Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.2 pp. 377~383

Large Sample Test for Independence in the Bivariate Pareto Model with Censored Data

Jang Sik Cho[1)] · Jea Man Lee[2)] · Woo Dong Lee[3)]

Abstract

In this paper, we consider two components system in which the lifetimes follow the bivariate Pareto model with random censored data. We assume that the censoring time is independent of the lifetimes of the two components. We develop large sample tests for testing independence between two components. Also we present simulated study which is the test based on asymptotic normal distribution in testing independence.

Keywords : Bivariate Pareto model; Random censored data; Independence; Maximum likelihood estimator.

1. Introduction

In many studies of two components system data, the component lifetimes were assumed to be statistically independent for the sake of simplicity of mathematical treatment. However, the assumption of independence is unrealistic as in many two components systems the component life lengths have a well-defined dependence structure.

Several bivariate models based on exponential distributions have been derived. The distribution of Freund(1961) is based on the joint survival of two components which initailly are independently on test with the exponential life distributions with

1) Corresponding Author. Associate Professor, Department of Statistical Information Science, Kyungsung University, Pusan, 608-736, Korea
E-mail : jscho@star.ks.ac.kr

2) Professor, Department of Information Statistics, Andong National University, Andong, 760-749, Korea

3) Department of Information Science, Kyungsan University, Kyungsan, 712-715, Korea

parameters α and β respectively. Failure of one component reduces the additional mean life of the remaining component by increasing either α to α' or β to β'. The distribution of Block and Basu(1974) is a reparameterization of the special case of the Freund family having a bivariate generalization of the memoryless property. It can also be derived in the context of the fatal shock model giving the bivariate exponential distribution of Marshall and Olkin(1967).

On the other hand, Lindley and Singpurwalla(1986) proposed bivariate Pareto(BVP) model in the modelling of lifetimes of two components systems working in a changing environment. They considered the distribution of life lengths measured in a laboratory environment as independent exponential distributions proved that, when they work in a different environment which may be harsher, the same or gentler than the original, the resulting density of life lengths has a BVP model. Bandyapadhyay and Basu(1990), and Veenus and Nair(1994) obtained some BVP models corresponding to some well known bivariate exponential models. Jeevanand(1997) obtained Bayes estimator of the reliability of stress-strength in BVP model. Hanagal(1996) introduced a new multivariate Pareto model including interesting properties. Cho, Cho and Cha(2003) obtained system reliability from stress-strength relationship.

Above all authors considered complete sample cases. Now we consider the following situations. Assume that there occure censored observations in lifetime. For example, there is a system which can not measure some characteristics of the lifetime of an item above a certain value. In this case, real lifetime may exceed the measured value, but we do not know the exact lifetime.

In this paper, we derive maximum likelihood estimators for the parameters in the bivariate Pareto model with random censored data as extension of complete data. And we construct large sample test for independence between the lifetimes of both components. Also we present simulated study which is the test based on the asymptotic normal distribution in test of independence.

2. Preliminaries

Let random variables (X, Y) be lifetimes of two components that follow a BVP model with parameter $(\theta_1, \theta_2, \theta_3, \beta)$. Then the joint probability density function is given as

$$f(x, y : \theta_1, \theta_2, \theta_3, \beta)$$
$$= \begin{cases} \theta_1(\theta_2+\theta_3)\beta^{\theta}x^{-\theta_1-1}y^{-(\theta_2+\theta_3)-1}, & \beta < x < y < \infty, \\ \theta_2(\theta_1+\theta_3)\beta^{\theta}x^{-(\theta_1+\theta_3)-1}y^{-(\theta_2+\theta_3)-1}, & \beta < y < x < \infty, \\ \theta_3\beta^{\theta}x^{-\theta-1}, \beta < x = y < \infty, \end{cases} \tag{1}$$

where $\theta_1, \theta_2, \theta_3 > 0$ and $\theta = \theta_1+\theta_2+\theta_3$.

We call (1) as a BVP type 2 if $\beta = 1$ and as a BVP type 1 elsewhere.

The above BVP model is not absolutely continuous with respect to Lebesgue measure on R^2. That is, there is provision for simultaneous failure of the both components $P[X=Y]=\theta_3/\theta$. And the random variables X and Y are independent if and only if $\theta_3=0$.

Suppose that there are n two components systems under study and ith pair of the components have life time (x_i, y_i) and a censoring time t_i. We assume that the censoring time t_i has the Pareto distribution with parameter (η, β) which is independent of $f(x,y)$ and the probability density function $f(t_i; \beta, \eta)=\eta\beta^{\eta}t_i^{-\eta-1}$, $t_i > \beta$. Then ith observed lifetime (x_i, y_i) is given by

$$(x_i, y_i) = \begin{cases} (x_i, y_i), & \text{if } \max(x_i, y_i) < t_i \\ (x_i, t_i), & \text{if } x_i < t_i < y_i \\ (t_i, y_i), & \text{if } y_i < t_i < x_i \\ (t_i, t_i), & \text{if } t_i < \min(x_i, y_i). \end{cases} \tag{2}$$

Let $I(\cdot)$ be an indicator function. And we define $n_j\ (j=1,\cdots,6)$ as follows;

$$n_1 = \sum_{i=1}^{n} I(x_i < y_i < t_i),\quad n_2 = \sum_{i=1}^{n} I(y_i < x_i < t_i),\quad n_3 = \sum_{i=1}^{n} I(x_i = y_i < t_i),$$

$$n_4 = \sum_{i=1}^{n} I(x_i < t_i < y_i),\quad n_5 = \sum_{i=1}^{n} I(y_i < t_i < x_i),\quad n_6 = \sum_{i=1}^{n} I(\min(x_i, y_i) > t_i).$$

Then the expected value of $n_j\ (j=1,\cdots,6)$ can be obtained as follows;

$$E(n_1)=n\left(\frac{\theta_1}{\theta}+\frac{\eta(\theta_2+\theta_3)}{\theta(\theta+\eta)}-\frac{\eta}{(\theta_2+\theta_3+\eta)}\right),$$

$$E(n_2)=n\left(\frac{\theta_2}{\theta}+\frac{\eta(\theta_1+\theta_3)}{\theta(\theta+\eta)}-\frac{\eta}{(\theta_1+\theta_3+\eta)}\right), E(n_3)=\frac{n\theta_3}{\theta+\eta},$$

$$E(n_4)=n\theta\left(\frac{1}{\eta+\theta_2+\theta_3}-\frac{1}{\theta+\eta}\right), E(n_5)=n\theta\left(\frac{1}{\eta+\theta_1+\theta_3}-\frac{1}{\theta+\eta}\right),\ E(n_6)=\frac{n\eta}{\theta+\eta}.$$

Let $m_1=n_1+n_4$, $m_2=n_2+n_5$, $m_3=n_3$ and $m_4=n_6$. Then (m_1, m_2, m_3, m_4) is multinomial distribution with parameter $\left(n, \frac{\theta_1}{\theta+\eta}, \frac{\theta_2}{\theta+\eta}, \frac{\theta_3}{\theta+\eta}, \frac{\eta}{\theta+\eta}\right)$.

Now the likelihood function of the sample of size n is given by

$$L = \theta_1^{n_1+n_4}\cdot\theta_2^{n_2+n_5}\cdot\theta_3^{n_3}\cdot(\theta_1+\theta_3)^{n_2}\cdot(\theta_2+\theta_3)^{n_1}\cdot\beta^{n\theta}\cdot\left[\prod_{i=1}^{n}x_i\right]^{-(\theta_1+1)}$$

$$\cdot\left[\prod_{i=1}^{n}y_i\right]^{-(\theta_2+1)}\cdot\left[\prod_{i=1}^{n}max(x_i, y_i)\right]^{-\theta_3}\cdot\left[\prod_{\{i\mid x_i=y_i\}}x_i\right]^{-1}$$

$$\cdot\ \eta^{n_4+n_5+n_6} \cdot \left[\prod_{i=1}^{n} \max(x_i, y_i) \right]^{-\eta}. \tag{3}$$

In this paper, we focus only on BVP type 2 model. Then the likelihood equations are given by

$$\frac{n_1+n_4}{\theta_1} + \frac{n_2}{\theta_1+\theta_3} - \sum_{i=1}^{n} \log(x_i) = 0, \tag{4}$$

$$\frac{n_2+n_5}{\theta_2} + \frac{n_1}{\theta_2+\theta_3} - \sum_{i=1}^{n} \log(y_i) = 0, \tag{5}$$

$$\frac{n_3}{\theta_3} + \frac{n_2}{\theta_1+\theta_3} + \frac{n_1}{\theta_2+\theta_3} - \sum_{i=1}^{n} \log(\max(x_i, y_i)) = 0, \tag{6}$$

$$\frac{n_4+n_5+n_6}{\eta} - \sum_{i=1}^{n} \log(\max(x_i, y_i)) = 0. \tag{7}$$

The likelihood equations (4-(6) are not easy to solve. But we can obtain MLE's ($\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$) by either Newton-Raphson procedure or Fisher's method of scoring. The Fisher information matrix is given by

$$I(\theta_1, \theta_2, \theta_3) = E\left[\frac{\partial^2 logL}{\partial\theta_i \partial\theta_j} \right] = n((I_{ij}))\ ;\ i,\ j = 1,2,3, \tag{8}$$

where $I_{11} = \left(\frac{E(n_1)+E(n_4)}{n\theta_1^2} + \frac{E(n_2)}{n(\theta_1+\theta_3)^2} \right),\ I_{12} = 0,\ I_{13} = \frac{E(n_2)}{n(\theta_1+\theta_3)^2},$

$$I_{22} = \left(\frac{E(n_2)+E(n_5)}{n\theta_2^2} + \frac{E(n_1)}{n(\theta_2+\theta_3)^2} \right),\ I_{23} = \frac{E(n_1)}{n(\theta_2+\theta_3)^2},$$

$$I_{33} = \left(\frac{E(n_1)}{n(\theta_2+\theta_3)^2} + \frac{E(n_2)}{n(\theta_1+\theta_3)^2} + \frac{E(n_3)}{n\theta_3^2} \right).$$

Thus $\sqrt{n}(\hat{\underline{\theta}} - \underline{\theta})$ has asymptotic trivariate normal distribution with mean vector zero and covariance matrix $I^{-1}(\underline{\theta}) = \frac{1}{n}((I^{ij}))\ ;\ i,j = 1,2,3$. Here, $\hat{\underline{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ and $\underline{\theta} = (\theta_1, \theta_2, \theta_3)$.

3. Large Sample Test of Independence

In this section, we consider two large sample tests of hypothesis for independence and symmetry between two components based on $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ and (m_1, m_2, m_3, m_4). That is, we construct a large sample test of null hypothesis $H_o: \theta_3 = 0$ for independence. The exact distribution of $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ is difficult to obtain but their asymptotic normal distribution can be obtained using the results of section 2.

Note that when $\theta_3 = 0$ we will have only one sided alternative $H_1: \theta_3 > 0$ and the test is equivalent to test for independence of X and Y. Now $\hat{\theta}_3$ has asymptotic normal distribution with mean θ_3 and variance I^{33}/n but I^{33} depends on the unknown parameters $(\theta_1, \theta_2, \theta_3)$. We estimate it from the MLE's of $(\theta_1, \theta_2, \theta_3)$ and construct the test statistic $T_1 = \sqrt{n} \cdot \hat{\theta}_3 / \sqrt{\hat{I}^{33}}$ which has asymptotic normal distribution. For $H_1: \theta_3 > 0$, we reject H_o with significance level α if

$$T_1 = \sqrt{n} \cdot \hat{\theta}_3 / \sqrt{\hat{I}^{33}} > z_{1-\alpha}, \quad (9)$$

where z_α is the $100 \cdot \alpha$ percentile of standard normal distribution.

We can also obtain a large sample test for $H_o: \theta_3 = 0$ based on m_3 which is binomial($n, \lambda_3/\lambda$) and use the studentized test statistic as follows;

$$T_2 = \sqrt{n} m_3 / \sqrt{m_3(n - m_3)}, \quad (10)$$

which is asymptotic normal distribution. Hence for $H_1: \theta_3 > 0$, we reject H_o with significance level α if

$$T_2 = \sqrt{n} m_3 / \sqrt{m_3(n - m_3)} > z_{1-\alpha}. \quad (11)$$

4. Numerical Example

In this section, we present a numerical example by giving a data set which is generated by computer. We generate a random samples of size 30 from BVP with parameter ($\theta_1 = 1.0$, $\theta_2 = 1.3$, $\theta_3 = 0.2$). Also we generate random censored data of size 30 corresponding the lifetimes from Pareto with parameters $\eta = 1$, $\beta = 0.4$. Then the generated data is not symmetric and not independent between X and Y. That is, true hypotheses is $H_1: \theta_3 > 0$ for independence. The data is given Table 1. In Table 1, * indicates censored data.

<Table 1> Generated sample (x, y) from BVP distribution

i	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	2.3039*	1.0785	11	1.1728	1.9171	21	1.2657	2.1435
2	1.0729	4.5355*	12	1.4271	1.3862	22	1.3025	1.2518
3	13.4415	1.2023	13	2.2609	1.6771	23	2.7157	1.0985
4	2.6248	22.1966	14	1.3947	1.0878	24	1.3517	1.2736
5	89.0278*	1.2402*	15	1.1333	7.1660	25	1.1741	1.1428
6	3.0337*	27.0862*	16	38.7218*	2.1062	26	4.6778*	1.8566
7	12.6393*	1.4266	17	9.3936	1.3264	27	1.0978	1.5769
8	1.3517	1.4697	18	1.0595	1.0595	28	1.1372	2.2099*
9	1.1918	1.2322	19	3.4351	3.4351	29	2.5370*	7.6270*
10	1.5415	1.2668	20	1.5701	1.6186	30	1.8018	2.0787

From <Table 1>, MLE's of the parameters in BVP model are given by

$\hat{\theta}_1 = 0.9435$, $\hat{\theta}_2 = 1.3935$, $\hat{\theta}_3 = 0.1701$ and m_1, m_2, m_3, m_4 are given by 11, 14, 2, and 3, respectively.

Then values of test statistics and p-values are given by <Table 2>.

<Table 2> The values of test statistics and p-values

Test statistics / Values	T_1	T_2
The value of test statistics	7.6655	1.4638
p-value	0.0000	0.0716

From <Table 2>, we reject $H_o: \theta_3 = 0$ for independence based on test statistics T_1 with significance level 0.01. But we can not reject $H_o: \theta_3 = 0$ for independence based on test statistics T_2 with significance level 0.01.

Hence, we note that test statistics T_1 based on MLE's perform better than T_2 based on m_1, m_2, m_3, m_4.

References

1. Bandyapadhyay, D. and Basu, A.P. (1990). On generalization of a model by Lindley and Singpurwallam, *Advanced Applied Probability*, 22, 498-500.
2. Block, H. W. and Basu, A. P. (1974). A Continuous Bivariate Exponential Extension, *Journal of the American Statistical Association*, 69, 1031-1037.
3. Cho, J. S., Cho, K. H. and Cha, Y. J. (2003), System Reliability From Stress-Strength Relationship in Bivariate Pareto Distribution, *Journal of the Korean Data & Information Science Society*, Vol. 14(1), 113-118.
4. Freund, J. E. (1961). A bivariate extension of the exponential distribution, *Journal of the American Statistical Association*, 971-977.
5. Hanagal, D.D. (1996). A multivariate Pareto distribution, *Communication in Statistics-Theory and Methods*, 25(7), 1471-1488.
6. Jeevanand, E.S. (1997). Bayes estimation of $P(X_2 < X_1)$ for a bivariate Pareto distribution, *The Statistician*, 46(1), 93-99.
7. Lehmann, E.L.(1983), *Theory of Point Estimation*, John Wiley and Sons. New York.
8. Lindley, D.V. and Singpurwalla, N.D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment, *Journal of Applied Probability*, 23, 418-431.

9. Marshall, A.W. and Olkin, I.(1967), A Multivariate Exponential Distribution, *Journal of the American Statistical Association*, 62, 30-44.
10. Veenus, P. and Nair, K.R.M. (1994). Characterization of a bivariate Pareto distribution, *Journal of Indian Statistical Association*, 32, 15-20

[received date : Mar. 2003, accepted date : May. 2003]