THE STEIN PARADOX IN THE PITMAN CLOSENESS

by

Pranab K. Sen
Department of Biostatistics
University of North Carolina at Chapel Hill

and

A.K. Md. Ehsanes Saleh
Department of Mathematics & Statistics
Carleton University, Ottawa, Canada

Institute of Statistics Mimeo Series No. 1823

April 1987

THE STEIN PARADOX IN THE PITMAN CLOSENESS

By PRANAB KUMAR SEN[1] and A.K.Md. EHSANES SALEH[2]

University of North Caorlina at Chapel Hill, and
Carleton University, Ottawa, Canada.

The dominance and related optimality properties of the usual Stein-rule estimators rest on the adaptation of appropriate quadratic loss functions. It is shown that in the light of Pitman-closeness, the Stein-rule estimators possess a similar dominance property when such quadratic loss functions are incorporated in the distance function. The impact of the Stein phenomenon under the Pitman closeness criterion is explored in the finite as well as asymptotic cases, and some answers to an open query of Rao (1981) are provided.

AMS Subject Classifications: 62C15, 62H12.

Key Words and Phrases: (Asymptotic and exact) Pitman closeness; dominance; loss function; maximum likelihood estimator; mean, median, mode; noncentral chi square distribution; nonparametric and robust estimators; risk; shrinkage estimators; Stein rule estimator; Hotelling T^2-statistic; U-statistics; Wishart matrices.

[1] Work supported by the Office of Naval Research, Contract N00014-83-K-0387.

[2] Work supported by NSERC of Canada, Grant No. A3088.

1. Introduction. For estimating the mean (vector) of a p-variate normal distribution, under a quadratic loss function, the inadmissibility of the sample mean (i.e., the classical maximum likelihood estimator (MLE)) was established by Stein (1956). Later on, James and Stein (1961), for $p \geq 3$, constructed a shrinkage estimator which dominates the MLE. The past twenty-five years have witnessed a phenomenal growth in the literature on this Stein-rule estimation theory in its diverse tributaries; for some systematic accounts of these related developments, we may refer to Anderson (1984), Arnold (1981) and Berger (1985), among others.

Pitman (1937) laid down the foundation of an important concept of "nearness" or "closeness" of an estimator, and the relationship of this "Pitman closeness" with other conventional measures of efficiency (of estimators) has been explored by a number of workers. However, the role of the Stein-rule estimation theory in the Pitman closeness has not yet been assessed fully. Rao (1981) considered some simple shrinkage estimators and showed that they need not be the Pitman closest ones. This led him to the basic query: Whether a Stein-rule (shrinkage) estimator is (Pitman-) closer than its usual counterpart (in the entire parameter space)? Rao (1981) has argued that the quadratic loss function places undue emphasis on large deviations which may occur with small probability, and minimising the mean square error may insure against large errors in estimator occurring more frequently rather than providing greater concentration of an estimator in neighbourhoods of the true value. This criticism is more justifiable for the Stein-rule estimators which, in general, may not have (multi-) normal distributions, even asymptotically. Sugiura (1984) has successfully incorporated the notion of Pitman closeness for deriving improved estimators of the normal covariance matrix, although his findings are restricted to the asymptotic case only.

The primary objective of the current study is to focus on the Pitman closeness as a suitable criterion for discriminating among competing estimators, and in the light of this, to discuss the appropriateness of the classical Stein-rule estimators. We shall confine ourselves to the estimation of the multinormal mean, where the covariance matrix may be specified, or known up to a multiplicative scalar factor, or even be totally unknown. In the light of the Pitman closeness, we shall show that, indeed, the Stein rule leads to improved estimation in the finite sample as well as in the asymptotic case. Along with the preliminary notions, the case of the known covariance matrix is treated in Section 2. In Section 3, these findings are extended to the case where the covariance matrix is of the form $\sigma^2 \underset{\sim}{V}$, where $\underset{\sim}{V}$ is known and σ^2 is an unknown parameter. The general case of an arbitrary (and unknown) covariance matrix is considered in Section 4. The results obtained in these sections hold for the finite sample size case (i.e., they are of exact nature, not asymptotic). However, these results are capable of being extended immediately to the asymptotic case and pertain to a much wider class of shrinkage estimators considered by the current authors in the past few years. As such, the concluding section is devoted to the role of the Pitman closeness in the characterization of improved estimation in the asymptotic case pertaining to a wider class of shrinkage estimators.

2. Pitman closeness of the James-Stein estimator. An estimator $\hat{\theta}$ of a parameter θ is said to be (Pitman-) closer than another one $\hat{\theta}'$ if

(2.1) $$P_\theta\{|\hat{\theta}-\theta| \leq |\hat{\theta}'-\theta|\} \geq 1/2 .$$

In the multi-parameter case, $\underset{\sim}{\theta}$, $\underset{\sim}{\hat{\theta}}$ and $\underset{\sim}{\hat{\theta}}'$ are all p-vectors, for some $p \geq 1$, so that in (2.1), we may as well consider the Euclidean norm $\|\underset{\sim}{\hat{\theta}}-\underset{\sim}{\theta}\|$ or more generally a norm

(2.2) $$\|\underset{\sim}{\hat{\theta}}-\underset{\sim}{\theta}\|_{\underset{\sim}{W}} = \{(\underset{\sim}{\hat{\theta}}-\underset{\sim}{\theta})'\underset{\sim}{W}(\underset{\sim}{\hat{\theta}}-\underset{\sim}{\theta})\}^{1/2}$$

where $\underset{\sim}{W}$ is a given positive-semi definite (p.s.d.) matrix. Thus, the Euclidean norm relates to $\|\underset{\sim}{\hat{\theta}}-\underset{\sim}{\theta}\|_{\underset{\sim}{I}}$. We say that $\underset{\sim}{\hat{\theta}}$ dominates $\underset{\sim}{\hat{\theta}}'$ in the light of the Pitman closeness, with respect to the norm $\|\cdot\|_{\underset{\sim}{W}}$, if

(2.3) $$P_{\underset{\sim}{\theta}}\{\|\underset{\sim}{\hat{\theta}}\ \underset{\sim}{\theta}\|_{\underset{\sim}{W}} \leq \|\underset{\sim}{\hat{\theta}}'-\underset{\sim}{\theta}\|_{\underset{\sim}{W}}\} \geq 1/2, \quad \forall\ \underset{\sim}{\theta} \in \Theta$$

with the strict inequality sign holding for at least some $\underset{\sim}{\theta} \in \Theta$.

Now, consider the specific case where

(2.4) $$\underset{\sim}{\hat{\theta}} \sim N_p(\underset{\sim}{\theta}, I) .$$

Note that here $E\|\underset{\sim}{\hat{\theta}}-\underset{\sim}{\theta}\|^2_{\underset{\sim}{I}} = p$. James and Stein (1961) considered the shrinkage estimator (for $p \geq 3$):

(2.5) $$\underset{\sim}{\hat{\theta}}^{JS}_c = \{1-c\|\underset{\sim}{\hat{\theta}}\|^{-2}\}\underset{\sim}{\hat{\theta}} \text{ , where } 0<c<2(p-2),$$

and they have shown that $E_{\underset{\sim}{\theta}}\|\underset{\sim}{\hat{\theta}}^{JS}_c-\underset{\sim}{\theta}\|^2_{\underset{\sim}{I}} \leq p$, $\forall\ \theta \in \Theta = E^p$, with more reduction of the risk for $\underset{\sim}{\theta}$ close to $\underset{\sim}{0}$ (i.e., for small $\|\underset{\sim}{\theta}\|$). In this context, the choice of $c = p-2$ is known to have some optimality properties. We present the following

<u>Theorem 2.1</u>. *For every* $p \geq 3$ *and* $0 < c \leq 2(p-2)$, $\hat{\underset{\sim}{\theta}}_c^{JS}$ *dominates* $\hat{\underset{\sim}{\theta}}$ *in the light of the Pitman closeness in* (2.3).

<u>Proof</u>. Note that by (2.5),

$$(2.6) \qquad \|\hat{\underset{\sim}{\theta}}_c^{JS}-\underset{\sim}{\theta}\|^2 = \|(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta})-c\|\hat{\theta}\|^{-2}\hat{\underset{\sim}{\theta}}\|^2$$
$$= \|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}\|^2 - c\|\hat{\underset{\sim}{\theta}}\|^{-2}\{2\|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}\|^2+2\underset{\sim}{\theta}'(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta})-c\},$$

so that $\{\|\hat{\underset{\sim}{\theta}}_c^{JS}-\underset{\sim}{\theta}\|^2 \leq \|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}\|^2\} \iff \{2\|\hat{\theta}-\underset{\sim}{\theta}\|^2+2\underset{\sim}{\theta}'(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta})-c \geq 0\}$.

First, consider the special case of $\underset{\sim}{\theta} = \underset{\sim}{0}$. Then, $\|\hat{\theta}\|^2 \sim \chi_p^2$, and for the central chi square distribution with p degrees of freedom (DF), it is known [viz., Johnson and Kotz (1970)] that

$$(2.7) \qquad \text{mode} = m_p^{(0)} = p-2 < \text{median} = M_p^{(0)} < p = \text{mean}, \ \forall\ p \geq 2.$$

Thus, $P\{2\|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}\|^2+2\theta'(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta})-c \geq 0 \mid \underset{\sim}{\theta}=\underset{\sim}{0}\} > \frac{1}{2}$, $\forall\ c/2 \leq m_p^{(0)} = p-2$.
In fact, here any $c/2 < M_p^{(0)}$ would suffice. Next, consider the case of $\underset{\sim}{\theta} \neq \underset{\sim}{0}$. We take an orthogonal matrix $\underset{\sim}{A}$ and write

$$(2.8) \qquad \underset{\sim}{Y} = \underset{\sim}{A}(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}); \ A' = (\underset{\sim}{a}_1', \underset{\sim}{A}_2'), \ \underset{\sim}{a}_1 = \|\underset{\sim}{\theta}\|^{-1}\underset{\sim}{\theta}', \ \underset{\sim}{A}_2 \perp \underset{\sim}{a}_1.$$

Then $\underset{\sim}{Y} \sim N(\underset{\sim}{0},\underset{\sim}{I})$ and $\underset{\sim}{\theta}'(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}) = \|\theta\|Y_1$. Therefore,

$$(2.9) \qquad \{2\|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}\|^2 + 2\theta'(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta})-c\}$$
$$= 2\|Y\|^2+2\|\underset{\sim}{\theta}\|Y_1-c$$
$$= 2\{\Sigma_{i=2}^p Y_i^2 + (Y_1+\tfrac{1}{2}\|\underset{\sim}{\theta}\|)^2 - (\tfrac{1}{2}c+\tfrac{1}{4}\|\underset{\sim}{\theta}\|^2)\}$$
$$= 2\{\chi_{p,\lambda}^2 - (\tfrac{1}{2}c+\lambda)\}; \quad \lambda = \tfrac{1}{4}\|\underset{\sim}{\theta}\|^2,$$

where $\chi_{p,\lambda}^2$ has the noncentral chi square distribution with p DF and noncentrality parameter λ (≥ 0). Thus, it suffices to show that for every $p \geq 3$, $0 < c \leq 2(p-2)$ and every $\lambda \geq 0$,

$$P\{\chi^2_{p,\lambda} \geq \tfrac{1}{2}c + \lambda\} \geq 1/2. \tag{2.10}$$

Let $q(r;m) = e^{-m}m^r/r!$, $r \geq 0$, be the Poisson probabilities with the parameter $m\ (\geq 0)$, and let $g_s(x) = \{2^{s/2}\Gamma(s/2)\}^{-1}e^{-\frac{1}{2}x}x^{\frac{1}{2}s-1}$, $s \geq 1$, $x \geq 0$ be the probability density functions for the central chi square distributions. Then, $g_p^{(\lambda)}(x)$, the density function of $\chi^2_{p,\lambda}$, is given by

$$g_p^{(\lambda)}(x) = \Sigma_{r \geq 0}\, q(r;\lambda/2)\, g_{p+2r}(x),\ x \geq 0,\ \lambda \geq 0. \tag{2.11}$$

Note that

$$(d/dx)\, g_p^{(\lambda)}(x) = \tfrac{1}{2}\{g_{p-2}^{(\lambda)}(x) - g_p^{(\lambda)}(x)\},\ \forall\ x \geq 0,\ p \geq 2. \tag{2.12}$$

Thus, if $m_p^{(\lambda)}$ stands for the mode of $\chi^2_{p,\lambda}$, we have

$$g_{p-2}^{(\lambda)}(m_p^{(\lambda)}) = g_p^{(\lambda)}(m_p^{(\lambda)}), \tag{2.13}$$

and $g_{p-2}^{(\lambda)}(x)$ is $\gtreqless g_p^{(\lambda)}(x)$ according as x is $\lesseqgtr m_p^{(\lambda)}$; the latter inequality rests on the well known unimodality of $g_p^{(\lambda)}(\cdot)$ [viz., Johnson and Kotz (1970); Ch. 28]. Writing $g_{p+2r-2}(x) = (p-2+2r)x^{-1}g_{p+2r}(x)$, $r \geq 0$, $x \geq 0$, it follows from (2.13) and some standard steps that

$$m_p^{(\lambda)} \leq p-2+\lambda,\ \forall\ \lambda \geq 0,\ p \geq 2. \tag{2.14}$$

In a similar manner, it follows that

$$m_{p+2}^{(\lambda)} \geq p-2+\lambda,\ \forall\ \lambda \geq 0,\ p \geq 2. \tag{2.15}$$

As a result, from (2.12) through (2.15), we conclude that

$$g_p^{(\lambda)}(p-2+\lambda) \geq g_{p+2}^{(\lambda)}(p-2+\lambda),\ \forall\ \lambda \geq 0,\ p \geq 2. \tag{2.16}$$

Let $\bar{G}_s(x)$ be the survival function corresponding to $g_s(x)$, so that

(2.17) $\bar{G}_p^{(\lambda)}(x) = P\{\chi^2_{p,\lambda} \geq x\} = \Sigma_{r\geq 0} q(r;\lambda/2)\bar{G}_{p+2r}(x), \ \forall\ x \geq 0,\ \lambda \geq 0.$

Note that $\bar{G}_s(x) - \bar{G}_{s-2}(x) = 2g_s(x),\ \forall\ s \geq 2,\ x \geq 0$, and hence,

$$\begin{aligned}(\partial/\partial\lambda)\bar{G}_p^{(\lambda)}(p-2+\lambda) &= \Sigma_{r\geq 0}\{\bar{G}_{p+2r}(p-2+\lambda)(\partial/\partial\lambda)q(r;\lambda/2) \\ &\quad + q(r;\lambda/2)(\partial/\partial\lambda)\bar{G}_{p+2r}(p-2+\lambda)\} \\ &= \Sigma_{r\geq 0}\ q(r;\lambda/2)\{\tfrac{1}{2}[\bar{G}_{p+2r+2}(p-2+\lambda)-\bar{G}_{p+2r}(p-2+\lambda)] \\ &\quad -g_{p+2r}(p-2+\lambda)\} \\ &= \Sigma_{r\geq 0}\ q(r;\lambda/2)\{g_{p+2r+2}(p-2+\lambda)-g_{p+2r}(p-2+\lambda)\} \\ &= g_{p+2}^{(\lambda)}(p-2+\lambda)-g_p^{(\lambda)}(p-2+\lambda) \leq 0, \text{ by (2.16)}.\end{aligned}$$

Therefore, we conclude that for every $p\ (\geq 2)$,

(2.19) $\bar{G}_p^{(\lambda)}(p-2+\lambda)$ is nonincreasing in $\lambda\ (\geq 0)$.

On the other hand, by the asymptotic (in λ) normality of $\chi^2_{p,\lambda}$ [viz., Johnson and Kotz (1970)],

(2.20) $$\begin{aligned}\lim_{\lambda\to\infty} \bar{G}_p^{(\lambda)}(p-2+\lambda) &= \lim_{\lambda\to\infty} P\{\chi^2_{p,\lambda}-p-\lambda \geq -2\} \\ &= \lim_{\lambda\to\infty} P\{(\chi^2_{p,\lambda}-p-\lambda)/\sqrt{2(p+2\lambda)} \geq -2/\sqrt{2(p+2\lambda)}\} \\ &= \tfrac{1}{2} + o,\end{aligned}$$

and $\bar{G}_p^{(0)}(p-2+0) = \bar{G}_p(p-2) > \bar{G}_p(m_p^{(0)}) = 1/2$. Therefore, by (2.19) and (2.20),

$$(2.21)\qquad \bar{G}_p^{(\lambda)}(p-2+\lambda) \geq 1/2, \quad \text{for every} \quad \lambda \geq 0.$$

Since $\bar{G}_p^{(\lambda)}(\frac{1}{2}c + \lambda) \geq \bar{G}_p^{(\lambda)}(p-2+\lambda)$, $\forall\ 0 < c \leq 2(p-2)$, the proof of (2.10) is complete.

Note that for the median $M_p^{(\lambda)}$ of $\chi^2_{p,\lambda}$, we have $\bar{G}_p^{(\lambda)}(M_p^{(\lambda)}) = 1/2$, for every $p \geq 1$, $\lambda \geq 0$. On the other hand, the median is sub-additive in λ, so that

$$(2.22)\qquad M_p^{(\lambda)} \leq M_p^{(0)} + \lambda, \ \forall\ \lambda \geq 0 \ \text{ and } \ p \geq 1.$$

Thus, instead of the upper bound $2(p-2)$ (for c), it may be possible to choose a sharper bound $2h(p)$, such that

$$(2.23)\qquad h(p)+\lambda \leq M_p^{(\lambda)}, \ \forall\ \lambda \geq 0,\ p \geq 1,$$

and (2.10) then holds for every $c \leq 2h(p)$. In particular, for $h(p) = p-1$, (2.23) holds, although for $h(p) \geq p-1 + .38$, it fails to do so. Thus, for $p = 2$, while there is no shrinkage according to the Stein rule [as $p-2 = 0$], it may still be possible to have some, by letting $0 < c < 2$. For $p \geq 3$, this refinement is of very little practical utility, as under (2.23), (2.10) will be so close to 1/2 that the improvement in the sense of Pitman-closeness will be hardly noticeable.

3. The case of unknown variance. As a natural generalization of the model in (2.4), we now assume that

$$(3.1)\qquad \hat{\theta} \sim N_p(\theta, \sigma^2 V),$$

where V is a known (p.d.) matrix and θ and σ^2 are both unknown. We also assume that there exist a nonnegative random variable S^2 and a positive integer m, such that

(3.2) mS^2/σ^2 has the central chi square distribution with

on DF, and $\hat{\underset{\sim}{\theta}}$ and S^2 are independent.

(3.1) and (3.2) hold typically for the usual (univariate) linear models. For (3.1), it is very natural to employ the usual Mahalanobis distance $\|\cdot\|_{\underset{\sim}{W}}$ which corresponds to $\underset{\sim}{W} = \underset{\sim}{V}^{-1}$. As such, we may, without any loss of generality, take $\underset{\sim}{V} = \underset{\sim}{I}$ and work with the usual Euclidean norm. The general case of an arbitrary $\underset{\sim}{W}$ will be considered in the next section. In this case, a Stein-rule estimator (for $p \geq 3$) is of the form

$$\hat{\underset{\sim}{\theta}}{}_c^S = \{1 - cS^2\|\hat{\underset{\sim}{\theta}}\|^{-2}\}\hat{\underset{\sim}{\theta}}\ , \tag{3.3}$$

where $c = c(p,m)$ satisfies the condition that

$$0 < c \leq 2(p-2)m/(m+2)\ . \tag{3.4}$$

For the usual dominance of $\hat{\underset{\sim}{\theta}}{}_c^S$ over $\hat{\underset{\sim}{\theta}}$, we may refer to Arnold (1981) and others.

<u>Theorem 3.1</u>. *For every* $p \geq 3$ *and* $0 < c \leq 2(p-2)m/(m+2)$, $\hat{\underset{\sim}{\theta}}{}^S$ *dominates* $\hat{\underset{\sim}{\theta}}$ *in the light of the Pitman closeness in* (2.3)

<u>Proof</u>. Parallel to (2.6), here, we have

$$\{\|\hat{\underset{\sim}{\theta}}{}^S-\underset{\sim}{\theta}\|^2 \leq \|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}\|^2\} \iff \{\|\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}\|^2 + \underset{\sim}{\theta}'(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}) \geq \tfrac{1}{2}\, cS^2\}. \tag{3.5}$$

We may set $\chi_m^2 = mS^2/\sigma^2$, so that defining $\chi_{p,\lambda}^2$ as in (2.9), we obtain that χ_m^2 and $\chi_{p,\lambda}^2$ are mutually independent; without any loss of generality, we set $\sigma^2 = 1$. Then, parallel to (2.9), our task is to verify that for every c satisfying (3.4),

$$(3.6) \qquad Q(c,\lambda;p,m) = P\{\chi^2_{p,\lambda} \geq (c/2m)\chi^2_m + \lambda\} \geq \frac{1}{2}, \ \forall\ \lambda \geq 0,$$

where the strict inequality holds for every finite λ (≥ 0). Note that

$$(3.7) \qquad Q(c,\lambda;p,m) = \int_0^\infty \bar{G}_p^{(\lambda)}(\lambda + cy/2m)\,dG_m(y).$$

First, consider the case of $\lambda = 0$. Then

$$(3.8) \quad Q(c,\lambda;p,m) = \int_{(c/2m)}^\infty \{\Gamma((p+m)/2)/\Gamma(p/2)\Gamma(m/2)\} y^{\frac{1}{2}p-1}(1+y)^{-\frac{1}{2}(p+m)}\,dy.$$

Again, using the unimodality of the density appearing in (3.8) [viz., Johnson and Kotz (1970), ch. 24] along with the related mean-median-mode inequality, we obtain that the mode is equal to (p-2)/(m+2)) and

$$(3.9) \qquad Q(c,0;p,m) > 1/2, \quad \text{for} \quad 0 < c \leq 2(p-2)m/(m+2).$$

Further, proceeding as in (2.18), we obtain that

$$(3.10) \quad (\partial/\partial\lambda)Q(c,\lambda;p,m) = \int_0^\infty g_{p+2}^{(\lambda)}(\lambda + cy/2m) - g_p^{(\lambda)}(\lambda + cy/2m)\}\,dG_m(y)$$

$$= -2\int_0^\infty \{(d/dx)\,g_{p+2}^{(\lambda)}(x)\Big|_{x=\lambda+cy/2m}\}\,dG_m(y)$$

[by (2.12)]

$$= c^{-1}4m\int_0^\infty \{(d/dy)\,g_{p+2}^{(\lambda)}(\lambda + cy/2m)\}\,d\bar{G}_m(y)$$

$$= \frac{4m}{c}\Big\{\Big[\bar{G}_m(y)\ \frac{d}{dy}\ g_{p+2}^{(\lambda)}(\lambda + \frac{cy}{2m})\Big]_0^\infty$$

$$- \int_0^\infty \bar{G}_m(y)\{\frac{d^2}{dy^2}\ g_{p+2}^{(\lambda)}(\lambda + \frac{cy}{2m})\}\,dy\Big\}$$

$$= -\frac{4m}{c}\{\frac{d}{dy}\ g_{p+2}^{(\lambda)}(\lambda + \frac{cy}{2m})\Big|_{y=0}\}$$

$$-\frac{4m}{c}\int_0^\infty \bar{G}_m(y)\ \frac{c^2}{4m^2}\{\frac{d^2}{dx^2}\ g_{p+2}^{(\lambda)}(x)\Big|_{x=\lambda+\frac{cy}{2m}}\}\,dy$$

$$= -2\left\{\frac{d}{dx}\, g_{p+2}^{(\lambda)}(x)\Big|_{x=\lambda}\right\} - \frac{c}{m}\int_0^\infty \bar{G}_m(y)\left\{\frac{d^2}{dx^2}\, g_{p+2}^{(\lambda)}(x)\Big|_{x=\lambda+cy/2m}\right\}dy.$$

Note that [viz., Johnson and Kotz (1970), Ch. 28] the noncentral density $g_{p+2}^{(\lambda)}(x)$ is (strongly) unimodal (with mode $m_{p+2}^{(\lambda)} \geq p-2+\lambda \geq \lambda$, $\forall\ p \geq 2$, by (2.15)), and hence,

(3.11) $$(d/dx)\, g_{p+2}^{(\lambda)}(x)\Big|_{x=\lambda} \geq 0,\ \forall\ \lambda \geq 0,\ p \geq 2,$$

(3.12) $$(d^2/dx^2)\, g_{p+2}^{(\lambda)}(x) = (d/dx)\left\{g_{p+2}^{(\lambda)}(d/dx)\ [\log g_{p+2}^{(\lambda)}(x)]\right\}$$

$$= g_{p+2}^{(\lambda)}(x)\left\{[(d/dx)\log g_{p+2}^{(\lambda)}(x)]^2 + (d^2/dx^2)\log g_{p+2}^{(\lambda)}(x)\right\}$$

$$\geq g_{p+2}^{(\lambda)}(x)\,(d/dx)\left\{(d/dx)\log g_{p+2}^{(\lambda)}(x)\right\}$$

$$\geq 0\ ,$$

as $(d/dx)\log g_{p+2}^{(\lambda)}(x)$ is monotone nondecreasing in x. Hence,

(3.13) $$(\partial/\partial\lambda)Q(c,\lambda;p,m) \leq 0,\ \forall\ \lambda \geq 0.$$

Finally, note that

(3.14) $$Q(c,\lambda;p,m) = P\left\{\frac{\chi^2_{p,\lambda}-p-\lambda}{\sqrt{2(p+2\lambda)}} \geq \frac{1}{\sqrt{2(p+2\lambda)}}\ [\tfrac{c}{2m}(\chi^2_m-m)-(p-\tfrac{c}{2})]\right\}$$

$$\to 1/2 \quad \text{as} \quad \lambda \to +\infty\ ,$$

where the last step follows from the fact that $p > c/2$, $m^{-1}\chi^2_m = O_p(1)$, and as $\lambda \to \infty$, $2(p+2\lambda) \to +\infty$, while the left hand side $(\chi^2_{p,\lambda} - p - \lambda)/\sqrt{2(p+2\lambda)}$ is asymptotically normal with 0 mean and unit variance. Thus, by (3.9), (3.13) and (3.14), we obtain that

(3.15) $$Q(c,\lambda;p,m) \geq \frac{1}{2}\ ,\ \forall\ \lambda \geq 0.$$

Thus (3.6) holds and the proof of the theorem is complete.

Note that here also, the upper bound $2(p-2)m/(m+2)$ (for c) may be replaced by $2h(p)m/(m+2)$, for $h(p) \leq p-1$, and this will enable the shrinkage to be effective, even for $p = 2$.

4. The case of an arbitrary covariance matrix. We consider the model:

$$\hat{\underset{\sim}{\theta}} \sim N_p(\underset{\sim}{\theta}, \underset{\sim}{\Sigma}) , \tag{4.1}$$

where $\underset{\sim}{\Sigma}$ is a p.d. matrix (unknown) and assume that there exists a p.s.d. stochastic matrix $\underset{\sim}{S}$ (independent of $\hat{\underset{\sim}{\theta}}$), such that

$$\underset{\sim}{S} \text{ has the Wishart distribution } W_p(m, \underset{\sim}{\Sigma}) , \tag{4.2}$$

for some positive integer m $(> p)$. With the quadratic loss function in (2.2), shrinkage estimators of $\underset{\sim}{\theta}$, considered by Berger et al. (1977), are of the form:

$$\hat{\underset{\sim}{\theta}}{}^S_c = [\underset{\sim}{I} - cdT^{-2}\underset{\sim}{W}^{-1}\underset{\sim}{S}^{-1}]\hat{\underset{\sim}{\theta}} , \tag{4.3}$$

where $p \geq 3$ and

$$T^2 = m\hat{\underset{\sim}{\theta}}'\underset{\sim}{S}^{-1}\hat{\underset{\sim}{\theta}}, \quad d = ch_{\min}(\underset{\sim}{W}\underset{\sim}{S}) , \tag{4.4}$$

$$0 < c \leq c(m,p);\ c(m,p) \text{ is } \uparrow \text{ in } m,\ \lim_{m\to\infty} c(m,p) = 2(p-2). \tag{4.5}$$

The dominance of $\hat{\underset{\sim}{\theta}}{}^S_c$ (over $\hat{\underset{\sim}{\theta}}$) and its minimax character were studied by Berger et al. (1977). We have the following

Theorem 4.1. *For every* $p \geq 3$ *and* $0 < c \leq 2(p-2)m/(m-p+3)$, $\hat{\underset{\sim}{\theta}}{}^S_c$ *dominates* $\hat{\underset{\sim}{\theta}}$ *in the light of Pitman closeness (and the norm* $\|\cdot\|_{\underset{\sim}{W}}$*).*

Proof. Note that

$$(4.6)\qquad \|\hat{\theta}^S-\theta\|_W^2 = \|\hat{\theta}-\theta\|_W^2 - 2cdT^{-2}(\hat{\theta}-\theta)'WW^{-1}S^{-1}\hat{\theta} + c^2d^2T^{-4}\hat{\theta}'S^{-1}W^{-1}WW^{-1}S^{-1}\hat{\theta}$$
$$= \|\hat{\theta}-\theta\|_W^2 - 2cdT^{-2}\{(\hat{\theta}-\theta)'S^{-1}\hat{\theta}\} + c^2d^2T^{-4}\hat{\theta}'S^{-1}W^{-1}S^{-1}\hat{\theta},$$

where by the Courant theorem

$$(4.7)\qquad (\hat{\theta}'S^{-1}W^{-1}S^{-1}\hat{\theta})/(\hat{\theta}'S^{-1}\hat{\theta}) \le ch_{max}(S^{-1}W^{-1}) = d^{-1},$$

for every $\hat{\theta} \in R^p$. Thus, to verify (2.3), it suffices to show that

$$(4.8)\qquad P_\theta\{(\hat{\theta}-\theta)'S^{-1}\hat{\theta} \ge c/2m\} \ge 1/2, \quad \forall\ \theta \in R^p .$$

First, consider the case of $\theta = 0$. Note that the left hand side of (4.8) is then

$$(4.9)\qquad P\{\hat{\theta}'S^{-1}\hat{\theta} \ge c/2m | \theta = 0\} = P\{T^2 \ge \tfrac{1}{2}c | \theta = 0\},$$

where using the spherical transformation, we readily obtain that

$$(4.10)\qquad m^{-1}T^2 \stackrel{D}{=} \chi_p^2/\chi_{m-p+1}^2 \quad (\text{under } \theta = 0)$$

[viz., Anderson (1984, p.162)], so that parallel to (3.9), we obtain that under $\theta = 0$,

$$(4.11)\qquad P_0\{(\hat{\theta}-\theta)'S^{-1}\hat{\theta} \ge c/2m\} > 1/2, \ \forall\ 0 < c \le 2(p-2)m/(m-p+3).$$

For the general case of $\theta \ne 0$, we write

$$(4.12)\qquad (\hat{\theta}-\theta)'S^{-1}\hat{\theta} = (\hat{\theta}-\theta)'S^{-1}(\hat{\theta}-\theta) + \theta'S^{-1}(\hat{\theta}-\theta)$$
$$= (\hat{\theta}-\theta)\Sigma^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}S^{-1}\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}(\hat{\theta}-\theta) + \theta'\Sigma^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}S^{-1}\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}(\hat{\theta}-\theta)$$
$$= Y'S^{*-1}Y + \eta'S^{*-1}Y ,$$

where $Y \sim N_p(0,I)$ and $mS^* \sim W_p(m,I)$. Further, we let

$$(4.13)\qquad \underset{\sim}{\eta} \to \underset{\sim}{\gamma} = \underset{\sim}{A}\underset{\sim}{\eta} \quad \text{where} \quad \underset{\sim}{A}' = (\|\underset{\sim}{\eta}\|^{-1}\underset{\sim}{\eta},\ \underset{\sim}{A}_2'),\ \underset{\sim}{A}_2' \perp \underset{\sim}{\eta}' ,$$

so that $\underset{\sim}{A}$ is an orthogonal matrix and

$$(4.14)\qquad \gamma_1 = \|\eta\| = (\underset{\sim}{\theta}'\Sigma^{-1}\underset{\sim}{\theta})^{\frac{1}{2}} \quad \text{and} \quad \underset{\sim}{\gamma}_2 = \underset{\sim}{A}_2\underset{\sim}{\eta} = \underset{\sim}{0} .$$

Also, we let $\underset{\sim}{Z} = \underset{\sim}{A}\underset{\sim}{Y}$ and use the spherical transformation on $\underset{\sim}{Z}$. Then (4.10) reduces to

$$(4.15)\qquad (\|\underset{\sim}{U}\|^2 + \gamma_1 U_1)V'';\quad \|U\|^2 \sim \chi_p^2,\ \frac{1}{V''} \sim \chi^2_{m-p+1} .$$

Thus, writing

$$(4.16)\qquad \|U\|^2 + \gamma_1 U_1 = \sum_2^p U_i^2 + (U_1+\gamma_1/2)^2 - \frac{1}{4}\gamma_1^2$$

$$= \chi^2_{p,\lambda} - \lambda,\ \text{where}\quad \lambda = \frac{1}{4}\gamma_1^2 = \frac{1}{4}(\underset{\sim}{\theta}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\theta}) ,$$

we obtain that (4.8) reduces to

$$(4.17)\qquad P_{\underset{\sim}{\theta}}\{(\chi^2_{p,\lambda}-\lambda)/\chi^2_{m-p+1} \geq c/2m\} = P_{\underset{\sim}{\theta}}\{\chi^2_{p,\lambda} \geq \lambda+(c/2m)\chi^2_{m-p+1}\},$$

so that the proof of theorem 3.1 can readily be adapted to verify that (4.17) is $\geq 1/2$ for every $\underset{\sim}{\theta} \in R^p$, whenever $0 < c < 2(p-2)m/(m-p+3)$, and this completes the proof of the theorem.

Comparing the range of c in Theorem 4.1 with that in Berger et al. (1977), we observe that we may choose a larger shrinkage factor in our case. Also, as in Section 3, the result applies to the case of $p = 2$ with $h(p) \leq p-1$.

5. <u>Asymptotic theory</u>. For underlying distributions not necessarily (multi-) normal, robust and nonparametric shrinkage estimation theory has been developed by the current authors in a variety of situations [viz., Sen (1984), Sen and Saleh (1985, 1987) and Saleh and Sen (1985,

1986), among others]. Also, for the MLE, the asymptotic theory of Stein-rule estimation has been studied by Sen (1986b). An asymptotic treatment of Pitman closeness of usual BAN estimators has been considered by Sen (1986a). The current study provides a natural extension of the latter to Stein-rule BAN estimators.

An essential feature of the asymptotic theory of shrinkage estimation, studied by these authors, is the proper identification of a (Pitman-) neighborhood of the pivot where shrinkage is effective; beyond this $O(n^{-\frac{1}{2}})$ neighborhood, the shrinkage version and the usual unrestricted version become asymptotically risk-equivalent. Further, the computation of the exact risk (or its asymptotic value) of a shrinkage estimator for non-normal distributions (or non-linear estimators) may become prohibitively laborious, and may also demand more stringent regularity conditions (to justify the existence of the limits). This difficulty can largely be avoided by incorporating the concept of asymptotic distributional risk (ADR), and this has also been effectively incorporated in the earlier studies. Dominance in the light of this ADR criterion is largely a distributional property, and, generally, may not require the verification of intricate moment-convergence results needed for the study of the (asymptotic or exact) risk. Study of the asymptotic dominance in the light of Pitman closeness (for Pitman-alternatives) also involves only the asymptotic distributional results, and hence, may not require regularity conditions more stringent than those pertaining to the ADR dominance. Besides, as in the earlier sections, the Pitman-closeness criterion may require less restrictive conditions on the shrinkage factor.

We consider the following model. Let $\{\hat{\theta}_n;\ n \geq n_0\}$ be a sequence of estimators admitting an asymptotic linear representation as in Sen (1986a), such that asymptotically

(5.1) $$n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \sim N_p(0, \Sigma) ,$$

where Σ is an unknown p.d. matrix (and may even be dependent on θ). Let $\{S_n;\ n \geq n_0\}$ be a sequence of stochastic matrices, such that

(5.2) $$S_n \to \Sigma \text{ , in probability, as } n \to \infty.$$

Further, corresponding to the null pivot for θ , consider a sequence $\{K_n\}$ of local alternatives

(5.3) $$K_n : \theta = \theta_{(n)} = n^{-\frac{1}{2}}\lambda,\ \lambda \text{ (fixed) } \in R^p.$$

Finally, consider a sequence $\{T_n^2;\ n \geq n_0\}$ of test statistics

(5.4) $$T_n^2 = n\hat{\theta}_n' S_n^{-1} \hat{\theta}_n \ , \quad n \geq n_0,$$

so that under $\{K_n\}$, T_n^2 has asymptotically the noncentral chi-square distribution with p DF and noncentrality parameter $\Delta = \lambda'\Sigma^{-1}\lambda$. [Alternative forms for T_n^2 are available for the likelihood ratio test or some other rank order tests, but these are asymptotically equivalent (under $\{K_n\}$) to (5.4).] Then, typically a Stein-rule version of $\hat{\theta}_n$ is given by

(5.5) $$\hat{\theta}_n^S = [I - cd_n T_n^{-2} W^{-1} S_n^{-1}]\hat{\theta}_n \ ,$$

where W is a given p.d. matrix and $d_n = ch_{min}(WS_n)$, $n \geq n_0$. In terms of the ADR of $n^{\frac{1}{2}}(\hat{\theta}_n^S - \theta)$ [under K_n], the dominance results have been studied earlier by the authors. It follows that the proof of Theorem 4.1

can readily be incorporated (with the further simplification that by (5.2) $\underset{\sim}{S}_n$ can be replaced by $\underset{\sim}{\Sigma}$) to show that for every $0 < c < 2(p-2)$, $p \geq 3$, under $\{K_n\}$ in (5.3),

$$\text{(5.6)} \quad \lim_{n\to\infty} P\{\sqrt{n}\|\hat{\underset{\sim}{\theta}}_n^S - \underset{\sim}{\theta}_{(n)}\|_{\underset{\sim}{W}} \leq \sqrt{n}\|\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_{(n)}\|_W \mid K_n\} \geq 1/2 ,$$

so that in the light of the (asymptotic) Pitman-closeness (for local alternatives), the Stein-rule version $\{\hat{\underset{\sim}{\theta}}_n^S\}$ dominates the usual version $\{\hat{\underset{\sim}{\theta}}_n\}$. This asymptotic dominance result applies to all the shrinkage estimators considered in the literature for which (5.1)-(5.2) hold. In particular, dealing with U-statistics [as in Sen (1984)], we observe that both (5.1) and (5.2) hold (for the latter, we use the jackknifed variance covariance estimators), so that identifying the sample covariance matrix as a special case of U-statistics, we see that our solution also applies to the case of shrinkage estimators of co-variance matrices for possibly non-normal distributions, and this extends the results of Sugiura (1984) to a wider class of distributions. This also provides a general answer to the open question of Rao (1981) in the asymptotic case, while answers for the finite sample case have been provided in the earlier sections for normal distributions. The counter examples of Rao (1981) correspond to a non-normal case for the variance estimators (for finite sample cases), but in the asymptotic case, for Pitman alternatives, they would satisfy (5.1) and (5.2), and hence, the Pitman-closeness holds.

REFERENCES

ANDERSON, T.W. (1984), An Introduction to Multivariate Statistical Analysis. 2nd Ed. John Wiley, New York.

ARNOLD, S.F. (1981). The Theory of Linear Models and Multivariate Analysis. John Wiley, New York.

BERGER, J.O. (1985). Statistical Decision Theory, 2nd Ed. Springer-Verlag, New York.

BERGER, J.O., BOCK, M.E., BROWN, L.D., CASELLA, G. and GLESER, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. Ann. Statist. 5, 763-771.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Prob. Univ. Calif. Press, 1, 361-379.

JOHNSON, N.L. and KOTZ, S. (1970). Distributions in Statistics: Continuous Univariate Distributions, I, II. John Wiley, New York.

JUDGE, G.G. and BOCK, M.E. (1978). Statistical Implications of Pre-test and Stein Rule Estimator in Econometrics. North Holland, Amsterdam.

PITMAN, E.J.G. (1937). The 'closest' estimates of statistical parameters. Proc. Cambridge Phil. Soc. 33, 212-217.

RAO, C.R. (1981). Some comments on the minimum mean square error as a criterion of estimation. Statistics and Related Topics (Eds. M. Csörgö, D.A. Dawson, J.N.K. Rao and A.K.M.E. Saleh), North Holland, Amsterdam, pp. 123-143.

SALEH, A.K.M.E. and SEN, P.K. (1985). On shrinkage M-estimators of location parameters. Commun. Statist. Theor. Meth. 24, 2313-2329.

SALEH, A.K.M.E. and SEN, P.K. (1986). On shrinkage R-estimation in a multiple regression model. Commun. Statist. Theory. Meth. 15, 2229-2244.

SEN, P.K. (1984). A James-Stein detour of U-statistics. Commun. Statist. Theor. Meth. A13, 2725-2747.

SEN, P.K. (1986a). Are BAN estimators the Pitman closest ones too? Sankhyā, Ser. A 48, 51-58.

SEN, P.K. (1986b). On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators. Sankhyā, Ser A 48, 354-372.

SEN, P.K. and SALEH, A.K.M.E. (1985). On some shrinkage estimators of multivariate location. Ann. Statist. 13, 272-281.

SEN, P.K. and SALEH, A.K.M.E. (1987). On preliminary test and shrinkage M-estimation in linear models. Ann. Statist.

STEIN, C. (1956). Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. Univ. Calif. Press, 1, 197-206.

SUGIURA, N. (1984). Asymptotically closer estimators for the normal covariance matrix. Jour. Japan Statist. Assoc. 14, 145-155.