

5-1-2006

Nonparametric Bayesian Multiple Comparisons for Dependence Parameter in Bivariate Exponential Populations

M. Masoom Ali

Ball State University, mali@bsu.edu

J. S. Cho

Kyungshung University

Munni Begum

Ball State University

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Ali, M. Masoom; Cho, J. S.; and Begum, Munni (2006) "Nonparametric Bayesian Multiple Comparisons for Dependence Parameter in Bivariate Exponential Populations," *Journal of Modern Applied Statistical Methods*: Vol. 5 : Iss. 1 , Article 7.

DOI: 10.22237/jmasm/1146456360

Nonparametric Bayesian Multiple Comparisons for Dependence Parameter in Bivariate Exponential Populations

M. Masoom Ali
Dept. of Math. Sciences
Ball State University

J. S. Cho
Dept. of Inform. Statistics
Kyungsung University

Munni Begum
Dept. of Math. Sciences
Ball State University

A nonparametric Bayesian multiple comparisons problem (MCP) for dependence parameters in I bivariate exponential populations is studied. A simple method for pairwise comparisons of these parameters is also suggested. The methodology by Gopalan and Berry (1998) is extended using Dirichlet process priors, 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 a Markov Chain Monte Carlo, Gibbs sampling, due to the intractability of analytic evaluation. The process of MCP for the dependent parameters of bivariate exponential populations is illustrated with a numerical example.

Key words: Bivariate exponential population; Dirichlet process prior; Gibbs sampler; mixture of Dirichlet processes; multiple comparison; nonparametric Bayes.

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 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 $\theta = (\theta_1, \dots, \theta_I)$ can be viewed as making inferences concerning relationships among the θ 's based on observations. This is tantamount to testing the following hypothesis,

$$H_0 : \theta_1 = \dots = \theta_I \text{ vs. } H_1 : \text{not } H_0.$$

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, the MCP for a set of geometric and negative binomial populations (Masoom, Cho, & Begum, 2005) was studied. In this article, the MCP for the dependence parameters of a set of

M. Masoom Ali is George & Frances Ball Distinguished Professor of Statistics. Email him at mali@bsu.edu. His research interests are in order statistics, Bayesian statistics, statistical inference and distribution problems. J. C. Cho is Associate Professor of Informational Statistics. His research interests are in Bayesian Statistics. Munni Begum is Assistant Professor of Statistics. Her research interests are in Bayesian Statistics, Statistical modeling and Biostatistical methods.

bivariate exponential populations along the same line was studied.

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 methods often 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.

The focus was on the Bayesian approach to the multiple comparisons problem for I bivariate exponential populations based on the nonparametric Dirichlet process priors in this article. The MCMC techniques, in particular Gibbs sampling, is adopted here to evaluate the posterior probabilities of the hypotheses.

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,

$$f(x, y) | \lambda, \theta = 2\theta\lambda^2 \exp(-2\lambda x - \lambda\theta y),$$

$$x, y > 0, \lambda, \theta > 0$$
(1)

with θ as the dependence parameter.

It is assumed that $(\mathbf{x}, \mathbf{y}) = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_I, \mathbf{y}_I)\}$ be a set of observations available on I populations, where $(\mathbf{x}_i, \mathbf{y}_i) = \{(x_{i1}, y_{i1}), \dots, (x_{in_i}, y_{in_i})\}$ is an $n_i \times I$ vector of conditionally independent observations on population i , $i = 1, 2, \dots, I$; $j = 1, 2, \dots, n_i$ and $\sum_{i=1}^I n_i = n$. Then the probability density function of (x_{ij}, y_{ij}) is,

$$f(x_{ij}, y_{ij} | \lambda_i, \theta_i) = 2\theta_i \lambda_i^2 \times \exp(-2\lambda_i x_{ij} - \lambda_i \theta_i y_{ij}),$$

$$x_{ij}, y_{ij} > 0, \lambda_i, \theta_i > 0.$$
(2)

Now a distribution function $G_0(\cdot)$ and a positive scalar precision parameter α together determine the Dirichlet process prior G . Here $G_0(\cdot)$ 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_0 , and therefore measures the strength of belief in G_0 . The DPP is usually denoted by $G \sim D(G / G_0, \alpha)$. For large values of α , G is very likely to be close to G_0 , while for small values of α , G is likely to put most of its probability mass on just a few atoms.

It is assumed that the θ_i 's come from G , and that $G \sim D(G / G_0, \alpha)$ 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 & MacQueen, 1973), the joint posterior distribution can be written as,

$$\theta_i | \mathbf{x}, \mathbf{y} \propto \prod_{i=1}^I f(\mathbf{x}_i, \mathbf{y}_i | \theta_i) \times \frac{\alpha G_0(\theta_i) + \sum_{k \neq i} \delta(\theta_i | \theta_k)}{\alpha + i - 1}, \quad (3)$$

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

$$\theta_i | \theta_k, k \neq i, \mathbf{x}, \mathbf{y} \propto q_0 G_b(\theta_i | \mathbf{x}_i, \mathbf{y}_i) + \sum_{k \neq i} q_k \delta(\theta_i | \theta_k), \quad (4)$$

where $G_b(\theta_i | \mathbf{x}_i, \mathbf{y}_i)$ is the baseline posterior distribution, $q_0 \propto \alpha \int f(\mathbf{x}_i, \mathbf{y}_i | \theta) dG_0(\theta)$, $q_k \propto f(\mathbf{x}_i, \mathbf{y}_i | \theta_k)$, and $1 = q_0 + \sum_{k \neq i} q_k$. Let

$\Theta = \{\theta = (\theta_1, \theta_2, \dots, \theta_I) : \theta_i \in R, i=1, 2, \dots, I\}$ 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.

$$\begin{aligned} H_0 : \theta_0 = \{\theta_i : \theta_1 = \theta_2 = \dots = \theta_I\}, \\ H_1 : \theta_1 = \{\theta_i : \theta_1 \neq \theta_2, \theta_2 = \theta_3 = \dots = \theta_I\}, \dots, \\ H_N : \theta_N = \{\theta_i : \theta_1 \neq \theta_2 \neq \theta_3 \neq \dots \neq \theta_N\}. \end{aligned}$$

The hypotheses $H_r : \theta_r, r = 0, 1, 2, \dots, N$, are disjoint, and $\bigcup_{r=0}^N \theta_r = \Theta$.

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 $(\theta^* = \theta_1^*, \theta_2^*, \dots, \theta_p^*)$. 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 $S = \{S_1, \dots, S_I\}$ determines a classification of the data $\Theta = \{\theta_1, \dots, \theta_I\}$ 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:

$$\theta_i | \theta_k, k \neq i, \mathbf{x}, \mathbf{y} \propto q_0 G_b(\theta_i | \mathbf{x}_i, \mathbf{y}_i) + \sum n_k q_k^* \delta(\theta_i | \theta_k^*), \quad (5)$$

with $q_k^* \propto f(\mathbf{x}_i, \mathbf{y}_i | \theta_k^*)$, 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.

Posterior Sampling In Dirichlet Process Mixtures

A gamma distribution with parameters $(\alpha_{0i}, \beta_{0i})$ is considered as baseline prior G_0 . This implies that $\theta_1, \theta_2, \dots, \theta_I$ are *i.i.d.* from G_0 . Then, a hierarchical set up for the Dirichlet process analysis as outlined above becomes,

$$\mathbf{x}_i, \mathbf{y}_i | \theta_i \sim BVE(\mathbf{x}_i, \mathbf{y}_i | \lambda_i, \theta_i), \quad (6)$$

$$\theta_i | G \sim G(\theta_i), \quad (7)$$

$$G | G_0, \alpha \sim D(G | G_0, \alpha), \quad (8)$$

$$G_0 | \alpha_{0i}, \beta_{0i} \sim Gam(\alpha_{0i}, \beta_{0i}), \quad (9)$$

$$\lambda_i | \alpha_{1i}, \beta_{1i} \sim \text{Gam}(\alpha_{1i}, \beta_{1i}), \quad (10)$$

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. A gamma prior for α with a shape parameter a and scale parameter b is considered, that is, $\alpha \sim \text{Gam}(a, b)$. Thus, the $\text{Gam}(a, b)$ becomes the reference prior if $a \rightarrow 0$ and $b \rightarrow 0$ and one has 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:

$$\begin{aligned} & (\theta_i | \mathbf{x}, \mathbf{y}, \theta_k, k \neq i, \alpha) \sim \\ & q_0 \text{Gam} \left(n_i + \alpha_{0i}, \lambda_i \sum_{j=1}^{n_i} y_{ij} + \beta_{0i} \right) + \sum_{k \neq i} q_k \delta(d\theta_i | \theta_k), \end{aligned} \quad (11)$$

$$\begin{aligned} & (\lambda_i | \mathbf{x}, \mathbf{y}, \theta_i, \alpha) \sim \\ & \text{Gam} \left(2n_i + \alpha_{1i}, 2 \sum_{j=1}^{n_i} x_{ij} + \theta_i \sum_{j=1}^{n_i} y_{ij} + \beta_{1i} \right), \end{aligned} \quad (12)$$

$$\begin{aligned} & (\theta_j^* | \mathbf{x}, \mathbf{y}, \mathbf{S}) \sim \\ & \text{Gam} \left(\sum_{i=1}^{I^*} n_i + \alpha_{0i}^*, \lambda_i \sum_{i=1}^{I^*} \sum_{j=1}^{n_i} y_{ij} + \beta_{0j}^* \right), \end{aligned} \quad (13)$$

$$\begin{aligned} & (\alpha | \eta, I^*) \sim \\ & \pi_\eta \text{Gam}(a + I^*, b - \log(\eta)) + \\ & (1 - \pi_\eta) \text{Gam}(a + I^* - 1, b - \log(\eta)), \end{aligned} \quad (14)$$

$$(\eta | \alpha, I^*) \sim \text{Beta}(\alpha + 1, I^*), \quad (15)$$

where

$$\begin{aligned} q_0 & \propto \alpha \lambda_i^{2n_i + \alpha_{0i} - 1} \exp \left(-2\lambda_i \sum_{j=1}^{n_i} x_{ij} \right) \\ & \frac{\Gamma(n_i + \alpha_{0i})}{\left[\lambda_i \left(\sum_{j=1}^{n_i} y_{ij} + \beta_{0i} \right) \right]^{n_i + \alpha_{0i}}}, \\ q_k & \propto \theta_k^{n_i} \lambda_k^{2n_i} \exp \left(-2\lambda_k \sum_{j=1}^{n_i} x_{ij} - \theta_k \lambda_k \sum_{j=1}^{n_i} y_{ij} \right) \end{aligned}$$

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, one takes

$$P(H_r | \mathbf{X}, \mathbf{Y}) \approx \frac{1}{L} \sum_{l=1}^L \delta_{S_l}(H_r), \quad (16)$$

where $\delta_{S_l}(H_r)$ denotes unit point mass for the case where l th draw of S , S_0 corresponds to H_r . The probability of equality for any two θ 's can be calculated from the posterior distributions on hypotheses, $P(H_r | \mathbf{X}, \mathbf{Y})$, $r = 1, 2, \dots, N$. This can be achieved by adding probabilities of those hypotheses in which the two θ_i and θ_j are equal. That is

$$\begin{aligned} P(\theta_i = \theta_j | \mathbf{X}, \mathbf{Y}) & \approx \frac{1}{L} \sum_{l=1}^L \delta_{S_l}(\theta_i = \theta_j) = \\ & \sum_{r=1}^N P(H_r | \mathbf{X}, \mathbf{Y}) \delta_{H_r}(\theta_i - \theta_j), i \neq j, \end{aligned}$$

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

Illustrative Example

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

It follows from Table 1, that the true hypothesis may be $H_{true} : \theta_1 = \theta_2 \neq \theta_3 = \theta_4$. For the precision parameter α , one considers 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 non-informative, giving reasonable mass to both high and low values of α . As well, each θ_i , $i=1, \dots, 4$ were set *a priori* following a gamma distribution with parameters $\alpha_{0i} = \alpha_{1i} = 2.0$ and $\beta_{0i} = \beta_{1i} = 0.001$ 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 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 (1.0000, 1.0000, 1.0000) for three cases of (a, b) respectively. This suggests that there is strong evidence in the equality $(\theta_1 = \theta_2)$ and $(\theta_3 = \theta_4)$.

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, it has been 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.

Table 1 The observed summary statistics for each populations

<u>Populations</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
$X_i = \sum_{j=1}^{n_i} X_{ij}$	1.500	1.560	0.700	0.720
$Y_i = \sum_{j=1}^{n_i} Y_{ij}$	6.500	6.000	1.300	1.130
$\hat{\theta}_{MLE}$	0.462	0.520	1.077	1.274

Table 2 Calculated posterior probabilities for each hypothesis with three cases of (a,b)

<u>Hypothesis</u>	<u>(1.0, 1.0)</u>	<u>(0.1, 0.1)</u>	<u>(0.01, 0.01)</u>
$\theta_1 = \theta_2 = \theta_3 = \theta_4$.2059	.2629	.2320
$\theta_1 = \theta_2 = \theta_3 \neq \theta_4$.0000	.0000	.0000
$\theta_1 = \theta_2 = \theta_4 \neq \theta_3$.0000	.0000	.0000
$\theta_1 = \theta_2 \neq \theta_3 = \theta_4$.7883	.7274	.7410
$\theta_1 = \theta_2 \neq \theta_3 \neq \theta_4$.0000	.0000	.0000
$\theta_1 = \theta_3 = \theta_4 \neq \theta_2$.0036	.0038	.0030
$\theta_1 = \theta_3 \neq \theta_2 = \theta_4$.0000	.0000	.0000
$\theta_1 = \theta_3 \neq \theta_2 \neq \theta_4$.0000	.0000	.0000
$\theta_1 = \theta_4 \neq \theta_2 = \theta_3$.0000	.0000	.0000
$\theta_1 = \theta_4 \neq \theta_2 \neq \theta_3$.0000	.0000	.0000
$\theta_1 \neq \theta_2 = \theta_3 = \theta_4$.0003	.0007	.0015
$\theta_1 \neq \theta_2 = \theta_3 \neq \theta_4$.0000	.0000	.0000
$\theta_1 \neq \theta_2 = \theta_4 \neq \theta_3$.0000	.0000	.0000
$\theta_1 \neq \theta_2 \neq \theta_3 = \theta_4$.0018	.0052	.0226
$\theta_1 \neq \theta_2 \neq \theta_3 \neq \theta_4$.0000	.0000	.0000

Table 3 Pairwise Posterior Probabilities with three cases of (a, b)

<u>Hypothesis</u>	<u>(1.0, 1.0)</u>	<u>(0.1, 0.1)</u>	<u>(0.01, 0.01)</u>
$\theta_1 = \theta_2$.9943	.9903	.9729
$\theta_1 = \theta_3$.2096	.2667	.2349
$\theta_1 = \theta_4$.2096	.2667	.2349
$\theta_2 = \theta_3$.2062	.2636	.2334
$\theta_2 = \theta_4$.2062	.2636	.2334
$\theta_3 = \theta_4$	1.0000	1.0000	1.0000

References

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*, 16(4), 1129-1140.

Ali, M. Masoom., Cho, J. S., Begum, M. (2006). Bayesian multiple comparisons with nonparametric dirichlet process priors for negative binomial populations. *Pakistan Journal of Statistics*, 22(2), 88-98.

Antoniak, C. E. (1974). Mixtures of dirichlet processes with applications to nonparametric problems. *The Annals of Statistics*, 2, 1152-1174.

Bush, C. A. & MacEachern, S. N. (1996). A semi-parametric bayesian model for randomized block designs. *Biometrika*, 83, 275-285.

Doss, H. (1994). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. *The Annals of Statistics*, 22, 1763-1786.

Escobar, M. D. (1988). Estimating the means of several normal populations by nonparametric estimation of the distribution of the means. *Unpublished dissertation*, Yale University.

Escobar, M. D. & West, M. (1997). Computing nonparametric hierarchical models. *ISDS Discussion Paper #97-15*, Duke University.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1, 209-230.

Gopalan, R. & Berry, D. A. (1998). Bayesian multiple comparisons using dirichlet process priors. *Journal of the American Statistical Association*, 90, 1130 - 1139.

MacEachern, S. N. & Müller, P. (1998). Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics*, 7, 223-239.

Weier, D. R. (1981). Bayes estimation for a bivariate survival model based on exponential distributions. *Communication in Statistics-Theory and Methods*, 10, 1415-1427.