QÜESTIIÓ, Vol 19, 1, 2, 3 pp. 65-91, 1995

A REVIEW OF THE RESULTS ON THE STEIN APPROACH FOR ESTIMATORS IMPROVEMENT

V.G. VOINOV* and M.S. NIKULIN†

Since 1956, a large number of papers have been devoted to Stein's technique of obtaining improved estimators of parameters, for several statistical models. We give a brief review of these papers, emphasizing those aspects which are interesting from the point of view of the theory of unbiased estimation.

Key words: Admissible Estimator, Berger's Estimator, Equivariant Estimator, James-Stein Estimator, Minimax Estimator, Quadratic Loss Function, Shrinkage Estimator.

1. INTRODUCTION

Let we need not an unbiased estimator but an estimator with the smallest risk for a prescribed loss function. Much help in the situation gives the well known Stein phenomenon discovered by Charles Stein (1956). He was the first showed that the best unbiased estimator for the mean of a multivariate normal distribution can be improved upon by so-called shrinkage estimators. Since that time, a large number of papers have been published, proposing classes of improved estimators for different probability models (see e.g. a monograph by Hoffmann (1992)). We shall give a brief account of these works emphasizing those questions which are related to the unbiased estimation point of view.

* Voinov, V.G. Physical Technical Institute, Alma-Ata, Kazakhstan.

† Nikulin, M.S. Mathématiques Stochastiques, Université Bordeaux 2, France.
Steklov Mathematical Institute, St Petersbourg, Russie.

– Article rebut el juny de 1994.

– Acceptat el setembre de 1994.

2. ESTIMATING THE MEAN VECTOR

Let us start with the problem of estimating the mean vector $\mu = (\mu_1, \cdots, \mu_p)^T$ of a p-dimensional normal distribution with known covariance matrix equal to identity, **I**, under the loss function

$$L(\mu, \hat{\mu}) = (\hat{\mu} - \mu)^T \mathbf{A} (\hat{\mu} - \mu), \tag{1}$$

where **A** is a positive-semidefinite matrix, from a single observation $\mathbf{X} = (X_1, X_2, \cdots, X_p)^T$.

It is known that the observation **X** itself is a maximum likelihood, minimum variance unbiased, invariant for a wide class of loss functions and minimax estimator for μ. Nevertheless, this estimator is inadmissible in the sense that there exist other estimators whose risks are everywhere smaller than the risk of **X**. To consider such estimators $\hat{\mu}(\mathbf{X})$ let us follow for a while the Stein's lectures (1977, 1986). Without loss of generality the matrix **A** may be considered to be a diagonal matrix $\mathbf{A} =$ diag $(\alpha_1, \cdots, \alpha_p)$, where

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_p \geq 0.$$

Estimators $\hat{\mu}(\mathbf{X})$ may be constructed with the help of the following identity

$$\mathbf{E}_\mu \sum_{i=1}^{p} \alpha_i [X_i + g_i(\mathbf{X}) - \mu_i]^2 = \mathbf{E}_\mu \sum_{i=1}^{p} \alpha_i [1 + g_i^2(\mathbf{X}) + 2 g_{ii}(\mathbf{X})], \tag{2}$$

where $\mathbf{g} = (g_1, \cdots, g_p)^T$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^n$,

$$g_{ii}(\mathbf{X}) = \frac{\partial}{\partial X_i} g_i(\mathbf{X}), \quad g_i : \mathbb{R}^n \to \mathbb{R}$$

and

$$\mathbf{E}_\mu |g_{ii}(\mathbf{X})| < \infty, \quad \mathbf{E}_\mu g_i^2(\mathbf{X}) < \infty.$$

A remarkable feature of equality (2) is that the expectand on the right hand side of it does not involve the unknown parameter μ.

The proof of identity (2) is based on

Lemma 1

Let Y be a standard normally distributed random variable and $g : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous on $\mathbb{R}$ function such that $\mathbf{E}|g'(Y)| < \infty$. Then

$$\mathbf{E} g'(Y) = \mathbf{E} Y g(Y). \tag{3}$$

Proof

$$
\begin{aligned}
\mathbf{E}g'(Y) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g'(y)e^{-\frac{y^2}{2}}dy = \\
&= \frac{1}{\sqrt{2\pi}}\left\{\int_0^{\infty} g'(y)dy\int_y^{\infty} xe^{-\frac{x^2}{2}}dx - \int_{-\infty}^{0} g'(y)dy\int_{-\infty}^{y} xe^{-\frac{x^2}{2}}dx\right\} = \\
&= \frac{1}{\sqrt{2\pi}}\left\{\int_0^{\infty} xe^{-\frac{x^2}{2}}dx\int_0^{x} g'(y)dy - \int_{-\infty}^{0} xe^{-\frac{x^2}{2}}dx\int_x^{0} g'(y)dy\right\} = \\
&= \frac{1}{\sqrt{2\pi}}\left\{\int_0^{\infty} (g(x)-g(0))xe^{-\frac{x^2}{2}}dx + \int_{-\infty}^{0} (g(x))-g(0))xe^{-\frac{x^2}{2}}dx\right\} = \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} xg(x)e^{-\frac{x^2}{2}}dx = EYg(Y).
\end{aligned}
$$

To prove (2) we have to prove that

$$\mathbf{E}_\mu[X_1 + g_1(\mathbf{X}) - \mu_1]^2 = 1 + \mathbf{E}_\mu[g_1^2(\mathbf{X}) + 2g_{11}(\mathbf{X})].$$

Assuming the function $g_1 : \mathbb{R}^n \to \mathbb{R}$ to be absolutely continuous by x_1 and

$$\mathbf{E}_\mu|g_{11}(\mathbf{X})\ | < \infty,$$

then by Lemma 1

$$\mathbf{E}_\mu g_{11}(\mathbf{X}) = \mathbf{E}_\mu(X_1 - \mu_1)g_1(\mathbf{X})$$

and

$$\mathbf{E}_\mu[X_1 + g_1(\mathbf{X}) - \mu_1]^2 = 1 + \mathbf{E}_\mu[g_1^2(\mathbf{X}) + 2g_{11}(\mathbf{X})]. \tag{4}$$

It follows from (2) that if

$$\sum_{i=1}^{p} \alpha_i[g_i^2(\mathbf{X}) + 2g_{ii}(\mathbf{X})] < 0, \tag{5}$$

then

$$\mathbf{E}_\mu \sum_{i=1}^{p} \alpha_i[X_i + g_i(\mathbf{X}) - \mu_i]^2 < \sum_{i=1}^{p} \alpha_i \le \mathbf{E}_\mu \sum_{i=1}^{p} \alpha_i(X_i - \mu_i)^2.$$

Since inequality (5) does not depend on the unknown parameter μ, the estimator

$$\hat{\mu} = \mathbf{X} + \mathbf{g}(\mathbf{X})$$

is everywhere in the parameter space better (in the sence of the smaller risk) than the usual estimator

$$\hat{\mu}^0 = \mathbf{X}.$$

Thus, the problem reduces to the appropriate choice of the function g satisfying differential inequality (5).

The first result in this direction has been obtained by James and Stein (1960).

Denoting

$$\mathbf{x}^T\mathbf{y} = \sum_{i=0}^{p} x_i y_i, \qquad |\mathbf{x}|^2 = \mathbf{x}^T\mathbf{x}$$

and

$$\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_p}\right)^T$$

for the special case $\mathbf{A} = \mathbf{I}$ we may rewrite identity (2) as follows

(6) $$\mathbf{E}_\mu|\mathbf{X} + \mathbf{g}(\mathbf{X}) - \mu|^2 = p + \mathbf{E}_\mu[|\mathbf{g}(\mathbf{X})| + 2\nabla^T\mathbf{g}(\mathbf{X})],$$

where

$$\nabla^T\mathbf{g}(\mathbf{X}) = \sum_{i=1}^{p} \frac{\partial g_i(\mathbf{X})}{\partial X_i}.$$

Consider the case when $\hat{\mu}$ is equivariant with respect to the group of orthogonal transformations of $\mathbb{R}^n$, i.e. there exists a function $h : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathbf{g}(\mathbf{X}) = h(|\mathbf{X}|^2)\mathbf{X}.$$

In this case (6) gives

(7) $$\begin{aligned} \mathbf{E}_\mu|\mathbf{X} + \mathbf{g}(\mathbf{X}) - \mu|^2 &= p + \mathbf{E}_\mu[h^2(|\mathbf{X}|^2)|\mathbf{X}|^2 + \\ &+ 2ph(|\mathbf{X}|^2) + 4|\mathbf{X}|^2 h'(|\mathbf{X}|^2)]. \end{aligned}$$

Let

$$h(t) = -\frac{c}{t},$$

where $c > 0$, then for $p \geq 3$

(8) $$\mathbf{E}_\mu\left|\left(1 - \frac{c}{|\mathbf{X}|^2}\right)\mathbf{X} - \mu\right|^2 = p + \mathbf{E}_\mu \frac{c^2 - 2(p-2)c}{|\mathbf{X}|^2}.$$

The right hand side of (8) is minimized by

$$c = p - 2$$

and the minimum value is

$$\mathbf{E}_\mu\left|\left(1-\frac{c}{|\mathbf{X}|^2}\right)\mathbf{X}-\mu\right|^2 = p-(p-2)^2\mathbf{E}_\mu\frac{1}{|\mathbf{X}|^2}.$$

This will be less than the constant risk p of the estimator $\mathbf{X}$ provided that $p>2$. Thus the James and Stein estimator reads

$$\hat{\mu} = \left(1-\frac{p-2}{|\mathbf{X}|^2}\right)\mathbf{X}. \tag{9}$$

James and Stein showed that the risk of $\hat{\mu}$ at $\hat{\mu}=0$ under the loss function

$$L(\hat{\mu},\mu) = |\hat{\mu}-\mu|^2$$

is equal to 2. For $p>2$ this is a substantial improvement over the constant risk p of the estimator $\hat{\mu}^0$. Even for $|\mu|^2/p = 2$ the estimator (9) gives the risk about 25 per cent lower than that of $\hat{\mu}^0 = \mathbf{X}$ at $\mu = 0$. Nevertheless, statisticians and users do not "rush to embrace this considerable improvement ... " as Efron (1975) said. He considered several reasons for this fact. The most interesting reason in our opinion is the following one. "If the different μ_i refer to obviously disjoint problems (e.g., μ_1 is the speed of light, μ_2 is the price of tea in China, μ_3 is the efficacy of new treatment for psoriasis, etc.) combining the data can produce a definitely uncomfortable feeling in the statistician". The situation seems to be paradoxical, but "if we really aren't interested in $\mu_2,\mu_3,\cdots,\mu_p$, just μ_1, then

$$L_1(\hat{\mu},\mu) = (\hat{\mu}_1-\mu_1)^2$$

seems to be a more reasonable loss function than $|\hat{\mu}-\mu|^2$. It is not true that

$$\mathbf{E}_\mu(\hat{\mu}_1-\mu_1)^2 < \mathbf{E}_\mu(X_1-\mu_1)^2$$

for all μ. As a matter of fact

$$\mathbf{E}_\mu(\hat{\mu}_1-\mu_1)^2/\mathbf{E}_\mu(X_1-\mu_1)^2$$

can be as large as about $\sqrt{p}$ for certain configurations of $\mu_2,\mu_3,\cdots,\mu_p$ (namely $\mu_1=\sqrt{p},\mu_2=\cdots=\mu_p=0$)". The paradox seems to be disappeared. Since it is not obvious that the loss function $(\hat{\mu_1}-\mu_1)^2$ is the best one for the situation, the "uncomfortable feeling" remains. That is why many attempts to resolve the paradox and to reduce the maximum component risk

$$R^*(\mu) = \max_i \mathbf{E}_\mu(\hat{\mu}_i-\mu_i)^2$$

have been undertaken.

3. THE BAYESIAN APPROACH

The best heuristic explanation of the paradox contains a Bayesian argument. If the μ_i are a priori independent $N(0,\tau^2)$, then (9) can be considered as an empirical Bayes estimator (Efron and Morris (1973)). More details about the relation of the Stein's effect to the empirical Bayes approach and related problems readers may find in Efron and Morris (1976b), Berger and Srinivasan (1977), Haff (1980), Berger (1982a), Chen (1988) and others. Another explanation interpretes (9) as a smoothed version of a weighted mean of 0 and $\hat{\mu}^0$, i.e. one uses $\hat{\mu} = 0$ or $\hat{\mu}_i = X_i$ depending on the outcome of the test verifying the null hypothesis that $\mu = 0$ (Lehmann (1983)). An interesting approach to the problem gave Stigler (1990) who had considered Stein estimation as a regression problem. A simple geometrical argument of the possibility to improve upon the best invariant estimator via shrinkage estimation have been given by Brandwein and Strawderman (1990). It is worth to note the work of Beran (1992) who showed that under quadratic loss the Stein and the positive-part Stein estimators of μ are both approximately admissible and approximately minimax on large compact balls about the shrinkage point as $p \to \infty$.

4. MINIMIZING THE RISK

To this end we would also like to say that actually one should choose such an estimator which corresponds a goal he pursues. This depends also on what one intends to do with an estimator. If he needs an unique point estimator it is preferable to use one which minimizes the risk. If on the contrary he will perform averaging of many estimators the unbiased one might be the best (see §1, Ch.3 in Voinov and Nikulin (1993)).

Let $\mathbf{X} \sim N(\hat{\mu}, \sigma^2 \mathbf{I})$, σ^2 known. In this case

(10) $$\hat{\mu} = \left(1 - \frac{(p-2)\sigma^2}{|\mathbf{X}|^2}\right)\mathbf{X}.$$

If σ^2 is unknown and one possesses an estimator S_n of σ^2 independent of $\mathbf{X}$ and distributed like $\sigma^2\chi^2_n$, then (James & Stein (1960))

(11) $$\hat{\mu} = \left(1 - \frac{(p-2)S_n}{(n+2)|\mathbf{X}|^2}\right)\mathbf{X}.$$

In the most realistic situation $\mathbf{X} \sim N(\mu, \Sigma)$, Σ being unknown. Let there is an independent estimator $\mathbf{S}$ for Σ, which is distributed as $W_{n-1}(\mathbf{s}; \Sigma)$, a Wishart distribution.

James & Stein proposed the estimator

$$\hat{\mu} = \left(1 - \frac{(p-2)}{(n-p+3)\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}}\right)\mathbf{X}. \tag{12}$$

An explicit formula for the risk of the James and Stein estimator is available in Egerton and Laycock (1982).
Baranchik (1964, 1970) showed that for $\mathbf{X} \sim N(\mu, \mathbf{I})$ the estimator

$$\hat{\mu} = \left(1 - \frac{r(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\right)\mathbf{X} \tag{13}$$

under the loss function

$$L(\hat{\mu}, \mu) = |\hat{\mu} - \mu|^2$$

dominates $\hat{\mu}^0$ if

$$0 < r(|\mathbf{X}|^2) < 2(p-2)$$

and $r(.)$ is a nondecreasing function.

A more general minimax condition for $r(.)$ has been given by Efron & Morris (1976a). Let $\mathbf{X}$ have a p-variate normal distribution $N(\mu, \mathbf{D})$, where μ is unknown and $\mathbf{D}$ is a known nonsingular covariance matrix. Let d_L be the largest eigenvalue of $\mathbf{D}$. Bock (1975) has shown that for $p > 2$

$$\hat{\mu} = \left(1 - \frac{cr(\mathbf{X}^T\mathbf{D}^{-1}\mathbf{X})}{\mathbf{X}^T\mathbf{D}^{-1}\mathbf{X}}\right)\mathbf{X} \tag{14}$$

is a minimax spherically symmetric estimator for μ if $0 < c \leq 2((tr\mathbf{D})d_L^{-1} - 2)$ and $r(.)$ is a monotone non-decreasing function. Some generalizations of the James and Stein estimator (9) have been summarized by Akai (1988). Considering estimators

$$\hat{\mu} = \left(1 - \frac{b}{|\mathbf{X}|^2}\right)\mathbf{X}, \tag{15}$$

where $0 < b < 2(p-2)$, Akai (1988) showed that the estimator

$$\hat{\mu} = \left(1 - \frac{b}{|\mathbf{X}|^2} + \frac{g(|\mathbf{X}|)}{|\mathbf{X}|^{2r}}\right)\mathbf{X} \tag{16}$$

has under the same loss function as above the smaller risk than that of (15), if r is an integer such that

$$r > 1 \text{ and } p > 2(2r-1) \geq b$$

and $g(s)$ satisfies either condition (i) or (ii):

(i) $g(s)s^{-1/\alpha}$ is nonincreasing for $\alpha > 0$, and $g(s)$ is nondecreasing and satisfies

$$0 < g(s) \leq 2^r(b-(p-2r+2/\alpha))\Gamma(\frac{p}{2}-r)/\Gamma(\frac{p}{2}-2r+1)$$

for $\alpha > 0$ and $b > p-2r+2/\alpha$,

(ii) $g(s)$ is nonincreasing and satisfies

$$0 > g(s) > 2^r/(b-(p-2r))\Gamma(\frac{p}{2}-r)/\Gamma(\frac{p}{2}-2r+1)$$

for $b < p-2r$.

Let, for example, $g(s) = d$, where d is a constant, and $b = p-2$. Then the risk of the estimator

$$\hat{\mu} = \left(1 - \frac{p-2}{|\mathbf{X}|^2} + \frac{d}{|\mathbf{X}|^{2r}}\right)\mathbf{X} \tag{17}$$

for

$$L(\hat{\mu},\mu) = |\hat{\mu}-\mu|^2$$

is uniformly smaller than that of $\hat{\mu}^0$ if

$$1 < r < (p+2)/4$$

and

$$0 < d \leq 2^{r+1}(r-1)\Gamma(\frac{p}{2}-r)/\Gamma(\frac{p}{2}-2r+1).$$

Consider the special form

$$\hat{\mu} = \left(1 - \frac{b}{a+|\mathbf{X}|^2}\right)\mathbf{X}, \qquad a > 0, \quad 0 < b < 2(p-2), \tag{18}$$

of Baranchik's estimator (13). Let $p > 6$ and $b = p-2$, then (Akai (1988)) the risk of (18) under the squared error loss function is uniformly smaller than that of $\hat{\mu}^0$ if

$$0 < a < 4(p-6)/(p-2).$$

Akai also showed that for $p > 6$ the risk of

$$\hat{\mu} = \left(1 - \frac{b}{a+|\mathbf{X}|^2} + \frac{c}{|\mathbf{X}|^4}\right)\mathbf{X} \tag{19}$$

for

$$L(\hat{\mu},\mu) = |\hat{\mu}-\mu|^2$$

is uniformly smaller than that of (18) if the constants a and c satisfy the conditions:

(i) for $b > p-4$,

$$0 \leq a < \frac{[b-(p-4)](p-6)}{p-4}$$

and

$$0 < c \leq 2\frac{[b-(p-4)](p-6)^2 - a(p-4)(p-6)}{a+p};$$

(ii) for $b < p-4$

$$a \geq 0 \text{ and } 0 > c \geq \frac{[b-(p-4)](p-6)^2 - a(p-4)(p-6)}{a+p}.$$

Suppose now that $p > 6$. Then the risk of

$$\hat{\mu} = \left(1 - \frac{b}{a+|\mathbf{X}|^2} + \frac{h(|\mathbf{X}|^2)}{(d_1+|\mathbf{X}|^2)^2}\right)\mathbf{X} \tag{20}$$

for

$$L(\hat{\mu},\mu) = |\hat{\mu}-\mu|^2$$

is uniformly smaller than that of (18) if $h(s)$, a and d_1 satisfy either condition (i) or (ii):

(i) $h(s)(d_1+s)^{-1/\alpha}$ is nonincreasing for $\alpha > 0$ and $d_1 > 0$ and $h(s)$ is nondecreasing and satisfies

$$0 < h(s) \leq 2\{(b-(p-4+2/\alpha))(p-6) - d_1 p\} \quad \text{for } \alpha > 0,$$

$$b > p-4+2/\alpha \quad \text{and} \quad a \leq d_1 < (b-(p-4+2/\alpha))(p-6)/p,$$

(ii) $h(s)$ is nonincreasing and satisfies

$$0 > h(s) \geq 2\{(b-(p-4))(p-6) - d_1 p\} \quad \text{for } a \geq d_1 \geq 0 \quad \text{and } b < p-4.$$

The estimator (9) modifies the usual estimator $\hat{\mu}^0$ by "shrinking" it towards the origin, the more so the closer $\mathbf{X}$ is itself to the origin. For $|\mathbf{X}|^2 < (p-2)$ the estimator $\hat{\mu}$ is actually passed the origin in the direction opposite $\mathbf{X}$. If such a behaviour is undesirable one may introduce the positive-part James & Stein estimator

$$\hat{\mu} = \left(1 - \frac{p-2}{|\mathbf{X}|^2}\right)_+ \mathbf{X}, \tag{21}$$

where $a_+ = \max(a,0)$. Baranchik (1964) showed that the estimator (21) dominates $\hat{\mu}$, defined by (9). The same is true for the estimator (see (11))

$$\hat{\mu}_1 = \left(1 - \frac{(p-2)S_n}{(n+2)|\mathbf{X}|^2}\right)_+ \mathbf{X}. \tag{22}$$

Both estimators are not admissible, but no estimators uniformly better than (21) and (22) are known. An explicit formula for the risk of the estimator (22) is available in Robert (1988).

Remark 1. Since all James-Stein like estimators are all inadmissible, there are different approaches for their improvement. The last example of such an improvement has been given e.g. by Berry (1994), who used for this the improved variance estimator of variance estimator. The estimator constructed is

$$\hat{\mu} = \left(1 - \frac{r(F)}{F}\right) X,$$

where

$$r(F) = \begin{cases} \frac{p-2}{n+2}, & F \geq \frac{p}{n+2}, \\ \frac{p-2}{n+p+2}(1+F), & F < \frac{p}{n+2} \end{cases}$$

and $F = \mid X \mid^2 / S_n$.

This estimator dominate the traditional James-Stein estimator (22), but the domination does not hold for the positive part version of these estimators.

5. OTHER ASPECTS ON THE IMPROVEMENT OF THE RISK

The improvement in the risk obtained by Stein estimators is significant only if μ_i are close to the point towards which these estimators shrink. For a heavy-tailed prior distribution the Stein estimators may give little improvement over $\hat{\mu}^0$. Stein (1981), Dey & Berger (1983) and Alam and Mitra (1986) gave some modifications of (10) which help to overcome the difficulty. For example, Alam and Mitra (1986) instead of (9) proposed the estimator which is given component-wise by

$$\eta_i(\mathbf{X}) = \left(1 - \frac{(p-2)}{T_i}\right) X_i, \tag{23}$$

where

$$T_i = (\alpha - 1)X_i^2 + |\mathbf{X}|^2, \quad \alpha \leq p/2.$$

This estimator is minimax and reduces to (9) for $\alpha = 1$. Due to the term $(\alpha - 1)X_i^2$ the estimator naturally reduces the maximum component risk

$$R^*(\mu) = \max_i \mathbf{E}_\mu(\eta_i(\mathbf{X}) - \mu_i)^2.$$

Alam and Mitra (1986) have shown that the positive part estimator

$$\eta_i^+(\mathbf{X}) = \left(1 - \frac{(p-2)}{T_i}\right)_+ X_i \tag{24}$$

dominates $\eta_i(\mathbf{X})$, component-wise.

When σ^2 is not known and there is an estimator S_n of σ^2 independent of $\mathbf{X}$ and distributed as $\sigma^2\chi_n^2$ the estimator

$$\eta_i(\mathbf{X}) = \left(1 - \frac{(p-2)S_n}{(n+2)T_i}\right) X_i$$

is again minimax for $\alpha \le p/2$. Alam and Mitra (1986) have given explicit expressions for component risks of (23) and (24).

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be i.i.d. $\mathbf{N}(\mu, \sigma^2\mathbf{V})$ random vectors, where μ and σ^2 are unknown and $\mathbf{V}$ is a $p \times p$ known positive definite matrix. Suppose the loss function to be

$$L(\hat{\mu}_n, \mu) = (\hat{\mu}_n - \mu)^T \mathbf{Q} (\hat{\mu}_n - \mu), \tag{25}$$

where $\mathbf{Q}$ is a known positive definite matrix and $\hat{\mu}_n = \hat{\mu}_n(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ is an estimator for μ. Hudson (1974) and Berger (1976) introduced the following class of James-Stein estimators

$$\hat{\mu}_n = \lambda + \left(\mathbf{I} - \frac{r(F_n)}{F_n} \mathbf{Q}^{-1}\mathbf{V}^{-1}\right) \left(\overline{\mathbf{X}}_n - \lambda\right), \tag{26}$$

where $\lambda \in \mathbb{R}^p$ is a preassigned constant to which $\hat{\mu}_n$ shrinks the estimator $\overline{\mathbf{X}}_n$,

$$\begin{aligned} F_n &= n(\overline{\mathbf{X}}_n - \lambda)^T \mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1} \left(\overline{\mathbf{X}}_n - \lambda\right) / S_n^2, \\ S_n^2 &= [(n-1)p+2]^{-1} \sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}}_n)^T \mathbf{V}^{-1} (\mathbf{X}_j - \overline{\mathbf{X}}_n), \quad n \ge 2. \end{aligned}$$

Nickerson (1989) has shown that if $r(.)$ is absolutely continuous,

$$0 < r(F_n) < 2(p-2)$$

and

$$\psi_n(F_n) = F_n^{(p-2)/2} r(F_n) / (2(p-2) - r(F_n))^{1+2b_n}, \quad b_n = (p-2)/((n-1)p+2),$$

is nondecreasing in F_n, then (26) dominates $\overline{\mathbf{X}}_n$ under the loss (25). For every absolutely continuous $r(.)$ he introduced the positive-part estimator

$$\hat{\mu}_{n1} = \lambda + \left(\mathbf{I} - \frac{r(F_n)}{F_n} \mathbf{Q}^{-1}\mathbf{V}^{-1}\right)_+ (\overline{\mathbf{X}}_n - \lambda), \tag{27}$$

where

$$\left(\mathbf{I}-\frac{r(F_n)}{F_n}\mathbf{Q}^{-1}\mathbf{V}^{-1}\right)_+ = \mathbf{P}^{-1}\left[\text{diag}\left(\left(1-\frac{r(F_n)}{F_n}d_1^{-1}\right)_+,\right.\right.$$
$$\left.\left.\cdots,\left(1-\frac{r(F_n)}{F_n}d_p^{-1}\right)_+\right)\right]\mathbf{P},$$

$\mathbf{P}$ is a nonsigular matrix such that $\mathbf{PQ}^{-1}\mathbf{P}^T=\mathbf{I}$ and $\mathbf{PVP}^T=\mathbf{D}=$ diag $(d_1,\cdots,d_p)$. The estimator (27) dominates $\overline{\mathbf{X}}_n$ under the loss function (25).

One may substitute λ in (26) and (27) by another estimator $\hat{\mu}_2$ of μ, then estimators (26) and (27) will shrink $\bar{\mathbf{X}}_n$ towards this arbitrary estimators. The problem of selecting the $\hat{\mu}_2$ has been considered among others by George (1986) and Sengupta (1991).

Let $\mathbf{X}\sim N(\mu,\Sigma)$, where $N(\mu,\Sigma)$ is a p-variate normal distribution with unknown μ and unknown positive-definite covariance matrix Σ. Let there is an independent estimator $\mathbf{S}$ for Σ, which is distributed as $W_{n-1}(\mathbf{s};\Sigma)$. Using the loss function

$$L(\hat{\mu},\mu)=(\hat{\mu}-\mu)^T\mathbf{Q}(\hat{\mu}-\mu),$$

where $\mathbf{Q}$ is a given positive definite matrix, Alam(1977) considered a class of estimators $\hat{\mu}$ of μ of the form

$$\hat{\mu}=\Phi(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})\mathbf{X}$$

for a certain class of functions Φ. In particular he has shown that an estimator $\hat{\mu}$ given by

$$\hat{\mu}=\frac{2\nu}{p}\frac{{}_2F_1(\nu+1,\frac{n}{2}+\frac{p}{2}+1;\frac{p}{2}+1;\frac{x}{1+x})}{{}_2F_1(\nu,\frac{n}{2}+\frac{p}{2}+1;\frac{p}{2};\frac{x}{1+x})}\mathbf{X},\quad 0<\nu<1,$$

where $x=\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}$, is minimax if

$$\nu\geq\frac{(n-p+2)n}{2(2\alpha_0 p+n-p-2)}-\frac{n-p}{2},\quad \alpha_0>\frac{1}{2}+\frac{p}{2},$$

with

$$\alpha_0=\frac{tr\Sigma^{1/2}\mathbf{Q}\Sigma^{1/2}}{p\max(\alpha_1,\ldots,\alpha_p)},$$

where $\alpha_1,\ldots,\alpha_p$ denote the characteristic roots of the matrix $\Sigma^{1/2}\mathbf{Q}\Sigma^{1/2}$.

Berger & Haff (1983) for the loss function

$$L(\hat{\mu},\mu)=(\hat{\mu}-\mu)^T\mathbf{Q}(\hat{\mu}-\mu)/tr(\mathbf{QS})$$

have considered a class of minimax estimators of μ

(28) $$\hat{\mu}=\left(\mathbf{I}-c\alpha(\mathbf{Q}^{1/2}\mathbf{SQ}^{1/2})h(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})\mathbf{Q}^{-1}\mathbf{S}^{-1}\right)\mathbf{X},$$

where α and h are positive real valued functions and $c \geq 0$. Berger & Haff (1983) have given conditions on c, α and h under which estimators (28) are minimax and hence

$$R(\hat{\mu},\mu) \leq R(\hat{\mu}^0,\mu).$$

Simple cases of the above problem under the same loss function have been considered by Menjoge & Rao (1985).

We see that the class of estimators better than $\hat{\mu}^0$ is very large, so it is difficult to choose the needed one. Berger (1980a) has shown that each estimator better than $\hat{\mu}^0$ is significantly better only for μ in a small region of the parameter space. Berger (1982b) proposed a rather simple minimax estimator which allows the user to select the region where significant improvement over $\hat{\mu}^0$ is achieved. He described the region of preferency by an ellipse

$$\{\mu : (\mu - \lambda)^T \mathbf{A}^{-1} (\mu - \lambda) \leq p\},$$

where λ is the center of the region and $\mathbf{A}$ determines axes and an orientation of the ellipse.

Let $\mathbf{X} = (X_1, \cdots, X_p)^T$ be distributed as $N(\mu, \Sigma)$, Σ being known. It is desired to estimate μ under the loss function

$$L(\hat{\mu},\mu) = (\hat{\mu} - \mu)^T \mathbf{Q} (\hat{\mu} - \mu),$$

where $\mathbf{Q}$ is a known positive-definite matrix. Assume for simplisity that $\mathbf{Q}, \Sigma$ and $\mathbf{A}$ are diagonal with elements q_i, σ_i^2 and A_i respectively. The proposed minimax estimator is given, coordinate-wise, by

$$\hat{\mu}_i = X_i - \frac{\sigma_i^2}{\sigma_i^2 + A_i}(X_i - \lambda_i)\left[\frac{1}{q_i^*}\sum_{j=1}^{p}(q_j^* - q_{j+1}^*)\min_j\left\{1, \frac{2(j-2)_+}{||\mathbf{x}^j - \lambda^j||^2}\right\}\right],$$

where

$$q_i^* = q_i\sigma_i^4/(\sigma_i^2 + A_i), ||\mathbf{x}^j - \lambda^j||^2 = \sum_{i=1}^{j}(x_i - \lambda_i)^2/(\sigma_i^2 + A_i), q_{p+1}^* = 0$$

and X_i are indexed so that $q_1^* \geq q_2^* \geq \cdots \geq q_p^*$.

Naturally, the point λ may be thought to be a prior for μ and $\mathbf{A}$ to be a prior covariance matrix for Σ.

Problems of estimating matrices of normal mean where the Stein's method may be applied have been considered by Zheng (1986) and Konno (1991).

It is worth to note that the minimax property of Stein's rule is preserved not only with respect to the quadratic loss function but with respect to a generalised loss function

$$L(\hat{\mu},\mu) = |\hat{\mu}-\mu|^{2m}$$

too (Alam and Hawkes (1979)).

The Stein effect of improving upon usual estimators of mean of a multivariate normal distribution is also valid under the classical Pitman closeness criterion. Let $\mathbf{X}$ be $N(\mu,\sigma^2\mathbf{V})$ random vector, where $\mathbf{V}$ is a known positive definite matrix while μ and σ^2 are both unknown. For a given positive-definite matrix $\mathbf{Q}$ define the norm

$$||\mathbf{x}-\mathbf{y}||_{\mathbf{Q}} = (\mathbf{x}-\mathbf{y})^T\mathbf{Q}(\mathbf{x}-\mathbf{y}).$$

One may say that an estimator $\hat{\mu}_1$ is closer to μ than $\hat{\mu}_2$ is in the Pitman sense (in the norm $||.||_{\mathbf{Q}}$) if

$$P\{||\hat{\mu}_1-\mu||_{\mathbf{Q}} \leq ||\hat{\mu}_2-\mu||_{\mathbf{Q}}\} \geq \frac{1}{2} \tag{29}$$

for all μ and σ^2.

Consider shrikage estimators of the form

$$\hat{\mu} = \mathbf{X} - \frac{\varphi(\mathbf{X},S_m)S_m\mathbf{Q}^{-1}\mathbf{V}^{-1}\mathbf{X}}{\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Q}^{-1}\mathbf{V}^{-1}\mathbf{X}}, \tag{30}$$

where φ is a nonegative function bounded from above by

$$(p-1)(3p+1)/(2p)$$

for every $(\mathbf{X},S_m)$ and a statistic S_m is independent of $\mathbf{X}$ and distributed like $m^{-1}\sigma^2\chi^2_m$. Sen, Kubokawa and Saleh (1989) have shown that the estimator (30) is closer to μ than $\mathbf{X}$ in the Pitman sense (29) (see also Keating and Mason (1988) and Strivastava (1993)). James and Stein (1992) have formuated the general problem of admissible estimation with quadratic loss and considered some problems of admissibility of Pitman's estimators. They have formulated also some unsolved problems.

The Stein approach is useful not only for estimating the unknown vector of means μ if $\mathbf{X} \sim N(\mu,\Sigma)$ but for estimating the unknown covariance matrix Σ as well. The usual best invariant unbiased and minimax estimator of Σ is $\hat{\Sigma}_0 = \mathbf{S}/(n-1)$. It is known that the sample eigenvalues of $\mathbf{S}$ tend to be more spread out than the population eigenvalues of Σ. This fact suggests to search estimators of Σ whose risks are smaller than that of $\hat{\Sigma}_0$. Dey and Srinivasan (1985, 1986) (see also reeferences there in) obtained some improved estimators of Σ under the Stein loss function

$$L(\hat{\Sigma},\Sigma) = tr(\hat{\Sigma}\Sigma^{-1}) - \log\det(\hat{\Sigma}\Sigma^{-1}) - p. \tag{31}$$

They considered the class of orthogonally invariant estimators

$$\hat{\Sigma} = R\varphi(\mathbf{L})R^T, \tag{32}$$

where $\mathbf{S} = R\mathbf{L}R^T$ with R the matrix of normalized eigenvectors ($RR^T = R^TR = \mathbf{I}$), $\mathbf{L} =$ diag $(l_1, l_2, \cdots, l_p)$ is the diagonal matrix of corresponding eigenvalues with

$$l_1 \geq l_2 \geq \cdots \geq l_p \text{ and } \varphi(\mathbf{L}) = \text{ diag } (\varphi_1(\mathbf{L}), \varphi_2(\mathbf{L}), \cdots, \varphi_p(\mathbf{L})).$$

For an estimator $\hat{\Sigma}$ of the form (32) the loss function (31) reduces to

$$L(\hat{\Sigma}, \Sigma) = tr(\hat{\Sigma}\Sigma^{-1}) - \sum_{i=1}^{p} \log \varphi_i(\mathbf{L}) + \log\det(\Sigma) - p. \tag{33}$$

Since the last two terms in (33) are constant it is sufficient to consider the risk as follows

$$R^*(\hat{\Sigma}, \Sigma) = \mathbf{E}_\Sigma \left\{ tr(\hat{\Sigma}\Sigma^{-1}) - \sum_{i=1}^{p} \log \varphi_i(\mathbf{L}) \right\}. \tag{34}$$

Having obtained the unbiased estimator of $R^*(\hat{\Sigma}, \Sigma)$ and its upper bound, Dey and Srinivasan (1985) have shown that the estimator (32) for $p \geq 3$ and

$$\varphi_i(\mathbf{L}) = \frac{l_i}{n} - \frac{(l_i log l_i)\tau(u)}{b+u}, \qquad i = 1, 2, \ldots, p,$$

where

$$u = \sum_{i=1}^{p} \log^2 l_i, \qquad b > 144(p-2)^2/25k^2,$$

and $\tau(u)$ is a function satisfying:

(i) $0 < \tau(u) < 2(p-2)/k^*$, $k^* = 5k^2/6$;

τ is monotone nondecreasing in u and $\mathbf{E}[\tau'(u)] < \infty$, dominates $\hat{\Sigma}_0$ for the loss function (33). Dey and Srinivasan (1985) gave another estimators of Σ whose risks are smaller than that of the considered one. For orthogonally invariant estimators (32) the order relation of components of $\mathbf{L}$ and $\varphi(\mathbf{L})$ is important. If their components follow the same order relation the estimator (32) will dominate an estimator which does not preserve this ordering (Sheena and Takemura (1992)). Perron (1992) proposed a minimax orthogonally equivariant estimator $\hat{\Sigma}$ of Σ which satisfies the ordering and the shrinkage properties. The two sample analogue of the above problem has been considered by Loh (1991b) (see Remark 1).

Kubokawa (1989) gave an improved estimator of Σ under the quadratic loss function

$$\mathbf{L}(\hat{\Sigma},\Sigma) = tr(\hat{\Sigma}\Sigma - \mathbf{I})^2. \tag{35}$$

He has shown that

$$\hat{\Sigma} = \begin{cases} b(\mathbf{S}+\mathbf{X}\mathbf{X}^T) & \text{if } \mathbf{X}^T\mathbf{S}^{-1}\mathbf{X} \leq (a_0 - b)/b, \\ a_0\mathbf{S} & \text{otherwise ,} \end{cases} \tag{36}$$

where

$$(n+2)a_0/(n+3) \leq b < a_0, \qquad a_0 = 1/(n+p+1),$$

dominates $\hat{\Sigma}_0$ under the loss function (35). Dey, Ghosh and Srinivasan (1990) have considered the problem under a loss introduced by Efron and Morris (1976b).

In the multiple linear regression model containing more than two explanatory variables with normally distributed disturbances, the least squares estimator for the coefficient vector is inadmissible under a quadratic loss function.

Srivastava & Srivastava (1993) using the Stein-rule proposed the estimator which dominates the least squares estimator under a general convex loss function.

The Stein's effect is useful not only when estimating parameters of a multivariate normal distribution but for estimating parameters of some univariate and many other multivariate distributions including their mixtures too.

Bravo and MacGibbon (1988a) proposed James-Stein estimators for the parameters of an inverse Gaussian distribution. Bravo and MacGibbon (1988b) and Srivastava and Bilodeau (1989) constructed minimax Stein's like estimators for the mean vector μ under loss functions of the type (1) if an observation $\mathbf{X}$ is distributed as a scale mixture of normal distributions. The spherically symmetric case has been considered by Brandwein and Strawderman (1990), Ralescu, Brandwein and Strawderman (1992) and Cellier and Fourdrinier (1992).

Consider the problem of estimation of a common location of several exponential distributions with unknown scale parameters. Let $\{X_{ij}\}$, $j = 1,2,\ldots,n_i$, $i = 1,2,\ldots,p$, be a sample where

$$X_{ij} \sim \frac{1}{\sigma_i} exp\left[-\frac{x_{ij}-\mu}{\sigma_i}\right], \qquad x_{ij} \geq \mu, \qquad \forall i,j. \tag{37}$$

The problem of estimating the common location μ of the model (37) arises, e.g., in life testing, survival analysis, etc. It is known that the set of complete sufficient statistics for $\mu,\sigma_1,\sigma_2,\ldots,\sigma_p$ is $Z,T_1,T_2,\ldots,T_p$, where

$$Z = \min_{\substack{1 \le i \le p \\ 1 \le j \le n_i}} X_{ij} = \min_{1 \le i \le p} X_{i(1)},$$

$$T_i = \sum_{j=1}^{n_i} (X_{ij} - Z), \quad i = 1,2,\ldots,p.$$

The **MLE** $\hat{\mu}$ and the **MVUE** $\tilde{\mu}$ are given by (Ghosh and Razmpour (1984), see also, Bordes, Nikulin, Voinov (1994))

$$\hat{\mu} = Z, \quad \tilde{\mu} = Z - \left(\sum_{i=1}^{p} \frac{n_i(n_i - 1)}{T_i} \right)^{-1}. \tag{38}$$

Introducing instead of T_i the statistics

$$T_i^* = \sum_{j=1}^{n_i} (X_{ij} - X_{i(1)}), \quad i = 1,2,\ldots,p,$$

assuming that the sample sizes $n_1,\ldots,n_p$ are all equal $n > 2$ and denoting

$$\hat{a}_c^* = nc \sum_{i=1}^{p} (1/T_i^*)$$

Pal and Sinha (1990) proved that the estimator

$$\hat{\mu}_c^* = Z - (\hat{a}_c^*)^{-1} \tag{39}$$

dominates the **MLE** $\hat{\mu} = Z$ for every $c \ge n/2$ under the squared loss

$$L(\hat{\mu}_c^*, \mu) = (\hat{\mu}_c^* - \mu)^2.$$

It dominates $\hat{\mu}$ whenever

$$p \ge 4 \text{ and } n > 5, \text{ or } p = 3 \text{ and } n > 4, \text{ or } p = 2 \text{ and } n > 3.$$

The same is true for the Pitman's criterion of nearness. All the same hold in the case of unequal sample sizes (Pal and Sinha (1990)).

Berger (1980b) developed a technique for improving upon inadmissible estimators of parameters for a class of continuous exponential distributions.

Let $\mathbf{X} = \{X_1, \dots, X_p\}$ be a sample, where $\mathbf{X} \sim f(\mathbf{x};\theta)$ and the loss function in estimating $\mathbf{g}(\theta)$ by $\delta(\mathbf{X})$ is $L(\delta(\mathbf{X}), \mathbf{g}(\theta))$. Following Stein Berger used for the risk

$$R(\delta(\mathbf{X}), \mathbf{g}(\theta)) = \mathbf{E}_\theta L(\delta(\mathbf{X}), \mathbf{g}(\theta))$$

the representation

$$R(\delta, \theta) = \int L(\delta(\mathbf{x}), \mathbf{g}(\theta)) f(\mathbf{x};\theta)\mathbf{dx} = \int \mathcal{D}(\delta(\mathbf{x})) f(\mathbf{x};\theta)\mathbf{dx}, \tag{40}$$

where a function $\mathcal{D}(\delta(\mathbf{x}))$ involves $\delta(\mathbf{x})$ and its derivatives but not θ. Comparing an estimator $\delta^*(\mathbf{X})$ with the usual one $\delta^0(\mathbf{X})$ one should solve the inequality

$$R(\delta^*, \theta) - R(\delta, \theta) = \mathbf{E}_\theta[\mathcal{D}(\delta^*(\mathbf{X})) - \mathcal{D}(\delta^0(\mathbf{X}))] < 0$$

for all θ. The problem thus reduces to solving the differential inequality

$$\mathcal{D}(\delta^*(\mathbf{x})) - \mathcal{D}(\delta^0(\mathbf{x})) < 0. \tag{41}$$

Let X_i, $i = 1, 2, \dots, p$, be independent random Gamma variables with probability density function

$$f(x_i;\theta_i) = \theta_i^{\alpha_i} x_i^{(\alpha_i - 1)} e^{-x_i\theta_i}/\Gamma(\alpha_i), \qquad \alpha_i > 0, 0 < \theta_i < \infty, \tag{42}$$

and we want to estimate the vector

$$(\theta_1^{-1}, \theta_2^{-1}, \dots, \theta_p^{-1})^T \text{ by } \delta(\mathbf{X}) = (\delta_1(\mathbf{X}), \delta_2(\mathbf{X}), \dots, \delta_p(\mathbf{X}))^T$$

with the loss function

$$L(\delta, \theta) = \sum_{i=1}^{p} \theta_i^m (1 - \delta_i(\mathbf{X})\theta_i)^2, \qquad m \in Z, \quad \delta = (\theta_1, \dots, \theta_p)^T. \tag{43}$$

Solving inequality (41) for $m = -2, -1, 0$ and 1, Berger (1980b) obtained estimators $\delta^*(\mathbf{X})$, which for $p \geq 2$ or 3 under the loss function (42) are uniformly better than the standard estimator

$$\delta_i^0(\mathbf{X}) = \frac{X_i}{\alpha_i + 1}$$

of θ_i^{-1} for the loss

$$\theta_i^m (1 - \delta_i \theta_i)^2, \qquad i = 1, 2, \dots, p.$$

In the case $m = -2$ e.g., which corresponds to the usual sum of squares error loss the estimator $\delta^*(\mathbf{X})$ is given component-wise by

(44)

$$\delta_i^*(\mathbf{X}) = \frac{X_i}{\alpha_i + 1}\left(1 + \frac{c(\alpha_i - 1)(\alpha_i + 1)^2/X_i^2}{[b + \sum_{j=1}^p (\alpha_j + 1)^4/X_j^2]} + \frac{2c(\alpha_i + 1)^6/X_i^4}{[b + \sum_{j=1}^p (\alpha_j + 1)^4/X_j^2]^2}\right),$$

$0 < c < 4(p-1)$, $p \geq 2$, $b \geq (c^2/4)(1 + 1/\min\alpha_i)$.

The Berger's estimators $\delta^*(\mathbf{X})$ for different values of m are completely different in their functional form. Das Gupta (1986) and Bilodeau (1988) obtained a new class of estimators for $(\theta_1^{-1}, \theta_2^{-1}, \ldots, \theta_p^{-1})^T$ under the more general weighted loss function

$$L(\delta,\theta) = \sum_{i=1}^{p} w_i \theta_i^{m_i} (1 - \delta_i(\mathbf{x})\theta_i)^2, \qquad (w_i > 0, m_i \neq 0).$$

All these estimators possess the same functional form. Bilodeau (1988) showed that for $p \geq 2$ the estimator

$$\delta_i(\mathbf{X}) = \frac{X_i}{\alpha_i + 1}(1 + \phi_i(\mathbf{X})), \tag{45}$$

where

$$\phi_i(\mathbf{X}) = -k(\text{sign}\Delta_i) X_i^{m_i/2} \prod_{j=1}^{p} X_j^{-m_j/2p}$$

and

$$\Delta_i = \frac{\Gamma(\alpha_i + 2 + \frac{m_i}{2} - \frac{m_i}{2p})}{(\alpha_i + 1)\Gamma(\alpha_i)} - \frac{\Gamma(\alpha_i + 1 + \frac{m_i}{2} - \frac{m_i}{2p})}{\Gamma(\alpha_i)}$$

uniformly dominates

$$\delta_i^0(\mathbf{X}) = \frac{X_i}{\alpha_i + 1} \text{ for } 0 < k < 2MDp/KB,$$

where

$$M = \prod_{i=1}^{p} \frac{\Gamma(\alpha_i - \frac{m_i}{2p})}{\Gamma(\alpha_i)}, \qquad D = \min_i \frac{w_i \mid \Delta_i \mid \Gamma(\alpha_i)}{(\alpha_i + 1)\Gamma(\alpha_i - \frac{m_i}{2p})},$$

$$K = \sum_{i=1}^{p} \frac{w_i \Gamma(\alpha_i + 2 + m_i - \frac{m_i}{p})}{(\alpha_i + 1)^2 \Gamma(\alpha_i - \frac{m_i}{p})} \text{ and } B = \prod_{i=1}^{p} \frac{\Gamma(\alpha_i - \frac{m_i}{p})}{\Gamma(\alpha_i)}.$$

Dey, Ghosh and Srinivasan (1987) using the Berger (1980b) technique obtained for the probability model (42) different classes of shrinkage estimators dominating for $p \geq 3$ the **UMVUE**

$$\delta^0(\mathbf{X}) = (X_1/\alpha_1, X_2/\alpha_2, \ldots, X_p/\alpha_p)^T$$

which is usual for the loss function

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

being taken by Dey *et al.*

A remarkable Stein's approach for estimators improvement proves to be useful for simultaneous estimation of location parameters of distributions with finite support too. Akai (1986) considered probability density functions which are positive on a finite interval, symmetric about θ_i and have the same variance. He obtained a class of shrinkage estimators of the location vector $\theta = (\theta_1, ..., \theta_p)^T$ under the squared error loss function. In particular, explicit dominating estimators where the distributions of X_i's are mixture of two uniform distributions were given.

A large amount of work is devoted to the application of Stein's ideas for shrinkage parameters estimation in multivariate discrete distributions. Ghosh, Hwang and Tsui (1983) constructed estimators which dominate the **UMVUE** under a weighted squared error loss function shrinking the **UMVUE** towards a prescribed nonzero point or a data-based point. Let, for example, X_i, $i = 1, 2, ..., p$, be independent negative binomial random variables and

$$P(X_i = x_i; \theta_i) = \binom{r_i + x_i - 1}{r_i - 1} \theta_i^{x_i}(1-\theta_i)^{r_i}, \qquad x_i = 0, 1, ...$$

The **UMVUE** $\delta^0(\mathbf{X}) = (\delta_1^0(\mathbf{X}), ..., \delta_p^0(\mathbf{X}))^T$ of vector $\theta = (\theta_1, ..., \theta_p)^T$ is (see §A25 in Voinov and Nikulin (1993)) the vector with components

$$\delta_i^0(\mathbf{X}) = X_i/(X_i + r_i - 1).$$

Denoting $N(\mathbf{X}) = \#\{i : X_i > \lambda_i\}$, where $\lambda = (\lambda_1, ..., \lambda_p)^T$ is above mentioned prescribed point, Ghosh, Hwang and Tsui (1983) showed that estimators

$$\delta_i(\mathbf{X}) = \delta_i^0(\mathbf{X}) - c[N(\mathbf{X}) - 2]_+ H_i(X_i)/D, \tag{46}$$

where

$$\begin{aligned} D &= \sum_{j=1}^{p} d_j(X_j), \quad h_i(X_i) = \sum_{j=1}^{X_i} (r_i + j - 1)/j, \quad H_i(X_i) = h_i(X_i) - h_i(\lambda_i), \\ d_i(X_i) &= \begin{cases} H_i^2(X_i) + b_i H_i(X_i) & \text{if } X_i \geq \lambda_i, \\ H_i^2(X_i) + a_i & \text{if } X_i < \lambda_i, \end{cases} \\ a_i &= r_i\{3h_i(\lambda_i)/2 - 1\}_+ \text{ and } b_i = (r_i + \lambda_i + 1)/(\lambda_i + 2), \end{aligned}$$

for $0 < c \leq 2$ and $p > 2$ dominate the **UMVUE** $\delta^0(\mathbf{X})$ under

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

Many results conserning the improvement upon **MVUEs** for discrete probability distributions parameters and functions of them including densities with dependent marginals may be found in Tsui (1986), Kant and Sharma (1986), Chou (1991) and Dey and Chung (1991).

In order to finish this review, we would like to add two more remarks concerning this topic.

Remark 2. The main feature of Stein's technique for estimators improvement consists of a representing a risk function as an expectation of a function which does not involve unknown parameters. There exists another approach to the problem with the same idea to work with expressions independent on parameters. If there exists a sufficient statistic for unknown parameters we may construct, if exists, the unbiased estimator for a risk function and then to use it for estimators improvement. This technique has been used among others by Haff (1980) and Loh (1991a). Loh (1991b) considered the following problem.

Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be two independent $p \times p$ Wishart matrices where $\mathbf{S}_1 \sim W_{n_1-1}(\mathbf{s}_1;\Sigma_1)$ and $\mathbf{S}_2 \sim W_{n_2-1}(\mathbf{s}_2;\Sigma_2)$. The problem is to estimate (Σ_1,Σ_2) under the loss function

$$L(\hat{\Sigma}_1,\hat{\Sigma}_2;\Sigma_1,\Sigma_2) = \sum_{i=1}^{2}\{tr(\Sigma_i^{-1}\hat{\Sigma}_i) - \log\det(\Sigma_i^{-1}\hat{\Sigma}_i) - p\}. \tag{47}$$

Loh (1991b) considered estimators invariant under the group of transformations

$$\Sigma_i \to \mathbf{A}\Sigma_i\mathbf{A}^T, \quad \mathbf{S}_i \to \mathbf{A}\mathbf{S}_i\mathbf{A}^T, \quad \forall \mathbf{A} \in GL(p,R), \quad i = 1,2, \tag{48}$$

where $GL(p,R)$ denotes the set of all $p \times p$ nonsingular matrices. He proved that the estimator $(\hat{\Sigma}_1,\hat{\Sigma}_2)$ is an equivariant under transformations (48) estimator of (Σ_1,Σ_2) iff

$$\begin{aligned} &\hat{\Sigma}_1(\mathbf{S}_1,\mathbf{S}_2) = \mathbf{B}^{-1}\Psi(\mathbf{I}-\mathbf{F})\mathbf{B}^{(T)^{-1}}, \\ &\hat{\Sigma}_2(\mathbf{S}_1,\mathbf{S}_2) = \mathbf{B}^{-1}\Phi(\mathbf{F})\mathbf{B}^{(T)^{-1}}, \\ &\mathbf{B}(\mathbf{S}_1+\mathbf{S}_2)\mathbf{B}^T = \mathbf{I}, \quad \mathbf{B}\mathbf{S}_2\mathbf{B}^T = \mathbf{F}, \end{aligned} \tag{49}$$

where Ψ and Φ are diagonal matrices, $\mathbf{F} = \mathrm{diag}(f_1,...,f_p)$ with $f_1 \geq ... \geq f_p$. Denote

$$\nabla^{(i)} = (\nabla^{(i)}_{jk})_{p\times p}, \quad 1 \leq j,k \leq p, \quad i = 1,2,$$

where

$$\nabla^{(i)}_{jk} = \frac{1}{2}(1+\delta_{jk})\frac{\partial}{\partial S^{(i)}_{jk}}, \quad \delta_{jk} = \begin{cases} 1, & j=k \\ 0, & j \neq k, \end{cases} \quad \mathbf{S}_i = (S^{(i)}_{jk}),$$

$S^{(i)}_{jk}$, $1 \leq j,\ k \leq p$, being elements of $\mathbf{S}_i$, $\quad i = 1,2$. If matrices

$$\Psi = \mathrm{diag}(\psi_1,...,\psi_p) \text{ and } \Phi = \mathrm{diag}(\phi_1,...,\phi_p)$$

satisfy the Wishart identity in the sense that

$$\mathbf{E}\ \mathrm{tr}\ (\Sigma_i^{-1}\hat{\Sigma}_i) = \mathbf{E}\ \mathrm{tr}\ [2\nabla^{(i)}(\hat{\Sigma}_i) + (n_i - p - 1)\mathbf{S}_i^{-1}\hat{\Sigma}_i], \qquad i = 1,2, \tag{50}$$

then the risk of $(\hat{\Sigma}_1, \hat{\Sigma}_2)$ with respect to the loss (47) is given by

$$R(\hat{\Sigma}_1,\hat{\Sigma}_2;\Sigma_1,\Sigma_2) = \mathbf{E}\{\sum_{i=1}^{p}[\frac{n_1-p-1}{1-f_i}\psi_i - 2\psi_i\sum_{i\neq j}\frac{f_j}{f_i-f_j} + 2\psi_i + 2f_i\frac{\partial\psi_i}{\partial(1-f_i)} -$$

$$\log\frac{\psi_i}{1-f_i} + \frac{n_2-p-1}{f_i}\phi_i + 2\phi_i\sum_{i\neq j}\frac{1-f_j}{f_i-f_j} + 2\phi_i + 2(1-f_i)\frac{\partial\phi_i}{\partial f_i} - \log\frac{\phi_i}{f_i} -$$

$$\log\chi^2_{n_1-i+1} - \log\chi^2_{n_2-i+1} - 2]\} = \mathbf{E}\hat{R}(\hat{\Sigma}_1,\hat{\Sigma}_2;\Sigma_1,\Sigma_2). \tag{51}$$

The estimator $\hat{R}(\hat{\Sigma}_1,\hat{\Sigma}_2;\Sigma_1,\Sigma_2)$ being dependent on complete sufficient for parameters Σ_1,Σ_2 statistic $(\mathbf{S}_1,\mathbf{S}_2)$ (see (49)) is the **UMVUE** of the risk $R(\hat{\Sigma}_1,\hat{\Sigma}_2;\Sigma_1,\Sigma_2)$.

To construct the Stein-type estimator for (Σ_1,Σ_2) one should now to minimize the **UMVUE** $\hat{R}(\hat{\Sigma}_1,\hat{\Sigma}_2;\Sigma_1,\Sigma_2)$ to obtain elements $\psi_1^*,...,\psi_p^*$ and $\phi_1^*,...,\phi_p^*$ of matrices Ψ and Φ respectively which give minimum to $\hat{R}(.)$. By ignoring the derivative terms in $\hat{R}(.)$ and solving equations

$$\frac{\partial\hat{R}(.)}{\partial\psi_i} = 0, \qquad \frac{\partial\hat{R}(.)}{\partial\phi_i} = 0, \qquad \forall i,$$

Loh (1991) obtained approximate values of ψ_i^*,ϕ_i^*, $\quad i = 1,2,...p$, as follows

$$\begin{aligned}
\psi_i^* &= (1-f_i)/\left[n_1 - p + 1 - 2\sum_{i\neq j}\frac{f_j(1-f_i)}{f_i-f_j}\right], \\
\phi_i^* &= f_i/\left[n_2 - p - 1 + 2\sum_{i\neq j}\frac{f_i(1-f_j)}{f_i-f_j}\right], \qquad i = 1,...,p,
\end{aligned}$$

where $0 \leq \psi_1 \leq ... \leq \psi_p$, $\phi_1 \geq ... \geq \phi_p \geq 0$. Substituting ψ_i^* and ϕ_i^*, $i = 1,...,p$, into (49) we obtain a Stein-type estimator $(\hat{\Sigma}_1(\mathbf{S}_1,\mathbf{S}_2),\hat{\Sigma}_2(\mathbf{S}_1,\mathbf{S}_2))$ for (Σ_1,Σ_2). Loh (1991a) have also shown that natural ordering of $\phi_1,...,\phi_p$ may be altered by Stein's isotonic regression technique.

Remark 3. The uniformly minimum variance unbiased estimation technique is useful not only for a total risk estimating to improve estimators, but for estimating the mean-squared-error matrix as well, since the last estimator is helpful in identifying the components of $\hat{\mu}$ which contribute most to the total risk.

Let $\mathbf{X} \sim N(\mu, \Sigma)$, μ and Σ being unknown. Let $\mathbf{S}$ be an estimator of Σ independent of $\mathbf{X}$ and $\mathbf{S} \sim W_{n-1}(\mathbf{s}; \Sigma)$. Considering the class of estimators

$$\hat{\mu}_s = \left(1 - \frac{k}{\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}}\right)\mathbf{X}, \quad 0 < k < \frac{2(p-2)}{n-p+3},$$

Bilodeau and Srivastava (1988) have shown that the **MVUE** of the mean-squared matrix

$$M(\hat{\mu}_s) = \mathbf{E}(\hat{\mu}_s - \mu)(\hat{\mu}_s - \mu)^T$$

is given by

$$\hat{M}(\hat{\mu}_s) = \left(1 - \frac{2k}{\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}}\right)\frac{\mathbf{S}}{n} + k\left(k + \frac{4(n+1)}{n(n-p+3)}\right)\frac{\mathbf{X}\mathbf{X}^T}{(\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X})^2}.$$

It is this estimator which allows to determine "risky" components of $\hat{\mu}_s$.

ACKNOWLEDGMENT

The authors wish to thank Professors R.Beran and C.Cuadras for their advice and encouragement and referees for their careful reading of the work.

REFERENCES

[1] **Akai, T.** (1988). "Further generalization of Stein estimators". *Keio Science and Technology reports*, **40**, 4, 41–54.

[2] **Alam, K.** (1977). "Minimax estimators of a multivariate normal mean". *Scand. J. Statist.*, **4**, 125–130.

[3] **Alam, K.** and **Mitra A.** (1986). "Component risk in multiparameter estimation". *Ann. Inst. Statist. Math.*, **38**, Part A, 399–410.

[4] **Alam, K.** and **Hawkes, J.S.** (1979). "Minimax property of Stein's estimator". *Commun. Statist.-Theor. Meth.*, **A8(6)**, 581–590.

[5] **Baranchik, A.J.** (1964). "Multiple regression and estimation of the mean of a multivariate normal distribution". *Stanford Univ. Technical Report #51.*

[6] **Baranchik, A.J.** (1970). "A family of minimax estimators of the mean of a multivariate normal distribution". *Ann. Math. Statist.*, **41**, 642–645.

[7] **Beran, R.** (1992). "Stein estimation in high dimensions and the bootstrap". *Preprint.* pp. 1–21.

[8] **Berger, J.** (1976). "Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss". *Ann. Statist.*, **4**, 223–226.

[9] **Berger, J.O.** and **Srinivasan, C.** (1977). "Generalized Bayes estimators in multivariate problems". *Purdue Univ., Depart. of Statist., Mimeograph Series #481.*

[10] **Berger, J.** (1980a). "A robust generalized Bayes estimator and confidence region for a multivariate normal mean". *Ann. Statist.*, **8**, 716–761.

[11] **Berger, J.** (1980b). "Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters". *Ann. Statist.*, **8**, 545–571.

[12] **Berger, J.** (1982a). "Bayesian robustness and the Stein effect". *JASA*, **77**, 358–368.

[13] **Berger, J.** (1982b). "Selecting a minimax estimator of a multivariate normal mean". *Ann. Statist.*, **10**, 81–92.

[14] **Berger, J.** and **Haff. L.R.** (1983). "A class of minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix". *Stat. & Decisions.*, **1**, 105–129.

[15] **Berry, J.C.** (1994). "Improving the James-Stein estimator using the Stein variance estimator". *Statist. & Prob. Letters.*, **20**, p. 241–245.

[16] **Bilodeau, M.** and **Srivastava, M.S.** (1982). "Estimation of the MSE matrix of the Stein estimator". *Can. J. Statist.*, **16**, 153–159.

[17] **Bilodeau, M.** (1988). "On the simultaneous estimation of scale parameters". *Can. J. Statist.*, **16**, 169–174.

[18] **Bock, M.E.** (1975). "Minimax estimators of the mean of a multivariate normal distribution". *Annals of Statist.*, **3**, 209–218.

[19] **Bordes, L., Nikulin, M.** and **Voinov, V.** (1994). "Unbiased estimators' tables of parameter functions for a several exponential with a common shift". *Preprint 94-07 de Bordeaux 2*, 22p.

[20] **Bravo, G.** and **Mac Gibbon, B.** (1988a). "Improved shrinkage estimators for the mean vector of a scale mixture of normals with unknown variance". *Can. J. Statist.*, **16**, 237–245.

[21] **Bravo, G.** and **Mac Gibbon, B.** (1988b). "Improved estimation for the parameters of an inverse Gaussian distribution". *Commun. Statist.- Theory Meth.*, **17**, 4285–4299.

[22] **Brandwein, A.C.** and **Strawderman, W.E.** (1990). "Stein estimation: The spherically symmetric case". *Statistical Science*, **5**, 356–369.

[23] **Cellier, D.** and **Fourdrinier, D.** (1992). "Estimation du paramètre de position d'une loi à symétrie sphérique: une condition générale de domination de l'estimateur des moindres carrés". *C.R. Acad. Sci. Paris*, **315**, 1203–1206.

[24] **Chen, S.Y.** (1988). "Restricted risk Bayes estimation for the mean of the multivariate normal distribution". *J. Multivariate Analysis.* **24**, 207–217.

[25] **Chou, J.P.** (1991). "Simultaneous estimation in discrete multivariate exponential families". *Ann. Statist.*, **19**, 314–328.

[26] **Das Gupta, A.** (1986). "Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses". *Ann. Statist.*, **14**, 206–219.

[27] **Dey, D.K.** and **Berger, J.O.** (1983). "On truncation of shrinkage estimators in simultaneous estimation of normal means". *JASA*, **78**, #384, 865–869.

[28] **Dey, D.K.** and **Srinivasan, C.** (1985). "Estimation of a covariance matrix under Stein's loss". *Annals of Statist.*, **13**, 1581–1591.

[29] **Dey, D.K.** and **Srinivasan, C.** (1986). "Trimmed minimax estimator of a covariance matrix". *Ann. Inst. Statist. Math.*, **38**, part A, 101–108.

[30] **Dey, D.K., Ghosh, M.** and **Srinivasan, C.** (1987). "Simultaneous estimation of parameters under entropy loss". *J. Statist. Plann. Imfer.*, **15**, 347–363.

[31] **Dey, D.K., Ghosh, M.** and **Srinivasan, C.** (1990). "A new class of improved estimators of a multinormal precision matrix". *Statistics & Decisions*, **8**, 141–151.

[32] **Dey, D.K.** and **Chung, Y.** (1991). "Multiparameter estimation in truncated power series distributions under the Stein's loss". *Commun. Statist. -Theory Meth.*, **20**, 309–326.

[33] **Efron, B.** and **Morris, C.** (1973). "Stein's estimation rule and its competitors. An empirical Bayes approach". *JASA*, **68**, 117–130.

[34] **Efron, B.** (1975). "Biased versus unbiased estimation". *Advances in Math.*, **16**, 259–277.

[35] **Efron, B.** and **Morris, C.** (1976a). "Families of minimax estimators of the mean of a multivariate normal distribution". *Ann. of Statist.*, **4**, 11–21.

[36] **Efron, B.** and **Morris, C.** (1976b). "Multivariate empirical Bayes and estimation of covariance matrices". *Ann. Statist.*, **4**, 22–32.

[37] **Egerton, M.** and **Laycock, P.** (1982). "An explicit formula for the risk of James-Stein estimators". *Canad. J. Statist.*, **10(3)**, 199–205.

[38] **Ghosh, M.** and **Razmpour, A.** (1984). "Estimation of the common location parameter of several exponentials". *Sankhyá*, **48A**, 383–394.

[39] **Ghosh, M., Hwang, J.T.** and **Tsui, K.W.** (1983). "Construction of improved estimators in multiparameter estimation for discrete exponential families". *Ann. Statist.*, **11**, 351–367.

[40] **George, E.** (1986). "Minimax multiple shrinkage estimators". *Ann. Statist.*, **14**, 188–205.

[41] **Haff, L.R.** (1980). "Empirical Bayes estimation of the multivariate normal covariance matrix". *Ann. Statist.*, **8**, 586–597.

[42] **Hoffmann, K.** (1992). "Improved estimation of distribution parameters: Stein-type estimators". *B.G. Teubner Verlagsgesellschaft mbH, Stuttgart.*

[43] **Hudson, H.M.** (1974). "Empirical Bayes estimation". *Technical Report #58, Departament of Statistics, Stanford University.*

[44] **James, W.** and **Stein, C.** (1960). "Estimation with quadratic loss". *Proc. 4rd Berkeley Symp. Math. Statist. Prob.* , **Vol 1**, *University of California Press, Berkeley, Calif.*, 361–379.

[45] **James, W.** and **Stein, C.** (1992). "Estimation with quadratic loss". *(in "Breakthroughs in Statistics.* **Vol.1**", *eds. S.Kotz & N.L.Johnson, Springer)*, 443–460.

[46] **Kant, V.** and **Sharma, D.** (1986). "Simultaneous estimation of λ_i^s for Poisson distributions". *Statistics*, **17**, 539–547.

[47] **Keating, J.P.** and **Mason, R.L.** (1988). "James-Stein estimation from an alternative perspective". *Amer. Statist.*, **42**, 160–164.

[48] **Konno, Y.** (1991). "On estimation of a matrix of normal means with unknown covariance matrix". *J. Multivariate Analysis*, **36**, 44–55.

[49] **Kubokawa, T.** (1989). "Improved estimation of a covariance matrix under quadratic loss". *Statist. & Prob. Letters*, **8**, 69–71.

[50] **Lehmann, E.L.** (1983). *Theory of Point Estimation.* Wiley, New York.

[51] **Loh, W.** (1991a). "Estimating covariance matrices". *Ann. Statist.*, **19**, 283–296.

[52] **Loh, W.** (1991b). "Estimating the common mean of two multivariate normal distributions". *Ann. Statist.*, **19**, 297–313.

[53] **Menjoge, S.S.** and **Rao, P.S.R.S.** (1985). "Simultaneous estimation of parameters using compound weighted loss". *Sankhyá*, **47B**, 338–345.

[54] **Nickerson, D.M.** (1989). "Dominance of the positive-part version of the James-Stein estimator". *Statist. & Prob. Letters*, **7**, 97–103.

[55] **Pal, N.** and **Sinha, B.K.** (1990). "Estimation of a common location of several exponentials". *Statist. & Decisions*, **8**, 27–36.

[56] **Perron, F.** (1992). "Minimax estimators of a covariance matrix". *J. Multivar. Anal.*, **43**, 16–28.

[57] **Ralescu, S., Brandwein, A.C.** and **Strawderman, W.E.** (1992). "Stein estimation for non-normal spherically symmetric location families in three dimensions". *J. Multivar. Analys.*, **42**, 35–50.

[58] **Robert, C.** (1988). "An explicit formula for the risk of the positive-part James-Stein estimator". *Canad. J. of Statist.*, **16**, 161–168.

[59] **Sen, P.K., Kubokawa, T.** and **Saleh, A.K.Md.E.** (1989). "The Stein paradox in the sense of the Pitman measure of closeness". *Ann. Statist.*, **17**, 1375–1386.

[60] **Sheena, Y.** and **Takemura, A.** (1992). "Inadmissibility of non-order-preserving orthogonally invariant estimators of the covariance matrix in the case of Stein's loss". *J. Multivariate Anal.*, **41**, 117–131.

[61] **Srivastava, M.S.** and **Bilodeau, M.** (1989). "Stein estimation under elliptical distributions". *J. of Multivariate Analysis*, **28**, 247–259.

[62] **Srivastava, A.K.** and **Srivastava, V.K.** (1993). "Pitman closeness for Stein-rule estimators of regression coefficients". *Statist. & Prob. Letters*, **18**, 85–89.

[63] **Stein, C.** (1956). "Inadmissibility of the usual estimator for a mean of multivariate normal distribution". *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, **Vol 1**, *University of California Press, Berkeley, Calif.*, 197–206.

[64] **Stein, C.** (1977). "Lectures on the theory of estimation of many parameters (in Russian)". *Studies in the Statistical Theory of Estimation, Part I* (I.A. Ibragimov and M.S.Nikulin. Eds.) Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division, **7**, pp.4–65.

[65] **Stein, C.** (1981). "Estimation of the mean of a multivariate normal distribution". *The Annals of Statistics*, **9**, 6, 1135–1151.

[66] **Stein, C.** (1986). "Lectures on the theory of estimation of many parameters". *J.Sov. Math.* **34**, 1, 1373–1403.

[67] **Stigler, S.M.** (1990). "The 1988 Neyman memorial lecture: a Galtonian perspective on shrinkage estimators". *Statistical Science* **5**, 1, 147–155.

[68] **Tsui, K.W.** (1986). "Multiparameter estimation for some multivariate discrete distributions with possibly dependent components". *Ann. Inst. Statist. Math.*, **38A**, 45–56.

[69] **Voinov, V.G.** and **Nikulin, M.S.** (1993). "Unbiased estimators and their applications". **V.1**, Univariate case, *Kluwer*.

[70] **Zheng, Z.** (1986). "On estimation of matrix of normal mean". *J. Multivariate Analysis*, **18**, 70–82.