Pak. J. Statist.
2006 Vol. 22(2) pp 89-98

BAYESIAN MULTIPLE COMPARISONS WITH NONPARAMETRIC DIRICHLET PROCESS PRIORS FOR NEGATIVE BINOMIAL POPULATIONS

M. Masoom Ali
Department of Mathematical Sciences
Ball State University
Muncie, IN 47306-0490, U.S.A.
Email: mali@bsu.edu

Cho, J.S
Department of Informational Statistics
Kyungsung University
Pusan, Korea, 608-736

and

Munni Begum
Department of Mathematical Sciences
Ball State University
Muncie, IN 47306-0490, U.S.A.

ABSTRACT

Bayesian multiple comparisons for Binomial and Normal populations were studied by Gopalan and Berry(1998) with nonparametric Dirichlet process priors. Here we consider the specific problem of multiple comparisons for the Negative Binomial populations. The family of Dirichlet process priors is applied in the form of baseline priors to compute posterior probabilities for various hypotheses on the parameters of Negative Binomial Distributions. The computation has been carried out using Gibbs algorithm due to intractability of analytic forms. An illustrative example with simulated data is presented to demonstrate the whole procedure.

Key Words:

Dirichlet Process Prior; Gibbs Sampler; Mixture of Dirichlet Processes; Negative Binomial Distribution; Multiple Comparison; Nonparametric Bayes.

AMS(2000) Subject Classification: 62G10.

1. INTRODUCTION

Negative Binomial distribution has a wide range of applicability in various fields. This distribution had been studied extensively in auto insurance for accident proneness and liability by the actuaries. Other field of applications include studying the number of bacteria per microscopic field, statistical quality control research and the studies on the inspection of defective items. In this paper, we consider I Negative Binomial populations with parameters $\theta = (\theta_1, \cdots, \theta_I)$ to compare of two or more populations in regard to comparing the probabilities of successes in respective populations.

In frequentist approach, the standard I-samples problem is to test for the equality of the distributions,

$$H_0 : \theta_1 = \cdots = \theta_I \quad vs. \quad H_1 : not\ H_0. \tag{1}$$

The multiple comparison problem (MCP) follows naturally if there is sufficient evidence from the sample data for rejecting the null hypothesis. The MCP implies making inferences concerning relationships among the θ's based on observations. For a set of Negative Binomial populations it is not straightforward to perform multiple comparisons due to computational difficulties. The multiple comparison problem using nonparametric priors 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 populations (2005).

In Bayesian approach, the posterior probabilities of respective hypotheses in multiple comparisons can be calculated with moderate effort. The prior information on the unknown parameters has to be quantified as a distribution. The selection of prior distribution could be tricky. One of the criticism Bayesian inferential methods often face is that 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 some 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.

Nonparametric Bayesian inference using the Dirichlet process prior (DPP) is one such objective prior distribution. The DPP is a prior specification on the family of the distributions, that is dense in the space of distribution functions. The family of DPPs is introduced by Ferguson(1973) and 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. The introduction of Markov chain Monte Carlo (MCMC) methods in nonparametric Bayesian modeling was begun with Escobar (1988). Novel computational techniques and the development of new 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 for different problems.

In this paper, we consider the Bayesian approach to the multiple comparisons problem for I negative binomial populations based on the hierarchical nonparametric family of Dirichlet process priors. The Calculation of posterior probabilities for the hypotheses is analytically intractable, but can be evaluated easily with Monte Carlo techniques such as Gibbs sampling. 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 described in section 4.

2. PRELIMINARIES

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.

We consider I negative binomial populations with the probabilities of success $\theta = (\theta_1, \theta_2, \cdots, \theta_I)$. Observations $\mathbf{Y} = (\mathbf{Y_1}, \mathbf{Y_2}, \cdots, \mathbf{Y_I})$ are available on these populations, where $\mathbf{Y_i} = (y_{i1}, \cdots, y_{in_i})$ is $n_i \times 1$ vector of conditionally independent observations on population i, $i = 1, 2, \cdots, I$; $j = 1, 2, \cdots, n_i$ and $\sum_{i=1}^{I} n_i = n$. Then the probability density function of y_{ij} can be written as,

$$f(y_{ij}|\theta_i) = \left(\begin{array}{c} r_i + y_{ij} - 1 \\ y_{ij} \end{array} \right) {\theta_i}^{r_i} . (1-\theta_i)^{y_{ij}},$$

where r_i denotes the number of successes at ith population. We assume that the θ_i's come from G, and that $G \sim D(G|G_0, \alpha)$. This structure results in a posterior distribution of so called the mixture of Dirichlet processes (Antoniak 1974). Following the Polya urn representation of the Dirichlet process (Blackwell and MacQueen 1973), the joint posterior distribution has the form,

$$\theta_i|\mathbf{Y} \ \propto \ \prod_{i=1}^{I} f(\mathbf{y_i}|\theta_i) \frac{\alpha G_0(\theta_i) + \sum_{k<i} \delta(\theta_i|\theta_k)}{\alpha + i - 1}, \tag{2}$$

where $\delta(\theta_i|\theta_k)$ is a probability distribution giving mass one to the point θ_k. For each $i = 1, \cdots, I$, the conditional posterior distribution of θ_i is given by,

$$\theta_i | \theta_k, k \neq i, \mathbf{Y} \ \propto \ q_0 G_b(\theta_i | y_i) + \sum_{k \neq i} q_k \delta(\theta_i | \theta_k), \tag{3}$$

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

The multiple comparison problem of I negative binomial populations is to make inferences concerning relationships among the θ based on $\mathbf{Y}$. Let $\Theta = \{\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_I) : \theta_i \in R, i = 1, 2, \cdots, I\}$ be the I-dimensional parameter space. Equality and inequality relationships among the θ's induce statistical hypotheses that are subsets of Θ, *i.e.*, $H_0 : \boldsymbol{\theta}_0 = \{\theta_i : \theta_1 = \theta_2 = \cdots = \theta_I\}$, $H_1 : \boldsymbol{\theta}_1 = \{\theta_i : \theta_1 \neq \theta_2 = \cdots = \theta_I\}$ and so on up to $H_N : \boldsymbol{\theta}_N = \{\theta_i : \theta_1 \neq \theta_2 \neq \cdots \neq \theta_I\}$. The hypotheses $H_r : \boldsymbol{\theta}_r, r = 0, 1, 2, \cdots, N$, are disjoint, and $\cup_{r=0}^{N} \boldsymbol{\theta}_r = \Theta$.

The elements of Θ behave as described in equation (3) and with positive probability, they will reduce to some $p \leq I$ distinct values. We use superscript '*' to denote the distinct values of the parameters. Then any realization of I parameters θ_i generated from G lies in a set of $p \leq I$ distinct values, denoted by $\theta^* = (\theta_1^*, \theta_2^*, \cdots, \theta_p^*)$. The computation of posterior probabilities of different hypotheses through Gibbs algorithm can be described easily through what is termed as *Configuration* 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 one-way classification of the data $Y = \{y_1, \cdots, y_I\}$ 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 $Y_{(j)} = \{Y_i : S_i = j\}$ be the corresponding group of $n_{I_j} = \sum_{i \in I_j} n_i$ observations.

It can be shown that there is a one-to-one correspondence between hypotheses and configurations. The required computations are reduced as the distinct θ_i's typically reduce to fewer than I due to the clustering of the θ_i's, inherent in the Dirichlet process. Therefore, (4) can be rewritten as,

$$\theta_i | \theta_k, k \neq i, \mathbf{y} \ \propto \ q_0 G_b(\theta_i | \mathbf{y_i}) + \sum_{k=1}^{I^*} n_k q_k^* \delta(\theta_i | \theta_k^*), \tag{4}$$

with $q_k^* \propto f(\mathbf{y_i}|\theta_k^*)$, and $1 = q_0 + \sum_k n_k q_k^*$. In addition to simplifying notations, the cluster structure of the θ_i can also be used to improve the efficiency of the algorithm.

3. POSTERIOR SAMPLING IN DIRICHLET PROCESS MIXTURES

A beta distribution with parameters $(\alpha_{oi}, \beta_{oi})$ is considered here as the baseline prior G_0. This implies that $\theta_1, \theta_2, \cdots, \theta_K$ are *i.i.d.* from G_0. Then a Dirichlet process analysis as outline in the above description results,

$$\mathbf{y_i}|\theta_i \sim NB(\mathbf{y_i}|\theta_i), \tag{5}$$
$$\theta_i|G \sim G(\theta_i), \tag{6}$$
$$G|G_0, \alpha \sim D(G|G_0, \alpha), \tag{7}$$
$$G_0|(\alpha_{oi}, \beta_{oi}) \sim Beta(\alpha_{oi}, \beta_{oi}). \tag{8}$$

Here, NB denotes negative binomial distribution. The choice of the precision parameter α in Dirichlet process is important for the model. We consider the gamma prior for α with a shape parameter a and scale parameter b. That is, $\alpha \sim Gamma(a, b)$. $Gamma(a, b)$ becomes the reference prior as $a \to 0$ and $b \to 0$. Then we can apply the data augmentation device for sampling α by proposed by Escobar and West(1995).

Using the definition of configuration one can conveniently describe the Gibbs sampling algorithm for the multiple comparison problem. Here with negative binomial populations we sample from the following conditional posterior distributions,

$$\begin{aligned}
(\theta_i|Y, \theta_k, k \neq i, \alpha, \alpha_{oi}, \beta_{oi}) &\sim q_0 Beta(n_i \cdot r_i + \alpha_{io}, \sum_{j=1}^{n_i} y_{ij} + \beta_{oi}) \\
&\quad + \sum_{k \neq i} q_k \delta(d\theta_i|\theta_k), && (9) \\
(\theta_j^*|Y, S, \alpha_{oi}, \beta_{oi}) &\sim Beta(\sum_{k \in J_j} n_k \cdot r_i + \alpha_{oj}{}^*, \sum_{k \in J_j} \sum_{l=1}^{n_i} y_{kl} + \beta_{oj}{}^*) && (10) \\
(\alpha|\eta, I^*) &\sim \pi_\eta Gamma(a + I^*, b - log(\eta)) \\
&\quad + (1 - \pi_\eta) Gamma(a + I^* - 1, b - log(\eta)), && (11) \\
(\eta|\alpha, I^*) &\sim Beta(\alpha + 1, I), && (12)
\end{aligned}$$

where

$$\begin{aligned}
q_0 &\propto \alpha \frac{\Gamma(\alpha_{oi} + \beta_{oi})}{\Gamma(\alpha_{oi}) \cdot \Gamma(\beta_{oi})} \cdot \frac{\Gamma(n_i \cdot r_i + \alpha_{oi}) \cdot \Gamma(\sum_{j=1}^{n_i} y_{ij} + \beta_{oi})}{\Gamma(n_i \cdot r_i + \sum_{j=1}^{n_i} y_{ij} + \alpha_{oi} + \beta_{oi})}, \\
q_k &\propto \theta_k{}^{n_i \cdot r_i} (1 - \theta_k)^{\sum_{j=1}^{n_i} y_{ij}}.
\end{aligned}$$

Gibbs sampling proceeds by simply iterating through (10) - (13) in order, sampling at each stage based on current values of all the conditioning variables.

The configuration gives the equality and inequality relationships among the θ's, which correspond to the partitions on the parameter space of Θ and in turn to the hypotheses of interest. To estimate the posterior probability of a hypothesis H_r from a large number(L) of sample drawings we use,

$$P(H_r|\mathbf{Y}) \approx \frac{1}{L}\sum_{l=1}^{L}\delta_{S_l}(H_r), \tag{13}$$

where $\delta_{S_l}(H_r)$ denotes unit point mass for the case where lth draw of S, or S_l corresponds to H_r.

The probability of equality for any two θ's can be calculated from the posterior distributions on hypotheses, $P(H_r|\mathbf{Y}),\ r = 1, 2, \cdots, N$. This can be achieved by adding probabilities of those hypotheses for which the two θ_i and θ_j are equal. Therefore,

$$P(\theta_i = \theta_j|\mathbf{Y}) \approx \frac{1}{L}\sum_{l=1}^{L}\delta_{S_l}(\theta_i = \theta_j) = \sum_{r=1}^{N}P(H_r|\mathbf{Y})\delta_{H_r}(\theta_i = \theta_j),\ i \neq j, \tag{14}$$

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 $\theta_i = \theta_j$, respectively.

4. ILLUSTRATIVE EXAMPLE

In this section, we use simulated data to illustrate the multiple comparisons for the parameters of negative binomial populations. Here, we consider 4 negative binomial populations with same $r_i = 3$ for all i and sample size of 15 from each negative binomial distributions. The observed summary statistics for each populations are given as Table 1. The multiple comparison problem is demonstrated for two cases.

Table 1 The observed summary statistics for each populations

	Case I				Case II			
Populations	1	2	3	4	1	2	3	4
$y_i = \sum_{j=1}^{n_i} y_{ij}$	30	31	76	71	29	27	31	75
r_i	3	3	3	3	3	3	3	3
n_i	15	15	15	15	15	15	15	15

The number of possible hypothesis is 15 and without loss of generality the true hypothesis may be taken as $H_{true} : \theta_1 = \theta_2 \neq \theta_3 = \theta_4$ for case I and $H_{true} : \theta_1 = \theta_2 = \theta_3 \neq \theta_4$ for case II. 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. But, the $Gamma(1.0,\ 1.0)$ prior favors relatively low values of α. Also we set $\alpha_{oi} = \beta_{oi} = 1.0$ for all i that the prior distributions of θ_i's are fairly noninformative.

Table 2 and Table 3 give the calculated posterior probabilities for all possible hypotheses approximated by the Gibbs sampling algorithm using 40,000 iterations with 20,000 burn-in iterations of cases I and II, respectively. It is evident from Table 2 that the hypotheses for "$\theta_1 = \theta_2 \neq \theta_3 = \theta_4$" have the largest posterior probabilities 0.6634, 0.5338, and 0.3194 for all priors of the precision parameter α, respectively. This suggests that 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 with three cases of (a, b) in Case I

Hypothesis	$(1.0, 1.0)$	$(0.1, 0.1)$	$(0.01, 0.01)$
$\theta_1 = \theta_2 = \theta_3 = \theta_4$	0.0167	0.0906	0.3036
$\theta_1 = \theta_2 = \theta_3 \neq \theta_4$	0.0007	0.0007	0.0003
$\theta_1 = \theta_2 = \theta_4 \neq \theta_3$	0.0000	0.0000	0.0000
$\theta_1 = \theta_2 \neq \theta_3 = \theta_4$	0.6634	0.5338	0.3194
$\theta_1 = \theta_2 \neq \theta_3 \neq \theta_4$	0.1103	0.1142	0.0860
$\theta_1 = \theta_3 = \theta_4 \neq \theta_2$	0.0091	0.0074	0.0042
$\theta_1 = \theta_3 \neq \theta_2 = \theta_4$	0.0000	0.0000	0.0000
$\theta_1 = \theta_3 \neq \theta_2 \neq \theta_4$	0.0014	0.0008	0.0008
$\theta_1 = \theta_4 \neq \theta_2 = \theta_3$	0.0000	0.0000	0.0000
$\theta_1 = \theta_4 \neq \theta_2 \neq \theta_3$	0.0001	0.0003	0.0002
$\theta_1 \neq \theta_2 = \theta_3 = \theta_4$	0.0152	0.0126	0.0071
$\theta_1 \neq \theta_2 = \theta_3 \neq \theta_4$	0.0016	0.0016	0.0009
$\theta_1 \neq \theta_2 = \theta_4 \neq \theta_3$	0.0003	0.0003	0.0002
$\theta_1 \neq \theta_2 \neq \theta_3 = \theta_4$	0.1426	0.1479	0.1154
$\theta_1 \neq \theta_2 \neq \theta_3 \neq \theta_4$	0.0387	0.0899	0.1620

Table 3 indicates that the hypotheses for $H_{true} : \theta_1 = \theta_2 = \theta_3 \neq \theta_4$ has the largest posterior probabilities of 0.7226, 0.6157, and 0.3916 for all priors of the precision parameter α, respectively. This suggests that the data lend greatest support to equalities for $\theta_1 = \theta_2 = \theta_3$ and being different from others.

Table 3 Calculated posterior probabilities for each hypothesis with three cases of (a, b) in Case II

Hypothesis	$(1.0, 1.0)$	$(0.1, 0.1)$	$(0.01, 0.01)$
$\theta_1 = \theta_2 = \theta_3 = \theta_4$	0.0084	0.0446	0.2629
$\theta_1 = \theta_2 = \theta_3 \neq \theta_4$	0.7226	0.6157	0.3916
$\theta_1 = \theta_2 = \theta_4 \neq \theta_3$	0.0008	0.0002	0.0002
$\theta_1 = \theta_2 \neq \theta_3 = \theta_4$	0.0086	0.0088	0.0041
$\theta_1 = \theta_2 \neq \theta_3 \neq \theta_4$	0.0712	0.0759	0.0586
$\theta_1 = \theta_3 = \theta_4 \neq \theta_2$	0.0018	0.0012	0.0008
$\theta_1 = \theta_3 \neq \theta_2 = \theta_4$	0.0025	0.0026	0.0010
$\theta_1 = \theta_3 \neq \theta_2 \neq \theta_4$	0.0237	0.0244	0.0203
$\theta_1 = \theta_4 \neq \theta_2 = \theta_3$	0.0050	0.0050	0.0019
$\theta_1 = \theta_4 \neq \theta_2 \neq \theta_3$	0.0009	0.0007	0.0007
$\theta_1 \neq \theta_2 = \theta_3 = \theta_4$	0.0016	0.0021	0.0007
$\theta_1 \neq \theta_2 = \theta_3 \neq \theta_4$	0.1209	0.1392	0.1151
$\theta_1 \neq \theta_2 = \theta_4 \neq \theta_3$	0.0007	0.0003	0.0003
$\theta_1 \neq \theta_2 \neq \theta_3 = \theta_4$	0.0030	0.0025	0.0014
$\theta_1 \neq \theta_2 \neq \theta_3 \neq \theta_4$	0.0288	0.0770	0.1403

In Table 4, the pairwise posterior probabilities for equality of pairs of θ's are presented. For the equality of $\theta_3 = \theta_4$, we see that the equality has the large posterior probabilities (0.8469, 0.7923, and 0.7497) for three cases of (a, b). This again suggests that there is strong evidence in the equality $\theta_3 = \theta_4$. For the equality of $\theta_1 = \theta_2$, we have the highest posterior probabilities (0.7911, 0.7393, and 0.7095) for all the priors of precision parameter when compared with other equalities. Again this is strong evidence in favor of the hypothesis, $\theta_1 = \theta_2$ and $\theta_3 = \theta_4$.

Table 4 Pairwise Posterior Probabilities with three cases of (a, b) in Case I

Hypothesis	$(1.0, 1.0)$	$(0.1, 0.1)$	$(0.01, 0.01)$
$\theta_1 = \theta_2$	0.7911	0.7393	0.7095
$\theta_1 = \theta_3$	0.0278	0.0995	0.3090
$\theta_1 = \theta_4$	0.0258	0.0984	0.3080
$\theta_2 = \theta_3$	0.0341	0.1054	0.3119
$\theta_2 = \theta_4$	0.0321	0.1035	0.3108
$\theta_3 = \theta_4$	0.8469	0.7923	0.7497

Table 5 demonstrates the pairwise posterior probabilities for equality of pairs of θ's in case II. We observe in Table 5, that the equalities $(\theta_2 = \theta_3)$, $(\theta_1 = \theta_2)$ and

($\theta_1 = \theta_3$) have the largest posterior probabilities of (0.8584, 0.8066, 0.7722), (0.8114, 0.7451, 0.7174), and (0.7588, 0.6884, 0.6766), respectively. This suggests that there is strong evidence for the equalities ($\theta_1 = \theta_2$), ($\theta_2 = \theta_3$) and ($\theta_1 = \theta_3$).

Table 5 Pairwise Posterior Probabilities with three cases of (a, b) in Case II

Hypothesis	$(1.0, 1.0)$	$(0.1, 0.1)$	$(0.01, 0.01)$
$\theta_1 = \theta_2$	0.8114	0.7451	0.7174
$\theta_1 = \theta_3$	0.7588	0.6884	0.6766
$\theta_1 = \theta_4$	0.0168	0.0516	0.2664
$\theta_2 = \theta_3$	0.8584	0.8066	0.7722
$\theta_2 = \theta_4$	0.0138	0.0498	0.2652
$\theta_3 = \theta_4$	0.0232	0.0590	0.2699

Up to this point, we have considered the problem of developing a Bayesian multiple comparisons for means of I negative binomial populations. 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 population means.

An extension of the method to the multiple comparison problems for the another populations would be accomplished straightforwardly. 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. Antoniak, C.E. (1974), Mixtures of Dirichlet Processes with Applications to Nonparametric Problems, *The Annals of Statistics*, 2, 1152-1174.

2. Blackwell, D. and MacQueen, J.B. (1973), Ferguson Distribution via Polya Urn Schemes, *The Annals of Statistics*, 1, 353-355.

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

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

5. 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. (1995), Bayesian Density Estimation and Inference using Mixtures, *Journal of the American Statistical Association*, 90, 577-588.

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

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

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

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

12. West, M., Müller, P. and Escobar, M.D. (1994), Hierarchical Priors and Mixture Models, with Application in Regression and Density Estimation, *Aspects of Uncertainty : A Tribute to D. V. Lindley* (eds: A.F.M. Smith and P.R. Freeman), London : John Wiley and Sons, 363-386.