Comparison of Parameters of Lognormal Distribution Based On the Classical and Posterior Estimates

Raja Sultan  
*University of Kashmir, Srinagar, India*, hamzasultan18@yahoo.com

S. P. Ahmad  
*University of Kashmir, Srinagar, India*, sprvz@yahoo.com

Follow this and additional works at: [http://digitalcommons.wayne.edu/jmasm](http://digitalcommons.wayne.edu/jmasm)

Part of the [Applied Statistics Commons](http://digitalcommons.wayne.edu/jmasm), [Social and Behavioral Sciences Commons](http://digitalcommons.wayne.edu/jmasm), and the [Statistical Theory Commons](http://digitalcommons.wayne.edu/jmasm)

Recommended Citation  
DOI: 10.22237/jmasm/1383279420  
Available at: [http://digitalcommons.wayne.edu/jmasm/vol12/iss2/18](http://digitalcommons.wayne.edu/jmasm/vol12/iss2/18)
Comparison of Parameters of Lognormal Distribution Based On the Classical & Posterior Estimates

Raja Sultan  
University of Kashmir  
Srinagar, India

S. P. Ahmad  
University of Kashmir  
Srinagar, India

Lognormal distribution is widely used in scientific field, such as agricultural, entomological, biology etc. If a variable can be thought as the multiplicative product of some positive independent random variables, then it could be modelled as lognormal. In this study, maximum likelihood estimates and posterior estimates of the parameters of lognormal distribution are obtained and using these estimates we calculate the point estimates of mean and variance for making comparisons.

Keywords: Lognormal distribution, maximum likelihood estimation, posterior estimates & R software

Introduction

Aitchison & Brown (1957) have given a very comprehensive treatment of lognormal distribution. The lognormal distribution arises in various different contexts such as in physics (distribution of particles due to pulverisation); economics (income distributions); biology (growth of organisms), etc. Epstein (1947), Brownlee (1949), Delaporte (1950), Moroney (1951) describes various applications of lognormal distribution to physical and industrial processes, textile research and quality control. In the context of life testing and reliability problems, the lognormal distribution answers a criticism sometimes raised against the use of normal distribution (ranging from $-\infty$ to $+\infty$) as a model for the failure time distribution which must range from 0 to $\infty$.

A random variable $X$ is said to have a lognormal distribution if $U = \log_e X$ has normal distribution with mean $\mu$ and variance $\sigma^2$. Thus, the pdf of lognormal distribution is given by
The likelihood function of the random sample \((x_1, x_2, x_3, ..., x_n)^T\) would be

\[
L(\mu, \sigma^2 | x) = \left(\frac{1}{\sqrt{2\pi} \sigma x}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (\log x_i - \mu)^2\right)
\]  

(2)

The mean and variance of the lognormal distribution are given by

\[
E(X) = \alpha_i = \exp\left(\mu + \frac{\sigma^2}{2}\right)
\]  

(3)

and

\[
V(X) = \beta_i = \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2) - 1\right)
\]  

(4)

**Maximum Likelihood Estimators**

Maximum Likelihood is a popular estimation technique for many distributions because it picks the values of the distribution's parameters that make the data "more likely" than any other values of the parameters would make them. This is accomplished by maximizing the likelihood function of the parameters given the data.

Consider the estimation of the parameters \(\alpha_1\) and \(\beta_1\). Let

\[
U_i = \log x_i, \; i = 1, 2, ..., n
\]

Then using the fact that \((U_1, U_2, ..., U_n)\) is a random sample from Normal distribution with parameters \((\mu, \sigma^2)\). The mle of \(\mu\) and \(\sigma^2\) first are given by

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} U_i = \bar{U}
\]  

(5)

and
COMPARISON OF PARAMETERS OF LOGNORMAL DISTRIBUTION

\[ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (U_i - \bar{U})^2 \]  

(6)

The mle of \( \alpha_1 \) and \( \beta_1 \) are given by

\[ \hat{\alpha}_i = \exp \left( \hat{\mu} + \frac{\hat{\sigma}^2}{2} \right) \]  

(7)

and

\[ \hat{\beta}_i = \exp \left( 2\hat{\mu} + \hat{\sigma}^2 \right) \left( \exp(\hat{\sigma}^2) - 1 \right) \]  

(8)

Posterior estimation of the parameter

Again, consider the estimation of the parameters \( \alpha_1 \) and \( \beta_1 \). First obtain the posterior estimates of \( \mu \) and \( \sigma^2 \) and then simultaneously the posterior estimates for \( \alpha_1 \) and \( \beta_1 \) will be obtained. Laplace (1774) found that it worked exceptionally well to simply always choose the prior probability distribution for the parameter(s) of the model to be constant on the parameter space.

The joint prior pdf for \( \mu \) and \( \sigma^2 \) considered is

\[ P(\mu, \sigma^2) \propto 1 \]  

(9)

According to Bayes theorem, Joint posterior density of \( \mu \) and \( \sigma^2 \) is given by

\[ \pi(\mu, \sigma^2 | x) \propto P(\mu, \sigma^2).P(\mu, \sigma^2 | x) \]

From equation (2) and (9) the joint posterior density of \( \mu \) and \( \sigma^2 \) is given by

\[ \pi(\mu, \sigma^2 | x) \propto \left[ \frac{1}{\sqrt{2\pi\sigma x}} \right]^n \exp \left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (\log x_i - \mu)^2 \right\} \]
\[
\pi(\mu, \sigma^2 \mid x) = \frac{c}{(\sigma^2)} \exp\left(\frac{-\beta}{2\sigma^2}\right) \exp\left(-\frac{n}{2\sigma^2} \left(\mu - \frac{\sum_{i=1}^{n} \log x_i}{n}\right)^2\right)
\]

(10)

where \( \beta = \sum_{i=1}^{n} (\log x_i)^2 - \frac{\left(\sum_{i=1}^{n} \log x_i\right)^2}{n} \) and \( c \) is a normalizing constant. Lindley (1961) explained if \( P(\theta) \) be the prior and \( P(x \mid \theta) \) be the likelihood, the posterior pdf \( P(\theta \mid x) \) is given by \( P(\theta \mid x) = c P(\theta). P(x \mid \theta) \), where \( c \) is the normalizing constant. Then the value of \( c \) is obtained by

\[
c = \left[ \int P(\theta). P(x \mid \theta) d\theta \right]^{-1}
\]

Therefore, \( c \) can be obtained by

\[
c^{-1} = \int_{0}^{\infty} \int_{-\infty}^{\infty} \pi(\mu, \sigma^2 \mid x) d\mu d\sigma^2
\]

Using the transformation

\[
t = \frac{\sqrt{n}}{\sigma} \left(\mu - \frac{\sum_{i=1}^{n} \log x_i}{n}\right)
\]

307
COMPARISON OF PARAMETERS OF LOGNORMAL DISTRIBUTION

\[ c^{-1} = \sqrt{\frac{2\pi}{n}} \int_0^{\infty} \frac{\exp\left(\frac{-\beta}{2\sigma^2}\right)}{(\sigma^2)^{n+1/2}} d\sigma^2 \]

\[ c^{-1} = \sqrt{\frac{2\pi}{n}} \frac{\Gamma\left(\frac{n-3}{2}\right)}{\left(\frac{\beta}{2}\right)^{n-3/2}} \]

\[ \Rightarrow c = \sqrt{\frac{n}{2\pi}} \frac{\left(\frac{\beta}{2}\right)^{n-3/2}}{\Gamma\left(\frac{n-3}{2}\right)} \] (11)

From the equation (10)

\[ \pi(\mu, \sigma^2 | x) = \sqrt{\frac{n}{2\pi}} \frac{\left(\frac{\beta}{2}\right)^{n-3/2}}{\Gamma\left(\frac{n-3}{2}\right)} \exp\left(\frac{-\beta}{2\sigma^2}\right) \exp\left(\frac{-n}{2\sigma^2} \left(\mu - \frac{\sum_{i=1}^{n} \log x_i}{n}\right)^2\right) \] (12)

**Marginal posterior densities of \( \mu \) and \( \sigma^2 \)**

The marginal density of \( \mu \) is obtained by integrating out \( \sigma^2 \) from (12) and is given as

\[ \pi(\mu | x) = \int_0^{\infty} \pi(\mu, \sigma^2 | x) d\sigma^2 \]
\[
\pi(\mu | x) = c \int_0^{\infty} \frac{1}{(\sigma^2)^{n/2}} \exp \left[-\frac{1}{2\sigma^2} \left(\beta + n \left(\mu - \frac{\sum_{i=1}^{n} \log x_i}{n}\right)^2\right)\right] d\sigma^2
\]

\[
\pi(\mu | x) = c \frac{\Gamma\left(\frac{n}{2} - 1\right) 2^{n-1}}{\beta \left[\beta + n \left(\mu - \frac{\sum_{i=1}^{n} \log x_i}{n}\right)^2\right]^{n/2 - 1}}
\]

\[
\pi(\mu | x) = \frac{\sqrt{n}}{\sqrt{\beta}} \frac{1}{B\left(\frac{1}{2}, \frac{n-3}{2}\right)} \left[1 + \frac{n}{\beta} \left(\mu - \frac{\sum_{i=1}^{n} \log x_i}{n}\right)^2\right]^{n/2 - 1}
\]

The marginal density of \(\sigma^2\) is obtained by integrating the joint posterior density of \(\mu\) and \(\sigma^2\) given in (12) over the range of \(\mu\). It is given as

\[
\pi(\sigma^2 | x) = c \int_{-\infty}^{\infty} \frac{1}{(\sigma^2)^{n/2}} \exp\left(-\frac{\beta}{2\sigma^2}\right) \exp\left\{-\frac{n}{2\sigma^2} \left(\mu - \frac{\sum_{i=1}^{n} \log x_i}{n}\right)^2\right\} d\mu
\]

\[
\pi(\sigma^2 | x) = \frac{c \exp\left(-\frac{\beta}{2\sigma^2}\right)}{(\sigma^2)^{n/2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{n}{2\sigma^2} \left(\mu - \frac{\sum_{i=1}^{n} \log x_i}{n}\right)^2\right\} d\mu
\]
COMPARISON OF PARAMETERS OF LOGNORMAL DISTRIBUTION

\[
\pi\left(\sigma^2 \mid x \right) = c \frac{\exp\left(\frac{-\beta}{2\sigma^2}\right) \sqrt{2\pi}}{(\sigma^2)^{n/2} \sqrt{n}}
\]

\[
\pi\left(\sigma^2 \mid x \right) = \frac{\exp\left(\frac{-\beta}{2\sigma^2}\right) \beta^{\frac{n-3}{2}}}{\left(\sigma^2\right)^{n/2} \sqrt{2\pi} \Gamma\left(\frac{n-3}{2}\right)}
\]  

(14)

**Posterior estimates of \( \mu \) and \( \sigma^2 \)**

The marginal density of \( \mu \) is given in (13) is a student’s t pdf. Thus the posterior estimates of \( \mu \) is given as

\[
\mu^* = E\left(\mu \mid x \right) = \sqrt{\frac{n}{\beta}} \frac{1}{B\left(\frac{1}{2}, \frac{n-3}{2}\right)} \int_{-\infty}^{\infty} \frac{\mu \, d\mu}{\left[1 + \frac{n}{\beta} \left(\mu - \frac{\sum \log x_i}{n}\right)^2\right]}^{n-2}
\]

Using the transformation \( t = \sqrt{\frac{n}{\beta}} \left(\mu - \frac{\sum \log x_i}{n}\right) \sqrt{n-3} \)

\[
\mu^* = \frac{\sum_{i=1}^{n} \log x_i}{n\sqrt{n-3} B\left(\frac{1}{2}, \frac{n-3}{2}\right)} \int_{-\infty}^{\infty} \frac{dt}{\left[1 + \frac{t^2}{n-3}\right]^{n-2}}
\]
\[ \mu^* = \frac{\sum_{i=1}^{n} \log x_i}{n} \]  \hspace{1cm} (15)

which is the posterior estimate for \( \mu \) under uniform prior. Now the posterior estimate of \( \sigma^2 \) can be obtained from equation (14) as

\[ \sigma^* = \int_{0}^{\infty} \frac{\sigma^2 \exp \left( \frac{-\beta}{2\sigma^2} \right) \beta^{n-3}}{(\sigma^2)^{n/2} \Gamma \left( \frac{n-3}{2} \right)} d\sigma^2 \]

\[ \sigma^* = \frac{\beta}{n-5} \]  \hspace{1cm} (16)

Thus, the posterior estimates of \( \alpha_i \) and \( \beta_i \) are given by

\[ \alpha_i^* = \exp \left[ \frac{\mu^* + \sigma^2}{2} \right] = \exp \left[ \frac{\sum_{i=1}^{n} \log x_i}{n} + \frac{\beta}{2(n-5)} \right] \]  \hspace{1cm} (17)

and

\[ \beta_i^* = \exp \left[ 2\mu^* + \sigma^2 \right] \left[ \exp \left( \sigma^2 \right) - 1 \right] \]
Simulation study and discussion

The estimates of the mean and variance using MLE and Bayesian estimation was obtained above. Next to obtain is the numerical relationship of point estimates using true value of the parameters, MLE and Bayesian estimation.

In this study, samples of 10, 20, 30, 40 and 50 observations were generated from lognormal pdf with parameters $\mu = 2$ and $\sigma = 1$. The simulations were done in R Software. The mean and variance were calculated to compare the methods of estimation. The results are presented in Table 1.

In Table 1, when point estimates of lognormal distribution are compared using true values of parameters with MLE and Bayesian estimation (by using uniform prior), the best estimator is the Maximum Likelihood (MLE) because it has the minimum variance.

Table 1. Point estimates of lognormal distribution compared using true values of parameters with MLE and Bayesian estimation

<table>
<thead>
<tr>
<th>n</th>
<th>True values</th>
<th>MLE</th>
<th>Posterior estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>($\alpha_1$)</td>
<td>($\beta_1$)</td>
<td>($\hat{\alpha}_1$)</td>
</tr>
<tr>
<td>20</td>
<td>12.6952</td>
<td>72.127</td>
<td>9.8447</td>
</tr>
<tr>
<td>30</td>
<td>12.1825</td>
<td>255.011</td>
<td>12.6655</td>
</tr>
<tr>
<td>40</td>
<td>12.6452</td>
<td>56.9317</td>
<td>10.2974</td>
</tr>
</tbody>
</table>

References


