AN IMPROVED BAYES EMPIRICAL BAYES ESTIMATOR

R. J. KARUNAMUNI and N. G. N. PRASAD

Consider an experiment yielding an observable random quantity X whose distribution F_θ depends on a parameter θ with θ being distributed according to some distribution G_0. We study the Bayesian estimation problem of θ under squared error loss function based on X, as well as some additional data available from other *similar* experiments according to an empirical Bayes structure. In a recent paper, Samaniego and Neath (1996) investigated the questions of whether, and when, this information can be exploited so as to provide a better estimate of θ in the *current* experiment. They constructed a *Bayes empirical Bayes* estimator that is superior to the original Bayes estimator, based only on the current observation X for sampling situations involving exponential families-conjugate prior pair. In this paper, we present an improved Bayes empirical Bayes estimator having a smaller Bayes risk than that of Samaniego and Neath's estimator. We further observe that our estimator is superior to the original Bayes estimator in more general situations than those of the exponential families-conjugate prior combination.

2000 Mathematics Subject Classification: 62C12.

1. Introduction. Suppose that an experiment yields an observable random variable X whose distribution is indexed by a parameter θ. Assume that θ is distributed according to some known (prior) distribution G_0. Consider the Bayesian estimation problem of θ based on X under squared error loss function. The researchers have investigated how to be a *better Bayesian* whenever some additional data are available from other *similar* experiments. This similarity of other experiments is described using an *empirical Bayes* structure (Robbins [8, 9]). More specifically, it is assumed that θ_i's are independent of

$$\theta_i \sim G, \quad i = 1,\ldots,k \tag{1.1}$$

and that, given θ_i, X_i is distributed according to F_{θ_i}, that is,

$$X_i \sim F_{\theta_i}, \quad i = 1,\ldots,k \tag{1.2}$$

(here and in what follows "$\sim$" means independent identically distributed (i.i.d.)). It is further assumed that the pairs $\{(X_i, \theta_i)\}$ are mutually independent and are

independent of (θ, X) as well. The above model is also known as the *compound sampling model*.

The Bayesian approach to inferences about θ depends on the assumed prior distribution G_0. This prior can depend on unknown parameters which in turn may follow some second-stage prior. This sequence of parameters and priors constitutes a hierarchical model. The hierarchy must stop at some point, with all remaining prior parameters assumed to be known. The Bayes estimators that are obtained with respect to such *hyperpriors* are known as *hierarchical Bayes estimators.* Alternatively, the basic empirical Bayes approach uses the observed data $(X_1, \ldots, X_k)$ from other similar experiments to estimate those final-stage parameters (or to estimate the Bayes rule itself) and proceeds as in a standard Bayesian analysis. The resulting estimators are generally known as *parametric (nonparametric) empirical Bayes estimators.* There is a huge litera-ture on empirical Bayes and hierarchical Bayes methods. An interested reader is referred to the monographs of Maritz and Lwin [7] and Carlin and Louis [2] for further details. On the other hand, the so-called *Bayes empirical Bayes* ap-proach (Deely and Lindley [4]) operates under the assumption that the prior G_0 (or the hyperprior) is completely known and seeks to combine subjective information about the unknown parameter (via G_0) with the available data $(X_1, \ldots, X_k, X)$ in the process of making inferences about the parameter of in-terest θ. For various developments on Bayes empirical Bayes methods, see the works of Rolph [10], Berry and Christensen [1], Deely and Lindley [4], Gilliland et al. [6], Walter and Hamedani [13, 14], Samaniego and Neath [12], and the references therein.

Like the Bayes empirical Bayes approach, hierarchical Bayes modeling is also a powerful tool for combining information from separate, but possibly related, experiments. The basic idea is to treat the unknown parameters from the indi-vidual experiments as realizations from a second-level distribution. The com-bined data can then be thought of as arising from a two-stage process: first, the parameters $\theta_1, \ldots, \theta_k$ are drawn from the second-level distribution, say $G_\eta(\theta)$, and then the data $X_1, \ldots, X_k$ are drawn from the resulting first-level distribu-tions $F_{\theta_1}, \ldots, F_{\theta_k}$. Under this formulation, the first-level model explains the vari-ation within experiments, and the second-level model explains the variation across experiments. For example, hierarchical Bayes methods form an ideal setting for combining information from several published studies of the same research area, a scientific discipline commonly referred to as *meta-analysis* (Cooper and Hedges [3]), though, in this context, primary interest is in the hyperparameters η rather than the parameters from individual studies, θ.

The appeal of the Bayes empirical Bayes or of the hierarchical Bayes mod-eling is that information from all the experiments can be used for inferences about the intermediate parameters $\theta_1, \ldots, \theta_k$ as well as θ. The probabilistic formulation is tailor-made for a fully Bayesian analysis. By putting prior prob-ability distributions on the third-level hyperparameters η and any nuisance

parameters, all inferences can be carried out with posterior probability statements. But, as noted by some authors, the danger lies in these approaches, however the misspecification of the prior and hyperprior distributions are. Most seriously, there is the possibility of false association. Perhaps the association of θ should not be combined with $\theta_1, \ldots, \theta_k$. See Efron [5] for more information on these kinds of criticisms and for an empirical Bayes-likelihood approach of data combining.

In a recent paper, Samaniego and Neath (hereafter S&N, [12]) presented an efficient method for exploiting the *past* data $(X_1, \ldots, X_k)$ along with the *current* observation X in the Bayesian estimation problem of θ, resulting in a Bayes empirical Bayes (BEB) estimator of θ. From a Bayesian perspective, they investigated under what circumstances their BEB estimator would offer improvement over the original Bayes estimator. They demonstrated that in the traditional empirical Bayes framework and in situations involving exponential families, conjugate priors, and squared error loss, their BEB estimator of θ (which is based on data $X_1, \ldots, X_k$ and X) is superior to the original Bayes estimator of θ, which is based only on X, showing a *better Bayesian* performance by combining some relevant empirical findings with the subjective information in the prior distribution G_0. The performance of BEB estimator of S&N depends on the choice of the prior G_0, and it is not known how good is their estimator from a frequentist perspective. Generally speaking, however, besides having good Bayesian properties, estimators derived using Bayesian methods such as BEB estimators can have excellent frequentist properties (such as a smaller frequentist risk) and produce improvements over estimators generated by frequentist- or likelihood-based approaches. See Carlin and Louis [2] for more elaborate discussion on this point.

In this paper, we present another potentially useful BEB estimator developed under the same setup of S&N. The proposed estimator of θ is shown to have better performance than that of S&N in the sense of having a smaller Bayes risk. We further observe that our estimator is superior to the original Bayes estimator for more general situations than those of the exponential families-conjugate prior combination, showing a wider applicability of the proposed estimator. It is not known whether our strategy is the optimal way of combining data from past similar experiments. Indeed, it is reasonable to expect that there may be other better methods of exploiting the data $(X_1, \ldots, X_k)$ from past experiments with the current experiment. The next section contains the main results of this article. Section 3 contains a numerical example comparing the proposed estimator with that of S&N estimator.

2. Bayes empirical Bayes estimator. For convenience of notation, we assume throughout this section that the *current* experiment is the $(k+1)$st-experiment, and thus the problem is Bayesian estimation of θ_{k+1} based on data $(\mathbf{X}_1, \ldots, \mathbf{X}_k)$ from k past experiments as well as the current data vector

$\mathbf{X}_{k+1} = (X_{k+1,1}, \ldots, X_{k+1,n_{k+1}})$, where $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,n_i})$, $i = 1, \ldots, k$, given θ_i,

$$X_{i,1}, \ldots, X_{i,n_i} \sim F_{\theta_i}, \quad i = 1, \ldots, k+1, \tag{2.1}$$

and that

$$\theta_i \sim G_0, \quad i = 1, \ldots, k+1. \tag{2.2}$$

It is further assumed that the pairs $\{(\mathbf{X}_i, \theta_i)\}_{i=1}^{k+1}$ are mutually independent. Let G be a prior distribution on θ_{k+1} such that the Bayes estimator $d_G(\mathbf{X}_{k+1})$ under squared error loss is given by

$$d_G(\mathbf{X}_{k+1}) = \alpha \widehat{\theta_{k+1}} + (1-\alpha)E_G(\theta), \tag{2.3}$$

where $\alpha \in [0,1)$ and $\widehat{\theta_{k+1}}$ denotes the uniformly minimum variance unbiased estimator (UMVUE) of θ_{k+1}. Based on data $(\mathbf{X}_1, \ldots, \mathbf{X}_k)$ from past experiments, let G_k denote the prior distribution with prior mean $c\widehat{\theta^*} + (1-c)E_G(\theta)$; that is, let G_k be the prior distribution on θ_{k+1} such that the Bayes estimator $d_{G_k}(\mathbf{X}_{k+1})$ of θ_{k+1} under squared error loss is

$$d_{G_k}(\mathbf{X}_{k+1}) = \alpha \widehat{\theta_{k+1}} + (1-\alpha)\left[c\widehat{\theta^*} + (1-c)E_G(\theta)\right], \tag{2.4}$$

where

$$\widehat{\theta^*} = \left(\sum_{i=1}^{k} n_i\right)^{-1} \sum_{i=1}^{k} n_i \widehat{\theta_i} \tag{2.5}$$

and for $i = 1, \ldots, k$, $\widehat{\theta_i}$ denotes the UMVUE of θ_i based on the observation vector $\mathbf{X}_i$. Then S&N showed that d_{G_k} has smaller Bayes risk than that of d_G given by (2.3) with respect to G_0 when estimating θ_{k+1}. That is,

$$r(G_0, d_{G_k}) < r(G_0, d_G) \tag{2.6}$$

for any value of the constant c satisfying

$$0 < c < \frac{2A^2}{A^2 + V(\widehat{\theta^*})}, \tag{2.7}$$

where $A = |E_{G_0}(\theta) - E_G(\theta)|$ and $r(G_0, d) = E(d - \theta_{k+1})^2$, with E denoting expectation with respect to all the random variables governed by (2.1) and (2.2). This notation of "E" is used in what follows without further mention.

For $k = 2$, (2.4) reduces to

$$d_{G_2}(\mathbf{X}_3) = \alpha\widehat{\theta}_3 + (1-\alpha)\left[c\left(\sum_{i=1}^{2}n_i\right)^{-1}\sum_{i=1}^{2}n_i\widehat{\theta}_i + (1-c)E_G(\theta)\right] \qquad (2.8)$$

as an estimator of θ_3.

We consider a class of estimators of θ_3 of the form

$$\delta_{G_2,\mathbf{w}}(\mathbf{X}_3) = \alpha\widehat{\theta}_3 + (1-\alpha)\left[w_2\widehat{\theta}_2 + w_1\widehat{\theta}_1 + (1-w_1-w_2)E_G(\theta)\right], \qquad (2.9)$$

where $\alpha \in [0,1)$, $0 \le w_i < 1$, $i = 1,2$ such that $0 \le w_1 + w_2 < 1$ and $\mathbf{w} = (w_1, w_2)$. Note that when $w_1 = cn_1/\sum_{i=1}^{2}n_i$ and $w_2 = cn_2/\sum_{i=1}^{2}n_i$, then $\delta_{G_2,\mathbf{w}}$ reduces to $d_{G_2}(\mathbf{X}_3)$, given by (2.8). Furthermore, when $w_1 = w_2 = 0$, then (2.9) reduces to $d_G(\mathbf{X}_3)$, given by (2.3) with $k = 2$. Thus, $d_{G_2}(\mathbf{X}_3)$ and $d_G(\mathbf{X}_3)$ are special cases of $\delta_{G_2,\mathbf{w}}(\mathbf{X}_3)$. We now find values $\mathbf{w} = (w_1, w_2)$ that minimize the Bayes risk of $\delta_{G_2,\mathbf{w}}(\mathbf{X}_3)$ with respect to G_0, and compare the respective Bayes risks of (2.3), (2.8), and (2.9).

THEOREM 2.1. *Let $r(G_0, \delta_{G_2,\mathbf{w}})$ denote the Bayes risk of $\delta_{G_2,\mathbf{w}}(\mathbf{X}_3)$, given by (2.9) with respect to G_0, that is, $r(G_0, \delta_{G_2,\mathbf{w}}) = E[\delta_{G_2,\mathbf{w}}(\mathbf{X}_3) - \theta_3]^2$. Then, the values of $\mathbf{w} = (w_1, w_2)$ that minimize $r(G_0, \delta_{G_2,\mathbf{w}})$ are $\mathbf{w}^* = (w_1^*, w_2^*)$ such that*

$$
\begin{aligned}
w_1^* &= \frac{(\mu - \mu_0)^2}{[V(\widehat{\theta}_1) + (\mu - \mu_0)^2 + E_{G_0}(\theta^2) - \mu_0^2][1 + (a_1 + a_2)(\mu - \mu_0)^2]}, \\
w_2^* &= \frac{(\mu - \mu_0)^2}{[V(\widehat{\theta}_2) + (\mu - \mu_0)^2 + E_{G_0}(\theta^2) - \mu_0^2][1 + (a_1 + a_2)(\mu - \mu_0)^2]},
\end{aligned}
\qquad (2.10)
$$

where $\mu = E_G(\theta)$, $\mu_0 = E_{G_0}(\theta)$, and

$$
\begin{aligned}
a_1 &= [V(\widehat{\theta}_1) + (\mu - \mu_0)^2 + E_{G_0}(\theta^2) - \mu_0^2]^{-1}, \\
a_2 &= [V(\widehat{\theta}_2) + (\mu - \mu_0)^2 + E_{G_0}(\theta^2) - \mu_0^2]^{-1}.
\end{aligned}
\qquad (2.11)
$$

With $\mathbf{w} = \mathbf{w}^*$ in (2.9), it is now clear that the following inequality holds:

$$r(G_0, \delta_{G_2,\mathbf{w}^*}) \le \min\{r(G_0, d_{G_2}), r(G_0, d_G)\}, \qquad (2.12)$$

where $r(G_0, d_{G_2})$ and $r(G_0, d_G)$ denote Bayes risks of d_{G_2} and d_G, given by (2.8) and (2.3), respectively, with respect to G_0; that is, $r(G_0, d_{G_2}) = E[d_{G_2}(\mathbf{X}_3) - \theta_3]^2$ and $r(G_0, d_G) = E[d_G(\mathbf{X}_3) - \theta_3]^2$. Inequality (2.12) follows easily from the facts that d_{G_2} and d_G are special cases of $\delta_{G_2,\mathbf{w}}$, and $\delta_{G_2,\mathbf{w}^*}$ attains the minimum

Bayes risk among all estimators of θ_3 of the form (2.9). Further, if c in (2.8) satisfies (2.7) with k = 2, then we have from (2.3) and (2.12) that

$$r(G_0, \delta_{G_2, \mathbf{w}^*}) \leq r(G_0, d_{G_2}) \leq r(G_0, d_G). \tag{2.13}$$

PROOF. The derivation of $\mathbf{w}^* = (w_1^*, w_2^*)$ is rather lengthy but straightforward. Therefore, we give only the main steps of the computation here. Let $w_0 = 1 - (w_1 + w_2)$. Then, $0 \leq w_0 < 1$, $w_0 + w_1 + w_2 = 1$, and the BEB estimator $\delta_{G_2, \mathbf{w}}(\mathbf{X}_3)$, given by (2.9), takes the form

$$\delta_{G_2, \mathbf{w}}(\mathbf{X}_3) = \alpha\widehat{\theta_3} + (1 - \alpha)[w_2\widehat{\theta_2} + w_1\widehat{\theta_1} + w_0\mu], \tag{2.14}$$

where $\mu = E_G(\theta)$. Subject to the condition $w_0 + w_1 + w_2 = 1$, we now minimize

$$\begin{aligned}
r(G_0, \delta_{G_2, \mathbf{w}}) &= E[\delta_{G_2, \mathbf{w}}(\mathbf{X}_3) - \theta_3]^2 \\
&= E[\alpha\widehat{\theta_3} + (1 - \alpha)(w_2\widehat{\theta_2} + w_1\widehat{\theta_1} + w_0\mu) - \theta_3]^2 \\
&= E[\alpha(\widehat{\theta_3} - \theta_3) + (1 - \alpha)(w_2\widehat{\theta_2} + w_1\widehat{\theta_1} + w_0\mu) - (1 - \alpha)\theta_3]^2 \\
&= \alpha^2 E(\widehat{\theta_3} - \theta_3)^2 + (1 - \alpha)^2 E[w_2\widehat{\theta_2} + w_1\widehat{\theta_1} + w_0\mu - \theta_3]^2 + \mathrm{CPT}_1,
\end{aligned} \tag{2.15}$$

(CPT stands for cross product terms) where

$$\begin{aligned}
\mathrm{CPT}_1 &= 2\alpha(1 - \alpha)E(\widehat{\theta_3} - \theta_3)[w_2\widehat{\theta_2} + w_1\widehat{\theta_1} + w_0\mu - \theta_3] \\
&= 2\alpha(1 - \alpha)EE_{X_3/\theta_3}(\widehat{\theta_3} - \theta_3)[w_2\widehat{\theta_2} + w_1\widehat{\theta_1} + w_0\mu - \theta_3] \tag{2.16} \\
&= 0.
\end{aligned}$$

Note that the second term on the right-hand side of (2.15) is equal to the product of $(1 - \alpha)^2$, and

$$\begin{aligned}
&E[w_2(\widehat{\theta_2} - \theta_2) + w_1\widehat{\theta_1} + w_0\mu - \theta_3 + w_2\theta_2]^2 \\
&= w_2^2 E(\widehat{\theta_2} - \theta_2)^2 + E[w_1\widehat{\theta_1} + w_0\mu - \theta_3 + w_2\theta_2]^2 + \mathrm{CPT}_2,
\end{aligned} \tag{2.17}$$

where $\mathrm{CPT}_2 = 2w_2 E(\widehat{\theta_2} - \theta_2)[w_1\widehat{\theta_1} + w_0\mu - \theta_3 + w_2\theta_2] = 0$. The second term on the right-hand side of (2.17) is equal to

$$\begin{aligned}
&E[w_1\widehat{\theta_1} + w_0\mu - \theta_3 + w_2\theta_2]^2 \\
&= E[w_1(\widehat{\theta_1} - \theta_1) + w_0\mu - \theta_3 + w_2\theta_2 + w_1\theta_1]^2 \tag{2.18} \\
&= w_1^2 E(\widehat{\theta_1} - \theta_1)^2 + E[w_0\mu + w_1\theta_1 + w_2\theta_2 - \theta_3]^2 + \mathrm{CPT}_3,
\end{aligned}$$

where $CPT_3 = 2w_1 E(\widehat{\theta}_1 - \theta_1)[w_0\mu + w_1\theta_1 + w_2\theta_2 - \theta_3] = 0$. Now, combining (2.15) to (2.18), we have

$$r(G_0, \delta_{G_2,\mathbf{w}}) = \alpha^2 E(\widehat{\theta}_3 - \theta_3)^2 + (1-\alpha)^2 \Big\{ w_2^2 E(\widehat{\theta}_2 - \theta_2)^2 + w_1^2 E(\widehat{\theta}_1 - \theta_1)^2 \\ + E[w_0\mu + w_1\theta_1 + w_2\theta_2 - \theta_3]^2 \Big\}. \tag{2.19}$$

We now minimize (2.19) subject to the restriction $w_0 + w_1 + w_2 = 1$. Let

$$r(\mathbf{w}) = r(G_0, \delta_{G_2,\mathbf{w}}) + \lambda(1 - w_0 - w_1 - w_2), \tag{2.20}$$

where λ denotes the Lagrangian multiplier and $r(G_0, \delta_{G_2,\mathbf{w}})$ is given by (2.19). By differentiating $r(\mathbf{w})$ with respect to w_0, w_1, and w_2 separately and setting equal to zero, the following three equations are obtained:

$$\mu(w_0\mu + w_2\mu_0 + w_1\mu_0 - \mu_0) = \lambda_1, \tag{2.21}$$
$$w_1 E(\widehat{\theta}_1 - \theta_1)^2 + w_0\mu_0\mu + w_2\mu_0^2 + w_1 E_{G_0}(\theta^2) - \mu_0^2 = \lambda_1, \tag{2.22}$$
$$w_2 E(\widehat{\theta}_2 - \theta_2)^2 + w_0\mu_0\mu + w_1\mu_0^2 + w_2 E_{G_0}(\theta^2) - \mu_0^2 = \lambda_1, \tag{2.23}$$

where $\lambda_1 = \lambda/2(1-\alpha)^2$. Now, from (2.21), (2.22), and (2.23) we obtain that

$$w_1 = \frac{\lambda_1(\mu_0 - \mu)}{\mu\mu_0^2 - \mu[E(\widehat{\theta}_1 - \theta_1)^2 + E_{G_0}(\theta^2)]}, \tag{2.24}$$

$$w_2 = \frac{\lambda_1(\mu_0 - \mu)}{\mu\mu_0^2 - \mu[E(\widehat{\theta}_2 - \theta_2)^2 + E_{G_0}(\theta^2)]}. \tag{2.25}$$

Substituting w_1 of (2.24) and w_2 of (2.25) into (2.21) and solving for w_0 gives

$$w_0 = \frac{\mu\mu_0 + \lambda_1[1 + (a_1 + a_2)(\mu_0^2 - \mu\mu_0)]}{\mu^2}, \tag{2.26}$$

where a_1 and a_2 are as defined in the theorem. Now, from $w_0 + w_1 + w_2 = 1$ with w_1, w_2, and w_0 given by (2.24), (2.25), and (2.26), respectively, we obtain the following solution for λ_1:

$$\lambda_1 = \frac{\mu^2 - \mu\mu_0}{1 + (a_1 + a_2)(\mu_0 - \mu)^2}. \tag{2.27}$$

The proof is now completed by substituting λ_1 of (2.27) in (2.24) and (2.25) and then using the facts that $E(\widehat{\theta}_1 - \theta_1)^2 = V(\widehat{\theta}_1) + (\mu - \mu_0)^2$ and $E(\widehat{\theta}_2 - \theta_2)^2 = V(\widehat{\theta}_2) + (\mu - \mu_0)^2$. □

For general k, we consider estimators of θ_{k+1} of the form

$$\delta_{G_k,\mathbf{w}}(\mathbf{X}_{k+1}) = \alpha\widehat{\theta_{k+1}} + (1-\alpha)\left[\sum_{i=1}^{k} w_i\widehat{\theta_i} + \left(1-\sum_{i=1}^{k} w_i\right)E_G(\theta)\right], \qquad (2.28)$$

where $\widehat{\theta_i}$ is the UMVUE of θ_i based on $\mathbf{X}_i$, $i = 1,\dots,k$, $\alpha \in [0,1)$ and $0 \leq w_i < 1$, $i = 1,\dots,k$ such that $0 \leq \sum_{i=1}^{k} w_i < 1$. Again, optimum values of $\mathbf{w} = (w_1,\dots,w_k)$ may be obtained by minimizing the Bayes risk $r(G_0,\delta_{G_k,\mathbf{w}}) = E[\delta_{G_k,\mathbf{w}}(\mathbf{X}_{k+1}) - \theta_{k+1}]^2$ with respect to $\mathbf{w} = (w_1,\dots,w_k)$ subject to $0 \leq \sum_{i=1}^{k} w_i < 1$. The solution is given in the next theorem. We state the theorem without proof since the proof is similar to that of Theorem 2.1.

THEOREM 2.2. *Let $r(G_0,\delta_{G_k,\mathbf{w}})$ denote the Bayes risk of $\delta_{G_k,\mathbf{w}}(\mathbf{X}_{k+1})$ given by (2.28) with respect to G_0, that is, $r(G_0,\delta_{G_k,\mathbf{w}}) = E[\delta_{G_k,\mathbf{w}}(\mathbf{X}_3) - \theta_{k+1}]^2$. Then, the values of $\mathbf{w} = (w_1,w_2,\dots,w_k)$ that minimize $r(G_0,\delta_{G_k,\mathbf{w}})$ are obtained as the solution to the system of equations given by*

$$\underline{A}\underline{W} = \underline{\mu}, \qquad (2.29)$$

where

$$\underline{A} = \begin{bmatrix} \mu^2 & \mu\mu_0 & \cdot & \cdot & \cdots & \mu\mu_0 \\ \mu\mu_0 & E(\widehat{\theta}_1-\theta_1)^2+E_0(\theta_1^2) & \mu_0^2 & \cdot & \cdots & \mu_0^2 \\ \mu\mu_0 & \mu_0^2 & E(\widehat{\theta}_2-\theta_2)^2+E_0(\theta_2^2) & \mu_0^2 & \cdots & \mu_0^2 \\ & & & & \cdots & \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ & & & & \cdots & \\ & & & & \cdots & \\ \mu\mu_0 & \mu_0^2 & \mu_0^2 & \cdot & \cdots & E(\widehat{\theta}_2-\theta_2)^2+E_0(\theta_2^2) \end{bmatrix},$$

$$\underline{W} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_k \end{bmatrix}, \qquad \underline{\mu} = \begin{bmatrix} \lambda_1 + \mu\mu_0 \\ \lambda_1 + \mu_0^2 \\ \vdots \\ \lambda_1 + \mu_0^2 \end{bmatrix}, \qquad (2.30)$$

$\lambda_1 = \lambda/2(1-\alpha)^2$ and $w_0 = 1 - \sum_{i=1}^{k} w_i$, where $\underline{A}$ is a $(k+1)\times(k+1)$ matrix and $\underline{W}$ and $\underline{\mu}$ are $(k+1)\times 1$ matrices.

TABLE 3.1. Efficiency of $\delta_{G_2,\mathbf{w}^*}(\mathbf{X}_3)$ relative to $d_{G_2}(\mathbf{X}_3)$.

n_1	n_2	n_3	Efficiency (%)				
			$\mu = -0.5$	$\mu = 0$	$\mu = 1.0$	$\mu = 2.0$	$\mu = 3.0$
28	35	73	100.99	101.56	101.80	102.56	110.30
40	35	73	100.99	105.18	113.95	119.54	123.36
60	35	73	100.93	104.83	112.18	115.58	116.94
28	50	73	100.92	104.79	111.94	115.04	116.03
15	35	100	100.88	104.44	110.08	110.86	109.20
28	35	100	100.99	109.52	100.23	100.23	109.52

3. A numerical example. In this section, we apply our proposed method on a real data set in order to see how much improvement is gained by using the proposed BEB estimator over that of S&N estimator. We employed the data provided by Efron [5, Table 1]. The preceding data is based on forty-one randomized trials of a new surgical treatment for stomach ulcers conducted between 1980 and 1989, Sacks et al. [11]. The kth experiment data are recorded as

$$(a_k, b_k, c_k, d_k) \quad (k = 1, \ldots, 41), \tag{3.1}$$

where a_k and b_k are the number of occurrences and nonoccurrences for the Treatment (the new surgery), and c_k and d_k are the occurrences and nonoccurrences for Control (an older surgery). The true log-odds ratio in the kth experimental population is given by

$$\theta_k = \log \left\{ \frac{P_k(\text{Occurrence}|\text{Treatment})}{P_k(\text{Nonoccurrence}|\text{Treatment})} \div \frac{P_k(\text{Occurrence}|\text{Control})}{P_k(\text{Nonoccurrence}|\text{Control})} \right\}. \tag{3.2}$$

An estimate of θ_k is given by

$$\tilde{\theta} = \log \left(\frac{a_k}{b_k} \div \frac{c_k}{d_k} \right), \tag{3.3}$$

$k = 1, \ldots, 41$. These values, along with approximate standard deviations $\widehat{SD_k}$ ($k = 1, \ldots, 41$) of the preceding estimates, are also given in [5, Table 1].

Applying this data to our computation, we acted as if estimates $\tilde{\theta} = \log(a_k/b_k \div c_k/d_k)$ are generated from a normal distribution with mean θ_k and standard deviation $\sigma_k = \widehat{SD_k}$, $k = 1, \ldots, 41$. The values of σ_1, σ_2, and σ_3 are given by $\sigma_1 = 0.86$, $\sigma_2 = 0.66$, and $\sigma_3 = 0.68$ from [5, Table 1]. We further assumed that the prior G_0 given by (2.2) is also normal with the mean μ_0 and standard deviation τ_0. Again, from Efron's paper [5], we obtain $\mu_0 = -1.22$ and $\tau_0 = 1.19$ (see [5, equation (3.8)]). A computational expression of $r(G_0, \delta_{G_2,\mathbf{w}})$ can be easily obtained from (2.19) with w_1 and w_2 replaced by w_1^* and w_2^*,

respectively. In all our computations, we assumed that the α used in expressions (2.4) and (2.9) is equal to $\alpha = 0.6$. Finally, we chose five values of $\mu = E(G)$, namely, $\mu = -0.5, 0, 1.0, 2.0,$ and 3.0, for our computation of $r(G_0, \delta_{G_2, \mathbf{w}^*})$ and $r(G_0, d_{G_2})$, where $r(G_0, d_{G_2})$ denotes the Bayes risk with respect to G_0 of S&N estimator d_{G_k} given by (2.4) with $k = 2$. The sample sizes of experiments 1, 2, and 3 in [5, Table 1] are given by $n_1 = 28$, $n_2 = 35$, and $n_3 = 73$, respectively. For the above specifications, we computed the efficiency of $\delta_{G_2, \mathbf{w}^*}(\mathbf{X}_3)$ relative to $d_{G_2}(\mathbf{X}_3)$, that is, the ratio $r(G_0, d_{G_2})/r(G_0, \delta_{G_2, \mathbf{w}^*})$. As percentages, the resulting values are 100.99, 101.56, 101.80, 102.56, and 110.30 (see the first row of Table 3.1), which shows a better performance of $\delta_{G_2, \mathbf{w}^*}(\mathbf{X}_3)$ relative to $d_{G_2}(\mathbf{X}_3)$. Various other choices of (n_1, n_2, n_3) were also considered, and the relative efficiency of the two estimators were computed. Our results are given in Table 3.1. Also, various other values of μ were investigated, and the results were similar to those in Table 3.1. From Table 3.1, it is clear that when the difference $\mu - \mu_0$ is large, the efficiency of $\delta_{G_2, \mathbf{w}^*}(\mathbf{X}_3)$ relative to $d_{G_2}(\mathbf{X}_3)$ is higher as we would expect. In al cases examined, we observe that $\delta_{G_2, \mathbf{w}^*}(\mathbf{X}_3)$ is relatively more efficient than $d_{G_2}(\mathbf{X}_3)$.

REFERENCES

[1]   D. A. Berry and R. Christensen, *Empirical Bayes estimation of a binomial parameter via mixtures of Dirichlet processes*, Ann. Statist. **7** (1979), no. 3, 558–568.

[2]   B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Monographs on Statistics and Applied Probability, vol. 69, Chapman & Hall, London, 1996.

[3]   H. Cooper and L. V. Hedges (eds.), *The Handbook of Research Synthesis*, Russell Sage Foundation, New York, 1994.

[4]   J. J. Deely and D. V. Lindley, *Bayes empirical Bayes*, J. Amer. Statist. Assoc. **76** (1981), no. 376, 833–841.

[5]   B. Efron, *Empirical Bayes methods for combining likelihoods*, J. Amer. Statist. Assoc. **91** (1996), no. 434, 538–565.

[6]   D. C. Gilliland, J. E. Boyer Jr., and H. J. Tsao, *Bayes empirical Bayes: finite parameter case*, Ann. Statist. **10** (1982), no. 4, 1277–1282.

[7]   J. S. Maritz and T. Lwin, *Empirical Bayes Methods*, Monographs on Statistics and Applied Probability, vol. 35, Chapman & Hall, London, 1989.

[8]   H. Robbins, *An empirical Bayes approach to statistics*, Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I, University of California Press, California, 1956, pp. 157–163.

[9]   _____, *The empirical Bayes approach to statistical decision problems*, Ann. Math. Statist. **35** (1964), 1–20.

[10]   J. E. Rolph, *Bayesian estimation of mixing distributions*, Ann. Math. Statist. **39** (1968), 1289–1302.

[11]   H. S. Sacks, T. C. Chalmers, A. L. Blum, J. Berrier, and D. Pagano, *Endoscopic hemostasis. An effective therapy for bleeding peptic ulcers*, J. Amer. Med. Assoc. **264** (1990), no. 4, 494–499.

[12] F. J. Samaniego and A. A. Neath, *How to be a better Bayesian*, J. Amer. Statist. Assoc. **91** (1996), no. 434, 733–742.

[13] G. G. Walter and G. G. Hamedani, *Bayes empirical Bayes estimation for discrete exponential families*, Ann. Inst. Statist. Math. **41** (1989), no. 1, 101–119.

[14] ———, *Bayes empirical Bayes estimation for natural exponential families with quadratic variance functions*, Ann. Statist. **19** (1991), no. 3, 1191–1224.

R. J. Karunamuni: Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

E-mail address: rkarunam@ualberta.ca

N. G. N. Prasad: Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

E-mail address: prasad@stat.ualberta.ca