The Korean Communications
in Statistics Vol. 6, No. 1, 1999
pp. 221-235

Hierarchical and Empirical Bayes Estimators of Gamma Parameter under Entropy Loss[1)]

Younshik Chung[2)]

Abstract

Let be $X_1, \cdots, X_p$, $p \geq 2$ independent random variables, where each X_i has a gamma distribution with k_i and θ_i. The problem is to simultaneously estimate p gamma parameters θ_i and θ_i^{-1} under entropy loss where the parameters are believed priori. Hierarchical Bayes(HB) and empirical Bayes(EB) estimators are investigated. And a preference of HB estimator over EB estimator is shown using Gibbs sampler(Gelfand and Smith, 1990). Finally, computer simulation is studied to compute the risk percentage improvements of the HB estimator and the estimator of Dey, Ghosh and Srinivasan(1987) compared to UMVUE estimator of θ^{-1}.

1. Introduction

This paper is devoted to hierarchical and empirical Bayesian estimation of the parameters of p independent gamma distributions under entropy loss. Suppose that $X=(X_1, \cdots, X_p)$ where $X_1, \cdots, X_p$ are p independent random variables, X_i having probability density function which is defined as

$$f_i(x_i) = f(x_i|\theta_i) = \exp(-\theta_i x_i) x_i^{k_i-1} \theta_i^{k_i} / \Gamma(k_i), \quad x_i > 0 \tag{1.1}$$

which is denoted by $\Gamma(k_i, \theta_i)$, where $\theta_i > 0$ unknown and $k_i > 0$ known. The improved estim ation of scale parameters from exponential families has also been studied recently. The major results in this direction are obtained by Hudson(1978), Berger(1980) and Ghosh, Hwang and T sui(1984). Das Gupta(1986) obtained improved estimators of gamma scale parameters without a variational argument by proposing an estimator which is a function of the geometric mean. In this paper, the loss is considered as the entropy distance (or the Kullback-Leibler information number) between two distributions of p independent gamma random variables. Assume that k_i's in (1.1) are known. Then the loss is given as

---

1) This reserach was supported by Research Institute of Computer and Information Communication, Pusan National University, KOREA, 1999.

2) Associate professor, Department of Statistics, Pusan National University, Pusan 609-735 KOREA

$$L(\theta,\delta) = E_\theta\left[\log\frac{\prod_{i=1}^{p}\frac{\theta_i^{k_i}}{\Gamma(k_i)}x_i^{k_i-1}e^{-\theta_i x_i}}{\prod_{i=1}^{p}\frac{\delta_i^{k_i}}{\Gamma(k_i)}x_i^{k_i-1}e^{-\delta_i x_i}}\right] = \sum_{i=1}^{p}k_i[\delta_i\theta_i^{-1}-\ln(\delta_i\theta_i^{-1})-1] \tag{1.2}$$

where ln denotes the natural logarithm. The loss (1.2) was first introduced in James and Stein(1961) for estimation of the variance-covariance matrix of a multinormal distribution. Dey and Chung(1991) and Chung and Dey(1994) investigated this loss(1.2) for estimation of parameters from truncated power series and power series distribution, respectively. Chung, Kim and Dey(1994) considered the weighted entropy loss for estimation of Poisson means. Since mean and the variance of the gamma distribution are the functions of $\frac{1}{\theta_i}$, we are interested in estimation of $\frac{1}{\theta}=(\frac{1}{\theta_1},\cdots,\frac{1}{\theta_p})$. So we consider the loss $L(\theta^{-1},\delta)$ for estimating θ^{-1} as

$$L(\theta^{-1},\delta)=\sum_{i=1}^{p}k_i(\delta_i\theta_i-\ln(\delta_i\theta_i)-1). \tag{1.3}$$

Under the loss (1.3), the best invariant estimator of θ^{-1} is $\delta^0(X)=(\frac{X_1}{k_1},\cdots,\frac{X_p}{k_p})$. It is also easy to check that $\delta^0(X)$ is the minimum variance unbiased estimator(MVUE) of θ^{-1}. And the best invariant UMVUE estimator of θ under the loss (1.2) is given by $\delta^+(X)=(\frac{k_1-1}{X_1},\cdots,\frac{k_p-1}{X_p})$. For, $p=1$ it follows from Brown(1966) that $\delta^0(X)$ and $\delta^+(X)$ are admissible for θ^{-1} and θ, respectively. Let $R(\theta^{-1},\delta(X))=E^X L(\theta^{-1},\delta(X))$ denote the risk function of a decision rule $\delta(X)$. Usually, $\delta^*(X)$ is R-better than $\delta(X)$ if $R(\theta^{-1},\delta^*(X))\le R(\theta^{-1},\delta(X))$ for all values of θ where strict inequality holds for some values of θ. Furthermore, one can say that δ^* is an improved estimator over δ. These losses in (1.2) and (1.3) are considered in Dey, Ghosh and Srinivasan(1987) for simultaneous estimation of p independent gamma scale parameters or their reciprocals. For $p\ge 3$, the improved estimator $\delta^D(X)=(\delta_1^D(X),\cdots,\delta_p^D(X))$ is componentwisely given as

$$\delta_i^D(X) = \frac{X_i}{k_i}\left(1-\frac{k_i}{r(S)}\; b + S(\ln X_i - a_i)\right) \tag{1.4}$$

where $S=\sum_{i=1}^{p}(\ln X_i-a_i)$ and $b>36(p-2)^2/25k^2$, $0<r(S)<6(p-2)/5k^2$ with $k=\max(k_1,\cdots,k_p)$. Although the above estimator $\delta^D(X)$ improves upon the MVUE

$\delta^0(X)$, they have no Bayesian interpretation. So in this paper hierarchical and empirical Bayes approach are considered.

In section 2, where μ in (2.4) and η in (2.15) are known or unknown, hierarchical Bayes(HB) and empirical Bayes(EB) estimators are investigated. Briefly, the hierarchical Bayes approach is a purely Bayesian approach using a hierarchical prior and the empirical Bayes approach estimates the unknown hyperparameter of the first stage prior (2.5) from the marginal density using maximum likelihood method.
If μ or η is unknown, the purely HB approach is avoided in this situation due to computational difficulty; however, a combination EB and HB approach is employed. In section 3, comparisons of two approaches are made. Then Gibbs sampler is applied to compute the HB estimator. Finally, in section 4, comparisons of the HB and the estimator $\delta^D(X)$ in (1.4) of Dey et al.(1987) with the MVUE of θ^{-1} are made according to computer simulation.

2. HB and EB estimators

Our focus is a fully Bayesian parametric approach. Assume that $\theta_1, \cdots, \theta_p$ are believed to be exchangeable. This prior information is represented by letting $\theta_1, \cdots, \theta_p$ a random sample from the conjugate prior

$$g(\theta_i|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \theta_i^{\alpha-1} e^{\theta_i/\beta}, \quad \alpha, \beta > 0 \tag{2.1}$$

with hyperparameters α and β.

Then, the prior density of $\theta = (\theta_1, \cdots, \theta_p)$ is given by

$$\pi(\theta|\alpha, \beta) = \prod_{i=1}^{p} \left\{ \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \theta_i^{\alpha-1} e^{-\frac{\theta_i}{\beta}} \right\}. \tag{2.2}$$

So, the posterior density of θ given $X = (X_1, \cdots, X_p)$, α and β is

$$\pi(\theta|X, \alpha, \beta) = \prod_{i=1}^{p} \left[\frac{(x_i + \frac{1}{\beta})^{\alpha+k_i}}{\Gamma(\alpha+k_i)} \theta_i^{\alpha+k_i-1} e^{-(x_i+\frac{1}{\beta})\theta_i} \right] \tag{2.3}$$

which is a product of p independent gamma distributions with $\alpha + k_i$ and $x_i + \frac{1}{\beta}$, denoted by $\Gamma(\alpha + k_i,\ x_i + \frac{1}{\beta})$, $i = 1, \cdots, p$.

2.1. Estimating θ^{-1} under loss (1.3).

It will be shown useful in sections 2.1.1 and 2.1.2 to reparameterize (α, β) into (γ, μ) where γ and μ are given respectively by

$$\gamma = \frac{1}{\alpha}, \quad \mu = \frac{1}{E(\theta_i|\alpha, \beta)} = \frac{1}{\alpha\beta}. \tag{2.4}$$

Remark that $\mu = \frac{1}{E(\theta_i|\alpha, \beta)}$ is the best prior estimate of θ_i^{-1} in the sense that it minimizes $E(L(\theta^{-1}, \delta)|\alpha, \beta)$. It will be of interest to consider the case when μ is known or unknown. If μ is known, the hierarchical model is as follow :

$$\theta_i|\gamma, \mu \sim \Gamma(\mu, \gamma), \quad \gamma \sim \pi_2(\gamma) \tag{2.5}$$

where the second stage prior π_2 is noninformative. On the other hand, if μ is unknown, the hierarchical model is as above (2.5) with replacement

$$(\mu, \gamma) \sim \pi_2(\mu, \gamma) \tag{2.6}$$

where π_2 is now a noninformative prior for hyperparameters μ and γ.

2.1.1. μ is known.

By reparametrization in (2.4), since $\mu = \frac{1}{\alpha\beta}$, replace $\frac{1}{\beta}$ by $\mu\alpha$. Then the posterior distribution of θ, denoted by $\pi(\theta|X, \alpha, \mu)$, is a product of p independent gamma distributions with $\alpha + k_i$ and $X_i + \alpha\mu$. Under the assumed model and given loss in (1.3), the estimate at the first stage of Bayesian analysis is $\widehat{\theta}^{-1} = (\widehat{\theta}_1^{-1}, \cdots, \widehat{\theta}_p^{-1})$ given componentwisely as

$$\widehat{\theta}_i^{-1} = \frac{1}{E(\theta_i|X, \alpha, \mu)} = \frac{k_i}{\alpha + k_i}\frac{X_i}{k_i} + \frac{\alpha}{\alpha + k_i}\mu. \tag{2.7}$$

Suppose a noninformative prior for θ is $\pi(\theta) = \prod_{i=1}^{p} \theta_i^{-1}$. This prior can be obtained from $\Gamma(\alpha, \frac{1}{\beta})$ by letting $\alpha \to 0$ and $\beta \to \infty$. Also by these limits, $\alpha \to 0$ implies $\widehat{\theta}_i^{-1} \to \frac{X_i}{k_i}$. If a noninformative prior is assumed, $\delta^{\ni}(X) = (\frac{X_1}{k_1}, \cdots, \frac{X_1}{k_p})$ is the derived estimator of θ^{-1} which is also UMUVE $\delta^0(X)$. On the other hand, precise prior beliefs can be modelled by fixing μ and letting $\alpha \to \infty$, and then $\widehat{\theta}_i^{-1} \to \mu$. It is desirable to consider estimating θ^{-1} by a convex combination of $\delta^{\ni}$ and μ. The Bayes estimate (2.7) is rewritten as

$$\hat{\theta}_i^{-1} = \frac{k_i}{\frac{1}{\gamma}+k_i}\,\overline{X}_i + \left(1-\frac{k_i}{\frac{1}{\gamma}+k_i}\right)\mu \tag{2.8}$$

where $\overline{X}_i = \frac{X_i}{k_i}$ and $\gamma = \frac{1}{\alpha}$. To derive HB estimator for θ^{-1}, note that marginally $X_1,\cdots,X_p$ are independent with X_i having the density

$$m(X_i|\gamma,\mu) = \int f(X_i|\theta_i)\pi(\theta_i|\gamma,\mu)\,d\theta_i = \frac{\Gamma(k_i+\frac{1}{\gamma})X_i^{k_i-1}(\frac{u}{\gamma})^{\frac{1}{\gamma}}}{\Gamma(k_i)\Gamma(\frac{1}{\gamma})(X_i+\frac{u}{\gamma})^{k_i+\frac{1}{\gamma}}}. \tag{2.9}$$

Under the second stage prior with $\pi_2(\gamma)$,

$$\delta_i^{HB}(X) = \frac{1}{E(\theta_i|X,\mu)} = \frac{1}{E_1E(\theta_i|X,\mu,\gamma)} = \frac{1}{E_1\left[\frac{\frac{1}{\gamma}+k_i}{X_i+\frac{1}{\gamma}\mu}\right]} \tag{2.10}$$

where E_1 denotes the expectation taken over the posterior $\pi(\gamma|X,\mu)$ of γ given X and μ, and $\pi(\gamma|X,\mu) \propto [\prod_{i=1}^{p} m(X_i|\gamma,\mu)]\pi_2(\gamma)$.

To derive EB estimator for $\hat{\theta}^{-1}$, it is necessary to maximize $m(X|\gamma,\mu) = \prod_{i=1}^{p} m(X_i|\gamma,\mu)$. Let $L(\gamma) = \ln m(X|\gamma,\mu)$. Then it follows that

$$\frac{\partial L}{\partial\gamma} = -\frac{1}{\gamma^2}\left\{\sum_{i=1}^{p}\frac{\Gamma'(\frac{1}{\gamma}+k_i)}{\Gamma(\frac{1}{\gamma}+k_i)} + p\ln\frac{\mu}{\gamma} + \frac{p}{1-\gamma} - \frac{p\Gamma'(\frac{1}{\gamma})}{\Gamma(\frac{1}{\gamma})} - \sum_{i=1}^{p}\ln(X_i+\frac{\mu}{\gamma}) - \mu\sum_{i=1}^{p}\frac{\frac{1}{\gamma}+k_i}{X_i+\frac{\mu}{\gamma}}\right\} \tag{2.11}$$

where $\Gamma'(\gamma)$ denotes $\int x^{\gamma}e^{-x}\log x\,dx$. After computing the optimal $\hat{\gamma}$,

$$\delta_i^{EB}(X) = \frac{k_i}{\frac{1}{\hat{\gamma}}+k_i}\,\overline{X}_i + \left[1-\frac{k_i}{\frac{1}{\hat{\gamma}}+k_i}\right]\mu. \tag{2.12}$$

2.1.2. μ is unknown.

Recall that the first stage Bayesian analysis under the loss (1.3) yields the estimate (2.7). When μ is unknown, to derive an HB estimator, it is first necessary to assign a prior distribution to the hyperparameter μ and to compute $E(\mu|X,\gamma)$. However, due to the complicated form of $m(X|\gamma,\mu)$, this expectation can be only calculated numerically for fixed X and γ. Thus the HB procedure would yield an estimator requiring the evaluation of two dimentional integrals. One alternative to the hierarchical Bayes approach is to estimate from data. Following the idea of ML-II prior (Berger, 1985), μ let $L(\mu) = \ln m(X|\gamma,\mu)$. Then,

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{p} \frac{1}{\gamma^2}\left[\gamma + \frac{\Gamma'(\frac{1}{\gamma})}{\Gamma(\frac{1}{\gamma})} - \gamma^2(k_i + \frac{1}{\gamma})\frac{1}{\mu + \gamma X_i}\right] \tag{2.13}$$

After computing the optimal $\tilde{\mu}$, then it follows from (2.8) that a combination of empirical Bayes and hierarchical Bayes estimator for θ^{-1} is given componentwisely as $\delta_i^{EH}(X) = \tilde{\gamma}_i \overline{X_i} + (1-\tilde{\gamma}_i)\tilde{\mu}$ where $\tilde{\gamma}_i = k_i E\left[(\frac{1}{\gamma} + k_i)^{-1}|X\right]$ and the above expectation is taken with respect to the posterior distribution of γ given by $\pi(\gamma|X) \propto \prod_{i=1}^{p} m(X_i|\gamma)\pi_2(\gamma)$ and $m(X_i|\gamma)$ is in (2.9) with μ replaced by $\tilde{\mu}$. An empirical Bayes estimator for θ^{-1} is derived by maximizing $m(X|\gamma) = \prod_{i=1}^{p} m(X_i|\gamma)$. After computing the optimal $\hat{\gamma}$, it follows from (2.8) that an EB estimator for θ^{-1} is given by $\delta^{EB}(X)$ defined componentwisely as

$$\delta_i^{EB}(X) = \frac{k_i}{\frac{1}{\hat{\gamma}} + k_i}\overline{X_i} + \left(1 - \frac{k_i}{\frac{1}{\hat{\gamma}} + k_i}\right)\tilde{\mu}.$$

2.2. Estimating θ under loss (1.2)

It will be shown useful in sections 2.2.1 and 2.2.2 to reparameterize (α, β) into (λ, η) where λ and η are given respectively by

$$\lambda = \frac{1}{\alpha - 1}, \quad \eta = \frac{1}{E(\theta_i^{-1}|\alpha,\beta)} = (\alpha-1)\beta. \tag{2.14}$$

assuming $\alpha > 1$. Remark that $\mu = \frac{1}{E(\theta_i^{-1}|\alpha,\beta)}$ is the best prior estimate of θ_i in the sense that it minimizes $E(L(\theta,\delta)|\alpha,\beta)$. It will be of interest to consider the case when η is known and unknown. If η is known, the hierarchical model is as follow:

$$\theta_i|\lambda, \quad \eta \sim \Gamma(\eta,\lambda), \quad \lambda \sim \pi_2(\lambda) \tag{2.15}$$

where the second stage prior π_2 is noninformative. On the other hand, if η is unknown, the hierarchical model is as above (2.16) with replacement

$$(\eta,\lambda) \sim \pi_2(\eta,\lambda) \tag{2.16}$$

where π_2 is now a noninformative prior for hyperparameters η and λ .

2.2.1. η is known.

By reparametrization in (2.15), since $\eta = (\alpha-1)\beta$, replace $\frac{1}{\beta}$ by $\frac{\alpha-1}{\eta}$. Then the posterior distribution of θ, denoted by $\pi(\theta|X, \alpha, \eta)$, is a product of p independent gamma distributions with $\alpha + k_i$ and $X_i + \frac{\alpha-1}{\eta}$. Under the assumed model and given loss in (1.2), the estimate at the first stage of Bayesian analysis is $\theta = (\theta_1, \cdots, \theta_p)$ given componentwisely as

$$\widehat{\theta}_i = \frac{1}{E(\theta_i^{-1}|X, \alpha, \eta)} = [\frac{k_i}{\alpha-1+k_i}\frac{X_i}{k_i} + \frac{\alpha-1}{\alpha-1+k_i}\frac{1}{\eta}]^{-1}. \tag{2.17}$$

Suppose a noninformative prior $\pi(\theta)$ for θ is propotion to constant. This prior can be obtained from $\Gamma(\alpha, \frac{1}{\beta})$ by letting $\alpha \to 1$, and $\beta \to \infty$. Also by these limits, $\alpha \to 1$ implies $\widehat{\theta}_i \to \frac{k_i}{X_i} = \delta^{\ni}(X_i)$. On the other hand, precise prior beliefs can be modelled by fixing η and letting $\alpha \to \infty$, and then $\widehat{\theta}_i \to \eta$ .It is desirable to consider estimating θ by inverse of a convex combination of $\delta^{\ni}$ and η. The Bayes estimate (2.18) is rewritten as

$$\widehat{\theta}_i = [\frac{k_i}{\frac{1}{\lambda}+k_i}\overline{X}_i + (1 - \frac{k_i}{\frac{1}{\lambda}+k_i})\frac{1}{\eta}]^{-1} \tag{2.18}$$

where $\overline{X}_i = \frac{X_i}{k_i}$ and $\lambda = \alpha - 1$. Note that marginally $X_1, \cdots, X_p$ are independent with X_i having the density

$$m(X_i|\lambda, \eta) = \int f(X_i|\theta_i)\pi(\theta_i|\lambda, \eta)d\theta_i = \frac{\Gamma(k_i + \frac{1}{\lambda}+1)X_i^{k_i-1}}{\Gamma(k_i)\Gamma(\frac{1}{\lambda}+1)(\lambda\eta)^{\frac{1}{\lambda}+1}(X_i + \frac{1}{\lambda\eta})^{k_i+\frac{1}{\lambda}+1}}. \tag{2.19}$$

Under the second stage prior with $\pi_2(\lambda)$,

$$\delta_i^{HB}(X) = \frac{1}{E(\theta_i^{-1}|X, \eta)} = \frac{1}{E_1E(\theta_i^{-1}|X, \eta, \lambda)} = \frac{1}{E_1\left[\frac{x_i+\lambda\eta}{k_i+\frac{1}{\lambda}}\right]} \tag{2.20}$$

where E_1 denotes the expectation taken over the posterior $\pi(\lambda|X, \eta)$ of λ given X and η, and $\pi(\lambda|X, \eta) \propto [\prod_{i=1}^{p} m(X_i|\lambda, \eta)]\pi_2(\lambda)$.

To derive EB estimator for $\hat{\theta}$, it is necessary to maximize $m(X|\gamma, \mu) = \prod_{i=1}^{p} m(X_i|\lambda, \eta)$ Let $L(\lambda) = \ln m(X|\lambda, \eta)$. Then it follows that

$$\frac{\partial L}{\partial \lambda} = -\frac{1}{\lambda^2}\sum_{i=1}^{p}\left[\frac{\Gamma'(k_i+\frac{1}{\lambda}+1)}{\Gamma(k_i+\frac{1}{\lambda}+1)}-\frac{\Gamma'(\frac{1}{\lambda}+1)}{\Gamma(\frac{1}{\lambda}+1)} - \log\lambda\,\eta\,(X_i+\frac{1}{\lambda\eta})+(1+\lambda)+(k_i+\frac{1}{\lambda}+1)\frac{\lambda}{1+\lambda\eta X_i}\right] \tag{2.21}$$

where $\Gamma'(\lambda)$ denotes $\int x^{\lambda}e^{-x}\log x\, dx$. After computing the optimal $\hat{\lambda}$,

$$\delta_i^{EB}(X)=\left[\frac{k_i}{\frac{1}{\hat{\gamma}}+k_i}\overline{X}_i+\left[1-\frac{k_i}{\frac{1}{\hat{\gamma}}+k_i}\right]\frac{1}{\eta}\right]^{-1}. \tag{2.22}$$

2.2.2. η is unknown.

Recall that the first stage Bayesian analysis under the loss (1.2) yields the estimate (2.18). When η is unknown, to derive an HB estimator, it is first necessary to assign a prior distribution to the hyperparameter η and to compute $E(\eta\,|X,\lambda)$. However, due to the complicated form of $m(X|\lambda,\eta)$, this expectation can be only calculated numerically for fixed X and λ. Thus the HB procedure would yield an estimator requiring the evaluation of two dimentional integrals. One alternative to the hierarchical Bayes approach is to estimate η from data. Following the idea of ML-II prior (Berger, 1985), let $L(\eta) = \ln m(X|\lambda,\eta)$. Then,

$$\frac{\partial L}{\partial \eta} = \sum_{i=1}^{p}\left[(k_i+\frac{1}{\lambda}+1)\frac{1}{\lambda+\lambda^2\eta X_i}-(\frac{1}{\lambda}+1)\frac{1}{\eta}\right].$$

After computing the optimal $\tilde{\eta}$, it then follows from (2.19) that a combination of empirical Bayes and hierarchical Bayes estimator for θ is given componentwisely as $\delta_i^{EH}(X)=E\left[\frac{\frac{1}{\lambda}+k_i}{X_i+\tilde{\eta}\lambda}\right]$ and the above expectation is taken with respect to the posterior distribution of γ given by $\pi(\lambda|X)\propto\prod_{i=1}^{p}m(X_i|\lambda)\,\pi_2(\lambda)$ and $m(X_i|\lambda)$ is in (2.20) with η replaced by $\tilde{\eta}$. An empirical Bayes estimator for θ is derived by maximizing $m(X|\lambda)=\prod_{i=1}^{p}m(X_i|\lambda)$. After computing the optimal $\tilde{\lambda}$, it follows from (2.19) that an EB estimator for θ is given by $\delta^{EB}(X)$ defined componentwisely as

$$\delta_i^{EH}(X)=\frac{\frac{1}{\tilde{\lambda}}+k_i}{X_i+\tilde{\eta}\tilde{\lambda}}=\left[\frac{k_i}{\frac{1}{\tilde{\lambda}}+k_i}\overline{X}_i+(1-\frac{k_i}{\frac{1}{\tilde{\lambda}}+k_i})\frac{1}{\tilde{\eta}}\right]^{-1}.$$

3. Comparison of HB and EB.

In this section, we deal with the comparison of HB and EB when μ and η are known.

3.1. The loss (1.3)

The ith component shrinkage of $\delta^{HB}(X)$ and $\delta^{EB}(X)$ away from the MVUE $\delta_i^0(X) = \overline{X}_i$ toward the best prior estimate μ are respectively given by

$$\frac{\delta_i^{HB}(X) - \overline{X}_i}{\mu - \overline{X}_i} \quad \text{and} \quad \frac{\delta_i^{EB}(X) - \overline{X}_i}{\mu - \overline{X}_i} \quad \text{for} \quad i = 1, \cdots, p. \tag{3.1}$$

Assume that the noninformative prior is $\pi_2(\gamma) = 1$ for $0 \le \gamma \le 1$. When computing

$\delta_i^{HB}(X) = \dfrac{1}{E(\theta_i|X, \mu)} = \left[E_1\left[\dfrac{\frac{1}{\gamma} + k_i}{X_i + \frac{1}{\gamma}\mu} \right] \right]^{-1}$ in (2.10) where E_1 denotes the expectation taken over the posterior $\pi(\gamma|X,\mu)$ of γ given X and μ, and $\pi(\gamma|X, \mu) \propto [\prod_{i=1}^{p} m(X_i|\gamma, \mu)]\pi_2(\gamma)$, it is impossible to find its analytic form. So, we use the approxmation methods, such as Laplace approximation and Gibbs sampler(Gelfand and Smith, 1990) etc. But in Laplace approximation, the mode of γ in $[\prod_{i=1}^{p} m(X_i|\gamma, \mu)]\pi_2(\gamma)$ is needed and its method is very complicated. Therefore we use Gibbs sampler to compute

$\delta_i^{HB}(X) = \left[E_1\left[\dfrac{\frac{1}{\gamma} + k_i}{X_i + \frac{1}{\gamma}\mu} \right] \right]^{-1}$. In this case, we need the univariate full conditional distribution(FCD) as follows:

For $i = 1, \cdots, p$,

$$[\theta_i|\gamma, x, \mu] \propto \theta_i^{k_i + \frac{1}{\gamma} - 1} \exp -(X_i + \frac{\mu}{\gamma})\theta_i \tag{3.2}$$

and

$$[\gamma|\theta, x, \mu] \propto \frac{(\frac{\mu}{\gamma})^{\frac{p}{\gamma}}}{\Gamma(\frac{1}{\gamma})^p} \prod_{i=1}^{p} \theta_i^{\frac{1}{\gamma} + k_i - 1} \exp((-\frac{\mu}{\gamma} + X_i)\theta_i). \tag{3.3}$$

When sampling from each FCD, since $[\theta_i|\gamma, x, \mu]$ is gamma distribution, its sampling is easy. But the exact form of $[\gamma|\theta, x, \mu]$ is very complicated and so we use Metropolis algorithm to sample from $[\gamma|\theta, x, \mu]$. In order to compute $\delta^{EB}(X)$, the optimal $\hat{\gamma}$ in the equation (2.11) is needed and so the Newton-Raphson method is used.

Table 1. Shrinkage of HB and EB for $k=3$ under (1.3)

$\overline{X}$			d_1	δ^{EB}	δ^{HB}
0.01	0.02	0.05	3.84	0.55330	0.40153
0.1	0.5	1.5	1.13	0.00903	0.39927
0.3	0.6	1.6	0.67	0.00420	0.40076
0.6	0.8	1.2	0.31	0.00062	0.40269
200	200	201	5.30	0.41131	0.39173
10	10	10.5	2.32	0.45274	0.39173
5	5	5.5	1.64	0.34755	0.39713
0.5	2	5	0.69	0.00860	0.39173
0.8	1.2	1.3	0.22	0.00800	0.39863

Table 2. Shrinkage of HB and EB for $k=10$ under (1.3)

$\overline{X}$			d_1	δ^{EB}	δ^{HB}
0.1	0.2	1.5	1.44	0.41390	0.16192
0.1	0.5	1.5	1.13	0.36515	0.16192
0.4	0.6	0.2	0.71	0.00548	0.16192
0.6	0.8	1.5	0.38	0.00771	0.16192
0.6	0.8	1.2	0.31	0.00575	0.16192
0.8	0.8	0.9	0.18	0.00630	0.16192
5	5	5.5	1.64	0.23200	0.16192
0.5	1.5	6	0.96	0.44700	0.16192
0.5	2	2	0.69	0.00988	0.16192
1.2	1.5	2	0.43	0.00503	0.16540
0.8	1.2	1.3	0.22	0.00870	0.16192

In tables 1 and 2, the simultaneous estimation of θ^{-1} is considered for $p=3$ and $k_1=k_2=k_3=k$. These tables 1 and 2 show the shrinkage of HB and EB estimators towards $\mu=1$ based on sample means of dimension $p=3$ for $k=3$ and 10 and the loss (1.3). The measure $d_1=\frac{1}{p}\sum_{i=1}^{p}|\log\overline{X}_i-\log\mu|$ is used to measure an average distance from $\overline{X}=(\overline{X}_1,\cdots,\overline{X}_p)$ to μ. For each n, the table is divided into two categories. An observation X is in the first categorywhen $\frac{1}{p}\sum_{i=1}^{p}\log\overline{X}_i<\log\mu$ and X is in the second category when $\frac{1}{p}\sum_{i=1}^{p}\log\overline{X_i} > \log\mu$. It appears that the shrinkage of $\delta^{EB}(X)$ approaches 0 faster than that of $\delta^{HB}(X)$ as d_1 becomes small. These observations suggest the preference of $\delta^{HB}(X)$ over $\delta^{EB}(X)$.

3.2. The loss(1.2)

The ith component shrinkage of $\delta^{HB}(X)$ and $\delta^{EB}(X)$ away from the MVUE $\delta_i^+(X)=\frac{k_i-1}{X_i}$ toward the best prior estimate η are respectively given by

$$\frac{\delta_i^{HB}(X)-\frac{k_i-1}{X_i}}{\eta-\frac{k_i-1}{X_i}} \quad\text{and}\quad \frac{\delta_i^{EB}(X)-\frac{k_i-1}{X_i}}{\eta-\frac{k_i-1}{X_i}} \quad \text{for } i=1,\cdots,p. \tag{3.4}$$

Assume that the noninformative prior is $\pi_{2(\gamma)}=1$ for $0\le\gamma\le 1$. When computing $\delta_i^{HB}(X)=\frac{1}{E(\theta_i^{-1}|X,\eta)}=\frac{1}{E_1\left[\frac{X_i+\lambda\eta}{k_i+\frac{1}{\lambda}}\right]}$ in (2.21) where E_1 denotes the expectation taken over the posterior $\pi(\lambda|X,\eta)$ of λ given X and η, and $\pi(\lambda|X,\eta)\propto[\prod_{i=1}^{p}m(X_i|\lambda,\eta)]\,\pi_2(\lambda)$, it is impossible to find its analytic form. Also we use Gibbs sampler to compute $\delta_i^{HB}(X)=\frac{1}{E_1\left[\frac{X_i+\lambda\eta}{k_i+\frac{1}{\lambda}}\right]}$. In this case, we need the univariate full conditional distribution(FCD) as follows:

For $i=1,\cdots,p$,

$$[\theta_i|\lambda,X,\eta]\propto\theta_i^{k_i+\frac{1}{\lambda}}\exp(-(X_i+\frac{1}{\lambda\eta})\,\theta_i) \tag{3.5}$$

and

$$[\lambda|\theta_1,\cdots,\theta_p,X,\eta]\propto\left(\frac{1}{(\lambda\eta)^{\frac{1}{\lambda}+1}\Gamma(\frac{1}{\lambda}+1)}\right)^p\prod_{i=1}^{p}[\theta_i^{\frac{1}{\lambda}}\exp(-(X_i+\frac{1}{\lambda\eta})\theta_i)]. \tag{3.6}$$

When sampling from each FCD, since $[\theta_i|\lambda,x,\eta]$ is gamma distribution, its sampling is easy. But the exact form of $[\lambda|\theta,x,\eta]$ is very complicated and so we use Metropolis algorithm to sample from $[\lambda|\theta,x,\eta]$. Then the estimate of $\delta_i^{HB}(X)$ can be expressed as

$\frac{1}{\frac{1}{N}\sum_{j=1}^{N}\frac{X_i+\lambda_j\eta}{k_i+\frac{1}{\lambda_j}}}$. Next to compute $\delta^{EB}(X)$, we use the Newton-Raphson method to find the optimal $\hat{\gamma}$ in the equation (2.22).

In tables 3 and 4, the simultaneous estimation of θ^{-1} is considered for $p=3$ and $k_1 = k_2 = k_3 = k$. These tables indicate the shrinkage of HB and EB estimators towards $\eta = 10$ based on sample means of dimension $p = 3$ for $k = 3$ and 10 and the loss (1.2). The measure $d_2 = \frac{1}{p}\sum_{i=1}^{p}\left|\log\frac{k-1}{X_i} - \log\eta\right|$ is used to measurean average distance from UMVUE $\frac{k-1}{X}$ to η. In this case, we only consider the category when $\frac{1}{p}\sum_{i=1}^{p}\log\frac{k-1}{X_i} < \log\eta$. It appears that the shrinkage of $\delta^{EB}(X)$ approaches 0 faster than that of $\delta^{HB}(X)$ as d_2 becomes large. These observations suggest the preference of $\delta^{HB}(X)$ over $\delta^{EB}(X)$.

Table 3. Shrinkage of HB and EB for $k=3$ under (1.2)

$\overline{X}$			d_2	δ^{EB}	δ^{HB}
0.1	0.1	0.1	0.4	1	1
0.2	0.2	0.3	1.2	0.7690	0.7180
0.6	0.65	0.7	2.2	0.2200	0.3970
0.8	0.8	1	2.5	0.1460	0.3260
1	1	1.5	2.7	0.1100	0.2790
15	15	10.5	5.4	0.0060	0.0029
100	100	100	7.3	0.0001	0.0006
200	200	201	8.0	0.0001	0.0003

Table 4. Shrinkage of HB and EB for unde k = 10 under (1.2)

$\overline{X}$			d_2	δ^{EB}	δ^{HB}
0.1	0.1	0.1	0.10	1	0.9910
0.2	0.2	0.3	0.19	0.5090	0.3930
0.6	0.65	0.7	1.97	0.0620	0.1590
0.8	0.8	1	2.25	0.0430	0.1230
1	1	1.5	2.54	0.0330	0.1010
15	15	10.5	4.99	0.0020	0.0091
100	100	100	7.00	0.0003	0.0004
200	200	201	7.00	0.00001	0.0002

4. Computer Simulation.

Assume that all $k_i = 1$ in (1.1) and (1.2) and that μ is known. In this section, we compute the percentage risk improvement of the proposed estimators $\delta(X)$ with $\delta^0(X) = X$ under the loss (1.3). The percentage of savings in risk using $\delta(X)$ are compared to $\delta^0(X)$ using the formula

$$\frac{R(\theta,\delta^0(X)) - R(\theta,\delta(X))}{R(\theta,\delta^0(X))} \times 100.$$

The estimator $\delta^D(X)$ of Dey et al.(1987) is given componentwisely by

$$\delta_i^D(x) = X_i\,(1-\frac{r(s)}{b+S}\ln(X_i)) \tag{4.1}$$

where $S = \sum_{i=1}^{k} \ln X_i$ and $b = 1.45(p-2)^2$ and $r(s) = 3(p-2)/5$. Estimates of the parameters are calculated according to the estimators $\delta^{HB}(X)$ and $\delta^D(X)$ given by (2.10) and (4.1), respectively. The next step is repeated 3000 times and the risks as the average of each loss are calculated for the different values of μ in (2.5) and the different ranges of θ.

In table 5 and table 6, it is observed that the improvements are always positive and $\delta^{HB}(X)$ and $\delta^D(X)$ dominate $\delta^0(X)$ in terms of risk according to computer simulation. Also we observe that the percentage improvement decreases as the magnitudes of the θ_i increase. These tables indicate that the percentage risk improvement of $\delta^{HB}(X)$ is better than that of $\delta^D(X)$ for variou values of μ.

Table 5. Percentage Risk Improvements for Range of $\theta : (0,1]^p$

	$p=3$		$p=5$		$p=10$	
μ	δ^D	δ^{HB}	δ^D	δ^{HB}	δ^D	δ^{HB}
1	0.03	72	0.01	72	0.06	73
1.2	0.03	72	0.01	72	0.06	73
1.4	0.03	71	0.02	72	0.05	74
1.6	0.02	72	0.01	74	0.05	75
1.8	0.02	73	0.01	72	0.06	75
2.0	0.02	72	0.01	72	0.06	71

Table 6. Percentage Risk Improvements for Range of $\theta : (0,2]^p$

	$p=3$		$p=5$		$p=10$	
μ	δ^D	δ^{HB}	δ^D	δ^{HB}	δ^D	δ^{HB}
1	0.3	60.2	0.07	61	0.03	62
1.2	0.3	60.7	0.07	61	0.03	63
1.4	0.3	60.4	0.08	62	0.04	62
1.6	0.3	60.7	0.05	62	0.04	58
1.8	0.3	60.5	0.05	62	0.04	65
2.0	0.3	60.5	0.07	62	0.03	58

REFERENCES

[1] Berger, J.O.(1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters, *Annals of Statistics,* 8, 545-571.

[2] Berger, J.O.(1985). *Statistical decision theory and Bayesian analysis*(2nd ed.) Springer - Verlag, New York.

[3] Brown, L.D.(1966). On the admissibility of invariant estimators of one or more location parameters. *Annals of Mathematical Statistics,* 37, 1087-1136.

[4] Chung, Y. and Dey, D.K.(1994). Simutaneous estimation of parameters from power series distribution under asymmetric loss. *Journal of Korean Statistical Society,* 23, 151-166.

[5] Chung, Y., Kim, C. and Dey, D.K.(1994). Simultaneous estimation of Poisson mean under weighted entropy loss. *Calcutta Statistical Association Bulletin,* 44, 165-174.

[6] DasGupta, A. (1986). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic loss. *Annals of Statistics,* 14, 206-219.

[7] Dey, D.K. and Chung, Y. (1991). Multiparameter estimation of truncated power series distribution under Stein's loss. *Communications in Statistics*, 20, 309-326.

[8] Dey, D.K., Ghosh, M. and Srinivasan, C.(1987). Simultaneous estimation of parameters under entropy loss. *Journal of Statistical Planning and Inference*, 15, 347-363.

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

[10] Ghosh, M., Hwang, J. and Tsui, K.W.(1984). Construction of improved estimators in multiparameter estimation for continuous exponential families, *Journal of Multivariate Analysis*, 14, 212-220.

[11] Hudson, H.M.(1978). A natural identity for exponential families with applications in multiparameter estimation, *Annals of Statistics*, 6, 473-84.

[12] James, W. and Stein, C.(1961). Estimation with quadratic loss. *Proceeding of Fourth Berkeley Symposium: Mathematical Statistics and Probability,* Vol.1 (1961), 361-379. University of California Press.