The Korean Communications
in Statistics Vol. 8, No. 2, 2001
pp. 427-434

Semiparametric Bayesian Multiple Comparisons for Poisson Populations

Jang Sik Cho[1], Dal Ho Kim[2] and Sang Gil Kang[3]

Abstract

In this paper, we consider the nonparametric Bayesian approach to the multiple comparisons problem for I Poisson populations using Dirichlet process priors. We describe Gibbs sampling algorithm for calculating posterior probabilities for the hypotheses and calculate posterior probabilities for the hypotheses using Markov chain Monte Carlo. Also we provide a numerical example to illustrate the developed numerical technique.

Keywords : Configuration; Gibbs Sampler; Mixture of Dirichlet Processes; Multiple Comparisons; Nonparametric Bayes; Poisson Populations.

1. Introduction

Poisson population occurs quite frequently as a model for many random phenomena which require a count of some sort. For example, one might consider the number of electronic components that fail per unit time, the number of radioactive particles emitted per unit time, or the number of telephone calls coming into a telephone switchboard per unit time.

In this paper, we consider I Poisson populations with means $(\theta_1, \theta_2, \cdots, \theta_I)$. The research for Poisson populations was provided by many authors. Ghosh and Parsian(1981) computed Bayes minimax estimation of multiple Poisson parameters. Albert(1981) considered the simultaneous estimation of means from independent Poisson populations. Albert(1983) obtained a pseudo-Bayes confidence region for I Poisson means. Albert (1985) discussed the simultaneous estimation of Poisson means under exchangeable and independent models. Ngai and Stroud(1994) provided the hierarchical Bayes simultaneous estimation of Poisson means.

The multiple comparisons problem(MCP) is to make inferences concerning relationships

1) Assistant Processor, Department of Statistical Information Science, Kyungsung University, Pusan, 608-736, Korea.
E-mail : jscho@star.kyungsung.ac.kr

2) Assistant Processor, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

3) Lecturer, Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

among the θ's based on data. In many cases, the multiple comparisons for Poisson populations is often important in statistical analyses. But, the study about the MCP of I Poisson population means has not been seen yet, in part because of the difficulty in handling the computations. So, we introduce Bayesian approach to resolving the MCP of I Poisson population means. But assessing a prior distribution and formulating a likelihood in the presence of large number N of hypotheses H_0 : $\theta_1 = \theta_2 = \cdots = \theta_I$, H_1 : $\theta_1 \neq \theta_2 = \cdots = \theta_I$ and so on up to H_N : $\theta_1 \neq \theta_2 \neq \cdots \neq \theta_I$ make the Bayesian approach difficult, since the number of hypotheses increase exponentially as the number of I of populations increases. Thus the MCP of I Poisson population means is tedious for moderate I and it is practically impossible for large I. Hence, we need to circumvent resolve this problems.

In this paper, we consider the nonparametric Bayesian approach to the multiple comparisons problem for I Poisson populations using Dirichlet process priors(DPPs) introduced by Ferguson(1973). And we develop a numerical technique to calculate the posterior probabilities of the hypotheses based on a hierarchical nonparametric family of DPP. Also we provide a numerical example to illustrate the developed numerical technique.

2. Mixture of Dirichlet Process Model

The DPP G is determined by two parameters: a distribution function $G_0(\cdot)$ and a positive scalar precision parameter α. Here $G_0(\cdot)$ defines the location of the DPP. So $G_0(\cdot)$ is called by 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. By way of notation we write $G \sim D(G \mid G_0, \alpha)$. For large values of α, a sampled G is very likely to be close to G_0. For small values of α, a sampled G is likely to put most of its probability mass on just a few atoms. Consider I Poisson populations with means $(\theta_1, \theta_2, \cdots, \theta_I)$. Observations $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_I)$ are available on these populations, where $Y_i = (y_{i1}, \cdots, y_{in_i})$ 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} is

$$f(y_{ij} \mid \theta_i) = \frac{\theta_i^{\,y_{ij}} \exp(-\theta_i)}{y_{ij}!}. \tag{2.1}$$

The MCP of I means is to make inferences concerning relationships among the $\underline{\theta}$ based on $\boldsymbol{Y}$. Let $\Theta = \{\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_I) : \theta_i \in R,\ i = 1, 2, \cdots, I\}$ 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 : $\underline{\theta_0}=\{\theta_i; \theta_1=\theta_2=\cdots=\theta_I\}$, H_1 : $\underline{\theta_1}=\{\theta_i; \theta_1\neq\theta_2=\cdots=\theta_I\}$ and so on up to H_N : $\underline{\theta_N}=\{\theta_i; \theta_1\neq\theta_2\neq\cdots\neq\theta_I\}$. The hypotheses $H_r; \underline{\theta_r}$, $r=0,1,2,\cdots,N$, are disjoint, and $\bigcup_{r=0}^{N}\underline{\theta_r}=\Theta$.

For the prior distribution of I Poisson population means, θ_i's, we use the family of DPPs introduced by Ferguson(1973) and extended to mixtures of DPP by Antoniak(1974).

We assume that the θ_i's come from G, and that $G \sim D(G \mid G_0, \alpha)$. This structure results in a posterior distribution which is a mixture of Dirichlet processes (Antoniak 1974). From Antoniak(1974)'s results, the joint posterior distribution has the form

$$\theta_i \mid \boldsymbol{Y} \propto \prod_{i=1}^{I} f(\boldsymbol{y}_i \mid \theta_i)\frac{\alpha G_0(\theta_i)+\sum_{k<i}\delta(\theta_i \mid \theta_k)}{\alpha+i-1}, \tag{2.2}$$

where $\delta(\theta_i \mid \theta_k)$ is the distribution which is a point mass on θ_k.

Also the conditional posterior distribution of θ_i is given by

$$\theta_i \mid \theta_k, k\neq i, \boldsymbol{Y} \propto q_0 G_b(\theta_i \mid y_i)+\sum_{k\neq i} q_k\delta(\theta_i \mid \theta_k), \tag{2.3}$$

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

The elements of Θ themselves behave as described by (2.3) and so with positive probability, they will reduce to some $p\le I$ distinct values.

Let superscript $*$ be distinct values of θ's. Then any realization of I parameters θ_i generated from G lies in a set of $p\le I$ distinct values, denoted by $\theta^*=(\theta_1^*, \theta_2^*, \cdots, \theta_p^*)$.

Definition (Configuration) The set of indices $S=\{S_1,\cdots,S_I\}$ determines a one-way classification of the data $\boldsymbol{Y}=\{y_1,\cdots,y_I\}$ into I^* distinct groups or clusters; the n_j is the number of the set $\{i : S_i=j\}$ 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.

Therefore, there is a one-to-one correspondence between hypotheses and configurations. And the required computations are reduced by the fact that 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 (Antoniak 1974). Then, the above formula can be rewritten as:

$$\theta_i \mid \theta_k, k\neq i, \boldsymbol{y} \propto q_0 G_b(\theta_i \mid \boldsymbol{y}_i)+\sum_{k=1}^{I^*} n_k q_k^*\delta(\theta_i \mid \theta_k^*), \tag{2.4}$$

with $q_k^* \propto f(\boldsymbol{y}_i \mid \theta_k^*)$, and $1=q_0+\sum_k n_k q_k^*$. Besides simplifying notation, the cluster structure of the θ_i can also be used to improve the efficiency of the algorithm.

3. Posterior Sampling In Dirichlet Process Priors

We take a baseline gamma prior G_0 which the θ_i are independent gamma with parameters $\underline{\lambda}_i = \{\lambda_{1i}, \lambda_{2i}\}$, where specified shape parameter is λ_{1i} and unknown scale parameter is λ_{2i}. Extending to a Dirichlet process analysis as outline in the above description results in

$$\boldsymbol{y}_i \mid \theta_i \sim Poisson(\boldsymbol{y}_i \mid \theta_i), \tag{3.1}$$

$$\theta_i \mid G \sim G(\theta_i \mid \underline{\lambda}_i), \tag{3.2}$$

$$G \mid G_0, \alpha \sim D(G \mid G_0, \alpha), \tag{3.3}$$

$$G_0 \mid \underline{\lambda}_i \sim Gamma(\lambda_{1i}, \lambda_{2i}), \tag{3.4}$$

$$\lambda_{2i} \sim Gamma(c, d). \tag{3.5}$$

Now the choice of the precision parameter α in Dirichlet process is extremely important for the model. Escobar and West(1995) and Liu(1996) considered various methods for the choice of α. Here, we consider the gamma prior for α with a shape parameter a and scale parameter b, that is, $\alpha \sim Gamma(a, b)$. Then the $Gamma(a, b)$ is to be the reference prior by $a \rightarrow 0$ and $b \rightarrow 0$.

By Escobar and West(1995), we have access to a neat data augmentation device for sampling α as followings;

$$\alpha \mid \eta, I^* \sim \pi_\eta Gamma(a + I^*, b - \log(\eta)) + (1 - \pi_\eta) Gamma(a + I^* - 1, b - \log(\eta)), \tag{3.6}$$

$$\eta \mid \alpha, I^* \sim \text{Beta}(\alpha + 1, I), \tag{3.7}$$

where the weights π_η are defined in odds form by

$$\frac{\pi_\eta}{(1 - \pi_\eta)} = \frac{a + I^* - 1}{I(b - \log(\eta))}. \tag{3.8}$$

These distributions are well defined for all gamma priors, all η in the unit interval and all $I^* > 1$.

Since $G_b(\theta_i \mid \boldsymbol{y}_i)$ is the baseline posterior distribution, $G_b(\theta_i \mid \boldsymbol{y}_i) \propto Poisson(\boldsymbol{y}_i \mid \theta_i) \times Gamma(\lambda_{1i}, \lambda_{2i})$. Hence, from (2.3), the conditional posterior distributions are given by

$$\theta_i \mid \boldsymbol{y}, \theta_k, k \neq i, \alpha, \underline{\lambda}_i \sim q_0 Gamma(\lambda_{1i} + \sum_{j=1}^{n_i} y_{ij}, \lambda_{2i} + n_i) + \sum_{k \neq i} q_k \delta(d\theta_i \mid \theta_k). \tag{3.9}$$

Also, the above formula can be rewritten using the configuration notations:

$$\theta_i \mid \boldsymbol{y}, \theta_k, k \neq i, \alpha, \underline{\lambda}_i \sim q_0 Gamma(\lambda_{1i} + \sum_{j=1}^{n_i} y_{ij}, \lambda_{2i} + n_i) + \sum_{k=1}^{I} n_k q_k^* \delta(\theta_i \mid \theta_k^*), \tag{3.10}$$

$$\theta_j^* \mid \boldsymbol{y}, S, \underline{\lambda}_j^* \sim Gamma(\sum_{r \in J_i} \sum_{j=1}^{n_r} y_{rj} + \lambda_{1i}, \sum_{r \in J_i} n_r + \lambda_{2j}^*). \tag{3.11}$$

Here since $q_0 \propto \alpha \int f(\boldsymbol{y}_i \mid \theta_i) \, dG_0(\theta_i)$, $q_k \propto f(\boldsymbol{y}_i \mid \theta_k)$ and $1 = q_0 + \sum_{k \neq i} q_k$, q_0 and q_k are given as following, respectively.

$$q_0 \propto \alpha \frac{(\sum_{j=1}^{n_i} y_{ij} + \lambda_{1i} - 1)!}{\prod_{j=1}^{n_i} y_{i}!(\lambda_{1i} - 1)!} \left(\frac{\lambda_{2i}}{n_i + \lambda_{2i}} \right)^{\lambda_{1i}} \left(\frac{1}{1 + \lambda_{2i}} \right)^{\sum_{j=1}^{n_i} y_{ij}},$$

and, for each $k>0$, $q_k \propto \theta_k^{* \sum_{j=1}^{n_i} y_{ij}} \exp(-\theta_k^*) / \prod_{j=1}^{n_i} y_{ij}!$.

$$\alpha \mid \eta, I^* \sim \pi_\eta Gamma(a + I^*, b - \log(\eta)) + (1 - \pi_\eta) Gamma(a + I^* - 1, b - \log(\eta)), \quad (3.12)$$

$$\eta \mid \alpha, I^* \sim Beta(\alpha + 1, I), \quad (3.13)$$

$$\lambda_{2j}^* \mid y, \theta_j^* \sim Gamma(\lambda_{1i} + c, \theta_j^* + d), \quad (3.14)$$

Gibbs sampling proceeds by simply iterating through (3.9) - (3.14) in order, sampling at each stage based on current values of all the conditioning variates.

The configuration gives the equality and inequality relations among the θ's, which correspond to the partitions on the parameter space Θ and in turn to the hypotheses of interest(multiple comparisons).

To estimate the posterior probability of a hypothesis H_r from a large number(L) of sample draws, use

$$P(H_r \mid Y) \approx \frac{1}{L} \sum_{l=1}^{L} \delta_{S_l}(H_r), \quad (3.15)$$

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

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

$$P(\theta_i = \theta_j \mid Y) \approx \frac{1}{L} \sum_{l=1}^{L} \delta_{S_l}(\theta_i = \theta_j) = \sum_{r=1}^{N} P(H_r \mid Y) \delta_{H_r}(\theta_i = \theta_j), \ i \neq j, \quad (3.16)$$

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. Numerical Example And Conclusion

In this section, an artificial data are used to illustrate the multiple comparisons for 4 Poisson populations and sample size of 4 from each populations with $\theta_1 = \theta_2 = 3.0$, $\theta_3 = 8.0$ and $\theta_4 = 1.0$, respectively. That is I=4 and $n_i = 4$ for all $i = 1, 2, \cdots, I$. Then true hypothesis is H_{true} : $\theta_1 = \theta_2 \neq \theta_3 \neq \theta_4$. In this case, the number of possible hypothesis is 15. The observed summary statistics for each populations are given in Table 1.

Table 1. The observed summary statistics for each populations

Populations	1	2	3	4
$\sum_{j=1}^{n_i} y_j$	11	12	31	3
M.L.E.	2.75	3.0	7.75	0.75

For the precision parameter α, we consider three priors: *Gamma*(1.0, 1.0), *Gamma*(0.1, 0.1) and *Gamma*(0.01, 0.01). The latter prior is 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.

Table 2 give the the calculated posterior probabilities for each of 15 possible hypotheses approximated by the Gibbs sampling algorithm using 10,000 iterations with 5,000 burn-in iterations and 5 replications. The hypothesis $\theta_1=\theta_2\neq\theta_3\neq\theta_4$ has most large posterior probabilities 0.4404, 0.4398 and 0.4101 for each prior of the precision parameter α. This suggests that the data lend greatest support to equality for θ_1, θ_2 and θ_3 and θ_4 being different from the others. And the hypothesis $\theta_1=\theta_2=\theta_4\neq\theta_3$ has secondly large posterior probabilities 0.4092, 0.3565 and 0.2632 for each prior of the precision parameter α.

Table 2. Calculated posterior probabilities for each hypothesis

Hypothesis	*Gamma*(1.0, 1.0)	*Gamma*(0.1, 0.1)	*Gamma*(0.01, 0.01)
$\theta_1=\theta_2=\theta_3=\theta_4$	0.0006	0.0042	0.0158
$\theta_1=\theta_2=\theta_3\neq\theta_4$	0.0325	0.0231	0.0186
$\theta_1=\theta_2=\theta_4\neq\theta_3$	0.4092	0.3565	0.2632
$\theta_1=\theta_2\neq\theta_3=\theta_4$	0.0000	0.0000	0.0000
$\theta_1=\theta_2\neq\theta_3\neq\theta_4$	0.4404	0.4398	0.4101
$\theta_1=\theta_3=\theta_4\neq\theta_2$	0.0000	0.0000	0.0000
$\theta_1=\theta_3\neq\theta_2=\theta_4$	0.0017	0.0011	0.0006
$\theta_1=\theta_3\neq\theta_2\neq\theta_4$	0.0033	0.0035	0.0044
$\theta_1=\theta_4\neq\theta_2=\theta_3$	0.0061	0.0055	0.0040
$\theta_1=\theta_4\neq\theta_2\neq\theta_3$	0.0242	0.0267	0.0260
$\theta_1\neq\theta_2=\theta_3=\theta_4$	0.0000	0.0000	0.0001
$\theta_1\neq\theta_2=\theta_3\neq\theta_4$	0.0091	0.0099	0.0077
$\theta_1\neq\theta_2=\theta_4\neq\theta_3$	0.0279	0.0271	0.0264
$\theta_1\neq\theta_2\neq\theta_3=\theta_4$	0.0000	0.0000	0.0000
$\theta_1\neq\theta_2\neq\theta_3\neq\theta_4$	0.0451	0.1025	0.2229

Table 3 indicates the pairwise posterior probabilities for equality of pairs of θ's. The hypothesis for $\theta_1 = \theta_2$ has most large posterior probability(0.8827, 0.8236 and 0.7078) for each choices of precision parameter. This suggests that there is strong evidence in the hypothesis for $\theta_1 = \theta_2$.

Table 3. Pairwise Posterior Probabilities

Hypothesis	*Gamma*(1.0, 1.0)	*Gamma*(0.1, 0.1)	*Gamma*(0.01, 0.01)
$\theta_1 = \theta_2$	0.8827	0.8236	0.7078
$\theta_1 = \theta_3$	0.0380	0.0320	0.0396
$\theta_1 = \theta_4$	0.4601	0.3929	0.3090
$\theta_2 = \theta_3$	0.0483	0.0427	0.0463
$\theta_2 = \theta_4$	0.4594	0.3889	0.3062
$\theta_3 = \theta_4$	0.0007	0.0042	0.0160

Until now, we have considered the problem of developing a Bayesian multiple comparisons for means of I Poisson 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 MCP 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] Albert, J. H. (1981), Simultaneous Estimation of Poisson Means, Journal of *Multivariate Analysis*, 11, 400-417.

[2] Albert, J. H. (1983), A Pseudo-Bayes Confidence Region for p Poisson Means, *Journal of Statistical Computation and Simulation*, 19, 11-29.

[3] Albert, J. H. (1985), Simultaneous Estimation of Poisson Means Under Exchangeable and Independence Models, *Journal of Statistical Computation and Simulation*, 23, 1-14.

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

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

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

[7] Ghosh, M. and Parsian, A. (1981), Bayes Minimax Estimation of Multiple Poisson Paramters, *Journal of Multivariate Analysis*, 11, 280-288.

[8] Liu, J. S. (1996), Nonparametric Hierarchical Bayes via Sequential Imputations, *The Annals of Statistics*, 24, 911-930.

[9] Ngai, H. M. and Stroud, T. W. F. (1994), Hierarchical Bayes Simultaneous Estimation of Poisson Means, *Communication Statistics-Theory and Method*, 23(10), 2965-2991.