Journal of Statistical Research
2001, Vol. 35, No. 2, pp 25-33
Bangladesh

ISSN 0256 - 422 X

TEST OF HOMOGENEITY OF PARALLEL SAMPLES FROM LOGNORMAL POPULATIONS WITH UNEQUAL VARIANCES

S. E. AHMED, R. J. TOMKINS and A. I. VOLODIN
University of Regine

SUMMARY

Consider $m(\geq 2)$ independent random samples from m lognormal populations with mean parameter $\theta_1, \cdots \theta_m$ respectively. A large sample test for the homogeneity of the mean parameters is developed. An estimator and a confidence interval are proposed for the common mean parameter. The asymptotic distribution of the proposed test-statistic under the null hypothesis as well as under local alternative is derived.

Keywords: Common mean, Combination of lognormal models, Asymptotic tests and confidence interval, Asymptotic power.

1 Introduction

Let the random variable Y be distributed normally with mean μ and variance σ^2; then $X = e^Y$ will have a lognormal distribution. The probability distribution function of X is given by

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

Evidently, the lognormal model is related to the normal distribution in the same way that the Weibull is related to the extreme value distribution. Both normal and lognormal models have received considerable use in lifetime and reliability problems. In medical and engineering applications, one often encounters random variables X such that logarithm of X has a normal distribution. This lognormal model is commonly used in medicine and economics, where the basic process under consideration leads to phenomena which are often a multiplication of factors. Also, as stated in Cheng (1977), reliability studies indicate that many semi-conductor devices follow lifetime distributions, which are well represented by the lognormal. We refer to an edited volume by Crow and Shimizu (1988) for a comprehensive review of the subject's theory and applications.

Suppose $X_{j1}, X_{j2}, \cdots, X_{jn_j}$ is a random sample from a two-parameter lognormal distribution with mean θ_j and variance τ_j^2, denoted by $\Lambda(\theta_j, \tau_j^2)$, corresponding to the normal distribution with mean μ_j and variance σ_j^2, $j = 1, 2, \cdots, m$. Thus,

$$\theta_j = e^{\mu_j + \frac{\sigma_j^2}{2}}, \quad \tau_j^2 = e^{2\mu_j + \sigma_j^2}\left(e^{\sigma_j^2} - 1\right).$$

Define $Y_{ji} = \ln(X_{ji})$, $1 \leq i \leq n_j$. If μ_j and σ_j^2 are unknown, then the maximum likelihood estimator (MLE) of (μ_j, σ_j^2) is $(\hat{\mu}_j, \hat{\sigma}_j^2)$ with

$$\hat{\mu}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ji}, \ \hat{\sigma}_j^2 = \frac{1}{n_j}\sum_{i=1}^{n_j}(Y_{ji} - \hat{\mu}_j)^2. \tag{1.1}$$

Hence, the *maximum likelihood estimators (MLE)* of (θ_j, τ_j^2) are

$$\hat{\theta}_j = e^{\hat{\mu}_j + \frac{1}{2}\hat{\sigma}_j^2}, \hat{\tau}_j^2 = e^{2\hat{\mu}_j + \hat{\sigma}_j^2}\left(e^{\hat{\sigma}_j^2} - 1\right). \tag{1.2}$$

The sampling distribution of $\hat{\theta}_j$ is given in the following lemma.

Lemma AT (Ahmed and Tomkins, 1995) For each $j = 1, 2, \cdots, m$,

$$\sqrt{n_j}\left(\hat{\theta}_j - \theta_j\right) \underset{D}{\longrightarrow} \mathcal{N}(0, \ \nu_j)$$

as n_j tends to ∞, where $\underset{D}{\longrightarrow}$ means convergence in distribution and $\nu_j = \sigma_j^2(1 + \sigma_j^2/2)\exp\{2\mu_j + \sigma_j^2\}$.

We consider the estimation and testing of the lognormal means when multiple-samples are available. For example, the data may have been acquired at a different time or space. Also, in many experimental situations it is a common practice to replicate an experiment. In these situations, we often encounter the problem of pooling independent estimates of a parameter obtained from different sources. Each such estimate is reported as a number with an estimated standard error. In this case two problems are to be considered. First, can all these estimates be considered to be homogeneous, i.e., are they estimating the same parameter? Second, if the estimates are homogeneous, what is the best way of combining them to obtain a single estimate? We will address these questions in an orderly manner. The focal point of this investigation is to develop an inference procedure when several parallel samples are available. In a large-sample classical setup, we propose a test of the homogeneity of the lognormal means. Further, a point estimator and interval estimator of the common parameter is given.

2 Multiple Sample Problems

In this section we discuss estimation and testing procedures to provide information about properties of the mean parameters based on several random samples taken from lognormal populations.

2.1 Pooling of Estimates

In many real occasions it is desirable to combine the individual sample estimates to obtain a combined or pooled estimate of common parameter θ. It is known that the variance of a linear combination $\sum_{j=1}^{m} l_j \xi_j$ of the random variables $\xi_1, \ldots, \xi_m$, subject to $\sum_{j=1}^{m} l_j = 1$ is minimized by choosing

$$l_j = \frac{\sigma_j^{-2}}{\sum_{j=1}^{m} \sigma_k^{-2}},$$

where σ_j^2 is the variance of the ξ_j, $j = 1, \ldots, m$.

Hence, the following theorem is proposed.

Theorem 2.1 Let $X_j \sim \Lambda(\theta, \tau_j^2)$, $j = 1, \cdots, m$, and suppose that observations from a sample of size n_j are available for each population. Then a combined sample estimate of θ which has minimum variance among the class of the unbiased estimators of θ which are linear functions of $\hat{\theta}_1, \ldots \hat{\theta}_m$ is given by

$$\tilde{\theta} = \frac{\sum_{j=1}^{m} \frac{n_j}{\hat{\nu}_j} \hat{\theta}_j}{\sum_{j=1}^{m} \frac{n_j}{\hat{\nu}_j}},$$

where

$$\hat{\nu}_j = \hat{\sigma}_j^2 (1 + \hat{\sigma}_j^2/2) \exp\{2\hat{\mu}_j + \hat{\sigma}_j^2\}.$$

Then, $\tilde{\theta}_n$ is approximately normally distributed with mean θ and asymptotic variance $\left[\sum_1^m n_j/\nu_j\right]^{-1}$

The proof of the above theorem follows from Lemma AT (stated in the previous section). It is concluded from Theorem 2.1 that $\tilde{\theta}$ provides a good estimate for a common θ based on several samples. More importantly, it is not necessary for the ν_j to be equal in this case. We refer to Rao (1981, pp. 389-391) for further discussions on the topic.

2.2 Test of Hypothesis

There are many situations when one must make a decision that is based on an unknown parameter's value(s). On such occasions a test of the hypothesis about the parameters may be more appropriate. We consider two classes of testing problems for lognormal data:

(i) simple null versus global alternative, and

(ii) test for homogeneity.

2.2.1 Test of simple null versus global alternative

Using familiar matrix notation, let $= (\theta_1, \cdots, \theta_m)'$ be a $m \times 1$ be vector of parameters and $= (\hat{\theta}_1, \cdots, \hat{\theta}_m)'$ be the maximum likelihood estimator vector of . Suppose it is desired to test the simple hypothesis

$$H_o :=^o, \quad ^o = (\theta_1^o, \theta_2^o, \cdots, \theta_m^o)' \tag{2.1}$$

against the global alternative

$$H_a :\neq^o .$$

Define $n = n_1 + n_2 + \cdots + n_m$. It is natural to construct a test statistic for the null hypothesis, which is defined by the normalized distance of from ^{o}. Hence, define

$$T_1 = n(-^o)'\mathbf{S}_1^{-1}(-^o),$$

where

$$\mathbf{S}_1 = Diag\left(\frac{n}{n_1}\hat{\nu}_1, \cdots, \frac{n}{n_m}\hat{\nu}_m\right).$$

Note that the test statistics T_1 can be rewritten as

$$T_1 = \sum_{j=1}^{m} \frac{n_j\left(\hat{\theta}_j - \theta_j^\circ\right)^2}{\hat{\nu}_j} \tag{2.2}$$

By using *Lemma AT* we arrive at the following theorem.

Theorem 2.2: Let $n \to \infty$ and also n_j/n approches to a constant for any $j = 1, \ldots, m$. Then under the null hypothesis in (2.1), the test statistic T_1 follows a chi-square distribution with m degrees of freedom.

Proof. Follows from the Lemma AT and the fact that T_1 is the sum of squares of independent asymptotically normal random variables.

Thus, when the null hypothesis is true, the upper α-level critical value of T_1, by $C_{n,\alpha}$, may be approximated by the central χ^2 distribution with m degrees of freedom. Note that σ_j^2 is a function of θ_j; i.e., $\sigma_j^2 = 2(\ln \theta_j - \mu_j)$. Hence, under the null hypothesis $\sigma_j^2 = \sigma_j^{o2} = 2(\ln \theta_j^o - \hat{\mu}_j)$ and in this case we will have

$$\hat{\nu}_j = \hat{\nu}_j^o = \sigma_j^{o2}(1 + \sigma_j^{o2}/2)\exp\{2\hat{\mu}_j + \sigma_j^{o2}\}.$$

Thus, we can define another test statistic for the problem at hand as follows:

$$T_1^* = n(-^o)'\mathbf{S}_1^{*-1}(-^o),$$

where

$$\mathbf{S}_1^* = Diag\left(\frac{n}{n_1}\hat{\nu}_1^o, \cdots, \frac{n}{n_m}\hat{\nu}_m^o\right).$$

Note that the test statistics T_2 can be rewritten as

$$T_1^* = \sum_{j=1}^{m} \frac{\left(\hat{\theta}_j - \theta_j^o\right)^2}{\hat{\nu}_j^o} n_j$$

Under H_o, for large n, T_1^* will have a χ^2 distribution with m degrees of freedom.

2.3 Test of Homogeneity

In this section we focus on developing a testing methodology for the homogeneity of m lognormal means when samples are pooled. Let us suppose that two or more samples are available for which a common value of θ is assumed. The statistical problem is to test the following null hypothesis of homogeneity of the mean parameters:

$$H_o : \theta_1 = \theta_2 \cdots = \theta_m = \theta, \tag{2.3}$$

where the common value of θ is unknown.

Define $= (\tilde{\theta} \cdots, \tilde{\theta})' = \tilde{\theta}\mathbf{1}_m$, where $\mathbf{1}_m = (1, \cdots, 1)'$. We propose the following test statistic to test the null hypothesis in relation (2.2):

$$T_2 = n(-\tilde{\theta}\mathbf{1}_m)'\mathbf{S}_m^{-1}(-\tilde{\theta}\mathbf{1}_m),$$

where

$$\mathbf{S}_m = Diag\left(\frac{n}{n_1}\tilde{\nu}_1, \cdots, \frac{n}{n_m}\tilde{\nu}_m\right),$$

and

$$\tilde{\nu}_j = \tilde{\sigma}_j^2(1 + \tilde{\sigma}_j^2/2)\exp\{2\hat{\mu}_j + \tilde{\sigma}_j^2\},$$

where $\tilde{\sigma}_j^2 = 2\ln(\tilde{\theta}_j - \hat{\mu}_j)$.

And again, we can rewrite test statistics T_2 without of matrix notations as

$$T_2 = \sum_{j=1}^{m} \frac{n_j\left(\hat{\theta}_j - \tilde{\theta}\right)^2}{\tilde{\nu}_j}, \tag{2.4}$$

In an effort to derive the null distribution of T_2 we consider the asymptotic distribution of some random variables related to the proposed test statistic. Define

$$\mathbf{U}_n = \sqrt{n}(-\tilde{\theta}\mathbf{1}_m).$$

It can be seen that $\mathbf{U}_n = \sqrt{n}\tilde{\mathbf{C}}$ where

$$\tilde{\mathbf{C}} = \mathbf{I} - \frac{1}{\hat{\omega}}\mathbf{J}\tilde{\mathbf{D}}, \quad \text{where}$$

$$\tilde{\mathbf{D}} = Diag\left(\frac{n_1}{\tilde{\nu}_1}, \cdots, \frac{n_m}{\tilde{\nu}_m}\right), \quad \mathbf{J} = \mathbf{I} - \mathbf{1}_m\mathbf{1}_m'\tilde{\mathbf{D}}, \text{ and } \hat{\omega} = \sum_{j=1}^{m}\left(\frac{n_j}{\tilde{\nu}_j}\right)$$

Theorem 2.3: For large n, under the null hypothesis in (2.2), the test statistic T_2 is distributed as a chi-square distribution with $(m-1)$ degrees of freedom.
Proof. Follows from the Lemma AT and the fact that T_2 is the sum of squares of asymptotically independent asymptotically normal random variables.

As the consequence of the above theorem, under the null hypothesis and for large n, for given α, the critical value of T_2 may be approximated by $\chi^2_{m-1,\alpha}$, the upper $100\alpha\%$ point of the chi-square distribution with $(m-1)$ degrees of freedom.

2.4 Power of the Tests

It is important to note that, for a fixed alternative that is different from the null hypothesis, the power of both test statistics proposed earlier will converge to one as $n \to \infty$. This follows from the fact that test statistics tends to infinity if $\theta \neq \theta_0$ (cf. the similar agument given in Sen and Singer (1993, pages 237-238)). Thus, to study the asymptotic power properties of T_1 and T_2, we must confine ourselves to a sequence of local alternatives $\{K_n\}$. When is the parameter of interest, such a sequence may be specified by

$$K_n := {}^{o} + \frac{}{n^{\frac{1}{2}}}.$$

where is a vector of fixed real numbers. Evidently, approaches ^{o} at a rate to $n^{-1/2}$. Stochastic convergence of to ensures that $\xrightarrow[p]{}$ under local alternatives as well. Hence, nonnull distributions and the power of the proposed test statistic can be determined under the local alternatives.

Theorem 2.4: Under the local alternatives and as $n \to \infty$ we have the following distributional result:

$$n^{1/2}\{ - {}^{o}\} \xrightarrow[D]{} \mathcal{N}_m(, {}_1),$$

where

$${}_1 = \text{Diag}\left(\frac{\nu_1}{\omega_1}, \cdots, \frac{\nu_m}{\omega_m}\right).$$

Theorem 2.5: Under the local alternatives and as $n \to \infty$ we have the following distributional result:

$$n^{1/2}\{ - \tilde{\theta}\mathbf{1}_m\} \xrightarrow[D]{} \mathcal{N}_m(\mathbf{J} , {}_2),$$

where

$${}_2 = \text{Diag}\left(\frac{\nu_1}{\omega_1}, \cdots, \frac{\nu_m}{\omega_m}\right)\mathbf{C},$$

and

$$\mathbf{C} = \mathbf{I} - \frac{1}{\omega}\mathbf{J}\mathbf{D},$$

where

$$\mathbf{D} = Diag\left(\frac{\omega_1}{\nu_1}, \cdots, \frac{\omega_m}{\nu_m}\right), \quad \mathbf{J} = \mathbf{I} - \mathbf{1}_m\mathbf{1}_m'\tilde{\mathbf{D}}, \text{ and } \omega = \sum_{j=1}^{m}\left(\frac{\omega_j}{\nu_j}\right).$$

Proof of the both theorems can be obtained using the general contiguity theory (cf. Roussas (1972)).

By thorems 2.4 and 2.5, a sum of squares of asymptotically independent and asymptotically normal random variables with the unit variance and nonzero means has the asymptotic distribution noncentral χ^2-square with the corresponding parameter of noncentrality. Thus, under local alternatives, test statistics T_1 and T_2 will have asymptotically a noncentral chi-square distribution with m and $m-1$ degree of freedom respectively, and noncentrality parameters

$$\Theta_1 =_1^{-1}, \quad \Theta_2 = (\mathbf{J})_2'^{-1}(\mathbf{J}),$$

respectively. Hence, using a noncentral chi-square distribution, one can do the power calculations of the proposed test statistics.

2.5 Interval Estimation

Let $z_{\frac{\alpha}{2}}$ be the usual percentile point such that $1 - \Phi\left(z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$, where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable. Noting that,

$$Pr\left\{\hat{\theta}_j - z_{\frac{\alpha}{2}}\left(\frac{\hat{\nu}_j}{n_j}\right)^{1/2} \leq \theta_j \leq \hat{\theta}_j + z_{\frac{\alpha}{2}}\left(\frac{\hat{\nu}_j}{n_j}\right)^{1/2}\right\}$$

converges to $1-\alpha$ as $n_j \to \infty$. Hence intervals having $1-\alpha$ coverage probability for θ_j can be expressed as

$$\hat{\theta}_j \pm z_{\frac{\alpha}{2}}\left(\frac{\hat{\nu}_j}{n_j}\right)^{1/2} \tag{2.5}$$

If the null hypothesis $H_o: \ \theta_1 = \theta_2 = \cdots = \theta_m$ is not rejected, it may be of interest to obtain a $100(1-\alpha)\%$ confidence interval about the common value of θ. A $100(1-\alpha)\%$ confidence interval about θ may be obtained by using the combined data. Note that,

$$Pr\left\{\tilde{\theta} - z_{\frac{\alpha}{2}}\left(\frac{1}{\sum_{j=1}^{m}(n_j/\tilde{\nu}_j)}\right)^{1/2} \leq \theta \leq \tilde{\theta} + z_{\frac{\alpha}{2}}\left(\frac{1}{\sum_{j=1}^{m}(n_j/\tilde{\nu}_j)}\right)^{1/2}\right\}$$

converges to $1-\alpha$ as $n \to \infty$. Thus, a $100(1-\alpha)\%$ confidence interval about common parameter θ may be obtained as follows:

$$\tilde{\theta} \pm z_{\frac{\alpha}{2}}\left(\frac{1}{\sum_{j=1}^{m}(n_j/\tilde{\nu}_j)}\right)^{1/2} \tag{2.6}$$

Clearly this interval will provide shorter confidence interval than that based on individual estimates, for any given α.

3 Concluding Remarks

A large sample analysis is presented when m lognormal means are combined. A test of the homogeneity of the means is presented, and a point and interval estimator of the common mean parameter is also provided. As a word of caution, the statistical procedures based on combined estimates are sensitive to departure from the null hypothesis. Therefore, some other alternatives to pooled estimator should be considered. Furthermore, the proposed procedures involve nonlinear functions of asymptotic normal estimators that may not be well approximated by a normal law unless the sample sizes are large.

4 Acknowledgements

The work was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

[1] Ahmed, S. E. and R. J. Tomkins (1995). Estimating lognormal means under uncertain prior information. *Pakistan Journal of Statistics,* **11**, *67-92.*

[2] Cheng, S. S. (1977). Optimal replacement rate of devices with lognormal failure distribution. *IEEE Trans. Reliability* **R-26**, 174-178.

[3] Crow, E. L. and Shimizu, K. (1988). Lognormal Distributions. *Marcel Dekker: New York.*

[4] Rao, C. R. (1981). Linear Statistical Inference and its Applications (second edition). *Wiley Eastern Limited: New Delhi.*

[5] Roussas, G. G. (1972) Contiguity of Probability Measures, Cambridge University Press.

[6] Sen, P.K. and Singer, J.M. (1993) Large Sample Methods in Statistics, Chapman and Hall: New York - London.