한국통계학회논문집 제2권 1호
1995년 4월 pp. 137-144

Hierarchical Bayes Estimators of Exchangeable Poisson Mean using Laplace Approximation

Younshik Chung[1)]

Abstract

Hierarchical Bayes estimations of exchangeable mean vector of a multivariate Poisson distribution are obtained. Since sophiscated analytic integration procedures are needed, the Laplace method is employed in order to compute these estimations approximately. An example is presented.

1. Introduction

When data are collected from many units that are some how similar, such as subjects, animals, cities, etcetra, statistcal problem is to combine the information from the various units to understand better the phenomenon under study. Usually there is substantial variability among units and a natural way to approach the problem is to build a two-stage "hierarchical model" and then use it to make inferences.

Suppose independent random variables $X_1, \cdots, X_p$ are observed, where X_i is distributed Poisson with mean $t_i \theta_i$ where t_i is known. Consider the problem of simultaneous estimating $\theta_1, \cdots, \theta_p$, where the $\theta_i's$ are believed a priori to satisfy an r-dimensional model, where $r < p$. The $\theta_i's$ are assumed to be exchangeable, which will be modelled via a hierarchical Bayesian approach. The hierarchical approach of constructing priors has also been shown useful in multinomial and binomial estimatation problems in Albert and Gupta (1983) and Albert (1984). Gaver (1985) and Gaver and O'Muircheartaigh (1987) considered a flat-tailed first stage priors for Poisson event rates.

In section 2, the necessary integrative operations are not analytically tractable, and so we are forced to use computational techniques and we introduce the Laplace method (Tierney and Kadane, 1986). Section 3 presents the hierarchical Bayesian formulation of the problem. At the first stage of the prior, the $\theta_i's$ are assigned independent gamma distributions. At the second stage, noninfomative prior densities are assigned to the hyperparameters. In section 4, the Laplace method is employed and these Bayesian methods are used for the pump failures data which was analyzed by Gaver and O'Muircheataigh (1987).

---

1) Department of Statistics, Research Institute of Information and Communication, Pusan National University, Pusan, 609-735, KOREA. E-mail : yschung@hyowon.pusan.ac.kr

2. Laplace Approximation

An important issue in Bayesian inference is the calculation of marginal posterior expectations of functions $g(\theta)$ of unknown parameter θ. In many applications, functions of θ, $g(\theta)$ are of interest, e.g. $\theta_i - \theta_j$, θ_i / θ_j, $\max \theta_i$. The technical problem encountered in attempting to carry out required numerical integrations to obtain these distributions and expectations have long served as an impediment to wilder application of the Baysian framework to real data. Substantial effort has been devoted to the development of analytic approximation for calculating expectations. In order to resolve these problems, there are many methods such as Laplace's method, importance sampling method and Gibbs sampler. The last two methods are applicable to researchers who are experienced to computer. But the Laplace's method is easier even to statisticians who are less familiar to computer because it needs the elementary mathematical computing(twice differentiation). So Most of these works involve application of Laplace's method (see e.g. DeBruijin, 1961). Writing these expectations as ratios of integrals and approximating numerator and denominator seperately, Tierney and Kadane (1986) obtain a second order approximation for the expectation of positive functions. We thus briefly review the method as follows:

Consider a set of random variables $\theta = (\theta_1, \cdots, \theta_p)$. Our interest is in calculating $Eg(\theta)$. Suppose $g > 0$ and suppose that the joint density of $\theta_1, \cdots, \theta_p$ is known upto normalizing constant, i.e. is proportional to $f(\theta_1, \cdots, \theta_p)$. Then

$$Eg(\theta) = \frac{\int g \cdot f}{\int f} = \frac{\int e^{-l^*}}{\int e^{-l}} \approx \left[\frac{|\Sigma^*|}{|\hat{\Sigma}|} \right]^{\frac{1}{2}} e^{l(\hat{\theta}) - l^*(\theta^*)} \tag{2.1}$$

where $l = -\log f$, $l^* = l - \log g$, $\hat{\theta}$ is the mode of l, θ^* is the mode of l^*, $\hat{\Sigma}$ and Σ^* are minus the inverse Hessians of l and l^* evaluated at $\hat{\theta}$ and θ^* respectively and $|A|$ denotes the determinant of matrix A. The form (2.1) first appeared in Tierney and Kadane (1986) who noted that when $\log f = O(p)$ approximation is then accurate in the order of p^{-2}.

3. Hierarchical Bayes estimation of exchangeable Poisson means

3.1. Hierarchical Prior and Preliminaries

Let $X=(X_1, \cdots, X_p)^T$ have a p-variate Poisson distribution with mean vector $t^T = (t_1\theta_1, \cdots, t_p\theta_p)^T$ where t_i is known. In addition, suppose that the parameters of interest $\theta_1 \cdot, \theta_p$, are believed to be exchangeable, which is typically modeled via a two stage prior

$$\theta | \alpha, \beta \sim \pi_1(\theta | \alpha, \beta) = \prod_{k=1}^{p} \pi_{1k}(\theta_k | \alpha, \beta), \quad \lambda = (\alpha, \beta) \sim \pi_2(\alpha, \beta).$$

Then the posterior mean of $g(\theta_j)$ can written as

$$E[\ g(\theta_j) | x\ ] = \frac{\int\int g(\theta_j | \alpha, \beta) \prod_{k=1}^{p} m(x_k | \alpha, \beta) \pi_2(\alpha, \beta)\, d\alpha\, d\beta}{\int\int \prod_{k=1}^{p} m(x_k | \alpha, \beta) \pi_2(\alpha, \beta)\, d\alpha\, d\beta} \tag{3.1}$$

where

$$g(\theta_j | \alpha, \beta) = \int g(\theta_j) \cdot \pi_1(\theta_j | \alpha, \beta, x_j)\, d\theta_j\,,\ m(x_k | \alpha, \beta) = \int f(x_k | \theta_k) \pi_1(\theta_k | \alpha, \beta)\, d\theta_k$$

$$\pi_1(\theta_k | \alpha, \beta, x_k) = \frac{f(x_k|\theta_k)\pi_1(\theta_k | \alpha, \beta)}{m(x_k | \alpha, \beta)} \quad \text{and} \quad f(x_k|\theta_k) = \frac{e^{-\theta_k}\theta_k^{x_k}}{x_k!}.$$

For proof, see Berger(1985).

A common choice for $\pi_1(\theta_k | \alpha, \beta)$ is a gamma density, while the noninformative choice for $\pi_2(\alpha, \beta)$ is $\pi_2(\alpha, \beta) = 1$ or $\pi_2(\alpha, \beta) = \beta^{-1}$. In this problem, this first stage informative prior will be modelled by assuming that $\theta_1, \cdots, \theta_p$ is a random sample from the conjugate gamma (α, β) distribution (with mean α / β and variance α / β^2). Typically, the hyperparameters α and β are difficult to specify : Priors of α and β is given as a noninformative density π_2 .

3.2. Poisson - Gamma

Theorem 3.1. Suppose that $X_i \sim P_o(t_i \theta_i)$ independently for $i = 1, \cdots, p$ and that $\theta_i \sim gam(\alpha,\ \beta^{-1})$ independently for $i = 1, \cdots, p$ where t_i, α and β are known. Then

$$\begin{aligned} m(x_j | \alpha, \beta) &= \text{marginal density of } x_j \text{ given } \alpha, \beta \\ &= \frac{\beta^{\alpha} t_j^{x_j}}{\Gamma(\alpha)\, x_j!} \Gamma(x_j + \alpha) \left(\frac{1}{\beta + t_j}\right)^{x_j + \alpha} \end{aligned} \tag{3.2}$$

$$E(\theta_j \mid \alpha, \beta) = \frac{x_j + \alpha}{t_j + \beta} \tag{3.3}$$

and

$$Var(\theta_j \mid \alpha, \beta) = \frac{x_j + \alpha}{(\beta + t_j)^2}. \tag{3.4}$$

Proof. It is a straightforward computation.

Note that the posterior covariance of θ_j and θ_k, given α and β, is equal to 0 for all $j \neq k$ since the X_j's are independent, as θ_j's are a priori.

3.3. Hierarchical Bayes Estimator

From equation (2.1), it is clear that $E(\theta_j|x)$ is obtained by integrating $E(\theta_j \mid \alpha, \beta)$ with respect to the density

$$\pi_2(\alpha, \beta \mid X) = \frac{\{\prod_{k=1}^{p} m(x_k \mid \alpha, \beta)\}\, \pi_2(\alpha, \beta)\, d\alpha d\beta}{\int_0^{\infty}\int_{\infty}^{-\infty} \{\prod_{k=1}^{p} m(x_k \mid \alpha, \beta)\}\, \pi_2(\alpha, \beta)\, d\alpha d\beta}.$$

Theorem 3.2. Let $X_i \sim P_o(\theta_j t_j)$ and $\theta_j \sim Gam(\alpha, \beta)$, $j = 1, \cdots, p$ where t_j is known but α, β are unknown. If $\pi_2(\alpha, \beta)$ is a prior density of (α, β), then, provided that all the integrals exist,

$$\begin{aligned} m(x) &= \text{marginal density of } (X_1, \cdots, X_p)^T \\ &= \int_0^{\infty}\int_0^{\infty} \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^p \prod_{j=1}^{p}\left[\frac{t_j^{x_j}}{x_j!}\Gamma(x_j+\alpha)\left(\frac{1}{\beta+t_j}\right)^{x_j+\alpha}\right] \pi_2(\alpha, \beta)\, d\alpha d\beta, \end{aligned} \tag{3.5}$$

$$E(\theta_j \mid x) = E^{\pi_2(\alpha, \beta \mid x)}\left[\frac{x_j + \alpha}{t_j + \beta}\right], \tag{3.6}$$

$$Var(\theta_j \mid x) = E^{\pi_2(\alpha, \beta \mid x)}\left[\frac{(x_j+\alpha+1)(x_j+\alpha)}{(t_j+\beta)^2}\right] - (E(\theta_j|x))^2. \tag{3.7}$$

$$\begin{aligned} V^{H_{j,k}} &= \text{posterior covariance of } \theta_j \text{ and } \theta_k \\ &= \int\int (\theta_j - E(\theta_j|x))(\theta_k - E(\theta_k|x))\, \pi(\theta_j, \theta_k \mid x)\, d\theta_j d\theta_k \end{aligned}$$

where $E^{\pi_2(\alpha, \beta \mid x)}$ stands for expectation over α, β with respect to

$$\pi_2(\alpha, \beta \mid x) = \frac{\left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^p \prod_{j=1}^{p}\left[t_j^{x_j}\Gamma(x_j+\alpha)\left(\frac{1}{\beta+t_j}\right)^{x_j+\alpha}\right] \pi_2(\alpha, \beta)}{m(x)}$$

Proof. Its proof is a straightforward computation using (3.1), (3.2) and Theorem 3.1.

4. Illustrative Example

We apply the hierarchical Bayes estimator of exchangeable Poisson model discussed in section 2. to pump failure data previously analyzed by Gaver and O'Muircheartaigh (1987). In table 1, there appears a small set of data representing failures of pumps in renewal systems of the nuclear plant Farley 1. These data may be found in an Electric Power Research Institute (EPRI) report (Worledge, Stringhan and McClymont 1982)and x_i denotes the number of failures and t_i is the length of time in thousands of hours.

Table 1. Pump Failures

system	x_i	t_i	$\rho_i(\times 10^2)$
1	5	94.320	5.3
2	1	15.720	6.4
3	5	62.880	8.0
4	14	25.760	11.1
5	3	5.240	57.3
6	19	31.440	60.4
7	1	1.048	95.4
8	1	1.048	95.4
9	4	2.096	191.0
10	22	10.480	209.09

We illustrated the hierarchical Bayes estimate given by (3.6) for this data set using the Laplace method in (2.1). In this problem, the first is due to the independence of $X_1, \cdots, X_p$ conditionally on λ. The compound sampling density of the data $X=(X_1, \cdots, X_p)$ becomes $f(x\mid\lambda)=\prod_{i=1}^{p} f(x_i\mid\lambda)$, where $f(x_i\mid\lambda) = \int f(x_i\mid\theta_i,\lambda)\,\pi_1(\theta_i\mid\lambda)\,d\theta_i$ and the likelihood function $L(\cdot)$ on Δ becomes $L(\lambda)=\prod_{i=1}^{k} L_i(\lambda)$ where $L_i(\lambda) = f(x_i\mid\lambda)$. Similarly, $\pi(\theta_i\mid x,\lambda) = \pi(\theta_i\mid x_i,\lambda)$. Now suppose that g is a real-valued function on Θ. Then $E[\,g(\theta_i)\mid x,\lambda\,] = E[\,g(\theta_i)\mid x_i,\lambda\,]$, so the positive expectation of $g(\theta_i)$ is $E[\,g(\theta_i)\mid x\,] = E_\lambda E[\,g(\theta_i)\mid x_i,\lambda\,]$ where E_λ denote the expectation with respect to the posterior distribution $\pi(\lambda | x)$ of λ. Also the posterior variance is $V(g(\theta_i) | y) = E_\lambda[V(g(\theta_i) | x_i, \lambda)] + V_\lambda[E(g(\theta_i) | x_i, \lambda)]$ where V_λ is the variance with respect to $\pi(\lambda | x)$. Let $G(\lambda) = E[g(\theta_i) | x_i, \lambda]$ be a function of λ. Then the positive expectaion of $G(\lambda)$ may be written in the form

$$E[G(\lambda) | x] = \frac{\int G(\lambda) L(\lambda) \pi(\lambda)\, d\lambda}{\int L(\lambda) \pi(\lambda)\, d\lambda} . \tag{4.1}$$

Due to Tierney and Kadane (1986), when G is a positive function,

$$E[G(\lambda) | x] = \frac{|\Sigma^*|^{\frac{1}{2}} \exp(l^* (\lambda^*))}{|\hat{\Sigma}|^{\frac{1}{2}} \exp(l(\hat{\lambda}))} (1 + O(p^{-2}))$$

where $l(\lambda) = \log(L(\lambda)\, \pi(\lambda))$, $l^*(\lambda) = l(\lambda) + \log(G(\lambda))$, $\hat{\lambda}$ and λ^* are modes of l and l^* respectively, $\hat{\Sigma}$ and Σ^* are minus the inverse Hessians of l and l^* evaluated at $\hat{\lambda}$ and λ^* respectively.

By the empirical Bayes argument for this data set mentioned by Gelfand and Smith (1990), let $\alpha = \overline{\rho}^2 / (S_\rho^2 - p^{-1} \overline{\rho} \sum_{i=1}^{p} t_i^{-1})$.

With such a value of α, $\pi_2(\alpha, \beta)$ is considered as a function of only β, say $\pi_2(\beta)$. Since we have no information about β, we consider $\pi_2(\beta) = 1$ or β^{-1} as diffuse and noninformative prior respectively. First of all, we compute $E(\theta_j | x)$ and $Var(\theta_j | x)$ given by (3.6) and (3.7) approximately under $\pi_2(\beta) = 1$. For computing

$$E(\theta_j | x) = E^{\pi_2(\beta|x)}[\frac{x_j + \alpha}{t_j + \beta}] ,$$

let $l(\lambda) = l(\beta) = \log(L(\beta) * \pi(\beta)) = \log[\{ \Gamma(\alpha)\, x_j! \}^{-1} \beta^\alpha t_j^{x_j} \Gamma(x_j + \alpha)(\beta + t_j)^{-(x_j + \alpha)}]$ and $l^*(\lambda) = l(\lambda) + \log G(\lambda) = l(\beta) + \log((x_j + \alpha)/(t_j + \alpha))$. Then $l(\beta)$ and $l^*(\beta)$ are maximized at $\hat{\beta} = t_j \alpha / x_j$ and $\beta^* = t_j \alpha / (x_j + 1)$ respectively. So using (4.1), $E(\theta_j | x)$ can be computed approximately by

$$\hat{E}(\theta_j | x) = \frac{\left| \frac{\partial^2 l^*(\beta^*)}{\partial^2 \beta} \right|^{\frac{1}{2}}}{\left| \frac{\partial^2 l(\hat{\beta})}{\partial^2 \beta} \right|^{\frac{1}{2}}} \frac{\exp(l^* (\beta^*))}{\exp(l(\hat{\beta}))} .$$

Since

$$Var(\theta_j|x) = E^{\pi(\beta|x)}[\ (x_j+\alpha+1)(x_j+\alpha)\ (t_j+\beta)^{-2}]\ -\ E^2(\theta_j|x),$$

let $G_1(\beta) = (x_j+\alpha+1)(x_j+\alpha)\ (t_j+\beta)^{-2}$. Then $l_1^*(\beta) = l(\beta) + \log(G_1(\beta))$ is maximized at $\beta_1^* = t_j\alpha\,/(x_j+2)$. So

$$\widehat{Var}(\theta_j|x) = \frac{\left|\frac{\partial^2 l_1^*(\beta_1^*)}{\partial^2\beta}\right|^{\frac{1}{2}}}{\left|\frac{\partial^2 l(\hat{\beta})}{\partial^2\beta}\right|^{\frac{1}{2}}}\ \frac{\exp(l_1^*(\beta_1^*))}{\exp(l(\hat{\beta}))}\ -\ [\ \hat{E}(\theta_j|x)]^2\ .$$

Similary, we can approximately compute $E(\theta_j|x)$ and $Var(\theta_j|x)$ under $\pi(\beta) = \beta^{-1}$.

According to $\pi_2(\beta) = 1$ or β^{-1}, table 2 shows the approximate values of $E(\theta_j|x)$ and $Var(\theta_j|x)$ for $j = 1, \ldots, p$. In table 2, the (second stage) prior $\pi_2(\beta) = \beta^{-1}$ is better than $\pi_2(\beta) = 1$ considering variability. Therefore $\pi_2(\beta) = \beta^{-1}$ seems to be a reasonable choice for β because β is a scale parameter.

Table 2. Approximate Moments

$\pi(\beta)$	$\pi(\beta)=1$		$\pi(\beta)=\frac{1}{\beta}$	
system	$\hat{E}(\theta_j\|x)$	$\widehat{Var}(\theta_j\|x)$(std.)	$\hat{E}(\theta_j\|x)$	$\widehat{Var}(\theta_j\|x)$(std.)
1	.06626	.00100 (.03165)	.07509	.00090 (.03010)
2	.18789	.02613 (.16166)	.20594	.01704 (.13054)
3	.09939	.00225 (.04748)	.11264	.00203 (.04516)
4	.12006	.00113 (.03363)	.12749	.00107 (.03282)
5	.83655	.25809 (.50803)	.97764	.22195 (.47111)
6	.63852	.02318 (.15227)	.66875	.02230 (.14933)
7	2.8183	5.88041 (2.42495)	3.08912	3.83447 (1.95818)
8	2.8183	5.88041 (2.42495)	3.08912	3.83447 (1.95818)
9	2.5294	1.81363 (1.34671)	2.90996	1.61060 (1.26909)
10	2.2009	.23610 (.48590)	2.29222	.22802 (.47752)

Acknowledgement

Thanks are due to the referees for the constructive comments which led substantial improvement over the original version of the paper.

References

[1] Albert, J. H. (1984). Empirical Bayes estimation of a set of binomial probabilities, *Journal of Statistical Computation and Simulation* 20, 129-144.

[2] Albert, J. H. and Gupta, A. K. (1983). Bayesian estimation methods for 2×2 contingency tables using mixtures of Dirichlet distributions, *Journal of American Statistical Association* 78, 708-717.

[3] DeBruijin, N. G. (1961). Sampling based approaches to calculating marginal densities, *Journal of American Statistical Association* 85, 398-409.

[4] Gaver, D. P. (1985). Discrepancy tolerant hierarchical Poisson event rate analysis, *Technical Report, Naval Postgraduate School, Operations Research Dept.*

[5] Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analysis of event rates, *Technometrics* 29, 1-15.

[6] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities, *Journal of American Statistical Association* 85, 398-409

[7] Tierney, L. and Kadane, J. B. (1986). Accurate approximations for positive moments and marginal densities, *Journal of American Statistical Association* 81, 82-86.

[8] Worledge, D. H., Strigham, R. S., and McClymont, A. S.(1982).PWR Power Plant Reliability Data, *Interium Report NP-2592, Electric Power Research Institute.*