Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 3, pp. 843 ~850

Bayesian Multiple Comparison of Bivariate Exponential Populations based on Fractional Bayes Factor

Jang Sik Cho[1)] · Kil Ho Cho[2)] · Seung Bae Choi[3)]

Abstract

In this paper, we consider the Bayesian multiple comparisons problem for K bivariate exponential populations to make inferences on the relationships among the parameters based on observations. And we suggest the Bayesian procedure based on fractional Bayes factor when noninformative priors are applied for the parameters. Also, we give a numerical examples to illustrate our procedure.

Keywords : Bayesian multiple comparison, Fractional Bayes factor, Noninformative priors, Posterior probability

1. Introduction

In many cases of two components system, the lifetimes of the components are assumed independent for convenience of computation. However it is more realistic to assume some form of dependence among components in many life testing situations. Let's consider a system which functions only as long as at least one of two identical or very similar components functions. Initially let the two components be independently on test with life distributions that are exponential with parameters λ. Failure of one changes the life distribution of the other to

---

1) (Corresponding Author) Associate Professor, Department of Informational Statistics, Kyungsung University, Busan, 608-736, Korea.
E-Mail : jscho@ks.ac.kr

2) Professor, Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea
E-mail : khcho@knu.ac.kr

3) Assistant Professor, Department of Information Statistics, Dongeui University, Busan, 614-714, Korea
E-mail : csb4851@deu.ac.kr

exponential with parameter $\lambda\zeta > 0$, where $\zeta = 1$ implies the independence of the two components lives. For $\zeta > 1$ the workload of the remaining component is increased, thereby decreasing the mean life. We call ζ as dependence parameter. For this bivariate exponential model, Weier(1981) and Lee et.al.(1998) obtained Bayes estimators of parameters and reliability.

For testing equality of dependence parameters in two independent bivariate exponential populations, classical tests such as approximate test are widely used. But the test of equality of dependence parameters more than three populations relies on likelihood ratio test statistic which is distributed as approximately χ^2-distribution. And classical tests only decide whether the null hypothesis, commonly the equality of the parameters, will be rejected or not. When the null hypothesis is rejected, we don't know which hypothesis is best for describing the equality of parameters. But Bayesian approach to resolve the multiple comparisons problem selects the model with the highest posterior probability. And we can compute all the posterior probabilities of the hypotheses under consideration.

In this paper, we focus on Bayesian multiple comparisons for K bivariate exponential populations based on Bayes factor. In many cases, noninformative priors for the parameters are used. Since noninformative priors are typically improper, the priors are only up to arbitrary constants which affects the values of Bayes factors.

Berger and Pericchi(1996) introduced the intrinsic Bayes factor(IBF) using a data splitting idea, which would eliminate the arbitrariness of improper priors. O'Hagan(1995) proposed the fractional Bayes factor(FBF) to remove the arbitrariness. These approaches have shown to be quite useful in several statistical areas. Cho(2006) proposed the Bayesian procedure for independence test based on FBF in bivariate exponentail model.

In this paper, we extend the results of Cho(2006) and consider Bayesian multiple comparisons for K bivariate exponential populations based on FBF. Also we propose the procedure for Bayesian multiple comparisons of dependence parameters based on FBF when improper priors are applied for the parameters. Finally, we give some numerical examples to illustrate our procedure.

2. Preliminaries

Let $M_1, \cdots, M_N$ be models under consideration. And let random sample $(\boldsymbol{X}, \boldsymbol{Y}) = ((X_1, Y_1), \cdots, (X_n, Y_n))$ have probability density function $f((\boldsymbol{x}, \boldsymbol{y}) \mid \Theta_i)$ under model M_i, $i = 1, \cdots, N$. The parameter vectors Θ_i are unknown. Let $\pi_i(\Theta_i)$ be the prior distribution of model M_i, and let p_i be the prior probabilities of model M_i. Then the posterior probability that the model M_i is true is given as

$$P(M_i \mid \boldsymbol{x}, \boldsymbol{y}) = \left(\sum_{j=1}^{N} \frac{p_j}{p_i} B_{ji} \right)^{-1},$$

where B_{ij} is the Bayes factor of model M_j to model M_i defined by

$$B_{ji} = \frac{m_j(\boldsymbol{x}, \boldsymbol{y})}{m_i(\boldsymbol{x}, \boldsymbol{y})} = \frac{\int_{\Theta_j} f(\boldsymbol{x}, \boldsymbol{y} \mid \Theta_j) \pi_j(\Theta_j) d\Theta_j}{\int_{\Theta_i} f(\boldsymbol{x}, \boldsymbol{y} \mid \Theta_i) \pi_i(\Theta_i) d\Theta_i}. \tag{1}$$

The B_{ji} interpreted as the comparative support of the data for the model j to i . The computation of B_{ji} needs specification of the prior distribution $\pi_i(\Theta_i)$ and $\pi_j(\Theta_j)$. Usually, one can use the noninformative prior which is improper. Let π_i^N be the noninformative prior for model M_i. Then the use of improper priors $\pi_i^N(\cdot)$ in (1) causes the B_{ji} to contain arbitrary constants.

To solve this problem, O'Hagan(1995) proposed the procedure for Bayesian testing and model selection problem based on FBF as follow. B_{ij} based on noninformative prior $\pi_i^N(\cdot)$ is given as

$$B_{ji}^N = \frac{m_j^N(\boldsymbol{x}, \boldsymbol{y})}{m_i^N(\boldsymbol{x}, \boldsymbol{y})} = \frac{\int_{\Theta_j} f(\boldsymbol{x}, \boldsymbol{y} \mid \Theta_j) \pi_j^N(\Theta_j) d\Theta_j}{\int_{\Theta_i} f(\boldsymbol{x}, \boldsymbol{y} \mid \Theta_i) \pi_i^N(\Theta_i) d\Theta_i}. \tag{2}$$

Hence the FBF of model M_j versus model M_i is given as

$$B_{ji}^F = \frac{q_j(b, \boldsymbol{x}, \boldsymbol{y})}{q_i(b, \boldsymbol{x}, \boldsymbol{y})}, \tag{3}$$

where $q_i(b, \boldsymbol{x}, \boldsymbol{y}) = \dfrac{\int_{\Theta_i} f_i(\boldsymbol{x}, \boldsymbol{y} \mid \Theta_i) \pi_i^N(\Theta_i) d\Theta_i}{\int_{\Theta_i} f_i^b(\boldsymbol{x}, \boldsymbol{y} \mid \Theta_i) \pi_i^N(\Theta_i) d\Theta_i}$ and b specifies a fraction of the likelihood which is to be used as a prior density. One frequently suggested choice is b = m/n, where m is the size of the minimal training sample.

Let's consider K populations with parameters $\Theta = (\Theta_1, \Theta_2, \cdots, \Theta_K)$. Let $(\boldsymbol{X_i}, \boldsymbol{Y_i}) = ((X_{i1}, Y_{i1}), \cdots, (X_{\in_i}, Y_{\in_i}))$ be a $n_i \times 1$ vector of independent observations on Θ_i with density $f(x_{ij}, y_{ij} \mid \Theta_i)$, $i = 1, \cdots, K$, $j = 1, \cdots, n_i$. Then the likelihood function for Θ given $(\boldsymbol{X}, \boldsymbol{Y}) = ((X_1, Y_1), \cdots, (X_n, Y_n))$ is

$$L(\Theta \mid \boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{K} \prod_{j=1}^{n_i} f(x_{ij}, y_{ij} \mid \Theta_i). \tag{4}$$

The multiple comparisons of K populations is to make inferences concerning relationships among the Θ_i's based on $(\boldsymbol{X}, \boldsymbol{Y})$.

Let $\Omega = \{(\Theta_1, \Theta_2, \cdots, \Theta_K) : \Theta_i \in R,\ i = 1, 2, \cdots, K\}$ be the K-dimensional parameter space. Equality and inequality relationships among the Θ_i's induce statistical hypotheses such that subsets of Θ, i.e., $M_1 : \Omega_1 = \{\Theta_i ; \Theta_1 = \Theta_2 = \cdots = \Theta_K\}$, $M_1 : \Omega_1 = \{\Theta_i ; \Theta_1 \neq \Theta_2 = \cdots = \Theta_K\}$ and so on up to $M_N : \Omega_N = \{\Theta_i ; \Theta_1 \neq \Theta_2 \neq \cdots \neq \Theta_K\}$. The hypotheses $M_r : \Omega_r,\ r = 1, 2, \cdots, N$, are disjoint, and $\cup_{r=1}^{N} \Omega_r = \Omega$.

Each hypothesis can classified $r(r = 1, \cdots, K)$ distinct groups. Let $\Theta_1^*, \cdots, \Theta_r^*$ denote the set of distinct Θ_i's, where r is the number of distinct elements in the vector Ω. We need to define the configuration notation.

Definition 1 ***(Configuration).*** The configuration $S = \{S_1, \cdots, S_K\}$ determines a classification of Θ into r distinct groups or clusters. Write K_j for the set of indices of parameters in group j, $K_j = \{i : S_i = j\}$. Let $n_{K_j} = \{n_i \mid i \in K_j\}$ be the index set of observations and Θ_j^* be the common parameter value for K_j.

There is a one to one correspondence between hypotheses and configurations. Therefore the Bayes factor for multiple comparisons can easily compute by this configuration notation.

Suppose that a model is classified r distinct groups. Then the likelihood function is given by

$$L(\Theta_1^*, \cdots, \Theta_r^* \mid \boldsymbol{x}, \boldsymbol{y}) = \prod_{t=1}^{r} \prod_{\{i : i \in K_t\}} \prod_{j=1}^{n_i} f(x_{ij}, y_{ij} \mid \Theta_t). \tag{5}$$

Since the noninformative prior for the model is $\pi_r^N(\Theta_1^*, \cdots, \Theta_r^*)$, the FBF is given by

$$q(b, \boldsymbol{x}, \boldsymbol{y}) = \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(\Theta_1^*, \cdots, \Theta_r^* \mid \boldsymbol{x}, \boldsymbol{y}) \cdot \pi_r^N(\Theta_1^*, \cdots, \Theta_r^*) d\Theta_1^* \cdots d\Theta_r^*}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L^b(\Theta_1^*, \cdots, \Theta_r^* \mid \boldsymbol{x}, \boldsymbol{y}) \cdot \pi_r^N(\Theta_1^*, \cdots, \Theta_r^*) d\Theta_1^* \cdots d\Theta_r^*}. \tag{6}$$

Thus if a model M_i is classified r_i distinct groups and a model M_j is classified r_j distinct groups then the FBF of M_j versus M_i is given by

$$B^F_{ji} = \frac{q_j(b, \boldsymbol{x}, \boldsymbol{y})}{q_i(b, \boldsymbol{x}, \boldsymbol{y})},$$

where $$q_i(b, \boldsymbol{x}, \boldsymbol{y}) = \frac{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(\Theta^*_1,\cdots,\Theta^*_{r_i} \mid \boldsymbol{x}, \boldsymbol{y}) \cdot \pi^N_{r_i}(\Theta^*_1,\cdots,\Theta^*_{r_i}) d\Theta^*_1 \cdots d\Theta^*_{r_i}}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L^b(\Theta^*_1,\cdots,\Theta^*_{r_i} \mid \boldsymbol{x}, \boldsymbol{y}) \cdot \pi^N_{r_i}(\Theta^*_1,\cdots,\Theta^*_{r_i}) d\Theta^*_1 \cdots d\Theta^*_{r_i}}.$$

Hence the FBF for all comparisons can computed by equation (2). Using these FBF's, we can calculate the posterior probability for model M_i, $i=1,\cdots,K$. Thus, we can select the hypothesis with highest posterior probability in Bayesian multiple comparisons based on FBF.

3. Bayesian Multiple Comparisons

Let $(\boldsymbol{X_i}, \boldsymbol{Y_i}) = ((X_{i1}, Y_{i1}),\cdots,(X_{\in_i}, Y_{\in_i}))$ be a random sample from a bivariate exponential population with parameter vector $\Theta_i = (\lambda_i, \zeta_i)$. Then the joint probability density function for (X_{ij}, Y_{ij}) is given as

$$f(x_{ij}, y_{ij} : \lambda_i, \zeta_i) = 2\zeta_i\lambda_i^2 \exp(-2\lambda_i x_{ij} - \lambda_i\zeta_i y_{ij}),\ \ x_{ij}, y_{il} > 0,\ \lambda_i, \zeta_i > 0.$$

Suppose that a model M_i classified r distinct groups. Then the noninformative prior for $(\Theta^*_1,\cdots,\Theta^*_r)$ is

$$\pi^N_r(\Theta^*_1,\cdots,\Theta^*_r) = \frac{1}{c},\ \ 0<\Theta^*_1<\infty,\cdots,0<\Theta^*_r<\infty, \tag{7}$$

where c is a fixed constant.
And likelihood function is

$$L(\Theta^*_1,\cdots,\Theta^*_r \mid \boldsymbol{x}, \boldsymbol{y}) = \prod_{t=1}^{r}\left\{(2\cdot\lambda_t^2\cdot\zeta_t)^{\mid K_t \mid} \cdot \exp\left(-\sum_{i\in K_t}\sum_{j=1}^{n_i}(2\cdot\lambda_i\cdot x_{ij} + \lambda_i\cdot\zeta_i\cdot y_{ij})\right)\right\}, \tag{8}$$

where $\mid K_t \mid$ is the number of the set of indices K_t.
Then the element of the FBF is computed as follows;

$$\int_0^{\infty}\cdots\int_0^{\infty} L(\Theta^*_1,\cdots,\Theta^*_r \mid \boldsymbol{x}, \boldsymbol{y})\pi^N_r(\Theta^*_1,\cdots,\Theta^*_r)d\Theta^*_1\cdots d\Theta^*_r$$
$$= c\cdot\prod_{t=1}^{r}\Gamma(\mid K_t\mid)\cdot\Gamma(\mid K_t \mid +1)\cdot\left(\sum_{i\in K_t}\sum_{j=1}^{n_i}x_{ij}\right)^{-\mid K_t\mid}\cdot\left(\sum_{i\in K_t}\sum_{j=1}^{n_i}y_{ij}\right)^{-\mid K_t\mid -1}$$

and

$$\int_0^{\infty}\cdots\int_0^{\infty} L^{b}(\Theta^*_1,\cdots,\Theta^*_r \mid \boldsymbol{x}, \boldsymbol{y})\pi^N_r(\Theta^*_1,\cdots,\Theta^*_r)d\Theta^*_1\cdots d\Theta^*_r$$

$$= c\cdot\prod_{t=1}^{r}\Gamma(b\cdot|K_t|)\cdot\Gamma(b\cdot|K_t|+1)\cdot\left(b\sum_{i\in K_t}\sum_{j=1}^{n_i}x_{ij}\right)^{-b\cdot|K_t|}\cdot\left(b\sum_{i\in K_t}\sum_{j=1}^{n_i}y_{ij}\right)^{-b\cdot|K_t|-1}.$$

Hence, $q(b,\boldsymbol{x},\boldsymbol{y})$ is given as

$$q(b,\boldsymbol{x},\boldsymbol{y})=\prod_{t=1}^{r}\frac{\Gamma(|K_t|)\cdot\Gamma(|K_t|+1)\cdot\left(b\sum_{i\in K_t}\sum_{j=1}^{n_i}x_{ij}\right)^{b\cdot|K_t|}\cdot\left(b\sum_{i\in K_t}\sum_{j=1}^{n_i}y_{ij}\right)^{b\cdot|K_t|+1}}{\Gamma(b\cdot|K_t|)\cdot\Gamma(b\cdot|K_t|+1)\cdot\left(\sum_{i\in K_t}\sum_{j=1}^{n_i}x_{ij}\right)^{|K_t|}\cdot\left(\sum_{i\in K_t}\sum_{j=1}^{n_i}y_{ij}\right)^{|K_t|+1}}. \quad (9)$$

Thus if a model M_i is classified r_i distinct groups and a model M_j is classified r_j distinct groups then the FBF of M_j versus M_i is given by

$$B^F_{ji}(\boldsymbol{x},\boldsymbol{y})=\frac{q_j(b,\boldsymbol{x},\boldsymbol{y})}{q_i(b,\boldsymbol{x},\boldsymbol{y})},$$

where

$$q_i(b,\boldsymbol{x},\boldsymbol{y})=\prod_{M_i:t=1}^{r_i}\frac{\Gamma(|K_t|)\cdot\Gamma(|K_t|+1)\cdot\left(b\sum_{i\in K_t}\sum_{j=1}^{n_i}x_{ij}\right)^{b\cdot|K_t|}\cdot\left(b\sum_{i\in K_t}\sum_{j=1}^{n_i}y_{ij}\right)^{b\cdot|K_t|+1}}{\Gamma(b\cdot|K_t|)\cdot\Gamma(b\cdot|K_t|+1)\cdot\left(\sum_{i\in K_t}\sum_{j=1}^{n_i}x_{ij}\right)^{|K_t|}\cdot\left(\sum_{i\in K_t}\sum_{j=1}^{n_i}y_{ij}\right)^{|K_t|+1}}$$

and

$$q_j(b,\boldsymbol{x},\boldsymbol{y})=\prod_{M_j:t=1}^{r_j}\frac{\Gamma(|K_t|)\cdot\Gamma(|K_t|+1)\cdot\left(b\sum_{i\in K_t}\sum_{j=1}^{n_i}x_{ij}\right)^{b\cdot|K_t|}\cdot\left(b\sum_{i\in K_t}\sum_{j=1}^{n_i}y_{ij}\right)^{b\cdot|K_t|+1}}{\Gamma(b\cdot|K_t|)\cdot\Gamma(b\cdot|K_t|+1)\cdot\left(\sum_{i\in K_t}\sum_{j=1}^{n_i}x_{ij}\right)^{|K_t|}\cdot\left(\sum_{i\in K_t}\sum_{j=1}^{n_i}y_{ij}\right)^{|K_t|+1}}. \quad (10)$$

4. Numerical Examples

In this section, we use a numerical data to illustrate the multiple comparisons for the dependence parameters of bivariate exponential populations based on FBF. Here, we consider 4 bivariate exponential populations and sample size of 10 from each populations. In this paper, we consider multiple comparisons for two cases, so that true hypothesis are $H_{true}:\zeta_1=\zeta_2\neq\zeta_3=\zeta_4$ for case I and $H_{true}:\zeta_1=\zeta_2=\zeta_3\neq\zeta_4$ for case II, respectively.

The observed summary statistics for each case are given as Table 1. And the numbers of possible hypothesis are 15 for each cases.

<Table 1> The observed summary statistics for each populations

populations	Case I				Case II			
	1	2	3	4	1	2	3	4
$x_i = \sum_{j=1}^{n_i} x_{ij}$	6.0	7.2	11.8	14.4	15.1	14.2	12.3	9.2
$y_i = \sum_{j=1}^{n_i} y_{ij}$	5.2	6.8	12.2	15.1	14.3	13.2	12.8	3.9
M.L.E. of ζ_i	0.6579	0.5952	0.4505	0.3984	0.4115	0.4310	0.4386	0.7194

Table 2 gives the calculated posterior probabilities for all possible hypotheses based on FBF of case I and case II, respectively.

<Table 2> Calculated posterior probabilities for each cases based on FBF

Hypothesis	Case I	Case II
$\zeta_1=\zeta_2=\zeta_3=\zeta_4$	0.0617	0.0460
$\zeta_1=\zeta_2=\zeta_3\neq\zeta_4$	0.1054	**0.4683**
$\zeta_1=\zeta_2=\zeta_4\neq\zeta_3$	0.0111	0.0091
$\zeta_1=\zeta_2\neq\zeta_3=\zeta_4$	**0.5865**	0.0429
$\zeta_1=\zeta_2\neq\zeta_3\neq\zeta_4$	0.1014	0.0945
$\zeta_1=\zeta_3=\zeta_4\neq\zeta_2$	0.0102	0.0134
$\zeta_1=\zeta_3\neq\zeta_2=\zeta_4$	0.0077	0.0281
$\zeta_1=\zeta_3\neq\zeta_2\neq\zeta_4$	0.0069	0.1035
$\zeta_1=\zeta_4\neq\zeta_2=\zeta_3$	0.0055	0.0181
$\zeta_1=\zeta_4\neq\zeta_2\neq\zeta_3$	0.0013	0.0028
$\zeta_1\neq\zeta_2=\zeta_3=\zeta_4$	0.0433	0.0216
$\zeta_1\neq\zeta_2=\zeta_3\neq\zeta_4$	0.0202	0.1196
$\zeta_1\neq\zeta_2=\zeta_4\neq\zeta_3$	0.0054	0.0050
$\zeta_1\neq\zeta_2\neq\zeta_3=\zeta_4$	0.0284	0.0084
$\zeta_1\neq\zeta_2\neq\zeta_3\neq\zeta_4$	0.0049	0.0186

For case I, it is evident that the hypotheses for $\zeta_1=\zeta_2\neq\zeta_3=\zeta_4$ has the most large posterior probabilities 0.5865. This suggests that the data lend greatest support to equalities for $\zeta_1=\zeta_2$ and $\zeta_3=\zeta_4$ being different from the others. Thus this example shows good performance of the Bayesian multiple comparisons method based on FBF.

For case II, it is evident that the hypotheses for $\zeta_1=\zeta_2=\zeta_3\neq\zeta_4$ has the most large posterior probabilities 0.4683. This suggests that the data lend greatest support to equalities for $\zeta_1=\zeta_2=\zeta_3$ and ζ_4 being different from the others. Also this example shows good performance of the Bayesian multiple comparisons method based on FBF.

Up to this point, we have considered the problem of developing a Bayesian multiple comparisons for dependence parameter in K bivariate exponential populations. Extension of the above approach to the multiple comparison problems for the another population is straightforward. The research topics pertaining to the extension of the method and the examination of its performance are worthy to study and are left as a future subject of research.

References

1. Ali, M.M., Cho, J.S. and Begum, M.(2005), Nonparametric Bayesian Multiple Comparisons for Geometric Populations, *Journal of the Korean Data & Information Science Society*, Vol. 16(4), 1129-1140.
2. Berge, C.(1971), *Principle of Combinatorics* , New York: Academic Press.
3. Berger, J. O. and Pericchi, L. R.(1996), The Intrinsic Bayes Factor for Model Selection and Prediction, *Journal of the American Statistical Association*, 91, 109-122.
4. Consonni, G. and Veronese, P.(1995), A Bayesian Method for Combining Results from Several Binomial Experiments, *Journal of the American Statistical Association*, 90, 935-944.
5. Fienberg, S. E.(1980). *The Analysis of Cross-Classified Categrical Data*, Cambridge, MA: The MIT Press.
6. Lee, I.S., Cho, J.S., Kang, S.G. and Ko, J.H.(1998), Bayes Computations for the Reliability in a Bivariate Expnential Model, *The Korean Communications in Statistics*, 5(1), 145-153.
7. O' Hagan, A.(1995), Fractional Bayes Factors for Model Comparison(with discussion), *Journal of Royal Statistical Society*, 56, 99-118.

[received date : Apr. 2006, accepted date : May. 2006]