Parametric Deconvolution for a Common Heteroscedastic Case

Justin USHIZE RUTIKANGA (justinushize@aims.ac.za)

Submitted in Partial Fulfilment of a Masters of Science in Statistics at the University of the Western Cape

Department of Statistics and Population Studies

UNIVERSITY of the WESTERN CAPE

Supervisor: Professor Chris Koen
Co-Supervisor: Mrs Rechelle Jacobs
University of the Western Cape, South Africa

February 19, 2016
Abstract

There exists an extensive statistics literature dealing with non-parametric deconvolution, the estimation of the underlying population probability density when sample values are subject to measurement errors. In parametric deconvolution, on the other hand, the data are known to be from a specific distribution. In this case the parameters of the distribution can be estimated by e.g. maximum likelihood.

In realistic cases the measurement errors may be heteroscedastic and there may be unknown parameters associated with the distribution. The specific realistic case is investigated in which the measurement error standard deviation is proportional to the true sample values. In this case it is shown that the method of moments estimation is particularly simple. Estimation by maximum likelihood is computationally very expensive, since numerical integration needs to be performed for each datapoint, for each evaluation of the likelihood function.

Method of moments estimation sometimes fails to give physically meaningful estimates. The origin of this problem lies in the large sampling variations of the third moment. Possible remedies are considered.

Due to the fact that a convolution integral needed to be calculated for each datapoint, and that this has to be repeated for each iteration towards the solution, maximum likelihood computing cost is very high. New preliminary work suggests that saddlepoint approximations could sometimes be used for the convolution integrals. This allows much larger datasets to be dealt with.

Application of the theory is illustrated with simulation and real data.

Keywords

Deconvolution
Heteroscedastic measurement errors
Lognormal distribution
Maximum likelihood
Methods of moments
Saddlepoint approximation
Declaration

I declare that “Parametric Deconvolution for a Common Heteroscedastic Case” is my own work, that it has not been submitted before for any degree or examination in any other university, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Signature:

Date: February 19, 2016
Acknowledgements

I would like to thank almighty God, source of my life for giving me power and strength of thinking and writing this project.

It is a great pleasure to thank my supervisor Professor Chris Koen for his assistance and supervision which made this project possible. Thank you Mrs. Rechelle Jacobs for your assistance until the completion of this work.

I highly appreciate support from UWC especially the department of Statistics and Population Studies for its consistent assistance during my study life at the university.

My greatest gratitude is extended to AIMS staff for the amazing program, experience and various assistance which gave me enough background to work on this topic.

The English usage in this thesis was substantially improved by an anonymous language editor. Thank you so much editor.

I am thankful to my family, friends for their support and encouragement.
Contents

Abstract i
Declaration ii

1 Introduction and Objectives 1
  1.1 Introduction ................................................. 1
  1.2 Measurement errors ......................................... 1
  1.3 Remarks ..................................................... 2
  1.4 Model assumptions .......................................... 2
  1.5 Objectives .................................................. 3
  1.6 Dissertation structure ...................................... 3

2 Background 4
  2.1 Deconvolution density estimation .......................... 4
  2.2 Low-order approximation in deconvolution ................. 5
  2.3 Lognormal distribution ...................................... 7
  2.4 Saddlepoint approximation .................................. 7
  2.5 Method of maximum likelihood estimation .................. 8
  2.6 Method of moments estimation .............................. 9

3 Model Fitting for the Lognormal Distribution 10
  3.1 Introduction .................................................. 10
  3.2 Convolved lognormal distribution and distribution of heteroscedastic measurement errors .................. 10
  3.3 Estimation method ............................................ 11
  3.4 Measure of estimated parameter accuracy ................. 14

4 Simulation 16
  4.1 The impact of measurement errors ......................... 16
  4.2 Testing of goodness of fit for the model .................. 17
  4.3 Effectiveness of bootstrapping sample size ............... 19
  4.4 Standard errors of the estimated parameters by MME and MLE .................................................. 20
  4.5 Cramér-Rao lower bound (CRLB) ............................ 28
  4.6 Computation time ............................................. 30
  4.7 Saddlepoint approximation method ......................... 30

5 Application of theory to real data 33
  5.1 Introduction .................................................. 33
  5.2 Masses of star clusters in galaxy M83 ..................... 33
  5.3 Zn in Chorizon data ......................................... 34
  5.4 Ba in Chorizon data ......................................... 35
  5.5 Mg_XRF in Chorizon data .................................. 35

6 Conclusion and Findings 37
  6.1 Conclusion ................................................... 37
  6.2 Findings ..................................................... 37

7 Appendices 38
  7.1 Derivation of the roots ..................................... 38
  7.2 Cramér-Rao lower bound .................................... 40

References 43
List of Figures

4.1 Histogram of the distribution of uncontaminated and contaminated with heteroscedastic measurement error with $\sigma = 0.5, \alpha = 0.3$. ............................................. 16
4.2 Histogram of the distribution of uncontaminated and contaminated with heteroscedastic measurement error $\sigma = 0.3, \alpha = 0.5$. ............................................. 16
4.3 Lognormal distribution with and without measurement errors. ............................................. 17
4.4 P-P plot for a simulated data has 200 observed value. ............................................. 18
4.5 Illustration of distribution of the estimated parameters for the data sets of sample size $n = 50$, from the 1000 bootstrap samples with $\sigma = 0.5, \alpha = 0.3$ and $\mu = 4$. ............................................. 19
4.6 Illustration of distribution of the estimated parameters for the data sets of sample size $n = 100$, from the 1000 bootstrap samples with $\sigma = 0.5, \alpha = 0.3$ and $\mu = 4$. ............................................. 20
4.7 Distribution of $\beta$ with $\sigma, \alpha$ and $\mu$ to be 0.5, 0.3 and 4 respectively. ............................................. 21
4.8 Panels (4.8a)–(4.8f): $\sigma = 0.2$ Panels (4.8g)–(4.8l): $\sigma = 0.5$ $n = 200, \alpha = 0.5$ and $\mu = 3$. ............................................. 23
4.9 Panels (4.9a)–(4.9f): $\mu = 3$ Panels (4.9g)–(4.9l): $\mu = 4$ $n = 500, \alpha = 0.1, \sigma = 0.2$. ............................................. 25
4.10 Panels (4.10a)–(4.10f): $\mu = 3$ Panels (4.10g)–(4.10l): $\mu = 4$ $n = 1000, \alpha = 0.1, 0.3, 0.5, \sigma = 0.2$. ............................................. 27
4.11 Distribution of $\beta$ with and without values removed in the sample of datapoints. ............................................. 28
4.12 Saddlepoint approximation of convolution of lognormal and Gaussian heteroscedastic measurement errors. ............................................. 32

5.1 Quantile-Quantile plots of masses of star clusters galaxy in M83. ............................................. 34
5.2 Quantile-Quantile plot of CDF of Zn. ............................................. 34
5.3 Quantile-Quantile plot of the CDF of concentration of Ba in Chorizon. ............................................. 35
5.4 Quantile-Quantile plot for Mg XRF concentration in the environment. ............................................. 36
# List of Tables

4.1 Table of the Kolmogorov-Smirnov calculated values and 5% critical value. .......... 18  
4.2 Standard errors of estimated parameters for parameter $\mu = 3$. ..................... 22  
4.3 Standard errors of estimated parameters for parameter $\mu = 4$. ..................... 24  
4.4 Shapiro-Wilk normality tests for distribution of the estimators $\hat{\alpha}$. .......... 26  
4.5 Biased of the estimated parameters by modified MME ..................................... 28  
4.6 Cramér-Rao lower bounds for $\mu = 3$ and the sample size $n = 1000$. .......... 29  
4.7 Computation time. ......................................................................................... 30
1. Introduction and Objectives

1.1 Introduction

Measurement errors are often present in areas of research and applied statistics. An unobserved “true” variable \( X \) is measured with error, giving the observed variable \( Y \). A convolution approach is used for the analysis. The convolution problem appears in many areas of research such as econometrics, astronomy, public health or biostatistics, finance and in communication (Stefanski, 2000).

It is often important to estimate the distribution properties of the population from samples of datasets which are contaminated by measurement errors. In statistical theory, a sample is defined as a subset of a population and population as all observations of interest in the research. Normally samples are drawn from the population and statistics are calculated using the samples to make inferences or decisions about the whole population.

In this dissertation, the sample values measured are subject to heteroscedastic measurement errors. An example includes density or distribution estimation of a variable \( X \) given an observed dataset \( Y \). The estimation of parameters of the distribution from the observed random variable \( Y \) is known to be problematic if the variable \( X \) is measured with errors, especially heteroscedastic measurement errors.

Previous studies in this field generally assumed that the measurement errors follow a homoscedastic Gaussian distribution, but in many applications the assumption is problematic (Delaigle and Meister, 2008; Nawarathna and Choudhary, 2015). However, in this dissertation the observed data points are taken to be contaminated by heteroscedastic measurement errors instead. According to Stefanski (2000) “all problems of this form of measurement errors are commonly called measurement errors problems and statistical models and methods for analysing such sample data are called measurement errors models” see Stefanski (2000).

Research has shown that there exists different methods to estimate parameters of the statistical model such as parametric and nonparametric estimation, where statistical models are the density obtained by the contaminated samples. Some relevant references are Fuller (1987); Carroll et al. (1995); Koen and Kondlo (2009); Dattner and Reiser (2013).

1.2 Measurement errors

Many datasets are contaminated by the mismeasured variables and the problem of measurement errors is most common. The presence of measurement errors causes bias and inconsistent parameter estimates and leads to erroneous conclusions in data analysis (Meister, 2009).

1.2.1 Description of the measurement error problem. The model is a combination of measurement errors and the true value which gives the observed value. The model for a variable \( X \) measured with errors is

\[
Y_i = X_i + \epsilon_i \quad i = 1, 2, 3, \cdots, n
\]

(1.2.1)

where \( Y \) presents the observed variable, \( X \) is the underlying variable of interest or target variable and \( \epsilon \) presents the difference between those two variables, commonly called measurement errors. In this research, the latter are assumed to be heteroscedastic, which means its variance is not constant. The variance of each value of the measurement error depends on the value of its corresponding target value.

In this research, the measurement errors have the form of \( \epsilon_i \sim N(0, \sigma^2_i) \), \( \sigma^2_i = h(x_i) \), where \( h(.) \) is a function of the “error free data”.

Generally, there are two statistical methods for estimating the underlying population properties. One is the parametric approach and the other is non-parametric.
1.2.2 Parametric methods. The random variables \(X\) and \(\epsilon\) have probability density functions \(f_X(.)\) and \(f_\epsilon(.)\) respectively. It is assumed that the general forms of both densities are known. The main goal is to estimate the parameters for both densities (Delaigle and Hall, 2015).

Chapter 3 contains more details about the case where \(X\) is drawn from a lognormal distribution and \(\epsilon\) follows the normal distribution. The method of moments and maximum likelihood estimation are recommended and performed in Chapters 3 and 4 for parameters involved in the density function of random variable \(Y\).

1.2.3 Non-parametric methods. In the case of the non-parametric approach \(f_X(.)\) is unspecified. Considering the random variables \(X\) with density function noted \(f_X(.)\) and \(\epsilon\) with density function \(f_\epsilon(.)\), then the probability density function \(f_Y(.)\) of random variable \(Y\) is obtained by the convolution integral (Delaigle and Hall, 2015; Meister, 2009; De Brabanter and De Moor, 2012),

\[
f_Y(y) = \int_{-\infty}^{\infty} f_X(x)f_\epsilon(y-x)dx,
\]

(1.2.2)

and cumulative distribution

\[
F_Y(y) = \int_{-\infty}^{\infty} F_X(x)f_\epsilon(y-x)dx.
\]

(1.2.3)

As mentioned above, our task is to estimate the probability density function of \(f_X(.)\) and \(\epsilon\) from the observed data \(y_1, y_2, \cdots, y_n\) presented in Equation (1.2.1).

The Fourier transform method has been used for statistical estimation for the deconvolution problem. The studies of Carroll and Hall (1988); Stefanski and Carroll (1990) introduced kernel deconvolution estimators, Carroll and Hall (2004) used the weighted kernel density for the method of density deconvolution; McIntyre and Stefanski (2011a) addressed the semi-parametric deconvolution for the density function of data measured with errors and McIntyre and Stefanski (2011b) proposed the deconvolution of the form of independent replicated data with measurement errors.

In Chapter (2) different methods of non-parametric deconvolution are discussed.

1.3 Remarks

In this research all distributional parameters are assumed to be unknown. The main objectives are focused on estimating parameters using maximum likelihood estimation and method of moments. The accuracy of each of these methods are illustrated. Simulation can be used to determine the distribution of estimated parameters for a given sample size and for certain parameters values.

1.4 Model assumptions

In order to construct any estimation method for analysis of the effect of measurement errors, one needs to make some assumptions about the process which generates the differences between the underlying variable of interest \(X\) and the observed value \(Y\).

The two general assumptions that underlie the measurement errors model are as follows:

i The random variables \(X\) and measurement errors \(\epsilon\) are uncorrelated.

ii The measurement errors follow a heteroscedastic Gaussian distribution.

These assumptions have been used in most of the studies related to deconvolution problems. Most of the previous authors assumed that the measurement errors are independent and identically distributed which is not always realistic. In this project the variance of the measurement errors is assumed to be proportional to each corresponding value of underlying variable \(X\). Therefore, except where otherwise stated, these assumptions will hold.
1.5 Objectives

The work presented in this project aims to:

1. Review the statistical literature on the deconvolution distribution function.
2. Develop the statistical model for deconvolution when the distribution function of a target random variable is known to be lognormal.
3. Estimate unknown parameters.

1.6 Dissertation structure

This dissertation is introduced in Chapter 1. Chapter 2 consists of a study of some of the relevant deconvolution literature. The selection of the methods proposed in the literature is primarily based on three criteria namely popularity, appealing logic, and simplicity. The various methods are summarised in Chapter 3.

Parametric estimation based on a sample drawn from a population with a lognormal distribution is dealt with in Chapter 4. Chapter 5 illustrates the application of the theory with real data. The conclusion and findings for future work is contained in Chapter 6. Extended results and some algebraic formula are presented in the appendix Chapter 7.
2. Background

Introduction

In this chapter, some existing deconvolution procedures applied to density estimation methods are discussed. The most popular and simplest methods to implement have been selected.

2.1 Deconvolution density estimation

The purpose of deconvolution is to estimate the density $f_X$ of the random variable $X$, given the observed data $y_1, \cdots , y_n$, as in Equation (1.2.1). The $X_i$'s and $\epsilon_i$'s are each identically independently distributed. In this case, the classical techniques of deconvolution can then be used to estimate $f_X$, assuming that the density function $f_\epsilon$ of $\epsilon$ is known (Delaigle and Meister, 2011).

2.1.1 Classical deconvolution. Let $\phi_g$ be the Fourier transform of the function $g = g(X)$:

$$\phi_g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(x)e^{-itx}dx . \quad (2.1.1)$$

If $g$ is the PDF of $X$ then $\phi_g$ is referred to as the characteristics function of $X$. It follows that $f_Y = f_X * f_\epsilon$, which is equivalent to $\varphi_Y = \varphi_X \varphi_\epsilon$, where $*$ is the deconvolution. If $\varphi_\epsilon \neq 0 \ \forall \ t \in R$, then the following expression holds:

$$\varphi_X = \frac{\varphi_Y}{\varphi_\epsilon} \quad \forall \ t. \quad (2.1.2)$$

The deconvolution kernel density estimator (Stefanski and Carroll, 1990) is then given by

$$f_X = \frac{1}{2\pi} \int e^{-itx}K(ht)\frac{\varphi_Y(t)}{\varphi_\epsilon(t)}dt \quad (2.1.3)$$

where

$$\varphi_Y(t) = \frac{1}{n} \sum_{j=1}^{n} e^{itY_j}, \quad \text{for} \quad j = 1, 2, \cdots , n \quad (2.1.4)$$

and $h > 0$ is a smoothing parameter called the bandwidth, the $K$ is a symmetric function known as the kernel.

If the integral in Equation (2.1.3) exists then $K(ht)\hat{\varphi}_Y(t)$ is an estimator of $\varphi_Y$ (Delaigle and Meister, 2011).

2.1.2 Deconvolution kernel estimation. The most useful approach to estimate the density distribution of data contaminated by measurement errors is the deconvolution kernel estimator method (see Stefanski and Carroll 1990; Xiao-Feng et al. 2009). This method consists of estimating the density function $f_X$ of the random variable $X$ by first finding the characteristic function $\varphi_X(t)$ of random variable $X$ and then the inverse of the Fourier transform which results in the estimate of $f_X$ (Stefanski and Carroll, 1990). Mathematically, the characteristic density function of $X$ is

$$\varphi_X(t) = E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx}f_X(x)dx \quad (2.1.5)$$

and has the inverse transform

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}\varphi_X(t)dt, \quad t \in \mathbb{R}, \quad i = \sqrt{-1} \quad (2.1.6)$$
provided $\varphi_X$ is absolutely integrable.

Since the random variables $X$ and $\epsilon$ are assumed to be independent, the characteristic function of density $f_Y$ is given by

$$\varphi_Y(t) = E(e^{itx})E(e^{it\epsilon}) = \varphi_X(t)\varphi_\epsilon(t).$$  \hfill (2.1.7)

The characteristic function of $X$

$$\varphi_X(t) = \frac{\varphi_Y(t)}{\varphi_\epsilon(t)}$$  \hfill (2.1.8)

being substituted in Equation (2.1.6) the density $f_X$ is expressed as

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\varphi_Y(t)}{\varphi_\epsilon(t)} dt$$  \hfill (2.1.9)

with estimator

$$\hat{f}_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \hat{\varphi}_n(t) \frac{\hat{\varphi}_Y(t)}{\hat{\varphi}_\epsilon(t)} dt.$$  \hfill (2.1.10)

The estimator of the characteristic function $\varphi_Y(t)$ is given by

$$\hat{\varphi}_Y(t) = \hat{\varphi}_n(t)\varphi_K(ht)$$  \hfill (2.1.11)

where

$$\hat{\varphi}_n(t) = \frac{1}{n} \sum_{j=1}^{n} e^{itY_j}$$  \hfill (2.1.12)

is the corresponding empirical characteristic function of $f_Y$, $h > 0$ is the bandwidth parameter and $\varphi_K(t)$ is the Fourier transform of the kernel function $K$. Then, the density estimator in (2.1.10) becomes

$$\hat{f}_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \hat{\varphi}_n(t) \frac{\varphi_K(ht)}{\varphi_\epsilon(t)} dt$$  \hfill (2.1.13)

provided the function $\frac{\varphi_K(ht)}{\varphi_\epsilon(t)}$ is integrable. Then, according to Carroll and Hall (1988); Stefanski and Carroll (1990), the deconvolution kernel density estimator of $f_X$ is given by

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^{n} K_u\left(\frac{x - y_i}{h}\right).$$  \hfill (2.1.14)

$$K_u(w) = \frac{1}{2\pi} \int e^{-itw} \frac{\varphi_K(t)}{\varphi_u(t/h)} dt.$$  \hfill (2.1.15)

\section*{2.2 Low-order approximation in deconvolution}

Based on the study by Carroll and Hall (2004), constructing consistent estimators for a given deconvolution problem is an unattainable objective. In practice, good results may be obtained by constructing a less ambitious low-order approximation of $f_X$, and accurately estimating the approximation rather than the density $f_X$. The motivation of the method suggested by Carroll and Hall (2004) is to express the expected value of kernel estimators of $f_X$ with the observation as a series expansion in expectations of kernel estimators of derivatives of $\epsilon$ (Carroll and Hall, 2004).
To illustrate this method, let $Y_j = X_i + \epsilon_i$ for $1 \leq i \leq n$ be the data sample size from the model of Equation (1.2.1). Assume that the lower order moments of the distribution of $\epsilon$ are known and $E(\epsilon) = 0$. Based on this, the density $f_X$ of $X$ can be estimated.

Let $f_Y, f_X$ denote the density of the distributions of $Y$ and $X$ respectively. The kernel density estimators of $f_Y$ and $f_X$ are given by,

$$
\hat{f}_Y(y) = \frac{1}{nB} \sum_{i=1}^{n} K \left( \frac{y - Y_i}{B} \right)
$$

$$
\hat{f}_X(x) = \frac{1}{nB} \sum_{i=1}^{n} K \left( \frac{x - X_i}{B} \right),
$$

(2.2.1)

where $K$ is kernel and $B$ is a bandwidth. Since the $X_i$ are not observed, $\hat{f}_X$ can not directly be calculated from data. A good approximation of $f_X$ and its expected value can be derived. This motivates the methodology for this study.

Suppose that $\mu'$ is a moment of the distribution of $\epsilon$, which is finite, and that the $K$ kernel is an analytic function, where all the derivatives of $K$ are well defined on the whole real line.

Let the $s$th derivative of $f$ be denoted by $f^{(s)}$. Then, the $s$th order approximation of the expected value of $\hat{f}_X^{(s)}$ for $1 \leq j$ is:

$$
E \left[ \hat{f}_X^{(s)}(x) \right] = E \left[ \hat{f}_Y^{(s)}(x) \right] + \sum_{j=1}^{\infty} \sum_{t_1 = 1}^{\infty} \cdots \sum_{t_j = 1}^{\infty} \frac{(-1)^{t_1 + \cdots + t_j + j}}{t_1! \cdots t_j!} \mu_{1}^{t_1} \cdots \hat{\mu}_{t_j}^{t_j} E \left[ \hat{f}_Y^{(t_1 + \cdots + t_j + s)}(x) \right].
$$

(2.2.2)

Therefore, the $s$th derivative of $f$ denoted by $f^{(s)}$ is given by

$$
\hat{f}_{X,v}^{(s)}(x) = \hat{f}_Y^{(s)}(x) + \sum_{j=1}^{\infty} A_{v} \hat{f}_Y^{(t_1 + \cdots + t_j + s)}(x),
$$

(2.2.3)

where

$$
A_{v} = \sum_{t_1 \cdots t_j \geq 1 : t_1 + \cdots + t_j \leq v} \frac{(-1)^{t_1 + \cdots + t_j + j}}{t_1! \cdots t_j!} \mu_{1}^{t_1} \cdots \hat{\mu}_{t_j}^{t_j}.
$$

(2.2.4)

It is assumed that the errors $\epsilon$ are normally distributed. Then in this case, $\hat{\mu}_{s} = 0$ for odd $s$. The approximation of even order is considered. Thus, Equation (2.2.3) becomes

$$
\hat{f}_{X,2v}^{(s)}(x) = \hat{f}_Y^{(s)}(x) + \sum_{j=1}^{\infty} A_{2v} \hat{f}_Y^{(2t_1 + \cdots + 2t_j + s)}(x)
$$

(2.2.5)

where

$$
A_{2v} = \sum_{t_1 \cdots t_j \geq 1 : t_1 + \cdots + t_j \leq 2v} \frac{(-1)^{t_1 + \cdots + t_j}}{2t_1 \cdots 2t_j} \mu_{1}^{2t_1} \cdots \hat{\mu}_{t_j}^{2t_j}.
$$

(2.2.6)

Hence, the higher order approximation can be determined by standard kernel estimators based on the adjusted kernels,

$$
K_{v}(x) = K(x) + \sum_{j=1}^{\infty} A_{v} K^{(t_1 + \cdots + t_j)}(x).
$$

(2.2.7)

Therefore,

$$
\hat{f}_{X,v}^{(s)}(x) = \frac{1}{nB^{s+v}} \sum_{i=1}^{n} K_{v}^{(s)} \left( \frac{x - Y_i}{B} \right).
$$

(2.2.8)

In the error-free case, the observations are of course assumed to be uncontaminated by measurement errors. This condition on $f_X$ is commonly used in kernel density estimation where it is usually assumed that $v = 2$. The finite moment is $\mu'_2 = \sigma^2$, and the estimator is simplified to

$$
\hat{f}_X(x) = \hat{f}_Y(x) - \frac{\sigma^2}{2} \hat{f}_Y^{(2)}(x)
$$

(2.2.9)
where $\hat{f}_Y$ is an error free kernel estimator of $f_Y$ which is estimated directly using the observed data. In model (2.2.9) $\hat{f}_Y^{(2)}$ is the second derivative of $\hat{f}_Y$ and $\sigma$ is one of the unknown parameters in the model that can be estimated using the empirical variance of the difference of replicated observations.

Delaigle (2008) shows that if the error variance is small, then the low-order approximation approach can outperform the deconvolution kernel estimator when it is not easy to solve the deconvolution problem.

However, the kernel density estimator works better than the low-order method of estimation if the error variance is large.

Finally, the applicability of the estimator $\hat{f}_X$ in Equation (2.2.9) is not limited to cases where the characteristic function of the errors does not vanish. This approach makes it possible to estimate of the derivative of $f_X$ itself.

### 2.3 Lognormal distribution

#### 2.3.1 Description of the lognormal distribution

A random variable $X$ is said to follow a lognormal distribution, with scale and location parameter $\mu$ and $\sigma$ when $\log(X)$ has a normal distribution with mean $\mu$ and variance $\sigma^2$.

The parameters $\mu$ and $\sigma$ are positive real numbers, while $\mu$ and $\sigma^2$ are the mean and variance of random variable $\log(X)$, not for random variable $X$.

Formally

$$f_X(x) = \frac{1}{x \sqrt{2\pi \sigma}} \exp \left[ -\frac{(\log(x) - \mu)^2}{2\sigma^2} \right]$$  \hspace{1cm} (2.3.1)

is the lognormal PDF.

### 2.4 Saddlepoint approximation

#### 2.4.1 Saddlepoint approximation for convolution

It is often required to approximate the distributions of some statistics whose exact distributions cannot be easily determined. Then, saddlepoint approximation is one of the common techniques used to obtain the approximate densities and distribution functions (Holly and Phillips, 1979; Goutis and Casella, 1999; Daniels, 1954).

One of the motivations for applying the saddlepoint approximation method is its link to the Taylor series expansion and Laplace approximation of integrals. In the case of convolution, the saddlepoint approximation is used because the computation of numerical integrals for each datapoint is time consuming.

#### 2.4.2 Derivation of the saddlepoint approximation

The saddlepoint approximation is derived from a natural sequence of approximations that becomes progressively more local (Xiong et al., 2005). Let $f(x)$ be a positive function that has an approximation at the point $x_0$. One way to approximate is to use the first few terms of a Taylor series expansion. This may not be applied to $f(x)$, but by using the function $g(x)$ such that $f(x) = \exp(g(x))$ (Goutis and Casella, 1999). The Taylor series expansion of $g(x)$ around a chosen point $x_0$ is given by

$$g(x) \approx g(x_0) + (x - x_0)g'(x_0) + \frac{(x - x_0)^2}{2!}g''(x_0) + \cdots.$$  \hspace{1cm} (2.4.1)

By considering the first three terms of the Taylor series expansion,

$$f(x) \approx \exp \left( g(x_0) + (x - x_0)g'(x_0) + \frac{(x - x_0)^2}{2!}g''(x_0) \right),$$  \hspace{1cm} (2.4.2)
the equation (2.4.2) is simplified to

\[ f(x) \approx \exp \left( g(\hat{x}) + \frac{(x - \hat{x})^2}{2} g''(\hat{x}) \right) \quad (2.4.3) \]

where \( \hat{x} = x_0 \), if the first derivative vanishes at point \( \hat{x} \).

The integral of the positive function \( f(x) \) is given by

\[
\int f(x) \, dx \approx \int \exp \left( g(\hat{x}) + \frac{(x - \hat{x})^2}{2} g''(\hat{x}) \right) \, dx \\
= \exp(g(\hat{x})) \int \exp \left( \frac{(x - \hat{x})^2}{2} g''(\hat{x}) \right) \, dx 
\quad (2.4.4)
\]

In the case where \( \hat{x} \) gives a maximum of \( g(x) \), then the second derivative of \( g(x) \) at point \( \hat{x} \) is negative. The integral can be easily evaluate by noting the normalisation of a normal density with mean \( \hat{x} \) and variance \( \frac{1}{g''(\hat{x})} \):

\[
\int f(x) \, dx \approx \exp(g(\hat{x})) \left( \frac{2\pi}{g''(\hat{x})} \right)^{\frac{1}{2}} \, dx .
\quad (2.4.5)
\]

As an illustration of the saddlepoint approximation, the distribution of the sum of two random variables is derived by approximating the convolution integral. See details of application of the saddlepoint approximation in Section 4.7.

### 2.5 Method of maximum likelihood estimation

The method of maximum likelihood estimation consists of estimating parameters of the statistical model of the given observation, \( x_1, \ldots, x_n \). The \( x_1, x_2, \ldots, x_n \) are from distributions that depend on the vectors of parameters \( \theta = (\theta_1, \theta_2, \theta_3, \ldots, \theta_k) \). In most cases the sample of \( n \) observations is drawn from the same distribution. The joint probability distribution of the observations \( x_i \) is based on a function of vectors \( \theta \) of unknown parameters and is called the “Likelihood Function” of the sample (Scholz, 1985). For independent \( x_i \) from PDF \( f_X \).

\[
\mathcal{L}(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} f_X(x_i, \theta) \quad (2.5.1)
\]

The maximum likelihood estimator is obtained by finding the roots of the first derivatives of the likelihood function with respect to the elements of the parameter vector \( \theta \), provided these exist:

\[
\mathcal{L}'(X, \theta) = \frac{\partial \mathcal{L}(X, \theta)}{\partial \theta} = 0. 
\quad (2.5.2)
\]

If more than one root is obtained, the value corresponding to the largest likelihood is selected as a global maximum. One of the sufficient conditions for a maximum is that the second derivative of likelihood function is negative,

\[
\mathcal{L}''(X, \theta) < 0. 
\quad (2.5.3)
\]

In practice, it is often easier to work with the logarithm of the likelihood function than with the function itself,

\[
\log (\mathcal{L}(X, \theta))' = \frac{\mathcal{L}'(X, \theta)}{\mathcal{L}} = 0 
\quad (2.5.4)
\]

provided that likelihood function is different to zero (Ruppert, 2004).
It is often simpler and easier to solve Equation (2.5.4) than Equations (2.5.2) and (2.5.1).

**Note:**
When the equations for the maximum likelihood estimation cannot be solved easily, this causes two problems, how to solve the equations and how to determine the standard errors of the maximum likelihood estimation.

Section (3.3.1) illustrates the application of maximum likelihood estimation to the convolution of the lognormal and normal distributions.

### 2.6 Method of moments estimation

The method of moments estimation is one of the oldest methods used to determine the unknown parameters of statistical and mathematical models. This method provides a consistent estimator, but is not efficient compared to the maximum likelihood method. It is often used because it has a very simple procedure for finding an estimator of one or more unknown parameters. Its algebraic computation is cheap also (Asquith, 2014).

Given the $q^{th}$ population moment of the random variable $X$ is written as $\mu_q' = E(X^q)$, there exists $m_q' = \frac{1}{n} \sum_{i=1}^{n} X_i^q$ as the corresponding sample moment, a consistent estimator of $\mu_q'$. Assuming that the distribution of the random variable $X$ has parameters $\zeta$ and $\eta$, the moments of $X$ will be a function of $\zeta$ and $\eta$ given by,

$$
\mu_1' = h_1(\zeta, \eta) \\
\mu_2' = h_2(\zeta, \eta).
$$

(2.6.1)

In general, the number of moment equations depends on the number of the unknown parameters which are in the distribution of $X$. Thus, the estimator of $\zeta$ and $\eta$ by method of moments is obtained by substituting the population moments $\mu_q'$ for $q = 1, 2$ in Equation (2.6.1) by the sample moments $m_q'$ for $q = 1, 2$ which leads to

$$
m_1' = h_1(\hat{\zeta}, \hat{\eta}) \\
m_2' = h_2(\hat{\zeta}, \hat{\eta}).
$$

(2.6.2)

Section (3.3.2) illustrates the application of method of moments estimation to the convolution of the lognormal and normal distributions.
3. Model Fitting for the Lognormal Distribution

3.1 Introduction

In many applications, an important problem in measurement error models is to estimate the unknown parameters, to draw conclusions about their accuracy and to recover the unknown density function of a variable. Most studies assume that the errors are independent and identically distributed for all observations which is not always realistic. Since the distribution function of errors varies with each observation, the measurement errors are heteroscedastic.

Let \( X \) be the target variable which we cannot observe directly. Assume \( Y_i = X_i + \epsilon_i \quad i = 1, \ldots, n \). One is interested in estimating the unknown density function of \( X \).

Most of the time the distribution of \( \epsilon \) is assumed known. The distribution of \( Y = X + \epsilon \) is obtained by the convolution integral (1.2.3).

3.2 Convolved lognormal distribution and distribution of heteroscedastic measurement errors

A model for a random variable contaminated by heteroscedastic measurement errors is,

\[
Y_i = X_i + \epsilon_i .
\]

Assume that the standard deviation is \( \sqrt{\text{Var}(\epsilon_i)} = h(x_i) \). This is a reasonable model for the covariance which introduces heteroscedasticity in the measurement error model. It shows that each measurement error depends on each subject or each observed value. Consider the model

\[
\begin{align*}
X_i & \sim \log N(\mu, \sigma) \\
\epsilon_i & \sim N(0, \sigma^2_{\epsilon}) \\
\sigma^2_{\epsilon_i} & = h(x_i) .
\end{align*}
\]

In this study, \( h(x) = \alpha x \) will be used to model the standard deviation of the heteroscedastic measurement errors of the model, which are normally distributed (Xiao-Feng and Bin, 2011). Assume that the variable \( X \) follows the lognormal distribution with PDF \( f_X(.) \), the measurement error \( \epsilon \) has PDF \( f_\epsilon(.) \) and the random variable \( Y \) has PDF \( f_Y(.) \) which is given by the convolution integral

\[
f_Y(y) = \int_0^\infty f_X(x)f_\epsilon(y-x)dx.
\]

The PDF of \( X \) is given by,

\[
f_X(x) = \frac{1}{\sqrt{2\pi}x\sigma} \exp\left\{ \frac{-1}{2\sigma^2} (\log(x) - \mu)^2 - \log(x) \right\},
\]

and the PDF of \( Y \) contaminated by zero mean Gaussian heteroscedastic measurement errors

\[
f_Y(y) = \int_0^\infty \frac{1}{\sqrt{2\pi}x\sigma} \exp\left\{ \frac{-1}{2\sigma^2} (\log(x) - \mu)^2 \right\} \frac{1}{\sqrt{2\pi}\alpha x} \exp\left\{ \frac{-1}{2\alpha^2x^2} (y-x)^2 \right\} dx.
\]

\[
= \frac{1}{\sqrt{2\pi}\alpha\sigma} \int_0^\infty \exp\left\{ \frac{-1}{2\sigma^2} (\log(x) - \mu)^2 - 2\log(x) - \frac{1}{2\alpha^2x^2} (y-x)^2 \right\} dx
\]
where \( \sigma \) is the scale parameter, \( \mu \) is the location parameter of the lognormal distribution and \( \alpha \) is the error standard deviation.

The CDF of \( X \) is given by integrating the probability function

\[
F_X(x) = \int_0^\infty f_X(x)dx = \Phi \left( \frac{\log(x) - \mu}{\sigma} \right)
\]

where \( \Phi(.) \) is the cumulative distribution function of the standard normal distribution. The corresponding CDF for the convolution integral \( f_Y(.) \) is given by

\[
F_Y(y) = \int_0^\infty F_X(x)f_x(y-x)dx = \int_0^{\infty} f_Y(x)dX.
\]

### 3.2.1 Proposition

If \( X \) and \( \epsilon \) respectively follow lognormal and normal distributions, they are uncorrelated.

**Proof:**

Let \( X \sim \log N(\mu, \sigma^2) \) and \( \epsilon \sim N(b, \alpha^2 X^2) \).

Then,

\[
E[X\epsilon] = \int_0^\infty \int_{-\infty}^{\infty} x\epsilon f_{X,\epsilon}(x, \epsilon) d\epsilon dx.
\]

\[
= \int_0^\infty \int_{-\infty}^{\infty} x\epsilon f_{X|x}(\epsilon|x)f_X(x) d\epsilon dx
\]

\[
= \int_0^\infty \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{1}{2\sigma^2}(\log(x) - \mu)^2 \right) \frac{1}{\sqrt{2\pi\alpha x}} \exp \left( -\frac{1}{2\alpha^2}(\epsilon - b)^2 \right) d\epsilon dx
\]

\[
(3.2.10)
\]

hence,

\[
= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2}(\log(x) - \mu)^2 \right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\alpha x}} \exp \left\{ -\frac{1}{2\alpha^2}(\epsilon - b)^2 \right\} d\epsilon dx.
\]

\[
= \exp \left[ \mu + \sigma^2/2 \right] b
\]

\[
E[X\epsilon] = E[X]E[\epsilon].
\]

### 3.3 Estimation method

There are many ways to estimate a parameter, given the same data set. A straightforward way of determining an estimate is called a natural estimate. Therefore, it is powerful to have a larger number of applicable methods of estimation which are easy and sensible to apply. The maximum likelihood method and method of moments are examples.
3.3.1 Maximum likelihood estimation. Assume that \( \phi = [\sigma, \alpha, \mu] \) is the vector of parameters and \( \hat{\phi} = [\hat{\sigma}, \hat{\alpha}, \hat{\mu}] \) the vector of parameter estimates. Therefore, for independent observed values \( \{y_1, y_2, \ldots, y_n\} \), the likelihood for the density \( f_Y(y; \phi) \) has the following form

\[
\ell(y_1, y_2, \ldots, y_n; \phi) = \prod_{i=1}^{n} f_Y(y_i; \phi) \tag{3.3.1}
\]

where

\[
f_Y(y) = \int_{0}^{\infty} \frac{1}{2\pi\sigma} \exp \left[ -\frac{1}{2\sigma^2} (\log(x) - \mu)^2 - 2\log(x) \right] \exp \left[ -\frac{1}{2\sigma^2} (y - x)^2 \right] dx. \tag{3.3.2}
\]

The likelihood function of the PDF is given by

\[
\ell(y; \phi) = \left[ \frac{1}{2\pi\sigma} \right]^n \prod_{i=1}^{n} \int_{0}^{\infty} \exp \left[ -\frac{1}{2\sigma^2} (\log(x) - \mu)^2 - 2\log(x) - \frac{1}{2\sigma^2} (y_i - x)^2 \right] dx. \tag{3.3.3}
\]

Often it is mathematically easier to maximise the log-likelihood function instead of the likelihood function itself. The log-likelihood function corresponding to (3.3.3) is given by

\[
\mathcal{L} = \log(\ell(y; \phi)) = n [-\log(2\pi) - \log(\sigma) - \log(\alpha)] + \sum_{i=1}^{n} \log \int_{0}^{\infty} \exp \left[ -\frac{1}{2\sigma^2} (\log(x) - \mu)^2 - 2\log(x) - \frac{1}{2\sigma^2} (y_i - x)^2 \right] dx. \tag{3.3.4}
\]

The partial derivative with respect to the unknown parameters is

\[
\frac{\partial \mathcal{L}}{\partial \phi_i} = 0 \quad \text{for} \quad i = 1, 2, 3. \tag{3.3.5}
\]

Considering the method of maximum likelihood, the estimates of \( \sigma, \alpha \) and \( \mu \) maximize the observed likelihood function in Equation (3.3.3) or equivalently its logarithm in Equation (3.3.4). The first and second derivatives of log-likelihood are illustrated in the Appendices.

To solve analytically for \( \sigma, \alpha \) and \( \mu \) is not possible. Alternatively a way to solve for the parameters is by direct numerical maximisation of the log-likelihood function in Equation (3.3.4).

3.3.2 Method of moment estimation.

3.3.2.1 Estimated parameters. Considering the \( r \)th population moment of random variable \( Y \) given by \( \mu'_r = E(Y^r) \). For given values \( Y_1, Y_2, \ldots, Y_n \) independent identically distributed, \( Y \) has \( r \)th sample moment \( M'_r = \frac{1}{n} \sum_{i=1}^{n} Y_{r}^{i} \).

Let \( Y = X + \epsilon \) where \( X \sim \log N(\mu, \sigma) \) and \( \epsilon \sim N(0, \alpha^2X^2) \). This implies that \( \epsilon = \alpha X Z \) where \( Z \) is the standard normal distribution \( Z \sim N(0, 1) \).

Therefore \( Y = X + \epsilon = X + \alpha X Z = X(1 + \alpha Z) = XW \)

where \( X \) has PDF \( f_X(.) \) and \( W \sim N(1, \alpha^2) \) with PDF \( f_W(.) \).

The moments of \( Y \) are given by the moment of product of the moments of random variables \( X \) and \( W \):

\[
\mu'_r = E[Y^r] = E[(XW)^r] = E[X^r]E[W^r]. \tag{3.3.6}
\]

The moments of the lognormal distribution are

\[
\begin{align*}
\mu'_1 &= \exp(\mu + \sigma^2/2), \\
\mu'_2 &= \exp(2\mu + 2\sigma^2), \\
\mu'_3 &= \exp(3\mu + \frac{3\sigma^2}{2}).
\end{align*} \tag{3.3.7, 3.3.8, 3.3.9}
\]
Raw moments of \( W \sim N(1, \alpha^2) \) are

\[
\begin{align*}
\mu'_1 & = 1, \\
\mu'_2 & = (1 + \alpha^2), \\
\mu'_3 & = (1 + 3\alpha^2).
\end{align*}
\] (3.3.10, 3.3.11, 3.3.12)

Hence, the sample moments of random variable \( Y \) are given by

\[
\begin{align*}
\mu'_1 & = \mu'_1 X, \\
\mu'_2 & = \mu'_2 X = \exp(\mu + \sigma^2/2), \\
\mu'_3 & = \mu'_3 X = \exp(3\mu + 9\sigma^2/2).
\end{align*}
\] (3.3.13)

Equating sample and population moments and taking logarithms:

\[
\begin{align*}
\log(M'_1) & = \mu + \sigma^2/2 \\
\log(M'_2) & = 2\mu + 2\sigma^2 + \log(1 + \alpha^2) \\
\log(M'_3) & = 3\mu + 9\sigma^2/2 + \log(1 + 3\alpha^2).
\end{align*}
\] (3.3.14)

Solving the algebraic expression for \( \alpha^2 \) gives

\[
\alpha^6 + 3\alpha^4 + 3\alpha^2 - \frac{3\alpha^2 M_2^3}{M_1^3 M_3^3} - \frac{M_2^3}{M_1^3 M_3^3} + 1 = 0.
\] (3.3.15)

Let \( x = \alpha^2 \) and \( \beta = \frac{M_2^3}{M_1^3 M_3^3} \)

\[
\alpha^6 + 3\alpha^4 + 3\alpha^2 - 3\alpha^2 \beta - \beta + 1 = 0.
\] (3.3.16)

Letting \( \alpha^2 = x \)

\[
x^3 + 3x^2 + 3x - 3x\beta - \beta + 1 = 0
\]

\[
x^3 + 3x^2 + 3(1 - \beta)x + (1 - \beta) = 0.
\] (3.3.17)

The transformation \( x = y + h \) leads to depressed cubic

\[
y^3 - 3\beta y + 2\beta = 0.
\] (3.3.18)

Realistically, \( \beta > 0 \). If \( \beta < 1 \), then \( \Delta > 0 \) and equation (3.3.18) has two complex conjugate and one negative root and hence no realistic solutions. If \( \beta > 1 \), \( \Delta < 0 \) and the roots of Equation (3.3.18) are given by

\[
x_k = y_k - 1
\]

\[
= 2\sqrt{\beta} \cos \left( \frac{\arccos(-\sqrt{1/\beta}) + 2\pi(k - 1)}{3} \right) - 1
\] (3.3.19)

where \( k = 1, 2, 3 \).

Hence, \( x = \alpha^2 \) where only the positive real root is considered in this situation.

Therefore,

\[
\alpha^2 = 2\sqrt{\beta} \cos \left( \frac{\arccos(-\sqrt{1/\beta})}{3} \right) - 1.
\] (3.3.20)
All the estimated parameters are positive real numbers if and only if the \( M'_3 = \beta > 1 \) condition holds,

\[
\hat{\alpha}^2 = 2\sqrt{\beta} \cos \left( \frac{\arccos(-\sqrt{1/\beta})}{3} \right) - 1
\]
\[
\hat{\sigma}^2 = \log(M'_2) - 2\log(M'_1) - \log \left( 2\sqrt{\beta} \cos \left( \frac{\arccos(-\sqrt{1/\beta})}{3} \right) \right)
\]
\[
\hat{\mu} = \log(M'_1) - \frac{1}{2} \left[ \log(M'_2) - 2\log(M'_1) - \log \left( 2\sqrt{\beta} \cos \left( \frac{\arccos(-\sqrt{1/\beta})}{3} \right) \right) \right].
\] (3.3.22)

More details of the derivation of the roots are presented in the appendix. The numerical calculation will be implemented in MATLAB.

### 3.4 Measure of estimated parameter accuracy

Accuracy of parameter estimates is mostly measured by approximate confidence intervals using the standard errors. There are several methods to calculate the standard errors. In this study three methods will be discussed: the bootstrap method, asymptotic errors based on the Fisher information matrix, and jackknife procedure.

#### 3.4.1 Bootstrap method.

Bootstrapping is based on sampling with replacement from the original sample of \( n \) elements. The number of bootstrap samples should be large enough to fully delineate the distributions of the parameters estimate. This study uses the parametric bootstrap.

In general, parametric bootstrapping supposes that a parametric model for the original data of the given model \( G_Y(y; \theta) \), is known and the vector of parameters \( \theta \) is unknown. The bootstrap data are drawn from \( G_Y(y; \hat{\theta}) \), where \( \hat{\theta} \) is as usual, the maximum likelihood estimate from the original data. The procedure for the parametric bootstrap requires the following steps:

- Assume \( \hat{\theta} \) is the estimate of \( \theta \) obtained from the data, for example, the maximum likelihood estimate. Let \( Y^* \) be a simulated sample of \( n \) observation from the model \( G_Y(\cdot; \hat{\theta}) \).
- Determine \( \hat{\theta}^* = \hat{\theta}(Y^*) \), in the same way \( \hat{\theta} \) was obtained.
- Repeat the preceding two steps \( B \) times to get an estimate of the parametric bootstrap distribution of \( \hat{\theta} \).

Let us assume that there are \( B \) bootstrap estimates of \( \theta \), the underlying parameter. For each element \( \theta_i \) of the vector \( \theta \), sort the bootstrap replicates into ascending order such that \( \hat{\theta}^*_i(1) \leq \cdots \leq \hat{\theta}^*_i(B) \). Further, estimating a confidence interval as usual requires an estimate at \( 100\alpha \) percentile of the bootstrap distribution. The upper and lower confidence bounds are \( B(1-\alpha) \) and \( B\alpha \) respectively for sorted elements (Davison and Hinkley, 1997). As a result the percent confidence interval of \( \hat{\theta} \) is given by

\[
\left[ \hat{\theta}^*_i(1-\alpha); \hat{\theta}^*_i(B(1-\alpha)) \right].
\] (3.4.1)

It is noted in passing that these confidence intervals may be inappropriate due to bias in parameter estimates and skewness of their sampling distributions. “Basic bootstrapping” confidence intervals may be better – see Davison and Hinkley 1997.

Estimate the standard error \( se(\hat{\theta}) \)

\[
\hat{se}(\hat{\theta}) = \left[ \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^*_b - \hat{\theta}^*_1)^2 \right]^{\frac{1}{2}}.
\] (3.4.2)
where
\[ \hat{\theta}^*_{(b)} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*_b. \] (3.4.3)

The estimated covariance matrix is
\[ \hat{C}_b(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^*_b - \hat{\theta}^*_{(b)})(\hat{\theta}^*_b - \hat{\theta}^*_{(b)}). \] (3.4.4)

### 3.4.2 Fisher information matrix
In most parameter estimation problems, the information about the parameters is obtained from a sample of data drawn from the underlying probability distribution.

The problem is to know how much information a sample of data can provide about the unknown parameters. The Fisher information matrix is used to measure information for estimators of the unknown parameters:

\[ F_{i,j} = -E \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} L \right], \quad i, j = 1, 2, \ldots, T \] (3.4.5)

where \( \theta \) is the vector of unknown parameters with dimension \( T \).

The approximation in the form of the “empirical information” follows by substituting directly the data into the second derivative. Hence, the empirical Fisher information matrix is given by,

\[ I_{i,j} = - \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} L \right], \quad i, j = 1, 2, \ldots, T \] (3.4.6)

The standard errors are the square roots of the diagonal elements of \( [I_{i,j}]^{-1} \).
4. Simulation

In this chapter, the theory is applied by simulating datasets. An experiment is used to demonstrate the recommended methodology.

4.1 The impact of measurement errors

The histogram in Figure 4.1a is for 1000 simulated data for the lognormal distribution with $\sigma = 0.5$ and $\mu = 4$. Gaussian heteroscedastic measurement errors with $\alpha^2 x^2$ variance were added to obtain the convolved distribution in the Figure 4.1b where $\alpha = 0.3$.

Data contaminated by measurement errors have wider ranges of observed values with even negative values appearing. The lognormal distribution is defined only on the positive range of observed values as illustrates in Figure 4.2.

![Figure 4.1: Histogram of the distribution of uncontaminated and contaminated with heteroscedastic measurement error with $\sigma = 0.5, \alpha = 0.3$.](image)

![Figure 4.2: Histogram of the distribution of uncontaminated and contaminated with heteroscedastic measurement error $\sigma = 0.3, \alpha = 0.5$.](image)
Figure 4.2 illustrates the histogram of the distribution contaminated and uncontaminated with heteroscedastic measurement errors with larger $\alpha$.

In this case there are negative values in the datasets in Figure 4.2b contaminated with measurement errors.

Figure 4.3: Lognormal distribution with and without measurement errors.

Figure 4.3 illustrates the PDFs (3.2.4) and (3.2.5) with $\sigma = 0.5$, $\alpha = 0.3$ and $\mu = 4$

4.2 Testing of goodness of fit for the model

It is important to show that the statistical model fitted to the dataset is adequate. There are many statistical tests for goodness of fit for the model, e.g. Kolmogorov-Smirnov, Anderson-Darling and Cramér-Von-Mises (Razali and Wah, 2010).

All these statistical tests are based on a measure of the distance between the theoretical and observed CDFs. These are quantitative methods.

Another approach is the Probability-Probability P-P plot or Quantile-Quantile Q-Q plot, which is an informal goodness of fit test (Oucherif, 2014).

Also, there is an alternative test, the Chi-squared $\chi^2$ which compares the observed and the predicted values of data in $N$ selected intervals.

Normally, the cumulative distribution function of the theoretical distribution values is obtained for a given set of parameters appearing in the theoretical CDF which are $(\sigma, \alpha, \mu)$. 
Table 4.1: Table of the Kolmogorov-Smirnov calculated values and 5% critical value.

<table>
<thead>
<tr>
<th>Model</th>
<th>sample size</th>
<th>K-S value</th>
<th>Critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model (3.2.4)</td>
<td>200</td>
<td>0.0624</td>
<td>0.0960</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0447</td>
<td>0.0607</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0287</td>
<td>0.0429</td>
</tr>
<tr>
<td>Model (3.2.5)</td>
<td>200</td>
<td>0.0494</td>
<td>0.0960</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0317</td>
<td>0.0607</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0149</td>
<td>0.0429</td>
</tr>
</tbody>
</table>

Table 4.1 illustrates the Kolmogorov-Smirnov calculated values for both model (3.2.4) of uncontaminated errors and model (3.2.5) which was contaminated by measurement errors for different sample sizes. It is convincing that the data follows a specific distribution.

The null hypothesis is not rejected for the theoretical distribution for the simulation with parameters values $\sigma = 0.5$, $\alpha = 0.1$ and $\mu = 3$, at 5% level of significance.

![Figure 4.4: P-P plot for a simulated data has 200 observed value.](image)

The P-P plot for simulated data has 200 observed values distributed according to model (3.2.6). The plot at the bottom panel of the Figure 4.4 is based on the model that was not contaminated by heteroscedastic measurement error. The plot in the top panel illustrates the model where the data were contaminated by measurement errors.
4.3 Effectiveness of bootstrapping sample size

The histograms in Figures (4.5) and (4.6) illustrate distributions of the estimated parameters for the different sample sizes from 1000 bootstrap samples.

Figure 4.5: Illustration of distribution of the estimated parameters for the data sets of sample size $n = 50$, from the 1000 bootstrap samples with $\sigma = 0.5, \alpha = 0.3$ and $\mu = 4$. 
4.4 Standard errors of the estimated parameters by MME and MLE

The estimated parameters by MME are real numbers when the ratio \( \beta = \frac{M_2}{M_1 M_3} \) is greater than 1, as was mentioned in section 3.3.2. Otherwise the estimated parameters are complex numbers. The value of \( \beta \) is close to one if the value of \( \sigma \) is larger than \( \alpha \).
Figure 4.7: Distribution of $\beta$ with $\sigma, \alpha$ and $\mu$ to be 0.5, 0.3 and 4 respectively.

Figure 4.7 Illustration of distribution of $\beta$ for $\sigma = 0.5, \alpha = 0.3$ and $\mu = 4$ of sample size $n = 300$, from 400 samples.

Table 4.2 illustrates simulated standard errors of the estimated parameters using the sample sizes $n = 200, 500$ and 1000.

MME sometimes fails to give physical meaning estimates.

In such cases it is assumed that the model reduces to the error-free case $\alpha = 0$. 
Table 4.2: Standard errors of estimated parameters for parameter $\mu = 3$.

<table>
<thead>
<tr>
<th>n</th>
<th>true value</th>
<th>MME</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma$</td>
<td>$\alpha$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>200</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0150</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.226</td>
<td>0.0470</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0599</td>
<td>0.0385</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0722</td>
<td>0.1449</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0949</td>
<td>0.1464</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0914</td>
<td>0.1058</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2079</td>
<td>0.1268</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.2059</td>
<td>0.1110</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2148</td>
<td>0.1192</td>
</tr>
<tr>
<td>500</td>
<td>0.2</td>
<td>0.0327</td>
<td>0.0704</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0130</td>
<td>0.0228</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0142</td>
<td>0.0238</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0588</td>
<td>0.1399</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0735</td>
<td>0.1413</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0855</td>
<td>0.1072</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2193</td>
<td>0.1587</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.2109</td>
<td>0.1482</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2217</td>
<td>0.1151</td>
</tr>
<tr>
<td>1000</td>
<td>0.2</td>
<td>0.0101</td>
<td>0.0640</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0086</td>
<td>0.0148</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0114</td>
<td>0.0177</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0479</td>
<td>0.1307</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0658</td>
<td>0.1375</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0691</td>
<td>0.0872</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2179</td>
<td>0.1937</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.2119</td>
<td>0.1752</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2069</td>
<td>0.1438</td>
</tr>
</tbody>
</table>

Table 4.2 presents the standard errors of estimated parameters by fixing the value of parameter $\mu = 3$ and changing the sample size and other values of parameters as illustrated in the table. Large sampling variations of the third moment can cause estimated parameters to be complex.

Some estimated parameters value are close to true parameter values and others equal to zero as presented in Figure 4.8. Zero-valued parameters suggest that the model may be inappropriate.
Figure 4.8: Panels (4.8a)–(4.8f): $\sigma = 0.2$ Panels (4.8g)–(4.8l): $\sigma = 0.5$ $n = 200$, $\alpha = 0.5$ and $\mu = 3$. This shows that as the values $\sigma$ increases as the range of distribution of the estimated parameters increases.
Table 4.3: Standard errors of estimated parameters for parameter $\mu = 4$.

<table>
<thead>
<tr>
<th>n</th>
<th>true value</th>
<th>MME</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma$</td>
<td>$\alpha$</td>
<td>S.e($\hat{\sigma}$)</td>
</tr>
<tr>
<td>200</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.0223</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0251</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1</td>
<td>0.0720</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.0914</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.1063</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1</td>
<td>0.1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.2061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.2071</td>
</tr>
<tr>
<td>500</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0321</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.0131</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0164</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1</td>
<td>0.0576</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.0782</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0795</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1</td>
<td>0.2173</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.2258</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.2092</td>
</tr>
<tr>
<td>1000</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0102</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1</td>
<td>0.0491</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.0636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0613</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1</td>
<td>0.2333</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.2443</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.2345</td>
</tr>
</tbody>
</table>
Figure 4.9: The distributions of the estimated parameters for fixed sample size $n = 500$, $\sigma = 0.2$ and $\alpha = 0.1$ with changing values of parameters $\mu = 3$ up to $\mu = 4$.

Based on the standard error, the method of maximum likelihood estimation is more accurate than method of moments.
### 4.4.1 Test of normality of distribution of estimated parameters

There are many tests of normality in statistical research. The most commonly used normality test procedures in computing software are the Kolmogorov-Smirnov, Anderson-Darling, Shapiro-Wilk and Lilliefors tests. According to the literature (Razali and Wah, 2010), the Shapiro-Wilk test is one of the best tests amongst those mentioned.

Table 4.4: Shapiro-Wilk normality tests for distribution of the estimators $\hat{\alpha}$.

<table>
<thead>
<tr>
<th>Parameter $\mu$</th>
<th>$\alpha$</th>
<th>Test statistics</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.1</td>
<td>0.824</td>
<td>1.0 e-16</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.9982</td>
<td>0.8957</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9973</td>
<td>0.5861</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.8244</td>
<td>1.0 e-16</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.9952</td>
<td>0.1269</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.998</td>
<td>0.8407</td>
</tr>
<tr>
<td>4</td>
<td>0.1</td>
<td>0.8062</td>
<td>1.0 e-16</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.9893</td>
<td>0.001069</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9981</td>
<td>0.8721</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.8035</td>
<td>1.0 e-16</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.9706</td>
<td>1.829e-08</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9972</td>
<td>0.5471</td>
</tr>
</tbody>
</table>

Test statistics and $p$-values of the estimators $\hat{\alpha}$ for maximum likelihood and method of moments estimation. Parameters $\sigma = 0.2$, $\mu = 3$ and $\mu = 4$ and $\alpha = 0.1, 0.3, 0.5$ with sample size $n = 1000$ are used. Distribution of $\hat{\alpha}$ approach normality as $\alpha$ increases.
Figure 4.10: Panels (4.10a)–(4.10f): $\mu = 3$ Panels (4.10g)–(4.10l) $\mu = 4$ $n = 1000, \alpha = 0.1, 0.3, 0.5, \sigma = 0.2$.

Figure 4.10 shows that, for the first two rows of figures in the top panel, the distribution of $\hat{\alpha}$ at values $\alpha = 0.1, 0.3, 0.5$ with parameter $\sigma = 0.2$ and $\mu = 3$ for both MLE and MME with sample size $n = 1000$.

The bottom panel presents the distribution of the $\hat{\alpha}$ estimator of $\alpha = 0.1, 0.3, 0.5$ with parameter $\sigma = 0.2$ and $\mu = 4$ for both MLE and MME and with sample size $n = 1000$. According to the test statistics illustrated in Table 4.4, as the values of $\alpha$ increases for both methods the maximum likelihood and the method of moments
estimation as the values of test statistics increases and close to 1. This illustrates that the distribution of the estimators \( \hat{\alpha} \) is approaching the normal distribution as \( \alpha \) increases.

### 4.4.2 Bias of the estimated parameters by a modification of MME.

The method of moments sometimes gives estimated parameters that have no physical meaning. This problem is caused by the large variation of the third sample moment, which again gives rise to \( \beta < 1 \) in Equation (3.3.20), so that the complex solutions for \( \hat{\alpha} \) are obtained from Equation (3.3.17). To deviate this, a percentage of the largest values in the sample of datasets is removed.

Consider the simulation using values for the parameters of \( \sigma, \alpha \) and \( \mu \) to be 0.5, 0.3 and 4 respectively with the sample size of \( n = 300 \) replicated 400 times. By removing 5\% of the largest values in the sample of datapoints, the estimated parameters are however biased. The estimated parameter \( \hat{\sigma} \) is particularly highly biased compared to other estimated parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True values:</th>
<th>Removed 5%:</th>
<th>Bias:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.3904</td>
<td>0.4399</td>
<td>12.68%</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.4120</td>
<td>0.2076</td>
<td>50.39%</td>
</tr>
<tr>
<td>( \mu )</td>
<td>4.0427</td>
<td>4.0099</td>
<td>0.81%</td>
</tr>
</tbody>
</table>

The figure below illustrates the distribution of \( \beta \) for the case where some of largest values in the sample of datapoints were deleted.

![Distribution of \( \beta \) with and without values removed in the sample of datapoints.](image)

For datasets that have the outliers present, the negative skewness is caused by the large value of the third moment as, is illustrated on Figure 4.11b.

### 4.5 Cramér-Rao lower bound (CRLB)

The Cramér-Rao Lower Bound (CRLB) is a minimum bound for the variance of any unbiased estimator of the parameter to be estimated for a statistical model. (Note though that some of parameter estimates below are biased.)

Assume that a random sample \( X_1, X_2, \cdots, X_n \) is drawn from the probability density function PDF \( f(x; \phi) \), such that the vector of parameters \( \phi \) is unknown.
To determine the CRLB of the variance of the estimated parameters $\hat{\phi}$, the Fisher information matrix can be used as has been discussed in the Section 3.4.2.

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables with density function $f(x, \phi)$, and assume that $\hat{\phi}$ is an unbiased estimator of parameter $\phi$. Then,

$$\text{Var}(\hat{\phi}) \geq \frac{1}{nE\left(\left[\frac{\partial}{\partial \phi} \log(f(x, \phi))\right]^2\right)},$$

(4.5.1)

or

$$\text{Var}(\hat{\phi}) \geq \frac{1}{-nE\left(\frac{\partial^2}{\partial \phi^2} \log(f(x, \phi))\right)}.$$

(4.5.2)

The expected value of the second derivative of the log-likelihood function is given by

$$E\left(\left[\frac{\partial^2}{\partial \phi^2} \log(f(x, \phi))\right]\right) = -\int_0^\infty \left[\frac{\partial}{\partial \phi} \log(f(x, \phi)) \frac{\partial}{\partial \phi} \left(f(x, \phi)\right)\right] dx,$$

(4.5.3)

where

$$f(x, \phi) = \frac{1}{2\pi\sigma \alpha} \int_0^\infty \exp\left\{-\frac{(\log(x) - \mu)^2}{2\sigma^2} - 2\log(x) - \frac{(y - x)^2}{2\alpha^2 x^2}\right\} dx,$$

(4.5.4)

and $\phi$ is a vector of parameters $\phi = (\sigma, \alpha, \mu)$.

Further details for the derivation of the minimum variance are illustrated in the appendices. The numerical computation procedure was implemented in MATLAB, and the results are illustrated in Table 4.6.

Table 4.6: Cramér-Rao lower bounds for $\mu = 3$ and the sample size $n = 1000$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$\alpha$</th>
<th>$S.e.(\hat{\sigma})$</th>
<th>$S.e.(\hat{\alpha})$</th>
<th>$S.e.(\hat{\mu})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>0.0342</td>
<td>0.0645</td>
<td>0.0097</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>0.0191</td>
<td>0.0128</td>
<td>0.0119</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.0251</td>
<td>0.0166</td>
<td>0.0177</td>
</tr>
<tr>
<td>0.2</td>
<td>0.1</td>
<td>0.1604</td>
<td>0.7596</td>
<td>0.0800</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>0.0220</td>
<td>0.0208</td>
<td>0.0200</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.0247</td>
<td>0.0186</td>
<td>0.0239</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.0194</td>
<td>0.3140</td>
<td>0.0182</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1</td>
<td>0.0301</td>
<td>0.0324</td>
<td>0.0357</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.0323</td>
<td>0.0205</td>
<td>0.0373</td>
</tr>
</tbody>
</table>

Table 4.6 shows some differences between standard errors obtained by simulation illustrated in Tables 4.2 and 4.3. There are at least two reasons to expect differences between the Cramér-Rao lower bound and the true standard errors as determined by simulations. For some combinations of parameter values, the distribution of estimated $\alpha$ values is bimodal. This means that the usual asymptotic theory does not apply, hence the Cramér-Rao lower bound is not appropriate. The second issue is that for small $\alpha$ the calculations of the Cramér-Rao lower bound are inaccurate since the Fisher information matrix has a high condition number.
4.6 Computation time

The performance of the MME and MLE is compared in terms of computation time. Simulation has been used to investigate the time spent to compute the standard error for each method as illustrated in Section 4.4. The time spent in computation may be measured in seconds and the time spent by every algorithm for every graph is summarised in Table 4.7.

<table>
<thead>
<tr>
<th>n</th>
<th>parameter value</th>
<th>MLE</th>
<th>MME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>µ</td>
<td>σ</td>
<td>α</td>
</tr>
<tr>
<td>1000</td>
<td>4</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>500</td>
<td>4</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>200</td>
<td>4</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
</tr>
</tbody>
</table>

Comparing the computation time for these two methods, there is a large difference caused by the numeric computational integral of the expression \( \exp(-\frac{(\log(x) - \mu)^2}{2\sigma^2}) - 2\log(x) - \frac{(y - x)^2}{2\alpha x^2} \) for each dataset. The maximum likelihood estimation method is very expensive as the results in Table 4.7 show.

4.7 Saddlepoint approximation method

In this dissertation the saddlepoint approximation is used to derive an approximation of the convolution of the lognormal distribution and the distribution of heteroscedastic measurement errors. Due to the fact that for MLE a convolution integral was needed to be calculated for each datapoint, and that this had to be repeated for each iteration towards the solution, the computing cost is very high.

The saddlepoint approximation of the convolution integral is derived. Simulated data will be used to apply the theory. The convolution PDF is

\[
f_y(y) = \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (\log(x) - \mu)^2 \right\} \frac{1}{\sqrt{2\pi}\alpha x} \exp \left\{ -\frac{1}{2\alpha^2 x^2} (y - x)^2 \right\} dx.
\]

Now, let \( h_y(y, x) \) be a positive function and assume that \( f_y(y) = \int_0^\infty g_y(y, x) dx \) such that \( h_y(y, x) = \log(g_y(y, x)) \) implies that \( g_y(y, x) = \exp(h_y(y, x)) \). The density function of \( Y \) can be expressed as,

\[
\int_0^\infty g_y(y, x) dx = \int_0^\infty \exp(h_y(y, x)) dx.
\]

Thus,

\[
h_y(y, x) = \frac{-1}{2} \left[ \frac{(\log(x) - \mu)^2}{\sigma^2} + 4\log(x) + \frac{(y - x)^2}{\alpha^2 x^2} \right].
\]
For fixes $y$ the Taylor series expansion of $h_y(y, x)$ is given by

$$h_y(y, x) = h_y(y, x_0) + (x - x_0)h'_y(y, x_0) + \frac{(x - x_0)^2}{2}(h''_y)(y, x_0). \quad (4.7.4)$$

Where primes denoted differentiation with respect to $x$. The first and second derivatives of $h_y(y, x)$ are respectively given by

$$h'_y(y, x) = \frac{\mu - \log(x)}{\sigma^2 x} - \frac{2}{x} + \frac{y(y - x)}{\alpha^2 x^3}$$

$$h''_y(y, x) = \frac{\log(x) - \mu - 1}{\sigma^2 x^2} + \frac{2}{x^2} + \frac{(2yx - 3y^2)}{\alpha^2 x^4}. \quad (4.7.5)$$

In this case, provided that the first derivative of $h_y(y, x)$ is vanished at $x = x_0$,

$$g_y(y, x) \approx \exp \left( h_y(y, x_0) + \frac{(x - x_0)^2}{2} h''_y(y, x_0) \right). \quad (4.7.6)$$

If $h_y(y, x)$ is a maximum in $x_0$ then, $h''_y(y, x_0)$ is negative, hence,

$$f_y(y) = \int_0^\infty g_y(y, x)dx \approx \exp(h_y(y, x_0)) \left[ \frac{-2\pi}{h''_y(y, x_0)} \right]^{1/2}$$

$$g_y(y, x) \left[ \frac{-2\pi}{h''_y(y, x_0)} \right]^{1/2}. \quad (4.7.7)$$

Therefore, the saddlepoint approximation of the PDF of the random variable $Y$ is given by

$$f_y(y) \approx g_y(y, x_0) \left[ \frac{-2\pi}{h''_y(y, x_0)} \right]^{1/2}. \quad (4.7.8)$$

By substituting the value of $g_y(y, x)$ and $h''_y(y, x_0)$ in (4.7.7),

$$f_y(y) \approx \frac{1}{2\pi\sigma\alpha} \exp \left[ -\frac{1}{2} \left( \left( \frac{\log(x) - \mu}{\sigma} \right)^2 + 4 \log(x) + \left( \frac{y - x}{\alpha x} \right)^2 \right) \right]$$

$$\left. \left[ \frac{-2\pi}{\left( \frac{\log(x) - \mu - 1}{\sigma^2 x^2} + \frac{2}{x^2} + \frac{(2yx - 3y^2)}{\alpha^2 x^3} \right)} \right]^{1/2} \right|_{x = x_0}. \quad (4.7.9)$$

All the numerical computations were done using MATLAB.

To illustrate the saddlepoint approximation, consider parameter values for $\sigma$, $\alpha$ and $\mu$ of 0.4, 0.3 and 3 respectively.

Figure 4.12 illustrates the saddlepoint approximation of the PDF of random variable $Y$. 
Figure 4.12 illustrates a convolution density of the lognormal and Gaussian heteroscedastic measurement errors (solid line) together with the saddlepoint approximation (dashes).

The saddlepoint approximation is inadequate near the peak of the PDF and can hence not be applied in the present context without carefully further study.

Note: Saddlepoint approximation is applicable if the expression \( h'_y(y, x) = \frac{\mu - \log(x)}{\sigma^2} \cdot \frac{2}{\sigma^2} + \frac{y(y-x)}{\alpha^2 x^3} \) has a real positive root for given values of \( \sigma, \alpha, \mu \) and \( y \).
5. Application of theory to real data

5.1 Introduction

In this chapter, the methodology is applied to the distribution of masses of star clusters in the galaxy M83 and to one environment datasets from the Kola project available in the R library StatDA. In most of cases the data sets from regional geochemical and environmental surveys are assumed to be either normally or lognormally distributed. “As a general rule, almost all variables (up to more than 50 analysed chemical elements per data set) show neither a normal nor a lognormal data distribution.” However it is not guaranteed that all data follow this assumption according to some authors e.g (Reimann and Filzmoser, 2000). It nevertheless seems worth considering a modification of the usually lognormal description of geochemical data, namely the incorporation of the measurement errors.

There exists informal and formal goodness of fit tests for verifying that the statistical model fitted to the data is appropriate. Examples of formal goodness of fit tests are Kolmogorov-Smirnov, Anderson-Darling and Cramér-Von-Misses, used a lot in practice. Q-Q or P-P plots are known as informal goodness of fit tests and are useful (Oucherif, 2014). In the cases of some form of the complex models for the example (3.2.4) and (3.2.6) the likelihood ratio test is recommended and therefore will be performed.

The null hypothesis involves testing whether a fewer model with parameters rather that the model of full form presents an appropriate model fitted to the data. The hypothesis can be tested by the likelihood ratio test statistic (Koen and Kondlo, 2009)

\[ \Lambda = 2[\max L(H_1) - \max L(H_0)] \sim \chi^2_p. \]  

(5.1.1)

\( H_1 \) and \( H_0 \) are the null and alternative hypotheses respectively, and the maxima of the log likelihoods are evaluated for each of these. The statistic has an asymptotic \( \chi^2 \) distribution with \( p \) degrees of freedom, where \( p \) is equal to number of additional parameters in the more complex model.

Since there are outliers in the datasets from the Kola project, available in the R library StatDA, in this research datasets will be used after removing the outliers.

5.2 Masses of star clusters in galaxy M83

There are \( N = 939 \) masses of star clusters in the M83 catalogue (Bastian et al., 2011). The parameters in the convolved full model (3.2.6), and in the simplified model of no measurement errors in (3.2.4) are estimated.

Estimated parameters using maximum likelihood estimation are given by \( \hat{\sigma} = 0.5747, \hat{\mu} = 3.8275 \) and \( \hat{\alpha} = 0.0748 \). The likelihood ratio (5.1.1) is \( \Lambda = 2[-4.4110E3 - (-4.4115E3)] = 1 \). Based on the value of \( \Lambda = 1 \) there is no significant difference between models (3.2.6) and (3.2.4).
Section 5.3. Zn in Chorizon data

The data analysed here were taken from Chorizon, Kola data set available in library StataDA (Reimann and Filzmoser, 2000). Zinc $Zn$ data have sample size $n = 605$, the estimated parameters appearing in the model $\hat{\sigma}, \hat{\alpha}$ and $\hat{\mu}$ are equal to $0.6417, 0.0171$ and $3.0765$ respectively with the maximum likelihood of $-2.4474E3$.

The result of the maximum log-likelihood ratio statistic (5.1.1) is $\Lambda = 2[-2.4474E3 - (-2.4474E3)] = 0$. Comparing the fits of the two distribution models, based on a value of $\Lambda = 0$, it is shown that there is no significant difference for models (3.2.6) and (3.2.4).

Quantile-Quantile plots have been used to check the goodness of fitted data for the model.
5.2a presents the Q-Q plot of the model contaminated by measurement errors.

## 5.4 Ba in Chorizon data

The $Ba$ data have sample size $n = 605$ and the estimated parameters appearing in the model $\hat{\sigma}, \hat{\alpha}$ and $\hat{\mu}$ are equal to $0.7079, 0.0223$ and $3.7576$ respectively with the maximum likelihood of $-2.9894E3$. The result of the maximum log-likelihood ratio (5.1.1) is $\Lambda = 2[-2.4474E3 - (-2.9038E3)] = 0$, there is no significant difference for models (3.2.6) and (3.2.4). The Quantile-Quantile plot has been used to check the goodness of fit for the data of model.

![Quantile-Quantile plot](image)

Figure 5.3: Quantile-Quantile plot of the CDF of concentration of Ba in Chorizon.

Figure 5.3 shows in Figure 5.3b the Q-Q plot of the model uncontaminated by measurement errors and Figure 5.3a presents the Q-Q plot of the model contaminated by measurement errors. The contaminated model is shown to be the better model to fit the chemical element Barium Ba data, even if the data does not have a significantly fit the two models although neither model fits well.

## 5.5 Mg XRF in Chorizon data

The transformed data set $Mg_{XRF}$ of chemical element Magnesium was taken in Chorizon with sample size $n = 617$. $Mg_{XRF}$ data sets obtained after the analytical technique of analysis of X-ray fluorescence has the abbreviation XRF (Reimann and Filzmoser, 2000). The estimated parameters appearing in the model $\hat{\sigma}, \hat{\alpha}$ and $\hat{\mu}$ are equal to $0.3237, 0.2988$ and $9.3732$ respectively with the maximum log-likelihood of $-5.9863E3$ for the full model, while the simplified model has maximum log-likelihood of $-5.9991E3$.

The likelihood ratio statistic is $\Lambda = 0.0256$. Dependent on those values, in this case the model contaminated by measurement errors (3.2.6) is the best fitted to the $Mg_{XRF}$ data, compared to uncontaminated model (3.2.4). A Quantile-Quantile plot has been used to check the goodness of fit for the data of model.
Figure 5.4 shows in Figure 5.4b the Q-Q plot of the model without measurement errors and Figure 5.4a presents the Q-Q plot of the model including measurement errors. The measurement errors model is shown to be the better model to fit the chemical element Mg\textsubscript{XRF} data compared to the uncontaminated measurement errors model.

**NOTE**

In the four data sets analysed in this chapter 5, all data sets do not fit the models properly or highly significantly. The full model (3.2.6) is usually the better fitted model when compared to the simplified model (3.2.4).
6. Conclusion and Findings

6.1 Conclusion

The statistical technique for estimating an unknown distribution or density in the presence of measurement errors is deconvolution. The usual method of nonparametric deconvolution estimation is the kernel deconvolution density estimation according to previous literature (Stefanski and Carroll, 1990). In this project, it was assumed that the measurement errors are heteroscedastic Gaussian distributed where the standard deviation of the measurement error is proportional to the corresponding values of the true value, since the assumption of homoscedasticity for measurement error is not always realistic in real life.

The assumption of the lognormal distribution is commonly used in biology, finance, population growth models and astronomy according to the literature (Stefanski, 2000). The methodology for deconvolution when the underlying variable distribution is known to be the lognormal distribution was discussed in this project in chapter 3.

In this research, the method of maximum likelihood and method of moment estimation have been used to estimate the parameters and their standard errors. The results show that the maximum likelihood estimation method is expensive compared to method of moments in terms of time consumed to determine the estimators. The method of moments is cheap, but sometimes gives estimated parameters that are non-physical or complex numbers. This problem lies in the large sampling variations of the third sample moment $M'_3$.

In addition the results of this research show that the saddlepoint approximation can be a good approximation for the density function.

Finally, in the application of the theory to the data, all four datasets analysed in this research illustrated that the models do not fit the data sets properly. The full model, which contained measurement errors (3.2.6), is the better model when fitted to the data and compared to the simplified model (3.2.4).

6.2 Findings

The method of maximum likelihood estimation gives better estimates than the method of moments, but the maximum likelihood estimate is computationally expensive. Unfortunately, the method of moment sometimes gives estimated parameters that are complex. This problem is caused by the large sampling variation of the third moment $M'_3$. This motivates us to look at other methods in the literature such as fractional moments (e.g. Zainodin and Amjad (1999)).
7. Appendices

7.1 Derivation of the roots

Proceeding from equation 3.3.18
\[ x^3 + 3x^2 + 3(1 - \beta)x + (1 - \beta) = 0, \]  
(7.1.1)
let \( x = y + h \), then Taylor expansion of \( f(y + h) \) is given by
\[ f(y + h) = f(h) + f'(h)y + \frac{f''(h)y^2}{2!} + \frac{f'''(h)y^3}{3!} \]
(7.1.2)
where
\[
\begin{align*}
  &f(h) = h^3 + 3h^2 + 3(1 - \beta)h + (1 - \beta) \\
  &f'(h) = 3h^2 + 6h + 3(1 - \beta) \\
  &f''(h) = 3h + 3 \\
  &\frac{f'''(h)}{6} = 1.
\end{align*}
\]
(7.1.3)
To eliminate the quadratic term in equation (7.1.1), let \( 3h + 3 = 0, h = -1 \), substitute \( h = -1 \) in Equation (7.1.1) to transform the equation to the form
\[ y^3 - 3\beta y + 2\beta = 0 \]
(7.1.4)
let \( y = z + v \), then substitute in (7.1.4) and arrange the two unknown variables. Hence,
\[ z^3 + v^3 + (-3\beta + 3zv)(z + v) + 2\beta = 0 \]
(7.1.5)
Equation (7.1.5) has only one solution if \(-3\beta + 3zv = 0\)
\[
\begin{align*}
  z^3 + v^3 &= -2\beta \\
  zv &= \beta
\end{align*}
\]
(7.1.6)
Thus,
\[ z^6 + 2\beta z^3 + \beta^3 = 0. \]
(7.1.7)
Let \( t = z^3 \), therefore,
\[ t^2 + 2\beta t + \beta^3 = 0 \]
(7.1.8)
\[ \Delta = 4\beta^2 - 4\beta^3 = 4\beta^2(1 - \beta). \]
(7.1.9)
From equation 7.1.6 \( z^3 + v^3 = -2\beta \) and \( z^3 v^3 = \beta^3 \) are sum and product of the two unknown quantities \( z^3 \) and \( v^3 \). These quantities are the roots of the equation quadratic 7.1.8, then equation 7.1.8 has the following roots,
\[
\begin{align*}
  t_1 &= -\beta + \beta \sqrt{1 - \beta} \\
  t_2 &= -\beta - \beta \sqrt{1 - \beta}.
\end{align*}
\]
(7.1.10)
Due to the symmetry between the terms \( z^3 \) and \( v^3 \) in the equation 7.1.6 the following are setted
\[
\begin{align*}
  z^3 &= t_1 = -\beta + \beta \sqrt{1 - \beta} \\
  v^3 &= t_2 = -\beta - \beta \sqrt{1 - \beta}
\end{align*}
\]
(7.1.11)
Then, (7.1.7) has the following solutions:

\[
\begin{align*}
z &= t_1^{\frac{1}{3}} \\
v &= t_2^{\frac{1}{3}}.
\end{align*}
\]

(7.1.12)

Three possible values of \(v\) and \(z\) are given by

\[
\begin{align*}
z &= t_1^{\frac{1}{3}}, z = \eta t_1^{\frac{1}{3}}, z = \eta^2 t_1^{\frac{1}{3}} \\
v &= t_2^{\frac{1}{3}}, v = \eta t_2^{\frac{1}{3}}, v = \eta^2 t_2^{\frac{1}{3}}
\end{align*}
\]

(7.1.13)

where \(\eta\) has unit magnitude. Therefore, the solution of equation (7.1.4) will have the following roots:

\[
\begin{align*}
y_1 &= \eta^{1/3} + t_1^{1/3} \\
y_2 &= \eta^{1/3} + \eta^2 t_2^{1/3} \\
y_3 &= \eta^{1/3} + \eta^2 t_2^{1/3}
\end{align*}
\]

(7.1.14)

### 7.1.1 Discussion of solution.

Obviously \(\Delta\) will be positive, zero or negative. Actually, there three possible values of \(x\) as the roots of (7.1.1). However, the values of \(x\) may be all real, repeated or one real and two conjugate complex roots (Uspensky, 1948).

The discriminant \(\Delta\) in (7.1.9) is given by

\[
\Delta = 4\beta^2(1 - \beta).
\]

(7.1.15)

The solution of the equation (7.1.8) is dependent on the sign of discriminant \(\Delta\). Since the interesting roots are real roots in this case, the roots are obtained for the case \(\Delta < 0\). This means that the real roots are obtained if and only if \(\beta > 1\).

Hence, the three roots of (7.1.8) are given by the following trigonometric functions:

\[
\begin{align*}
t_1 &= -\beta + i\beta\sqrt{-1 + \beta} \\
t_2 &= -\beta - i\beta\sqrt{-1 + \beta}.
\end{align*}
\]

(7.1.16)

The modulus \(\rho\) and argument \(\varphi\) of \(t_1\) and \(t_2\) are given by

\[
\begin{align*}
\rho^2 &= \beta^2 + \beta^2(-1 + \beta) = \beta^3 \\
\rho &= \beta^{3/2}
\end{align*}
\]

(7.1.17)

\[
\begin{align*}
\cos \varphi &= \frac{-\beta}{\beta^{3/2}} \implies \varphi = \arccos(-\beta^{-1/2}) \\
\tan \varphi &= \sqrt{-1 + \beta} \implies \varphi = \arctan(\sqrt{-1 + \beta}).
\end{align*}
\]

(7.1.18)

Then,

\[
\begin{align*}
z &= t_1^{1/3} = \sqrt[3]{\rho} \cos\left(\frac{\varphi}{3} + i \sin\frac{\varphi}{3}\right) \\
&= \sqrt[3]{\beta} \left(\cos\frac{\varphi}{3} + i \sin\frac{\varphi}{3}\right) \\
v &= t_2^{1/3} = \sqrt[3]{\rho} \cos\left(\frac{\varphi}{3} - i \sin\frac{\varphi}{3}\right) \\
&= \sqrt[3]{\beta} \left(\cos\frac{\varphi}{3} - i \sin\frac{\varphi}{3}\right)
\end{align*}
\]

(7.1.19)
Therefore, all the estimated parameters are real numbers if and only if the
\[ x_k = \frac{\eta \sqrt{\beta} \left( \cos \left( \frac{\varphi}{3} \right) + i \sin \left( \frac{\varphi}{3} \right) \right) + \sqrt{\beta} \left( \cos \left( \frac{\varphi}{3} \right) - i \sin \left( \frac{\varphi}{3} \right) \right)}{1} \]

where
\[ \eta = \cos \frac{3\pi}{2} + i \sin \frac{3\pi}{2}. \]  

Therefore, the roots of the equation (7.1.1) are given by
\[ x_k = y_k - 1 \]
\[ = 2\sqrt{\beta} \cos \left( \frac{\arccos(-\sqrt{1/\beta}) + 2\pi k - (k - 1)}{3} \right) - 1 \]

where \( k = 1, 2, 3 \).

Hence, \( x = \alpha^2 \) where only the positive real root is considered in this situation.

Therefore,
\[ \alpha^2 = 2\sqrt{\beta} \cos \left( \frac{\arccos(-\sqrt{1/\beta})}{3} \right) - 1 \]

All the estimated parameters are real numbers if and only if the \( \left( \frac{M^2}{\beta^2} = \beta > 1 \right) \) condition holds.

The numerical calculation were implemented in different numerical computations, according to the software used such as MATLAB and MATHEMATICA.

### 7.2 Cramér-Rao lower bound

The partial differentiation of (4.5.4) with respect to \( \sigma \) is given by

\[
-I_2(\phi) = \int_0^\infty \frac{\partial}{\partial \sigma} \log(f(x, \phi)) \frac{\partial}{\partial \sigma} f(x, \phi) \, dy
\]

\[
= \int_0^\infty \left[ \frac{1}{2\pi \sigma^2 \alpha} \int_0^{\infty} \exp \left\{ -\frac{(\log(x) - \mu)^2}{2\sigma^2} - 2\log(x) + \frac{-(y-x)^2}{2\sigma^2} \right\} \, dx \right] dy
\]

\[
+ 2 \int_0^\infty \frac{1}{2\pi \sigma^2 \alpha} \int_0^\infty \frac{(\log(x) - \mu)^2}{2\sigma^2} \exp \left\{ -\frac{(\log(x) - \mu)^2}{2\sigma^2} - 2\log(x) + \frac{-(y-x)^2}{2\sigma^2} \right\} \, dx \int_0^\infty \exp \left\{ -\frac{(\log(x) - \mu)^2}{2\sigma^2} - 2\log(x) + \frac{-(y-x)^2}{2\sigma^2} \right\} \, dy d\sigma
\]

\[
= \int_0^{\infty} \left[ \frac{1}{2\pi \sigma^2 \alpha} \int_0^\infty \exp \left\{ -\frac{(\log(x) - \mu)^2}{2\sigma^2} - 2\log(x) + \frac{-(y-x)^2}{2\sigma^2} \right\} \, dx \right] dy
\]

\[
+ \int_0^{\infty} \frac{1}{2\pi \sigma^2 \alpha} \int_0^\infty \exp \left\{ -\frac{(\log(x) - \mu)^2}{2\sigma^2} - 2\log(x) + \frac{-(y-x)^2}{2\sigma^2} \right\} \, dx \, dy
\]  

(7.2.2)
The partial differentiation of \((4.5.4)\) with respect to \(\alpha\) is given by

\[
-I_2(\phi) = \int_0^\infty \left[ \frac{\partial}{\partial \alpha} \log(f(x, \phi)) \cdot \frac{\partial}{\partial \alpha} (f(x, \phi)) \right] \, dy
\]

\[
= \int_0^\infty \left( -\frac{1}{2\pi\sigma^2} \int_0^\infty \exp \left\{ \frac{-(\log(x)-\mu)^2}{2\sigma^2} - 2\log(x) \right\} \, dx \right) \, dy
\]

\[
+ 2 \int_0^\infty \frac{1}{2\pi\sigma^2} \int_0^\infty \exp \left\{ \frac{-(\log(x)-\mu)^2}{2\sigma^2} - 2\log(x) \right\} \, dx \, dy
\]

\[
+ \frac{1}{2\pi\sigma^2} \int_0^\infty \exp \left\{ \frac{-(\log(x)-\mu)^2}{2\sigma^2} - 2\log(x) \right\} \, dx \, dy
\]

\[
(7.2.3)
\]

The partial differentiation of \((4.5.4)\) with respect to \(\mu\) is given by

\[
-I_2(\phi) = \int_0^\infty \left[ \frac{\partial}{\partial \mu} \log(f(x, \phi)) \cdot \frac{\partial}{\partial \mu} (f(x, \phi)) \right] \, dy
\]

\[
= \int_0^\infty \left( -\frac{1}{2\pi\sigma^2} \int_0^\infty \frac{1}{\sigma} \exp \left\{ \frac{-(\log(x)-\mu)^2}{2\sigma^2} - 2\log(x) + \frac{-(y-x)^2}{2\alpha^2 \sigma^2} \right\} \, dx \right) \, dy
\]

\[
(7.2.5)
\]
References


