DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 12 | Issue 2 Article 18

11-1-2013

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

Journal of Modern Applied Statistical Methods
November 2013, Vol. 12, No. 2, 304-313.

ISSN 1538 − 9472

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 -∞ to +∞) as a model for the failure time distribution which must range from 0 to ∞.

A random variable X is said to have a lognormal distribution if $U = \log_e X$ has normal distribution with mean μ and variance σ^2. Thus, the pdf of lognormal distribution is given by

Dr. Sultan is a research scholar in the Department of Statistics. Email him at: hamzasultan18@yahoo.com. Dr. Ahmad is an Assistant Professor of in the Department of Statistics. Email him at: sprvz@yahoo.com.

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma x} \exp(-\frac{1}{2\sigma^2}(\log x - \mu)^2) \quad , -\infty < \mu < \infty,\ \sigma^2 > 0,\ \ 0 \le x < \infty \quad (1)$$

The likelihood function of the random sample $(x_1, x_2, x_3, ..., x_n)^T$ would be

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

The mean and variance of the lognormal distribution are given by

$$E(X) = \alpha_1 = \exp\left(\mu + \frac{\sigma^2}{2} \right) \qquad (3)$$

and

$$V(X) = \beta_1 = \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2) - 1\right) \qquad (4)$$

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 α1 and β1. Let $U_i = \log x_i$, $i = 1, 2, \ldots, n$. Then using the fact that $(U_1, U_2, ..., U_n)$ is a random sample from Normal distribution with parameters (μ, σ^2). The mle of μ and σ^2 first are given by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} U_i = \overline{U} \qquad (5)$$

and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(U_i - \bar{U})^2 \tag{6}$$

The mle of α_1 and β_1 are given by

$$\hat{\alpha}_1 = \exp\left(\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right) \tag{7}$$

and

$$\hat{\beta}_1 = \exp\left(2\hat{\mu} + \hat{\sigma}^2\right)\left(\exp\left(\hat{\sigma}^2\right) - 1\right) \tag{8}$$

Posterior estimation of the parameter

Again, consider the estimation of the parameters α_1 and β_1. First obtain the posterior estimates of μ and σ^2 and then simultaneously the posterior estimates for α_1 and β_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 μ and σ^2 considered is

$$P\left(\mu, \sigma^2\right) \propto 1 \tag{9}$$

According to Bayes theorem, Joint posterior density of μ and σ^2 is given by

Posterior density $\propto$ prior density* likelihood

$$\pi\left(\mu, \sigma^2 \mid \underset{-}{x}\right) \propto P\left(\mu, \sigma^2\right).P\left(\mu, \sigma^2 \mid \underset{-}{x}\right)$$

From equation (2) and (9) the joint posterior density of μ and σ^2 is given by

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

$$\pi\left(\mu,\sigma^2 \mid \underset{-}{x}\right)=\frac{c}{\left(\sigma^2\right)}\exp\left(\frac{-\beta}{2\sigma^2}\right)\exp\left\{\frac{-n}{2\sigma^2}\left(\mu-\frac{\sum_{i=1}^{n}\log x_i}{n}\right)^2\right\} \tag{10}$$

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

Therefore, c can be obtained by

$$c^{-1}=\int_{0}^{\infty}\int_{-\infty}^{\infty}\pi\left(\mu,\sigma^2\mid\underset{-}{x}\right)d\mu\, d\sigma^2$$

$$c^{-1}=\int_{o}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\left(\sigma^2\right)^{n/2}}\exp\left(\frac{-\beta}{2\sigma^2}\right)\exp\left\{\frac{-n}{2\sigma^2}\left(\mu-\frac{\sum_{i=1}^{n}\log x_i}{n}\right)^2\right\}d\mu\ d\sigma^2$$

Using the transformation $t=\frac{\sqrt{n}\left(\mu-\frac{\sum_{i=1}^{n}\log x_i}{n}\right)}{\sigma}$

$$c^{-1} = \sqrt{\frac{2\pi}{n}} \int_{o}^{\infty} \frac{\exp\left(\frac{-\beta}{2\sigma^2}\right)}{\left(\sigma^2\right)^{n-1/2}} d\sigma^2$$

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

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

From the equation (10)

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

Marginal posterior densities of μ and σ^2

The marginal density of μ is obtained by integrating out σ^2 from (12) and is given as

$$\pi\left(\mu \mid \underset{-}{x}\right) = \int_{0}^{\infty} \pi\left(\mu,\sigma^2 \mid \underset{-}{x}\right) d\sigma^2$$

$$\pi\left(\mu \mid \underset{-}{x}\right)=c\int_{0}^{\infty}\frac{1}{\left(\sigma^{2}\right)^{n/2}}\exp\left[-\frac{1}{2\sigma^{2}}\left\{\beta+n\left(\mu-\frac{\sum_{i=1}^{n}\log x_{i}}{n}\right)^{2}\right\}\right]d\sigma^{2}$$

$$\pi\left(\mu \mid \underset{-}{x}\right)=c\frac{\Gamma\left(\frac{n}{2}-1\right)2^{\frac{n}{2}-1}}{\left[\beta+n\left(\mu-\frac{\sum_{i=1}^{n}\log x_{i}}{n}\right)^{2}\right]^{\frac{n}{2}-1}}$$

$$\pi\left(\mu \mid \underset{-}{x}\right)=\sqrt{\frac{n}{\beta}}\frac{1}{B\left(\frac{1}{2},\frac{n-3}{2}\right)\left[1+\frac{n}{\beta}\left(\mu-\frac{\sum_{i=1}^{n}\log x_{i}}{n}\right)^{2}\right]^{\frac{n-2}{2}}} \tag{13}$$

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

$$\pi\left(\sigma^{2} \mid \underset{-}{x}\right)=c\int_{-\infty}^{\infty}\frac{1}{\left(\sigma^{2}\right)^{n/2}}\exp\left(\frac{-\beta}{2\sigma^{2}}\right)\exp\left\{-\frac{n}{2\sigma^{2}}\left(\mu-\frac{\sum_{i=1}^{n}\log x_{i}}{n}\right)^{2}\right\}d\mu$$

$$\pi\left(\sigma^{2} \mid \underset{-}{x}\right)=\frac{c\exp\left(\frac{-\beta}{2\sigma^{2}}\right)}{\left(\sigma^{2}\right)^{n/2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{n}{2\sigma^{2}}\left(\mu-\frac{\sum_{i=1}^{n}\log x_{i}}{n}\right)^{2}\right\}d\mu$$

$$\pi\left(\sigma^2 \mid \underset{-}{x}\right) = c\,\frac{\exp\left(\dfrac{-\beta}{2\sigma^2}\right)\sqrt{2\pi}}{\left(\sigma^2\right)^{n/2}\sqrt{n}}$$

$$\pi\left(\sigma^2 \mid \underset{-}{x}\right) = \frac{\exp\left(\dfrac{-\beta}{2\sigma^2}\right)\beta^{\frac{n-3}{2}}}{\left(\sigma^2\right)^{n/2} 2^{\frac{n-3}{2}}\,\Gamma\left(\dfrac{n-3}{2}\right)} \tag{14}$$

Posterior estimates of μ and σ^2

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

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

Using the transformation $t = \sqrt{\dfrac{n}{\beta}}\left(\mu - \dfrac{\sum_{i=1}^{n}\log x_i}{n}\right)\sqrt{n-3}$

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

$$\mu^* = \frac{\sum_{i=1}^{n} \log x_i}{n} \tag{15}$$

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

$$\sigma^{*2} = \int_0^\infty \frac{\sigma^2 \exp\left(\frac{-\beta}{2\sigma^2}\right)\beta^{\frac{n-3}{2}}}{\left(\sigma^2\right)^{\frac{n-1}{2}} 2^{\frac{n-3}{2}} \Gamma\left(\frac{n-3}{2}\right)} d\sigma^2$$

$$\sigma^{*2} = \frac{\beta^{\frac{n-3}{2}}}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-3}{2}\right)} \int_0^\infty \frac{\exp\left(\frac{-\beta}{2\sigma^2}\right)}{\left(\sigma^2\right)^{\frac{n-3}{2}}} d\sigma^2$$

$$\sigma^{*2} = \frac{\beta}{n-5} \tag{16}$$

Thus, the posterior estimates of α_1 and β_1 are given by

$$\alpha_1^* = \exp\left[\mu^* + \frac{\sigma^{*2}}{2}\right] = \exp\left[\frac{\sum_{i=1}^{n} \log x_i}{n} + \frac{\beta}{2(n-5)}\right] \tag{17}$$

and

$$\beta_1^* = \exp\left[2\mu^* + \sigma^{*2}\right]\left[\exp\left(\sigma^{*2}\right) - 1\right]$$

$$\beta_1^* = \exp\left[2\frac{\sum_{i=1}^{n}\log x_i}{n} + \frac{\beta}{n-5}\right]\left[\exp\left(\frac{\beta}{n-5}\right) - 1\right] \quad (18)$$

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

n	True values		MLE		Posterior estimates	
	Mean (α_1)	Variance (β_1)	Mean ($\hat{\alpha}_1$)	Variance ($\hat{\beta}_1$)	Mean (α_1^*)	Variance (β_1^*)
10			9.9004	21.336	9.1225	62.1358
20			12.6952	72.127	9.8447	130.857
30	12.1825	255.011	12.6655	52.1804	10.5913	70.6613
40			12.6452	56.9317	10.2974	71.6757
50			20.4039	267.461	12.5356	356.339

References

Aitchison, J., & Brown, J. A. C. (1957). *The lognormal distribution: With special reference to its uses in economics*. Cambridge: University Press.

Brownlee, K. A. (1949). *Industrial experimentation*. 2nd ed. Brooklyn: Chemical Pub. Co.

Delaporte, P. (1950). Etude statistique sur les proprietes des fontes. *Revue De L'institut International De Statistique / Review Of The International Statistical Institute*, (3/4), 161. doi:10.2307/1401035

Epstein. B. (1947). The mathematical description of certain breakage mechanism leading to the logarithmic-normal distribution. *Journal of Franklin Institute*, *244*, 471-477.

Laplace, P. S. (1774). Mémorie sur la probabilitié des causes par les événemens. *Académie royale des sciences (France), 6*, 621-656.

Lindley, D. V. (1961). *Introduction to probability and statistics from a Bayesian viewpoint: Part 2, Inference*, Aberystwyth: University College of Wales.

Moroney, M. J. (1951). *Facts from figures*. Baltimore, MD: Penguin Books.