International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Two-Piece Multivariate Lognormal Distribution

Dr. Parag Shah[a]*, Dr. C. D. Bhavsar[b]

[a] *Department of Statistics, H L College of Commerce,Ahmedabad,380009,India*

[b]*Dr. C. D. Bhavsar, Department of Statistics, Gujarat University, Ahmedabad, 380009, India*

[a] *Email: pbs1374@yahoo.co.in*

[b] *Email: chetna_bhavsar@yahoo.co.in*

Abstract

We have extended the univariate bilognormal distribution to two piece multivariate lognormal distribution. The density function of two piece multivariate lognormal distribution is given and the maximum likelihood estimates of the parameters when the variance - covariance matrix of two component multivariate lognormal distributions are proportional to each other are obtained. Its truncated distribution and estimation of parameters are also considered in this paper.

Keywords: Two piece multivariate lognormal distribution; maximum likelihood estimates; variance - covariance matrix.

1. Introduction

The Statistical theory based on the normal distribution has the advantage that the multivariate methods based on it are extensively developed and can be studied in an organized and systematic way. This is due not only to the need for such methods because they are of practical use, but also to the fact that normal theory is amenable to exact mathematical treatment. The suitable methods of analysis are mainly based on standard operations of matrix algebra; the distributions of many statistics involved can be obtained exactly or characterized, and in many cases optimum properties of procedures can be deduced. Although the bivariate normal distribution has been studied at the beginning of the nineteenth century, interest in multivariate distributions remained at a low level until it was stimulated by the work of [1] in the last quarter of the century.

---

* Corresponding author.

E-mail address: pbs1374@yahoo.co.in.

He did not, himself, introduce new firms of joint distribution, but he developed the idea of correlation and regression and focused attention on the need for greater knowledge of possible forms of multivariate distribution. [2] enunciated the theory of the multivariate normal distribution as a generalization of observed properties of samples.

We know that log transformation reduces the skewness and kurtosis of the distribution. Tarmast, Ghasem has defined multivariate lognormal distribution and its mean and covariance matrix are obtained and their estimates are calculated. He has shown that multivariate lognormal distribution can be applied in reliability study.

Similar to two-piece normal distribution, there are many advantages of the two-piece multivariate normal (TPMN) distribution over the other distributions those are used to handle asymmetry in data. One such advantage is that it can handle a wide range of skewness both positive and negative for more than one correlated variables. [3], [4] and [5] have studied some important properties of two-piece normal distribution.

In section 2, we have defined two piece multivariate lognormal (TPMLN) distribution. In section 3, we have given m.l.e's of the parameters. Truncated two piece multivariate lognormal (TTPMLN) distribution is defined in section 4 and its m.l.e's are given in section 5 of this paper.

2. The density function

The density function of two piece multivariate lognormal (TPMLN) distribution is given by

$$f(\underline{x})=\begin{cases}\left(\sqrt{\dfrac{2}{\pi}}\right)^{p}\dfrac{1}{\left(\prod_{i=1}^{p}x_i\right)}\left(|V_1|^{1/2}+|V_2|^{1/2}\right)^{-1}\exp\left\{-\dfrac{1}{2}\left(\log\underline{x}-\underline{\mu}\right)' V_1^{-1}\left(\log\underline{x}-\underline{\mu}\right)\right\},\log\underline{x}\le\underline{\mu}\\ \left(\sqrt{\dfrac{2}{\pi}}\right)^{p}\dfrac{1}{\left(\prod_{i=1}^{p}x_i\right)}\left(|V_1|^{1/2}+|V_2|^{1/2}\right)^{-1}\exp\left\{-\dfrac{1}{2}\left(\log\underline{x}-\underline{\mu}\right)' V_2^{-1}\left(\log\underline{x}-\underline{\mu}\right)\right\},\log\underline{x}>\underline{\mu}\end{cases}$$

……(2.1)

If V_2 is proportional to V_1 i.e. $V_1= V$ and $V_2 = kV_1, (k \neq 0)$ then the density function (2.1) reduces to

$$f(\underline{x})=\begin{cases}\left(\sqrt{\dfrac{2}{\pi}}\right)^{p}\dfrac{1}{\left(\prod_{i=1}^{p}x_i\right)}\left(1+k^{P}\right)^{-1}|V|^{-\frac{1}{2}}\exp\left\{-\dfrac{1}{2}\left(\log\underline{x}-\underline{\mu}\right)' V_1^{-1}\left(\log\underline{x}-\underline{\mu}\right)\right\},\log\underline{x}\le\underline{\mu}\\ \left(\sqrt{\dfrac{2}{\pi}}\right)^{p}\dfrac{1}{\left(\prod_{i=1}^{p}x_i\right)}\left(1+k^{P}\right)^{-1}|V|^{-\frac{1}{2}}\exp\left\{-\dfrac{1}{2}\left(\log\underline{x}-\underline{\mu}\right)' (kV)^{-1}\left(\log\underline{x}-\underline{\mu}\right)\right\},\log\underline{x}>\underline{\mu}\end{cases}$$

……..(2.2)

Taking $\underline{y} = \log \underline{x}$, (2.2) can be rewritten as

$$f(\underline{y}) = \begin{cases} \left(\sqrt{\frac{2}{\pi}}\right)^p \left(1+k^P\right)^{-1} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\underline{y}-\underline{\mu}\right)' V^{-1}\left(\underline{y}-\underline{\mu}\right)\right\}, \underline{y} \le \underline{\mu} \\ \left(\sqrt{\frac{2}{\pi}}\right)^p \left(1+k^P\right)^{-1} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\underline{y}-\underline{\mu}\right)' (kV)^{-1}\left(\underline{y}-\underline{\mu}\right)\right\}, \underline{y} > \underline{\mu} \end{cases}$$

……..(2.3)

3. Estimation of Parameters

According to the suggestion of [6] we define concomitants of multivariate order statistics as follows

Let $\left(y_{h1}, y_{h2}, \ldots, y_{hp-1}, y_{hp}\right)_{h=1}^{N}$ be an i id random sample from a p-variate distribution, where the random variables $y_1, y_2 \ldots\ldots y_{p-1}$ are absolutely continous. Denote the order of y_{h1} among $y_{11}, y_{21} \ldots\ldots y_{N1}$ by $R^1_{h:N}$, denote the order of y_{h2} among $y_{12}, y_{22} \ldots\ldots y_{N2}$ by $R^2_{h:N}$ and so on, denote the order of $y_{h_{p-1}}$ among $y_{1p-1}, y_{2p-1}, \ldots\ldots y_{Np-1}$ by $R^2_{h:N}$ and so on, denote the order of $y_{h_{p-1}}$ among $y_{1p-1}, y_{2p-1}, \ldots\ldots y_{Np-1}$ by $R^{p-1}_{h:N}$. We consider the random variable y_p given the ranks $R^1_{h:N} = r_1$, $R^2_{h:N} = r_2, \ldots\ldots, R^{p-1}_{h:N} = r_{p-1}$ as the concomitant of the r_1^{th} order statistics of y_1, r_2^{th} order of y_2 and so on r_{p-1}^{th} statistics of y_{p-1} and is denoted by $y_{ph\left[r_1, r_2, \ldots r_{p-1}, N\right]}$. For simplicity, we ignore the subscripts N and h and denote concomitant as $y_{p\left(r_1, r_2, \ldots r_{p-1}\right)}$. Let $\underline{y}_{(1)} = \underline{y}_{p1}\left(r_1^1, r_2^1, \ldots, r_{p-1}^1\right)$, $\underline{y}_{(2)} = \underline{y}_{p2}\left(r_1^2, r_2^2, \ldots, r_{p-1}^2\right), \ldots\ldots, \underline{y}_{(n)} = \underline{y}_{pn}\left(r_1^n, r_2^n, \ldots, r_{p-1}^n\right)$ be n concomitants of a multivariate order statistics from TPMLN distribution. On the assumption that $\underline{y}_{(t)} < \underline{\mu} < \underline{y}_{(t+1)} = \underline{y}_{(pt)}\left(r_1^t, r_2^t, \ldots r_{p-1}^t\right) < \underline{\mu} < \underline{y}_{pt+1}\left(r_1^{t+1}, r_2^{t+1}, \ldots r_{p-1}^{t+1}\right)$. For some $t\left(t = 1, 2 \ldots n-1\right)$, the likelihood function can be written as

$$\log L = \frac{np}{2}\log\left(\frac{2}{\pi}\right) - n\log\left(1+k^P\right) - \frac{n}{2}\log|V|$$
$$-\frac{1}{2}\sum\nolimits_1 \left(\underline{y}-\underline{\mu}\right)' V^{-1}\left(\underline{y}-\underline{\mu}\right) - \frac{1}{2}\sum\nolimits_2 \left(\underline{y}-\underline{\mu}\right)' \left(kV\right)^{-1}\left(\underline{y}-\underline{\mu}\right)$$

…….. (3.1)

Where $\sum_1 \left(\sum_2\right)$ denotes summation over all observations less than or equal to (greater than) $\underline{\mu}$

Differentiating $\log L$ with respect to $\underline{\mu}$ and equating it to zero, we get

$$V^{-1}\sum_{1}\left(\underline{y}-\underline{\mu}\right)+k^{-1}V^{-1}\sum_{2}\left(\underline{y}-\underline{\mu}\right)=0 \quad \text{........ (3.2)}$$

$$\therefore k\sum_{1}\left(\underline{y}-\underline{\mu}\right)+\sum_{2}\left(\underline{y}-\underline{\mu}\right)=0 \quad \text{....... (3.3)}$$

Let $\hat{\underline{\mu}}$ denote the m.l.e of μ and

let $$G\left(\underline{\mu}\right)=k\sum_{1}\left(\underline{y}-\underline{\mu}\right)+\sum_{2}\left(\underline{y}-\underline{\mu}\right) \quad \text{........(3.4)}$$

We can observe that,

(1) $G\left(\underline{y}_{(1)}\right)>0$ and $G\left(\underline{y}_{(n)}\right)<0$. Also $G\left(\underline{y}_{(t)}\right)>G\left(\underline{y}_{(t+1)}\right)$ for $t=1,2,....n-1$

(2) $\dfrac{\partial G\left(\underline{\mu}\right)}{\partial\underline{\mu}}=-kt-(n-t)<0$ for $\underline{y}_{(t)}<\underline{\mu}<\underline{y}_{(t+1)}$

As $G\left(\underline{\mu}\right)$ is continous, from (1) and (2) we conclude that $G\left(\underline{\mu}\right)$ is a decreasing functions of $\underline{\mu}$ and there exists only one real vector $\hat{\underline{\mu}}$, between $\underline{y}_{(1)}$ and $\underline{y}_{(n)}$ at which $G\left(\underline{\mu}\right)=0$. This implies that $\partial\log L/\partial\underline{\mu}$ is a decreasing function of $\underline{\mu}$ and $\hat{\underline{\mu}}$ is the only solution of the likelihood equation. One can note that even though $\partial\log L/\partial\underline{\mu}$ is not differentiable with respect to $\underline{\mu}$ at sample points, it is continous.

As $\log L$ at $\underline{\mu}=\underline{y}_{(1)}-\underline{h}$ is less than $\log L$ at $\underline{\mu}=\underline{y}_{(1)}$ and $\log L$ at $\underline{\mu}=\underline{y}_{(n)}+h$ is less than $\log L$ at $\underline{\mu}=\underline{y}_{(n)}$, the search for $\hat{\underline{\mu}}$ has to be restricted between $\underline{y}_{(1)}$ and $\underline{y}_{(n)}$.

We adopt the following procedure for obtaining $\hat{\underline{\mu}}$. Sequentially, one can start with t =1 and $\underline{\mu}=\underline{y}_{(1)}$, if $G\left(\underline{y}_{(1)}\right)>0$, then take t = 2 with $\underline{\mu}=\underline{y}_{(2)}$ again if $G\left(\underline{y}_{(2)}\right)>0$, then take t = 3 and so on. We can continue in this manner till $G\left(\underline{y}_{(t)}\right)<0$ for some t say t_1 Hence $\underline{y}_{(t_1-1)}<\hat{\underline{\mu}}<\underline{y}_{(t_1)}$.

Using the value of $t = t_1 - 1$ in (3.3), $\hat{\underline{\mu}}$ the value of $\underline{\mu}$ can be obtained as

$$\hat{\underline{\mu}} = \frac{k\Sigma_1 \log \underline{x} + \Sigma_2 \log \underline{x}}{k\ t + (n-t)} \quad \text{……..(3.5)}$$

Differentiating log likelihood function with respect to elements of *V* and equating it to zero, we get

$$\hat{V} = \frac{k\Sigma_1 \left(\log \underline{x} - \hat{\underline{\mu}}\right)\left(\log \underline{x} - \hat{\underline{\mu}}\right)' + \Sigma_2 \left(\log \underline{x} - \hat{\underline{\mu}}\right)\left(\log \underline{x} - \hat{\underline{\mu}}\right)'}{nk} \quad \text{……..(3.6)}$$

4. The density of its truncated distribution

The density function of truncated two piece multivariate lognormal (TTPMLN) distribution for $V_2 = KV_1$ is

$$f(\underline{y}) = \begin{cases} f_1\left(\underline{y}\right) = C_1^{-1}\left(\sqrt{\frac{2}{\pi}}\right)^p \left(1+k^P\right)^{-1} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\underline{y}-\underline{\mu}\right)' V^{-1}\left(\underline{y}-\underline{\mu}\right)\right\},\ \underline{A} \le \underline{y} \le \underline{\mu} \\ f_2\left(\underline{y}\right) = C_2^{-1}\left(\sqrt{\frac{2}{\pi}}\right)^p \left(1+k^P\right)^{-1} |V|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\underline{y}-\underline{\mu}\right)' (kV)^{-1}\left(\underline{y}-\underline{\mu}\right)\right\},\ \underline{\mu} \le \underline{y} < \underline{B} \end{cases}$$

…….. (4.1)

where $\underline{A}: p\times 1$ and $\underline{B}: p\times 1$ are vectors of truncation. C_1, C_2 are constants due to truncation and are given as

$$C_1 = 2^{p+1}\left(1+k^p\right)^{-1}\prod_{i=1}^{p}\left(\frac{1}{2} - \Phi\left(A_i^*\right)\right)$$

$$C_2 = 2^{p+1}\left(1+k^p\right)^{-1} k^{p/2}\prod_{i=1}^{p}\left(\Phi\left(\frac{B_i^*}{\sqrt{k}}\right) - \frac{1}{2}\right)$$

…….. (4.2)

with $\Phi\left(x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2 dt}$

using (4.2), (4.1) can be rewritten as

$$f(\underline{y}) = \begin{cases} C_1^* |V|^{-1/2} \exp\left\{-\frac{1}{2}\left(\underline{y}-\underline{\mu}\right)' V^{-1}\left(\underline{y}-\underline{\mu}\right)\right\}, \underline{A} \le y \le \underline{\mu} \\ C_2^* |V|^{-1/2} \exp\left\{-\frac{1}{2k}\left(\underline{y}-\underline{\mu}\right)' V^{-1}\left(\underline{y}-\underline{\mu}\right)\right\}, \underline{\mu} \le y \le \underline{B} \end{cases}$$

…….. (4.3)

Where $\left(C_1^*\right)^{-1} = 2\left(\sqrt{2\pi}\right)^p \prod_{i=1}^{p}\left(\frac{1}{2} - \Phi\left(A_i^*\right)\right)$ and

$$\left(C_2^*\right)^{-1} = 2\left(\sqrt{2\pi}\right)^p \prod_{i=1}^{p}\left(\Phi\left(\frac{B_i^*}{\sqrt{K}}\right) - \frac{1}{2}\right)$$

5. Estimation of parameters of TTPMLN distribution

Let $\underline{y}_{(1)}, \underline{y}_{(2)} .....\underline{y}_{(n)}$ be concomitants multivariate ordered sample from TTPMLN distribution. On the assumption that $\underline{y}_{(t)} < \underline{\mu} < \underline{y}_{(t+1)}$ for some t, (t=1,……..n-1) the likelihood function can be written as

$$LogL = t\log C_1^* + (n-t)\log C_2^* - \frac{n}{2}\log|V| \\ -\frac{1}{2}\Sigma_1\left(\underline{y}-\underline{\mu}\right)' V^{-1}\left(\underline{y}-\underline{\mu}\right) - \frac{1}{2k}\Sigma_2\left(\underline{y}-\underline{\mu}\right)' V^{-1}\left(\underline{y}-\underline{\mu}\right)$$

…….. (5.1)

where $\Sigma_1\left(\Sigma_2\right)$ denotes summation over all observations lying between $\underline{A}$ and $\underline{\mu}$ (lying between $\underline{\mu}$ and $\underline{B}$).

Differentiating logL with respect to $\underline{\mu}$ and equating it to zero, we get

$$V^{-1}\Sigma_1\left(\underline{y}-\underline{\mu}\right) + k^{-1}V^{-1}\Sigma_2\left(\underline{y}-\underline{\mu}\right) = 0$$

$$\therefore k\Sigma_1\left(\underline{y}-\underline{\mu}\right) + \Sigma_2\left(\underline{y}-\underline{\mu}\right) = 0$$

…….. (5.2)

Applying the same procedure as TPMLN distribution, we can choose some $t = t_1$ such that $\underline{y}_{(t_1-1)} < \hat{\underline{\mu}} < \underline{y}_{(t_1)}$ and similar to section 3, we get

$$\hat{\underline{\mu}} = \frac{\Sigma_1 \log \underline{x} + \Sigma_2 \log \underline{x}}{k\ t + (n-t)}$$

…….. (5.3)

Differentiating log likelihood function with respect to element of V and equating it to zero, we get

$$\hat{V} = \frac{\sum_1 \left(\log \underline{x} - \underline{\hat{\mu}}\right)\left(\log \underline{x} - \underline{\hat{\mu}}\right)' + \sum_2 \left(\log \underline{x} - \underline{\hat{\mu}}\right)\left(\log \underline{x} - \underline{\hat{\mu}}\right)'}{nk} \quad \text{........ (5.4)}$$

where $\sum_1 \left(\sum_2\right)$ denotes summation over all observations lying between $\underline{A}$ and $\underline{\mu}$ (lying between $\underline{\mu}$ and $\underline{B}$) which are different from those in equation 3.5.

References:

[1] Galton F. "Typical laws of Heredity in Man", *Lecture at the Royal Institution of Great Britain,* 1877.

[2] Galton, F. "Natural Inheritance", *Mac Milan, London,* pp *1.2 - 2.5, 1889.*

[3] John, S. "The parameter two piece normal family of distribution and its fitting", *Communications in Statistic theory and Methods, 11, 8,* pp *879-885*, 1982.

[4] Kimber, A.C. "Methods for the two-piece normal distribution", *Communications in Statistics, Theory and Methods, 14, 1,* pp *235-245*, 1985.

[5] Nabar, S.P. and Deshmukh, S.C. "On estimation of parameters of bilognormal distribution", *JISA, 40, 1,* pp *59-70*, 2001.

[6] Xinlie Wang, Lynne Stokes, Johan Lim and Min Chew "Concomitants of Multivariate order statistics with applications to Judgment Post stratification", *JASA, 101, 476,* pp *1693-1704, 2006.*