A note on Minimax Estimators of the Mean of a Multivariate Normal Distribution

Ik-Chan Kim

多變量 正規分布의 平均의 Minimax 推定量에 關한 小考

金 益 贊

Summary

In this paper, we treat the condition of minimax estimators for the mean vector of p-variate normal distribution and consider the class of admissible for arbitrary definite and covariance matrix.

1. Indtoruction.

Consider the problem of estimating the mean of $p \geq 3$ variate normal distribution. Stein has shown that the best invariant estimator of p-dimensional normal mean was inadmissible for $p \geq 3$ and it became admissible for p=2. In this paper, we treat good classes of minimax estimators for location vectors of certain symmetric densities and a class of admissible minimax estimators for arbitrary definite matrix Q and covariance matrix Σ.

Let $a_i > 0$ $(i=1,2, \ldots, p)$ be known positive constants. If $x=(x_1, \ldots, x_p)^t$ is a p-dimensional vector, uefine $|x|_1^2 = \sum_{i=1}^{p} x_i^2$, $|x|_2^2 = \sum_{i=1}^{p} x_i^2/a_i$, and $|x|_3^2 = \sum_{i=1}^{p} x_i^2/a_i^2$.

Let $X=(X_1, \ldots, X_p)^t$ be an observation from a p-dimensional density of the form $f(|x-\theta|_2^2)$, here $\theta=(\theta_1, \ldots, \theta_p)^t$ is an unknown location vector. Assume the loss incurred in estimating θ by δ is $|\delta-\theta|^2$

For a measurable estimator $\delta(X)=(\delta_1(X), \ldots, \delta_p(X))^t$, define the risk function $R(\delta, \theta) = E_\theta |\delta(X)-\theta|_1^2$, where E_θ is the expectation under θ. Then the best invariant estimator of θ is $\delta_0(X)=X$ and it is known that δ_0 is a minimax estimator under squred error loss. Furthermore, δ_0 has constant risk. An estimator δ is minimax if $\Delta_\delta(\theta) = R(\delta_0, \theta) - R(\delta, \theta) \geq 0$.

2. A class of minimax estimators.

Let consider the estimators are given componentwise by

(1) $\delta_i(X) = (1 - r(|X|_3^2)/|X|_3^2 \, a_i) X_i$

and density of the form with respect to Lebesque measure

(2) $f(|x-\theta|_2^2)=\int_0^{\infty}(2\pi)^{-p/2}\sigma^{-p}\,(\prod_{i=1}^{p} a_i^{-1/2})\exp[-|x-\theta|_2^2/(2\sigma^2)]\,dF(\sigma)$, where F is a cdf on $(0,\infty)$.

Theorem 1. A density $f(|x-\theta|_2^2)$ is of the form (2) if and only if f is completely monotonic in $(0,\infty)$.

Proof. Define $t=1/(2\sigma^2)$, $s=|x-\theta|_2^2$, and $G(t)=-\int_0^t c(2v)^{p/2}\,dF(1/(2v)^{1/2})$. Then G is positive and nondecreasing t and $f(s)=\int_0^{\infty} e^{-st}\,dG(t)$ is the form of Laplace transforms if and only if it is completely monotone.

Theorem 2. Let $f(|x-\theta|_2^2)$ be a density and X an observation frcm p-dimensional density of form (2), where $p\geq 3$. Assume $E_0|X|_1^2$ and $E_0|X|_1^{-2}$ are finite. If f satisfies

i) The set points W in $(0,\infty)$ at which $f(\cdot)$ is discontinuous has Lebesque measure zero.

ii) $c=\inf_{s\in U}\dfrac{\int_s^{\infty} f(v)dv}{f(s)}>0$, where $U=\{s\notin W: f(s)>0\}$.

Let δ be an estimator of the form (1) where $r(\cdot)$ is nondecreasing and $0\leq r\leq c(p-2)$. Then δ is a minimax estimator of θ under squred error loss.

Proof. Assume that r is differentiable.

(3) $\Delta_\delta(\theta)=E_\theta[|X-\theta|_1^2-|\delta(X)-\theta|_1^2]$

$$=2\sum_{i=1}^{p}E_\theta\left[\frac{r(|X|_3^2)\,X_i(X_i-\theta_i)}{|X|_3^2\,a_i}\right]-E_\theta\left[\frac{r^2(|X|_3^2)}{|X|_3^2}\right].$$

If $\epsilon>0$ then $\int_\epsilon^{\infty} f(v)dv<\infty$, hence if $|x-\theta|_2^2$ is a positive point of continuity of f,

$$\frac{\partial}{\partial x_i}\left(-1/2\int_{|x-\theta|_2^2}^{\infty} f(v)dv\right)=f(|x-\theta|_2^2)(x_i-\theta_i)/a_i.$$

An integration by parts gives

$$(4)\; E_\theta\left[\frac{X_i\,r(|X|_3^2)}{|X|_3^2}\cdot\frac{(X_i-\theta_i)}{a_i}\right]$$

$$=\int\left[\frac{r(|x|_3^2)}{|x|_3^2}-\frac{2x_i^2\,r(|x|_3^2)}{|x|_3^4\cdot a_i^2}+\frac{2x_i^2\,r'(|x|_3^2)}{|x|_3^2\,a_i^2}\right]\left(1/2\int_{|x-\theta|_2^2}^{\infty} f(v)dv\right)dx.$$

Since $r'\geq 0$ by assumption

$$\Delta_\delta(\theta)\geq\int\frac{r(|x|_3^2)}{|x|_3^2}\left[(p-2)\int_{|x-\theta|_2^2}^{\infty} f(v)dv-r(|x|_3^2)\,f(|x-\theta|_2^2)\right]dx.$$

Since $0\leq\gamma\leq c\,(p-2)$,

$$[(p-2)\int_{|x-\theta|_2^2}^{\infty} f(v)dv-r(|x|_3^2)f(|x-\theta|_2^2)]\geq 0,$$

except on $\{x:|x|_3^2=0 \text{ or } |x-\theta|_2^2=0 \text{ or } |x-\theta|_2^2\in W\}$ of measure 0.

Thus $\Delta_\delta(\theta)\geq 0$ and δ is minimax.

If r is not differentiable, the proof completes using Riemann integration, that the terms $r'(|x|_3^2)$ be placed by $dr_i(x_i)$.

For the general situation, let $X=(X_1,\dots,X_p)^t$ be an observation from a p-dimensional normal population of the density of the form $f((x-\theta)^t\Sigma^{-1}(x-\theta))$ with mean vector $\theta=(\theta_1,\dots\theta_p)^t$ and known positive definite covariance matrix Σ.

Then the loss is the quadratic loss $(\delta-\theta)^tQ(\delta-\theta)$ and Q is positive definite and $p\geq 3$.

In terms of the general problem, with arbitrary Σ and Q, the estimator of (1) corresponds to

$$(5)\;\delta(X)=\left(I_p-\frac{r(X^t\Sigma^{-1}Q^{-1}\Sigma^{-1}X)\,Q^{-1}\Sigma^{-1}}{X^t\Sigma^{-1}Q^{-1}\Sigma^{-1}X}\right)X,$$

where r is a measurable function from $R^1\rightarrow R^1$ and I_p the p x p identity matrix. For simplicity let define $\|x\|^2=x^t\Sigma^{-1}Q^{-1}\Sigma^{-1}x$.

Theorem 3. The estimator δ given by (5) is minimax if

i) $0 \leqslant r(\cdot) \leqslant 2(p-2)$ and

ii) $r(\cdot)$ is nondecreasing.

Theorem 3 is a special case of theorem 2 in such of $c=\inf_s [\int_s^\infty f(v)dv/f(s)]=2$. For the choice of r in (5), let α be the smallest characteristic root of the matrix ΣQ, and consider;

$$(6)\quad r_c(t) = \frac{t\int_0^\alpha \lambda^{(p/2-c+1)} \exp(-\lambda t/2)d\lambda}{\int_0^\alpha \lambda^{(p/2-c)} \exp(-\lambda t/2)d\lambda},\ c<1+p/2.$$

Then integration by parts gives

$$(7)\quad r_c(t)=(p-2c+2) - \frac{2\alpha^{(p/2-c+1)} \exp(-\alpha t/2)}{\int_0^\alpha \lambda^{(p/2-c)} \exp(-\lambda t/2)d\lambda}.$$

When $Q=\Sigma^{-1}$ ($\alpha=1$), r_c gives rise to the admissible minimax estimator and the question of choosing c arises.

3. Admissiblity for minimaxity.

Note that in estimating a multivariate normal mean under a quadratic loss, any admissible estimator must be generalized Bayes.

Suppose that X has $N_p(\theta, \Sigma)$, Σ known and that it is desired to estimate θ under a quadratic loss. Then an estimator $\delta(X)$ is generalized Bayes (and hence potentially admissible) if and only if

1) The vector function $g(X)=\Sigma^{-1}\delta(X)$ is continuously differentiable, and the p x p matrix of first partial derivatives of g is symmetric;

2) exp (h(X)) is a Laplace transform (of some generalized prior), where h(X) is the real-valued function which has g(X) as a gradient. As a simple example, consider the estimator $\delta(X)=AX$, where A is a p x p matrix. Clearly $g(X)=\Sigma^{-1}AX$ is continuously differentiable and its matrix of first partial derivatives is $\Sigma^{-1}A$. Hence for above 1) to be satisfied, it must be true that A is of the form ΣB, where B is a symmetric p x p matrix. For 2), h(X) can be seen to be $h(X)=X^tBX/2$, and exp((h(X)) is a Laplace transform of a generalized prior if and only if B is positive definite.

Theorem 4. Assume that δ is given by (5), with $r=r_c$.

a) If $3-p/2 \leqslant c < 1+p/2$, then δ is minimax.

b) If $3-p/2 \leqslant c < 2$, then δ is admissible.

c) If $3-p/2 \leqslant c < 1$, then δ is proper Bayes.

Proof. a). To prove a), it is necessary to verify i) and ii) of theorem 3. From (6), $r_c(\cdot)>0$, since $c \geqslant 3-p/2$ and (7), $r_c(\cdot) < 2(p-2)$.

Hence i) is satified. From (7) and the fact that $\exp((\alpha-\lambda)t/2)$ is nondecreasing in t for $0 \leqslant \lambda \leqslant \alpha$, it follows that $r_c(t)$ is nondecreasing in t. Thus ii) is satisfied.

b). To prove b) and c), δ must be shown to be a generalized Bayes estimator. Let simplify by considering only the case $Q = I_p$ and $\Sigma = A$ where A is a p x p diagonal matrix with diagonal elements $a_i>0$.

Since Q and Σ are positive definite, it can check always to be transformed into this diagonal case. Note that $\|X\|^2 = \sum_{i=1}^p X_i^2/a_i^2$ and $\alpha = \min(a_i)$ and let define $b_i(\lambda) = a_i(a_i-\lambda)/\lambda$ for notation.

For $c<1+k/2$, the generalized prior density

$$(8)\quad g_c(\theta) = \int_0^\alpha [\prod_{i=1}^p b_i(\lambda)^{-1/2}] \exp(-1/2 \sum_{i=1}^p \theta_i^2/b_i(\lambda))\, \lambda^{-c}d\lambda.$$

Since $b_i(\lambda)$ behaves like a_i^2/λ near $\lambda = 0$, $g_c(\cdot)$ is a bounded for the given c. Note also g_c has finite mass if $c<1$.

The generalized Bayes of θ with respect to g_c is given componentwise by

$$(9)\quad \delta_i^c(X) = \frac{\int \theta_i \exp(-1/2 \sum_{i=1}^p (X_i-\theta_i)^2/a_i)\, g_c(\theta)d\theta}{\int \exp(-1/2 \sum_{i=1}^p (X_i-\theta_i)^2/a_i)\, g_c(\theta)d\theta}$$

Put the identities.

$$a_j + b_j(\lambda) = a_j + a_j(a_j-\lambda)/\lambda = a_j^2/\lambda$$

$$(1+a_jb_j(\lambda)^{-1})^{-1} = 1-\lambda/a_j,\ (b_j(\lambda)\ a_j^{-1}+1)^{-1/2} = (\lambda/a_j)^{1/2}$$

Integrating out over θ and using above identities, the numerator of (9) is

(10)

$$\int_0^\alpha (1-\lambda/a_i)X_i \exp(-\lambda\|X\|^2/2)\lambda^{(p/2-c)}\ [\prod_{j=1}^{P} a_j^{-1/2}]\,d\lambda$$

Simarly the denominator of (9) equals

(11)

$$\int_0^\alpha (1-\lambda/a_i)X_i \exp(-\lambda\ \|X\|^2/2)\lambda^{(P/2-c)}\ [\prod_{j=1}^{P} a_j^{-1/2}]\,d\lambda$$

$$= \int_0^\alpha \exp(-\lambda\|X\|^2/2)\ \lambda^{(P/2-c)}\ [\prod_{j=1}^{P} a_j^{-1/2}]\,d\lambda$$

Hence from (9)

(12)

$$\delta_i^c(X) = \frac{\int_0^\alpha (1-\lambda/a_i)X_i \exp(-\lambda\|X\|^2/2)\lambda^{(P/2-c)}\ d\lambda}{\int_0^\alpha \exp(-\lambda\|X\|^2/2)\lambda^{(P/2-c)}\ d\lambda}$$

$$= [\,1-r_c(\|X\|^2)\,/\,(a_i\|X\|^2)]\,X_i\,.$$

Thus $\delta=\delta^c$ is the generalized Bayes estimator with respect to g_c and g_c has finite mass if $c<1$. Hence (c) of theorem 4 follows.

(c). The condition for determing whether or not a generalized Bayes estimator is admissible are based on the behavior of $\delta(X)$ for large $|X|$ or alternatively, on the behavior of the generalized prior for large $|\theta|$ and it was discovered in Brown, i.e.

Suppose that X has $N_p(\theta, I_p)$ and that it is desired to estimate θ under sum of squres error loss. Then a generalized Bayes estimator of the form $\delta(X) = (1\text{-}h(|X|))X$ is admissible if there exist $K_1<\infty$ and $K_2<\infty$ such that $|X|h(|X|)\leq K_1$ for all X and

$$h(X) \geq -\frac{(2-p)}{|X|^2}\ \text{for}\ |X| > K_2\ .$$

Above fact, (11) and a change of variable give

$$(13)\quad \int_{R^P} [\prod_{i=1}^{P}(2\pi a_i)^{-1/2}]\ \exp(-1/2\sum_{i=1}^{P}(X_i-\theta_i)^2/a_i)$$

$$g_c(\theta)d\theta$$

$$= (2\pi)^{-P/2}[\prod_{i=1}^{P} a_i^{-1}]\int_0^\alpha \exp(-\lambda\|X\|^2/2)\lambda^{(P/2-c)}d\lambda$$

$$= (2\pi)^{-P/2}\ [\prod_{i=1}^{P} a_i^{-1}]\ \|X\|^{-(P-2c+2)}\int_0^{\alpha\|X\|^2}$$

$$\exp(-\lambda/2)\lambda^{(P/2-c)}\ d\lambda \leq K|X|^{(2c-P-2)}$$

where $|X|$ denote the norm of X.

Then with the assumption that $c<2$, it follows b) of theorem 4.

Literatures Cited

[1] Berger, J.O. 1976. Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. Ann. Statist. 4.223-226.

[2] Berger, J.O. 1980. Statistical Decision Theory.

[3] Bickel, P.S. and Doksum, K.A. 1977. Mathematical Statistics.

[4] Efron, B. and Morris, C. 1976. Families of minimax estomators of the mean of a multivariate normal distirubtion. Ann. Statist. 4. 11-21.

[5] Giri, N.C. 1977. Multivariate Statistical Inference.

[6] Johnson, R.A. and Wichern, D.W. 1982. Applied Multivariate Statistical Analysis.

國 文 抄 錄

多變量 正規分布의 平均의 Minimax 推定量에 關한 小考

本 論文에서는 多變量正規分布에서의 母集團의 平均 推定을 위한 Minimax의 條件과 그에 수반된 Admissible class에 관하여 정리 요약하였다.