Journal of the Korean Statistical Society (2005), **34***: 2, pp 99–107*

ASYMPTOTIC MEAN SQUARED ERROR OF POSITIVE PART JAMES-STEIN ESTIMATORS

MYUNG JOON KIM[1], YEONG-HWA KIM[2]

ABSTRACT

In this paper we consider the asymptotic mean squared error of positive part James-Stein estimators. In the normal-normal example, estimators of the mean squared error of these estimators are provided which are correct asymptotically up to $O(m^{-1})$. Asymptotic estimators of the MSE's which correct up to $O(m^{-1})$ are also provide. Here, m denotes the number of strata. A simulation study is undertaken to evaluate the performance of these estimators.

AMS 2000 subject classifications. Primary 62F12; Secondary 62F15.
Keywords. Positive part James-Stein, mean squared error, asymptotic.

1. INTRODUCTION

The beauty of the Bayesian approach is its ability to structure complicated models, inferential goals and analyses , and also Bayesian techniques are widely used for simultaneous estimation of several parameters in compound decision problems. A well-known example is a small area estimation where interest lies in simultaneous estimation of means or other parameters of interest, say, for counties, census tracks or other local areas. Under any quadratic loss, the Bayes estimates turn out to be the posterior means of the parameters of interest.

Bayesian and empirical Bayesian methods have become quite popular in the theory and practice of statistics in the last two decades. In particular, empirical Bayesian methods are very suitable in the context of simultaneous estimation when there is a genuine need for borrowing strength.

Received November 2004; accepted May 2005.
[1]Automobile Insurance Product Pricing Department, SAMSUNG Fire & Marine Insurance Co., Ltd., Seoul 100–782, Korea (e-mail : myungjoon.kim@samsung.com)
[2]Department of Statistics, Chung-Ang University, Seoul 156-756, Korea (email : gogators@cau.ac.kr)

James-Stein estimators (James and Stein, 1961) have long been popular among statisticians. The theoretical interest in these estimators stems from their minimaxity and other related properties. On the other hand, practitioners have found these estimators quite appealing in the context of simultaneous estimation of parameters when there is a clear need for borrowing strength from neighbors. These estimators have interesting empirical Bayes (EB) interpretation (Efron and Morris, 1973).

The main objective of our research is to find measures of precision associated with the positive part James-Stein estimators which is a little different version of usual EB estimators. We focus on the shrinkage with same direction only. Also we consider these two estimators for the comparison. For all practical purposes, it is not just enough to report only the estimates, but report the margins of error associated with these estimates as well. The estimated mean squared error (MSE) will be of use in constructing confidence sets for parameters of interest, but it seems impossible to find exact MSE's of estimators. One needs at least, some asymptotic approximation of its MSE. We provide instead the asymptotic MSE's of these estimators which are correct up to order $O(m^{-1})$. In detail, we only considered the contribution of lower order terms such as constant and m^{-1} terms, but ignore higher order terms such as m^{-2}, m^{-3}, and etc. We find also estimators of these MSE's which are also valid up to $O(m^{-1})$. Our results are thus similar in spirits to those of Prasad and Rao (1990), Lahiri and Rao (1995) and Datta and Lahiri (2000).

In Section 2 of this paper, we introduce the James-Stein estimators which have interesting empirical Bayes (EB) interpretation. Section 3 provides an estimator of the mean squared error (MSE) which is asymptotically valid upto $O(m^{-1})$. Section 4 contains approximate MSE estimators of the positive-part James-Stein estimator which is also asymptotically valid up to $O(m^{-1})$. In Section 5, we provide some simulation results demonstrating the accuracy of all the approximations.

2. James-Stein Estimators

Efron and Morris (1973) showed how James-Stein estimators arise naturally in an empirical Bayes context. We begin with the situation where $\boldsymbol{X}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \boldsymbol{I}_m)$. Suppose that the prior distribution for θ_i is $N(\mu_i, A)$. Assuming the squared error loss, the Bayes estimator of θ_i is given by

$$\begin{aligned} E(\theta_i|X_i = x_i) &= \frac{x_i + \mu_i/A}{1+1/A} = \frac{A}{A+1}x_i + \frac{1}{A+1}\mu_i \\ &= \mu_i + \left(1 - \frac{1}{A+1}\right)(x_i - \mu_i) \\ &= \mu_i + (1-B)(x_i - \mu_i) \ , \end{aligned}$$

where $B = (1+A)^{-1}$.
Suppose that $\boldsymbol{\mu}$ is known, but A is unknown, and it has to be estimated from the data. Note that $\boldsymbol{X} \sim N(\boldsymbol{\mu}, B^{-1}\boldsymbol{I}_m)$.
Thus, marginally, $||\boldsymbol{X} - \boldsymbol{\mu}||^2 \sim B^{-1}\chi^2_m$. Hence,

$$E\left[\frac{m-2}{||\boldsymbol{X} - \boldsymbol{\mu}||^2}\right] = B.$$

Substituting this estimator for B, one gets the estimator for $\boldsymbol{\theta}$ as

$$\boldsymbol{\delta}(\boldsymbol{X}) = \boldsymbol{\mu} + \left(1 - \frac{m-2}{||\boldsymbol{X} - \boldsymbol{\mu}||^2}\right)(\boldsymbol{X} - \boldsymbol{\mu}).$$

Thus, the estimator $\boldsymbol{\delta}$ shrinks $\boldsymbol{X}$ towards an arbitrary point $\boldsymbol{\mu}$.

Suppose now that $\boldsymbol{\mu}' = (\mu, ..., \mu)$ and μ and A are unknown. Now, marginally $\boldsymbol{X}$ is $N(\mu\boldsymbol{I}_m, B^{-1}\boldsymbol{I}_m)$. In this case $(\bar{X}, \sum(X_i - \bar{X})^2)$ is complete sufficient for (μ, A). Since $\sum(X_i - \bar{X})^2 \sim B^{-1}\chi^2_{m-1}$, the UMVUE for B is given by $\frac{m-3}{\sum(X_i - \bar{X})^2}$ when $m \geq 4$. Also, $\bar{X}$ is the UMVUE for μ. Since, in this case the Bayes estimate of θ_i is

$$\begin{aligned} E(\theta_i|X_i = x_i) &= \frac{x_i + \mu/A}{1+1/A} = (1-B)x_i + B\mu \\ &= \mu + (1-B)(x_i - \mu) \ , \end{aligned}$$

substituting the UMVUE estimators of μ and B, it follows that the empirical Bayes estimator of $\boldsymbol{\theta}$ is

$$\begin{aligned} \hat{\boldsymbol{\theta}}^{EB} &= \bar{X}\mathbf{1}_m + \left(1 - \frac{m-3}{(m-1)S}\right)(\boldsymbol{X} - \bar{X}\mathbf{1}_m) \\ &= (1-\hat{B})\boldsymbol{X} + \hat{B}\bar{X}\mathbf{1}_m \ , \end{aligned}$$

where $\hat{B} = \frac{m-3}{(m-1)S}$, and $S = \frac{1}{m-1}\sum(X_i - \bar{X})^2$. It can be shown that $\hat{\boldsymbol{\theta}}^{EB}$ also dominates $\boldsymbol{X}$ for $m > 4$. The estimator $\hat{\boldsymbol{\theta}}^{EB}$ shrinks the usual estimator $\boldsymbol{X}$ of $\boldsymbol{\theta}$

towards $\bar{X}$, and was proposed by Lindley (1962). In the next section, we will find the asymptotic MSE of EB estimator considering with positive shrink part of the estimator. Since we are trying to consider the same direction of shrinkness, we also calculate the MSE of EB estimator when $\hat{B} = \frac{m-3}{(m-1)S}$ in the next section for the comparison. Thus our $\hat{B}$ becomes $min(1, \frac{m-3}{(m-1)S})$.

3. The Asymptotic Mean Squared Error

Consider the usual normal-normal model where $X_i|\theta_i \sim indep.\ N(\theta_i, 1)$, while $\theta_i \sim iid\ N(\mu, A)$, $i = 1, \cdots, m$. Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$. For an estimate $\boldsymbol{d} = (d_1, \cdots, d_m)^T$ of $\boldsymbol{\theta}$, consider the loss $\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{d}) = m^{-1}||\boldsymbol{\theta} - \boldsymbol{d}||^2$, where $||\cdot||$ denotes the Euclidean norm. Then the Bayes estimator of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}^B = (\hat{\theta}_1^B(\boldsymbol{X}), ..., \hat{\theta}_m^B(\boldsymbol{X}))^T, \tag{3.1}$$

where $\hat{\theta}_i^B(\boldsymbol{X}) = (1 - B)X_i + B\mu$, $i = 1, \cdots, m$ and $B = (1 + A)^{-1}$.

In an EB scenario, typically both μ and A are unknown, and need to be estimated from the marginal distributions of the X_i's. Marginally, X_i's are *iid* $N(\mu, B^{-1})$, where $B = (1 + A)^{-1}$. Based on these marginals, $\bar{X} = m^{-1}\sum_{i=1}^m X_i$, and $S = \sum_{i=1}^m (X_i - \bar{X})^2/(m-1)$ is complete sufficient for (μ, B). Following Morris (1981), we estimate μ and B respectively by $\hat{\mu} = \bar{X}$ and $\hat{B} = min(1, \frac{m-3}{(m-1)S})$. We now have the EB estimator of $\boldsymbol{\theta}$ from (3.1) as

$$\hat{\boldsymbol{\theta}}^{EB} = (1 - \hat{B})\boldsymbol{X} + \hat{B}\bar{X}\mathbf{1}_m\ . \tag{3.2}$$

Under the assumed loss, the MSE of $\hat{\boldsymbol{\theta}}^{EB}$ is given by

$$\begin{aligned} \mathrm{MSE}(\hat{\boldsymbol{\theta}}^{EB}) &= m^{-1}E||\hat{\boldsymbol{\theta}}^{EB} - \boldsymbol{\theta}||^2 \\ &= m^{-1}E||\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}}^B + \hat{\boldsymbol{\theta}}^B - \boldsymbol{\theta}||^2 \\ &= m^{-1}\{E||\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}}^B||^2 + E||\hat{\boldsymbol{\theta}}^B - \boldsymbol{\theta}||^2\}\ . \end{aligned} \tag{3.3}$$

It is well-known that

$$m^{-1}E||\hat{\boldsymbol{\theta}}^B - \boldsymbol{\theta}||^2 = m^{-1}\sum_{i=1}^m E(\hat{\theta}_i^B - \theta_i)^2 = 1 - B. \tag{3.4}$$

The following lemma provides an asymptotic result of $m^{-1}E||\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}}^B||^2$ correct up to $O(m^{-1})$.

LEMMA 3.1. *Under the given model and the loss,*

$$m^{-1}E||\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}}^{B}||^2 = \frac{3B}{m} + o(m^{-1}). \tag{3.5}$$

PROOF. First by (3.1) and (3.2) we get

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}}^{B} &= (1-B)\boldsymbol{X} + B\mu\mathbf{1}_m - (1-\hat{B})\boldsymbol{X} - \hat{B}\bar{X}\mathbf{1}_m \\ &= (B-\hat{B})\boldsymbol{X} + \hat{B}\bar{X}\mathbf{1}_m - B\mu\mathbf{1}_m \\ &= (B-\hat{B})(\boldsymbol{X}-\bar{X}) + B(\bar{X}-\mu)\mathbf{1}_m \end{aligned} \tag{3.6}$$

Hence, by the independence of $\bar{X}$ and $\boldsymbol{X} - \bar{X}\mathbf{1}_m$,

$$E||\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}}^{B}||^2 = B + E[(\hat{B}-B)^2(m-1)S]. \tag{3.7}$$

Also,

$$\begin{aligned}E[(\hat{B}-B)^2(m-1)S] &= E[(\hat{\hat{B}}-B)^2(m-1)S] + E[(\hat{B}-\hat{\hat{B}})^2(m-1)S] \\ &\quad + 2E[(\hat{\hat{B}}-B)(\hat{B}-\hat{\hat{B}})(m-1)S]. \end{aligned} \tag{3.8}$$

Here, let $\hat{\hat{B}} = \frac{m-3}{(m-1)S}$, then

$$\begin{aligned}E[(\hat{\hat{B}}-B\}^2(m-1)S] &= E[\hat{\hat{B}}^2(m-1)S] + E[B^2(m-1)S] \\ &\quad - 2BE[\hat{\hat{B}}(m-1)S] \\ &= E[\frac{(m-3)^2}{(m-1)S}] - 2(m-3)B + (m-1)B \\ &= (m-3)B - 2(m-3)B + (m-1)B \\ &= 2B , \end{aligned} \tag{3.9}$$

by $E[S] = B^{-1}$ and $E[1/S] = \frac{(m-1)B}{m-3}$. And also

$$\begin{aligned}E[(\hat{B}-\hat{\hat{B}})^2(m-1)S] &= E[(1-\hat{\hat{B}})^2(m-1)SI_{[\frac{m-3}{(m-1)S}>1]}] \\ &\leq E^{1/2}[(1-\hat{\hat{B}})^4(m-1)^2S^2]P^{1/2}\left\{\frac{m-3}{(m-1)S} > 1\right\}. \end{aligned} \tag{3.10}$$

But, by the c_δ-inequality $(a+b)^{1+\delta} \leq 2^\delta(a^{1+\delta}+b^{1+\delta})$ for $a > 0$, $b > 0$ and $\delta > 0$,

$$\begin{aligned}E[(1-\hat{\hat{B}})^4(m-1)^2S^2] &\leq 8E[(1+\hat{\hat{B}}^4)(m-1)^2S^2] \\ &= 8E\left[(m-1)(m+1)B^{-2} + \frac{(m-3)^4B^2}{(m-3)(m-5)}\right] \\ &= O(m^2) , \end{aligned} \tag{3.11}$$

while,

$$\begin{aligned} P\left\{\frac{m-3}{(m-1)S} > 1\right\} &= P\{(m-1)S < m-3\} \\ &= P\{B^{-1}\chi^2_{m-1} - B^{-1}(m-1) < m-3-B^{-1}(m-1)\} \\ &= P\{\chi^2_{m-1} - (m-1) < (m-3)B - (m-1)\} \\ &\le P\{\ |\chi^2_{m-1} - (m-1)| > m-1-(m-3)B\} \\ &\le E\left[\{\chi^2_{m-1} - (m-1)\}^{2r}/\{m-1-(m-3)B\}^{2r}\right] \\ &= O(m^{-r}). \end{aligned} \tag{3.12}$$

Choosing $r > 4$, one gets from (3.10) – (3.12),

$$E[(\hat{B}-\hat{\hat{B}})^2(m-1)S] = o(m^{-1}). \tag{3.13}$$

Next, by (3.9), (3.13) and the Schwarz inequality,

$$\begin{aligned} E[\ |\hat{\hat{B}}-B||\hat{B}-\hat{\hat{B}}|(m-1)S] &\le E^{\frac{1}{2}}[(\hat{\hat{B}}-B\}^2(m-1)S]E^{\frac{1}{2}}[(\hat{B}-\hat{\hat{B}})^2(m-1)S] \\ &= (2B)^{1/2}o(m^{-1}) \\ &= O(1)o(m^{-1}) = o(m^{-1}). \end{aligned} \tag{3.14}$$

Now by (3.13), (3.14) and (3.7), one gets

$$m^{-1}E||\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}}^{B}||^2 = \frac{3B}{m} + o(m^{-1}). \tag{3.15}$$

□

By the previous lemma, one gets the following theorem which provides an asymptotic result of $MSE(\hat{\boldsymbol{\theta}}^{EB})$ correct up to $O(m^{-1})$.

THEOREM 3.1. *Under the given model and the loss function*

$$MSE(\hat{\boldsymbol{\theta}}^{EB}) = 1 - B + \frac{3B}{m} + o(m^{-1}). \tag{3.16}$$

PROOF. By (3.3), we know

$$\text{MSE}(\hat{\boldsymbol{\theta}}^{EB}) = m^{-1}\{E||\hat{\boldsymbol{\theta}}^{EB} - \hat{\boldsymbol{\theta}}^{B}||^2 + E||\hat{\boldsymbol{\theta}}^{B} - \boldsymbol{\theta}||^2\}\ , \tag{3.17}$$

and also by Lemma 3.1, the first term in the RHS is $\frac{3B}{m} + o(m^{-1})$. Also by (3.4), the second term in the RHS is $1-B$. Thus, combining these results, the proof is completed. □

Also, using (3.4), (3.5) and (3.9) with $\hat{\hat{B}}$ one can get the MSE of the usual EB estimators as $1 - B + \frac{3B}{m}$.

4. ESTIMATION OF THE MSE

This section is devoted to estimation of the MSE derived in Section 3. To estimate the MSE expression given in (3.16), we need only to find $E(\hat{B})$. With the same argument as before

$$\begin{aligned} E[\hat{B}] &= E[\hat{B} - \hat{\hat{B}}] + E[\hat{\hat{B}}] \\ &= E\left[(1 - \hat{\hat{B}}) I_{[\frac{m-3}{(m-1)S} > 1]}\right] + E\left[\frac{m-3}{(m-1)S}\right] \ ; \end{aligned}$$

$$E\left[\frac{m-3}{(m-1)S}\right] = B \ ;$$

$$E\left[(1 - \hat{\hat{B}}) I_{[\frac{m-3}{(m-1)S} > 1]}\right] \le E^{1/2}\left[(1 - \hat{\hat{B}})^2\right] P^{1/2}\left\{\frac{m-3}{(m-1)S} > 1\right\}.$$

But,

$$E\left[(1 - \hat{\hat{B}})^2\right] \le 4E\left[1 + \hat{\hat{B}}^2\right] = 4\left(1 + \frac{(m-3)B^2}{m-5}\right) = O(1).$$

Hence, from (3.3)-(3.5) and by choosing $r > 2$,

$$E[\hat{B} - \hat{\hat{B}}] = o(m^{-1}).$$

Now by (3.1),(3.2) and (3.5), one gets

$$E[\hat{B}] = B + o(m^{-1}).$$

Hence,

$$E\left[1 - \hat{B} + \frac{3\hat{B}}{m}\right] = 1 - B + \frac{3B}{m} + o(m^{-1}).$$

Thus $O(m^{-1})$ bias-corrected estimator of the MSE of $\hat{\boldsymbol{\theta}}^{EB}$ is given by $1 - \hat{B} + \frac{3\hat{B}}{m}$.

5. NUMERICAL CALCULATIONS

In this section, we report the results of a simulation study to demonstrate the accuracy of the MSE approximation as described in the previous section. For

TABLE 5.1 *Simulated MSE's of EB estimates as well as asymptotic MSE of EB estimates for selected values of A and m*

		MSE		
A	m	Exact(B)	Simulated (EB)	Asymptotic (EB)
1	10	0.5000	0.6313	0.6500
	30		0.5469	0.5500
	50		0.5289	0.5300
	100		0.5145	0.5150
	300		0.5046	0.5050
2	10	0.6667	0.7600	0.7667
	30		0.6971	0.7000
	50		0.6851	0.6867
	100		0.6759	0.6767
	300		0.6694	0.6700
3	10	0.7500	0.8208	0.8250
	30		0.7719	0.7750
	50		0.7632	0.7650
	100		0.7566	0.7575
	300		0.7518	0.7525

the sake of comparison, we also provide the exact value of the MSE of Bayes estimator, which is $1 - B$.

We now discuss the simulation. For illustration, we consider a simple normal-normal model with $\mu = 0$. We investigate the performance of the simulated MSE corresponding to (3.2) as well as the asymptotically estimated MSE of the EB estimators for several m. The simulated MSE's of the EB estimators are also calculated. Different values of $A = 1,\ 2,\ 3$ are considered.

Details of our simulation study are described below :

(a) First we generate θ_i $(i = 1, \cdots, m)$ from the $N(0, A)$ distribution with fixed A value.

(b) For given θ_i $(i = 1, \cdots, m)$, we generate the data x_i, $i = 1, \cdots, m$ from the $N(\theta_i,\ 1)$ distribution. We repeat steps (a) and (b) $R = 10,000$ times. Then we calculate EB estimates for each simulated data set.

(c) Finally we compute the simulated MSE's as follows ;

$$(mR)^{-1} \sum_{i=1}^{m} \sum_{r=1}^{R} (\hat{\theta}_{ir}^{EB} - \theta_{ir})^2$$

for different values of m after $R = 10,000$ repetitions of the experiment. In addition, we calculate asymptotically estimated MSE of the EB estimates for the same m values.

Table 5.1 reports the values of the simulated MSE of EB estimates as well as the asymptotically estimated MSE of the EB estimates for $m = 10, 30, 50, 100, 300$ and for selected values of A. Not surprisingly, Table 1 shows that the simulated MSE and asymptotic MSE for the EB estimates are fairly close even for $m = 50$. We note also the first order optimality of the EB estimator noting that its Bayes risk tends to $1 - B$ as $m \to \infty$. Also, very slight difference indicates the higher order term in positive part James-Stein estimators.

ACKNOWLEDGEMENTS

The authors are grateful to the referees for their careful reading and valuable suggestions.

REFERENCES

DATTA, G.S. AND LAHIRI, P. (2000) . "A unified measure of uncertainty of estimated best linear unbiased predictors in small-area estimation problems", *Statistica Sinica,* **10**, 613–628.

EFRON, B. AND MORRIS, C. (1973) . " Stein's estimation rule and its competitors - an empirical Bayes approach", *Journal of the American Statistical Association,* **68**, 117–130.

JAMES, W. AND STEIN, C. (1961) . " Estimation with quadratic loss", *Procedings of 4th Berkeley Symposium Math, Statistics and Probability,* Univ. California Press, Berkeley **1**, 361–380.

LAHIRI, P. AND RAO, J.N.K. (1995) . " Robust estimation of mean squared error of small area estimators", *Journal of the American Statistical Association,* **90**, 758–766.

LINDLEY, D.V. (1962) . "Discussion of Professor Stein's paper", *Journal of the Royal Statistical Society,* **B24**, 265–296.

MORRIS, C. (1981) . "Parametric empirical Bayes confidence intervals", In *Scientific Inference, Data Analysis, and Robustness,* eds. G.E.P. Box, T. Leonard and C.F. Jeff Wu, Academic Press, 25–50.

PRASAD, N.G.N. AND RAO, J.N.K. (1990) . "The estimation of the mean squared error of small-area estimators", *Journal of the American Statistical Association,* **85**, 163–171.