Communications for Statistical Applications and Methods
2013, Vol. 20, No. 4, 321–327
DOI: http://dx.doi.org/10.5351/CSAM.2013.20.4.321

Constrained Bayes and Empirical Bayes Estimator Applications in Insurance Pricing

Myung Joon Kim[a], Yeong-Hwa Kim[1,b]

[a]Department of Business Statistics, Hannam University
[b]Department of Applied Statistics, Chung-Ang University

Abstract

Bayesian and empirical Bayesian methods have become quite popular in the theory and practice of statistics. However, the objective is to often produce an ensemble of parameter estimates as well as to produce the histogram of the estimates. For example, in insurance pricing, the accurate point estimates of risk for each group is necessary and also proper dispersion estimation should be considered. Well-known Bayes estimates (which is the posterior means under quadratic loss) are underdispersed as an estimate of the histogram of parameters. The adjustment of Bayes estimates to correct this problem is known as constrained Bayes estimators, which are matching the first two empirical moments. In this paper, we propose a way to apply the constrained Bayes estimators in insurance pricing, which is required to estimate accurately both location and dispersion. Also, the benefit of the constrained Bayes estimates will be discussed by analyzing real insurance accident data.

Keywords: Bayes estimate, constrained Bayes estimate, insurance pricing.

1. Introduction

Bayesian techniques are widely used for the simultaneous estimation of several parameters in compound decision problems. Under the quadratic loss, Bayes estimates are the posterior means of the parameters of interest which is minimizing the squared error loss focusing on the precision of estimation. However, the goal is to often produce an ensemble of parameter estimates as well as ensure the histogram of estimates somewhat close to the histogram of the population parameters. The twin objectives mentioned the above are usually conflicting in the sense that one is often achieved at the expenses of the other.

The posterior means of the parameters are optimal estimates under quadratic loss, but it can be shown that the histogram of the posterior means of parameters is underdispersed as an estimate of the histogram of parameters. Accordingly, the histogram of posterior means is clearly inappropriate to estimate the parameter histogram. There is a need to find a set of suboptimal estimates to compromise between the two criteria. Louis (1984) proposed a constrained Bayes method that matches the first two empirical moments of Bayes estimates of the normal means with minimizing the squared distance of the parameters and estimates subject to these constraints. After that Ghosh (1992) generalized the Louis' result for any arbitrary distribution and not necessarily normal distribution.

This paper was supported by the 2013 Hannam University Research Fund.
[1] Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, 221 Heuksuk-dong, Dongjak-gu, Seoul 156-756, Korea. E-mail: gogators@cau.ac.kr

Consider a situation where the data are denoted by $\boldsymbol{x}$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ is the parameter of interest. The basic idea is to find an estimate of parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ that minimizes

$$E\left\{\sum_{i=1}^{m}(\theta_i - t_i)^2 \middle| \boldsymbol{x}\right\}, \tag{1.1}$$

within the class of all estimates $\boldsymbol{t}(\boldsymbol{x}) = \boldsymbol{t} = (t_1, \dots, t_m)^T$ of $\boldsymbol{\theta}$ that satisfy these two conditions

$$(1)\ \ E\left(\bar{\theta}|\boldsymbol{x}\right) = m^{-1}\sum_{i=1}^{m} t_i(\boldsymbol{x}) = \bar{t}(\boldsymbol{x}), \tag{1.2}$$

$$(2)\ \ E\left\{\sum_{i=1}^{m}(\theta_i - \bar{\theta})^2 \middle| \boldsymbol{x}\right\} = \sum_{i=1}^{m}\{t_i(\boldsymbol{x}) - \bar{t}(\boldsymbol{x})\}^2. \tag{1.3}$$

Let $\boldsymbol{\theta}^B$ denote Bayes estimate of $\boldsymbol{\theta}$ under any quadratic loss, it is easy to to see that Bayes estimate satisfies (1.2); however, regarding to the (1.3) can be calculated as:

$$\sum_{i=1}^{m}\left\{\theta_i^B(\boldsymbol{X}) - \bar{\theta}^B(\boldsymbol{X})\right\}^2 < E\left\{\sum_{i=1}^{m}\left(\theta_i - \bar{\theta}\right)^2 \middle| \boldsymbol{X}\right\}. \tag{1.4}$$

This clearly shows the limitation of usual Bayes estimates to estimate the true variation of parameter $\boldsymbol{\theta}$ and the reason why constrained Bayes estimates should be considered in the case of the twin objectives.

In Section 2, the constrained Bayes and constrained empirical Bayes estimates are described in detail. The insurance pricing application with described estimates is discussed in Section 3. After that the benefit of the those estimates are derived with numerical results in Section 4 using real insurance data.

2. Constrained and Constrained Empirical Bayes Estimate

2.1. Constrained Bayes estimate

Constrained Bayes estimate(CB) of parameter θ can be found which minimizes (1.1) subject to (1.2) and (1.3) and the main result is proved in Ghosh (1992). For stating the theorem require a few notions and the expected value of second moment can be denoted as following. To see this we define $\boldsymbol{I}_m$ as the identity matrix of order m, $\mathbf{1}_m$ as the m-component column vector with each element equal to 1 and $\boldsymbol{J}_m = \mathbf{1}_m\mathbf{1}_m^T$.

$$\begin{aligned}
E\left\{\sum_{i=1}^{m}\left(\theta_i - \bar{\theta}\right)^2 \middle| \boldsymbol{X}\right\} &= \operatorname{tr}\left\{\left(\boldsymbol{I}_m - \frac{1}{m}\boldsymbol{J}_m\right)E\left(\boldsymbol{\theta}\boldsymbol{\theta}^T \middle| \boldsymbol{X}\right)\right\} \\
&= \operatorname{tr}\left\{\left(\boldsymbol{I}_m - \frac{1}{m}\boldsymbol{J}_m\right)\left\{V(\boldsymbol{\theta}|\boldsymbol{X}) + E(\boldsymbol{\theta}|\boldsymbol{X})E(\boldsymbol{\theta}|\boldsymbol{X})^T\right\}\right\} \\
&= \operatorname{tr}\left\{\left(\boldsymbol{I}_m - \frac{1}{m}\boldsymbol{J}_m\right)V(\boldsymbol{\theta}|\boldsymbol{X})\right\} + \sum_{i=1}^{m}\left\{\theta_i^B(\boldsymbol{X}) - \bar{\theta}^B(\boldsymbol{X})\right\}^2 \\
&= \operatorname{tr}\left\{V\left(\boldsymbol{\theta} - \bar{\theta}\mathbf{1}_m|\boldsymbol{X}\right)\right\} + \sum_{i=1}^{m}\left\{\theta_i^B(\boldsymbol{X}) - \bar{\theta}^B(\boldsymbol{X})\right\}^2.
\end{aligned} \tag{2.1}$$

The second term of (2.1) is the squared distance of the Bayes estimate and clearly shows the under-dispersion of the Bayes estimate. Let the first term of (2.1) be $H_1(\boldsymbol{x})$, and the second term be $H_2(\boldsymbol{x})$, then the CB estimate $\boldsymbol{\theta}^{CB}(\boldsymbol{x}) = (\theta_1^{CB}(\boldsymbol{x}), \ldots, \theta_m^{CB}(\boldsymbol{x}))^T$ can be stated as follows.

$$\theta_i^{CB}(\boldsymbol{x}) = a\theta_i^B(\boldsymbol{x}) + (1-a)\bar{\theta}^B(\boldsymbol{x}), \quad i = 1, \ldots, m, \ a \equiv a(\boldsymbol{x}) = \left[1 + \frac{H_1(\boldsymbol{x})}{H_2(\boldsymbol{x})}\right]^{\frac{1}{2}}. \tag{2.2}$$

The Bayes estimate for the exponential family can be derived as follows. Suppose that $X_1, \ldots, X_m$ are m independent random variables, where X_i has the pdf given by

$$f_{\phi_i}(x_i) = \exp\{n\phi_i x_i - n\psi(\phi_i)\}, \quad i = 1, \ldots, m.$$

Each X_i can be viewed as each having a pdf belonging to a one-parameter exponential family. Assume the independent conjugate priors

$$g(\phi_i) = \exp(\nu\phi_i\mu - \nu\psi(\phi_i)),$$

for the ϕ_i's. Then under the quadratic loss, the Bayes estimates of θ_i's are given by

$$\theta_i^B(\boldsymbol{x}) = E(\theta_i|\boldsymbol{x}) = E\left[\psi'(\phi_i)|\boldsymbol{x}\right] = (1-B)x_i + B\mu, \tag{2.3}$$

where $B = \nu/(n+\nu)$.

Equation (2.2) appearance looks like convex combinations of the Bayes estimates and their average, but it is deceptive expressing because function $a(\boldsymbol{x})$ exceeds 1. Ghosh and Kim (2002) have extended the result when the parameters are vector-valued and can be stated similarly.

2.2. Constrained empirical Bayes estimate

When the Bayesian model is true, the Bayes estimators have the smallest Bayes risks. The CB estimators cannot claim any risk improvement over the Bayes estimators. The same phenomenon is reflected in the comparison of the empirical Bayes(EB) and the constrained empirical Bayes(CEB) estimators. The CEB estimators are not designed to improve on the EB estimators by producing smaller MSE's. They are constructed to meet the twin objectives mentioned earlier more satisfactorily than the EB estimators.

Louis (1984) proposed the CEB estimators in the original James-Stein framework. Ghosh (1992) developed the CB estimators in a more general framework when the distribution was not necessarily normal. In this section, for more clear understanding, we will discuss under the assumption of normal distribution. Consider the usual normal model where $X_i|\theta_i$ are independent $N(\theta_i, 1), i = 1, \ldots, m$ while θ_i are iid $N(\mu, A)$. Then the CB estimator of $\boldsymbol{\theta}$ can be expressed as following by Equation (2.2).

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{CB} &= a_B[(1-B)\boldsymbol{X} + B\mu\boldsymbol{1}_m] + (1-a_B)\left[(1-B)\bar{X} + B\mu\right]\boldsymbol{1}_m \\ &= (1-B)\left[a_B\boldsymbol{X} + (1-a_B)\bar{X}\boldsymbol{1}_m\right] + B\mu\boldsymbol{1}_m, \end{aligned}\tag{2.4}$$

where $\boldsymbol{1}_m$ is an m-component column vector with each element equal to 1, $\bar{X} = m^{-1}\sum_{i=1}^m X_i$, $a_B^2 = 1 + 1/\{(1-B)S\}$, and $S = \sum_{i=1}^m (X_i - \bar{X})^2/(m-1)$.

In an EB scenario, typically both μ and A are unknown, and are estimated from the marginal distribution of $\boldsymbol{X} = (X_1, \ldots, X_m)^T$. Marginally, $\boldsymbol{X} \sim N(\mu\boldsymbol{1}_m, B^{-1}\boldsymbol{I}_m)$. Hence, marginally, $S \sim$

$B^{-1}\chi^2_{m-1}/(m-1)$. We estimate μ by $\bar{X}$, and B by $\hat{B} = 1/S$. The CEB estimator of $\boldsymbol{\theta}$ is then given by

$$\hat{\boldsymbol{\theta}}^{CEB} = \left(1-\hat{B}\right)\left[a_{EB}\boldsymbol{X} + (1-a_{EB})\bar{X}\mathbf{1}_m\right] + \hat{B}\bar{X}\mathbf{1}_m, \tag{2.5}$$

where a_{EB} replaces B by $\hat{B}$ in a_B.

2.3. Estimation with ANOVA model

Insurance pricing focuses on the estimation for the risk of each group. The key idea to segment the group is similar to the ANOVA model. By trying to minimize the within variance for each group and also to maximize the between variance, each group risk could be evaluated properly. Higher risk characteristics are segmentized with same group and lower risk characteristics are categorized with same group for better segmentation in insurance pricing. Therefore, we will introduce the CB estimators and the CEB estimators in the ANOVA model, which results expanded under different loss functions by Ghosh *et al.* (2008). In the next section, this model assumption and numerical analysis will be implemented with auto insurance accident data to estimate the each group risk.

Consider the ANOVA model with $Y_{ij} = \theta_i + e_{ij}$ and $\theta_i = \mu + \alpha_i$ $(i = 1, \ldots, m;\ j = 1, \ldots, k)$. Here the α_i and the e_{ij} are mutually independent with $\alpha_i \overset{iid}{\sim} N(0, \tau^2)$ and $e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$. Alternatively, in a Bayesian framework, these amounts to saying that $Y_{ij}|\theta_i \overset{iid}{\sim} N(\theta_i, \sigma^2)$, $i = 1, \ldots, m$ and $\theta_i \overset{iid}{\sim} N(\mu, \tau^2)$.

Minimal sufficiency consideration allows us to restrict to $(X_1, \ldots, X_m, \text{SSW})$, where $X_i = 1/k \sum_{j=1}^{k} Y_{ij} = \bar{Y}_i$ and $\text{SSW} = \sum_{i=1}^{m}\sum_{j=1}^{k}(Y_{ij} - \bar{Y}_i)^2$. We may note that marginally $X_1, \ldots, X_m$ and SSW are mutually independent with $X_i \overset{iid}{\sim} N(\mu, \tau^2 + \sigma^2/k)$, *i.e.*, $N(\mu, \sigma^2/(kB))$, where $B = (\sigma^2/k)/(\sigma^2/k + \tau^2) = \sigma^2/(\sigma^2 + k\tau^2)$ and $\text{SSW} \sim \sigma^2\chi^2_{m(k-1)}$.

From the results of the previous section, the CB estimator of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$ is given by

$$\hat{\boldsymbol{\theta}}^{CB} = a(\boldsymbol{X})(1-B)\left(\boldsymbol{X} - \bar{X}\mathbf{1}_m\right) + \left\{(1-B)\bar{X} + B\mu\right\}\mathbf{1}_m, \tag{2.6}$$

where $\boldsymbol{X} = (X_1, \ldots, X_m)^T$, $\bar{X} = m^{-1}\sum_{i=1}^{m} X_i$, and $\mathbf{1}_m$ is an m-component column vector with each element equal to 1. Also $a^2(X) = 1 + H_1(X)/H_2(X)$, where $H_1(X) = (m-1)(1-B)\sigma^2/k$ and $H_2(X) = (1-B)^2\sum_{i=1}^{m}(X_i - \bar{X})^2 = (1-B)^2(\text{SSB}/k)$, (say). Then, on simplification,

$$a^2(X) = 1 + \frac{\sigma^2}{(1-B)\text{MSB}}, \tag{2.7}$$

where $\text{MSB} = \text{SSB}/(m-1)$ and $\text{MSW} = \text{SSW}/(k-1)$.

At the result of Equation (2.7), by substitution of μ by $\bar{X}$, B by $\hat{B}$ and σ^2 by MSW. Here $\hat{B} = \{(m-3)\text{MSW}\}/\{(m-1)\text{MSB}\}$, the CEB estimator of $\boldsymbol{\theta}$ can be derived as

$$\hat{\boldsymbol{\theta}}^{CEB} = a_{EB}(\boldsymbol{X})\left(1-\hat{B}\right)\left(\boldsymbol{X} - \bar{X}\mathbf{1}_m\right) + \bar{X}\mathbf{1}_m, \tag{2.8}$$

where $a_{EB}(\boldsymbol{X}) = 1 + \text{MSW}/\{(1-\hat{B})\text{MSB}\}$.

Based on the previous result, the numerical analysis implementation will be performed in the next section to show the advantage of the CB estimators in the real business area.

Table 1: Data structure (n denotes the sample size of each group and n = 1,000)

N	Group			
	Small size car	Medium size car	Large size car	SUV & Van type
1	y_{11}	y_{21}	y_{31}	y_{41}
2	y_{12}	y_{22}	y_{32}	y_{42}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{1n}	y_{2n}	y_{3n}	y_{4n}

3. Data Analysis and Result

Insurance companies estimate premiums based on the paid loss amount caused by accidents using various customer information; therefore, pricing is the historical trial for converging to the true risk. For the proper and accurate estimation of the true risk, there exist three main issues. The first one is what kind of customer information should be used for the explanatory factor to estimate the risk (known as the rating factor). Secondly, how well segmented and assign the rating factor as groups for same group customers with similar risk characteristics. The last issue is how to estimate risk of each group precisely for corresponding rating factors. In this paper, the first two issues are not an interesting topic; however, there is a focus on the last issue that provides an accurate estimation of the risk.

3.1. Data descriptions

Auto insurance companies use approximately thirty to forty rating factors to differentiate each customer premium and apply the different segmentation group for each rating factor. In this section, the one main and basic factor in auto insurance, car type-small size car, medium size car and other factors will be considered among several factors. The data used for the analysis is real accident results in regards to property damage liability coverage from one property and casualty company in Korea. For each group, one thousand samples are randomly selected and Table 1 illustrates the data structures. In Table 1, y_{ij} indicates the loss amount during the policy period by accident for each policy and for the normality assumption log transformation is implemented.

From the data structure described the above, by assuming the parameter empirically, numerical analysis regarding the ANOVA model explained in the previous section will be implemented. Our suggesting estimator, constrained and empirical Bayes estimator and its second moment will be derived and the well known and widely used estimator Bayes estimator and its second moment also will be calculated for the comparison.

3.2. Analysis result and comparison

To indicate and evaluate the advantage of the estimators, we need to develop and define the index with some equations that reflect the research purpose. In this paper, the purpose of estimation has twin objectives by matching the first and second moments to produce histogram estimates close to the histogram of population parameters. Therefore, the idea for the developed index derives from the these objectives and the equation is:

$$\text{Index} = w\left[E\left(\bar{\theta}|\boldsymbol{x}\right) - m^{-1}\sum_{i=1}^{m} t_i(\boldsymbol{x})\right]^2 + (1-w)\left\|\left[E\left[\sum_{i=1}^{m}\left(\theta_i - \bar{\theta}\right)^2 \middle|\boldsymbol{x}\right] - \sum_{i=1}^{m}\left[t_i(\boldsymbol{x}) - \bar{t}(\boldsymbol{x})\right]^2\right]\right\|, \tag{3.1}$$

where $w = 1/2$ to assign the equal weight for each moments, and $m = 4$ indicates the number of group.

Table 2: Numerical analysis result.

Estimates	Group				Index
	Small size car	Medium size car	Large size car	SUV & Van type	
θ	488,942	540,365	597,196	660,003	
$\hat{\theta}^B$	481,275	523,352	580,225	548,308	1.0180
$\hat{\theta}^{EB}$	483,791	523,788	577,585	547,430	1.0190
$\hat{\theta}^{CB}$	478,134	522,789	583,517	549,386	1.0167
$\hat{\theta}^{CEB}$	480,277	523,161	581,236	548,629	1.0176

The index reflects the weighted sum of distance both first moment and second moment. The reason why is that the second term considers not square distance but that distance of absolute value is equalizing the size of distance with the first term.

Table 2 provides the summary of the data analysis. The result shows the index of the CB estimator is closer than the Bayes estimator in regards to the distance of the two parameters; in addition, the CEB estimator is closer to the EB estimator. This proves the benefit of the constrained Bayes estimators, not just minimizing the MSE, but trying to simultaneously match the second moments. The popular Bayes estimates minimize the errors; however, they do not satisfy the properness of simultaneous estimation for the point and dispersion together. In the table the advantage of the CB estimators totally comes from the second term of the index that calculates the distance of the second moments. The difference of index between Bayes and constrained Bayes estimators in regards to loss amount aspect, Bayes estimators shrink toward a mean of about 28,400 won compared to the constrained Bayes estimators and empirical Bayes estimators that shrink to 30,700 won compared to the constrained empirical Bayes estimators. The amount is approximately a 5.0% and 5.4% deviation from the total loss mean.

The insured period of auto insurance is a one year base in Korea; therefore, a consistent pricing technique is required every year. An accurate risk point estimation and the dispersion of the each group should also be considered. With regard to this reason and business needs, applying the constrained Bayes estimates for the auto insurance pricing can be a meaningful recommendations and this research suggests alternative tools for the auto insurance business field.

4. Further Study

In this paper, we assume the loss function as a quadratic loss function which is widely used for the estimation. However, quadratic losses are geared solely towards the precision of estimation. Sometimes there is a need to provide a framework that considers the tradeoff between goodness of fit and precision of estimation. Balanced loss function, suggested by Zellner (1988, 1992), meets this need; subsequently, the analysis with different loss function should be further studied for better estimation.

References

Louis, T. A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes method, *Journal of the American Statistical Association*, **79**, 393–398.

Ghosh, M. (1992). Constrained Bayes estimation with applications, *Journal of the American Statistical Association*, **87**, 533–540.

Ghosh, M. and Kim, D. (2002). Multivariate constrained Bayes estimation, *Pakistan Journal of Statistics*, **18**, 143–148.

Ghosh, M., Kim, M. and Kim, D. (2008). Constrained Bayes and empirical Bayes estimation under random effects normal ANOVA model with balanced loss function, *Journal of Statistical*

Planning and Inference, **138**, 2017–2028.
Zellner, A. (1988). Bayesian analysis in econometrics, *Journal of Econometrics*, **37**, 27–50.
Zellner, A. (1992). Bayesian and non-Bayesian estimation using balanced loss functions, *Statistical Decision Theory and Related Topics V*, Springer-Verlag, New York, 377–390.

Received April 29, 2013; Revised May 28, 2013; Accepted June 21, 2013