Journal of the Korean Data & Information Science Society
2014, **25**(6), 1549–1555

http://dx.doi.org/10.7465/jkdi.2014.25.6.1549
한국데이터정보과학회지

An approach to improving the James-Stein estimator shrinking towards projection vectors†

Tae Ryong Park[1] · Hoh Yoo Baek[2]

[1]Department of Computer Engineering, Seokyeong University
[2]Division of Mathematics and Informational Statistics, Wonkwang University
Received 23 August 2014, revised 12 October 2014, accepted 17 October 2014

Abstract

Consider a p-variate normal distribution $(p-q \geq 3,\ q = rank(P_V)$ with a projection matrix $P_V)$. Using a simple property of noncentral chi square distribution, the generalized Bayes estimators dominating the James-Stein estimator shrinking towards projection vectors under quadratic loss are given based on the methods of Brown, Brewster and Zidek for estimating a normal variance. This result can be extended the cases where covariance matrix is completely unknown or $\sum = \sigma^2 \boldsymbol{I}$ for an unknown scalar σ^2.

Keywords: Generalized Bayes estimator, James-Stein estimator, normal distribution, projection vectors, quadratic loss.

1. Introduction

Let $\boldsymbol{X} = (X_1, \cdots, X_p)'$ be a p-variate random vector normally distributed with unknown mean θ and the identity covariance matrix $\boldsymbol{I}$. Then we consider the problem of estimating θ by $\delta(\boldsymbol{X})$ relative to the quadratic loss function $\|\delta(\boldsymbol{X})-\theta\|^2 = (\delta(\boldsymbol{X})-\theta)'(\delta(\boldsymbol{X})-\theta)$. Every estimator will be evaluated by the risk function $R(\theta, \delta(\boldsymbol{X})) = E[\|\delta(\boldsymbol{X})-\theta\|^2]$.

Stein (1956) showed that the usual estimator $\boldsymbol{X}$ is inadmissible for $p \geq 3$ and James and Stein (1961) constructed the improved estimator, $\delta_1^{JS} = (1-(p-2)\ /\|\boldsymbol{X}\|^2\ )\,\boldsymbol{X}$. Also, Casella and Hwang (1987) has proposed the another improved estimator $\delta_1^{SH} = P_V\boldsymbol{X} + (1-(p-q-2)/||\boldsymbol{X}-P_V\boldsymbol{X}||^2)(\boldsymbol{X}-P_V\boldsymbol{X})$ where P_V is an idempotent and projection matrix and $rank(P_V) = q$. $\boldsymbol{X}$ is dominated by δ_1^{SH} for $p-q \geq 3$. With the similar process of Baranchick (1964), we can construct the positive part estimator $\delta_1^{+SH} = P_V\boldsymbol{X}$ if $\|\boldsymbol{X}-P_V\boldsymbol{X}\|^2 \leq p-q-2$; $\delta_1^{+SH} = \delta_1^{SH}$, otherwise, and we can show that δ_1^{+SH} has a smaller risk than δ_1^{SH} by Baranchik's (1964) method. This is known as an estimator eliminating undesirable properties of δ_1^{SH} that it has singularity at $(P_V\boldsymbol{X})_i$ and changes the sign of each $X_i - (P_V\boldsymbol{X})_i$ for $\|X - P_V\boldsymbol{X}\|^2 \leq p-q-2$. However, δ_1^{+SH} itself is unsatisfactory for θ must be estimated by a projection vector $P_V\boldsymbol{X}$ when $\|X-P_V\boldsymbol{X}\|^2 \leq p-q-2$. Of course, it is known that such a truncated estimator is inadmissible.

† This work was supported by Wonkwang University Research Fund 2013.
[1] Professor, Department of Computer Engineering, Seokyeong University, Seoul 136-704, Korea.
[2] Corresponding author: Professor, Division of Mathematics and Informational Statistics, Wonkwang University, Jeonbuk 570-749, Korea. E-mail: hybaek@wonkwang.ac.kr

In this paper we propose a generalized Bayes estimator dominating δ_1^{SH} based on the ideas used in Brown (1968), Brewster and Zidek (1974), and Park and Baek(2011) for estimating a normal variance. In Section 2, such a smooth estimator is derived and it is shown to be admissible. It should be noted that this admissible estimator dominating δ_1^{SH} is just identical to the generalized Bayes estimator given by Strawderman (1971), Casella and Hwang(1987), and Berger (1976) with $a(=c)=2$. Section 3 discusses the cases where the covariance matrix $\sum$ of $\boldsymbol{X}$ is fully unknown or $\sum=\sigma^2\boldsymbol{I}$ for an unknown scalar σ^2.

2. Admissible estimator dominating δ_1^{SH}

To improve on δ_1^{SH}, we consider the estimator

$$\delta_1(c,r)=\begin{cases}P_V\boldsymbol{X}+(1-c/||\boldsymbol{X}-P_V\boldsymbol{X}||^2)(\boldsymbol{X}-P_V\boldsymbol{X}), & \text{if } ||\boldsymbol{X}-P_V\boldsymbol{X}||^2\le r\\ \delta_1^{SH}, & \text{otherwise,}\end{cases} \tag{2.1}$$

where c and r are positive constants. For a fixed r, we shall find the best $c=c(r)$ in the sense of minimizing the risk. Such an idea is due to Brown (1968) which constructed an improved estimator for a normal variance. Let $\lambda=||\boldsymbol{\theta}-P_V\boldsymbol{\theta}||/2$ and $f_{p-q}(t;\lambda)$ denote the density of a noncentral chi square random variable with the degrees of freedom $p-q$ and the noncentrality λ. Letting

$$c_1(r,\lambda)=p-q-2-2f_{p-q}(r;\lambda)\Big/\int_0^r t^{-1}f_{p-q}(t;\lambda)\,dt, \tag{2.2}$$

we can obtain the following lemma which will be proved later.

Lemma 2.1 (i) The risk function of $\delta_1(c,r)$ is quadratic with respect to c and is minimized at $c=c_1(r,\lambda)$.
(ii) $c_1(r,\lambda)\le c_1(r;0)=c_1(r)$, where $c_1(r)$ is expressed as
$c_1(r)=p-q-2-2[\int_0^1 t^{(p-q)/2-2}\exp\{\frac{1}{2}(1-t)r\}dt]^{-1}$.
(iii) $c_1(r)$ is increasing in r and $0<c_1(r)<p-q-2$.

Lemma 2.1 implies that for all λ, $c_1(r)$ is closer to minimizing value of the risk $R(\theta,\delta_1(c,r))$ than $p-q-2$, so that we obtain the following theorem.

Theorem 2.1 The estimator $\delta_1(c_1(r),r)$ dominates $\delta_1(p-q-2,r)$ or δ_1^{SH}.

Further select $0<r'<r$. By the property (iii) of Lemma 2.1 and a similar manner, it can be seen that $\delta_1(c_1(r),r)$ is dominated by another estimator of the form

$$\delta_1'(c_1,r',r)=\begin{cases}P_V\boldsymbol{X}+\left(1-c_1(r')/\|\boldsymbol{X}-P_V\boldsymbol{X}\|^2\right)(\boldsymbol{X}-P_V\boldsymbol{X}), & \text{if } \|\boldsymbol{X}-P_V\boldsymbol{X}\|^2\le r'\\ P_V\boldsymbol{X}+\left(1-c_1(r)/\|\boldsymbol{X}-P_V\boldsymbol{X}\|^2\right)(\boldsymbol{X}-P_V\boldsymbol{X}), & \text{if } r'<\|\boldsymbol{X}-P_V\boldsymbol{X}\|^2\le r\\ \delta_1^{SH}, & \text{otherwise.}\end{cases} \tag{2.3}$$

Now from the innovative idea of Brewster and Zidek (1974), we select a finite partition of $[0,\infty)$ represented by $0=r_{i,0}<\cdots<r_{i,n_i-1}<r_{i,n_i}=\infty$ for each $i=1,2\cdots$ and a

corresponding estimator

$$\delta_1^{(i)} = P_V\boldsymbol{X} + (1 - c_1(r_{ij})/||\boldsymbol{X} - P_V\boldsymbol{X}||^2)(\boldsymbol{X} - P_V\boldsymbol{X}) \text{ if } r_{i,j-1} < ||\boldsymbol{X} - P_V\boldsymbol{X}||^2 \le r_{ij}.$$

Then, providing $\max_j |r_{i,j} - r_{i,j-1}| \to 0$ and $r_{i,n_{i-1}} \to \infty$ as $i \to \infty$, the sequence $\delta_1^{(i)}$ will converge pointwise to δ_1^*, where

$$\delta_1^* = P_V\boldsymbol{X} + (1 - c_1(||\boldsymbol{X} - P_V\boldsymbol{X}||^2)/||\boldsymbol{X} - P_V\boldsymbol{X}||^2)(\boldsymbol{X} - P_V\boldsymbol{X}). \tag{2.4}$$

It should be noted that δ_1^* is the generalized Bayes estimator given by Strawderman (1971), Berger (1976) with a $(= c) = 2$, and Casella and Hwang's (1987) method against the prior density

$$\pi^*(\theta) = \int_0^1 (2\pi)^{-\frac{p-q}{2}}\lambda^{-2}(\lambda/(1-\lambda))^{\frac{p-q}{2}} \exp\left\{-\frac{1}{2}(\lambda/(1-\lambda))||\theta - P_V\boldsymbol{\theta}||^2\right\} d\lambda.$$

Theorem 2.2 The estimator δ_1^* is an admissible estimator dominating δ_1^{SH}.

Proof Since $\delta_1^{(i)}$ has uniformly smaller risk than δ_1^{SH} for each i, applying Fatou's lemma gives that δ_1^* dominates δ_1^{SH}. The admissibility follows from the result of Brown and Hwang (1982) for the prior density $\pi^*(\theta)$ which satisfies the conditions of (ii) in page 213 of their paper. Hence we get the desired conclusion. □

Proof of Lemma 2.1 Let $W = ||\boldsymbol{X} - P_V\boldsymbol{X}||^2$ and $I(\cdot)$ denote the indicator function. Then for a fixed r, $R(\theta, \delta_1(c, r))$ is minimized at

$$c = \frac{E[\ ||\boldsymbol{X} - P_V\boldsymbol{X}||^{-2}(\boldsymbol{X} - P_V\boldsymbol{X})'(\boldsymbol{X} - \theta)I(||\boldsymbol{X} - P_V\boldsymbol{X}||^2 \le r)\ ]}{E[\ ||\boldsymbol{X} - P_V\boldsymbol{X}||^{-2}\ I(||\boldsymbol{X} - P_V\boldsymbol{X}||^2 \le r)\ ]}$$

$$= E\left[\ \left(1 - \frac{\theta'(\boldsymbol{X} - P_V\boldsymbol{X})}{W}\right) I(W \le r)\ \right] \Big/ \ E\left[\ \frac{1}{W} I(W \le r)\ \right] = c_1^*, \tag{2.5}$$

so that we shall demonstrate that c_1^* given by (2.5) is expressed as $c_1(r, \lambda)$ given in (2.2). Using the similar calculation by Kim *et al.* (2002) and Bock (1975) c_1^* can be represented as

$$c_1^* = \frac{E^J\left[\ I_r(p-q+2J) - \dfrac{2J}{p-q-2+2J}\ I_r(p-q-2+2J)\ \right]}{E^J\left[\ \dfrac{1}{p-q-2+2J}\ I_r(p-q-2+2J)\ \right]}, \tag{2.6}$$

where J is a random variable having a Poisson distribution with mean λ and $I_r(X) = \int_0^r f_\alpha(x)dx$ for a central chi square density $f_\alpha(x)$ with degrees of freedom α. Since $I_r(\alpha+2) = -2f_{\alpha+2}(r) + I_r(\alpha)$, we observe that

$$c_1^* = p - q - 2 - 2E^J[f_{p-q+2J}(r)]/E^J[(p-q-2+2J)^{-1}I_r(p-q-2+2J)],$$

which can be rewritten as $c_1(r;\lambda)$ given by (2.2), and we obtain part(i). For part(ii), It is sufficient to show that

$$f_{p-q}(r, \lambda)/\int_0^r t^{-1}f_{p-q}(t;\lambda)dt \ge f_{p-q}(r)/\int_0^r t^{-1}f_{p-q}(t)dt, \tag{2.7}$$

which follows from the fact that $f_{p-q}(t;\lambda)/f_{p-q}(t)$ is increasing in t. Part(iii) can be easily checked and Lemma 2.1 is proved. □

3. The cases of unknown covariance matrices

In this section we extend the result derived in section 2 to the case where covariance matrix is completely unknown or $\Sigma = \sigma^2 I$ for an unknown scalar σ^2. At first, the case of $\Sigma = \sigma^2 I$ is treated.

Let $\boldsymbol{X}$ and S be a independent random variables with $\boldsymbol{X} \sim N_p(\theta, \sigma^2 I)$ and $S \sim \sigma^2\ \chi_n^2$. Here we want to estimate θ under the loss $\|\widehat{\theta} - P_V\boldsymbol{\theta}\|^2/\sigma^2$. For positive constants c and r, a corresponding estimator to (2.1) is of the form

$$\delta_2(c,r) = \begin{cases} P_V\boldsymbol{X} + \left(1 - cS/\|\boldsymbol{X} - P_V\boldsymbol{X}\|^2\right)(\boldsymbol{X} - P_V\boldsymbol{X}), & \text{if } \dfrac{\|\boldsymbol{X} - P_V\boldsymbol{X}\|^2}{S} \le r \\ \delta_2^{SH}, & \text{otherwise,} \end{cases} \tag{3.1}$$

where, in this case, the James-Stein estimator shrinking towards a projection vector is given by

$$\delta_2^{SH} = P_V\boldsymbol{X}\left\{\left(1 - \frac{p-q-2}{n+2}\right) S/\left\|\boldsymbol{X} - P_V\boldsymbol{X}\right\|^2\right\}(\boldsymbol{X} - P_V\boldsymbol{X}).$$

Define $c_2\,(r)$ by

$$c_2\,(r) = \frac{p-q-2}{n+2} - \frac{2}{n+2}\left[\int_0^1 \frac{(1+r)^{(p+n-q)/2}}{(1+rz)^{(p+n-q)/2}} z^{\frac{p-q}{2}-2} dz\right]^{-1}. \tag{3.2}$$

Theorem 3.1 The estimator $\delta_2(c_2(r), r)$ dominates δ_2^{SH}.

Proof Let $\lambda = \|\theta - P_V\boldsymbol{\theta}\|^2/(2\sigma^2)$. Note that the risk function of $\delta_2(c,r)$ is minimized at

$$c_2(r;\lambda) = \frac{E\left[\left((S/\sigma^2)\left\{1 - (\boldsymbol{X} - P_V\boldsymbol{X})'\,\theta\right\}/\|P_V\boldsymbol{X}\|^2\right) I\left(\left\|\boldsymbol{X} - P_V\boldsymbol{X}\right\|^2/S \le r\right)\right]}{E\left[(S/\sigma^2)^2\left(\sigma^2/\left\|\boldsymbol{X} - P_V\boldsymbol{X}\right\|^2\right) I\left(\left\|\boldsymbol{X} - P_V\boldsymbol{X}\right\|^2/S \le r\right)\right]},$$

which from (2.6), can be expressed by

$$c_2(r;\lambda) = \frac{E^J\left[\int_0^\infty v^{\frac{n}{2}} e^{-\frac{v}{2}}\left\{I_{rv}(p-q+2J) - \dfrac{2J}{p-q-2+2J}\, I_{rv}(p-q-2+2J)\right\} dv\right]}{E^J\left[\int_0^\infty v^{\frac{n}{2}+1} e^{-\frac{v}{2}} \dfrac{1}{p-q-2+2J}\, I_{rv}(p-q-2+2J) dv\right]}$$

$$= \frac{\int_0^\infty v^{\frac{n}{2}} e^{-\frac{v}{2}}\left\{(p-q-2)\int_0^{rv} w^{-1} f_{p-q}\,(w;\lambda)\, dw - 2 f_{p-q}\,(rv;\lambda)\right\} dv}{\int_0^\infty v^{\frac{n}{2}+1} e^{-\frac{v}{2}} \int_0^{rv} w^{-1} f_{p-q}\,(w;\lambda)\, dw dv}. \tag{3.3}$$

By integration by parts,

$$\int_0^\infty e^{-\frac{v}{2}} \left\{ v^{\frac{n}{2}+1} \int_0^{rv} w^{-1} f_{p-q}(w;\lambda)\, dw \right\} dv$$

$$= (n+2) \int_0^\infty v^{\frac{n}{2}} e^{-\frac{v}{2}} \int_0^{rv} w^{-1} f_{p-q}(w;\lambda)\, dw dv + 2 \int_0^\infty v^{\frac{n}{2}} e^{-\frac{v}{2}} f_{p-q}(rv;\lambda)\, dv,$$

so that

$$c_2(r;\lambda) = (p-q-2-2H(\lambda)) / (n+2+2H(\lambda)), \tag{3.4}$$

where

$$H(\lambda) = \frac{\int_0^\infty v^{\frac{n}{2}} e^{-\frac{\nu}{2}}\ f_{p-q}(r\nu;\lambda) dv}{\int_0^\infty v^{\frac{n}{2}} e^{-\frac{\nu}{2}}\ f_0^{rv}(w;\lambda) dw dv}.$$

Let $A(\alpha) = 2^{-\frac{\alpha}{2}} (\Gamma(\frac{\alpha}{2}))^{-1}$ and let

$$g_{p,n}(z,\lambda) = E^J \left[\frac{A(p+2J-q)}{A(n+p+2J-q)} z^{(p+2J-q)/2-1} (1+z)^{-(n+p+2J-q)/2} \right].$$

Then $H(\lambda)$ can be rewritten as $H(\lambda) = g_{p,n}(r;\lambda) / \int_0^r z^{-1} g_{p,n}(z,\lambda) dz$.

Similar to (2.7), we can show that $H(\lambda) \geq H(0)$, so that from (3.4), $c_2(r;\lambda) \leq c_2(r;0)$. Here we can verify that $c_2(r;0)$ is equal to $c_2(r)$ given by (3.2), and that $c_2(r)$ is increasing in r and $0 < c_2(r) < (p-q-2)/(n+2)$. Therefore the proof of Theorem 3.1 is completed. □

As a limiting form corresponding to (2.4), we can take the estimator

$$\delta_2^* = P_V \boldsymbol{X} + \left\{ 1 - c_2 \left(\|\boldsymbol{X} - P_V \boldsymbol{X}\|^2 / S \right) S / \|\boldsymbol{X} - P_V \boldsymbol{X}\|^2 \right\} (\boldsymbol{X} - P_V \boldsymbol{X}),$$

which is identical to the generalized Bayes estimator derived from Park and Baek (2011) when $P_V = \frac{1}{p} J$ and J is the $p \times p$ matrix all entries are $1'$ s. By the same arguments as in Section 2, we can prove the following theorem.

Theorem 3.2 The estimator δ_2^* is the generalized Bayes estimator dominating δ_2^{SH}.

Proof For the case where Σ is fully unknown, the above discussions are directly applied. Let X and S be independent random variables with $\boldsymbol{X} \sim N_p(\theta, .\boldsymbol{\Sigma})$ and $S \sim W_p(n, \boldsymbol{\Sigma})$. Assume that we want to estimate θ under the loss $(\widehat{\theta} - \theta)' \boldsymbol{\Sigma}^{-1} (\widehat{\theta} - \theta)$. Define $c_3(r)$ by

$$c_3(r) = \frac{p-q-2}{n-p+q+3} - \frac{2}{n-p+q+3} \int_0^1 \frac{(1+r)^{\frac{n+1}{n}}}{(1+rt)^{\frac{n+1}{n}+1}} t^{\frac{p-q}{2}+1}. \tag{3.5}$$

The estimator $\delta_3^* = P_V \boldsymbol{X} + [1 - \frac{c_3\{(\boldsymbol{X}-P_V\boldsymbol{X})' S^{-1} (\boldsymbol{X}-P_V\boldsymbol{X})\}}{(\boldsymbol{X}-P_V\boldsymbol{X})' S^{-1} (\boldsymbol{X}-P_V\boldsymbol{X})}](\boldsymbol{X} - P_V \boldsymbol{X})$ is the generalized Bayes estimator modified from Park and Baek (2011) and Lin and Tsa (1973). Note that

$(\boldsymbol{X}-P_V\boldsymbol{X})'\Sigma^{-1}(\boldsymbol{X}-P_V\boldsymbol{X})/(\boldsymbol{X}-P_V\boldsymbol{X})'S^{-1}(\boldsymbol{X}-P_V\boldsymbol{X})$ is distributed as $\chi^2_{n-p+q+1}$ independent of X. Then from Theorem 3.2, it is seen the δ_3^* dominates James-Stein estimator shrinking towards a projection vector which is given by

$$\delta_3^{SH} = P_V\boldsymbol{X} + \left[1 - \frac{p-q-2}{n-p+q+3}\left\{(\boldsymbol{X}-P_V\boldsymbol{X})' S^{-1}(\boldsymbol{X}-P_V\boldsymbol{X})\right\}^{-1}\right](\boldsymbol{X}-P_V\boldsymbol{X}).$$

This completes the proof. □

4. Concluding remarks

There are some special cases of P_V. Let the $O_{p\times p}$ and J be the $p\times p$ matrices all entries are 0's and 1's, respectively. The estimators in Kubokawa (1991) and Park and Baek (2011) are the cases of $P_V = O_{p\times p}$ and $P_V = \frac{1}{p}J$. Another case is $P_V = T(T'T)^{-1}T'$ when $T = \begin{pmatrix} 1 & 1\cdots 1 \\ t_1 & t_2\cdots t_p \end{pmatrix}'$ and $\theta_i = \alpha + \beta t_i$ for known t_i and unknown α and β (Lehmann and Casella, 1999), this is the case of $rank(P_V) = 2$. More general case would be represented as follows. When

$$T = [(1\ 1\cdots\ \cdots 1), (t_{11}\ t_{12}, \cdots\ \cdots t_{1p}), \cdots (t_{h1}\ t_{h2}\cdots\ \cdots t_{hp})]'$$

and $\theta_i = \alpha + \beta_1 t_{1i} + \beta_2 t_{2i} + \cdots + \beta_h t_{hi}$ for known $t_{1i}, t_{2i}, \cdots, t_{hi}$ and unknown α, and $\beta_1, \beta_2, \cdots, \beta_h$, such projection matrices $P_V = T(T'T)^{-1}T'$ are symmetric and idempotent of rank $h+1$.

References

Baranchick, A. (1964). *Multiple regression and estimation of the mean of a multivariate normal distribution*, Technical Report **51**, Department of Statistics, Stanford University, California

Berger, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Annals of Statistics*, **4**, 223-226.

Bock, M. E. (1975). Minimax estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, **3**, 209-218.

Brewster, J. F. and Zidek, J. V. (1974). Improving on equivariant estimators. *Annals of Statistics*, **2**, 21-38.

Brown, L. D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Annals of Mathematical Statistics*, **39**, 29-48.

Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics*, edited by S. S. Gupta and J. Berger, Academic Press, New York.

Casella G. and Hwang J. T. (1987). Employing vague prior information in the construction of confidence sets. *Journal of the Multivariate Analysis*, **21**, 79-104.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 361-379.

Kim, B. H., Baek, H. Y. and Chang, I. H., (2002). Improved estimators of the natural parameters in continuous multiparameter exponential families. *Communications in Statistics*, **31**, 11-29.

Kubokawa, T. (1991). An approach to improving the James-Stein estimator. *Journal of Multivariate Analysis*, **36**, 121-126.

Lehmann, E. L. and Casella G. (1999). *Theory of point estimation*, second edition, Springer. New York.

Park, T. R. and Baek, H. Y. (2011). An approach to improving the Lindley estimator *Journal of the Korean Data & Information Science Society*, **22**, 1251-1256.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 197-206.

Strawderman, W. E. (1971). Proper Bayes minimax estimator for the mean of multivariate normal mean. *Annals of Mathematical Statistics*, **42**, 385-388.