Florida International University
FIU Digital Commons

FIU Electronic Theses and Dissertations | University Graduate School

3-30-2012

Estimating the Parameters of the Three-Parameter Lognormal Distribution

Rodrigo J. Aristizabal
Florida International University, rodrigojr55@hotmail.com

Follow this and additional works at: http://digitalcommons.fiu.edu/etd

Recommended Citation
Aristizabal, Rodrigo J., "Estimating the Parameters of the Three-Parameter Lognormal Distribution" (2012). *FIU Electronic Theses and Dissertations.* Paper 575.
http://digitalcommons.fiu.edu/etd/575

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

ESTIMATING THE PARAMETERS OF THE THREE-PARAMETER LOGNORMAL DISTRIBUTION

A thesis proposal submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

by

Rodrigo J. Aristizabal

2012

To: Dean Kenneth Furton
College of Art and Sciences

This thesis, written by Rodrigo J. Aristizabal, and entitled Estimating the Parameters of the Three-Parameter Lognormal Distribution, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Florence George

Gauri L. Ghai

Zhenmin Chen, Major Professor

Date of defense: March 30, 2012

The Thesis of Rodrigo J. Aristizabal is approved.

Dean Kenneth Furton
College of Arts and Sciences

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2012

DEDICATION

I dedicate this thesis to my wife, Lilyana, and my two children, Rodrigo and Katherin; they have inspired me to serve others.

ACKNOWLEDGMENTS

I thank God, who gave me the strength and settled everything to accomplish this thesis. To him, be the glory forever. I would like to thank the members of my committee, Dr. Zhenmin Chen, Dr. Florence George, and Dr. Gauri Ghai for their support and patience. I am especially thankful to my Major Professor, Dr. Zhenmin Chen, for his guidance and encouragement.

I would like to express my gratitude to all other professors in the Department of Mathematics and Statistics who have taught me about this wonderful field of science and inspired me to teach it.

I wish to thank to all my friends at FIU for their encouragement and unconditional support, especially to my friend Zeyi Wang.

Finally, I am very grateful to my wonderful wife who has been my companion during countless long nights in this journey.

ABSTRACT OF THE THESIS

ESTIMATING THE PARAMETERS OF THE

THREE-PARAMETER LOGNORMAL DISTRIBUTION

by

Rodrigo J. Aristizabal

Florida International University, 2012

Miami, Florida

Professor Zhenmin Chen, Major Professor

The three-parameter lognormal distribution is widely used in many areas of science. Some modifications have been proposed to improve the maximum likelihood estimator. In some cases, however, the modified maximum likelihood estimates do not exist or the procedure encounters multiple estimates.

The purpose of this research is focused on estimating the threshold or location parameter γ, because when $\hat{\gamma}$ is known, then the other two estimated parameters are obtained from the first two MLE equations. In this research, a method for constructing confidence intervals, confidence limits, and point estimator for the threshold parameter γ is proposed. Monte-Carlo simulation, bisection method, and SAS/IML were used to accomplish this objective. The bias of the point estimator and mean square error (MSE) criteria were used throughout extensive simulation to evaluate the performance of the proposed method. The result shows that the proposed method can provide quite accurate estimates.

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION AND BACKGROUD.....1
1.1 Introduction ……..... 1
1.2 Background ...…….2
1.3 Research Hypothesis……...3
1.4 Method and Statistical Analysis ……6
1.5 Two-Parameter Lognormal Distribution Review ……..7
1.6 Three-Parameter Lognormal Distribution Review ...……...12

2. INTERVAL AND POINT ESTIMATION FOR γ …….23
2.1 Pivotal Quantity …….….23
2.2 Critical Values of ξ for Any Sample Size n ……..25
2.3 Confidence Interval for the Parameter γ …….28
2.4 $\xi(\boldsymbol{X}, \gamma)$ As an Increasing Function of γ …….31
2.5 Upper Confidence Limit for the Parameter γ …….……....34
2.6 Point Estimator of the Parameter γ …….....35

3. SIMULATION RESULTS AND DISCUSSION ……...…….….38

4. CONCLUSION …….……44

LIST OF REFERENCES …….…….46

LIST OF TABLES

TABLE PAGE

Table 2.2 Critical Values for the Pivotal Quantity ξ ……..26

Table 2.3.1 Sample from a Lognormal Distribution $(\mu = 4,\ \sigma = 2, \text{and}\ \gamma = 100)$.……...30

Table 2.3.2 Values of ξ for a fix sample when γ increase from 0 to $X_{(1)}$.…….31

Table 2.5.1 Percentage of Samples from a Lognormal Distribution for which the Confidence Limits cannot be calculated ……..…..35

Table 2.6.1 Random Sample from a Lognormal Distribution ($\mu = \ln(50),\quad \sigma = 0.40,$ and $\gamma = 100$) ……..…….36

Table 2.6.2 Summary of Estimates of 14 different methods …...…….…....37

Table 3.1 Effect of n on the Performance of MLE and Proposed Methods ……..…40

Table 3.2 Effect of the Shape Parameter σ on the Performance of the MLE and Proposed Methods …….…41

Table 3.3 Effect of the Scale Parameter μ on the Performance of the Two Methods.…..41

Table3.4 Effect of the Threshold Parameter γ on the Performance of the Two Methods. 42

Table 3.5 Performance of the Modified Method ...……...42

1. INTRODUCTION AND BACKGROUD

1.1 Introduction

The three-parameter lognormal distribution is a skewed distribution that is useful for modeling continuous positive random variables with support set $[\gamma, \infty)$ for some $\gamma \geq 0$. Limpert, Stahel and Abbt (2001) illustrated how the log-normal distributions are widespread through the science. In this article they presented a table which summarizes some fields where the lognormal distribution is applied: in Geology (Concentration of elements Co, Cu, Cr, ^{226}Ra, Au, and U), in medicine (latency periods of infectious diseases, survival times after cancer diagnosis), in environmental science (rainfall, air pollution, the distribution of particles, chemicals and organism in the environment), food technology (size of unit), ecology (species abundance), linguistics (length of spoken words in phone conversations, length of sentences), social sciences (age of marriage, farm size in England and Wales, Incomes), in operation research (time distribution in queuing applications), in finance (a long term discount factor), and it has been used in other diverse areas of the sciences. They concluded: "We feel that doing so would lead to a general preference for the log-normal or multiplicative normal, distribution over the Gaussian distribution when describing original data."

Although the lognormal distribution is used for modeling positively skewed data, depending on the values of its parameters, the lognormal distribution can have various shapes including a bell-curve similar to the normal distribution. The maximum likelihood method is the most popular estimation technique for many distributions because it selects the values of the distribution's parameters that give the observed data the greatest

probability, and it has been used for estimating the parameters μ *and* σ of the two-parameter lognormal distribution. The maximum likelihood estimator of the location or threshold parameter γ for the three-parameter lognormal distribution is the smallest order statistic $X_{(1)}$, which is always greater than γ. The inclusion of the third parameter, the location parameter γ, may create some trouble in the estimation of the three-parameter lognormal distribution. Some modification has been proposed to improve the maximum likelihood estimator, but the iterative procedure that they suggest, for some samples the estimates did not exist. This research proposes a new approach for constructing exact confidence intervals of the location or threshold parameter γ.

1.2 Background

The probability density function of the three-parameter lognormal distribution is

$$f(x;\mu,\sigma,\gamma) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}} \exp\left\{-\frac{[\ln(x-\gamma)-\mu]^2}{2\sigma^2}\right\}, \qquad (1.2.1)$$

where $0 \le \gamma < x$, $-\infty < \mu < \infty$, $\sigma > 0$. μ, σ and γ are the parameters of the distribution. The distribution becomes the two-parameter lognormal distribution when $\gamma = 0$.

The three-parameter lognormal distribution has been studied extensively by Yuan (1933), Cohen (1951), Hill (1963), Harter and Moore (1966), Munro and Wixley (1970), Giesbrecht and Kempthome (1976), Kane (1978, 1982), Cohen and Whitten (1980, 1981), Wingo (1984), and Cohen and White and Ding (1985), and by others. In 1933 Yuan obtained moment estimators of the three parameters of the lognormal distribution

by equating the first three sample moments to corresponding population moments and used iterative procedure. In 1963 Hill demonstrated that the global maximum likelihood led to the inadmissible estimates $\hat{\gamma}=x_{(1)}$, $\hat{\mu}=-\infty$, $\hat{\sigma}=+\infty$, regardless of the sample. The solution of the likelihood equation led to local-maximum-likelihood estimates, but in some cases the local maximums do not exist. Harter and Moore (1966) proposed an interactive procedure for obtaining maximum-likelihood estimates of the parameters based on complete, singly censored, and doubly censored samples. Munro and Wixley (1970) examined the estimator based on order statistics of small samples from a three-parameter lognormal distribution. Wingo (1984) suggested an improve algorithm for computing local-maximum likelihood estimators (LMLE), which involves only a single numerical maximization of a suitably transformed conditional log-likelihood function (LLF). Cohen, Whitten and Ding (1985) presented modification of the moment estimators (MME) and modification of the maximum likelihood estimators (MMLE) in which the third moment and the third local maximum likelihood estimating equations were respectively replaced by functions of the first order statistic.

1.3 Research Hypothesis

In this research, all parameters (μ, σ, and γ) are assumed to be unknown. Let $X_1, X_2, X_3, \dots, X_n$ be the observations of a sample from the three-parameter lognormal distribution and $X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)}$ the corresponding order statistics. Let $\mathbf{X}$ = $(X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)})$ a vector that represents a sample from the three-parameter lognormal distribution.

Our attention is focused on estimating parameter γ, because when $\hat{\gamma}$ is known, then it follows from the first two maximum likelihood estimation equations that

$$\hat{\mu} = \frac{1}{n}\sum_1^n ln(X_i - \hat{\gamma}), \tag{1.3.1}$$

and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_1^n ln^2(X_i - \hat{\gamma}) - \left[\frac{1}{n}\sum_1^n (X_i - \hat{\gamma})\right]^2. \tag{1.3.2}$$

In order to find a $1 - \alpha$ confident interval for the parameter γ, a pivotal quantity, $\xi(\boldsymbol{X}, \gamma)$, is defined as

$$\xi(X,\gamma) = \frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} \ln(x_i - \gamma)}{n - 2[n/3]} - \dfrac{\sum_{i=1}^{[n/3]} \ln(x_i - \gamma)}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n} \ln(x_i - \gamma)}{[n/3]} - \dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} \ln(x_i - \gamma)}{n - 2[n/3]}}, \tag{1.3.3}$$

where $[n/3]$ represents the integer part of $n/3$. $\xi(\boldsymbol{X}, \gamma)$ can be rewritten as

$$\xi(\boldsymbol{X},\gamma) = \frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} \dfrac{\ln(x_{(i)} - \gamma) - \mu}{\sigma}}{n - 2[n/3]} - \dfrac{\sum_{i=1}^{[n/3]} \dfrac{\ln(x_{(i)} - \gamma) - \mu}{\sigma}}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n} \dfrac{\ln(x_{(i)} - \gamma) - \mu}{\sigma}}{[n/3]} - \dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} \dfrac{\ln(x_{(i)} - \gamma) - \mu}{\sigma}}{n - 2[n/3]}}. \tag{1.3.4}$$

Since $\frac{\ln(x_{(i)} - \gamma) - \mu}{\sigma} = z_{(i)} \sim N(0,1)$, the pivotal quantity can be written as

$$\xi(\boldsymbol{X}, \gamma) = \frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} z_{(i)}}{n-2[n/3]} - \dfrac{\sum_{i=1}^{[n/3]} z_{(i)}}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n} z}{[n/3]} - \dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} z_{(i)}}{n-2[n/3]}} \tag{1.3.5}$$

The pivotal quantity $\xi(\boldsymbol{X}, \gamma)$, as a mathematical function, is strict monotonic function of γ, and as a random variable, the probability distribution of $\xi(\boldsymbol{X}, \gamma)$ does not depend on any parameters of the population. Then $\xi(\boldsymbol{X}, \gamma)$ has the same probability distribution for all values of γ, μ, and σ.

Monte Carlo simulation can be used to find quantiles of $\xi(\boldsymbol{X}, \gamma)$. For fixed value of n and α, we can find numbers $\xi_{\alpha/2}$ and $\xi_{1-\alpha/2}$ satisfying

$$P\left(\xi_{\alpha/2} < \xi(\boldsymbol{X}, \gamma) < \xi_{1-\alpha/2}\right) = 1 - \alpha \quad , \tag{1.3.6}$$

where $\xi_{\alpha/2}$ represents the empirical $\alpha/2$ quintile of the random variable $\xi(\boldsymbol{X}, \gamma)$. The definition of the $\alpha/2$ th quintile of $\xi(\boldsymbol{X}, \gamma)$ is

$$P\left(\xi(\gamma) < \xi_{\alpha/2}\right) = \alpha/2 \ \ and \ P\left(\xi(\gamma) > \xi_{\alpha/2}\right) = 1 - \alpha/2. \tag{1.3.7}$$

Since $\xi(\boldsymbol{X}, \gamma)$ is an increasing function of γ, then the confidence interval of γ can be constructed from the Equations (1.3.8) and (1.3.9), which were obtained from the equation (1.3.6).

$$\xi(\boldsymbol{X},\gamma) = \xi_{\alpha/2} \tag{1.3.8}$$

$$\xi(\mathbf{X},\gamma) = \xi_{1-\alpha/2} \quad . \tag{1.3.9}$$

Equations (1.3.8) and (1.3.9) can be solved to find γ_L and γ_U. Both γ_L and γ_U must be smaller than $X_{(1)}$.

Therefore, the hypothesis for this research is that a table for empirical quantiles of the random variable $\xi(\boldsymbol{X},\gamma)$ and the equations (1.3.8) and (1.3.9) can be used to find confidence interval for the parameter γ. Confidence interval of γ will be squeezed to obtain point estimate of γ.

1.4 Method and Statistical Analysis

Monte Carlo simulations will be used to establish the quantiles of the random variable $\xi(\boldsymbol{X},\gamma)$ for a given sample size n and α. A large number of pseudo-random samples were generated from a normal distribution with parameters $\mu = 0$ and $\sigma = 1$.The pseudo random sample of size n drawn from the normal distribution N(0,1) will be sorted in ascending order, and the value of $\xi(\boldsymbol{X},\gamma)$ is calculated for each sample using equation (1.3.5). This procedure is repeated 10,000,000 times for each sample size. Then all $\xi(\boldsymbol{X},\gamma)$ values will be sorted in ascending order to find the percentiles of the pivotal quantity for each n.

In order to use the percentiles or critical values $\xi_{\alpha/2}$ and $\xi_{1-\alpha/2}$ to find confidence interval for γ, it can be shown that $\xi(\boldsymbol{X},\gamma)$ is a strictly increasing function of γ.

A random sample $\mathbf{X}$ will be drawn from a population that follows a lognormal distribution with known parameters $\gamma, \mu,$ and σ using the following formula in SAS/IML

$$x = \exp(\mu + \sigma * rannor(repeat(0, n)) + \gamma .$$

The $(1-\alpha)\%$ confidence interval for γ will be calculated using the critical values $\xi_{\alpha/2}$ and $\xi_{1-\alpha/2}$ for a given n and α using the equations (1.3.8) and (1.3.9). The equations can be solved for γ using the bisection method.

Statistical simulation will be used to evaluate the performance of the estimation method proposed, and the proposed method will be compared with the existing methods in the literature.

1.5 Two-Parameter Lognormal Distribution Review

The probability density function (pdf) of the two-parameter lognormal distribution is:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left\{-\frac{(lnx-\mu)^2}{2\sigma^2}\right\}, \quad x > 0, \quad -\infty < \mu < \infty, \quad \sigma > 0. \qquad (1.5.1)$$

If X is a random variable that has a log-normal probability distribution, then Y = lnX is normally distributed with scale parameters μ and shape parameter σ, which are the mean and standard deviation of the random variable Y = lnX.

Casella and Berger presented in Statistical Inference (2002) page 81 an interesting property of the lognormal distribution that all moments exist and are finite. The rth moment of X about the origin is

$$E(X^r) = e^{(r\mu + r^2\sigma^2/2)} \qquad (1.5.2)$$

However, the lognormal distribution does not have a moment generating function.

The expected value of a random variable X that follows a two-parameter lognormal distribution can be calculated using (1.5.2).

$$E(X) = e^{\left(\mu + \frac{\sigma^2}{2}\right)}$$

Using (1.5.2) again to calculate the second moment of X, the variance of X is given by

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= e^{2(\mu+\sigma^2)} - \left(e^{\left(\mu + \frac{\sigma^2}{2}\right)}\right)^2 \\ &= e^{(2\mu+\sigma^2)}\left[e^{\sigma^2} - 1\right]. \end{aligned}$$

If the expected value and the variance of X are known, we can obtain equivalent expression for the parameters $\mu \ and \ \sigma^2$

$$\mu = ln(E(X)) - \frac{1}{2} ln\left(1 + \frac{Var(X)}{(E(X))^2}\right)$$

$$\sigma^2 = ln\left(1 + \frac{Var(X)}{(E(X))^2}\right) \ .$$

The cumulative distribution function of the log-normal distribution (1.5.1) with two parameters is

$$F_X(x;\ \mu, \sigma)\ ; = \Phi\left(\frac{lnx - \mu}{\sigma}\right) \ .$$

Since the median of X is such a point where $F_X(x;\ \mu, \sigma) = 0.5$, then it is not difficult to prove that $Median\ (X) = e^{\mu}$. The mode of X is the point of global maximum of the

probability function (1.5.1) which satisfy the equation . Then the mode of X is . The relationship among the mean, the median and the mode is

The geometric mean of the two-parameter log-normal distribution is .

= Geometric-Mean(X).

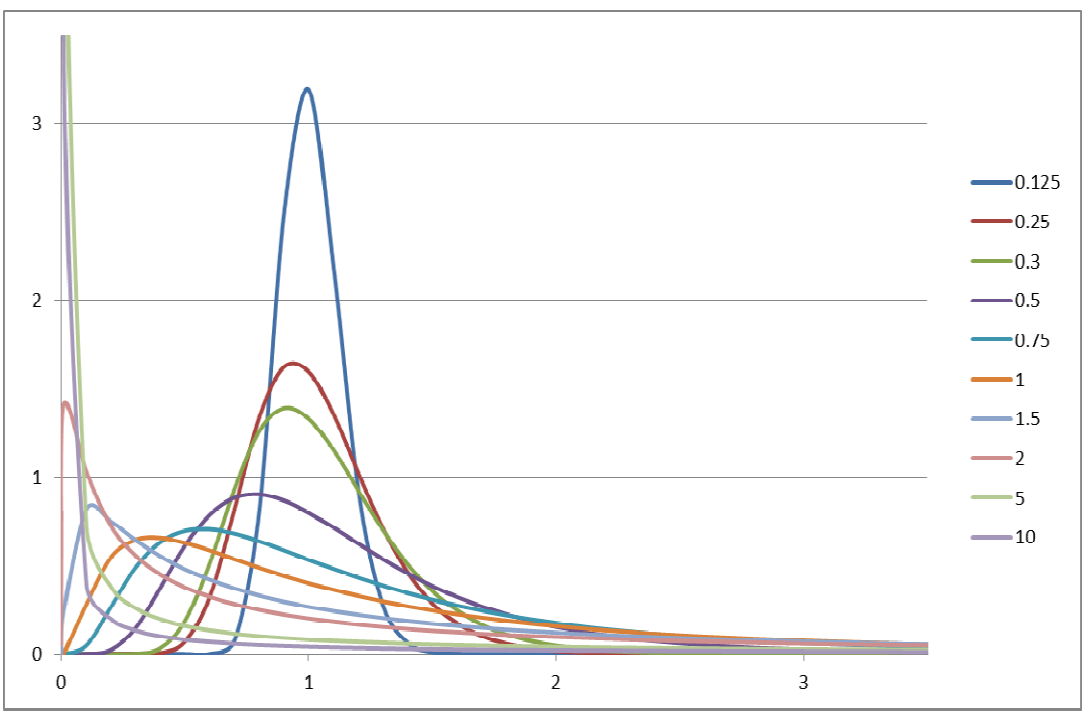


Figure 1.5.1 Lognormal Density Plots for $\mu = 0$, $\gamma = 0$, and some values of σ.

An important property of the log-normal distribution is that, for very small values of the shape parameter σ $(\sigma < 1/3)$, the shape of the lognormal distribution is nearly close to a normal distribution. Figure 1.5.1 shows this property.

The method of maximum likelihood is the most commonly used technique of parameter estimation for the two-parameter lognormal distribution, because it finds the parameter estimators that make the data "more likely" than any other value of the parameter that would make them. The likelihood function of the two-parameter lognormal distribution for a given sample is derived by taking the product of the probability densities of the individual observations X_i.

$$L(\mu, \sigma^2|\boldsymbol{X}) = \prod_{i=1}^{n}[f(x_i|\mu, \sigma^2]$$

$$= \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \prod_{i=1}^{n} \frac{1}{x_i} exp\left[\sum_{i=1}^{n} \frac{-(\ln(x_i)-\mu)^2}{2\sigma^2}\right]$$

$$\mathcal{L}(\mu, \sigma^2|\boldsymbol{X}) = ln[L(\mu, \sigma^2|\boldsymbol{X})] = ln\left[\frac{1}{\sqrt{(2\pi\sigma^2)^n}} \prod_{i=1}^{n} \frac{1}{x_i} exp\left[\sum_{i=1}^{n} \frac{-(\ln(x_i)-\mu)^2}{2\sigma^2}\right]\right]$$

$$= -\frac{n}{2} ln(2\pi\sigma^2) - \sum_{i=1}^{n} ln(x_i) - \frac{\sum_{i=1}^{n} ln(x_i)^2}{2\sigma^2} + \frac{\sum_{i=1}^{n} \mu ln(x_i)}{\sigma^2} - \frac{n\mu^2}{2\sigma^2},$$

The values of μ and σ^2 that maximize $L(\mu, \sigma^2|\boldsymbol{X})$ also maximize $\mathcal{L}(\mu, \sigma^2|\boldsymbol{X})$. To find $\hat{x}$ and $\hat{\sigma}^2$ the maximum likelihood estimators of μ and σ^2, the gradient of $\mathcal{L}$ with respect to μ and σ^2 is calculated, and equate to 0.

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{\sum_{i=1}^{n} ln(x_i)}{\hat{\sigma}^2} - \frac{n\hat{\mu}}{\hat{\sigma}^2} = 0$$

$$\rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} ln(x_i)}{n}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\hat{\sigma}^2} - \frac{\sum_{i=1}^{n}(ln(X_i) - \hat{\mu})^2}{2}(-\hat{\sigma}^2)^{-2} = 0$$

$$\rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(\ln(X_i) - \hat{\mu})^2}{n}$$

Then, the maximum likelihood estimators for $\mu \; and \; \sigma^2$ are

$$\hat{\mu} = \frac{\sum_{i=1}^{n} ln(x_i)}{n} \tag{1.5.3}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(\ln(X_i) - \frac{\sum_{i=1}^{n} ln(X_i)}{n}\right)^2}{n} \quad . \tag{1.5.4}$$

To verify that these estimators maximize $\mathcal{L}(\mu, \sigma^2|\boldsymbol{X})$, the second derivative matrix of $\mathcal{L}$ is calculated.

$$\frac{\partial^2 \mathcal{L}}{\partial \mu^2} = \frac{\partial}{\partial \mu}\left[\frac{\sum_{i=1}^{n} ln(x_i)}{\hat{\sigma}^2} - \frac{n\hat{\mu}}{\hat{\sigma}^2}\right] = -\frac{n}{\hat{\sigma}^2}.$$

$$\frac{\partial^2 \mathcal{L}}{\partial (\sigma^2)^2} = \frac{\partial \mathcal{L}}{\partial \sigma^2}\left[-\frac{n}{2}\frac{1}{\hat{\sigma}^2} - \frac{\sum_{i=1}^{n}(ln(X_i) - \hat{\mu})^2}{2}(-\hat{\sigma}^2)^{-2}\right] = -\frac{1}{2(\hat{\sigma}^2)^3}\left[\sum_{i=1}^{n}\left(ln(X_i - \hat{\mu})\right)^2\right].$$

$$\frac{\partial \mathcal{L}}{\partial \mu}\left[\frac{\partial \mathcal{L}}{\partial \sigma^2}\right] = \frac{\partial \mathcal{L}}{\partial \mu}\left[-\frac{n}{2}\frac{1}{\hat{\sigma}^2} - \frac{\sum_{i=1}^{n}(ln(X_i) - \hat{\mu})^2}{2}(-\hat{\sigma}^2)^{-2}\right] = 0.$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2}\left[\frac{\partial \mathcal{L}}{\partial \mu}\right] = \frac{\partial \mathcal{L}}{\partial \sigma^2}\left[\frac{\sum_{i=1}^{n} ln(x_i)}{\hat{\sigma}^2} - \frac{n\hat{\mu}}{\hat{\sigma}^2}\right] = 0.$$

The Hessian or second derivative matrix $\mathcal{L}$ is given by

$$H = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \mu^2} & \frac{\partial \mathcal{L}}{\partial \sigma^2}\left[\frac{\partial \mathcal{L}}{\partial \mu}\right] \\ \frac{\partial \mathcal{L}}{\partial \mu}\left[\frac{\partial \mathcal{L}}{\partial \sigma^2}\right] & \frac{\partial^2 \mathcal{L}}{\partial (\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{1}{2(\hat{\sigma}^2)^3}\left[\sum_{i=1}^{n}\left(ln(X_i - \hat{\mu})\right)^2\right] \end{bmatrix}.$$

The determinant of the Hessian matrix is

$$|H| = \frac{n}{2(\hat{\sigma}^2)^4}\left[\sum_{i=1}^{n}\left(ln(X_i - \hat{\mu})\right)^2\right] > 0.$$

On the other hand for all non-zero vector $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$

$$Z^T H Z = [Z_1 \;\; Z_2] \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{1}{2(\hat{\sigma}^2)^3}\left[\sum_{i=1}^{n}\left(ln(X_i - \hat{\mu})\right)^2\right] \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

$$Z^T H Z = -\frac{n{Z_1}^2}{\hat{\sigma}^2} - \frac{{Z_2}^2}{2(\hat{\sigma}^2)^3}\left[\sum_{i=1}^{n}\left(ln(X_i - \hat{\mu})\right)^2\right] < 0$$

$Z^T H Z < 0$ because $n, {Z_1}^2, {Z_2}^2, \hat{\sigma}^2$ *and* $\sum_{i=1}^{n}\left(ln(X_i - \hat{\mu})\right)^2$ are all positive. Since |H|>0 and the Hessian is negative-definite, the log-likelihood function $\mathcal{L}$ and the likelihood function L have a local maximum when μ *and* σ^2 take the values in equation (1.5.3) and (1.5.4) respectively.

1.6 Three-Parameter Lognormal Distribution Review

The probability density function (pdf) of the three-parameter lognormal distribution is:

$$f(x;\mu,\sigma,\gamma) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}} \exp\left\{-\frac{[\ln(x-\gamma)-\mu]^2}{2\sigma^2}\right\}, \qquad (1.6.1)$$

where $x > \gamma \geq 0$, $-\infty < \mu < \infty$, $\sigma > 0$, and γ is the threshold parameter or location parameter that defines the point where the support set of the distribution begins; μ is the scale parameter that stretch or shrink the distribution and σ is the shape parameter that affects the shape of the distribution. If X is a random variable that has a three-parameter log-normal probability distribution, then $Y = \ln(X-\gamma)$ has a normal distribution with mean μ and variance σ^2. The two-parameter lognormal distribution is a special case of the three-parameter lognormal distribution when $\gamma = 0$.

In 1933, Yuan derived the mean, variance, third standard moment (skewness), and forth standard moment (kurtosis) of the lognormal distribution as a function of μ, σ and γ.

$$\text{Mean (X)} = \text{E(X)} = \gamma + \exp\left(\mu + \frac{\sigma^2}{2}\right) \qquad (1.6.2)$$

$$\text{Var(X)} = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1). \qquad (1.6.3)$$

$$\alpha_3 = \sqrt{\exp(\sigma^2) - 1} * (\exp(\sigma^2) + 2) \qquad (1.6.4)$$

$$\alpha_4 = \exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 3. \qquad (1.6.5)$$

If the expected value and the variance of X are known, we can solve these two equations to obtain equivalent expression for the parameters μ and σ^2.

$$\mu = \ln(E(X)) - \frac{1}{2}\ln\left(1 + \frac{Var(X)}{(E(X))^2}\right)$$

$$\sigma^2 = \ln\left(1 + \frac{Var(X)}{(E(X))^2}\right)$$

The cumulative distribution function of the three-parameter log-normal distribution is

$$F_X(x;\mu,\sigma,\gamma) = \Phi\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right).$$

The median of X is a value that satisfies the equation $F_X(x;\mu,\sigma,\gamma) = 0.5$, then

$$\frac{\ln(x-\gamma)-\mu}{\sigma} = 0$$

$$\rightarrow Median(X) = \gamma + e^{\mu}.$$

The mode of X is the point of global maximum of the probability function (1.6.1) which satisfies the equation $f'(x;\mu,\sigma,\gamma) = 0$.

$$Mode(X) = \gamma + e^{(\mu-\sigma^2)}.$$

The relationship among the mean, the median and the mode is

$$Mode(X) < Median(X) < E(X).$$

Estimating the parameters of the three-parameter lognormal distribution is more complicated than estimating the parameters of the two-parameter lognormal distribution. The likelihood function of the three-parameter lognormal distribution for a given sample $\mathbf{X} = \{x_1, x_2, ... x_n\}$ is derived by taking the product of the probability densities of the individual observations Xi.

$$L(\mu, \sigma, \lambda \mid X) = \prod_{i=1}^{n} [f(x_i \mid \mu, \sigma, \gamma)]$$

$$L(\mu, \sigma, \gamma \mid X) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . \prod_{i=1}^{n} (x_i - \gamma)^{-1} . \exp\left[-\frac{\sum_{i=1}^{n} (\ln(x_i - \gamma) - \mu)^2}{2\sigma^2}\right] . I_{\min\{x_i, ..., x_n\} > \gamma} \quad (1.6.6)$$

If $\mathcal{L}(\mu, \sigma, \gamma \mid \mathbf{X}) = \ln(L(\mu, \sigma, \gamma \mid \mathbf{X})$, the values of μ, σ, and γ that maximize $L(\mu, \sigma, \gamma \mid \mathbf{X})$ also maximize $\mathcal{L}(\mu, \sigma, \gamma \mid \mathbf{X})$. Cohen (1951) obtained local maximum likelihood estimators (LMLE) equating the partial derivatives of the log-likelihood to zero. Differentiating the log-likelihood function and equating to zero, we obtain the local maximum likelihood estimating equations.

$$\ln L(\mu, \sigma, \gamma \mid \mathbf{X}) = -n \ln \sigma - n \ln \sqrt{2\pi} - \sum_{i=1}^{n} \ln(x_i - \gamma) - \frac{\sum_{i=1}^{n} (\ln(x_i - \gamma) - \mu)^2}{2\sigma^2}$$

i) $$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\mu\sigma^2} \sum_{i=1}^{n} [\ln(x_i - \gamma) - \mu] = 0$$

$$\rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} \ln(x_i - \hat{\gamma})}{n} \tag{1.6.7}$$

$$ii)\ \frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} \left[\ln(x_i - \gamma) - \mu\right]^2 = 0$$

$$\rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\ln(x_i - \hat{\gamma}) - \hat{\mu}\right)^2 \tag{1.6.8}$$

$$iii)\quad \frac{\partial \ln L}{\partial \gamma} = \sum_{i=1}^{n} \frac{1}{x_i - \gamma} + \frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{\ln(x_i - \gamma) - \mu}{x_i - \gamma} = 0 \tag{1.6.9}$$

It can be seen that the maximum likelihood estimator (MLE) of γ is $\min\{x_1,...,x_n\} = x_{(1)}$. Hill (1963) demonstrated the existence of paths along which the likelihood function of any ordered sample, $x_{(1)},...,x_{(n)}$, tends to ∞ as (γ, μ, σ^2) approach ($x_{(1)}$, $-\infty$, ∞). This global maximum thereby leads to the inadmissible estimates $\hat{\gamma} = X_{(1)}$, $\hat{\mu} = -\alpha$, and $\hat{\sigma}^2 = +\alpha$, regardless of the sample. See Cohen and Whitten (1980). However, Cohen found a local maximum likelihood estimate for γ by replacing (1.6.7) and (1.6.8) in the last equation (1.6.9) to obtain an equation in γ.

$$\left[\sum_{i=1}^{n} \frac{1}{x_i - \gamma}\right]\left[\sum_{i=1}^{n} \ln(x_i - \gamma) - \sum_{i=1}^{n} \left(\ln(x_i - \gamma)\right)^2 + \frac{1}{n}\left(\sum_{i=1}^{n} \ln(x_i - \gamma)\right)^2\right] - n\sum_{i=1}^{n} \frac{\ln(x_i - \gamma)}{x_i - \gamma} = 0 \tag{1.6.10}$$

Equation (1.6.10) can be solved iteratively to obtain the local maximum likelihood estimator of γ. Cohen and Whitten (1980) presented a modified maximum

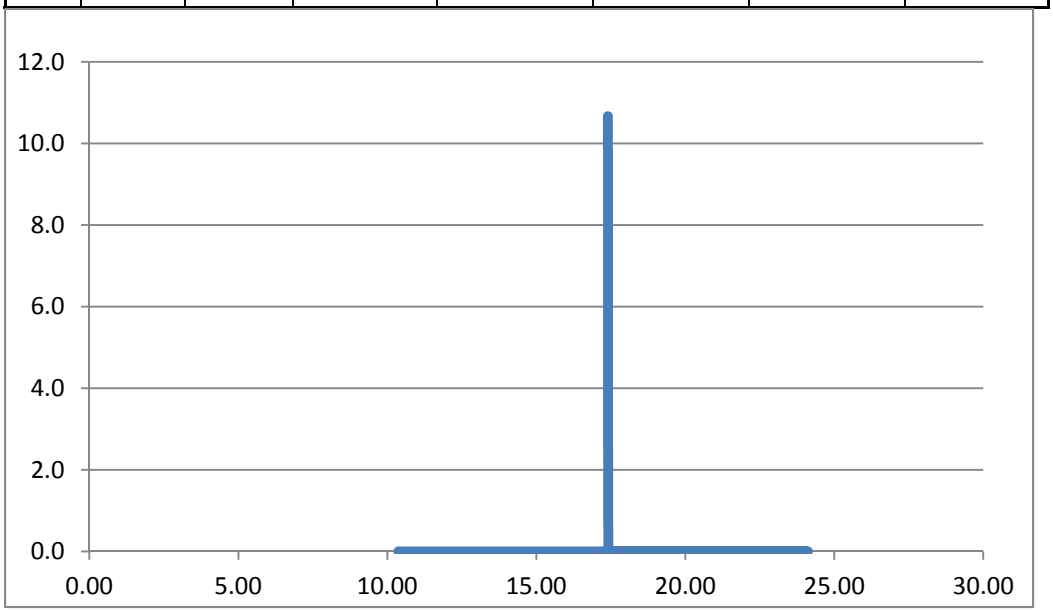
likelihood estimator by replacing the third equation (1.6.10) for another equation of γ a little less complicated; however, when multiple admissible roots occur, Cohen and Whitten suggest choosing as LMLE for γ, the root which results in the closest agreement between the sample mean $\bar{x}$ and $E(X) = \hat{\mu}_x = \hat{\gamma} + \exp(\hat{\mu} + \hat{\sigma}^2/2)$.

In this research, all the parameters μ, σ and γ, are assumed to be unknown. Our attention will be focused on estimating parameter γ because when γ is known, equations (1.6.7) and (1.6.3) can be used to estimate μ and σ. The objective of this research is to replace the third equation of the maximum likelihood with another equation which has only one root for γ. Confidence intervals and one-side confidence limits can also be obtained.

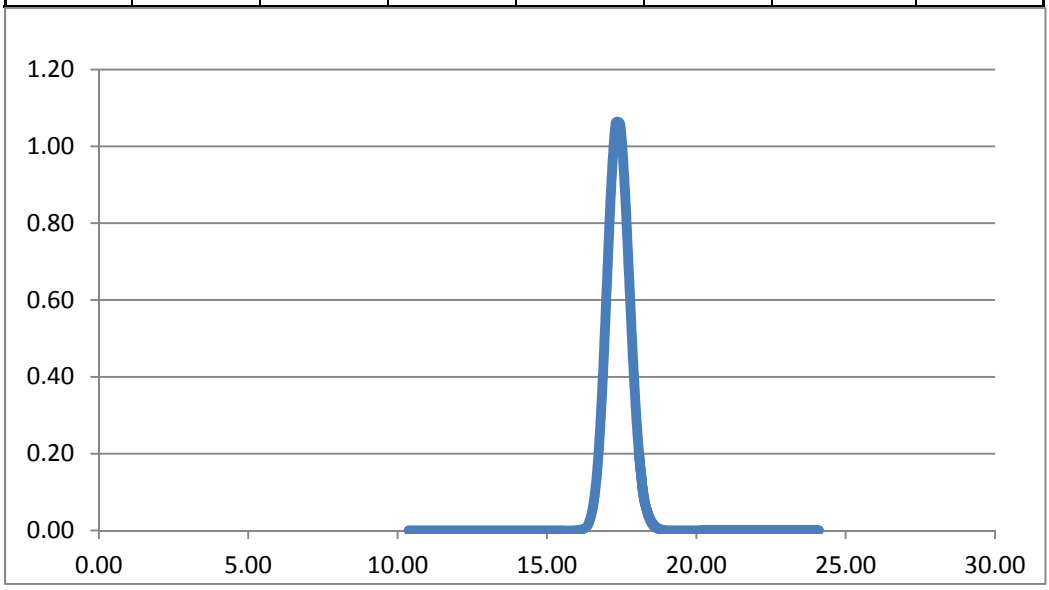
To find a good estimator and to construct a confidence interval of γ is a difficult problem. The log-normal pdf's shape varies from almost symmetric, as an upturned T, to highly skewed, as an L. Figure 1.6 shows the diversity of shape of the three-parameter log-normal distribution. The shape parameter σ affects the skewness and peakedness of the pdf. Equation (1.6.4) and (1.6.5) confirm that the skewnees (α_3) and peakedness (α_4) only depend on σ. On the other hand, when the scale parameter μ is increased to 8, the dispersion of the random variable also increases, but the shape of the pdf does not vary. The last two graphs show the same pdf, but with a different scale in the vertical axis. The pdf is very close to the horizontal axis, and the value of this pdf function is 0.0022 when x=mode. The larger the scale parameter μ, the more spread out the distribution. The threshold parameter λ basically shifts the pdf.

Figure 1.6 Different shapes of the Three-Parameters Log-Normal Distribution.

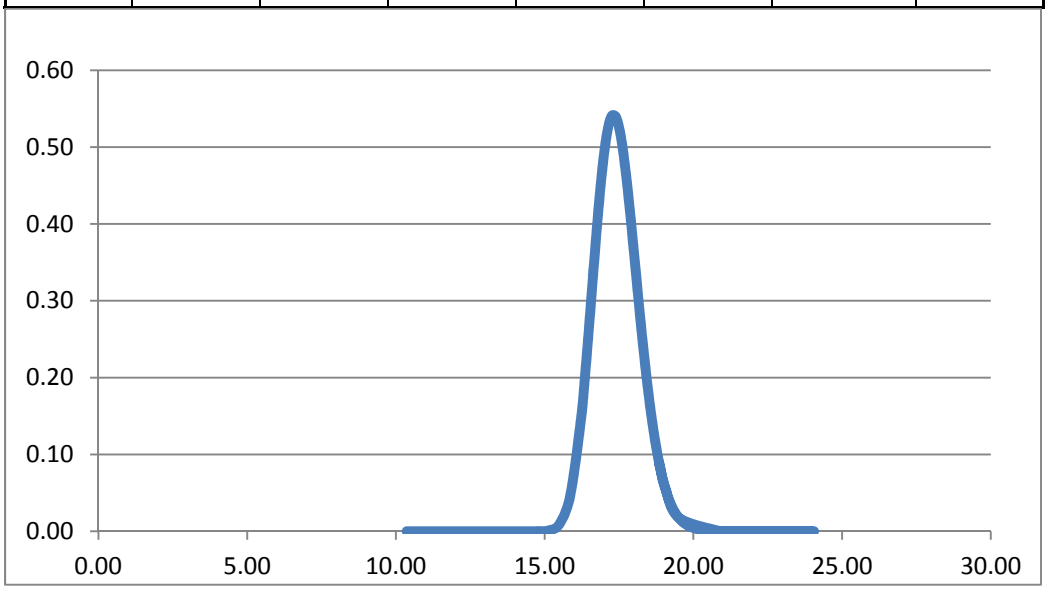
μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
2	0.001	10	17	17.39	17.39	0.003	3.000



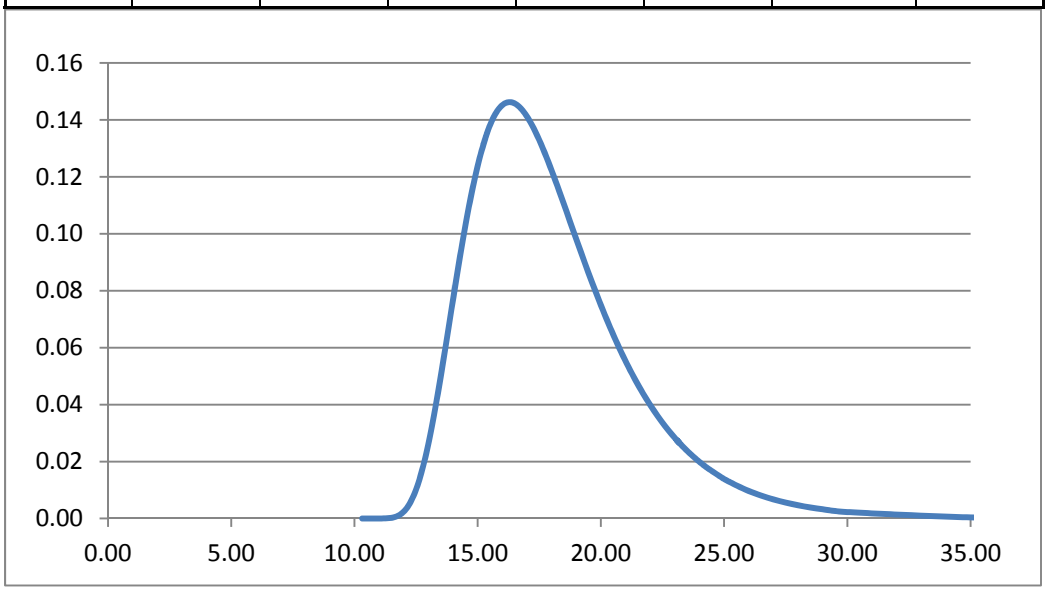
μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
2	0.05	10	17	17.39	17.40	0.15	3.04



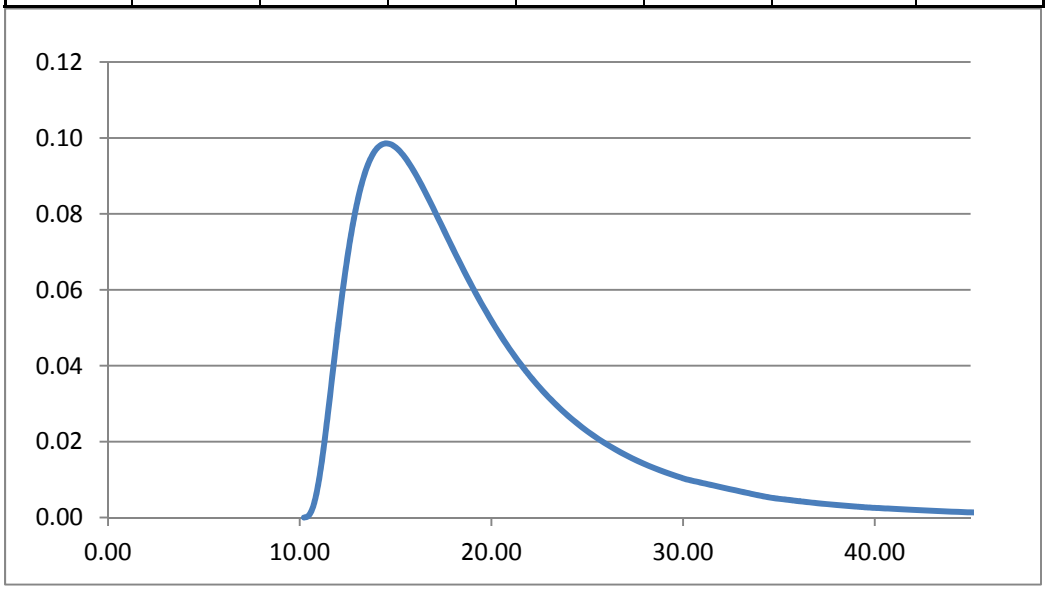
μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
2	0.1	10	17	17.39	17.43	0.30	3.16



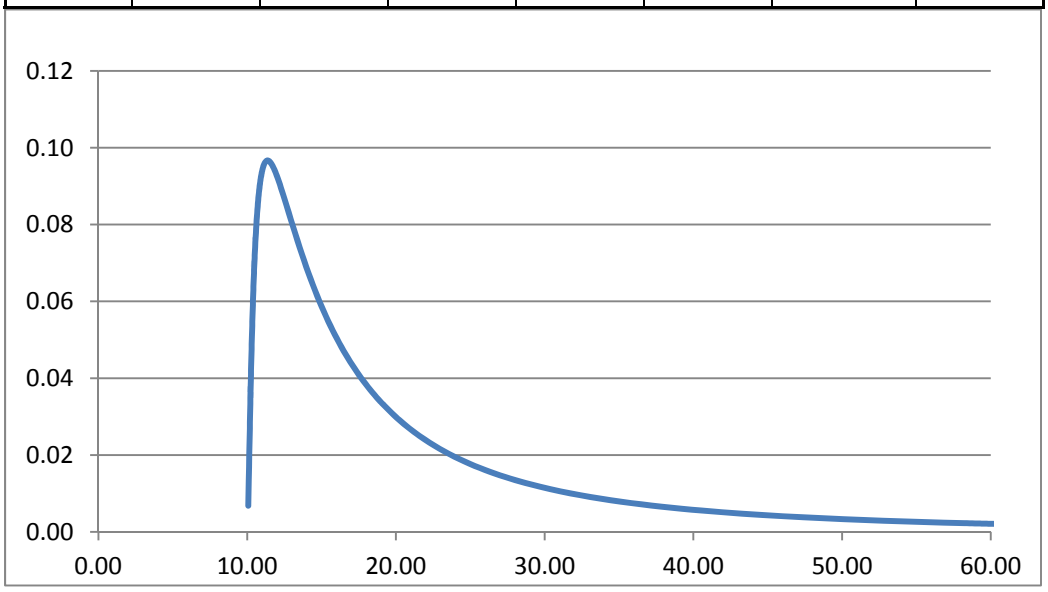
μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
2	0.4	10	16	17.39	18	1.32	6.26



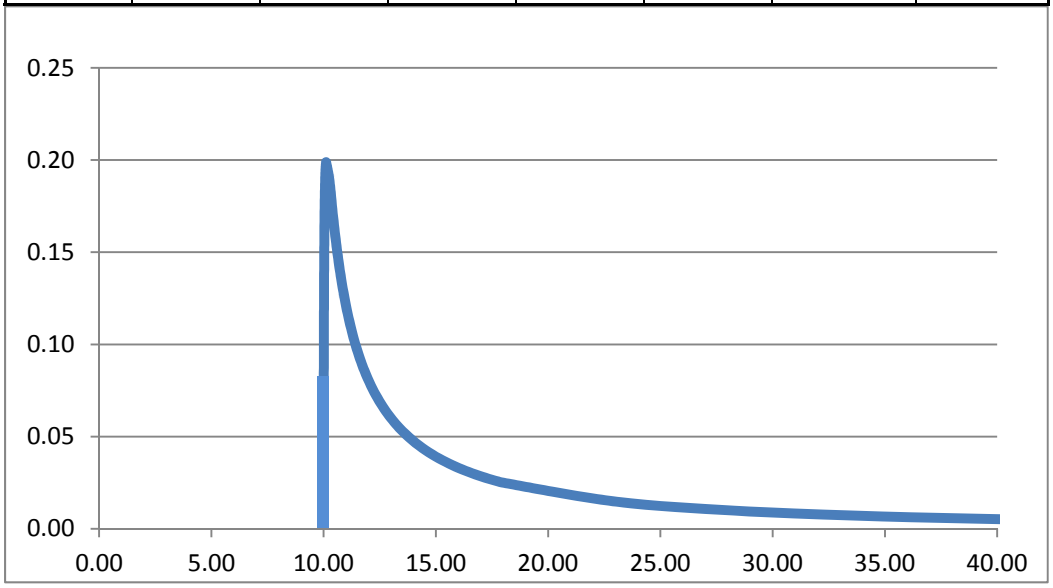
μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
2	0.7	10	15	17.39	19.44	2.89	20.79



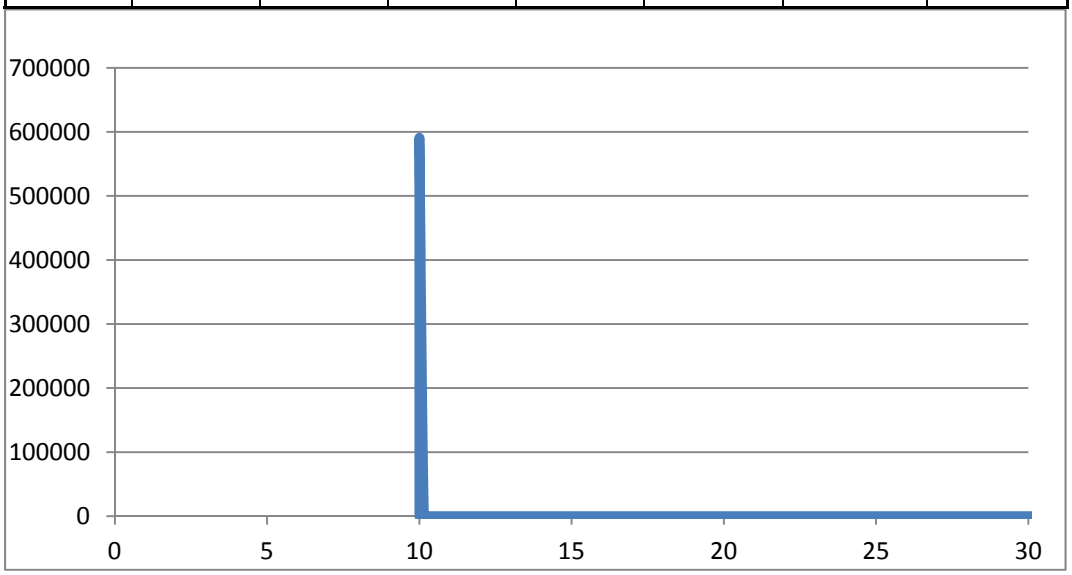
μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
2	1.3	10	11	17.39	27.20	15.60	1266.10



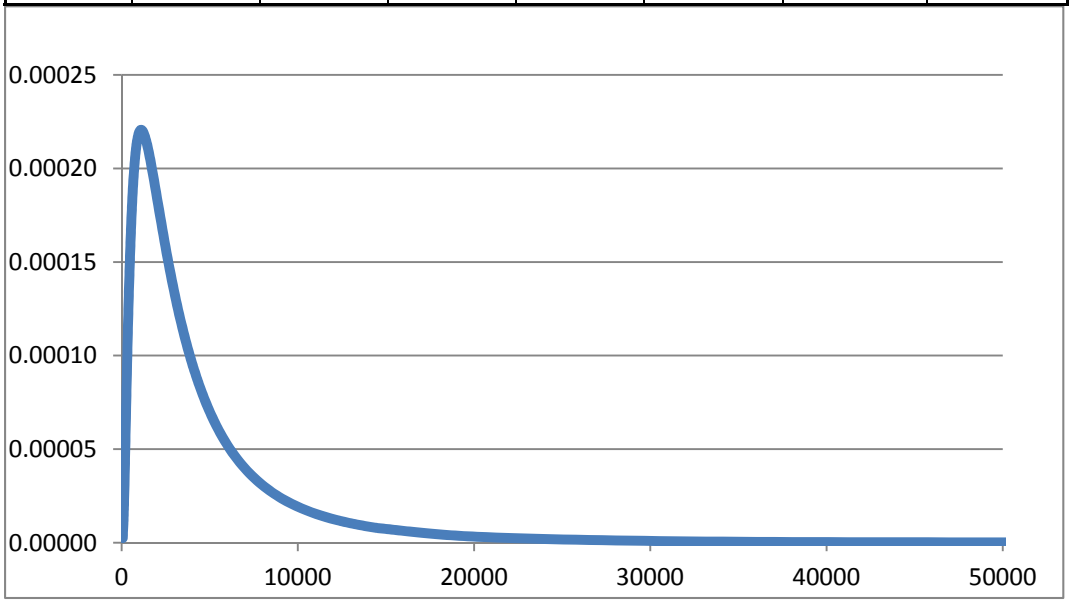
μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
2	2	10	10	17.39	64.60	414.36	9220560



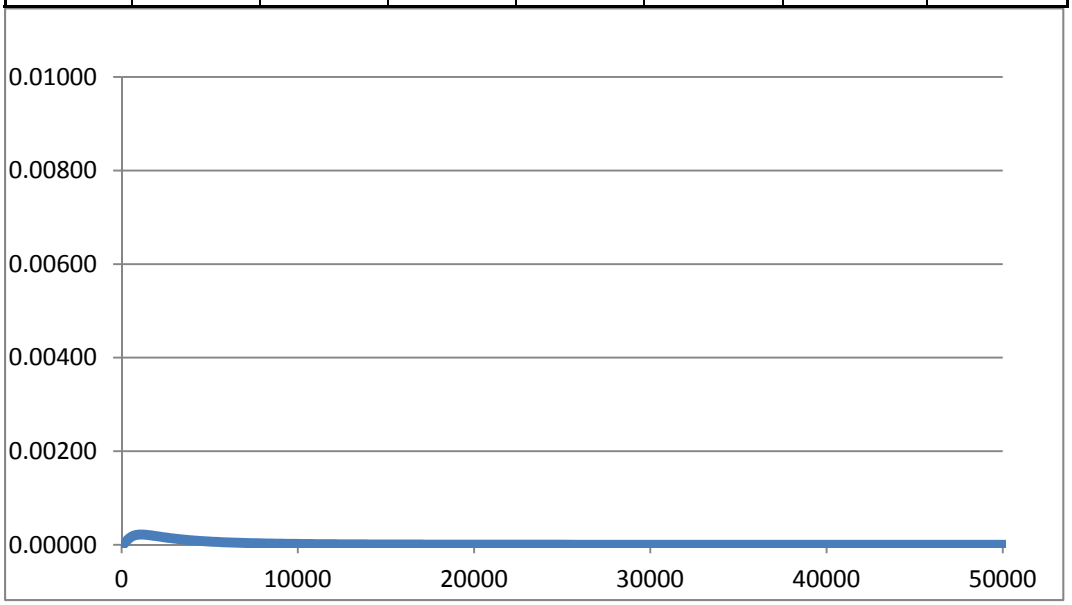
μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
2	6	10	10	17.39	4.85E+08	2.83E+23	3.45E+62



μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
8	1	10	1107	2990	4.95E+03	6.18E+23	3.45E+62



μ	σ	γ	**Mode**	**median**	**mean**	Skewness	Kurtosis
8	1	10	1107	2990	4.95E+03	6.18E+23	3.45E+62



2. INTERVAL AND POINT ESTIMATION FOR γ

Let $x_1, x_2, x_3, ..., x_n$ be the observations of a sample from a three-parameter lognormal distribution and $x_{(1)}, x_{(2)}, x_{(3)}, ..., x_{(n)}$ the corresponding order statistics. Let $X = \left(x_{(1)}, x_{(2)}, x_{(3)}, ..., x_{(n)}\right)$ be a vector that represents a sample point of a lognormal distributed population.

2.1 Pivotal Quantity

As mentioned in Chapter 1, a pivotal quantity ξ can be defined in order to estimate parameter γ.

$$\xi(\boldsymbol{X}, \gamma) = \frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} \ln(x_{(i)} - \gamma)}{n - 2[n/3]} - \dfrac{\sum_{i=1}^{[n/3]} \ln(x_{(i)} - \gamma)}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n} \ln(x_{(i)} - \gamma)}{[n/3]} - \dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} \ln(x_{(i)} - \gamma)}{n - 2[n/3]}} \quad , \qquad (2.1.1)$$

where $[n/3]$ represents the integer part of $n/3$. A random variable $Q(\boldsymbol{X}, \theta) = Q(X_1, X_2, ..., X_n, \theta)$ is said to be a pivotal quantity (or pivot) if the distribution of $Q(\boldsymbol{X}, \theta)$ is independent of all parameters. That is, if $\boldsymbol{X} \sim F(\boldsymbol{x} \mid \theta)$, then $Q(\boldsymbol{X}, \theta)$ has the same distribution for all values of θ. See G. Casella and R. Berger, Statistical Inference (2002). The expression of $\xi(\mathbf{X}, \gamma)$ in (2.1.1) can be rewritten as

$$\xi(\boldsymbol{X},\gamma)=\frac{\dfrac{\displaystyle\sum_{i=[n/3]+1}^{n-[n/3]}\frac{\ln(x_{(i)}-\gamma)-\mu}{\sigma}}{n-2[n/3]}-\dfrac{\displaystyle\sum_{i=1}^{[n/3]}\frac{\ln(x_{(i)}-\gamma)-\mu}{\sigma}}{[n/3]}}{\dfrac{\displaystyle\sum_{i=n-[n/3]+1}^{n}\frac{\ln(x_{(i)}-\gamma)-\mu}{\sigma}}{[n/3]}-\dfrac{\displaystyle\sum_{i=[n/3]+1}^{n-[n/3]}\frac{\ln(x_{(i)}-\gamma)-\mu}{\sigma}}{n-2[n/3]}}. \tag{2.1.2}$$

Since $\ln(x_i-\gamma)\sim N(\mu,\sigma^2)$ then $\dfrac{\ln(x_i-\gamma)-\mu}{\sigma}=Z_i \sim N(0,1)$ and $\dfrac{\ln(x_{(i)}-\gamma)-\mu}{\sigma}=Z_{(i)}$. The pivotal quantity ξ can also be express as a function of the order statistics,

$$\xi(\boldsymbol{X},\gamma)=\frac{\dfrac{\displaystyle\sum_{i=[n/3]+1}^{n-[n/3]}Z_{(i)}}{n-2[n/3]}-\dfrac{\displaystyle\sum_{i=1}^{[n/3]}Z_{(i)}}{[n/3]}}{\dfrac{\displaystyle\sum_{i=n-[n/3]+1}^{n}Z_{(i)}}{[n/3]}-\dfrac{\displaystyle\sum_{i=[n/3]+1}^{n-[n/3]}Z_{(i)}}{n-2[n/3]}} \tag{2.1.3}$$

The pdf of the order statistics $Z_{(i)}$ is

$$f_{Z(i)}(z)=\frac{n!}{(i-1)!(n-i)!}f_Z(z)\left[F_Z(z)\right]^{i-1}\left[1-F_Z(z)\right]^{n-i} \tag{2.1.3}$$

where $f_Z(z)$ and $F_Z(z)$ are the pdf and cdf respectively of the standard normal distribution which are independent of all parameters. The probability distribution of the pivot ξ does not depend on μ,σ or γ.

Another important property of the pivotal quantity $\xi(\boldsymbol{X},\gamma)$ is that, for a fix sample $\boldsymbol{X}$, the pivotal quantity ξ is a strictly increasing function of γ. This property of the pivotal quantity allows us to construct a confident interval for γ. Theoretical proof of this property can be obtained, but we will discuss this property in the Section 2.4 using a fix sample.

2.2 Critical Values of ξ for Any Sample Size n

It has been shown above that ξ satisfy the definition of pivotal quantity. Monte Carlo simulation method is used to find the critical values of the pivotal quantity ξ for different sample sizes from n=5 to n=150. To accomplish this objective, 10,000,000 pseudo-samples of size n were generated each time from a log-normal distributed population. Since the critical values of ξ and its probability distribution does not depend on any parameters, the following parameter combination $\mu = 0$, $\sigma = 1$, and $\gamma = 0$ was used in order to simplify the computation. Each pseudo-sample was sorted in ascending order, and $\xi(\boldsymbol{X},\gamma)$ was calculated for each sample using Equation (2.1.1). Then, 10,000,000 pivotal quantities ξ were obtained and sorted in ascending order. The following percentiles of ξ were calculated

$$\xi_{0.01}, \xi_{0.025}, \xi_{0.05}, \xi_{0.10}, \xi_{0.25}, \xi_{0.30}, \xi_{0.50}, \xi_{0.70}, \xi_{0.75}, \xi_{0.90}, \xi_{0.95}, \xi_{0.975}, \xi_{0.99}.$$

The critical values of ξ were calculated as

$$\xi_{\alpha/2} = \begin{cases} \xi_{(m\alpha/2)} & \text{if } (m\alpha)/2 \text{ is an integer} \\ \dfrac{\xi_{([m\alpha/2])} + \xi_{([m\varepsilon/2]+1)}}{2} & \text{otherwise} \end{cases}$$

and

$$\xi_{1-\alpha/2} = \begin{cases} \xi_{(m(1-\alpha/2))} & \text{if } m(1-\alpha/2) \text{ is an integer} \\ \dfrac{\xi_{([m(1-\alpha/2)])} + \xi_{([m(1-\varepsilon/2)]+1)}}{2} & \text{otherwise} \end{cases}$$

This procedure was repeated for each size n of samples. Table (2.2.) shows the critical values of ξ for sample sizes from n = 5 to n = 150.

Table 2.2 Critical Values for the Pivotal Quantity ξ.

n	$\xi_{0.005}$	$\xi_{0.01}$	$\xi_{0.025}$	$\xi_{0.05}$	$\xi_{0.50}$	$\xi_{0.95}$	$\xi_{0.975}$	$\xi_{0.99}$	$\xi_{0.995}$
5	0.122	0.156	0.219	0.288	1.000	3.471	4.562	6.410	8.210
6	0.119	0.153	0.216	0.286	1.000	3.498	4.621	6.540	8.435
7	0.176	0.215	0.283	0.353	1.000	2.834	3.539	4.656	5.677
8	0.222	0.262	0.330	0.399	1.000	2.502	3.027	3.818	4.506
9	0.218	0.258	0.326	0.396	1.000	2.524	3.063	3.877	4.593
10	0.258	0.298	0.366	0.435	1.000	2.297	2.724	3.347	3.876
11	0.290	0.330	0.399	0.466	1.000	2.147	2.508	3.020	3.446
12	0.287	0.327	0.396	0.463	1.000	2.162	2.528	3.054	3.488
13	0.315	0.356	0.424	0.490	1.000	2.043	2.360	2.802	3.162
14	0.341	0.382	0.448	0.512	1.000	1.954	2.234	2.621	2.930
15	0.338	0.378	0.444	0.509	1.000	1.964	2.249	2.643	2.958
16	0.361	0.401	0.466	0.529	1.000	1.889	2.145	2.495	2.776
17	0.381	0.420	0.485	0.547	1.000	1.829	2.061	2.378	2.627
18	0.378	0.418	0.482	0.544	1.000	1.838	2.075	2.395	2.646

Table 2.2 Critical Values for the Pivotal Quantity ξ.

n	$\xi_{0.005}$	$\xi_{0.01}$	$\xi_{0.025}$	$\xi_{0.05}$	$\xi_{0.50}$	$\xi_{0.95}$	$\xi_{0.975}$	$\xi_{0.99}$	$\xi_{0.995}$
19	0.396	0.436	0.499	0.560	1.000	1.785	2.002	2.295	2.521
20	0.413	0.452	0.515	0.574	1.000	1.741	1.943	2.211	2.419
21	0.411	0.450	0.512	0.572	1.000	1.748	1.952	2.225	2.436
22	0.426	0.464	0.526	0.585	1.000	1.709	1.900	2.152	2.348
23	0.440	0.478	0.539	0.597	1.000	1.676	1.855	2.091	2.273
24	0.438	0.476	0.537	0.595	1.000	1.681	1.861	2.100	2.281
25	0.451	0.489	0.549	0.606	1.000	1.650	1.821	2.045	2.217
26	0.464	0.501	0.560	0.616	1.000	1.623	1.785	1.996	2.157
27	0.461	0.499	0.558	0.614	1.000	1.628	1.791	2.004	2.165
28	0.473	0.510	0.569	0.624	1.000	1.603	1.758	1.962	2.115
29	0.484	0.520	0.578	0.632	1.000	1.580	1.729	1.920	2.066
30	0.482	0.518	0.577	0.631	1.000	1.585	1.735	1.928	2.074
31	0.493	0.529	0.586	0.639	1.000	1.565	1.707	1.892	2.031
32	0.502	0.537	0.594	0.647	1.000	1.546	1.683	1.860	1.993
33	0.500	0.536	0.593	0.645	1.000	1.550	1.688	1.867	2.001
34	0.509	0.545	0.601	0.653	1.000	1.532	1.665	1.836	1.964
35	0.518	0.553	0.608	0.660	1.000	1.517	1.645	1.809	1.932
36	0.515	0.551	0.607	0.658	1.000	1.519	1.648	1.814	1.938
37	0.524	0.559	0.614	0.665	1.000	1.505	1.629	1.789	1.907
38	0.532	0.567	0.621	0.671	1.000	1.491	1.611	1.765	1.881
39	0.531	0.565	0.619	0.669	1.000	1.493	1.614	1.769	1.885
40	0.538	0.572	0.626	0.676	1.000	1.481	1.598	1.748	1.858
41	0.545	0.579	0.632	0.681	1.000	1.468	1.582	1.727	1.834
42	0.544	0.577	0.631	0.680	1.000	1.471	1.586	1.732	1.840
43	0.550	0.584	0.637	0.685	1.000	1.460	1.571	1.712	1.815
44	0.557	0.590	0.642	0.690	1.000	1.449	1.557	1.695	1.796
45	0.556	0.589	0.641	0.689	1.000	1.451	1.560	1.699	1.799
46	0.562	0.595	0.647	0.694	1.000	1.441	1.547	1.682	1.780

Table 2.2 Critical Values for the Pivotal Quantity ξ.

n	$\xi_{0.005}$	$\xi_{0.01}$	$\xi_{0.025}$	$\xi_{0.05}$	$\xi_{0.50}$	$\xi_{0.95}$	$\xi_{0.975}$	$\xi_{0.99}$	$\xi_{0.995}$
47	0.568	0.601	0.652	0.699	1.000	1.432	1.535	1.665	1.761
48	0.567	0.599	0.650	0.698	1.000	1.433	1.537	1.668	1.765
49	0.572	0.605	0.656	0.702	1.000	1.424	1.526	1.653	1.747
50	0.578	0.610	0.660	0.706	1.000	1.416	1.515	1.639	1.730
55	0.591	0.623	0.672	0.716	1.000	1.396	1.489	1.606	1.691
60	0.603	0.634	0.682	0.725	1.000	1.378	1.466	1.577	1.657
65	0.619	0.649	0.695	0.737	1.000	1.356	1.439	1.542	1.617
70	0.629	0.658	0.703	0.745	1.000	1.343	1.422	1.520	1.591
75	0.638	0.667	0.711	0.751	1.000	1.331	1.407	1.501	1.569
80	0.650	0.678	0.721	0.760	1.000	1.316	1.388	1.477	1.541
85	0.657	0.685	0.727	0.765	1.000	1.307	1.376	1.462	1.524
90	0.664	0.691	0.733	0.771	1.000	1.298	1.365	1.448	1.508
95	0.674	0.700	0.741	0.777	1.000	1.287	1.351	1.430	1.487
100	0.680	0.706	0.746	0.782	1.000	1.280	1.342	1.419	1.474
110	0.693	0.718	0.757	0.792	1.000	1.264	1.322	1.394	1.446
120	0.702	0.727	0.765	0.798	1.000	1.253	1.309	1.377	1.425
130	0.713	0.737	0.773	0.806	1.000	1.241	1.294	1.358	1.404
140	0.723	0.746	0.781	0.813	1.000	1.231	1.281	1.342	1.386
150	0.730	0.752	0.787	0.818	1.000	1.223	1.272	1.331	1.372

2.3 $\xi(\boldsymbol{X}, \gamma)$ As an Increasing Function of γ

The method of finding confident interval for γ using a pivotal quantity is possible only if the pivotal quantity $\xi(\boldsymbol{X}, \gamma)$ is a strictly monotonic function of γ. It can be shown that the pivotal quantity function

$$\xi(\boldsymbol{X},\gamma)=\frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln(x_{(i)}-\gamma)}{n-2[n/3]}-\dfrac{\sum_{i=1}^{[n/3]}\ln(x_{(i)}-\gamma)}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n}\ln(x_{(i)}-\gamma)}{[n/3]}-\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln(x_{(i)}-\gamma)}{n-2[n/3]}}$$

is a strictly increasing function of γ when we fix the sample $\boldsymbol{X}$.

Pivotal quantity ξ increase infinitely when γ approaches from the left to the first order statistic $X_{(1)}$. In fact, the numerator of $\xi(\boldsymbol{X},\gamma)$ contains a term that has the form $-[\ln(X_{(1)}-\gamma)]$ then $\lim_{\gamma\to X_{(1)}^{-}}\xi(X,\gamma)=\infty$.

On the other hand, the pivotal quantity $\xi(\boldsymbol{X},\gamma)$ reaches its minimum value when γ approaches from the right to its minimum value 0 $\left(0<\gamma<X_{(1)}\right)$.

$$\xi_{\min}=\lim_{\gamma\to 0^{+}}\xi(X,\gamma)=\lim_{\gamma\to 0^{+}}\frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln(x_{(i)}-\gamma)}{n-2[n/3]}-\dfrac{\sum_{i=1}^{[n/3]}\ln(x_{(i)}-\gamma)}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n}\ln(x_{(i)}-\gamma)}{[n/3]}-\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln(x_{(i)}-\gamma)}{n-2[n/3]}}.$$

$$\to\xi_{\min}=\frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln x_{(i)}}{n-2[n/3]}-\dfrac{\sum_{i=1}^{[n/3]}\ln x_{(i)}}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n}\ln x_{(i)}}{[n/3]}-\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln x_{(i)}}{n-2[n/3]}} \tag{2.3.1}$$

Since $\xi(\boldsymbol{X},\gamma)$ is independent of all parameters, we can select the fix sample $\boldsymbol{X}$ from a lognormal distributed population with any combination of parameters μ , σ and γ. A sample of 30 observations was generated from a lognormal distribution with parameters $\mu = 4,\ \sigma = 2,$ and $\gamma = 100$, and its data are listed below.

Table 2.3.1 Sample from a Lognormal Distribution $(\mu = 4,\ \sigma = 2,$ and $\gamma = 100)$.

100.207	116.143	153.838	239.768	566.972
102.256	118.922	162.861	287.369	629.204
103.724	120.047	164.751	326.786	643.492
105.271	128.778	165.674	425.179	692.266
105.358	136.122	166.460	548.583	1063.562
110.966	144.123	201.961	551.102	1250.333

The first order statistic of the sample is $X_{(1)} = 100.207$ and the 30th order statistic is $X_{(30)} = 1250.333$. The minimum value for $\xi(\boldsymbol{X},\gamma)$ is

$$\xi_{\min} = \frac{\dfrac{\sum_{i=11}^{20} \ln x_{(i)}}{10} - \dfrac{\sum_{i=1}^{10} \ln x_{(i)}}{10}}{\dfrac{\sum_{i=21}^{30} \ln x_{(i)}}{10} - \dfrac{\sum_{i=11}^{20} \ln x_{(i)}}{10}} = 0.3751$$

We can graph the pivotal quantity $\xi(\boldsymbol{X},\gamma)$ versus γ using the data of this example. The results are presented in Table 2.3.2 and in Figure 2.3.

Table 2.3.2 Values of ξ for a fix sample when γ increase from 0 to $X_{(1)}$.

γ	ξ
0	0.375
3	0.377
6	0.382
9	0.387
12	0.392
15	0.397
18	0.402
21	0.408
24	0.414
27	0.420
30	0.427
33	0.434

γ	ξ
36	0.443
39	0.451
42	0.460
45	0.469
48	0.479
51	0.489
54	0.500
57	0.512
60	0.526
63	0.540
66	0.556
69	0.574

γ	ξ
72	0.593
75	0.616
78	0.641
81	0.670
84	0.704
87	0.744
90	0.795
93	0.860
96	0.951
99	1.112
100	1.246
100.1	1.289

γ	ξ
100.15000	1.324
100.20000	1.436
100.20300	1.467
100.20600	1.555
100.20660	1.697
100.20663	1.767
100.206639	1.885
100.2066399	2.002
100.20663999	2.120

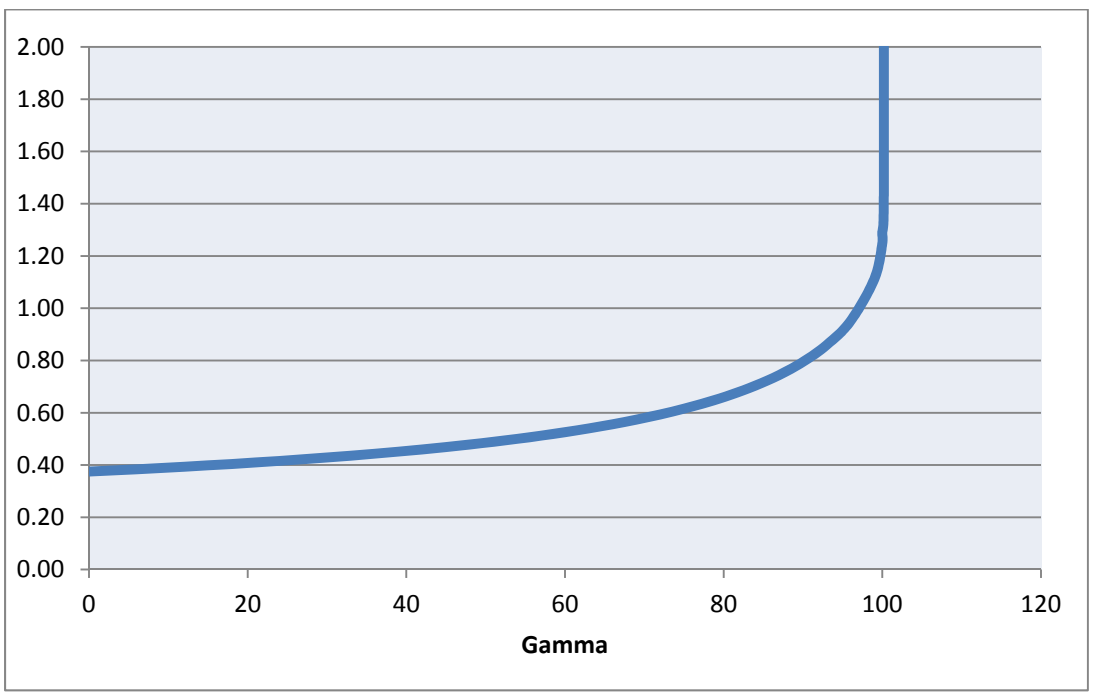


Figure 2.3 Graphic of ξ as a strictly increasing function of γ.

We can find through Table 2.3.1 and Figure 2.3 that $\xi(\boldsymbol{X},\gamma)$ is a strictly increasing function of γ. Additionally we can verify the tendency of $\xi(\boldsymbol{X},\gamma)$ when γ approaches to its boundaries 0 and $X_{(1)} = 100.20664$

$$\lim_{\gamma \to 0^+} \xi(X,\gamma) = \xi_{\min} = 0.3751 \quad \text{and} \quad \lim_{\gamma \to X_{(1)}^-} \xi(X,\gamma) = \infty \quad .$$

When γ is very close to the first order statistic $X_{(1)}$, the graph of the pivotal quantity $\xi(\boldsymbol{X},\gamma)$ is almost a vertical line.

Since pivotal quantity $\xi(\boldsymbol{X},\gamma)$ is a strictly increasing function of γ, we can construct the confident interval for γ .

2.4 Confidence Interval for the Parameter γ

For a given sample size n and confidence level $1-\alpha$ we can find corresponding critical values $\xi_{\alpha/2}$ and $\xi_{1-\alpha/2}$.

$$P(\xi_{\alpha/2} < \xi(\boldsymbol{X},\gamma) < \xi_{1-\alpha/2}) = 1-\alpha .$$

We are $(1-\alpha)100\%$ confident that the true value of $\xi(\boldsymbol{X},\gamma)$ is at least $\xi_{\alpha/2}$ but not greater than $\xi_{1-\alpha/2}$. Since $\xi(\boldsymbol{X},\gamma)$ is a strictly increasing function of γ, we can obtain a $(1-\alpha)100\%$ confident interval for γ,which has the form

$$L(\boldsymbol{X},\xi_{\alpha/2}) < \gamma < U(\boldsymbol{X},\xi_{1-\alpha/2}).$$

Here $L(\boldsymbol{X},\xi_{\alpha/2})=\gamma_L$ and $U(\boldsymbol{X},\xi_{1-\alpha/2})=\gamma_U$ are the lower and upper confident limits for γ . In fact, γ_L and γ_U are the solutions of the equations

$$\frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln(x_{(i)}-\gamma)}{n-2[n/3]}-\dfrac{\sum_{i=1}^{[n/3]}\ln(x_{(i)}-\gamma)}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n}\ln(x_{(i)}-\gamma)}{[n/3]}-\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln(x_{(i)}-\gamma)}{n-2[n/3]}}=\xi_{\alpha/2} \tag{2.4.1}$$

and

$$\frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln(x_{(i)}-\gamma)}{n-2[n/3]}-\dfrac{\sum_{i=1}^{[n/3]}\ln(x_{(i)}-\gamma)}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n}\ln(x_{(i)}-\gamma)}{[n/3]}-\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]}\ln(x_{(i)}-\gamma)}{n-2[n/3]}}=\xi_{1-\alpha/2} \tag{2.4.2}$$

Equations (2.4.1) and (2.4.2) can be solved to find γ_L and γ_U. Both, the lower and upper confident limits of γ must be smaller than $X_{(1)}$, since the first order statistic of the random variable X is greater than the threshold parameter γ.

Now, we can calculate the confidence interval of γ for the random sample presented in Section 2.3 using the Equations 2.4.1 and 2.4.2. To solve these equations, the bisection method and SAS/IML was used. We can solve the Equations 2.4.1 and 2.4.2 because $\xi_{\min}$ is smaller than the critical values of ξ.

$$\xi_{\min} = 0.3751 < 0.577 = \xi_{0.025} \quad \text{and} \quad \xi_{\min} = 0.3751 < 1.735 = \xi_{0.975}.$$

The 95% confident interval for γ, obtained using SAS/IML and the bisection method, is

(69.5083 , 100.2066).

If $\xi_{\min}$ (2.3.1) is greater than $\xi_{\alpha/2}$ or $\xi_{1-\alpha/2}$, then Equations (2.4.1) or (2.4.2) cannot be solved, and two-side confidence interval for γ cannot be constructed in those cases.

2.5 Upper Confidence Limit for the Parameter γ

It was mentioned above that if $\xi_{\min}$ is greater than $\xi_{\alpha/2}$ or $\xi_{1-\alpha/2}$, then Equations (2.4.1) or (2.4.2) have no solution in the interval $[0, X_{(1)})$. Monte Carlo simulation and SAS/IML were used to estimate the percentage of samples for which we cannot calculate the upper and lower confidence limits. For several sample sizes n, 1,000,000 samples were generated from a lognormal distribution with parameters ($\mu = 2$, $\sigma = 4$, and $\gamma = 10$).

It was found that the lower confidence limit cannot be calculated for more than 1% of the samples when n is smaller than 30; in contrast, the upper confidence limit can be calculated for more than 99% $(1 - 0.99\%)$ of the sample when n is greater than 7. The results were tabulated in Table 2.5.1.

Since the two-side confidence cannot always be constructed, a one-side upper confidence limit of γ is recommended. In fact, the proposed method is good for finding upper confidence limits of γ, not for lower confidence limits.

Table 2.5.1 Percentage of Samples from a Lognormal Distribution for which the Confidence Limits cannot be calculated.

N	Replications	μ	σ	γ	%No Low	%No Up
50	1000000	2	4	10	0.11%	0.00%
40	1000000	2	4	10	0.34%	0.00%
33	1000000	2	4	10	0.77%	0.00%
30	1000000	2	4	10	1.08%	0.00%
20	1000000	2	4	10	2.83%	0.00%
15	1000000	2	4	10	6.05%	0.03%
10	1000000	2	4	10	10.89%	0.20%
8	1000000	2	4	10	14.00%	0.39%
7	1000000	2	4	10	17.10%	0.99%
6	1000000	2	4	10	21.42%	2.15%
5	1000000	2	4	10	24.52%	2.56%

2.6 Point Estimator of the Parameter γ

The point estimator of the threshold or location parameter γ can be obtained by squeezing the confidence interval described in the Section 2.4. A significance level $\alpha = 100\%$ produces 0% confidence interval which is a point estimator of the median of γ because $\gamma_{\alpha/2} = \gamma_{1-\alpha/2} = \gamma_{0.50}$.

The pivotal quantity $\xi(\boldsymbol{X}, \gamma)$ can be calculated as a function of γ using (2.1.1) for each sample. The critical value $\xi_{0.50}$ corresponding to the sample size n can be found in Table 2.3.1. Then, the threshold parameter γ can be estimated by solving the Equation

$$\frac{\dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} \ln(x_{(i)} - \hat{\gamma})}{n - 2[n/3]} - \dfrac{\sum_{i=1}^{[n/3]} \ln(x_{(i)} - \hat{\gamma})}{[n/3]}}{\dfrac{\sum_{i=n-[n/3]+1}^{n} \ln(x_{(i)} - \hat{\gamma})}{[n/3]} - \dfrac{\sum_{i=[n/3]+1}^{n-[n/3]} \ln(x_{(i)} - \hat{\gamma})}{n - 2[n/3]}} = \xi_{0.50} \qquad (2.6.1)$$

The proposed method can be illustrated using the example considered in a previous study by Cohen and Whitten (1980). The data set consists of 20 observations from a population with $\mu = \ln(50)$, $\sigma = 0.40$, and $\gamma = 100$. The sample data are listed in the Table 2.6.1.

Table 2.6.1 Random Sample from a Lognormal Distribution ($\mu = \ln(50)$, $\sigma = 0.40$, and $\gamma = 100$).

148.290	144.328	174.800	1687.554
184.101	166.475	131.375	145.788
135.880	137.338	164.304	155.369
127.211	132.971	128.709	201.415
133.143	155.680	153.070	157.238

Summary statistics of this data set are: $\bar{x} = 152.302$, $s = 19.954$, median=150.68, $x_{(1)} = 127.2112$, $x_{(20)} = 201.4125$, $\xi_{\min} = 0.781$. The critical values of the pivotal quantity when n=20 are $\xi_{0.025} = 0.515$, $\xi_{0.50} = 1.00$, and $\xi_{0.975} = 1.943$.

Using SAS/IML and the bisection method, Equations 2.4.1, 2.4.2 and 2.6.1 can be solved. In this case, the lower confidence limit cannot be found because Equation 2.4.1 has no solution in the interval [0, 127.211). The upper confidence limit of γ is 125.427 and the point estimate of γ is $\hat{\gamma} = 94.758$.

Table 2.6.2 summarizes the point estimates obtained for 14 different methods mentioned by Cohen and Witten in their article. The estimate, obtained by the proposed method, is among the three best estimates.

Table 2.6.2 Summary of Estimates of 14 different methods

Method	Estimator $\hat{\gamma}$	Bias of $\hat{\gamma}$
Maximum Likelihood Estimator	127.2112	27.2112
Local Maximum Likelihood Estimator	117.7206	17.7206
Modified Maximum Likelihood Estimator I_1	107.0445	7.0445
Modified Maximum Likelihood Estimator I_2	19.1641	-80.8359
Modified Maximum Likelihood Estimator I_3	126.477	26.4777
Modified Maximum Likelihood Estimator II_1	118.0314	18.0314
Modified Maximum Likelihood Estimator II_2	86.8243	-13.1757
Modified Maximum Likelihood Estimator II_3	125.9937	25.9937
Moment Estimator	70.70367	-29.2933
Modified Moment Estimator I_1	105.1667	5.1667
Modified Moment Estimator I_2	75.9154	-24.0846
Modified Moment Estimator I_3	-0.7037	-100.7037
Modified Moment Estimator II_1	112.5581	12.5581
Modified Moment Estimator II_2	96.0955	-3.9045
Modified Moment Estimator II_3	82.2976	-17.7024
Pivotal Quantity Method (Proposed)	94.7580	-5.2420
Population	**100**	

3. SIMULATION RESULTS AND DISCUSSION

The performance of the proposed method to estimate the threshold parameter γ is evaluated using the Mean Square Error (MSE) criteria. Monte Carlo simulation was used to generate *r* samples from a three-parameter lognormal distribution and $\hat{\gamma}$ was estimated for each sample using the procedure described above.

If $\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_k$ are point estimates of γ obtained from the *k* samples, the estimates of the mean, the variance, and the MSE of the random variable $\hat{\gamma}$ are

$$\hat{\mu}_{\hat{\gamma}} = \frac{\sum_{i=1}^{k} \hat{\gamma}_i}{k}, \tag{3.1}$$

$$\hat{\sigma}^2_{\hat{\gamma}} = \frac{\sum_{i=1}^{k} \left(\hat{\gamma}_i - \hat{\mu}_{\hat{\gamma}}\right)^2}{k-1}, \tag{3.2}$$

$$Bias_{\gamma}\hat{\gamma} = \mu_{\hat{\gamma}} - \gamma \tag{3.3}$$

and

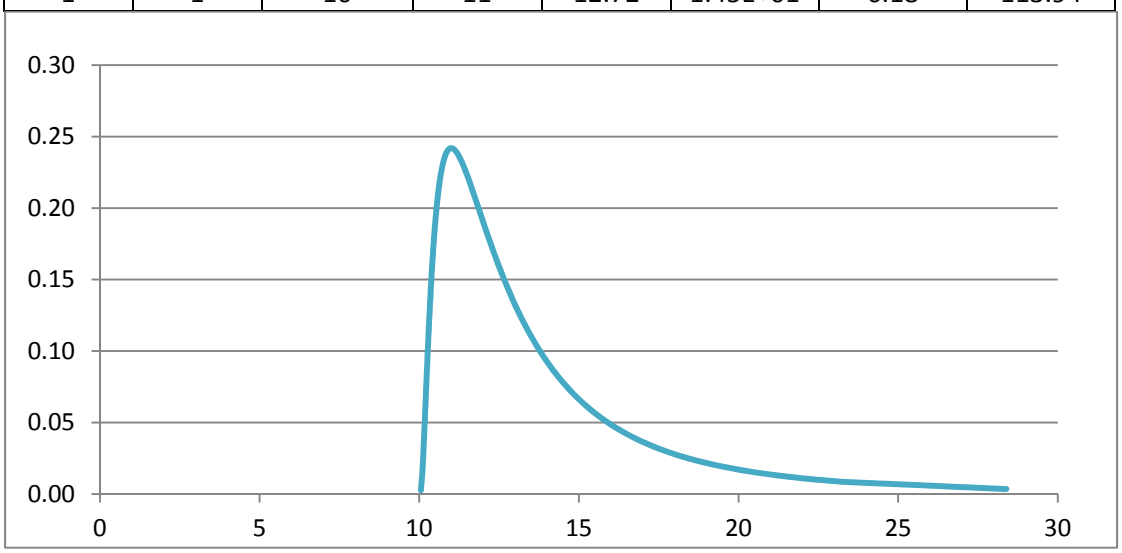
$$MSE = \frac{\sum_{I=1}^{k} \left(\hat{\gamma}_i - \gamma\right)^2}{k} = \hat{\sigma}^2_{\hat{\gamma}} + (Bias_{\gamma}\hat{\gamma})^2 \tag{3.4}$$

where γ is the true value of the threshold parameter and *k* is the number of samples with size *n* taken from the population. The bias of the point estimator $\hat{\gamma}$ is the difference between the expected value of $\hat{\gamma}$ and the parameter γ.

Statistical simulation was conducted to investigate the effect of different sample sizes on the performance of the method proposed in this research, and it was compared with the performance of the MLE method. It was mentioned, in Section 1.6, that the MLE estimator of γ is the first order statistic $x_{(1)}$. Monte Carlo simulation and SAS/IML were used to generate 1,000,000 samples for each sample size n considered from a population with parameters $\mu = 1,\ \sigma = 1,$ and $\gamma = 10$. The pdf of this population is presented in Figure 3.1.

Figure 3.1 Pdf of the Population with Parameters $\mu = 1,\ \sigma = 1,$ and $\gamma = 10$.

mu	sig	gam	**mode**	**median**	**mean**	Skewness	Kurtosis
1	1	10	11	12.72	1.45E+01	6.18	113.94



The mean, the variance, the MSE, and the bias for $\hat{\gamma}$ were calculated for several sample sizes n using both methods. The proposed method has smaller bias, even for small samples. When n is 200 or bigger, the MSE of the proposed method is better than MLE method. Table 3.1 shows the results.

Table 3.1 Effect of *n* on the Performance of MLE Method and Proposed Method

		PARAMETERS			MLE			PROPOSED		
N	***K***	μ	σ	γ	$\hat{\mu}_{\hat{\gamma}}$	BIAS $(\hat{\gamma})$	MSE $(\hat{\gamma})$	$\hat{\mu}_{\hat{\gamma}}$	BIAS $(\hat{\gamma})$	MSE $(\hat{\gamma})$
500	1000000	1	1	10	10.14	0.14	0.02	9.98	-0.01	0.01
200	1000000	1	1	10	10.19	0.19	0.04	9.96	-0.04	0.04
100	1000000	1	1	10	10.24	0.24	0.07	9.93	-0.07	0.09
50	1000000	1	1	10	10.32	0.32	0.12	9.87	-0.13	0.30
40	1000000	1	1	10	10.35	0.35	0.14	9.82	-0.18	0.48
30	1000000	1	1	10	10.39	0.39	0.19	9.75	-0.25	0.90
20	1000000	1	1	10	10.48	0.48	0.28	9.66	-0.34	1.56
15	1000000	1	1	10	10.55	0.55	0.39	9.56	-0.44	2.44
10	1000000	1	1	10	10.69	0.69	0.62	9.53	-0.47	3.22
7	1000000	1	1	10	10.84	0.84	0.98	9.60	-0.40	3.81
5	1000000	1	1	10	11.05	1.05	1.60	9.79	-0.21	4.24

When the shape parameter σ increases, the skewness and the kurtosis of the distribution also increase. The skewness measures the lack of symmetry of the pdf and the kurtosis measures the peakedness and heaviness of the tail of the pdf. The proposed method performs better than MLE method when the data shows low skew and no heavy tail; this occurs when the pdf looks more like a normal distribution. However, the MLE method performs better when the data shows high skewness and heavy tail or the pdf looks like an L-shape. Table 3.2 shows the effect of σ (shape parameter) on the performance of the two methods.

When scale parameter μ increases, the dispersion of three-parameter lognormal random variable increases. The proposed method performs better than MLE when the data is more dispersed. When $\mu = 8$, the MSE of MLE method is 12,439,494 and the

estimate of γ is 3,210, which is 321 times bigger than the real parameter. Table 3.3 shows the results.

Table 3.2 Effect of the Shape Parameter σ on the Performance of the MLE and Proposed Method.

		PARAMETERS			MLE			PROPOSED		
N	***k***	μ	σ	γ	$\mu_{\hat{\gamma}}$	BIAS $(\hat{\gamma})$	MSE $(\hat{\gamma})$	$\hat{\mu}_{\hat{\gamma}}$	BIAS $(\hat{\gamma})$	MSE $(\hat{\gamma})$
30	1000000	1	0.1	10	12.22	2.22	4.93	10.62	0.62	4.69
30	1000000	1	0.5	10	11.01	1.01	1.07	9.51	-0.49	3.33
30	1000000	1	1	10	10.39	0.39	0.19	9.75	-0.25	0.90
30	1000000	1	2	10	10.07	0.07	0.01	9.90	-0.10	0.12
30	1000000	1	4	10	10.00	0.00	0.00	9.97	-0.03	0.02
30	1000000	1	8	10	10.00	0.00	0.00	9.99	-0.01	0.01
30	1000000	1	10	10	10.00	0.00	0.00	10.00	0.00	0.01

Table 3.3 Effect of the Scale Parameter μ on the Performance of the Two Methods.

		PARAMETERS			MLE			PROPOSED		
n	***k***	μ	σ	γ	$\hat{\mu}_{\hat{\gamma}}$	BIAS $(\hat{\gamma})$	MSE $(\hat{\gamma})$	$\hat{\mu}_{\hat{\gamma}}$	BIAS $(\hat{\gamma})$	MSE $(\hat{\gamma})$
30	1000000	0	1	10	10.15	0.15	0.03	9.90	-0.10	0.19
30	1000000	0.5	1	10	10.24	0.24	0.07	9.83	-0.17	0.42
30	1000000	1	1	10	10.39	0.39	0.19	9.75	-0.25	0.90
30	1000000	1.5	1	10	10.65	0.65	0.51	9.63	-0.37	1.88
30	1000000	2	1	10	11.07	1.07	1.40	9.50	-0.50	5.18
30	1000000	4	1	10	17.92	7.92	76.31	11.99	1.99	36.29
30	1000000	6	1	10	442.99	432.99	227765.42	45.10	35.10	2076.18
30	1000000	8	1	10	3209.83	3199.83	12439494.6	296.73	286.73	123341.6

Variations in the threshold parameter γ do not significantly affect the performance of both methods. Table 3.4 shows the results. The MSE of the proposed method could be improved if we solve Equation (2.6.1) for $\hat{\gamma}$ starting the bisection method at a value closer to the data. The samples of Table 3.4 were selected from a population with

parameters $\mu = 1$ and $\sigma = 1$, and 99% of the values of this population are in the interval ($e^{-1.5758} + \gamma$, $e^{3.5758} + \gamma$). When $\gamma = 1000$, 99% of the values of this population are in the interval (1000.21, 1035.72). It does not make sense to start the bisection method from 0. When 1,000,000 samples with n=30 are generated, we expect about 150,000 observations less than 1000.21 and some of these samples could have smaller skewness than the population. For this kind of sample, the bisection method obtains estimates very far from 1,000. Only one of these estimates increases the mean square error significantly.

Table3.4 Effect of the threshold parameter γ on the performance of the two methods.

		PARAMETERS			MLE			PROPOSED		
n	***k***	μ	σ	γ	$\hat{\mu}_{\hat{\gamma}}$	BIAS ($\hat{\gamma}$)	MSE ($\hat{\gamma}$)	$\hat{\mu}_{\hat{\gamma}}$	BIAS ($\hat{\gamma}$)	MSE ($\hat{\gamma}$)
30	1000000	1	1	0	0.39	0.39	0.19	0.27	0.27	0.10
30	1000000	1	1	10	10.39	0.39	0.19	9.75	-0.25	0.90
30	1000000	1	1	20	20.39	0.39	0.19	19.72	-0.28	1.22
30	1000000	1	1	100	100.39	0.39	0.19	99.70	-0.30	2.48
30	1000000	1	1	200	200.39	0.39	0.19	199.69	-0.31	3.64
30	1000000	1	1	1000	1000.39	0.39	0.19	999.67	-0.33	10.25

Table 3.5 Performance of the modified method.

n	***K***	μ	σ	γ	$\hat{\mu}_{\hat{\gamma}}$	BIAS ($\hat{\gamma}$)	MSE ($\hat{\gamma}$)	**% NO** $\hat{\gamma}$	**% NO** $\hat{\gamma}_U$	Start Bisection
30	1000000	1	1	1000	999.72	-0.28	1.42	0.001	0.000	$x_{(1)} - 3(range)$
30	1000000	1	1	1000	999.72	-0.28	1.25	0.002	0.000	$x_{(1)} - 2(range)$
30	1000000	1	1	1000	999.74	-0.25	0.94	0.001	0.000	$x_{(1)} - 1(range)$

The proposed method could be improved if we start the bisection method at a point equivalent to $x_{(1)} -$ (three, two or one time the data range). Simulations were performed with this modification and the MSE was reduced significantly. If we start bisection method for each sample at $x_{(1)} - 1$(1000.72 - 1035.21), the MSE is 0.94 and it is possible to estimate γ for almost all random samples as you can see in the column %NO $\hat{\gamma}$ in Table 3.5.

4. CONCLUSION

The three-parameter lognormal distribution is widely used in many areas of science. Some statisticians have manifested their preference for this distribution when describing original data. Finding a good estimator and constructing a confidence interval of this distribution is a difficult problem. Some modifications have been proposed to improve the maximum likelihood estimator. In some cases, however, the modified maximum likelihood estimates do not exist or the procedure encounters multiple estimates. The three-parameter lognormal pdf's shape varies from almost symmetric, as an upturned T, to highly skewed, as an L. The purpose of this research is focused on estimating the threshold parameter γ, because when $\hat{\gamma}$ is known, then the estimation of the other two parameters can be obtained from the first two maximum likelihood estimation equations. Consequently, better estimate of γ leads to a better estimate of the other two parameters.

In this research a method for constructing confidence intervals, confidence limits, and point estimators for the threshold or location parameter γ is proposed. In order to find a $1-\alpha$ confident interval for the parameter γ, a pivotal quantity, $\xi(\boldsymbol{X},\gamma)$, was defined. Two important properties of $\xi(\boldsymbol{X},\gamma)$ were shown. The first one is that the probability distribution of $\xi(\boldsymbol{X},\gamma)$ is independent of all parameters, which allows us to construct a table for critical values of the pivotal quantity. The second one is that the pivotal quantity function $\xi(\gamma)$ is a strictly increasing function of γ, which allows us to construct confident intervals for γ. Monte-Carlo simulation was used to find the critical values $\xi_{\alpha/2}$ and $\xi_{1-\alpha/2}$. Equations $\xi(\boldsymbol{X},\gamma)=\xi_{\alpha/2}$ and $\xi(\boldsymbol{X},\gamma)=\xi_{1-\alpha/2}$ were solved using

the bisection method in order to find the confidence interval of γ. The proposed method is good for finding upper confidence limits of γ, not for lower confidence limits. The point estimator of the threshold or location parameter γ can be obtained by squeezing the confidence interval and using the critical value $\xi_{0.50}$.

The bias of the point estimator and mean square error (MSE) criteria were used throughout extensive simulation to evaluate the performance of the proposed method. Compared with other methods, the pivotal quantity method performs quite well obtaining point estimators of γ with lower bias than most of them, even in cases where other methods confront difficulties such as small samples and population with low skewness. In fact, the proposed method produces a lower bias of $\hat{\gamma}$ than the maximum likelihood estimator (MLE) method does in all simulations summarized in Table 3.1 through 3.5, where a wide combination of the three parameters was considered. Using the MSE criteria, the proposed method shows better performance than the MLE method when the population skewness is low and the spread of the distribution is high. The MSE of the proposed method can be improved if we do not consider the few atypical samples, which are much less than 1% of the samples, for which the bisection method produces a big bias. Moreover, the pivotal quantity could be slightly modified in order to get better MSE.

The method introduced in this research provides a more accurate estimate of the location parameter in most of the original data, especially when the population skewness is low and the spread of the distribution is high.

LIST OF REFERENCES

Casella, G., and R. L. Berger, Statistical Inference, (2002).

Cohen, A Clifford, Estimating Parameters of Logarithmic Normal Distributions by Maximum likelihood, Journal of the American Statistical Association, 46, (1951), 206-212.

Cohen, A.C., B.J. Whitten, Estimation in the Three-Parameter Lognormal Distribution, Journal of the American Statistical Society, 75 (1980) 399-404.

Cohen, A.C., B.J. Whitten and Y. Ding, Modified Moment Estimation for the Three Parameter Lognormal Distribution, Journal of Quality Technology, 17 (1985) 92-99.

Harter, H.L. and A.L. Moore, Local-Maximum-Likelihood Estimation of the Parameter of Three-Parameter Lognormal Population From Complete and Censored Samples, Journal of the American Statistical Society, 61 (1966) 842-851.

Hill, M.B., The Three-Parameter Log-Normal Distribution and Bayesian Analysis of a Point-Source Epidemic, Journal of the American Statistical Association, 58, (1963) 72-84.

Limpert, E., W. Stahel and M. Abbt, Log-Normal Distribution Acroos the Sciences: Keys and Clues, BioScience, 51 (5), (2001) 341-352.

Munro A.H. and R.A.J. Wixley, Estimation on Order Statistics of Small Samples From a Three-Parameter Lognormal Distribution, Journal of the American Statistical Society, 65 (1970) 212-225.

Wingo, D.R., Fitting Three-Parameter Log-Normal Models by Numerical Global Optimization – an Improved Algorithm, Computation Statistical Data Analysis, 2 (1984) 13-25.

Yuan, Pae-Tsi, On the Logarithmic Frequency Distribution and Semi-Logarithmic Correlation Surface, Annals of Mathematical Statistics, 4, (1933), 30-74.