

Optimal Estimation within Class of James-Stein Type Decision Rules on the Known Norm

Hoh Yoo Baek[†]

Abstract

For the mean vector of a p -variate normal distribution ($p \geq 3$), the optimal estimation within the class of James-Stein type decision rules under the quadratic loss are given when the underlying distribution is that of a variance mixture of normals and when the norm $\|\underline{\theta}\|$ is known. It also demonstrated that the optimal estimation within the class of Lindley type decision rules under the same loss when the underlying distribution is the previous type and the norm $\|\theta - \bar{\theta}\underline{1}\|$ with $\bar{\theta} = \frac{1}{p} \sum_{i=1}^n \theta_i$ and $\underline{1} = (1, \dots, 1)'$ is known.

Key words : James-Stein Type Decision Rule, Mean Vector, Quadratic Loss, Optimal Estimation, Underlying Distribution

1. Introduction

The problem considered in that of estimating with quadratic loss function the mean vector of a compound multinormal distribution when the norm $\|\underline{\theta}\|$ is known. The class of estimation rules considered will consist of James-Stein type estimators only. Such a class was introduced by James-Stein (1961) and Baranchick (1964) in order to prove that some of into members dominate the sample mean in the multinormal case.

Strawderman (1974) also derived a similar result for the more general case considered in this paper of a compound multinormal distribution.

The problem of estimation of a mean vector under constraint has an old origin and recently focussed again in the context of curved model in the works of Efron (1978), Hinckley (1977), Amari (1982), Kariya (1989), Perron and Giri (1990), Marchand and Giri (1993), Park and Baek (2011), among others. A study of compound multinormal distributions and the estimation of their location vectors was carried out by Berger (1975) and Kubokawa (1991).

In Section 2, the general setting of this problem and necessary notations are presented. In section 3, the best

James-Stein type estimator of a mean vector in the viewpoint of Branchick's method is derived when it's norm is known and an example is given. Concluding remarks are presented in Section 4.

2. Results and Discussion

2.1. Notation and Preliminaries

Let $\underline{x} = (x_1, \dots, x_p)'$, $p \geq 3$, be an observation from a compound multinormal distribution with unknown location parameter $\underline{\theta} (p \times 1)$ and mixture parameter $H(\cdot)$, where $H(\cdot)$ represents a known c.d.f. defined on the interval $(0, \infty)$.

In other words, It is assumed that the random variable \underline{X} generating the observation \underline{x} admits the representation,

$$\mathcal{L}(\underline{X} | Z = z) = N_p(\underline{\theta}, zI_p), \forall z > 0 \quad (2.1)$$

Z being the positive random variable with c.d.f. $H(\cdot)$.

Then we consider the problem of estimating $\underline{\theta}$ by $\delta(\underline{X})$ relative to the quadratic loss function

$$L(\underline{\theta}, \delta(\underline{x})) = (\delta(\underline{x}) - \underline{\theta})' (\delta(\underline{x}) - \underline{\theta}),$$

with $\underline{\theta} \in \Theta_\lambda = \{\underline{\theta} \in R^p \mid \|\underline{\theta}\| = \lambda, 0 \leq \lambda < \infty\}$ and the decision rule $\delta, \delta(\cdot) : R^p \rightarrow R^p$, is of the form

Division of Mathematics and Informational Statistics, Wonkwang University, Iksan, Jeonbuk 570-749, Korea

[†]Corresponding author : hybaek@wonkwang.ac.kr

(Received : August 29, 2012, Revised : September 20, 2012,

Accepted : September 25, 2012)

$$\delta^c(\underline{x}) = (1 - \frac{c}{\underline{x}'\underline{x}})\underline{x}, \quad c \in R$$

Let another decision rule δ_1 , be of the truncated form

$$\delta^{c_1}(\underline{x}) = \begin{cases} (1 - \frac{c_1}{\underline{x}'\underline{x}})\underline{x}, & \text{if } \|\underline{X}\| \leq r, \\ \delta^c(\underline{x}), & \text{otherwise,} \end{cases}$$

where c_1 and r are positive constants. For a fixed r , it will be found the test $c_1 = c_1(r)$ in the sense of which constructed an improved estimator for a normal variance. Restated in terms of the family of probability functions of \underline{X} , the distributional assumption given by expression (2.1) and the restriction on the location parameter $\underline{\theta}$ indicate that the p.d.f. of \underline{X} is

$$p_{\underline{\theta}}(\underline{x}) = \int_0^\infty (2\pi z)^{-p/2} \exp\left(\frac{-\|\underline{x} - \underline{\theta}\|^2}{2z}\right) dH(z), \quad (2.2)$$

$\underline{x} \in R^p$ and $\underline{\theta} \in \Theta_\lambda$. It will be also assumed that $E(Z) < \infty$ which will guarantee the existence of the covariance matrix $\Sigma = Cov(\underline{X}) = E(Z)I_p$ and the mean vector $E(\underline{X}) = \underline{\theta}$. The performance of the estimator δ_1 will be measured by its risk function

$$R(\underline{\theta}, \delta_1) = E_{\underline{\theta}}[L(\underline{\theta}, \delta_1(\underline{X}))] = E_{\underline{\theta}}[(\delta(\underline{X}) - \underline{\theta})'(\delta_1(\underline{X}) - \underline{\theta})], \quad \underline{\theta} \in \Theta_\lambda.$$

2.2. Optimal James-Stein Estimation with Truncated Type when the Norm $\|\underline{\theta}\|$ in Known

In this section, the best estimation in derived within

$$\mathcal{D}_1 = \left\{ \delta: R^p \rightarrow R^p \mid \delta(\underline{x}) = \delta^{c_1}(\underline{x}) = \begin{cases} (1 - \frac{c_1}{\underline{x}'\underline{x}})\underline{x}, & \text{if } \underline{x}'\underline{x} \leq r, \\ \delta^c(\underline{x}), & \text{otherwise} \end{cases}, r \geq 0 \right\},$$

where c and c_1 are in R and the parameter space in of the form

$$\Theta_\lambda = \{\underline{\theta} \in R^p \mid \|\underline{\theta}\| = \lambda\}, \quad \lambda \geq 0.$$

If $c_1 = c$, the class \mathcal{D}_1 is coincided with the type of Merchand and Giri(1993) as a special case.

The following lemmas will prove useful in the

evaluation of the risk function of the decision rule δ^{c_1} , $c_1 \in R$.

Lemma 2.2.1. Let \underline{X} be a random multivariate vector $N_p(\underline{\theta}, I_p)$, $p \geq 3$ and $\underline{\theta} \in R^p$ and $I(\cdot)$ denote the indicator function. Then

$$(i) E_{\underline{\theta}}\left[\frac{1}{\underline{X}'\underline{X}}I(\|\underline{X}\|^2 \leq r)\right] = E^L\left[\frac{1}{p+2L-2}I_r(p+2L-2)\right]$$

and

$$(ii) E_{\underline{\theta}}\left[\frac{(\underline{X}-\underline{\theta})'\underline{X}}{\underline{X}'\underline{X}}I(\|\underline{X}\|^2 \leq r)\right] = E^L\left[I_r(p+2L) - \frac{1}{p+2L-2}I_r(p+2L-2)\right],$$

where L is Poisson random variable with mean $\underline{\theta}'\underline{\theta}/2$ and $I_r(\alpha) = \int_0^r g_\alpha(x)dx$ for a central chi square density $g_\alpha(x)$ with degrees of freedom α .

Proof. Using the similar calculation by Bock (1975) and Kim et. al. (1995), this lemma can be proved.

Lemma 2.2.2. Let \underline{X} be a compound multinormal vector with location parameter $\underline{\theta}$, $p \leq 3$ and $\underline{\theta} \in R^p$, and known mixture parameter $H(\cdot)$ with p.d.f. of the form given in (2.2). Then with $\lambda = \|\underline{\theta}\|$,

$$(i) E_{\underline{\theta}}\left[\frac{1}{\underline{X}'\underline{X}}I(\|\underline{X}\|^2 \leq r)\right] = \int_0^\infty E^L\left[\frac{1}{p+2L-2}I_r(p+2L-2)\right]\frac{dH(z)}{z},$$

and

$$(ii) E_{\underline{\theta}}\left[\frac{(\underline{X}-\underline{\theta})'\underline{X}}{\underline{X}'\underline{X}}I(\|\underline{X}\|^2 \leq r)\right] = \int_0^\infty E^L\left[I_r(p+2L) - \frac{2L}{p+2L-2}I_r(p+2L-2)\right]dH(z),$$

where L is Poisson random variable with mean $\frac{\lambda^2}{2z}$ and

$I_r(\alpha) = \int_0^r g_\alpha(x)dx$ for a central chi square density $g_\alpha(x)$ with degrees of freedom α .

Proof. (i) Using both the representation gives in (2.1)

and part (i) of Lemma 3.1, the following expression can be obtained;

$$\begin{aligned}
 & E_{\theta} \left(\frac{1}{\underline{X}' \underline{X}} I(\|\underline{X}\|^2 \leq r) \right) \\
 &= E^Z \left\{ Z^{-1} E_{\theta}^{X|Z} \left[\frac{Z}{\underline{X}' \underline{X}} I\left(\frac{\|\underline{X}\|^2}{Z} \leq \frac{r}{Z}\right) \right] \right\} \\
 &= \int_0^{\infty} z^{-1} \sum_{l=0}^{\infty} \frac{e^{-\frac{\lambda^2}{2z}} \left(\frac{\lambda^2}{2z}\right)^l}{l!(p+2l-2)} \times \\
 & \quad \left(\int_0^{\frac{r}{z}} \frac{v^{\frac{p+2l-2}{2}-1} e^{-\frac{v}{z}}}{\Gamma\left(\frac{p+2l-2}{2}\right) 2^{\frac{p+2l-2}{2}}} dv \right) \frac{dH(z)}{z} \\
 &= \int_0^{\infty} E^L \left[\frac{1}{p+2L-2} I_r(p+2L-2) \right] \frac{dH(z)}{z}
 \end{aligned}$$

(ii) Again, combining the representation given in (2.1) and part (ii) of Lemma 3.1, the following expression can be obtained;

$$\begin{aligned}
 & E_{\theta} \left(\frac{(\underline{X}-\theta)' \underline{X}}{\underline{X}' \underline{X}} I(\|\underline{X}\|^2 \leq r) \right) \\
 &= E^Z \left\{ E_{\theta}^{X|Z} \left[\frac{\left(\frac{X-\theta}{\sqrt{Z}}\right)' \left(\frac{X}{\sqrt{Z}}\right)}{\left(\frac{X}{\sqrt{Z}}\right)' \left(\frac{X}{\sqrt{Z}}\right)} I\left(\frac{\|\underline{X}\|^2}{Z} \leq \frac{r}{Z}\right) \right] \right\} \\
 &= E^Z \left\{ E^{X|Z} \left[\left(1 - \frac{\left(\frac{\theta}{\sqrt{Z}}\right)' \left(\frac{X}{\sqrt{Z}}\right)}{\left(\frac{X}{\sqrt{Z}}\right)' \left(\frac{X}{\sqrt{Z}}\right)}\right) I\left(\frac{\|\underline{X}\|^2}{Z} \leq \frac{r}{Z}\right) \right] \right\} \\
 &= \int_0^{\infty} E^L \left[I_r(p+2L) - \frac{2L}{p+2L-2} I_r(p+2L-2) \right] dH(z)
 \end{aligned}$$

Assume that the function $f_p(\cdot, \cdot) : [0, \infty) \rightarrow (0, \infty)$ is defined by the relation $f_p(\lambda, z)$

$$= \sum_{l=0}^{\infty} \frac{e^{-\frac{\lambda^2}{2z}} \left(\frac{\lambda^2}{2z}\right)^l}{l!(p+2l-2)} \int_0^{\frac{r}{z}} \frac{v^{\frac{p+2l-2}{2}}}{\Gamma\left(\frac{p+2l-2}{2}\right) 2^{\frac{p+2l-2}{2}}} dv.$$

Then the main result of this section now follows.

Theorem 2.2.3. Let \underline{x} be a single observation from a p -dimensional location parameter with p.d.f. of the form gives by (2.2). Under the assumption $\theta \in \Theta_{\lambda}$, $p \geq 3$ and

$E[Z] < \infty$, the unique best estimator within the class \mathcal{S}_1 is given by $\delta^{c_1^*}$ where $c_1^* =$

$$\begin{aligned}
 & (p-2) \frac{\int_0^{\infty} f_p(\lambda, z) dH(z)}{\int_0^{\infty} f_p(\lambda, z) \frac{dH(z)}{z}} \\
 & - 2 \frac{\int_0^{\infty} E^L(g_{p+2L}(\frac{r}{2})) dH(z)}{\int_0^{\infty} f_p(\lambda, z) \frac{dH(z)}{z}}
 \end{aligned} \tag{3.1}$$

Proof. It can be shown that for a fixed r , the risk function of $\delta_1(\underline{X})$ is minimized at

$$c_1 = c_1^* = \frac{E\left[\frac{(\underline{X}-\theta)' \underline{X}}{\underline{X}' \underline{X}} I(\|\underline{X}\|^2 \leq r)\right]}{E\left[\frac{1}{\underline{X}' \underline{X}} I(\|\underline{X}\|^2 \leq r)\right]}$$

By Lemma 3.2

$$c_1^* = \frac{\int_0^{\infty} E^L \left[I_r(p+2L) - \frac{2L}{p+2L-2} I_r(p+2L-2) \right] dH(z)}{\int_0^{\infty} E^L \left[\frac{1}{p+2L-2} I_r(p+2L-2) \right] \frac{dH(z)}{z}}$$

Since $I_r(\alpha+2) = I_r(\alpha) - 2g_{\alpha+2}(r)$, it can be proved this theorem.

Example 2.2.1. For $\underline{X} \sim N_p(\theta, \sigma^2 I_p)$, $p \geq 3$, (i.e. $H(z) = 1_{(\sigma^2, \infty)}(z)$ with $1_A(\cdot)$ being the indicator function the set A it can be deduced from Theorem 3.1 that

$$\begin{aligned}
 c_1^* &= (p-2)\sigma^2 - \frac{2E^L(g_{p+2L}(\frac{r}{\sigma^2}))}{f_p(\lambda, \sigma^2)/\sigma^2} \\
 &= \sigma^2 \left\{ (p-2) - \frac{2E^L[g_{p+2L}(\frac{r}{\sigma^2})]}{f_p(\lambda, \sigma^2)} \right\}
 \end{aligned}$$

we can show that and the best estimator

$$\delta^{(p-2)\sigma^2}(x) = \left(1 - \frac{(p-2)\sigma^2}{\underline{x}' \underline{x}}\right) \underline{x}$$

within \mathcal{S} (See Merchand and Giri (1993)) is improved by $\delta_1^{c_1^*}(\underline{x})$ in \mathcal{S}_1 , which is shown in Kubokawa (1991).

3. Conclusion

In the case of $H(z) = 1_{[0, \infty)}(z)$ the previous result is coincided with that of Kobokawa (1991), irregardless of the value of the norm $\lambda = \|\theta\|$. It can be extended to the case when the norm is restricted to an interval. We can also consider another case of Stein type

$$\delta^c(\underline{x}) = \begin{cases} \bar{x}\underline{1} + (1 - \frac{c_1}{\|\underline{x} - \bar{x}\underline{1}\|^2})(\underline{x} - \bar{x}\underline{1}), & \text{if } \|\underline{x} - \bar{x}\underline{1}\|^2 < r, \\ \delta^c, & \text{otherwise} \end{cases},$$

where $\delta^c = \bar{x}\underline{1} + (1 - \frac{c}{\|\underline{x} - \bar{x}\underline{1}\|^2})(\underline{x} - \bar{x}\underline{1})$ and $p \geq 4$.

(See Park and Baek (2011)). This case shrinks \underline{x} toward $\bar{x}\underline{1}$, where $\bar{x} = (1/p) \sum_{i=1}^p x_i$ and $\underline{1}$ is column vector of ones.

By using the similar method, the estimator which minimizes it's risk can be derived.

Acknowledgements

This study was supported by Wonkwang University Research Fund 2011.

Reference

- [1] S. Amari, "Differential geometry of curved exponential families, curvature and information loss," *Ann. Statist.*, Vol.10, pp.357-385, 1982.
- [2] A. Baranchick, "Multiple regression and estimation of the mean of a multivariate normal distribution," Technical Report, Department of Statistics, Stanford university, USA, Vol.51, 1964.
- [3] J. O. Berger, "Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss," *Ann. Statist.*, Vol.4, pp.223-226, 1976.
- [4] M. E. Bock, "Minimax estimation of the mean of multivariate normal distribution," *Ann. Statist.*, Vol.3, pp.209-218, 1975.
- [5] L. D. Brown, "Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters," *Ann. Math. Statist.*, Vol.39, pp.29-48, 1968.
- [6] B. Efron, "The geometry of exponential families," *Ann. Statist.*, Vol.6, pp.362-376, 1978.
- [7] D. V. Hinkley, "Conditional inference about a normal mean with known coefficient of variation," *Biometrika*, Vol.64, pp.105-109, 1977.
- [8] W. James and C. Stein, "Estimation with quadratic loss," In *Roc. 4 th Berkeley Symp. Math. Statist. Probability*, Vol.1, University of California Press, Berkeley, pp.361-380, 1961.
- [9] T. Kariya, "Equivariant estimation in a model with ancillary statistics," *Anal. Statist.*, Vol.17, pp.920-928, 1989.
- [10] B. H. Kim, To W. Koh, and H. Y. Baek, "Estimators with nondecreasing risk in a multivariate normal distribution," *J. Korean Stat. Soc*, Vol.24, pp.257-266, 1995.
- [11] T. Kubokawa, "An approach to improving the James-Stein Estimator," *J. Multivariate Anal.*, Vol.36, pp.121-126, 1991.
- [13] T. R. Park and H. Y. Baek, "An approach to improving the Lindley estimator," *J. Korean Data Inform. Sci. Soc.*, Vol.22, pp.1251-1256, 2011.
- [14] F. Perron and N. Giri, "On the best equivariant estimator of mean of a multivariate normal population," *J. Multivariate Anal.*, Vol.32, pp. 1-16, 1989.
- [15] W. E. Strawderman, "Minimax estimation of location parameters for certain spherically symmetric distributions," *J. Multivariate Anal.*, Vol.4, pp.255-264, 1974.