Improving on the James-Stein Estimator

Yuzo Maruyama

The University of Tokyo

Abstract: In the estimation of a multivariate normal mean for the case where the unknown covariance matrix is proportional to the identity matrix, a class of generalized Bayes estimators dominating the James-Stein rule is obtained. It is noted that a sequence of estimators in our class converges to the positive-part James-Stein estimator.

AMS 2000 subject classifications: Primary 62C10, 62C20; secondary 62A15, 62H12, 62F10, 62F15.
Keywords and phrases: minimax, multivariate normal mean, generalized Bayes, James-Stein estimator, quadratic loss, improved variance estimator.

1. Introduction

Let X and S be independent random variables where X has the p-variate normal distribution $N_p(\theta, \sigma^2 I_p)$ and S/σ^2 has the chi square distribution χ^2_n with n degrees of freedom. We deal with the problem of estimating the mean vector θ relative to the quadratic loss function $\|\delta - \theta\|^2/\sigma^2$. The usual minimax estimator X is inadmissible for $p \geq 3$. James and Stein (1961) constructed the improved estimator

$$\delta^{JS} = \left[1 - ((p-2)/(n+2))S/\|X\|^2\right] X,$$

which is also dominated by the James-Stein positive-part estimator $\delta^{JS}_+ = \max(0, \delta^{JS})$ as shown in Baranchik (1964). We note that shrinkage estimators such as James-Stein's procedure are derived by using the vague prior information that $\lambda = \|\theta\|^2/\sigma^2$ is close to 0. It goes without saying that we would like to get significant improvement of risk when the prior information is accurate. Though δ^{JS}_+ improves on δ^{JS} at $\lambda = 0$, it is not analytic and thus inadmissible. Kubokawa (1991) showed that one minimax generalized Bayes estimator derived by Lin and Tsai (1973) dominates δ^{JS}. This estimator, however, does not improve on δ^{JS} at $\lambda = 0$. Therefore it is desirable to get analytic improved estimators dominating δ^{JS} especially at $\lambda = 0$. In Section 2, we show that the estimators in the subclass of Lin-Tsai's minimax generalized Bayes estimators $\delta^M_\alpha = (1 - \psi^M_\alpha(Z)/Z)X$, where $Z = \|X\|^2/S$ and

$$\psi^M_\alpha(z) = z \frac{\int_0^1 \lambda^{\alpha(p-2)/2}(1+\lambda z)^{-\alpha(n+p)/2-1} d\lambda}{\int_0^1 \lambda^{\alpha(p-2)/2-1}(1+\lambda z)^{-\alpha(n+p)/2-1} d\lambda}$$

with $\alpha \geq 1$, dominate δ^{JS}. It is noted that δ^M_α with $\alpha = 1$ coincides with Kubokawa's estimator. Further we demonstrate that δ^M_α with $\alpha > 1$ improves on δ^{JS} especially at $\lambda = 0$ and that δ^M_α approaches the James-Stein positive-part estimator when α tends to infinity.

2. Main results

We first represent $\psi_\alpha^M(Z)$ through the hypergeometric function

$$F(a,b,c,x) = 1 + \sum_{n=1}^{\infty} \frac{(a)_n(b)_n}{(c)_n}\frac{x^n}{n!} \qquad \text{for} \qquad (a)_n = a\cdot(a+1)\cdots(a+n-1).$$

The following facts about $F(a,b,c,x)$, from Abramowitz and Stegun (1964), are needed;

$$\int_0^x t^{a-1}(1-t)^{b-1}dt = x^a(1-x)^b F(1,a+b,a+1,x)/a \quad \text{for} \quad a,b>1, \tag{2.1}$$

$$c(1-x)F(a,b,c,x) - cF(a,b-1,c,x) + (c-a)xF(a,b,c+1,x) = 0, \tag{2.2}$$

$$(c-a-b)F(a,b,c,x) - (c-a)F(a-1,b,c,x) + b(1-x)F(a,b+1,c,x) = 0, \tag{2.3}$$

$$F(a,b,c,1) = \infty \qquad \text{when } \; c-a-b \le -1. \tag{2.4}$$

Making a transformation and using (2.1), we have

$$\begin{aligned}&\frac{\int_0^1 \lambda^{\alpha(p/2-1)}(1+\lambda z)^{-\alpha(n+p)/2-1}d\lambda}{\int_0^1 \lambda^{\alpha(p/2-1)-1}(1+\lambda z)^{-\alpha(n+p)/2-1}d\lambda}\\ &= \frac{\alpha(p/2-1)}{\alpha(p/2-1)+1}\frac{F\left(1,\alpha(n+p)/2+1,\alpha(p/2-1)+2,z/(z+1)\right)}{F\left(1,\alpha(n+p)/2+1,\alpha(p/2-1)+1,z/(z+1)\right)}.\end{aligned}$$

Moreover by (2.2) and (2.3), we obtain

$$\psi_\alpha^M(z) = \frac{p-2}{n+2}\left[1 - \frac{n+p}{(n+2)F(1,\alpha(n+p)/2,\alpha(p-2)/2+1,z/(z+1)) + p-2}\right]. \tag{2.5}$$

Making use of (2.5), we can easily prove the theorem.

Theorem 2.1. *The estimator δ_α^M with $\alpha \ge 1$ dominates δ^{JS}.*

Proof. We shall verify that $\psi_\alpha^M(z)$ with $\alpha \ge 1$ satisfies the condition for dominating the James-Stein estimator derived by Kubokawa (1994): for $\delta_\psi = (1 - \psi(Z)/Z)X$, $\psi(z)$ is nondecreasing, $\lim_{z\to\infty}\psi(z) = (p-2)/(n+2)$, and $\psi(z) \ge \psi_1^M(z)$. Since $F(1,\alpha(n+p)/2,\alpha(p-2)/2+1,z/(z+1))$ is increasing in z, $\psi_\alpha^M(z)$ is increasing in z. By (2.4), it is clear that $\lim_{z\to\infty}\psi_\alpha^M(z) = (p-2)/(n+2)$. In order to show that $\psi_\alpha^M(z) \ge \psi_1^M(z)$ for $\alpha \ge 1$, we have only to check that $F(1,\alpha(n+p)/2,\alpha(p-2)/2+1,z/(z+1))$ is increasing in α, which is easily verified because the coefficient of each term of the r.h.s. of the equation

$$\begin{aligned}&F(1,\alpha(n+p)/2,\alpha(p-2)/2+1,z/(z+1))\\ &= 1 + \frac{p+n}{p-2+2/\alpha}\frac{z}{1+z} + \frac{(p+n)(p+n+2/\alpha)}{(p-2+2/\alpha)(p-2+4/\alpha)}\left(\frac{z}{1+z}\right)^2 + \cdots\end{aligned} \tag{2.6}$$

is increasing in α. We have thus proved the theorem. $\square$

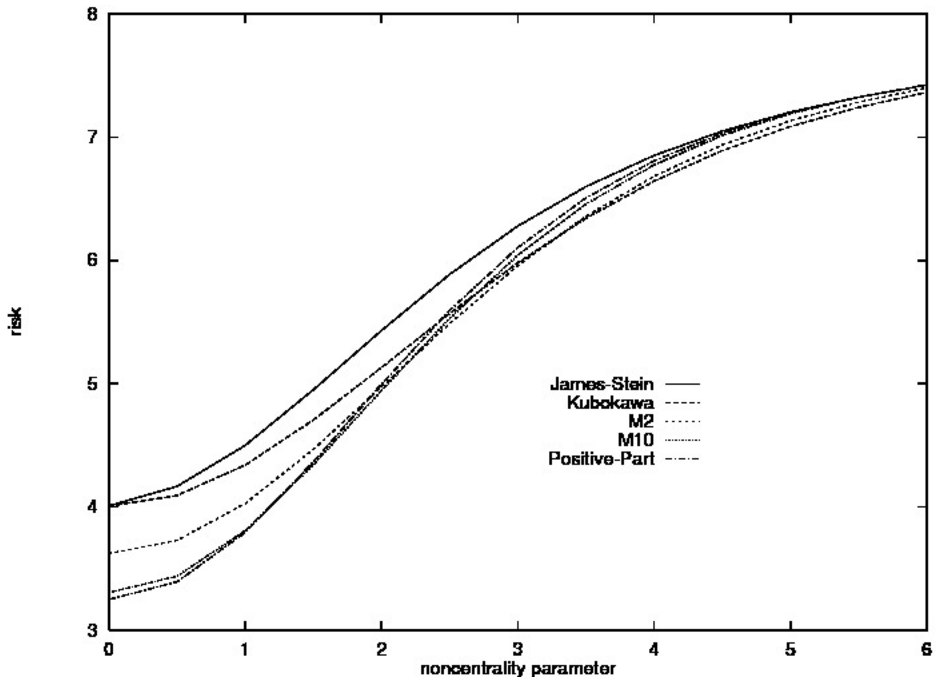

FIG 1. *Comparison of the risks of the estimators δ^{JS}, δ_+^{JS}, δ_α^M with $\alpha = 1, 2, 10$*

Now we investigate the nature of the risk of δ_α^M with $\alpha \geq 1$. By using Kubokawa's (1994) method, the risk difference between the James-Stein estimator and δ_α^M at $\lambda = 0$ is written as

$$R(0, \delta^{JS}) - R(0, \delta_\alpha^M) = 2(n+p) \int_0^\infty \frac{d}{dz} \psi_\alpha^M(z) \left(\frac{\psi_\alpha^M(z) - \psi_1^M(z)}{1 + \psi_1^M(z)} \right) \int_0^z s^{-1} h(s) ds dz,$$

where $h(s) = \int_0^\infty v f_n(v) f_p(vs) dv$ and $f_p(t)$ designates a density of χ_p^2. Therefore we see that Kubokawa's (1991) estimator (δ_α^M with $\alpha = 1$) does not improve upon the James-Stein estimator at $\lambda = 0$. On the other hand, since $\psi_\alpha^M(z)$ is strictly increasing in α, δ_α^M with $\alpha > 1$ improves on the James-Stein estimator especially at $\lambda = 0$ and the risk $R(\lambda, \delta_\alpha^M)$ at $\lambda = 0$ is decreasing in α. Figure 1 gives a comparison of the respective risks of the James-Stein estimator, its positive-part rule and δ_α^M with $\alpha = 1, 2$ and 10 for $p = 8$ and $n = 4$. In the figure 'noncentrality parameter' denotes $\|\theta\|/\sigma$ and Mα denotes the risk of δ_α^M. This figure reveals that the risk behavior of δ_α^M with $\alpha = 10$ is similar to that of the James-Stein positive-part estimator. In fact, we have the following result.

Proposition 2.1. δ_α^M approaches the James-Stein positive-part estimator when α tends to infinity, that is, $\lim_{\alpha \to \infty} \delta_\alpha^M = \delta_+^{JS}$.

Proof. Since $F(1, \alpha(n+p)/2, \alpha(p-2)/2 + 1, z/(z+1))$ is increasing in α, by the monotone convergence theorem this function converges to $\sum_{i=0}^\infty \{((p-2)(z+$

$1))^{-1}(n+p)z\}^i$. Considering two cases: $(n+p)z < (\geq)(p-2)(z+1)$, we obtain $\lim_{\alpha\to\infty}\psi_\alpha^M(z) = z$ if $z < (p-2)/(n+2)$; $= (p-2)/(n+2)$ otherwise. This completes the proof. □

We, here, consider the connection between the estimation of the mean vector and the variance. It is noted that the James-Stein estimator is expressed as $(1-((p-2)/\|X\|^2)\hat{\sigma}_0^2)X$, for $\hat{\sigma}_0^2 = (n+2)^{-1}S$ and that $\hat{\sigma}_0^2$ is the best affine equivariant estimator for the problem of estimating the variance under the quadratic loss function $(\delta/\sigma^2-1)^2$. Stein (1964) showed that $\hat{\sigma}_0^2$ is inadmissible. George (1990) suggested that it might be possible to use an improved variance estimator to improve on the James-Stein estimator. In the same way as Theorem 2.1, we have the following result, which includes Berry's (1994) and Kubokawa *et al.*'s (1993).

Theorem 2.2. *Assume that $\hat{\sigma}_{IM}^2 = \phi_{IM}(Z)S$ is an improved estimator of variance, which satisfies Brewster and Zidek's (1974) condition, that is, $\phi_{IM}(z)$ is nondecreasing and $\phi^{BZ}(z) \leq \phi_{IM}(z) \leq 1/(n+2)$, where*

$$\phi^{BZ}(z) = \frac{1}{n+p+2}\frac{\int_0^1 \lambda^{p/2-1}(1+\lambda z)^{-(n+p)/2-1}d\lambda}{\int_0^1 \lambda^{p/2-1}(1+\lambda z)^{-(n+p)/2-2}d\lambda}.$$

Then $\delta_{IM} = (1-\{(p-2)/\|X\|^2\}\hat{\sigma}_{IM}^2)X$ dominates δ^{JS}.

It should be, however, noticed that the estimator derived from Theorem 2.2 is obviously inadmissible, since $(1-\{(p-2)/\|X\|^2\}\hat{\sigma}_{IM}^2)$ sometimes takes a negative value.

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables. National Bureau of Standards Applied Mathematics Series* **55**. MR0167642 (29 ##4914)

Baranchik, A. J. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. Technical Report, Department of Statistics, Stanford University.

Berry, J. C. (1994). Improving the James-Stein estimator using the Stein variance estimator. *Statist. Probab. Lett.* **20** 241–245. MR1294111

Brewster, J. F. and Zidek, J. V. (1974). Improving on equivariant estimators. *Ann. Statist.* **2** 21–38. MR0381098 (52 ##1995)

George, E. I. (1990). Comment on "Developments in decision-theoretic variance estimation" by Maatta, Jon M. and Casella, George. *Statist. Sci.* **5** 107–109. MR1054858 (91h:62008)

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, Calif. MR0133191 (24 ##A3025)

Kubokawa, T. (1991). An approach to improving the James-Stein estimator. *J. Multivariate Anal.* **36** 121–126. MR1094273 (92b:62070)

Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Ann. Statist.* **22** 290–299. MR1272084 (95h:62011)

Kubokawa, T., Morita, K., Makita, S. and Nagakura, K. (1993). Estimation of the variance and its applications. *J. Statist. Plann. Inference* **35** 319–333. MR1229247 (94d:62076)

Lin, P. E. and Tsai, H. L. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *Ann. Statist.* **1** 142–145.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160. MR0171344 (30 ##1575)