Nonparametric Bayesian Multiple Comparisons for Dependence Parameter in Bivariate Exponential Populations

Jang Sik Cho[1], M. Masoom Ali[2], Munni Begum[3]

Abstract

A nonparametric Bayesian multiple comparisons problem (MCP) for dependence parameters in I bivariate exponential populations is studied here. A simple method for pairwise comparisons of these parameters is also suggested. Here we extend the methodology studied by Gopalan and Berry (1998) using Dirichlet process priors. The family of Dirichlet process priors is applied in the form of baseline prior and likelihood combination to provide the comparisons. Computation of the posterior probabilities of all possible hypotheses are carried out through Markov Chain Monte Carlo method, namely, Gibbs sampling, due to the intractability of analytic evaluation. The whole process of MCP for the dependent parameters of bivariate exponential populations is illustrated through a numerical example.

Keywords : Bivariate exponential population; Dirichlet Process Prior; Gibbs Sampler; Mixture of Dirichlet Processes; Multiple Comparison; Nonparametric Bayes

1. Introduction

In reliability studies of mechanical components, dependence between two components occurs quite often. A system, which functions as long as at least one of the two identical components functions, has a functional correlation between the

---

1) First Author, Associate Professor, Department of Informational Statistics, Kyungsung University, Busan, 608-736, Korea.
   E-Mail : jscho@ks.ac.kr
2) Professor, Department of Mathematical Science, Ball State University, Muncie, IN, USA, 47306.
3) Assistant Professor, Department of Mathematical Science, Ball State University, Muncie, IN, USA, 47306.

system components. Initially let the two components be independently on test with life distributions that are exponential with parameter λ, denoted as $\exp(\lambda)$. Failure of one changes the life distribution of the other to $\exp(\lambda\theta)$, $\theta > 0$.

When $\theta = 1$, the two components function independently. For $\theta > 1$, the workload of the remaining component is increased, thereby decreasing the mean life. Here θ is called the dependence parameter. Weier (1981) provided the Bayes estimators of the parameters and reliability using a conjugate prior for such problems.

The multiple comparison problem (MCP) for I bivariate exponential populations with dependence parameters $\theta = (\theta_1, \cdots, \theta_I)$ can be viewed as making inferences concerning relationships among the θ's based on observations. This is tantamount to testing the following hypothesis,

$$H_o : \theta_1 = \cdots = \theta_I \quad v.s. \quad H_1 : not\, H_o.$$

For bivariate exponential populations, the frequentist approach of multiple comparison is not very straightforward. This is partly due to the difficulty in handling the distributional aspects and associated computations. The multiple comparison problem using nonparametric priors in a Bayesian inferential setup was studied by Gopalan and Berry (1998) providing specific applications to the Binomial and Normal populations. Following similar approach we studied the MCP for a set of geometric and negative binomial populations (2005). In this paper, we studied the MCP for the dependence parameters of a set of bivariate exponential populations along the same line.

In a Bayesian approach, the posterior probabilities of respective hypotheses in MCP can be calculated with moderate effort. The prior information on the unknown parameters has to be quantified as a distribution. However, the selection of the prior distribution could be tricky. One of the criticisms Bayesian inferential methodsoften face is the subjectivity in prior specification. In real data analysis prior specification could be based on scientific knowledge about the parameters. Non-informative prior specification is optimal in cases when there is little known about the background information. It is very important that prior distributions be as objective as possible while doing Bayesian inference. A typical objective prior distribution is the Dirichlet process prior (DPP) that leads to nonparametric Bayesian inference.

The DPP is a prior distribution on the family of distributions that is dense in the space of distribution functions. The family of DPPs was introduced by Ferguson (1973) and was extended to mixtures of DPP by Antoniak (1974) in order to treat problems including the estimation of a mixing distribution, bio-assay, empirical Bayes problems and discrimination problems. Escobar (1988) started the application of Markov chain Monte Carlo (MCMC) methods in nonparametric Bayesian modeling. Novel computational techniques and developments of MCMC schemes, including key contributions by Doss (1994), Bush and MacEachern (1996), Escobar and West (1997), MacEachern and Müller (1998), West, Müller and Escobar (1994) made it possible to study nonparametric Bayesian methods widely.

We focused on the Bayesian approach to the multiple comparisons problem for I bivariate exponential populations based on the nonparametric Dirichlet process priors in this paper. The MCMC techniques, in particular, Gibbs sampling is adopted here to evaluate the posterior probabilities of the hypotheses. Reviews on the DPP are presented in Section 2, while Section 3 presents the calculation of posterior probabilities for the hypotheses in MCP. A numerical example illustrating the procedure is presented in Section 4.

2. Preliminaries

Let (X, Y) denote the lifetimes of the two components that have a bivariate exponential model. The joint probability density function of (X, Y) can be written as,

$$f(x,y \mid \lambda,\theta) = 2\theta\lambda^2 \exp(-2\lambda x - \lambda\theta y), \quad x,y > 0, \quad \lambda, \theta > 0 \tag{1}$$

with θ as the dependence parameter.

We assume that $(x,y) = \{(x_1,y_1),\cdots,(x_I,y_I)\}$ be a set of observations available on I populations, where $(x_i,y_i) = \{(x_{i1},y_{i1}),\cdots,(x_{in_i},y_{in_i})\}$ is an $n_i \times 1$ vector of conditionally independent observations on population i, $i = 1,\cdots,I$; $j = 1,\cdots,n_i$ and $\sum_{i=1}^{I} n_i = n$. Then the probability density function of (x_{ij},y_{ij}) is,

$$f(x_{ij},y_{ij} \mid \lambda_i,\theta_i) = 2\theta_i\lambda_i^2 \times \exp(-2\lambda_i x_{ij} - \lambda_i\theta_i y_{ij}), \quad x_{ij},y_{ij} > 0, \lambda_i,\theta_i > 0. \tag{2}$$

Now a distribution function $G_o(\cdot)$ and a positive scalar precision parameter α together determine the Dirichlet process prior G. Here $G_o(\cdot)$ that defines the location of the DPP, is sometimes called prior "guess" or baseline prior. The precision parameter α determines the concentration of the prior for G around the prior guess G_o, and therefore measures the strength of belief in G_o. The DPP is usually denoted by $G \sim D(G \mid G_o, \alpha)$. For large values of α, G is very likely to be close to G_o, while for small values of α, G is likely to put most of its probability mass on just a few atoms.

We assume that the θ_i's come from G, and that $G \sim D(G \mid G_o, \alpha)$ as stated above. This structure results in a posterior distribution which is a mixture of Dirichlet processes (Antoniak 1974). Now following the Polya urn representation of the Dirichlet process (Blackwell andMacQueen, 1973), the joint posterior distribution can be written as,

$$\theta_i \mid x, y \propto \prod_{i=1}^{I} f(x_i, y_i \mid \theta_i) \times \frac{\alpha G_o(\theta_i) + \sum_{k<i} \delta(\theta_i \mid \theta_k)}{\alpha + i - 1} \tag{3}$$

where $\delta(\theta_i \mid \theta_k)$ is the distribution putting a point mass on θ_k. For each $i = 1, \cdots, I$, the conditional posterior distribution of θ_i is given by,

$$\theta_i \mid \theta_k, k \neq i, x, y \propto q_o G_b(\theta_i \mid x_i, y_i) + \sum_{k \neq i} q_k \delta(\theta_i \mid \theta_k), \tag{4}$$

where $G_b(\theta_i \mid x_i, y_i)$ is the baseline posterior distribution, $q_o \propto \alpha \int f(x_i, y_i \mid \theta) dG_o(\theta_i)$, $q_k \propto f(x_i, y_i \mid \theta_k)$, and $1 = q_o + \sum_{k \neq i} q_k$.

Let $\Theta = \{\theta = (\theta_1, \cdots, \theta_I) : \theta_i \in R, i = 1, \cdots, I\}$ be the I-dimensional parameter space. Equality and inequality relationships among θ's induce statistical hypotheses that are subsets of Θ. Thus the MCP becomes testing the following hypotheses.

$H_o : \theta_o = \{\theta_i : \theta_1 = \theta_2 = \cdots = \theta_I\}$, $H_1 : \theta_1 = \{\theta_i : \theta_1 \neq \theta_2 = \theta_3 = \cdots = \theta_I\}$, $\cdots$,

$H_N : \theta_N = \{\theta_i : \theta_1 \neq \theta_2 \neq \cdots \neq \theta_I\}$.

The hypotheses $H_r : \theta_r, r = 0, 1, 2, \cdots, N$ are disjoint, and $\cup_{r=0}^{N} \theta_r = \Theta$. The elements of Θ themselves behave as described by (3) and so with positive probability, they will reduce to some $p < I$ distinct values. Let superscript * denote distinct values of the parameters. Then any realization of I parameters θ_i generated from G lies in a set of $p < I$ distinct values, denoted by $\theta^* = (\theta_1^*, \cdots, \theta_p^*)$. The computation of posterior probabilities for different hypotheses through Gibbs

algorithm becomes manageable using the notion of *Configuration* as termed by Gopalan and Berry (1998). Their definition of *Configuration* is restated here,

Definition (*Configuration*): The set of indices $S = \{S_1, \cdots, S_I\}$ determines a classification of the data $\Theta = (\theta_1, \cdots, \theta_I)$ into I^* distinct groups or clusters; the $n_j = \#\{S_i = j\}$ observations in group j share the common parameter value θ_j^*. Now, define I_j as the set of indices of observations in group j That is, $I_j = \{i : S_i = j\}$. Let $(X, Y)_{(j)} = \{(X_i, Y_i) : S_i = j\}$ be the corresponding group of $n_{I_j} = \sum_{i \in I_j} n_i$ observations. Thus a one-to-one correspondence between hypotheses and configurations follows. And the required computations are reduced by the fact that the distinct θ_i's are typically reduced to fewer than I due to the clustering of the θ_i's inherent in the Dirichlet process. Hence, (4) can be rewritten as:

$$\theta_i \mid \theta_k, k \neq i, x, y \propto q_o G_b(\theta_i \mid x_i, y_i) + \sum_{i \neq k} n_k q_k^* \delta(\theta_i \mid \theta_k^*), \tag{5}$$

with $q_k^* \propto f(x_i, y_i \mid \theta_k^*)$, and $1 = q_0 + \sum_{k \neq i} n_k q_k^*$. In addition to the simplification of notations, the cluster structure of the θ_i also improves the efficiency of the algorithm.

3. Posterior Sampling In Dirichlet Process Mixtures

A gamma distribution with parameters $(\alpha_{oi}, \beta_{oi})$ is considered as baseline prior G_o. This implies that $\theta_1, \cdots, \theta_I$ are *i.i.d.* from G_o. Then a hierarchical set up for the Dirichlet process analysis as outlined above becomes,

$$x_i, y_i \mid \theta_i \sim BVE(x_i, y_i \mid \lambda_i, \theta_i) \tag{6}$$

$$\theta_i \mid G \sim G(\theta_i) \tag{7}$$

$$G \mid G_o, \alpha \sim D(G \mid G_o, \alpha) \tag{8}$$

$$G_o \mid \alpha_{oi}, \beta_{oi} \sim Gam(\alpha_{oi}, \beta_{oi}) \tag{9}$$

$$\lambda_i \mid \alpha_{1i}, \beta_{1i} \sim Gam(\alpha_{1i}, \beta_{1i}) \tag{10}$$

Here, *BVE* and *Gam* stand for bivariate exponential and gamma distributions, respectively. Now the choice of the precision parameter α in Dirichlet process is

extremely important for the model. We consider a gamma prior for α with a shape parameter a and scale parameter b, that is, $\alpha \sim Gam(a,b)$. Thus the $Gam(a,b)$ becomes the reference prior if $a \to 0$ and $b \to 0$. And we have access to a neat data augmentation device for sampling α by Escobar and West (1995).

The configuration notation is more convenient to use in describing the Gibbs sampling algorithm as the full conditionals can be written in closed form as under:

$$\theta_i \mid x,y,\theta_k, k \neq i, \alpha \sim q_o Gam\left(n_i + \alpha_{oi}, \lambda_i \sum_{j=1}^{n_i} y_{ij} + \beta_{oi}\right) + \sum_{k \neq i} q_k \delta(d\theta_i \mid \theta_k) \quad (11)$$

$$\lambda_i \mid x,y,\theta_i,\alpha \sim Gam\left(2n_i + \alpha_{1i}, \; 2\sum_{j=1}^{n_i} x_{ij} + \theta_i \sum_{j=1}^{n_i} y_{ij} + \beta_{1i}\right) \quad (12)$$

$$\theta_j^* \mid x,y,S \sim Gam\left(\sum_{i=1}^{I^*} n_i + \alpha_{oi}^*, \; \lambda_i \sum_{i=1}^{I^*} \sum_{j=1}^{n_i} y_{ij} + \beta_{oj}^*\right) \quad (13)$$

$$\alpha \mid \eta, I^* \sim \pi_\eta Gam(\alpha + I^*, \; b - \log(\eta)) + (1 - \pi_\eta) Gam(\alpha + I^* - 1, \; b - \log(\eta)) \quad (14)$$

$$\eta \mid \alpha, I^* \sim Beta(\alpha + 1, I^*) \quad (15)$$

where

$$q_o = \alpha \lambda_i^{2n_i + \alpha_{oi} - 1} \exp\left(-2\lambda_i \sum_{j=1}^{n_i} x_{ij}\right) \cdot \frac{\Gamma(n_i + \alpha_{oi})}{\left[\lambda_i\left(\sum_{j=1}^{n_i} y_{ij} + \beta_{oi}\right)\right]^{n_i + \alpha_{oi}}} \text{ and}$$

$$q_k \propto \theta_k^{n_i} \lambda_k^{2n_i} \exp\left(-2\lambda_k \sum_{j=1}^{n_i} x_{ij} - \theta_k \lambda_k \sum_{j=1}^{n_i} y_{ij}\right).$$

Gibbs sampling proceeds by simply iterating through (11) – (15) in order, sampling at each stage based on the current values of all the conditioning variables.

The configuration induces the equality and inequality relationships among the θ's, that corresponds to the partitions on the parameter space Θ and in turn to the hypotheses of interest. In order to estimate the posterior probability of a hypothesis H_r from a large number (L) of sample draws, we take

$$P(H_r \mid x,y) \approx \frac{1}{L} \sum_{l=1}^{L} \delta_{S_l}(H_r) \quad (16)$$

where $\delta_{S_l}(H_r)$ denotes unit point mass for the case where lth draw of S, S_l corresponds to H_r. The probability of equality for any two θ's can be calculated from the posterior distributions on hypotheses, $P(H_r \mid x,y)$, $r = 0,1,2,\cdots,N$. This

can be achieved by adding probabilities of those hypotheses in which the two θ_i and θ_j are equal. That is

$$P(\theta_i = \theta_j \mid x,y) \approx \frac{1}{L}\sum_{l=1}^{L}\delta_{S_l}(\theta_i = \theta_j), i \neq j \tag{17}$$

where $\delta_{S_l}(\theta_i = \theta_j)$ and denote unit point mass for the case where S_l indicate $\theta_i = \theta_j$.

4. Numerical Examples

A numerical illustration of the multiple comparisons for the dependence parameters in bivariate exponential populations is presented in this section using simulated data. We consider 4 bivariate exponential populations each with size $n_i = 20$. Then the numbers of possible hypotheses for multiple comparisons are 15. The observed summary statistics for these data are given in Table 1.

Table 1   The observed summary statistics for each populations

populations	1	2	3	4
$\sum_{j=1}^{n_i} x_{ij}$	1.500	1.560	0.700	0.720
$\sum_{j=1}^{n_i} y_{ij}$	6.500	6.000	1.300	1.130
$\hat{\theta}_{MLE}$	0.462	0.520	1.077	1.274

It follows from the above table, that the true hypothesis may be $H_{True} : \theta_1 = \theta_2 \neq \theta_3 = \theta_4$. For the precision parameter α, we consider three Gamma priors with parameters $(a,b)=(1.0, 1.0)$, $(0.1, 0.1)$ and $(0.01, 0.01)$ in order to have equal mean 1 and different variances 1, 10, and 100, respectively. This also facilitates that the latter prior be fairly noninformative, giving reasonable mass to both high and low values of α. We also set *a priori* that each $\theta_i, i = 1,\cdots,4$ follows a gamma distribution with parameters $\alpha_{oi} = \alpha_{1i} = 2$ and $\beta_{oi} = \beta_{1i} = 0.001$ to reflect vagueness of the prior knowledge.

The posterior probabilities for all possible hypotheses are approximated by the Gibbs sampling algorithm using 20,000 iterations with 10,000 burn-ins and 5 replications and are presented in Table 2. It is to be noted that the hypothesis ′

$\theta_1 = \theta_2 \neq \theta_3 = \theta_4{}'$ has the largest posterior probabilities 0.7883, 0.7274 and 0.7410 for all priors of the precision parameter α. Thus the data lend greatest support to equalities for $\theta_1 = \theta_2$ and $\theta_3 = \theta_4$ being different from the others.

Table 2   Calculated posterior probabilities for each hypothesis

Hypothesis	(1.0, 1.0)	(0.1, 0.1)	(0.01, 0.01)
$\theta_1 = \theta_2 = \theta_3 = \theta_4$	0.2059	0.2629	0.2320
$\theta_1 = \theta_2 = \theta_3 \neq \theta_4$	0.0000	0.0000	0.0000
$\theta_1 = \theta_2 = \theta_4 \neq \theta_3$	0.0000	0.0000	0.0000
$\theta_1 = \theta_2 \neq \theta_3 = \theta_4$	0.7883	0.7274	0.7410
$\theta_1 = \theta_2 \neq \theta_3 \neq \theta_4$	0.0000	0.0000	0.0000
$\theta_1 = \theta_3 = \theta_4 \neq \theta_2$	0.0036	0.0038	0.0030
$\theta_1 = \theta_3 \neq \theta_2 = \theta_4$	0.0000	0.0000	0.0000
$\theta_1 = \theta_3 \neq \theta_2 \neq \theta_4$	0.0000	0.0000	0.0000
$\theta_1 = \theta_4 \neq \theta_2 = \theta_3$	0.0000	0.0000	0.0000
$\theta_1 = \theta_4 \neq \theta_2 \neq \theta_3$	0.0000	0.0000	0.0000
$\theta_1 \neq \theta_2 = \theta_3 = \theta_4$	0.0003	0.0007	0.0015
$\theta_1 \neq \theta_2 = \theta_3 \neq \theta_4$	0.0000	0.0000	0.0000
$\theta_1 \neq \theta_2 = \theta_4 \neq \theta_3$	0.0000	0.0000	0.0000
$\theta_1 \neq \theta_2 \neq \theta_3 = \theta_4$	0.0018	0.0052	0.0226
$\theta_1 \neq \theta_2 \neq \theta_3 \neq \theta_4$	0.0000	0.0000	0.0000

Table 3 presents the pairwise posterior probabilities for the equalities in pairs of θ's. The equalities of $(\theta_1 = \theta_2)$ and $(\theta_3 = \theta_4)$ have the largest posterior probabilities (0.9943, 0.9903, 0.9729) and (0.9999, 0.9999, 0.9999) for three cases of (a, b) respectively. This suggests that there is strong evidence in the equality $(\theta_1 = \theta_2)$ and $(\theta_3 = \theta_4)$.

Table 3 Calculated Posterior Probabilities for the equalities in pairs of θ's

Hypothesis	(1.0, 1.0)	(0.1, 0.1)	(0.01, 0.01)
$\theta_1 = \theta_2$	0.9943	0.9903	0.9729
$\theta_1 = \theta_3$	0.2096	0.2667	0.2349
$\theta_1 = \theta_4$	0.2034	0.2555	0.2232
$\theta_2 = \theta_3$	0.2062	0.2636	0.2334
$\theta_2 = \theta_4$	0.2142	0.2526	0.2316
$\theta_3 = \theta_4$	0.9999	0.9999	0.9999

The Bayesian approach using nonparametric Dirichlet process priors facilitates studying the problem of multiple comparisons in a number of different distributions. So far, the MCP was carried out for a univariate distribution. Here we have shown that the method can be extended to a bivariate distribution as well, with moderate effort. As an alternative to a formal Bayesian analysis of a mixture model that usually leads to intractable calculations, the DPP is used to provide a nonparametric Bayesian method for obtaining posterior probabilities for various hypotheses of equality among the dependence parameters of bivariate exponential populations.

References

1. Ali, M. Masoom, Cho, J.S., Begum, M. (2005) Nonparametric Bayesian Multiple Comparisons for Geometric Populations, *Journal of the Korean Data and Information Science Society*. Accepted for publication.

2. Ali, M. Masoom., Cho, J.S., Begum, M. (2005) Bayesian Multiple Comparisons with Nonparametric Dirichlet Process Priors for Negative Binomial Populations. *Pakistan Journal of Statistics*. Under review.

3. Antoniak, C.E. (1974), Mixtures of Dirichlet Processes with Applications to Nonparametric Problems, *The Annals of Statistics*, 2, 1152-1174.

4. Bush, C.A. and MacEachern, S.N. (1996), A Semi-parametric Bayesian Model for randomized Block Designs, *Biometrika*, 83, 275-285.

5. Doss, H. (1994), Bayesian Nonparametric Estimation for Incomplete Data via Successive Substitution Sampling, *The Annals of Statistics*, 22, 1763-1786.

6. Escobar, M. D. (1988), Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means. *Unpublished dissertation*, Yale University.

7. Escobar, M. D. and West, M. (1997), Computing Nonparametric Hierarchical Models, *ISDS Discussion Paper #97-15*, Duke University.

8. Ferguson, T.S. (1973), A Bayesian Analysis of Some Nonparametric Problems, *The Annals of Statistics*, 1, 209-230.

9. Gopalan, R. and Berry, D.A. (1998), Bayesian Multiple Comparisons Using Dirichlet Process Priors, *Journal of the American Statistical Association*, 90, 1130 - 1139.

10. MacEachern, S.N. and Müller, P. (1998), Estimating Mixture of Dirichlet Process Models, *Journal of Computational and Graphical Statistics*, 7, 223-239.

11. Weier,D.R. (1981), Bayes Estimation for a Bivariate Survival Model Based on Exponential Distributions, *Communication in Statistics-Theory and Methods*, 10, 1415-1427.