t) = E[e^{-tx}] = e^{\frac{t^2\sigma^2}{2}}.$$

As previously described, the asymptotic distribution is given as

$$\hat{\sigma^2} \xrightarrow{D} N\left(\sigma^2, (h^{'}(E[M_n(t)]))^2 \text{Var}[M_n(t)]\right) = N\left(\sigma^2, \frac{1}{n} v(\sigma^2, t)\right),$$

where

$$v(\sigma^2, t) = 4t^{-4}\left(\exp\left(t^2\sigma^2\right) - 1\right). \tag{2.26}$$

We need to find the value of t which minimizes $V(\sigma^2, t)$. Letting $\frac{\partial}{\partial t} v(\sigma^2, t) = 0$ yields

$$t = \frac{\sqrt{2 + \text{W}(-2e^{-2})}}{\sigma}. \tag{2.27}$$

However, σ is unknown. To estimate σ, we can start from arbitrary t_0 and obtain $\sigma_0 = \frac{\sqrt{2\log(M_n(t_0))}}{|t_0|}$. The next value t_1 is given as $t_1 = \frac{\sqrt{2+\text{W}(-2e^{-2}))}}{\sigma_0}$. It can be shown that σ_i converges to σ by repeating this procedure. The optimal t is given by substituting $\sigma_i = \frac{\sqrt{2\log(M_n(t_i))}}{|t_i|}$ into $t_{i+1} = \frac{\sqrt{2+\text{W}(-2e^{-2}))}}{\sigma_i}$ and letting $i \to \infty$. The optimal t is the solution of the following formula

$$2\log(M_n(t)) = 2 + \text{W}(-2e^{-2}).$$

Solving this for t gives

$$t = M_n^{-1}\left(exp\left(1 + \frac{\text{W}(-2e^{-2})}{2}\right)\right). \tag{2.28}$$

Simulation Result

Here we show the simulation result of the adaptive estimator. The initial value of t is set 10.0 and the number of samples is 1000. Table 1 shows the value of $\hat{\sigma}^2$ at each iteration. $\sigma^2 = 3$.

Figure 2.4 shows the plot of the logarithm of Equation (2.26) with different sigma values. The red line corresponds to $\sigma^2 = 3$. From this graph it can be easily seen that t_i converges to the value which minimizes the variance of the asymptotic distribution. Note that it is unnecessary to do this iterative calculation since the best value of t is given by Equation (2.28). This example just illustrates how the adaptive estimator works.

2.7 Empirical Minimum Variance Estimator

Another similar alternative approach is the empirical minimum variance estimator. This estimator has the same asymptotic property with an adaptive estimator (see detail [3]).

Table 1: Adaptive Iteration

ith iteration	t_i	σ_i^2
0	10	0.9166257
1	1.3185511	2.9122434
2	0.7397400	2.9786549
3	0.7314470	2.9787706
4	0.7314328	2.9787708
5	0.7314328	2.9787708
6	0.7314328	2.9787708
7	0.7314328	2.9787708
8	0.7314328	2.9787708
9	0.7314328	2.9787708

Example - Normal Scale Problem

We estimate a Normal distribution$N(0, \sigma^2)$. The estimator of σ^2 is given as

$$\hat{\sigma^2} = \frac{2 \log \hat{\phi}(t)}{t^2}.$$

By substituting this formula into Equation (2.26), we have

$$v(\sigma^2, t) = 4t^{-4} \left(\exp(2 \log \hat{\phi}(t)) - 1 \right).$$

Simulation Result

The optimal value of t was estimated by the empirical minimum variance estimator. The simulation was done ten times. The number of samples is 10000 and $\sigma^2 = 3$. The obtained values are 0.7314136, 0.7289867, 0.7178052, 0.7289310, 0.7396936, 0.7345791, 0.7042096, 0.7178947, 0.7349337, 0.7169946. These values are similar to those obtained by the adaptive estimator (see Table 1).

The adaptive estimator and empirical minimum variance estimator have been introduced. The results of the simulations show both estimators give a similar Laplace transform parameter as the optimal value. However, the performance might be different for the small number of samples. Therefore, we compared the performance of these two methods for different number of samples $n = 10, 100, 1000$. The number of iteration is fixed to 1000 for all experiments. The results are given in Table 3. The result of the MLE is also given in this table. The performance of an adaptive estimator and empirical variance estimator are quite similar and both are about three times lower than that of the MLE. It is worth to remark that the adaptive estimator outperforms an empirical minimum variance estimator if n is small when

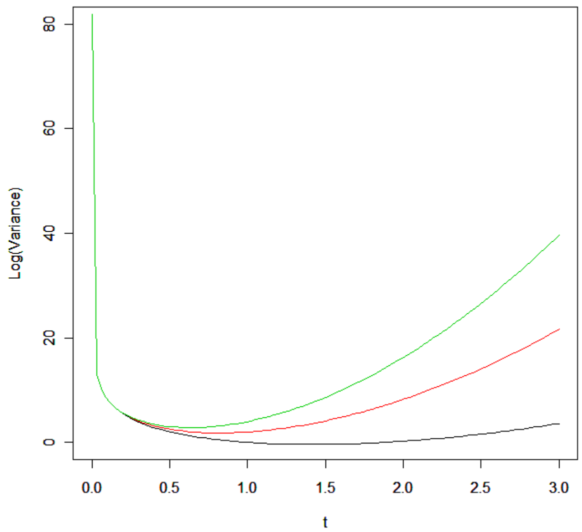

Figure 2.4: Logarithm of the variance of the asymptotic distribution of $\hat{\sigma}$ against t. $\hat{\sigma}$ is a convex function of t. $(\sigma^2 = 1(black), 3(red), 5(green))$

$\sigma^2 = 1$. As Figure 2.4 shows, the variance curve of the asymptotic distribution is not steep for small values of σ^2. In this case, the adaptive estimator is preferable.

3 Comparison with Other Type Estimators

We derived the Laplace transform based estimator and its asymptotic distribution in the previous section. However, no performance comparison has been made with other estimators. In order to use the Laplace transform based estimator instead of the MLE, the performance of both estimators should be compared. Therefore, we start this section by considering the asymptotic distribution of the MLE. There exist many methods to estimate a parameter of a statistical model. In this section we limit our attention to the estimation methods for the stable law and introduce a couple of methods. Their performance is compared with that of the Laplace transform based estimator. Throughout this section, the Normal scale problem is used for the comparison.

3.1 Comparison with Maximum Likelihood Estimator

Let $\boldsymbol{\theta}$ be parameters of a statistical model. The asymptotic distribution of $\boldsymbol{\theta}$ is given as

$$\hat{\boldsymbol{\theta}} \xrightarrow{D} N(\boldsymbol{\theta}, I(\boldsymbol{\theta})^{-1}),$$

where I is the Fisher information matrix which is defined as

$$I(\boldsymbol{\theta})_{ij} = -E\left[\frac{\partial^2 l}{\partial\theta_i \partial\theta_j}\right].$$

The performance of the Laplace transform based estimator can be evaluated by comparing the determinant of the matrix $I(\boldsymbol{\theta})^{-1}$ with that of the covariance matrix of the asymptotic distribution of the Laplace transform based estimator.

Example - Normal Scale Problem

Now the performance of the Laplace transform based estimator is compared with that of the MLE for the Normal scale problem. Let X be a random variable which follows the Normal distribution with mean 0 and variance σ^2 (i.e. $X \sim N(0, \sigma^2)$. The score function and the Fisher information are given as

$$U(\sigma^2) = \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} x_i^2$$

and

$$I(\sigma^2) = -E\left[\frac{\partial^2 l}{\partial \sigma^2}\right] = \frac{n}{2\sigma^4}.$$

Therefore, the asymptotic distribution of $\hat{\sigma^2}$ is given as

$$\hat{\sigma^2} \xrightarrow{D} N\left(\sigma^2, \frac{2\sigma^4}{n}\right).$$

Table 2: Comparison with Maximum Likelihood Estimator

σ^2	Laplace Transform Based Estimator	Maximum Likelihood Estimator
1.7	0.0050185	0.0015997
3.19	0.0622216	0.0198332

On the other hand, as we have seen, the asymptotic distribution of $\hat{\sigma^2}$ obtained by the Laplace transform based estimator is given as

$$\hat{\sigma^2} \xrightarrow{D} N\left(\sigma^2, \frac{1}{n}v(\sigma^2, t)\right).$$

The function $v(\sigma^2, t)$ is given by Equation (2.26). Substituting t given by Equation (2.27) minimizes the variance.

$$v(\sigma^2, t) = \frac{4\sigma^4}{(2 + \mathrm{W}(-2e^{-2}))^2}\left(e^{2+\mathrm{W}(-2e^{-2})} - 1\right).$$

If the coefficients of σ^4 are compared

$$2 < \frac{4}{(2 + \mathrm{W}(-2e^{-2}))^2}\left(e^{2+\mathrm{W}(-2e^{-2})} - 1\right) = 6.17655. \tag{3.1}$$

The variance of the Laplace transform based estimator is roughly three times greater than that of the MLE. This means the MLE is superior to the Laplace transform based estimator for the Normal scale problem.

Simulation Result

Table 2 shows the variance of estimates for each method. The variance of the Laplace transform based estimator is roughly three times greater than that of the MLE as we have shown in Equation (3.1). The value of t used in this simulation is obtained by Equation (2.27). In practical, we have to use the adaptive estimator or empirical minimum variance estimator since we have to estimate the optimal value of t. The comparison of three methods: the adaptive estimator, the empirical minimum variance estimator and the MLE is found in Table 3.

3.2 Comparison with Characteristic Function Based Estimator

In this subsection we introduce two methods, the method of Press[21] and method of Koutrouvelis[17] (the regression-type estimator). Both were developed to estimate the four parameters of the stable law and utilize the characteristic function. The method of Koutrouvelis is a standard method of estimating the parameters of the stable law. We give the definition of the stable law first and then briefly describe their derivation.

3.2.1 α-Stable Distribution

The characteristic function is given by the Fourier-Stieltjes transform

$$\phi(t;\boldsymbol{\theta}) = \int_{-\infty}^{\infty} e^{-itx} dF(x;\boldsymbol{\theta}).$$

The stable law is a distribution whose characteristic function is given as

$$\begin{aligned} \log\phi(t) &= i\delta t - |ct|^{\alpha}(1 - i\beta sgn(t)\tan(\frac{\Pi\alpha}{2})) && (\alpha \neq 1) \\ \log\phi(t) &= i\delta t - |ct|^{\alpha}(1 - i\beta sgn(t)\frac{2}{\Pi}\log|t|) && (\alpha = 1). \end{aligned}$$

3.2.2 Method of Press

The method of Press can be derived in the following way. The logarithm of the characteristic function of the stable law is

$$\log\phi(t) = -\sigma^{\alpha}|t|^{\alpha}\left(1 - i\beta\mathrm{sign}(t)\tan(\frac{\pi\alpha}{2})\right)$$

Therefore,

$$-\log|\phi(t)| = \sigma^{\alpha}|t|^{\alpha}.$$

Since

$$|\phi(t)| = \exp(-\sigma^{\alpha}|t|^{\alpha}).$$

By using t_1 and t_2 ($t_1 \neq t_2$),

$$\begin{array}{ll} -\log|\phi(t_1)| & = \sigma^{\alpha}|t_1|^{\alpha} \\ -\log|\phi(t_2)| & = \sigma^{\alpha}|t_2|^{\alpha} \end{array} . \tag{3.2}$$

From the above equation, we have

$$\hat{\alpha} = \log\frac{\log|\phi(\hat{t_2})|}{\log|\phi(\hat{t_1})|} / \log\left|\frac{t_2}{t_1}\right|.$$

Substituting this into Equation(3.2) yields

$$\log\hat{\sigma} = \frac{\log|t_2|\log(-\log|\hat{\phi}(t_1)|) - \log|t_1|\log(-\log|\hat{\phi}(t_2)|)}{\log(-\log|\hat{\phi}(t_2)|) - \log(-\log|\hat{\phi}(t_1)|)}.$$

The imaginary part of $\log\phi(t)$ is

$$\Im(\log\phi(t)) = \mu t + \sigma^{\alpha}|t|^{\alpha}\beta\mathrm{sign}(t)\tan(\frac{\pi\alpha}{2}).$$

Dividing both sides by t yields

$$\frac{\Im(\log\phi(t))}{t} = \mu + \sigma^\alpha |t|^{\alpha-1}\beta \text{sign}(t)\tan(\frac{\pi\alpha}{2}).$$

By using t_3 and t_4 ($t_3 \neq t_4$), we have

$$\frac{\Im(\log\phi(t_3))}{t_3} = \mu + \sigma^\alpha |t_3|^{\alpha-1}\beta \text{sign}(t_3)\tan(\frac{\pi\alpha}{2})$$
$$\frac{\Im(\log\phi(t_4))}{t_4} = \mu + \sigma^\alpha |t_4|^{\alpha-1}\beta \text{sign}(t_4)\tan(\frac{\pi\alpha}{2}).$$

Solving the above formula for β and μ, we obtain

$$\hat{\beta} = \frac{T_3 - T_4}{U_3 - U_4},$$

and

$$\hat{\mu} = \frac{V_4 T_3 - V_3 T_4}{V_4 - V_3},$$

where

$$T_i = \frac{\Im(\log\phi(t_i))}{t_i},$$

$$U_i = \sigma^\alpha |t_i|^{\alpha-1}\text{sign}(t_i) tan(\frac{\pi\alpha}{2})$$

and

$$V_i = \sigma^\alpha |t_i|^{\alpha-1}\beta\text{sign}(t_i) tan(\frac{\pi\alpha}{2}),$$

for $i = 3, 4$.

Example - Normal Scale Problem

Here we show how to estimate the scale parameter of the Normal distribution. The characteristic function of $N(0, \sigma^2)$ is given as

$$\phi(t) = \exp\left(-\frac{\sigma^2 t^2}{2}\right). \tag{3.3}$$

The e.c.f is similar with the empirical Laplace transformation, which is given as

$$\phi_n(t) = \frac{1}{n}\sum_{i=1}^{n} e^{-itx_i}.$$

The estimator of σ^2 is given as

$$\hat{\sigma^2} = \frac{-2\log|\hat{\phi}(t)|}{t^2},$$

which is exactly the same with the Laplace transform based estimator since

$$|\phi_n(t)| = \exp\left(-\frac{\sigma^2 t^2}{2}\right). \tag{3.4}$$

3.2.3 Regression-Type Estimator

The regression-type estimator is derived as follows. In the case of $\alpha \neq 1$,

$$\log(-\log|\phi(t)|^2) = \log(2\sigma^\alpha) + \alpha\log|t|.$$

By regressing $\log(-\log|\phi(t)|^2)$ on $\log|t|$, we can estimate α and σ. The real and imaginary part of the characteristic function $\phi(t)$ are

$$\begin{aligned}\Re\phi(t) &= \exp(-|\sigma t|^\alpha)\cos\left(\mu t + |\sigma t|^\alpha \beta \mathrm{sign}(t)\tan(\frac{\pi\alpha}{2})\right)\\ \Im\phi(t) &= \exp(-|\sigma t|^\alpha)\sin\left(\mu t + |\sigma t|^\alpha \beta \mathrm{sign}(t)\tan(\frac{\pi\alpha}{2})\right).\end{aligned}$$

Therefore,

$$\tan^{-1}\left(\frac{\Im\phi(t)}{\Re\phi(t)}\right) = \mu t + \beta\sigma^\alpha \tan(\frac{\pi\alpha}{2}\mathrm{sign}(t))|t|^\alpha.$$

Since

$$\phi(t) = \exp\left(-\sigma^\alpha|t|^\alpha + i\left(\sigma^\alpha|t|^\alpha\beta\mathrm{sign}(t)\tan(\frac{\pi\alpha}{2} + \mu t)\right)\right).$$

By using estimated $\hat{\mu}$ and $\hat{\beta}$ and regressing $\tan^{-1}\left(\frac{\Im\phi(t)}{\Re\phi(t)}\right)$ on t and $\mathrm{sign}(t))|t|^\alpha$, we obtain $\hat{\beta}$ and $\hat{\mu}$. The asymptotic distribution of the regression-type estimator is rather hard to know for the general case because this method consists of two steps.

Example - Normal Scale Problem

The characteristic function of $N(0,\sigma^2)$ is Equation (3.3).
We have

$$\log(-\log(|\phi(t)|^2)) = \log(\sigma^2) + 2\log(|t|).$$

Substituting Equation (3.4) into the above formula shows the estimator is exactly the same with the Laplace transform based estimator if only one transform parameter is used. As originally defined in his paper, the method of Koutrouvelis estimates the parameter using many transform parameters to improve the performance. However, the use of regression degrades the performance for the Normal scale problem. This is because that the asymptotic variance is a convex function of the transform parameter t (see Figure 2.4) and only a single point minimizes the asymptotic variance. Thus increasing the number of the transform parameters gives no gain.

4 Improving the Efficiency of Estimator with Multiple Transform Parameters

In section 3 we compared Laplace transform based estimator with other methods. We have seen the MLE is the best for the Normal scale problem. For the Normal scale problem, the efficiency of the Laplace transform based estimator is three times lower than that of the MLE. This example implies that there might be room to improve the performance. If it is possible, the Laplace transform based estimator will be a more attractive method as an alternative of MLE.

It is important to note that the number of the transform parameters and the number parameters of a statistical model were identical so far. At the end of the previous section, we introduced the method of Koutrouvelis. It uses many transform variables in a regression style and achieves better performance than the method of Press (see detail [17]). However, this method can be used only for estimating the parameters of the stable law.

The minimum rth distance method is a more general method. Paulson[20] proposed the method to estimate the parameters of the stable law by minimizing the following integral

$$I = \int_{-\infty}^{\infty} |M_n(t) - M_\theta(t)| \, e^{-t^2} dt.$$

Heatcote[15] generalized his method and suggested to minimize

$$I = \int_{-\infty}^{\infty} |M_n(t) - M_\theta(t)|^2 \, dG(t),$$

where $G(t)$ is a weight function. The above methods can be thought of as a special case ($r = 2$) of the minimum rth distance method (a good overview is found in [23]), which is formulated as

$$I = \int_{-\infty}^{\infty} |M_n(t) - M_\theta(t)|^r \, W_\theta(t) dt. \tag{4.1}$$

Clearly the choice of a weight function is crucial to obtain the most effective estimator. Many researchers tried to find the optimal weight function, however, as Jun [26] stated, it is still an open problem if the maximum likelihood efficiency can be achieved or not.

Another general and more frequently used method is the GMM estimator. The GMM was developed by Hansen [13] to deal with over-identified models. Over-identified means that the number of conditional equations is greater than the number of parameters. It is literally a generalization of the method of moments. As shown by many researchers, the empirical transform based estimator can be used in the GMM framework. Essentially, the GMM estimator uses a finite number of the transform parameters. However, Carrasco [6] derived the continuous GMM estimator which exploits a continuum of moment conditions. She also showed her estimator is asymptotically as efficient as the MLE (see detail [7]).

Now two methods, the minimum rth distance estimator and the GMM estimator, are examined with an example. We focus more on the GMM estimator rather than the minimum

$r$the distance method.

4.1 Minimum rth Distance Method

As we have shown the estimates can be obtained by minimizing the integral given by Equation (4.1). Here we show a simple example of Heatcote's method.

Example - Normal Scale Problem

The integrated squared error estimation is applied to the Normal scale problem. The estimate can be given as

$$\hat{\boldsymbol{\sigma^2}} = \arg\min_{\boldsymbol{\sigma^2}} \int_{-\infty}^{\infty} \left| M_n(t) - \exp(\frac{\sigma^2 t^2}{2}) \right|^2 dG(t). \tag{4.2}$$

In this example, the standard Normal distribution is used as a weight function ($G(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$).

Simulation Result

The scale parameter was estimated with different ranges of t. The simulation result is in Table 4. The result shows that the integrated squared error estimation performs worse than any other estimators for the Normal scale problem. The reason is the same with our discussion given at the end of the section 3. This reveals the difficulty of the choice of a weight function and also indicates that this method does not necessarily improve the performance of the estimator.

4.2 Generalized Method of Moments Estimator

Many studies on the GMM estimator have been done with the characteristic function. Those results can be used for the Laplace transform based estimator. As described before, there are two kinds of the GMM estimators. One is discrete and the other is continuous. Firstly, we derive the discrete GMM estimator and show an example of the joint estimation of location and scale parameters. Then the derivation of the continuous GMM estimator is briefly described.

4.2.1 Generalized Method of Moments

As a preliminary step, we briefly review the GMM. The procedure of the GMM is as follows. Let $E[X^i]$ be the ith moment and m_i be ith sample moment. The problem is to find the parameter $\boldsymbol{\theta}$ of a statistical model which minimizes the quantity

$$\boldsymbol{g}^T S^{-1} \boldsymbol{g},$$

where

$$g_i = m_i - E[X^i], \tag{4.3}$$

and S is a weight matrix. Let $\bar{S}$ be the optimal weight matrix which gives the smallest asymptotic variance. The ij-element of the optimal weight matrix is given as $\bar{S}_{ij} = cov(g_i, g_j)$. This is intuitively plausible since the weight should be small if the variance is large. In practice $\bar{S}$ is not obtainable since the true values of the parameters $\boldsymbol{\theta}_0$ are unknown. Thus we need to to estimate the optimal weight matrix. The estimation can be done in the following way.

By repeating the below two step procedure until $\boldsymbol{\theta}_{i+1} \simeq \boldsymbol{\theta}_i$, the estimate $\hat{S}$ can be computed. Before starting the procedure, S_0 must be initialized to the identity matrix I.

1 Estimate $\boldsymbol{\theta}_i$ with S_i

2 Obtain S_{i+1} with $\boldsymbol{\theta}_i$

The estimate $\hat{S}$ is given as S_{i+1} when $\boldsymbol{\theta}$ converges. Similar with the multiple parameters case described in section 2, the asymptotic distribution of the estimate of $\boldsymbol{\theta}$ is given as

$$\hat{\boldsymbol{\theta}} \xrightarrow{D} N\left(\boldsymbol{\theta}, (D^T S^{-1} D)^{-1}\right),$$

where

$$D = \frac{\partial E\left[\boldsymbol{X}\right]}{\partial \boldsymbol{\theta}^T}, \ \boldsymbol{X} = (X^1, X^2, \ldots, X^p)^T.$$

In this way, we can exploit an arbitrary numbers of different orders of moments.

4.2.2 Discrete GMM Estimator

The derivation of the discrete GMM estimator is rather easy. By replacing the moment condition (Equation (4.3)) with the following formula, the Laplace transform based estimator can be extended so that it can handle the over-identified case.

$$g_i = M_n(t_i) - M(t_i).$$

The estimator is given as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(\boldsymbol{g}^T S^{-1} \boldsymbol{g}\right). \tag{4.4}$$

When the number of unknowns and that of ts are identical, the estimator is equivalent to the multiple-parameter case which we have shown in section 2. Additionally, it does not depend on S (see [26]).

Example - Exponential Distribution

Let $X = X_0 + \omega$, where X_0 is a random variable which follows the Exponential distribution with the scale parameter θ. We show how to estimate the location and scale parameters of X by the GMM estimator (The setting is identical with the example in section 2.3.3.). Let p denote the number of ts. The estimator is given as

$$(\hat{\omega}, \hat{\theta}) = \arg\min_{\omega,\theta} \left(\boldsymbol{g}^T S^{-1} \boldsymbol{g} \right).$$

The each element of the $p \times p$ covariance matrix S is given as

$$S_{ij} = E_{ij} - E_i E_j,$$

where

$$\begin{aligned}
E_i &= exp(-t_i \omega) \frac{\theta}{\theta + t_i} \\
E_j &= exp(-t_j \omega) \frac{\theta}{\theta + t_j} \\
E_{ij} &= exp(-(t_i + t_j)\omega) \frac{\theta}{\theta + (t_i + t_j)},
\end{aligned}$$

for $i = 1, \ldots, p$ and $j = 1, \ldots, p$. The asymptotic distribution of $(\hat{\omega}, \hat{\theta})$ is given as

$$\begin{pmatrix} \hat{\omega} \\ \hat{\phi} \end{pmatrix} \xrightarrow{D} N \left(\begin{pmatrix} \omega \\ \phi \end{pmatrix}, (D^T S^{-1} D)^{-1} \right),$$

where

$$D = \left(\frac{\partial M_{\omega,\theta}(\boldsymbol{t})}{\partial \omega}, \frac{\partial M_{\omega,\theta}(\boldsymbol{t})}{\partial \theta} \right) = \begin{pmatrix} -t_1 e^{-\omega t_1} \frac{\theta}{\theta + t_1} & t_1 e^{-\omega t_1} \frac{1}{(\theta + t_1)^2} \\ \vdots & \vdots \\ -t_p e^{-\omega t_p} \frac{\theta}{\theta + t_p} & t_p e^{-\omega t_p} \frac{1}{(\theta + t_p)^2} \end{pmatrix}.$$

Simulation Result

In this simulation both of the location ω and the scale parameter θ and are set to 1. These parameters are estimated by the GMM estimator. The number of iteration is fixed to 1000 for each simulation. The performance of a course grid and fine grid. The step size of a course grid is 1.0 (i.e. $t_1 = 1.0$, $t_2 = 2.0$, $t_3 = 3.0, \ldots$) and 0.1 is used for a fine grid (i.e. $t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3, \ldots$). The number of ts ranges from 2 to 10 for both grids. The result of a course grid is given in Table 5 and of a fine grid is given in 6. From these tables we can see that the efficiency is improved by increasing the number of ts (This facts leads us to the continuous case). Unfortunately, the estimates cannot be obtained when many ts are used. This is because the matrix inversion cannot be computed in the estimation procedure because of a singularity. Furthermore, when n is small, the GMM estimator does not work properly. On the other hand, the MLE is robust and always performs better. Especially for the location estimation, the MLE outperforms the GMM estimator. However, it is worth to note that for the scale parameter the GMM estimator is as efficient as the MLE when several transform parameters are used.

A singularity occurs if the number of ts is large or the locations of ts are too close. To avoid this problem, the Tikhonov regularization can be used. Suppose the linear system $A\boldsymbol{x} = \boldsymbol{b}$ is ill-posed. We can solve this problem by minimizing the quantity

$$|A\boldsymbol{x} - \boldsymbol{b}|^2 + |\Gamma \boldsymbol{x}|^2.$$

The solution is given as

$$\hat{\boldsymbol{x}} = (A^T A + \Gamma^T \Gamma)^{-1} A^T \boldsymbol{b}.$$

Normally, αI is used for Γ, where I is the identity matrix. Talbe 7 and 8 show that the performance of the estimator depends on the choice of α. As these tables suggest, a small α value gives better performance. However, if it is too small, correct estimates cannot be obtained (see the bottom row of Table 8). Thus, the choice of α is problematic. Furthermore, introducing the regularization parameter makes it difficult to derive the exact asymptotic distribution.

4.3 Continuous GMM Estimator

Carrasco[6] presented the GMM based estimator which uses a continuum of moment conditions. Her method uses the characteristic function, however, it can be applied to the Laplace transform estimator. The continuous GMM estimator is similar with the discrete one. In the continuous case, the covariance operator K is used instead of the covariance matrix S. Let $g(t) = M_n(t) - M(t)$ and $k(s,t) = E\left[g(s)g(t)\right] = M(s+t) - M(s)M(t)$. Then the estimator is given as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \int\int g(s)k(s,t)g(t)dsdt.$$

The function $k(s,t)$ can be thought of a counterpart of the covariance matrix S of the discrete GMM estimator. Let define the covariance operator K as

$$(Kf)(t) = \int k(s,t)f(s)ds.$$

Furthermore, let ϕ be the eigenfunction and λ be the eigenvector of K (i.e. $K\phi = \lambda\phi$). As a similar reason with the discrete case, K^{-1} does not exist always. Therefore, Carrasco derived the estimator by using the Tikhonov approximation. By replacing the eigenvalues λ with $\frac{\lambda^2+\alpha}{\lambda}$, the inverse of the new covariance operator K' satisfies

$$(K'^{-1}f)(t) = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i^2 + \alpha}(f, \phi_i)\phi_i(t),$$

where α is a smoothing parameter. It functions as the parameter α introduced for the discrete GMM estimator to avoid a singularity problem. The optimal estimator can be obtained as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i^2 + \alpha}(\phi_i, g(\boldsymbol{\theta}))^2.$$

More detail can be found in [6] and [7]. We do not show the example since this method is computationally very expensive.

5 Conclusion and Discussion

In this paper, we introduced the Laplace transform based estimator and discussed its property. As we have shown, the MLE always performs better than the Laplace transform based estimator. However, the efficiency can be improved in the GMM framework by increasing the number of Laplace transform parameters. In some cases a comparable efficiency can be achieved. However, We have also seen the interesting fact that the increase of the transform parameters does not always improve the performance of estimators. In some instances, only one transform variable gives the best efficiency. If the form of the asymptotic distribution is not available or too complicated, a choice of methods or choice of weight function could be hard. Further investigation is needed to address this problem.

There is a trade-off between efficiency and computational complexity. The computational time becomes longer as the number of the transform parameters becomes large. The computational cost is not negligible also in the continuous case. We showed the example of Heatcote's method in section 4. The computation time is considerably high because of the numerical integration which is required for the estimation. We only described the estimation method of the continuous GMM estimator because the estimates cannot be obtained in a feasible time without any approximation.

There is another trade-off between efficiency and robustness. As we have showed in the example of the Exponential distribution in section 4, the regularization is necessary to derive a stable estimator. The choice of the regularization parameter α is difficult because a small value gives an accurate result, however, if it is too small, the estimator becomes unstable. Another problem of introducing a regularization is that it makes difficult to derive the exact asymptotic distribution. To derive a robust estimator is rather hard.

It should be mentioned that the Laplace transform based estimator is not always available. Since the moment generating function(m.g.f.) is undefined for some probability distributions. For instance, the m.g.f. is not defined for the Levy distribution meanwhile its characteristic function exists. In such case, the e.c.f based estimator must be used.

One concern about implementation is numerical error. We used the statistical package R for all experiments. Through the experiments, we found the numerical integration routine could give a large absolute error. Although this is a minor issue, it should be taken into account or avoided if possible.

There are many alternative estimation methods besides those described in this paper. For instance, the empirical likelihood estimator is considered to have an advantage over the e.c.f based estimator (see for example [16]). Other methods are also have to be studied to understand the advantage of the Laplace transform based estimator.

Table 3: Comparison of Performance - Normal Distribution $N(0, \sigma^2)$

($\sigma^2 = 1$)

Estimator		The number of samples (n)		
		10	100	1000
Adaptive Estimator	$mean$	0.9657395	0.9967026	0.999392
	var	0.0508228	0.00599233	0.000598098
	MSE	0.0519458	0.00599721	0.000597870
Empirical Minimum Variance Estimator	$mean$	0.9371843	1.000238	0.9993202
	var	0.2558396	0.00595167	0.000671987
	MSE	0.2595295	0.00594578	0.000671777
Maximum Likelihood Estimator	$mean$	1.000782	1.001351	1.000505
	var	0.0201489	0.00196527	0.000186033
	MSE	0.0201294	0.00196513	0.000186102

($\sigma^2 = 3$)

Estimator		The number of samples (n)		
		10	100	1000
Adaptive Estimator	$mean$	2.943984	3.000098	3.002261
	var	0.54054	0.05207458	0.005314705
	MSE	0.5431372	0.05202251	0.005314503
Empirical Minimum Variance Estimator	$mean$	2.919569	2.992783	2.999604
	var	0.5436256	0.04962659	0.005257955
	MSE	0.5495511	0.04962905	0.005252855
Maximum Likelihood Estimator	$mean$	2.990838	2.99759	2.997937
	var	0.1752654	0.0183039	0.001720202
	MSE	0.1751741	0.01829141	0.001722737

($\sigma^2 = 5$)

Estimator		The number of samples (n)		
		10	100	1000
Adaptive Estimator	$mean$	4.907138	5.002849	4.998482
	var	1.400281	0.1512927	0.01556097
	MSE	1.407504	0.1511496	0.01554771
Empirical Minimum Variance Estimator	$mean$	4.840697	4.986414	4.995339
	var	1.366790	0.1542322	0.01545801
	MSE	1.390801	0.1542625	0.01546427
Maximum Likelihood Estimator	$mean$	5.011264	5.000778	5.002633
	var	0.4674835	0.0488539	0.005043829
	MSE	0.4671429	0.04880565	0.005045717

Table 4: Comparison of Performance - Normal Distribution $N(0, \sigma^2)$

($\sigma^2 = 1$)

Estimator		The number of samples (n)		
		10	100	1000
Heatcote $t = [-0.01, 0.01]$	$mean$	0.9714372	1.003914	0.9989877
	var	0.1183390	0.01888060	0.001950277
	MSE	0.1190365	0.01887704	0.001949352
Heatcote $t = [-0.1, 0.1]$	$mean$	0.9759225	1.001336	1.000244
	var	0.1113935	0.02000950	0.002092236
	MSE	0.1118618	0.01999128	0.002090204
Heatcote $t = [-1.0, 1.0]$	$mean$	0.9471491	0.997741	0.9976314
	var	0.1098501	0.01983562	0.002007995
	MSE	0.1125334	0.01982089	0.002011597

($\sigma^2 = 3$)

Estimator		The number of samples (n)		
		10	100	1000
Heatcote $t = [-0.01, 0.01]$	$mean$	2.939282	3.010854	2.998804
	var	0.6141537	0.167843	0.01821945
	MSE	0.6172262	0.1677930	0.01820266
Heatcote $t = [-0.1, 0.1]$	$mean$	2.917967	2.968488	2.994382
	var	0.6036712	0.1707516	0.01632724
	MSE	0.6097969	0.1715739	0.01634247
Heatcote $t = [-1.0, 1.0]$	$mean$	2.714301	2.953053	2.992911
	var	0.5044807	0.1948728	0.02646382
	MSE	0.5856	0.1968819	0.02648761

($\sigma^2 = 5$)

Estimator		The number of samples (n)		
		10	100	1000
Heatcote $t = [-0.01, 0.01]$	$mean$	4.868089	4.992571	4.99834
	var	2.14304	0.4759336	0.04927537
	MSE	2.158297	0.4755128	0.04922885
Heatcote $t = [-0.1, 0.1]$	$mean$	4.752257	4.993198	4.989874
	var	2.165844	0.5030996	0.04904361
	MSE	2.225055	0.5026427	0.0490971
Heatcote $t = [-1.0, 1.0]$	$mean$	4.298646	4.810637	4.98253
	var	1.562511	0.6311191	0.1257302
	MSE	2.052847	0.6663464	0.1259097

Table 5: Comparison of Performance - Exponential Distribution $NE(\theta)$
(scale $\theta = 1$, shift $\omega = 1$)

Estimator			n = 10	100	1000
Maximum Likelihood Estimator	ω	*mean*	1.101993	1.01008	1.000920
		var	0.01049800	9.582203e-05	7.88985e-07
		MSE	0.02089009	0.0001973332	1.635356e-06
	θ	*mean*	0.9017085	0.9866779	0.9998756
		var	0.08757485	0.01040339	0.0009782439
		MSE	0.0971485	0.01057046	0.000977281
GMM ($t_1 = 1$, $t_2 = 2$)	ω	*mean*	1.037423	1.003510	1.000086
		var	0.0163785	0.001360857	0.0001276980
		MSE	0.01776259	0.001371819	0.0001275777
	θ	*mean*	1.162105	1.017325	1.001015
		var	0.2695025	0.01353516	0.001142398
		MSE	0.2955111	0.01382178	0.001142285
GMM ($t_1 = 1$, $t_2 = 2$, $t_3 = 3$)	ω	*mean*	N/A	1.002803	0.9999362
		var	N/A	0.0007480634	7.657846e-05
		MSE	N/A	0.000755173	7.650595e-05
	θ	*mean*	N/A	1.013786	0.9994717
		var	N/A	0.01344781	0.001182703
		MSE	N/A	0.01362441	0.001181799
GMM ($t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$)	ω	*mean*	N/A	1.002019	1.000495
		var	N/A	0.0004226337	4.611024e-05
		MSE	N/A	0.0004262862	4.63095e-05
	θ	*mean*	N/A	1.011348	1.001584
		var	N/A	0.01213074	0.001061635
		MSE	N/A	0.0122474	0.001063081
GMM ($t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$, $t_5 = 5$)	ω	*mean*	N/A	1.003044	1.000495
		var	N/A	0.0003811196	2.99353e-05
		MSE	N/A	0.0003900016	3.015044e-05
	θ	*mean*	N/A	1.011561	1.002198
		var	N/A	0.01167366	0.001036173
		MSE	N/A	0.01179564	0.001039967

(scale $\theta = 1$, shift $\omega = 1$)

Estimator			n		
			10	100	1000
GMM ($t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$, $t_5 = 5$, $t_6 = 6$)	ω	*mean*	N/A	N/A	1.000028
		var	N/A	N/A	2.074229e-05
		MSE	N/A	N/A	2.072234e-05
	θ	*mean*	N/A	N/A	1.000316
		var	N/A	N/A	0.001046751
		MSE	N/A	N/A	0.001045804
GMM ($t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$, $t_5 = 5$, $t_6 = 6$, $t_7 = 7$)	ω	*mean*	N/A	N/A	1.000352
		var	N/A	N/A	1.630260e-05
		MSE	N/A	N/A	1.640999e-05
	θ	*mean*	N/A	N/A	1.001796
		var	N/A	N/A	0.0009959811
		MSE	N/A	N/A	0.0009982103
GMM ($t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$, $t_5 = 5$, $t_6 = 6$, $t_7 = 7$, $t_8 = 8$)	ω	*mean*	N/A	N/A	1.000345
		var	N/A	N/A	1.327175e-05
		MSE	N/A	N/A	1.337728e-05
	θ	*mean*	N/A	N/A	1.001202
		var	N/A	N/A	0.001074440
		MSE	N/A	N/A	0.001074810
GMM ($t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$, $t_5 = 5$, $t_6 = 6$, $t_7 = 7$, $t_8 = 8$, $t_9 = 9$)	ω	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A
	θ	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A
GMM ($t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$, $t_5 = 5$, $t_6 = 6$, $t_7 = 7$, $t_8 = 8$, $t_9 = 9$, $t_{10} = 10$)	ω	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A
	θ	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A

Table 6: Comparison of Performance - Exponential Distribution $NE(\theta)$
(scale $\theta = 1$, shift $\omega = 1$)

Estimator			n		
			10	100	1000
Maximum Likelihood Estimator	ω	*mean*	1.101993	1.01008	1.000920
		var	0.01049800	9.582203e-05	7.88985e-07
		MSE	0.02089009	0.0001973332	1.635356e-06
	θ	*mean*	0.9017085	0.9866779	0.9998756
		var	0.08757485	0.01040339	0.0009782439
		MSE	0.0971485	0.01057046	0.000977281
GMM ($t_1 = 0.1$, $t_2 = 0.2$)	ω	*mean*	1.096489	1.011473	1.001758
		var	0.04636605	0.006167253	0.0006036219
		MSE	0.05562987	0.006292711	0.0006061099
	θ	*mean*	1.305690	1.028978	1.004625
		var	0.3886073	0.01655707	0.001673643
		MSE	0.4816649	0.01738024	0.001693362
GMM ($t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$)	ω	*mean*	1.037863	1.005285	1.001179
		var	0.03312287	0.003149285	0.0003119856
		MSE	0.03452336	0.003174067	0.0003130646
	θ	*mean*	1.218294	1.019855	1.002162
		var	0.285755	0.01430266	0.001322806
		MSE	0.3331214	0.01468257	0.001326159
GMM ($t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$, $t_4 = 0.4$)	ω	*mean*	N/A	1.005367	1.000094
		var	N/A	0.002203662	0.000195303
		MSE	N/A	0.002230258	0.0001951165
	θ	*mean*	N/A	1.021793	1.001999
		var	N/A	0.01468262	0.001244161
		MSE	N/A	0.01514287	0.001246913
GMM ($t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$, $t_4 = 0.4$, $t_5 = 0.5$)	ω	*mean*	N/A	1.002485	1.000214
		var	N/A	0.001412908	0 .0001312621
		MSE	N/A	0.001417668	0.0001311764
	θ	*mean*	N/A	1.014813	0.9994036
		var	N/A	0.01208797	0.00119775
		MSE	N/A	0.01229532	0.001196908

(scale $\theta = 1$, shift $\omega = 1$)

Estimator			n		
			10	100	1000
GMM ($t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$, $t_4 = 0.4$, $t_5 = 0.5$, $t_6 = 0.6$)	ω	*mean*	N/A	1.001435	1.000478
		var	N/A	0.001488437	0.0001143814
		MSE	N/A	0.001489008	0.0001144960
	θ	*mean*	N/A	1.017509	1.002856
		var	N/A	0.01280900	0.001154941
		MSE	N/A	0.01310276	0.001161942
GMM ($t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$, $t_4 = 0.4$, $t_5 = 0.5$, $t_6 = 0.6$, $t_7 = 0.7$)	ω	*mean*	N/A	N/A	1.000257
		var	N/A	N/A	8.47471e-05
		MSE	N/A	N/A	8.472853e-05
	θ	*mean*	N/A	N/A	1.002155
		var	N/A	N/A	0.001120226
		MSE	N/A	N/A	0.001123747
GMM ($t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$, $t_4 = 0.4$, $t_5 = 0.5$, $t_6 = 0.6$, $t_7 = 0.7$, $t_8 = 0.8$)	ω	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A
	θ	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A
GMM ($t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$, $t_4 = 0.4$, $t_5 = 0.5$, $t_6 = 0.6$, $t_7 = 0.7$, $t_8 = 0.8$, $t_9 = 0.9$)	ω	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A
	θ	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A
GMM ($t_1 = 0.1$, $t_2 = 0.2$, $t_3 = 0.3$, $t_4 = 0.4$, $t_5 = 0.5$, $t_6 = 0.6$, $t_7 = 0.7$, $t_8 = 0.8$, $t_9 = 0.9$, $t_{10} = 1$)	ω	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A
	θ	*mean*	N/A	N/A	N/A
		var	N/A	N/A	N/A
		MSE	N/A	N/A	N/A

Table 7: The affect of α

Course grid $t_1 = 1.0$, $t_2 = 2.0$, …, $t_{10} = 10.0$ (scale $\theta = 1$, shift $\omega = 1$)

Estimator			n		
			10	100	1000
GMM ($\alpha = 0.1$)	ω	*mean*	1.042449	1.004651	1.000179
		var	0.01563071	0.001266222	0.0001243514
		MSE	0.01741698	0.001286590	0.0001242593
	θ	*mean*	1.185241	1.015554	1.001795
		var	0.2377245	0.01316198	0.001243342
		MSE	0.2718012	0.01339076	0.001245321
GMM ($\alpha = 0.01$)	ω	*mean*	1.048557	1.004171	1.000277
		var	0.01733618	0.001245997	0.0001104469
		MSE	0.01967662	0.001262148	0.0001104132
	θ	*mean*	1.157193	1.018828	1.000458
		var	0.2252576	0.01340716	0.001118598
		MSE	0.2497419	0.01374825	0.001117689
GMM ($\alpha = 0.001$)	ω	*mean*	1.045719	1.002821	1.000735
		var	0.01635575	0.0007588366	6.866032e-05
		MSE	0.01842960	0.0007660371	6.913257e-05
	θ	*mean*	1.166014	1.010516	1.005206
		var	0.2414703	0.01224280	0.001138455
		MSE	0.2687894	0.01234114	0.001164416

Table 8: The affect of α

Fine grid $t_1 = 0.1$, $t_2 = 0.2$, ..., $t_{10} = 1.0$ (scale $\theta = 1$, shift $\omega = 1$)

Estimator			n		
			10	100	1000
GMM ($\alpha = 0.1$)	ω	*mean*	1.060009	1.006262	1.001021
		var	0.03391439	0.003155521	0.0002681172
		MSE	0.03748157	0.003191576	0.0002688924
	θ	*mean*	1.213653	1.022032	1.003856
		var	0.2757228	0.01330890	0.001344458
		MSE	0.3210945	0.01378102	0.00135798
GMM ($\alpha = 0.01$)	ω	*mean*	1.055865	1.005280	1.000732
		var	0.02856582	0.002784277	0.0002754582
		MSE	0.03165814	0.002809373	0.0002757185
	θ	*mean*	1.226661	1.015048	1.002967
		var	0.2974118	0.01276048	0.001279088
		MSE	0.3484898	0.01297418	0.001286614
GMM ($\alpha = 0.001$)	ω	*mean*	1.042044	1.005541	1.000207
		var	0.02211322	0.001762504	0.000158223
		MSE	0.02385882	0.001791444	0.0001581075
	θ	*mean*	1.84459	1.020550	1.004128
		var	378.2733	0.01326644	0.001078444
		MSE	378.6083	0.01367546	0.001094410

References

[1] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. Statistical Models Based on Counting Processes, 1993, Springer-Verlag, New York.

[2] Ball, F. G., Davies, S. S. and Sansom, M. S. P. Ion Channel Gating and Time Interval Omission: Statistical Inference for a Two-State Markov Model Biological Sciences, Vol. 255, No. 1344, Mar., 22, 1994, pp. 267-272

[3] Ball, F. G. and Milne, R. K. On Choosing Values of the Transform Variables in Empirical Transform Based Inference. 25/1996, pp 1-17, Perth, Western Australia

[4] Bates, C. E. and White, H. Determination of Estimators with Minimum Asymptotic Covariance Matrices Econometric Theory, Vol. 9, No. 4, Dec., 1993, pp. 633-648

[5] Besbeas, P. and Morgan, B.J.T. Improved estimation of the stable laws Statistics and Computing archive Vol. 18, Issue 2, June, 2008, pp. 219 - 231

[6] Carrasco, M. and Florens, J. P. Generalization of GMM to a Continuum of Moment Conditions. Econometric Theory, Vol. 16, No. 6, Dec., 2000, pp. 797-834

[7] Carrasco, M. and Florens, J. P. Efficient GMM Estimation Using the Empirical Characteristic Function

[8] Corless, R.M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J. and Knuth, D. E. On the Lambert W Function. Advances in Computational Mathematics, Vol. 5, 1996, pp. 329-359

[9] Feuerverger, A. and McDunnough, P. On the Efficiency of Empirical Characteristic Function Procedures Journal of the Royal Statistical Society. Series B (Methodological), Vol. 43, No. 1, 1981, pp. 20-27

[10] Feuerverger, A. and McDunnough, P. On some Fourier methods for inference Journal of the American Statistical Association, Vol. 76, No. 374, Jun., 1981, pp. 379-387

[11] Feuerverger, A. and McDunnough, P. On Statistical Transform Methods and Their Efficiency. The Canadian Journal of Statistics / La Revue Canadienne de Statistique, Vol. 12, No. 4, Dec., 1984, pp. 303-317

[12] Garthwaite, P. H., Jolliffe, I. T. and Jones, B. Statistical Inference. Second edition, 2002, Oxford University Press, Oxford.

[13] Hansen, L. P. Large Sample Properties of Generalized Method of Moments Estimators Econometrica, Vol. 50, No. 4, Jul., 1982, pp. 1029-1054

[14] Hansen, L. P. A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators Journal of Econometrics, Vol. 30, No. 1-2, 1985, pp. 203-238

[15] Heatcote, C. R. The Integrated Squared Error Estimation of Parameters. Biometrika, Vol. 64, No. 2 (Aug., 1977), pp. 255-264

[16] Imbens, G. W. Generalized Method of Moments and Empirical Likelihood. Journal of Business and Economic Statistics, Vol. 20, No 4, 1, Oct., 2002, pp. 493-506(14)

[17] Koutrouvelis, I. A. Regression-Type Estimation of the Parameters of Stable Laws. Journal of the American Statistical Association, Vol. 75, No. 372, Dec., 1980, pp. 918-928

[18] Koutrouvelis, I. A. Estimation of Location and Scale in Cauchy Distributions Using the Empirical Characteristic Function. Biometrika, Vol. 69, No. 1 (Apr., 1982), pp. 205-213

[19] Parzen, E. On Estimation of a Probability Density Function and Mode. The Annals of Mathematical Statistics, Vol. 33, No. 3, Sep., 1962, pp. 1065-1076

[20] Paulson, A. S., Holcomb, E. W. and Leitch, R. A. The Estimation of the Parameters of the Stable Laws. Biometrika, Vol. 62, No. 1 (Apr., 1975), pp. 163-170

[21] Press, S. J. Estimation in Univariate and Multivariate Stable Distribution. Journal of the American Statistical Association, 67, 1972, pp. 842-846

[22] Robert, C. P. and Casella, G. Monte Carlo Statistical Method. Springer, New York, 1999.

[23] Weron, R. Performance of the Estimators of Stable Law Parameters. Tech. Report, HSC/95/1, Wroclaw University of Technology.

[24] Yao, Q. and Morgan, B. J. T. Empirical Transform Estimation for Tables in Empirical Transform Based Inference

[25] Yao, Q. and Morgan, B. J. T. Empirical transform estimation for indexed stochastic models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 61, No. 1, 1999, pp. 127-141(15)

[26] Yu, J. Empirical Characteristic Function Estimation and Its Applications. Econometric Reviews, Vol. 23, Issue 2, Dec., 2004, pp. 93-123