AD-A048 125 CLEMSON UNIV S C DEPT OF MATHEMATICAL SCIENCES F/G 12/1

COMPONENT-WISE MINIMAX PROPERTY OF STEIN'S RULE.(U)

JUL 77 K ALAM N00014-75-C-0451

UNCLASSIFIED N87 NL

1 OF 1

ADA 048125

END

DATE FILMED

1-78

DDC

(12)

(See back page for 1473)

DEPARTMENT OF MATHEMATICAL SCIENCES

CLEMSON UNIVERSITY
Clemson, South Carolina

COMPONENT-WISE MINIMAX PROPERTY OF STEIN'S RULE

BY

KHURSHEED ALAM

N87

JULY 1977

TECHNICAL REPORT #255

Component-Wise Minimax Property of Stein's Rule

Khursheed Alam

Clemson University

ABSTRACT

For estimating the mean of a p-variate normal distribution a family of estimators known as Stein's estimators are known to dominate the maximum likelihood estimator and are therefore minimax when $p \geq 3$, under quadratic loss, equal to the sum of squared errors of the components of the estimator. It is shown in this paper for a subfamily, the vector consisting of any K components of the estimator is also minimax for estimating the corresponding components of the mean vector, where $3 \leq K \leq p$.

1. Introduction. Let the p-component vector $X = (X_1,\dots,X_p)'$ be normally distributed with mean $\theta = (\theta_1,\dots,\theta_K)'$ and convariance $\sigma^2 I$ where $p \geq 3$, I denotes the identity matrix and σ is known. For estimating θ let the loss be equal to the sum of squared errors, given by

$$L(\delta,\theta) = \sum_{i=1}^{p} (\delta_i - \theta_i)^2$$

where $\delta(X) = (\delta_1(X),\dots,\delta_p(X))$ is an estimate of θ, based on X. The maximum likelihood estimator X is minimax but inadmissible for $p \geq 3$ with respect to the given loss function. The inadmissibility of X was first proved by Stein (1955). An estimator which dominates X and is known as Stein's estimator, is given by

(1.1) $$\delta^*(X) = \left(1 - \frac{\nu(p-2)\sigma^2}{S}\right) X$$

where $S = \sum_{i=1}^{p} X_i^2$ and $0<\nu<2$. The risk of δ^* is given by

(1.2) $$\begin{aligned} R(\delta^*,\theta) &= E\, L(\delta^*,\theta) \\ &= p\sigma^2(1-\nu(2-\nu)(p-2)^2 E(\tfrac{\sigma^2}{pS})). \\ &< p\sigma^2 = R(X,\theta). \end{aligned}$$

We are interested in assessing the performance of δ^* component-wise. Let

(1.3) $$R_i^* = E(\delta_i^* - \theta_i)^2$$

denote the contribution to the risk of δ^* from the ith component. It is shown in the following section that for any integer K, such that $3 \le K \le p$

$$\sum_{i=1}^{K} R_i^* < K\sigma^2 \tag{1.4}$$

for $0 < \nu < \frac{2(K-2)}{p-2}$. The above inequality shows that δ^* is minimax for estimating three or more components of θ. We also generalize the above result for the case in which the variances of the components of X are different but known or unknown up to a constant factor. The result is extended to another class of estimators.

2. Main results. We shall compute the value of R_i^*, the contribution to the risk of δ^* from the ith component. First we consider the case where σ is known. We let $\sigma = 1$, without loss of generality. Let $f_m(y)$ denote the density function of a chi-square random variable with m degrees of freedom and

$$g(x) = (2\pi)^{-p/2} \exp(-\tfrac{1}{2}(X-\theta)'(X-\theta))$$

denote the normal density function. For any integrable function ϕ, we have

$$\begin{aligned} E\, X_i \phi(X'X) &= e^{-\theta_i^2/2} \frac{\partial}{\partial \theta_i} (e^{\theta_i^2/2} \int \phi(x'x) g(x) dx) \\ &= \theta_i\, e^{-\theta'\theta/2} \sum_{r=1}^{\infty} \frac{(\theta'\theta/2)^{r-1}}{(r-1)!} \int_0^\infty \phi(y) f_{p+2r}(y) dy \\ &= \theta_i E\phi(\chi^2_{p+2,\theta'\theta}) \end{aligned} \tag{2.1}$$

where $\chi^2_{m,\lambda}$ denotes a non-central chi-square random variable with m degrees of freedom and noncentrality parameter equal to λ. Similarly

$$(2.2) \qquad E\, X_i^2 \phi(X'X) = E\, \phi(\chi^2_{p+2,\theta'\theta}) + \theta_i^2\, E\, \phi(\chi^2_{p+4,\theta'\theta}).$$

Let $\lambda = \theta'\theta$. We have $E(\chi^2_{p+2,\lambda})^{-1} = \frac{1}{p} e^{-\lambda/2} \Phi(\frac{p}{2}, \frac{p}{2} + 1; \frac{\lambda}{2})$ where

$$\Phi(a,b;y) = 1 + \frac{a}{b} y + \frac{a(a+1)}{b(b+1)} \cdot \frac{y^2}{2!} + \cdots$$

denotes the confluent hypergeometric function. From (1.3), (2.1) and (2.2) we get by direct computation

$$(2.3) \quad R_i^* = (1 - \frac{\nu(p-2)}{\chi^2_{p+2,\lambda}})^2 + \theta_i^2[(1 - \frac{\nu(p-2)}{\chi^2_{p+4,\lambda}})^2 - 2(1 - \frac{\nu(p-2)}{\chi^2_{p+2,\lambda}}) + 1]$$

$$= 1 + (p-2)\, e^{-\lambda/2}[\frac{\nu}{p}(\nu\Phi(\frac{p}{2} - 1, \frac{p}{2} + 1; \frac{\lambda}{2}) - 2\Phi(\frac{p}{2}, \frac{p}{2} + 1; \frac{\lambda}{2}))$$

$$+ \theta_i^2\{\frac{\nu^2(p-2)}{p(p+2)} \Phi(\frac{p}{2}, \frac{p}{2} + 2; \frac{\lambda}{2}) + \frac{2\nu}{p} \Phi(\frac{p}{2}; \frac{p}{2} + 1; \frac{\lambda}{2})$$

$$- \frac{2\nu}{p+2} \Phi(\frac{p}{2} + 1, \frac{p}{2} + 2; \frac{\lambda}{2})\}].$$

An upper bound on the value of R_i^* is obtained below, from the following formulas, where prime denotes derivative with respect to y.

$$a(\Phi(a+1,b;y) - \Phi(a,b;y)) = y\Phi'(a,b;y) = \frac{ay}{b} \Phi(a+1,b+1;y)$$

$$(b-a)\,\Phi\,(a,b+1;y) = b\Phi(a,b;y) - b\Phi'(a,b;y).$$

Since $0<\nu<2$; we have

$$(2.4)\qquad \nu\Phi(\tfrac{p}{2}-1,\tfrac{p}{2}+1;\tfrac{\lambda}{2}) - 2\Phi(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2})$$

$$< 2(\Phi(\tfrac{p}{2}-1,\tfrac{p}{2}+1;\tfrac{\lambda}{2}) - \Phi(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2}))$$

$$= -\frac{2\lambda}{p+2}\,\Phi(\tfrac{p}{2},\tfrac{p}{2}+2;\tfrac{\lambda}{2}).$$

The quantity inside the braces on the right side of (2.3) is by the given formulas, equal to

$$(2.5)\qquad \left(\frac{\nu^2(p-2)}{2p} + \frac{2\nu}{p}\right)(\Phi(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2}) - \Phi'(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2}))$$

$$= \frac{\nu}{p(p+2)}\,(\nu(p-2)+4)\Phi(\tfrac{p}{2},\tfrac{p}{2}+2;\tfrac{\lambda}{2}).$$

From (2.3), (2.4) and (2.5) we have

$$(2.6)\qquad R_i^* < 1 + \frac{\nu(p-2)}{p(p+2)}\,e^{-\lambda/2}\Phi(\tfrac{p}{2},\tfrac{p}{2}+2;\tfrac{\lambda}{2})((\nu(p-2)+4)\theta_i^2 - 2\lambda).$$

Therefore, for $3 \leq K \leq p$

$$(2.7)\qquad \sum_{i=1}^{K} R_i^* < K + \frac{\nu\lambda(p-2)}{p(p+2)}\,e^{-\lambda/2}\Phi(\tfrac{p}{2},\tfrac{p}{2}+2;\tfrac{\lambda}{2})(\nu(p-2)+4-2K)$$

$$< K = \sum_{i=1}^{K} R_i$$

for $\nu < \frac{2(K-2)}{p-2}$, where R_i denotes the contribution to the risk of X from the ith component.

Using in (2.3) the asymptotic property of the confluent hypergeometric function, given by

$$\Phi(a,b;y) = \frac{\Gamma(b)}{\Gamma(a)} e^{y} y^{a-b}(1+0(y^{-1})), \quad y>0$$

we have that for large values of λ

(2.8) $$R_i^* = 1 + \frac{\nu(p-2)}{\lambda}((\nu(p-2) + 4)\frac{\theta_i^2}{\lambda} - 2) + 0(\lambda^{-2}).$$

From (2.8) it follows that $\nu<2(K-2)/(p-2)$ is a necessary condition for the inequality (1.4) to hold uniformly in θ. Moreover, it is known that X is admissible for $p\leq 2$. Therefore, the inequality (2.7) cannot be true for $K<3$. Hence, we have shown the following result.

Theorem 2.1. The inequality $\sum_{i=1}^{K} R_i^* < \sum_{i=1}^{K} R_i$ holds for all θ if and only if $3\leq K\leq p$. and $0<\nu<2(K-2)/(p-2)$.

Now we consider the case when σ^2 is unknown but there is available an estimate T, say, such that $T \overset{d}{\sim} \sigma^2\chi_m^2/m$, independent of X. In this case, substituting T for σ^2 in (1.1), Stein's estimator is given by

(2.9) $$\delta^{**}(X) = (1-\frac{\nu(p-2)T}{S})X$$

$$= (1-\frac{\hat{\nu}(p-2)}{S})X$$

where $\hat{\nu} = \nu T$. To compare δ^{**} with the maximum likelihood estimator X we let $\sigma = 1$, as before, without loss of generality.

Substituting $\hat{\nu}$ for ν (1.2) and taking expectation with respect to the distribution of T we find that

$$R(\delta^{**},\theta) < R(X,\theta)$$

for $\nu<2m/(m+2)$. Similarly, we substitute $\hat{\nu}$ for ν in (2.3) and taking expectation with respect to the distribution of T, we obtain the risk of δ^{**} due to the ith component, given by

$$(2.10) \qquad R_i^{**} = 1 + (p-2)e^{-\lambda/2}\Big[\frac{\nu}{p}\Big(\frac{\nu(m+2)}{m}\,\Phi\,(\tfrac{p}{2}-1,\tfrac{p}{2}+1;\tfrac{\lambda}{2}) - 2\Phi(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2})\Big) + \theta_i^2\Big\{\frac{\nu^2(m+2)(p-2)}{mp(p+2)}\,\Phi\,(\tfrac{p}{2},\tfrac{p}{2}+2;\tfrac{\lambda}{2}) + \frac{2\nu}{p}\,\Phi\,(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2}) - \frac{2\nu}{p+2}\,\Phi\,(\tfrac{p}{2}+1,\,\tfrac{p}{2}+2;\tfrac{\lambda}{2})\Big\}\Big].$$

Let $\nu<2m/(m+2)$. Applying the method used in deriving (2.6) we get

$$(2.11) \qquad R_i^{**} = < 1 + (p-2)e^{-\lambda/2}\Big[\frac{2\nu}{p}(\Phi(\tfrac{p}{2}-1,\tfrac{p}{2}+1;\tfrac{\lambda}{2})-\Phi(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2})] + \theta_i^2\,\Big(\frac{\nu^2(m+2)(p-2)}{2mp} + \frac{2\nu}{p}\Big)(\Phi(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2})-\Phi'(\tfrac{p}{2},\tfrac{p}{2}+1;\tfrac{\lambda}{2})]$$

$$= 1 + \frac{\nu(p-2)}{p(p+2)}\,e^{-\lambda/2}\,\Phi\,(\tfrac{p}{2},\tfrac{p}{2}+2;\tfrac{\lambda}{2})\Big[\theta_i^2\Big(\frac{\nu(m+2)(p-2)}{m}+4\Big)-2\lambda\Big].$$

Therefore, for 3 K p

$$(2.12) \quad \sum_{i=1}^{K} R_i^{**} < K + \frac{\nu\lambda(p-2)}{p(b+2)} e^{-\lambda/2} \Phi\left(\frac{p}{2}, \frac{p}{2} + 2; \frac{\lambda}{2}\right)\left(\frac{\nu(m+2)(p-2)}{m} + 4 - 2K\right)$$

$$< K$$

for $\nu < \dfrac{2m(K-2)}{(m+2)(p-2)}$.

From (2.12) we obtain the following theorem extending the result of Theorem 2.1. The necessary part of the theorem is obtained by considering the asymptotic property of the confluent hypergeometric function as in the proof of Theorem 2.1.

Theorem 2.2. The inequality $\sum_{i=1}^{K} R_i^{**} < \sum_{i=1}^{K} R_i$ holds for all θ if and only if $3 \le K \le p$ and $0 < \nu < 2m(K-2)/(m+2)(p-2)$.

If the loss function is suitably modified then the results given above hold also for the case in which the components of X are independent but have unequal variances, as shown below. Let $X_i \overset{d}{\sim} N(\theta_i, \sigma_i^2)$, and let the loss function be given by

$$(2.13) \qquad \hat{L}(\delta,\theta) = \sum_{i=1}^{p} (\delta_i - \theta_i)^2/\sigma_i^2.$$

First suppose that the variances are known. Let $\hat{S} = \sum_{i=1}^{p} X_i^2/\sigma_i^2$. Consider the estimator

$$(2.14) \qquad \hat{\delta} = (1 - \nu(p-2)/\hat{S})X$$

It is easy to see that the conclusion of Theorem 2.1 is valid for $\hat{\delta}$ with respect to the loss given by (2.13). Next, suppose that the variances are known up to a constant factor, that is, let $\sigma_i^2 = \alpha V_i$ where α is unknown but the V_i's are known. If an estimate T is available for α, such that, $T \underset{\sim}{d} \alpha \chi_m^2/m$ then the conclusion of Theorem 2.2 is valid with respect to the loss (2.13) for the estimator.

(2.15) $$\hat{\hat{\delta}} = (1-\nu(p-2)T/\hat{\hat{S}})X .$$

where $$\hat{\hat{S}} = \Sigma_{i=1}^{p} X_i^2/V_i .$$

On the other hand, if the loss function is not modified, that is, we consider the original loss function equal to the sum of squared errors, then we can use the estimators $\hat{\eta}$ and $\hat{\hat{\eta}}$ for $\hat{\delta}$ and $\hat{\delta}$, respectively, in the two cases considered in the preceding paragraph, where

(2.16) $$\hat{\eta}_i = (1-\nu(p-2)/\sigma_i^2 S^*)X_i$$

$$\hat{\hat{\eta}}_i = (1-\nu(p-2)T/\sigma_i^2 S^*)X_i$$

where $S^* = \sum_{i=1}^{p} X_i^2/\sigma_i^4$. It follows from Theorem 1 of Berger (1976) that $\hat{\eta}$ dominates the maximum likelihood estimator X. By direct computation we have for the contribution to the risk of $\hat{\eta}$ from the ith component

(2.18) $$\hat{R}_i = E(\hat{\eta}_i-\theta_i)^2$$

$$= \sigma_i^2 + 2\nu(p-2)E(\frac{2Y_i^2}{\sigma_i^2 V^2} - \frac{1}{V}) + \nu^2(p-2)^2 EY_i^2/\sigma_i^2 V^2$$

where $Y_i = \frac{X_i}{\sigma_i} \underset{\sim}{d} N(\frac{\theta_i}{\sigma_i},1)$ and $V = \sum_{i=1}^{p} Y_i^2/\sigma_i^2$. From (2.18) we have for $3 \le K \le p$

$$(2.19)\quad \sum_{i=1}^{K} \hat{R}_i \le \sum_{i=1}^{K} \sigma_i^2 + 2\nu(p-2)E(\frac{2}{V} - \frac{K}{V}) + \nu^2(p-2)^2 E\frac{1}{V}$$

$$< \sum_{i=1}^{K} \sigma_i^2 = \sum_{i=1}^{K} R_i$$

for $0<\nu< \frac{2(K-2)}{p-2}$. Similarly, for the risk of $\hat{\hat{\eta}}$ we have

$$(2.20)\qquad \sum_{i=1}^{K} \hat{\hat{R}}_i < \sum_{i=1}^{K} R_i \ .$$

for $0<\nu< \frac{2m(K-2)}{(m+2)(p-2)}$. It is easily shown that the sufficient conditions given above for the inequalities (2.19) and (2.20) are also necessary. Therefore, the conclusions of Theorem 2.1 and Theorem 2.2 hold for the estimators $\hat{\eta}$ and $\hat{\hat{\eta}}$, respectively.

Remark: A wide class of estimators for the mean of a multivariate normal distribution is known (see e.g. Bock (1975)) to dominate the maximum likelihood estimator. It would be interesting to investigate whether some of these estimators have the component-wide minimax property of Stein's estimator, shown above. It is also interesting to investigate similar property for the estimator δ^+ , obtained by substituting for the quantity inside the paranthesis on the right side of (1.1) by its positive part, the estimator δ^+ is known to dominate δ^*.

(14) N87, TR-255

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N-87	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Component-Wise Minimax Property of Stein's Rule.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical rept., 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Khursheed/Alam		8. CONTRACT OR GRANT NUMBER(s) (15) NØØØ14-75-C-Ø451
9. PERFORMING ORGANIZATION NAME AND ADDRESS Clemson University Dept. of Mathematical Sciences Clemson, South Carolina 29631		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-271
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Code 436 Arlington, Va. 22217		12. REPORT DATE (11) Jul 77 13. NUMBER OF PAGES (12) 13p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Minimax, Maximum Likelihood Estimator, Stein's Rule.

407 183

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

For estimating the mean of a p-variate normal distribution it is known that Stein's rule is minimax. It is shown for a subfamily of Stein's estimators that the vector consisting of any $K \geq 3$ components of the estimator is also minimax for estimating corresponding components of the mean vector.

> or =

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

References

[1] Berger, J.O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. Ann. Stat. (4) 223-226.

[2] Bock, M.E. (1975). Minimax estimators of the mean of a multivariate normal distribution. Ann. Stat. (3) 209-218.

[3] Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob (1) 197-206.