Journal of Statistical and Econometric Methods, vol.1, no.2, 2012, 125-131
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2012

Inference on Difference of Means of two Log-Normal Distributions

A Generalized Approach

K. Abdollahnezhad[1], M. Babanezhad[*,2] and A.A. Jafari[3]

Abstract

Over the past decades, various methods for comparing the means of two log-normal have been proposed. Some of them are differing in terms of how the statistic test adjust to accept or to reject the null hypothesis. In this study, a new method of test for comparing the means of two log-normal populations is given through the generalized measure of evidence to have against the null hypothesis. However calculations of this method are simple, we find analytically that the considered method is doing well through comparing the size and power statistic test. In addition to the simulations, an example with real data is illustrated.

Mathematics Subject Classification: 49J21, 49K21
Keywords: Log-Normal distribution, Null hypothesis, Test statistics, Generalized test

1 Introduction

One often encounters with random variables that are inherently positive in some real life applications such as analyzing biological, medical, and industrial

[1] Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Iran
[2] Department of Statistics, Faculty of Sciences, Golestan University, Gorgan, Golestan, Iran. * Corresponding Author, e-mail: m.babanezhad@gu.ac.ir
[3] Department of Statistics, Faculty of Sciences, Yazd University, Yazd, Iran.

Article Info: *Received* : May 1, 2012. *Revised* : June 30, 2012
Published online : July 30, 2012

data. In this regards the normal distribution is applied in most of applications. In the family of normal distribution the Log normal distribution has a long term applications. In probability theory, a log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Further, a variable might be modeled as log-normal if it can be thought of as the multiplicative product of many independent random variables each of which is positive. The suitability of the log-normal random variable has been investigated by some researchers (Crow and Shimizu, 1988). There are also some recent articles regarding the statistical inference of parameters of several log-normal distributions. For example, one-sided test have been investigated for two distributions with a large sample under the homogeneity of the mean parameters for m log-normal populations (Zhou et al., 1997; Ahmed et al., 2002). Further, exact confidence interval test for the ratio or difference of the means of two log-normal distributions using the generalized variable and generalized p-values through a modified likelihood ratio has been done (Krishnamoorthy and Mathew, 2003; Gill, 2004; Gupta and Li, 2005). In this paper, we consider random samples from two lognormal populations and our interest is to present a test of difference of the means of these two populations. In Section 2, the theory of generalized p-value is introduced. Section 3 is devoted to an exact one-sided test or two-sided test for two log-normal distributions. We compare the size and power of different proposed methods to test of the means of two log-normal populations in Section 4 through simulation. We examine them by a numerical example with real data set. A brief discussion is given in Section 5.

2 Generalized p-value

The concept of generalized p-value is applied to deal with the statistical testing problem in which nuisance parameters are presented (Tsui and Weerahandi, 1989). It is difficult or impossible to obtain a nontrivial test with a fixed level of significance. To go further, let $\boldsymbol{X}$ be a random variable with density function $f(\boldsymbol{x}|\boldsymbol{\zeta})$, where $\boldsymbol{\zeta} = (\theta, \boldsymbol{\eta})$ is a vector of unknown parameters, θ is the parameter of interest, and $\boldsymbol{\eta}$ is a vector of nuisance parameters (Figure 1).

Suppose we are interested to test

$$H_\circ : \theta \leqslant \theta_\circ \quad vs \quad H_1 : \theta > \theta_\circ, \tag{1}$$

where $\theta_\circ$ is a specified value.

Let $\boldsymbol{x}$ denote the observed value of $\boldsymbol{X}$ and consider a variable $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\zeta})$, by the name of generalized variable. We assume that $T(\boldsymbol{X}; \boldsymbol{x}, \boldsymbol{\zeta})$ satisfies the following conditions:

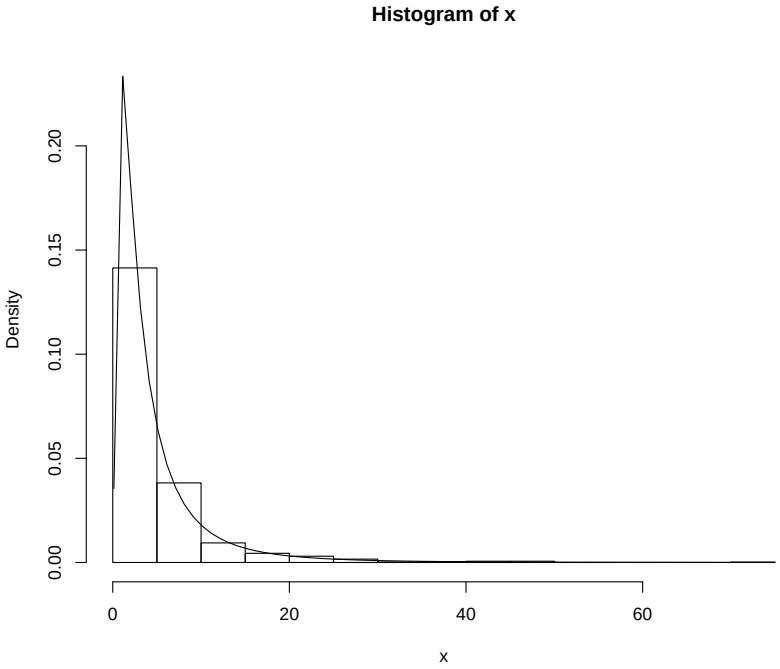

Figure 1: X is a random variable with a Lognormal distribution.

(i) For fixed $\boldsymbol{x}$, the distribution of $T(\boldsymbol{X};\boldsymbol{x},\boldsymbol{\zeta})$ is free from the nuisance parameters $\boldsymbol{\eta}$.

(ii) $t_{obs} = T(\boldsymbol{x};\boldsymbol{x},\boldsymbol{\zeta})$ is free from any unknown parameters.

(iii) For fixed $\boldsymbol{x}$ and $\boldsymbol{\eta}$, $T(\boldsymbol{X};\boldsymbol{x},\boldsymbol{\zeta})$ is either stochastically increasing or decreasing in θ for any given t.

Under the above conditions, if $T(\boldsymbol{X};\boldsymbol{x},\boldsymbol{\zeta})$ is stochastically increasing in θ, then the generalized p-value for testing the hypothesis in (1) can be defined as

$$p = \sup_{\theta \leqslant \theta_\circ} P(T(\boldsymbol{X};\boldsymbol{x},\theta,\boldsymbol{\eta}) \geq t^*) = P(T(\boldsymbol{X};\boldsymbol{x},\theta_\circ,\boldsymbol{\eta})) \geq t^*), \qquad (2)$$

where $t^* = T(\boldsymbol{X};\boldsymbol{x},\theta_\circ,\boldsymbol{\eta})$. For further details and for several applications based on the generalized p-value, we refer to the book by Weerahandi (1995).

3 A Generalized Test Variable

Let $Y_{ij} = \ln(X_{ij}) \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$, $j = 1, 2, ..., n_i$ be independent random samples from two log-normal populations.
We know that $M_i = E(X_{ij}) = \exp(\mu_i + 0.5\sigma_i^2)$. The problem of our interest is one sided and two sided test hypothesis about $\eta = M_1 - M_2$.

In this section, using the concept of generalized p-value, we test

$$H_\circ : M_1 \leqslant M_2 \quad vs \quad H_1 : M_1 > M_2, \tag{3}$$

which is equivalent to

$$H_\circ : \theta \ \leqslant 0 \quad vs \quad H_1 : \theta \ > 0, \tag{4}$$

where $\theta = \ln M_1 - \ln M_2$.

The MLE's for μ_i and σ_i^2 $(i = 1, 2)$ are $\bar{Y}_i$ and S_i^2, respectively, where

$$\bar{Y}_i = \frac{1}{n_i}\sum_{i=1}^{n} Y_{ij} \quad , \quad S_i^2 = \frac{1}{n_i}\sum_{i=1}^{n}(Y_{ij} - \bar{Y}_i)^2.$$

Now, consider

$$\begin{aligned} T &= \bar{y}_{1.} - \bar{y}_{2.} + \frac{\bar{Y}_{2.} - \bar{Y}_{1.} - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sqrt{\frac{\sigma_1^2 s_1^2}{n_1 S_1^2} + \frac{\sigma_2^2 s_2^2}{n_2 S_2^2}} + \frac{\sigma_1^2 s_1^2}{2S_1^2} - \frac{\sigma_2^2 s_2^2}{2S_2^2} - \theta \\ &= \bar{y}_{1.} - \bar{y}_{2.} + Z\sqrt{\frac{s_1^2}{U_1} + \frac{s_2^2}{U_2}} + \frac{n_1 s_1^2}{2U_1} - \frac{n_2 s_2^2}{2U_2} - \theta, \end{aligned}$$

where

$$Z = \frac{\bar{Y}_{2.} - \bar{Y}_{1.} - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1),$$

and

$$U_i = \frac{n_i S_i^2}{\sigma_i^2} \sim \chi^2_{(n_i - 1)} \ , \ \ i = 1, 2,$$

are three independent random variables, and $\bar{y}_i$ and s_i^2 are observed values of $\bar{Y}_i$ and S_i^2, respectively.Then, T is a generalized variable for θ because

i) $t_{obs} = 0$

ii) Distribution of T is free from the nuisance parameters μ_i and σ_i^2.

iii) Distribution of T is an increasing function with respect to θ.

Thus the generalized p-value for the null hypothesis (3) is given by

$$p = P(T \leqslant t_{obs} | \theta = 0) = E(\Phi(\frac{\bar{y}_{2.} - \bar{y}_{1.} + \frac{n_2 s_2^2}{2U_2} - \frac{n_1 s_1^2}{2U_1}}{\sqrt{\frac{s_1^2}{U_1} + \frac{s_2^2}{U_2}}})), \tag{5}$$

where $\Phi(.)$ is the standard normal distribution function and the expectation is taken with respect to independent chi-square random variables, U_1 and U_2.

This generalized p-value can be well approximated by a Monte Carlo simulation using the following algorithm:

Algorithm 1. For a given random sample $x_{i1}, ..., x_{in_i}$, let $y_{ij} = \ln(x_{ij})$, $i = 1, ..., k$, $j = 1, 2.$, and compute $\bar{y}_{1.}$, $\bar{y}_{2.}$, s_1^2, s_2^2.

Algorithm 2. For $l = 1$ to m, generate

$$U_1 \sim \chi^2_{(n_1-1)}, \;\; U_2 \sim \chi^2_{(n_2-1)},$$

and calculate

$$T_l = \Phi\left(\frac{\bar{y}_{2.} - \bar{y}_{1.} + \frac{n_2 s_2^2}{2U_2} - \frac{n_1 s_1^2}{2U_1}}{\sqrt{\frac{s_1^2}{U_1} + \frac{s_2^2}{U_2}}}\right).$$

$\frac{1}{m}\sum_{l=1}^{m} T_l$ is a Monte Carlo estimation of generalized p-value for the null hypothesis (3).

The generalized p-value in (5) is used for one sided test hypothesis but we can use this generalized p-value for two sided test hypothesis by

$$p = 2\min\{p, 1-p\},$$

where p is the generalized p-value in (5).

4 Simulation Study

To investigate the power of the considered test statistics in finite samples, we conducted a simulation experiment. To do so, several data set from two log-normal distributions with $\mu_2 = 0$ were generated. For each scenarios 10000 sample size are performed. The size and the power of the considered test statistics are summarized in Table 1. These tests are:
(a) generalized p-value in (5)
(b) generalized p-value by Krishnamoorthy and Mathew (2003)
(c) Z-score test by Zhou et al. (1997).

The simulation study indicates that
(i) The size for (a) and (b) are close to 0.05 and the powers are close to each other.
(ii) The size of (c) is very larger than nominal level, 0.05.

The numerical examples data is the amount of rainfall (in acre-feet) from 52 clouds. From this 26 clouds were chosen at random and seeded with silver nitrate. We can show that the considered data follow the log-normal distribution. The summary statistics for the log-transformed data are given in Table 1. In order to understand the effect of silver nitrate seeding, we like to test

Table 1: The summary statistics for the log-transformed data of rainfall

Clouds	n_i	$\bar{y}_{i.}$	s_i^2
seeded clouds	26	5.134	2.46
unseeded clouds	26	3.990	2.60

$$H_\circ : M_1 = M_2 \quad vs \quad H_1 : M_1 > M_2, \tag{6}$$

where $M_i = \exp(\mu_i + 0.5\sigma_i^2)$, $i = 1, 2$.

The p-values for our generalized approach, Krishnamoorthy and Mathew approach and Z-score test are 0.078, 0.075, and 0.060, respectively.

5 Discussion

This paper investigated the inference on difference of means of two Log-Normal distribution true testing a hypothesis. The lognormal distribution is widely used to describe the distribution of positive random variables. In this regards, we have investigated the impact of test statistic for comparing the means of two log-normal populations through the generalized measure of evidence to have against the null hypothesis. We in fact derived the exact inference procedures (hypotheses tests) concerning the difference mean of two single lognormal distribution. Our interest in this stems from the fact that we anticipated the different test statistics are doing differently by critical regions. In terms to our anticipation, we found that we cannot reject H_0 at the level of 0.05, using all 3 methods in the rainfall data. However, the problem of statistical inference concerning the mean of the lognormal distribution might be appeared.

References

[1] S.E. Ahmed, R. J. Tomkins and A.I. Volodin, Test of homogeneity of parallel samples from lognormal populations with unequal variances, *Journal of Statistical Research*, **35**(2), (2001), 25-33.

[2] E.L Crow and K. Shimizu, *Lognormal distribution*, Marcel Dekker, New York, 1998.

[3] P.S. Gill, Small sample inference for the comparison of means of lognormal distribution, *Biometrics*, **60**(2), (2004), 525-527.

[4] R.C. Gupta and X. Li, Statistical inferences on the common mean of two log-normal distributions and some applications in reliability, *Computational Statistics and Data Analysis*, **50**(11), (2006), 3141-3164.

[5] K. Krishnamoorthy and T. Mathew, Inferences on the means of lognormal distributions using generalized p-values and generalized confidence interval, *Journal of Statistical Planning and Inference*, **115**, (2003), 103-121.

[6] K. Krishnamoorthy and L. Yong, Inferences on the common mean of several normal populations based on the generalized variable method, *Biometrics*, **59**, (2003), 237-247.

[7] K.W. Tsui and S. Weerahandi, Generalized p-values in significance testing of hypothesis in the presence of nuisance parameters, *Journal of the American Statistical Association*, **84**, (1989), 602-607.

[8] S. Weerahandi, Generalized confidence intervals, *Journal of the American Statistical Association*, **88**, (1993), 899-905.

[9] S. Weerahandi, *Exact Statistical Methods for Data Analysis*, Springer, NewYork, 1995.

[10] S. Weerahandi and V.W. Berger, Exact inference for growth curves with interclass correlation structure, *Biometrics*, **55**, (1999), 921-924.