CIRJE-F-6

Shrinkage and Modification Techniques in Estimation of Variance and the Related Problems: A Review

Tatsuya KUBOKAWA
The University of Tokyo

June 1998

Shrinkage and Modification Techniques in Estimation of Variance and the Related Problems: A Review

Tatsuya KUBOKAWA
University of Tokyo

One of the surprising decision-theoretic results Charles Stein discovered is the inadmissibility of the uniformly minimum variance unbiased estimator(UMVUE) of the variance of a normal distribution with an unknown mean. Some methods for deriving estimators better than the UMVUE were given by Stein, Brown, Brewster and Zidek. Recently Kubokawa established a novel approach, called the *IERD method*, by use of which one gets a unified class of improved estimators including their previous procedures. This paper gives a review for a series of these decision-theoretical developments as well as surveys the study of the variance-estimation problem from various aspects. Related to this issue, the paper enumerates several topics with the situations where the usual plain estimators are required to be shrunken or modified, and gives reasonable procedures improving the usual ones through the IERD method.

AMS 1991 subject classifications. Primary 62F11; secondary 62C12, 62C15, 62H12, 62J05.

Key words and phrases: Shrinkage procedure, statistical decision theory, empirical and generalized Bayes rules, estimation of variance, variance components, noncentrality parameter, linear calibration problem, estimation of restricted parameters.

1 Introduction

The decision theory of the parametric point estimation in a small sample has been remarkably developed since Charles Stein established breakthroughs in some problems. One of them is the surprising, important and seminal work given by Stein(1956), called the *Stein phenomenon* (or *Stein problem*), that the maximum likelihood estimator(MLE) is inadmissible and is improved on by a shrinkage rule in simultaneous estimation of the mean vector of a multivariate normal distribution when the dimension of the mean is larger than or equal to three. Stein(1973) developed an integration-by-parts approach, referred to as the *Stein identity*, which was very powerful and useful for deriving improved estimators, so that it brought about a huge amount of benefits and developments in this field. For the good accounts, see Judge and Bock(1978), Berger(1985), Brandwein and Strawderman(1990), Robert(1994) and Rukhin(1995). Also see Kubokawa(1997a) for an extensive survey including recent developments. The Stein identity produced the Haff identity in a Wishart distribution, which caused substantial development of research in estimation of a covariance matrix. This issue also stems from James and Stein(1961) having proved non-minimaxity and inadmissibility of the UMVUE of the covariance matrix.

The other topic of Stein's breakthroughs in point estimation is the amazing inadmissibility result discovered by Stein(1964), that the usual UMVUE of the variance of the normal distribution with an unknown mean is improved on by a shrinkage (modified) estimator by using the information contained in the sample mean. The proposed shrinkage estimator is a truncated procedure, and is still inadmissible since it cannot be expressed as a generalized Bayes rule. Brewster and Zidek(1974) developed, on the basis of Brown(1968)'s method, a smooth estimator better than the unbiased one, and showed the generalized Bayesness of it. The admissibility was established by Proskin(1985). Both of the Stein and Brewster-Zidek methods have been applied to provide improved confidence intervals of the variance and to get superior procedures in estimation of the scale parameter of an exponential distribution. Recently Kubokawa(1994a) and Takeuchi(1991) gave a novel technique for deriving improved procedures, which is based on an idea of expressing the difference of the risk functions via an integral, and we call it the *IERD (Integral Expression of Risk Difference) method.* The IERD method unifies the above two methods and gives a class of better procedures.

The major purpose of this paper is to survey a series of the decision-theoretical results concerning the above issue of estimating the variance from various aspects. This review is given in Section 2, including the Stein method, the Brown-Brewster-Zidek method, the IERD method, extensions to general distributions, applications to the interval estimation, extensions to estimation of multidimensional parameters and other methods proposed by Strawderman(1974) and Shinozaki(1995). Also the related problem about estimation of ratio of variances is dealt with and double shrinkage improved procedures are given.

Decision-theoretic studies on the shrinkage procedures have produced a large amount of beautiful and wonderful benefits in the above mentioned problems of Stein's breakthroughs. Beyond these familiar issues, we are really faced with lots of situations where the usual plain estimators are required to be shrunken or modified. In such situations, it is of importance to clarify the extent of how much usual procedures should be shrunken or modified. Empirical Bayes and hierarchical Bayes rules, cross-validation (or leaving-one-out), bootstrap and penalized methods and Wavelet analysis are standard techniques useful for specifying appropriate extents of shrinkage or modification. In a decision-theoretic sense, it may be reasonable to provide the extent of shrinkage such that the shrinkage estimator has a uniformly smaller risk than the usual plain estimator. The IERD method may be expected as one of tools useful for the purpose.

A second purpose of the paper is to enumerate several problems where the shrinkage or modification of usual estimators are required. These are given in Section 3, including estimation of variance components, noncentrality parameters and restricted parameters and the linear calibration, statistical control problems. Reasonable shrinkage procedures derived through the IERD method are there presented. Other problems enumerated include the estimation after a selection, multicollinearity, shrinkage procedures towards a null hypothesis, discriminant analysis, estimation of error rates and the problem of improper solutions in factor analysis.

The IERD method, in some estimation problems, can provide classes of estimators which include empirical and generalized Bayes rules improving on usual ones. The derivation of such Bayesian improved procedures is meaningful from a robust-Bayesian point of view. In general, the proper Bayes estimator depends on the prior knowledge completely while the knowledge is neglected in the usual procedure such as MLE and UMVUE. The proper Bayes rule is sensitive to the prior information and it has a large frequentist's risk-reduction when the prior information is true, while it gives a poor estimate otherwise. To the contrary, the merit of the Bayesian improved procedure is to guarantee the superiority to the usual one even if one can not suppose any exact prior information. The Bayesian improved procedure thus incorporates parts of the prior information and yields no actual harm from the frequentist viewpoint, namely, possessing the robust Bayesian property.

2 Estimation of Variance

Stein(1964) discovered the amazing inadmissibility result that the usual unbiased estimator of the variance of the normal distribution with an unknown mean is improved on by a shrinkage (modified) estimator by using the information contained in the sample mean. This section gives a review of the developments concerning this estimation problem. One can also see Maatta and Casella(1990) for good accounts of this issue.

2.1 The problem

We treat a canonical form which appears in analyses of experimental designs, linear regression models and others: Let S and $\boldsymbol{X}$ be independent random variables having

$$S/\sigma^2 \sim \chi_n^2, \quad \boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I}_p), \tag{2.1}$$

where χ_n^2 and $\mathcal{N}_p(\boldsymbol{\theta}, \sigma^2\boldsymbol{I}_p)$ designate, respectively, a chi-squared distribution with n degrees of freedom and a p-variate normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\sigma^2\boldsymbol{I}_p$ for identity matrix $\boldsymbol{I}_p$. It is supposed that $\boldsymbol{\theta}$ and σ^2 are unknown parameters and that we want to estimate the variance σ^2 based on $\boldsymbol{X}$ and S. Every estimator $\delta = \delta(\boldsymbol{X}, S)$ is evaluated by the risk function $R(\sigma^2, \boldsymbol{\theta}, \delta) = E[L(\delta/\sigma^2)]$ through the *Kullback-Leibler loss function*

$$L(\delta/\sigma^2) = \delta/\sigma^2 - \log(\delta/\sigma^2) - 1,$$

which is also called the Stein loss or entropy loss. When random variable y has density function $f(y;\omega)$, the Kullback-Leibler loss in estimating ω by δ is defined by the Kullback-Leibler distance between two densities $f(y;\delta)$ and $f(y;\omega)$, given by $\int \log\{f(y;\delta)/f(y;\omega)\} f(y;\delta)dy$. Another popular measure is the squared errored loss $L_s(\delta/\sigma^2) = (\delta/\sigma^2 - 1)^2$. It may be, however, inappropriate to employ $L_s(\delta/\sigma^2)$ in the estimation of the variance σ^2, because $L_s(\delta/\sigma^2)$ penalizes the underestimate less than the overestimate as seen from the fact that $\lim_{t\to 0} L_s(t) = 1$ and $\lim_{t\to\infty} L_s(t) = \infty$. Thus the Kullback-Leibler loss $L(\delta/\sigma^2)$ is adopted here.

A natural estimator of σ^2 is $\delta_0 = n^{-1}S$, which is UMVUE and RMLE (restricted maximum likelihood estimator). This is also optimal in the following sense: Let $O(p)$ be a class of $p \times p$ orthogonal matrices and consider the affine transformation $S \to c^2 S$, $\boldsymbol{X} \to c\Gamma\boldsymbol{X} + \boldsymbol{d}$, $\sigma^2 \to c^2\sigma^2$, $\boldsymbol{\theta} \to c\Gamma\boldsymbol{\theta} + \boldsymbol{d}$, $c \in \boldsymbol{R}$, $\boldsymbol{d} \in \boldsymbol{R}^p$, $\Gamma \in O(p)$. The estimation problem remains invariant under the affine transformation when an estimator of σ^2 is equivariant, that is, $\delta(c^2S, c\Gamma\boldsymbol{X} + \boldsymbol{d}) = c^2\delta(S, \boldsymbol{X})$, which is equivalently written by $\delta(S, \boldsymbol{X}) = aS$, $a > 0$. The estimator minimizing the risk among the equivariant estimators is called BEE (best equivariant estimator), just being δ_0.

Stein(1964) discovered the interesting result that the natural estimator δ_0 is improved on by using the information contained in $\boldsymbol{X}$. For finding a superior procedure, he considered a class of estimators equivariant under the scale transformation group $S \to c^2S$, $\boldsymbol{X} \to c\Gamma\boldsymbol{X}$, $\sigma^2 \to c^2\sigma^2$, $\boldsymbol{\theta} \to c\Gamma\boldsymbol{\theta}$, which is a subgroup of the affine group. The scale equivariant estimator is of the form

$$\delta_\phi = S\phi(W), \quad W = ||\boldsymbol{X}||^2/S,$$

where S and W are, respectively, equivariant and maximal invariant. Since $||\boldsymbol{X}||^2/\sigma^2$ has a noncentral chi square distribution $\chi_p^2(\lambda)$ with unknown noncentrality parameter $\lambda = ||\boldsymbol{\theta}||^2/\sigma^2$, there does not exist the best estimator among the scale equivariant δ_ϕ, but it is possible to find out an estimator dominating δ_0 within the class. Two approaches to the purpose are well known: the Stein and Brown-Brewster-Zidek methods.

2.2 The Stein method

The approach by Stein(1964) is to minimize the conditional expectation given $W = w$, $E_\lambda[L(\phi(w)S/\sigma^2)|W = w]$ with respect to ϕ. Let $u = ||\boldsymbol{X}||^2/\sigma^2$, $v = S/\sigma^2$ and denote their density functions by $f_p(u;\lambda)$ and $f_n(v)$, respectively. Then the optimal function $\phi^*_\lambda(W)$ is given by

$$\begin{aligned}\phi^*_\lambda(w) &= \frac{1}{E[S/\sigma^2|W=w]}\\ &= \frac{\int v f_p(vw;\lambda)f_n(v)dv}{\int v^2 f_p(vw;\lambda)f_n(v)dv}\\ &\le \frac{\int v f_p(vw;0)f_n(v)dv}{\int v^2 f_p(vw;0)f_n(v)dv}\\ &= \frac{\int v^{(n+p)/2-1}e^{-(1+w)v/2}dv}{\int v^{(n+p)/2}e^{-(1+w)v/2}dv} = \phi^*_0(w),\end{aligned}$$

where the above inequality follows from the fact that $f_p(u;\lambda)/f_p(u;0)$ is increasing in u. Since $\phi^*_0(w) = (1+w)/(n+p)$, letting

$$\phi_T(W) = \min\left\{\frac{1}{n}, \frac{1+W}{n+p}\right\} \tag{2.2}$$

guarantees the inequalities $\phi^*_\lambda(W) \le \phi_T(W) \le n^{-1}$. Hence from the convexity of the loss, we can assert that $E_\lambda[L(\phi_T(W)S/\sigma^2)|W] \le E_\lambda[L(n^{-1}S/\sigma^2)|W]$, yielding the inadmissibility result that the truncated estimator

$$\delta^{ST} = \delta_{\phi_T} = \min\left\{\frac{S}{n}, \frac{S+||\boldsymbol{X}||^2}{n+p}\right\}$$

has a uniformly smaller risk than δ_0. Such a truncated procedure obtained with this *Stein method* is here called the *Stein type*.

The Stein estimator δ^{ST} is interpreted as a preliminary-test estimator for the hypothesis $H : \boldsymbol{\theta} = \boldsymbol{0}$, that is, we have the estimator $(S + ||\boldsymbol{X}||^2)/(n+p)$ if H is accepted, and otherwise we take δ_0. Since the risk of δ_0 gets larger for smaller n, the estimator δ^{ST} is more effective for smaller n and larger p. The risk of δ^{ST} has the minimum value at $\lambda = 0$ and approaches that of δ_0 when λ tends to infinity.

It is interesting to indicate that δ^{ST} can be derived as an empirical Bayes estimator from the Bayesian aspect (Kubokawa, *et al.*(1992)). Let $\eta = 1/\sigma^2$ and let us suppose that $\boldsymbol{\theta}$ and η are random variables, θ having a conditional distribution given η such as

$$\boldsymbol{\theta}|\eta \sim \mathcal{N}_p(\boldsymbol{\theta}_0, \frac{1}{a\eta}\boldsymbol{I}_p),$$

where a is an unknown parameter and $\boldsymbol{\theta}_0$ is a known vector to be chosen beforehand. Also suppose that η has the noninformative prior $\eta^{-1}d\eta$. Then the Bayes estimator of σ^2 against this prior is given by

$$\hat{\sigma}^2_B(\boldsymbol{\theta}_0) = \frac{1}{E[\eta|\boldsymbol{X}, S]} = \frac{\tau||\boldsymbol{X}-\boldsymbol{\theta}_0||^2 + S}{n+p}$$

for $\tau = a/(1+a)$. Since τ is unknown, it is needed to be estimated from the marginal distribution of $(\boldsymbol{X}, S)$, which is written as

$$(const.) \times s^{n/2-1} \frac{\tau^{p/2}}{(\tau||\boldsymbol{x} - \boldsymbol{\theta}_0||^2 + s)^{(n+p)/2}}.$$

Taking the restriction $0 < \tau < 1$ into account, we see that MLE of τ is given by

$$\hat{\tau} = \min\left\{\frac{pS}{n||\boldsymbol{X} - \boldsymbol{\theta}_0||^2},\ 1\right\}.$$

By substituting $\hat{\tau}$ for τ in the Bayes rule $\hat{\sigma}^2_B(\boldsymbol{\theta}_0)$, we get the empirical Bays estimator

$$\hat{\sigma}^2_{EB}(\boldsymbol{\theta}_0) = \frac{\hat{\tau}||\boldsymbol{X} - \boldsymbol{\theta}_0||^2 + S}{n+p} = \min\left\{\frac{S}{n},\ \frac{||\boldsymbol{X} - \boldsymbol{\theta}_0||^2 + S}{n+p}\right\}$$

and $\hat{\sigma}^2_{EB}(\mathbf{0})$ is identical to δ^{ST}.

The value of $\boldsymbol{\theta}_0$ is given based on prior information and $\hat{\sigma}^2_{EB}(\boldsymbol{\theta}_0)$ has a large risk-reduction for $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$, so that in the case where one can guess or take the prior information about $\boldsymbol{\theta}$, $\hat{\sigma}^2_{EB}(\boldsymbol{\theta}_0)$ brings a good estimate. Even if one can not suppose any exact prior information, the risk of $\hat{\sigma}^2_{EB}(\boldsymbol{\theta}_0)$ is always less than that of δ_0 and it does not yield any actual harm from the frequentist's viewpoint, that is, $\hat{\sigma}^2_{EB}(\boldsymbol{\theta}_0)$ is robust for the prior information. The Bayes estimator depends on the prior knowledge completely while the knowledge is neglected in the usual procedures such as UMVUE and MLE. The empirical Bayes estimator is thus interpreted as an intermediate of the Bayes and usual ones such that the drawbacks of both estimators are made up for.

2.3 The Brown-Brewster-Zidek(BBZ) method

Another approach to improvement on δ_0 stems from Brown(1968), who divided the half line $[0, \infty)$ at $r > 0$ into two parts and considered the truncated estimator of the form

$$\hat{\sigma}^{2BR}(c) = \begin{cases} cS, & \text{if } W < r, \\ n^{-1}S, & \text{if } W \geq r. \end{cases}$$

When the constant c minimizing the conditional expectation $E_\lambda[L(cS/\sigma^2)|W < r]$ is denoted by $c = c^*_\lambda(r)$, we observe tat

$$\begin{aligned} c^*_\lambda(r) &= \frac{\int F_p(rv;\lambda) f_n(v)dv}{\int vF_p(rv;\lambda) f_n(v)dv} \\ &\leq \frac{\int F_p(rv;0) f_n(v)dv}{\int vF_p(rv;0) f_n(v)dv} \\ &= \frac{1}{n+p} \frac{\int_0^r \lambda^{p/2-1}/(1+\lambda)^{(n+p)/2} d\lambda}{\int_0^r \lambda^{p/2-1}/(1+\lambda)^{(n+p)/2+1} d\lambda} = c^*_0(r), \end{aligned} \tag{2.3}$$

where $F_p(x;\lambda) = \int_0^x f_p(t;\lambda)dt$, and the inequality in (2.3) follows from the monotonicity of $F_p(x;\lambda)/F_p(x;0)$. Note that $c^*_0(r)$ is increasing in r and that $c^*_0(r) \leq \lim_{r\to\infty} c^*_0(r) = n^{-1}$. From this fact and the convexity of the loss, it is seen that Brown's estimator $\hat{\sigma}^{2BR}(c^*_0(r))$ is better than δ_0.

Brewster and Zidek(1974) presented, on the basis of Brown's approach, an innovative idea of partitioning the half line $[0, \infty)$ into lots of parts. For a sequence of partitions $0 = r_{i,0} < r_{i,1} <$

$\cdots < r_{i,n_i-1} < r_{i,n_i} = \infty$ such that $\lim_{i\to\infty} r_{i,1} = 0$, $\lim_{i\to\infty} r_{i,n_i-1} = \infty$ and $\lim_{i\to\infty}\sup_j |r_{i,j} - r_{i,j-1}| = 0$, one can consider the corresponding sequence of estimators of the form

$$\hat{\sigma}_i^2 = \{c_0^*(r_{i,j})S;\quad r_{i,j-1} \le W < r_{i,j}\}_{j=1,\cdots,n_i},$$

which is guaranteed to dominate δ_0 since $c_0^*(r)$ is increasing in r.

As the number of the partitions tends to infinity, $\hat{\sigma}_i^2$ converges to

$$c_0^*(W)S = \delta^{BZ}, \quad \text{say},$$

which is referred to as the *Brewster-Zidek estimator*. From Fatou's lemma, it follows that

$$\begin{aligned} E[L(\delta^{BZ}/\sigma^2)] &= E[L(\lim_{i\to\infty} \hat{\sigma}_i^2/\sigma^2)] \\ &\le \underline{\lim}_{i\to\infty} E[L(\hat{\sigma}_i^2/\sigma^2)] \\ &\le E[L(\delta_0/\sigma^2)], \end{aligned}$$

which implies that the limiting value δ^{BZ} improves on δ_0. We here call this approach the *BBZ(Brown-Brewster-Zidek) method*, and the derived estimator the *BBZ type*.

Brewster and Zidek(1974) verified that δ^{BZ} is the generalized Bayes estimator against the hierarchical prior distribution

$$\boldsymbol{\theta}|\eta,\lambda \sim \mathcal{N}_p(\mathbf{0}, \frac{1-\lambda}{\lambda}\eta^{-1}\boldsymbol{I}_p),\ \lambda \sim \frac{1}{\lambda}I_{(0,1)}(\lambda)d\lambda,\ \eta \sim \frac{1}{\eta}d\eta$$

and that it is admissible within the class δ_ϕ. Proskin(1985) established its admissibility beyond the class. Thus δ^{BZ} is an admissible and minimax estimator improving on δ_0.

As pointed out by Rukhin(1987), the risk gain of δ^{BZ} is quite small for $p = 1$ while it is more effective for larger p. In contrast to δ^{ST}, the estimator δ^{BZ} has no risk gain at $\lambda = 0$ while the maximum of the risk gain is attained a bit far from zero.

2.4 A new unified approach

The unified method in derivation of the Stein type and BBZ type estimators was proposed by Kubokawa(1994a) and Takeuchi(1991). Since this method is based on an idea of expressing the difference of the risk functions via an integral, and we call it the *IERD (Integral Expression of Risk Difference) method.*

Let us suppose that the shrinkage function $\phi(\cdot)$ satisfies $\lim_{w\to\infty}\phi(w) = n^{-1}$. The difference of the risk functions of the estimators δ_0 and δ_ϕ is expressed, by using the definite integral argument, as follows:

$$\begin{aligned} &R(\theta,\sigma^2,\delta_0) - R(\theta,\sigma^2,\delta_\phi) \\ &= E[L(\phi(\infty)S/\sigma^2)] - E[L(\phi(W)S/\sigma^2)] \\ &= E[L(\phi(tW)S/\sigma^2)\Big|_{t=1}^{t=\infty}] \\ &= E[\int_1^\infty \frac{d}{dt}\{L(\phi(tW)S/\sigma^2)\}dt] \\ &= E[\int_1^\infty L'(\phi(tW)S/\sigma^2)\phi'(tW)WS/\sigma^2 dt] \\ &= \int_0^\infty\int_0^\infty\int_1^\infty L'(\phi(t\frac{u}{v})v)\phi'(t\frac{u}{v})u\,dt\,f_p(u;\lambda)f_n(v)du\,dv. \end{aligned} \tag{2.4}$$

Making the transformations $w = (t/v)u$, $x = (wv)/t$ in order with $dw = (t/v)du$, $dx = (wv)/t^2 dt$, the r.h.s. of the extreme equation in (2.4) is rewritten by

$$\begin{aligned} & R(\theta, \sigma^2, \delta_0) - R(\theta, \sigma^2, \delta_\phi) \\ = & \int_0^\infty \int_0^\infty \int_1^\infty L'(\phi(w)v)\phi'(w)\frac{wv^2}{t^2} f_p(\frac{wv}{t};\lambda) f_n(v) dt dv dw \\ = & \int_0^\infty \int_0^\infty \int_0^{wv} L'(\phi(w)v)\phi'(w) v f_p(x;\lambda) f_n(v) dx dv dw \\ = & \int_0^\infty \phi'(w) \int_0^\infty L'(\phi(w)v) v F_p(wv;\lambda) f_n(v) dv dw. \end{aligned}$$

The following inequality due to Strawderman(1974) is here useful for getting conditions on ϕ.

Lemma 2.1. *Let $h(x)$ be a nondecreasing and integrable function on interval (a,b), and let $\nu(\cdot)$ be a finite measure on (a,b). If for integrable function $K(x)$ on (a,b), there exists a point x_0 on (a,b) such that $K(x) \le 0$ for $x \le x_0$ and $K(x) \ge 0$ for $x > x_0$, then*

$$\int_a^b K(x)h(x)\nu(dx) \ge h(x_0) \int_a^b K(x)\nu(dx),$$

where the equality holds if and only if $h(x)$ is a constant almost everywhere.

Supposing that $\phi'(w) \ge 0$, from Lemma 2.1 and the monotonicity of $F_p(x;\lambda)/F_p(x;0)$, we get that

$$\begin{aligned} & R(\theta, \sigma^2, \delta_0) - R(\theta, \sigma^2, \delta_\phi) \\ \ge & \int_0^\infty \phi'(w) \frac{F_p(wv_0;\lambda)}{F_p(wv_0;0)} \left\{ \int_0^\infty L'(\phi(w)v) v F_p(wv;0) f_n(v) dv \right\} dw. \end{aligned}$$

For the sake of simplicity, let $F_p(x) = F_p(x;0)$. It is thus concluded that δ_ϕ improves on δ_0 if $\phi(w)$ satisfies the following conditions:

(a) $\phi(w)$ is nondecreasing and $\lim_{w\to\infty}\phi(w) = n^{-1}$,

(b) $\int_0^\infty L'(\phi(w)v) v F_p(wv) f_n(v) dv \ge 0$, that is, $\phi(w) \ge \phi_0(w)$, where

$$\phi_0(w) = \frac{\int_0^\infty F_p(wv) f_n(v) dv}{\int_0^\infty v F_p(wv) f_n(v) dv}.$$

Since $\phi_0(w) = c_0^*(w)$ for $c_0^*(r)$ given by (2.3), $\phi_0(W)S$ is just the Brewster-Zidek estimator δ^{BZ}. It can be checked that $\phi_0(w)$ and $\phi_T(w)$ given by (2.2) satisfy the conditions (a) and (b), so that this class includes δ^{BZ} and the Stein estimator δ^{ST}. Two types of the improved estimators δ^{ST} and δ^{BZ} derived separately in the previous subsections are thus obtained at once through the IERD method.

A weak point of δ^{BZ} is that it has the same risk as δ_0 at $\lambda = 0$, that is, $R(0, \sigma^2, \delta_0) = R(0, \sigma^2, \delta^{BZ})$, while δ^{ST} gives a substantial risk gain at $\lambda = 0$. For settling the weakness, Ghosh (1994) and Maruyama(1996) gave an alternative generalized Bayes estimator of the form

$$\delta_k^{GB} = \frac{S}{n+p+2(1-k)} \frac{\int_0^W \lambda^{p/2-k}/(1+\lambda)^{(n+p)/2-k+1} d\lambda}{\int_0^W \lambda^{p/2-k}/(1+\lambda)^{(n+p)/2-k+2} d\lambda}$$

against the prior distribution

$$\boldsymbol{\theta}|\eta,\lambda \sim \mathcal{N}_p(\mathbf{0}, \frac{1-\lambda}{\lambda}\eta^{-1}\boldsymbol{I}_p),\ \lambda \sim \frac{1}{\lambda^k} I_{(0,1)}(\lambda) d\lambda,\ \eta \sim \frac{1}{\eta^k} d\eta \quad \text{for} \quad \eta = 1/\sigma^2.$$

It is demonstrated that δ_k^{GB} belongs to our class of improved estimators for $1 \le k < p/2 + 1$. When $k = 1$, δ_1^{GB} is identical to δ^{BZ}.

We here provide a numerical comparison of risks for the above improved estimators. This is done on the basis of simulation with 50,000 replications in the case of $n = 4$ and $p = 10$. Figure 1 gives the risk performances of the estimators δ_0, δ^{ST}, δ^{BZ}, δ_3^{GB} and δ_4^{GB} for $k = 3, 4$, which are indicated there by UB, ST, BZ, GB(3) and GB(4), respectively. The axes of ordinate and abscissa in the figure designate, respectively, the values of the risk and the square root of the noncentrality parameter $||\boldsymbol{\theta}||/\sigma$. From the figure, it is revealed that the improvement of δ^{ST} is small for $||\boldsymbol{\theta}||/\sigma > 6$ and that δ^{BZ} has no risk gain at the origin. Also the figure indicates that the generalized Bayes estimators δ_3^{GB} and δ_4^{GB} eliminate these drawbacks and give substantial risk reductions in a wide range of the noncentrality parameter while their maximum risk reductions are smaller than those of δ^{ST} and δ^{BZ}.

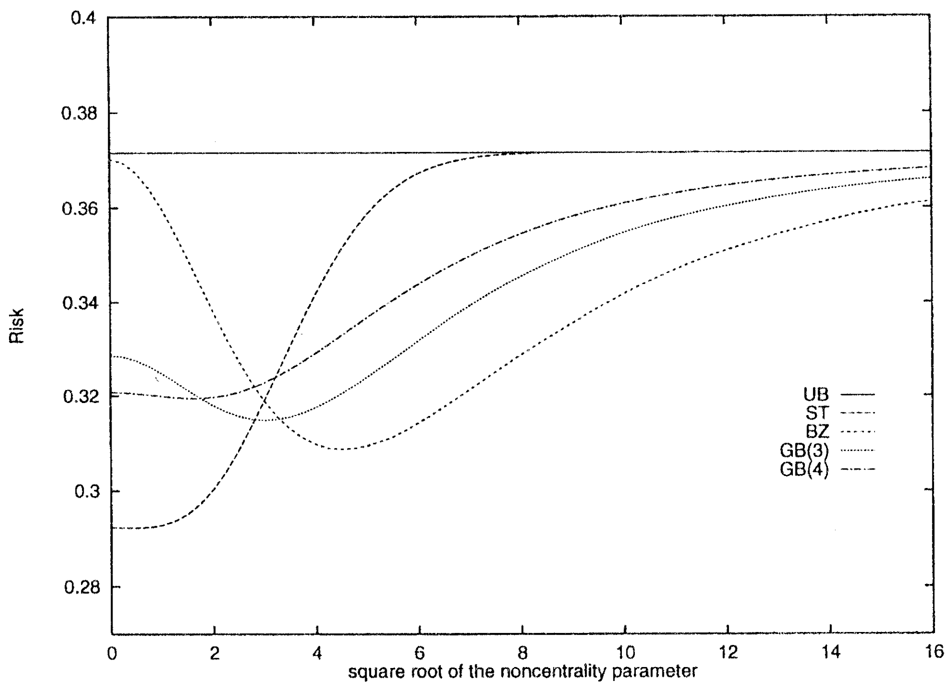

Figure 1. Risks of the estimators δ_0, δ^{ST}, δ^{BZ}, δ_3^{GB} and δ_4^{GB} for $n = 4$ and $p = 10$

It may be interesting to note that the problem of improving on the UMVUE of the variance is related to that of improving on the James-Stein estimator in the simultaneous estimation of the mean vector of the multivariate normal distribution, where the Stein type and BBZ type estimators, respectively, correspond to the positive-part Stein and Strawderman-Berger estimators (Kubokawa (1991, 94a)). Rukhin(1992b) showed that their two problems are asymptotically equivalent to that of estimating a positive mean of a normal distribution. Kubokawa *et al.*(1993b) demonstrated that the use of better estimators of the variance leads to the improvement on the James-Stein estimator.

2.5 Extensions to general distributions and applications to interval estimation

The use of the IERD method enables us to extend the result of the previous subsection to the Bowl-shaped loss functions and general distributions with monotone likelihood ratio properties, including normal, lognormal, exponential, Pareto and inverse Gaussian distributions. It is also applicable to construct improved confidence intervals.

Let S and T be mutually independent random variables whose density functions are given by

$$S/\sigma \sim g(v)I_{[v>0]}, \quad T/\sigma \sim h(u;\lambda)I_{[u>k(\lambda)]},$$

where $k(\cdot)$ is a real-valued function, σ is an unknown scale parameter, λ is an unknown real parameter and $I_{[\cdot]}$ is the indicator function. In the case of the exponential distribution $\sigma^{-1}\exp\{-(x-\mu)/\sigma\}I_{[x>\mu]}$, for instance, we have $\lambda = \mu/\sigma$ and $k(\lambda) = \lambda$, and for the normal distribution, $\lambda = ||\boldsymbol{\theta}||^2/\sigma^2$ and $k(\lambda) = 0$. Let $L(t)$ be a continuous and bowl-shaped function, that is, it is decreasing for $t < 1$ and increasing for $t > 1$. We address the issue of estimating the scale parameter σ by estimator $\delta = \delta(S,T)$ relative to the bowl-shaped loss function $L(\delta/\sigma)$.

Let $\delta_0 = c_0 S$ be the best estimator among multiples of S, and for improving on it, consider estimators of the form

$$\delta_\phi = \begin{cases} \phi(W)S, & \text{if } W > 0, \\ c_0 S, & \text{if } W \le 0, \end{cases}$$

where $W = T/S$. Denote $H(x;\lambda) = \int_0^x h(u;\lambda)I_{[u>k(\lambda)]}du$, $H(x) = H(x;0)$ and $h(x) = h(x;0)$, and assume that

(A.1) $H(x;\lambda)/H(x)$ is nondecreasing in $x > 0$.

Then applying the IERD method gives that δ_ϕ dominates δ_0 if

(a) $\phi(w)$ is nondecreasing and $\lim_{w\to\infty}\phi(w) = c_0$,

(b) $\int_0^\infty L'(\phi(w)v)vH(wv)g(v)dv \ge 0$.

Assuming further that for $c_1 > c_2 > 0$,

(A.2) $h(c_2 x)/h(c_1 x)$ is nondecreasing in x,

we can see that the BBZ type and Stein type estimators belong to this class (Kubokawa(1994a)). The assumptions (A.1) and (A.2) are satisfied by normal, lognormal, exponential and Pareto distributions. For the exponential distribution, the Stein type estimator was proposed by Arnold(1970), Zidek(1973), and the BBZ type estimator was given by Brewster(1974).

For an inverse Gaussian distribution, the Stein type and BBZ type estimators are derived by Pal and Sinha(1989) and MacGibbon and Shorrock(1994), respectively. Kourouklis(1997) recently demonstrated the monotone likelihood ratio property for the inverse Gaussian distribution, so that (A.1), (A.2) are satisfied and it belongs to the general framework given in this section. An improved truncated estimator of scale in a uniform distribution was derived by Rukhin *et al.*(1990). Rukhin(1991) demonstrated that the issue of estimating the scale in the general location-scale family is asymptotically equivalent to that of estimating a function including a quadratic form of a random variable, and provided the Stein type and BBZ type estimators with asymptotical improvements.

For interval estimation of the variance, we can obtain the results corresponding to the case of the point estimation. Confidence intervals proposed by Tate and Klett(1959) are of the form $I_0 = [aS, bS]$, where constants a and b satisfy the requirement $P[aS < \sigma < bS] = 1-\alpha$ for confidence coefficient $0 < 1-\alpha < 1$. Several criteria are employed in order to specify the constants uniquely. One is to minimize the length $b-a$ and the resulting confidence interval is refereed to as the *Minimum Length Confidence Interval (MLCI)*. The other is to minimize the ratio b/a and the optimal values of a and b satisfy the equation $a^{-1}g(a^{-1}) = b^{-1}g(b^{-1})$, giving the *Minimum Ratio Confidence Interval (MRCI)*. We here address the issue of constructing a confidence interval improving on the MRCI.

When a confidence interval of the form

$$I_\phi = \begin{cases} [a\phi(W)S, b\phi(W)S], & \text{if } W > 0, \\ I_0, & \text{if } W \le 0 \end{cases}$$

is considered, it has the same ratio of the endpoints as I_0.

Using the IERD method, Kubokawa(1994a) showed that I_ϕ improves on the MRCI I_0 in the sense of maximizing the coverage probability if we assume (A.1) and if

(a) $\phi(w)$ is nondecreasing and $\lim_{w\to\infty}\phi(w)=1$,

(b) $\frac{1}{a}g(\frac{1}{a\phi})H(\frac{w}{a\phi}) \geq \frac{1}{b}g(\frac{1}{b\phi})H(\frac{w}{b\phi})$.

The essential inequality utilized in the proof is given in the following lemma which corresponds to Lemma 2.1.

Lemma 2.2. *Assume that $h(x)/g(x)$ is nondecreasing for positive functions $g(x)$ and $h(x)$. Then for c_1, $c_2 > 0$ and $0 < x_1 < x_2$,*

$$c_2h(x_2) - c_1h(x_1) \geq \frac{h(x_1)}{g(x_1)}\{c_2g(x_2) - c_1g(x_1)\}.$$

The Stein type and BBZ type confidence intervals can be obtained under the assumption (A.2). The Stein type confidence intervals in normal and exponential distributions were derived by Nagata(1989, 91). In the normal case, for instance, it is given by

$$\begin{aligned} I^{ST} &= [a\phi_T(W)S,\ b\phi_T(W)S] \\ \phi_T(W) &= \min\left\{1,\ \frac{a^{-1}-b^{-1}}{\log(b/a)}\frac{1+W}{n+p}\right\}. \end{aligned}$$

On the other hand, the improvements on the MLCI was stemmed from Cohen(1972), and the BBZ type confidence interval and its generalized Bayesness was studied by Shorrock(1990). Goutis and Casella(1991) obtained the BBZ type confidence interval improving on the MLCI in both senses of the length and the coverage probability and verified its generalized Bayesness.

2.6 Other improved estimators

Besides the Stein and BBZ methods, there are other methods for deriving improved estimators. In this subsection, we state briefly the methods given by Strawderman(1974) and Shinozaki(1995).

A shrinkage estimator Strawderman(1974) treated is of the smooth form

$$\delta_\phi^{STD} = \frac{1}{n}\left\{1 - \phi\left(\frac{1}{1+W}\right)\frac{1}{(1+W)^\varepsilon}\right\}S$$

for nondecreasing function $\phi(\cdot)$. The conditions on ϕ and ε for the estimator δ_ϕ^{STD} being better than δ_0 can be derived. Mathew *et al.*(1992) applied this method to the estimation problem of variance components in mixed effects models.

Another approach was tried by Shinozaki(1995), who rewrote the Stein estimator δ^{ST} as

$$\delta^{ST} = \frac{S}{n} - \frac{p}{n+p}\left(\frac{S}{n} - \frac{||\boldsymbol{X}||^2}{p}\right)^+, \quad a^+ = \max(0,a),$$

and, beyond the class of equivariant estimators, proposed three types of non-equivariant estimator shrinking S/n toward $\max(1, ||\boldsymbol{X}||^2/p)$. One of them is given by

$$\delta_\phi^{SN} = \frac{S}{n} - \frac{p}{n+p}\phi\left(\frac{S}{S+||\boldsymbol{X}||^2}\right)\left\{\frac{S}{n} - \max\left(1, \frac{||\boldsymbol{X}||^2}{p}\right)\right\}^+$$

where $a^+ = \max(a, 0)$. The estimator δ_ϕ^{SN} has a uniformly smaller risk than δ_0 under some conditions on ϕ. Also it is revealed that the actual risk-gains of the non-equivariant estimators are much larger than those of the Stein estimator. In the case of $n = 3$, $p = 1$ and $\sigma^2 = 1$, for instance, the relative risk improvements of δ^{ST} and δ_ϕ^{SN} over δ_0 are, respectively, 4.23% and 14.22% for $\theta = 0$ and 3.98% and 11.03% for $\theta = 0.5$. While the Stein estimator shrinks too much $n^{-1}S$ towards $p^{-1}||\boldsymbol{X}||^2$ when $\sigma^2 > 1 > ||\boldsymbol{X}||^2/p$, the proposed estimator δ_ϕ^{SN} shrinks it to $\max(1, p^{-1}||\boldsymbol{X}||^2)$ and this explains why δ_ϕ^{SN} is more efficient.

2.7 Double shrinkage estimation of ratio of variances

The problem related to the variance estimation is to derive double shrinkage procedures for the ratio of the variances. Let $\boldsymbol{X}_1$, S_1, $\boldsymbol{X}_2$ and S_2 be mutually independent random variables where for $i = 1, 2$,

$$\boldsymbol{X}_i \sim \mathcal{N}_{p_i}(\boldsymbol{\theta}_i, \sigma_i^2 \boldsymbol{I}_{p_i}), \quad S_i \sim \sigma_i^2 \chi_{n_i}^2.$$

When it is supposed that we want to estimate the ratio of the variances $\rho = \sigma_2^2/\sigma_1^2$ relative to the Kullback-Leibler loss $\hat{\rho}/\rho - \hat{\rho}/\rho - 1$, the best estimator among multiples of S_2/S_1 is given by

$$\hat{\rho}_0 = \frac{S_2}{n_2} \frac{n_1 - 2}{S_1}.$$

For improving on $\hat{\rho}_0$ by using $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, Gelfand and Dey(1988) considered the estimators

$$\begin{aligned} \hat{\rho}_1 &= \frac{S_2}{n_2} \max\left\{ \frac{n_1 - 2}{S_1}, \frac{n_1 + p_1 - 2}{S_1 + ||\boldsymbol{X}_1||^2} \right\}, \\ \hat{\rho}_2 &= \min\left\{ \frac{S_2}{n_2}, \frac{S_2 + ||\boldsymbol{X}_2||^2}{n_2 + p_2} \right\} \frac{n_1 - 2}{S_1}, \end{aligned}$$

which may be called *single shrinkage estimators* since they use either of the statistics $||\boldsymbol{X}_1||^2$ and $||\boldsymbol{X}_2||^2$. Although it had been desired to develop a *double shrinkage improved procedure* utilizing both of them, it was known to be technically difficult, for it shrinks in opposite directions. Recently Kubokawa(1994b) resolved this issue by utilizing the IERD method. The resulting double shrinkage estimator is

$$\begin{aligned} \hat{\rho}_3^{DS} &= \hat{\rho}_1 + \hat{\rho}_2 - \hat{\rho}_0 \\ &= \hat{\rho}_1 + \min\left\{ 0, \frac{n_1 - 2}{n_2 + p_2} \frac{S_2 + ||\boldsymbol{X}_2||^2}{S_1} - \hat{\rho}_0 \right\} \\ &= \hat{\rho}_2 + \max\left\{ 0, \frac{n_1 + p_1 - 2}{n_2} \frac{S_2}{S_1 + ||\boldsymbol{X}_1||^2} - \hat{\rho}_0 \right\}, \end{aligned}$$

which dominates both of $\hat{\rho}_1$ and $\hat{\rho}_2$. The second terms in the r.h.s. of the second and the third equations are interpreted as modification parts against over-shrinkage of $\hat{\rho}_1$ and $\hat{\rho}_2$, respectively. Kubokawa and Srivastava(1996) established the natural double shrinkage estimator improving the single shrinkage ones, given by

$$\hat{\rho}_4^{DS} = \min\left\{ \frac{S_2}{n_2}, \frac{S_2 + ||\boldsymbol{X}_2||^2}{n_2 + p_2} \right\} \times \max\left\{ \frac{n_1 - 2}{S_1}, \frac{n_1 + p_1 - 2}{S_1 + ||\boldsymbol{X}_1||^2} \right\},$$

and also gave the improved and generalized Bayes estimator $\hat{\rho}^{GB} = \phi_2^*(W_2)S_2/\{\phi_1^*(W_1)S_1\}$ where for $i = 1, 2$, $W_i = ||\boldsymbol{X}_i||^2/S_i$ and

$$\phi_i^*(w_i) = \frac{1}{n_i + p_i + 2(i-2)} \frac{\int_0^{w_i} z^{\frac{p_i}{2}-1}/(1+z)^{\frac{n_i+p_i}{2}+(i-2)} dz}{\int_0^{w_i} z^{\frac{p_i}{2}-1}/(1+z)^{\frac{n_i+p_i}{2}+(i-1)} dz}.$$

We provide Monte Carlo simulation results for the risk functions of the estimators $\hat{\rho}_0$, $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3^{DS}$, $\hat{\rho}_4^{DS}$, $\hat{\rho}^{GB}$. The simulation experiments are done in the cases of $n_i = 3; p_i = 10; \sigma_i^2 = 1; \lambda_i = ||\boldsymbol{\theta}_i||^2/\sigma_i^2 = 0.0, 0.5, 1.0, 5.0, 10.0$ for $i = 1, 2$. Table 1 reports the average values of the risks based on 50,000 replications. From the table, it is revealed that $\hat{\rho}_4^{DS}$ and $\hat{\rho}^{GB}$ have smaller risks than others, and that $\hat{\rho}^{GB}$ is the best estimator with significant improvements. It is also indicated that the risk gain of $\hat{\rho}_1$ is much greater than that of $\hat{\rho}_2$. This may arise from the unstableness of the denominator of $\hat{\rho}_0$ in comparison with the numerator. That is, the simulation result for $\hat{\rho}_1$ and $\hat{\rho}_2$ implies that stabilizing the denominator yields a more improvement than stabilizing the numerator.

Table 1. Expected kullback-Leibler losses of $\hat{\rho}_0$, $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3^{DS}$, $\hat{\rho}_4^{DS}$ and $\hat{\rho}^{GB}$ for $n_1 = n_2 = 3$ and $p_1 = p_2 = 10$

λ_1	λ_2	$\hat{\rho}_0$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3^{DS}$	$\hat{\rho}_4^{DS}$	$\hat{\rho}^{GB}$
	0.0	1.075	0.749	1.016	0.593	0.566	0.530
	0.5	1.075	0.749	1.016	0.600	0.572	0.536
0.0	1.0	1.075	0.749	1.016	0.606	0.578	0.541
	5.0	1.075	0.749	1.025	0.651	0.624	0.584
	10.0	1.075	0.749	1.039	0.689	0.669	0.635
	0.0	1.075	0.750	1.016	0.596	0.574	0.530
	0.5	1.075	0.750	1.016	0.603	0.580	0.535
0.5	1.0	1.075	0.750	1.016	0.609	0.585	0.540
	5.0	1.075	0.750	1.025	0.653	0.629	0.581
	10.0	1.075	0.750	1.039	0.691	0.672	0.629
	0.0	1.075	0.751	1.016	0.600	0.582	0.530
	0.5	1.075	0.751	1.016	0.606	0.587	0.535
1.0	1.0	1.075	0.751	1.016	0.612	0.593	0.539
	5.0	1.075	0.751	1.025	0.655	0.634	0.579
	10.0	1.075	0.751	1.039	0.692	0.676	0.625
	0.0	1.075	0.778	1.016	0.645	0.650	0.541
	0.5	1.075	0.778	1.016	0.650	0.654	0.544
5.0	1.0	1.075	0.778	1.016	0.654	0.657	0.547
	5.0	1.075	0.778	1.025	0.691	0.686	0.574
	10.0	1.075	0.778	1.039	0.724	0.717	0.608
	0.0	1.075	0.829	1.016	0.713	0.730	0.566
	0.5	1.075	0.829	1.016	0.717	0.732	0.567
10.0	1.0	1.075	0.829	1.016	0.721	0.733	0.569
	5.0	1.075	0.829	1.025	0.751	0.754	0.586
	10.0	1.075	0.829	1.039	0.779	0.779	0.610

2.8 Extensions to estimation of multidimensional parameters

As one of extensions to estimation of multidimensional parameters, this subsection briefly deals with estimation of a covariance matrix and the generalized variance in a multivariate linear regression model whose canonical form is given by $p \times p$ matrix $\boldsymbol{S}$ and $p \times r$ matrix $\boldsymbol{X}$, which are mutually independent random variables and

$$\boldsymbol{S} \sim \mathcal{W}_p(n, \Sigma), \quad \boldsymbol{X} \sim \mathcal{N}_{p\times r}(\Theta, \Sigma \otimes \boldsymbol{I}_r)$$

where $\mathcal{W}_p(n, \Sigma)$ designates a Wishart distribution with mean $n\Sigma$ and $\otimes$ denotes the Kronecker product.

For the point estimation of the generalized variance $|\Sigma|$, Shorrock and Zidek(1976) derived the Stein type improved estimator with using the Zonal polynomials and Sinha(1976) provided another proof of it without the Zonal polynomials. A series expression of its risk function and numerical comparisons were given by Sugiura and Konno(1988). To the contrary, it is not easy to get the BBZ type estimator for $r \geq 2$, since the maximal invariant statistic $\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X}$ is not one-dimensional. The Stein type procedure in the interval estimation was given by Sarkar(1989).

It is interesting to note that $|\boldsymbol{S}| = \prod_{i=1}^{p} y_i$ where $y_1, \ldots, y_p$ are mutually independent random variables, $y_i \sim \sigma_i \chi^2_{n-i+1}$ for $|\Sigma| = \prod_{i=1}^{p} \sigma_i$. This implies that the estimation of $|\Sigma|$ is expressed by estimation of the product of scale parameters of independent gamma distributions. Using the results of Berger(1976a,b), Rukhin and Sinha(1991) proved the interesting fact that the usual estimator of $|\Sigma|$ can be improved without the information in $\boldsymbol{X}$ when $p \geq 4$, and that for $p \leq 3$, it is admissible within the subclass of estimators depending only on $y_1, \ldots, y_p$.

For estimation of the covariance matrix Σ, the Stein type estimator was derived by Sinha and Ghosh(1987). In the special case of $r = 1$, Perron(1990) proposed other types of improved estimators, Kubokawa *et al.*(1992) gave an empirical Bayes rule improving the Sinha-Ghosh estimator, and Kubokawa *et al.*(1993a) clarified the structure of the estimation of Σ and obtained a generalized Bayes improved estimator. Similar to the case of $|\Sigma|$, it is recognized as a hard work to derive BBZ type minimax or generalized Bayes minimax estimators when $r \geq 2$, and so some important problems remain still open. Inadmissibility results of usual or minimax estimators of Σ without using $\boldsymbol{X}$ are of course well known, so that it will be of great interest to construct minimax procedures with incorporating $\boldsymbol{X}$.

3 Related Problems Where Modifications of Estimators Are Required

Studies on the shrinkage procedures, from the decision-theoretic viewpoint, have produced a huge amount of beautiful and wonderful benefits especially in the estimation problems of the variance, the covariance matrix and the mean vector. From the practical aspects, on the other hand, the concept of shrinkage has been recognized as an important tool to stabilize estimates in various small-samples problems such that the data smoothing, small-area estimation, estimation of mortality rates and others. Beyond some familiar theoretical or applied issues, we are really faced with lots of situations where the usual plain estimators are required to be shrunken or modified. In such situations, it is of importance to clarify how much usual procedures should be shrunken or modified. Empirical and hierarchical Bayes rules, cross-validation (or leaving-one-out), bootstrap and penalized methods and Wavelet analysis are standard techniques useful for giving appropriate extents of the shrinkage or modification. In the decision-theoretic sense, it may be reasonable to provide the extent of shrinkage such that the shrinkage estimator has a uniformly smaller risk than the usual plain estimator. The IERD method may be expected as one of tools useful for the purpose. We here enumerate several problems where the usual estimators should be shrunken or modified, and for a few issues, we present reasonable shrinkage procedures derived through the IERD method.

3.1 Estimation of variance components

Mixed linear models or variance components models have been effectively and extensively employed in practical data-analysis. When the statistical inference for regression coefficients is implemented, estimators of the variance components are used to get two-stage procedures such that two-stage generalized least squares (2GLS) estimators and 2GLS tests.

For simplicity, consider the one-way random effects model with equal replications:

$$\begin{aligned} y_{ij} &= \mu + a_i + e_{ij}, \quad i = 1, \dots, k, \; j = 1, \dots, n \\ a_i &\sim \mathcal{N}(0, \sigma_A^2), \quad e_{ij} \sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

where $\{a_i\}$ and $\{e_{ij}\}$ are mutually independent and μ, σ_A^2, σ^2 are unknown parameters. Let $\overline{Y}_i = \Sigma_{j=1}^n Y_{ij}/n$, $\overline{Y} = \Sigma_{i=1}^k \Sigma_{j=1}^n Y_{ij}/(nk)$, $S_1 = \Sigma_{i=1}^k \Sigma_{j=1}^n (Y_{ij} - \overline{Y}_i)^2$ and $S_2 = n\Sigma_{j=1}^k (\overline{Y}_i - \overline{Y})^2$. $\overline{Y}$, S_1 and S_2 are mutually independent and the minimal sufficient statistics having

$$\begin{aligned} \overline{Y} &\sim \mathcal{N}(\mu, (\sigma^2 + n\sigma_A^2)/(nk)), \\ S_1 &\sim \sigma^2 \chi_{\nu_1}^2, \qquad \nu_1 = k(n-1), \\ S_2 &\sim (\sigma^2 + n\sigma_A^2)\chi_{\nu_2}^2, \qquad \nu_2 = k - 1. \end{aligned}$$

The UMVU(ANOVA) estimators of σ_A^2 and σ^2 is, respectively, given by

$$\hat{\sigma}_A^{2U} = \frac{1}{n}\left(\frac{S_2}{\nu_2} - \frac{S_1}{\nu_1}\right)$$

and $\hat{\sigma}^{2U} = \nu_1^{-1} S_1$, $\hat{\sigma}_A^{2U}$ possessing a critical drawback of taking negative values with a positive probability. Much effort has been devoted to this issue and reasonable procedures eliminating this undesirable property have been proposed. Of these, LaMotte (1973) showed that unbiased nonnegative quadratic estimators of σ_A^2 do not exist. Kleffe and Rao (1986) demonstrated that nonnegative biased quadratic estimators of σ_A^2 fail the minimum condition of consistency as n remains fixed, but $k \to \infty$. These persuade us to pay our attention to nonnegative estimators other than the quadratic ones.

The Kullback-Leibler loss function in estimating (σ^2, σ_A^2) simultaneously is given by

$$L(\hat{\sigma}^2, \hat{\sigma}_A^2, \sigma^2, \sigma_A^2) = \nu_1\left\{\frac{\hat{\sigma}^2}{\sigma^2} - \log\frac{\hat{\sigma}^2}{\sigma^2} - 1\right\} + \nu_2\left\{\frac{\hat{\sigma}^2 + n\hat{\sigma}_A^2}{\sigma^2 + n\sigma_A^2} - \log\frac{\hat{\sigma}^2 + n\hat{\sigma}_A^2}{\sigma^2 + n\sigma_A^2} - 1\right\}.$$

Using the IERD method, Kubokawa *et al.* (1996) proved that estimators of the forms

$$\begin{aligned} \hat{\sigma}^2(\psi) &= S_1 \psi\left(\frac{S_2}{S_1}\right) \\ \hat{\sigma}_A^2(\phi, \psi; n) &= \frac{1}{n}\left\{S_2 \phi\left(\frac{S_1}{S_2}\right) - S_1 \psi\left(\frac{S_2}{S_1}\right)\right\} \end{aligned}$$

improve on $\hat{\sigma}^{2U}$, $\hat{\sigma}_A^{2U}$ relative to the Kullback-Leibler loss if the following conditions hold:

(a) $\psi(w)$ is nondecreasing and $\lim_{w\to\infty}\psi(w) = \nu_1^{-1}$,

(b) $\psi(w) \geq \psi_0(w)$ where

$$\psi_0(w) = \frac{1}{\nu_1 + \nu_2} \frac{\int_0^w x^{\nu_2/2-1}/(1+x)^{(\nu_1+\nu_2)/2} dx}{\int_0^w x^{\nu_2/2-1}/(1+x)^{(\nu_1+\nu_2+2)/2} dx},$$

(c) $\phi(w)$ is nondecreasing and $\phi(0) = \nu_2^{-1}$,

(d) $\phi(w) \leq \phi_0(w)$ where

$$\phi_0(w) = \frac{1}{\nu_1 + \nu_2} \frac{\int_w^\infty x^{\nu_1/2-1}/(1+x)^{(\nu_1+\nu_2)/2} dx}{\int_w^\infty x^{\nu_1/2-1}/(1+x)^{(\nu_1+\nu_2+2)/2} dx}.$$

This class includes improved estimators $(\hat{\sigma}^{2EB}, \hat{\sigma}_A^{2EB})$, $(\hat{\sigma}^{2PT}, \hat{\sigma}_A^{2PT})$ and $(\hat{\sigma}^{2GB}, \hat{\sigma}_A^{2GB})$, where

$$\begin{aligned}
\hat{\sigma}^{2EB} &= \min\left\{\frac{S_1}{\nu_1}, \frac{S_1+S_2}{\nu_1+\nu_2}\right\}, \\
\hat{\sigma}_A^{2EB} &= \frac{1}{n}\max\left\{\frac{S_2}{\nu_2}-\frac{S_1}{\nu_1}, 0\right\}, \\
\hat{\sigma}^{2PT} &= \min\left\{\frac{S_1}{\nu_1}, \frac{S_1+\nu_2 S_2/(\nu_2+2)}{\nu_1+\nu_2}\right\}, \\
\hat{\sigma}_A^{2PT} &= \frac{1}{n}\left[\max\left\{\frac{S_2}{\nu_2}, \frac{S_1+S_2}{\nu_1+\nu_2-2}\right\} - \hat{\sigma}^{2PT}\right], \\
\hat{\sigma}^{2GB} &= \frac{S_1}{\nu_1+\nu_2}\frac{\int_0^{S_2/S_1} x^{\nu_2/2-1}/(1+x)^{(\nu_1+\nu_2)/2}dx}{\int_0^{S_2/S_1} x^{\nu_2/2-1}/(1+x)^{(\nu_1+\nu_2)/2}dx}. \\
\hat{\sigma}_A^{2GB} &= \hat{\sigma}_A^2(\phi_0, \psi_0; n) = \frac{1}{n}\left[\frac{S_2}{\nu_1+\nu_2}\frac{\int_{S_1/S_2}^{\infty} x^{\nu_1/2-1}/(1+x)^{(\nu_1+\nu_2)/2}dx}{\int_{S_1/S_2}^{\infty} x^{\nu_1/2-1}/(1+x)^{(\nu_1+\nu_2+2)/2}dx} - \hat{\sigma}^{2GB}\right].
\end{aligned}$$

It is interesting to note that the estimator $(\hat{\sigma}^{2EB}, \hat{\sigma}_A^{2EB})$ is not only given as a restricted (residual) maximum likelihood estimator but also obtained as an empirical Bayes rule from a Bayesian aspect. Also $(\hat{\sigma}^{2GB}, \hat{\sigma}_A^{2GB})$ is derived as a generalized Bayes rule. $\hat{\sigma}_A^{2GB}$ and $\hat{\sigma}_A^{2PT}$ are improved procedures taking positive values almost everywhere. Although the above results are limited to the cases of equal replications, they can be extended to general unbalanced cases (Kubokawa *et al.*(1996)).

When the mean squared error is taken as a criterion for comparing estimators, Kubokawa *et al.*(1993d) and Kubokawa(1995) applied the IERD method to get improved and positive estimators of σ_A^2 in the balanced and unbalanced cases. Especially the conjecture of Portnoy(1971) concerning the estimation of σ^2 was resolved in Kubokawa *et al.*(1993d).

An analogized problem is arisen in estimation of a covariance component in a multivariate mixed linear model, and non-negative definite estimators of the covariance component have been derived by Amemiya(1985), Mathew *et al.*(1994) and Remadi and Amemiya(1994). Calvin and Dykstra(1991) gave an algorithm for computing the restricted MLE. However a lot of problems remain to be settled from the decision-theoretic view point.

3.2 Estimation of the noncentrality parameter or the SN ratio

The SN(Signal-Noise) ratio is a criterion for comparing measuring instruments or measurement techniques in quality control. When the normality of error terms and the homogeneity of variances can be supposed, Miwa(1979) indicated that the estimation of the SN ratio is essentially equivalent to estimation of the noncentrality parameter, which is $\lambda = ||\boldsymbol{\theta}||^2/\sigma^2$ in the mode (2.1). The UMVU estimator is $\delta_0 = (n-2)||\boldsymbol{X}||^2/S - p$, having a drawback of taking negative values similar to the previous subsection. Some devises for this issue have been made in the literature (Neff and Strawderman(1976)). As shown by Kubokawa *et al.*(1993), using the IERD method gives that a shrinkage estimator of the form

$$\delta_\phi = (n-2)\frac{||\boldsymbol{X}||^2}{S} - \phi\left(\frac{||\boldsymbol{X}||^2}{S}\right)$$

improves on δ_0 if

(a) $\phi(w)$ is nondecreasing and $\lim_{w\to\infty}\phi(w) = p$,

(b) $\phi(w) \geq \phi_0(w)$, where

$$\phi_0(w) = (n-2)\frac{\int_0^w \int vtf_p(vt)f_n(v)dvdt}{\int_0^w \int vf_p(vt)f_n(v)dvdt}.$$

This class includes the positive and improved estimator

$$\delta^{PT} = \max\left\{\delta_0, \frac{2(n-2)}{p+2}\frac{||\boldsymbol{X}||^2}{S}\right\}.$$

When the variance σ^2 is known, the object is to estimate the noncentrality of a noncentral chi square distribution. Some topics have been studied by Perlman and Rasmussen(1975) for an empirical Bayes approach, by Chow(1987) for a complete class theorem, by Saxena and Alam(1982) for inadmissibility of the MLE. Kubokawa *et al.*(1993c) provided the improved and positive estimator $\max\{\delta_0, 2(p+2)^{-1}||X||^2\sigma^{-2}\}$ through the IERD method. Recently Shao and Strawderman(1995) succeeded in derivation of an estimator dominating $\max(\delta_0, 0)$ by using the same arguments as in the proof of improvement on the positive-part Stein estimator. A related issue is an estimation of the multiple correlation coefficient and it was studied by Muirhead(1985), Leung and Muirhead(1987) and others.

3.3 The linear calibration and statistical control problems

The linear calibration is the problem of inverse estimation in regression and one of useful statistical procedures. It is supposed that explanatory variable x and explained variable $\boldsymbol{y}$ have the linear relationship such that $E[\boldsymbol{y}|x] = \boldsymbol{\alpha} + \boldsymbol{\beta}x$. First, data $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ are observed for known scalar values $x_1, \dots, x_n$ with the linear regression model

$$\boldsymbol{y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}x_i + \boldsymbol{e}_i, \quad i = 1, \dots, n.$$

Next for unknown scalar value x_0, let us get k observations as

$$\boldsymbol{y}_{0j} = \boldsymbol{\alpha} + \boldsymbol{\beta}x_0 + \boldsymbol{e}_{0j}, \quad j = 1, \dots, k.$$

Here $\boldsymbol{e}_i$ and $\boldsymbol{e}_{0j}$ are mutually independent and have $\mathcal{N}_p(\boldsymbol{0}, \sigma^2\boldsymbol{I}_p)$ with unknown error variance σ^2, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are p-dimensional unknown vectors and $\boldsymbol{y}_i$ and $\boldsymbol{y}_{0j}$ are p-dimensional vectors. Then the *linear calibration* is to estimate x_0 inversely from the data $\boldsymbol{y}_i$ and $\boldsymbol{y}_{0j}$. The problem of estimating x_0 in the above model can be reduced to that of estimating x in the followings: Let $\boldsymbol{y}$, $\boldsymbol{z}$ and $\boldsymbol{T}$ be mutually independent random variables having

$$\boldsymbol{y} \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2\boldsymbol{I}_p), \quad \boldsymbol{z} \sim \mathcal{N}_p(\boldsymbol{\theta}x, \sigma^2\boldsymbol{I}_p), \quad \boldsymbol{T} \sim \mathcal{W}_p(n+k-3, \sigma^2\boldsymbol{I}_p),$$

where $\boldsymbol{\theta}$, x and σ^2 are unknown parameters. Further $\boldsymbol{T}$ can be reduced to $s = \mathrm{tr}\boldsymbol{T}$ with $s \sim \sigma^2\chi^2_q$, $q = p(n+k-3)$.

Two opposed procedures have been in the controversy in the inverse estimation of x. One is given by x minimizing $||\boldsymbol{z} - \boldsymbol{y}x||^2$, that is,

$$\delta_C = \frac{\boldsymbol{y}'\boldsymbol{z}}{||\boldsymbol{y}||^2},$$

called the *classical estimator*. Although this is consistent, it has the undesirable property that there does not exist its mean squared error for $p = 1, 2$. Against this estimator, Krutchkoff(1967) considered to derive $\boldsymbol{\gamma}$ minimizing $\Sigma_{i=1}^n\{\boldsymbol{\gamma}'(\boldsymbol{y}_i - \overline{\boldsymbol{y}}) - (x_i - \overline{x})\}^2$ in the inverse regression and proposed

$$\delta_I = \frac{\boldsymbol{y}'\boldsymbol{T}^{-1}\boldsymbol{z}}{1 + \boldsymbol{y}'\boldsymbol{T}^{-1}\boldsymbol{y}},$$

called the *inverse regression estimator*. However, δ_I is not consistent while its moments are finite. Hoadley(1970) verified the generalized Bayesness of δ_I for $p = 1$, and for $p \geq 1$, Kubokawa and Robert(1994) showed

$$\delta_B = \frac{\boldsymbol{y}'\boldsymbol{z}}{s + ||\boldsymbol{y}||^2}$$

is proper Bayes, so that admissible. For eliminating the inconsistency of δ_B, Miwa(1985) proposed the generalized inverse regression estimator

$$\delta^{MW}(d) = \frac{\boldsymbol{y}'\boldsymbol{z}}{ds/q + ||\boldsymbol{y}||^2}$$

satisfying the consistency, and gave the best d when the SN ratio $||\boldsymbol{\theta}||^2/\sigma^2$ is large.

It is interesting to point out the relationship between the linear calibration and the statistical control problems. Note that a mean squared error of the estimator of the general form $\delta_g = g(||\boldsymbol{y}||^2/s)\boldsymbol{y}'\boldsymbol{z}/s$ for nonnegative function $g(\cdot)$ is written as

$$E[(\delta_g - x)^2] = E\left[g^2\left(\frac{||\boldsymbol{y}||^2}{s}\right)\frac{||\boldsymbol{y}||^2}{s^2}\sigma^2\right] + x_0^2 E\left[\left\{g\left(\frac{||\boldsymbol{y}||^2}{s}\right)\frac{\boldsymbol{y}'\boldsymbol{\theta}}{s} - 1\right\}^2\right],$$

so that the problem is decomposed into two parts:

(I) minimizing $g(t)$,

(II) minimizing $E\left[\{g\left(||\boldsymbol{y}||^2/s\right)\boldsymbol{y}'\boldsymbol{\theta}/s - 1\}^2\right]$.

The part (II) is called the statistical control problem, which has been studied by Takeuchi(1968), Zaman(1981), Berliner(1983), Berger *et al.*(1982) and others.

When $\boldsymbol{y}$ is near $\mathbf{0}$, the classical estimator is unstable and should be modified. Kubokawa and Robert(1994) treated the shrinkage estimator

$$\delta_\phi = \left\{1 - \phi\left(\frac{||\boldsymbol{y}||^2}{s}\right)\right\}\frac{\boldsymbol{y}'\boldsymbol{z}}{||\boldsymbol{y}||^2}$$

and provided the conditions for δ_ϕ improving on δ_C through the IERD method. The resulting improved estimator is given by

$$\delta^{KR} = \min\left\{\frac{1}{||\boldsymbol{y}||^2}, \frac{q+p-2}{s+||\boldsymbol{y}||^2}\right\}\boldsymbol{y}'\boldsymbol{z},$$

which possesses the finite moments and the consistency even for $p \geq 1$.

A multivariate linear calibration has been studied by Brown(1982), Nishii and Krishnaiah(1988), Fujikoshi and Nishii(1986) and others. Many issues in a small sample remain to be resolved from a decision-theoretic viewpoint. For a good survey concerning the linear calibration, see Osborne(1991).

3.4 Estimation of restricted parameters

When a parameter space is restricted, unrestricted procedures such as unbiased estimators are needed to be modified. A simple situation illustrating the problem is to estimate the positive mean $\mu > 0$ by X, having $\mathcal{N}(\mu, 1)$. The unrestricted estimator X is minimax, but unreasonable, and this is why the shrinkage estimator $\delta_\phi = X - \phi(X)$ is considered. Applying the IERD method gives that δ_ϕ is better than X if

(a) $\phi(w)$ is nondecreasing and $\lim_{w\to\infty}\phi(w) = 0$,

(b) $\phi(w) \geq \phi_0(w)$, where

$$\phi_0(w) = \frac{\int_{-\infty}^{w} ze^{-z^2/2}dz}{\int_{-\infty}^{w} e^{-z^2/2}dz}.$$

The conditions (a), (b) are satisfied by $\phi_T(w) = \min(0, w)$ and $\phi_0(w)$, respectively, which yields the Stein type estimator $\delta^{ST} = \max(X, 0)$ and the BBZ type estimator $\delta^{BZ} = X - \phi_0(X)$. δ^{BZ}

is generalized Bayes against the non-informative prior distribution $d\mu I_{[\mu>0]}$. Rukhin(1992b) showed the interesting relationship that the improvement on the James-Stein estimator and the estimation of variance discussed in Section 2 are equivalent to the estimation of the positive mean in the asymptotic theory as $n \to \infty$, $||\boldsymbol{\theta}||^2/(p\sigma^2) \to \mu$ and $p \to \infty$.

In the case where the parameter space is restricted from both sides, namely $\mu \in [-m, m]$ for $m > 0$, the unrestricted estimator X fails the minimaxity. Casella and Strawderman(1981) showed that the Bayes estimator against a prior distribution with probabilities at the endpoints $-m$, m, given by

$$\delta_B(m) = m \ \tanh(mX) = m\frac{e^{mX} - e^{-mX}}{e^{mX} + e^{-mX}},$$

is minimax for $m < 1.05$ and better than the MLE

$$\delta^{ML} = XI_{[|X|<m]} - mI_{[X\leq -m]} + mI_{[X\geq m]}$$

and X. Other topics were studied by Bickel(1981) for second order minimaxity about m, by Iwasa and Moritani (1996) for the proper Bayesness of the MLE, and by Johnstone and MacGibbon(1992) for discussions in a Poisson distribution.

Moors(1981) indicated that, in estimation of a restricted parameter, taking the symmetry into account gives a further improvement. For instance, consider a coin tossing with probability p of the head (H) and let us suppose that $0.2 < p < 0.8$. Let us consider the simple case where p is estimated with one trial. When the head appears, we consider the estimator $\hat{p}(H) = z$ and otherwise, we take the estimator $\hat{p}(T) = 1 - \hat{p}(H) = 1 - z$ by the symmetry consideration. The mean squared error $E[(\hat{p} - p)^2]$ is then calculated as

$$\begin{aligned} E[(\hat{p} - p)^2] &= p(z - p)^2 + (1 - p)((1 - z) - p)^2 \\ &= z^2 + 2(2p(1 - p) - 1)z + 1 - 3p(1 - p), \end{aligned}$$

so that when we choose z such as

$$z > \max_{0.2<p<0.8}\{1 - 2p(1 - p)\} = 0.68,$$

the MSE is increasing in z for all p. If the head appears, $\hat{p}(H) = 0.8$ may be taken as the usual estimate since $p < 0.8$. The above symmetry consideration implies that the truncated procedure $\min(\hat{p}(H), 0.68) = 0.68$ can reduce the MSE. Moors(1981) applied such a discussion to estimate parameters in a linear regression model $y_i = \alpha + \beta x_i + e_i$, $0 < \beta < 1$.

In some applications, one can suppose an order restriction among means or variances. Let $Y_1, \ldots, Y_k$ be mutually independent random variables having $Y_i \sim \mathcal{N}(\mu_i, \sigma^2/n_i)$, and consider the case where the simple order restriction $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$ is supposed. In this case, the unrestricted estimator Y_i of μ_i is inadmissible. Lee(1981) demonstrated that Y_i has a uniformly larger MSE than the MLE, which is given by

$$\hat{\mu}_i = \min_{s\geq i} \max_{r\leq i} \frac{\Sigma_{j=r}^s Y_j n_j}{\Sigma_{j=r}^s n_j},$$

and it is derived as an isotonic regression estimator with minimizing $\Sigma_{i=1}^k (Y_i - \theta_i)^2 n_i$ under the order restriction $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$. Hwang and Peddada(1994) recently provided decision-theoretic results concerning confidence intervals under the simple order. In the case of the simple tree order $\mu_1 \leq \mu_i$, $i = 2, \ldots, k$, the generalized Bayes estimator of μ_1 improving on Y_1 was obtained by Kubokawa and Saleh(1994) through the IERD method. For the statistical inference in the restricted parameter space, see Barlow *et al.*(1972) and Robertson *et al.*(1988).

Other topics have been studied by Chang(1982), Sengupta and Sen(1991) for the Stein problem in order restrictions of means, by Blumenthal and Cohen(1968a,b), Cohen and Sackrowitz(1970) for decision-theoretic researches in the case of $k = 2$.

When functional equality constraints are embedded in a parameter space, the differential geometric structures in the statistical inference have been studied by Amari(1982) and others. The Nile problem and the estimation with a known coefficient of variation are representative examples. Appropriate group structures can be embedded in the both examples, ancillary statistics are maximal invariant and the best equivariant estimators improving on the MLE are obtained (Kariya(1989)). The latter is the issue of estimating $\boldsymbol{\theta}$ under the constraint $\boldsymbol{\theta}'\boldsymbol{\theta}/\sigma^2 = const.$ in the model (2.1), and the details of derivation of the best equivariant estimator and the extensions to more general models are discussed in Kariya *et al.*(1988), Marchand(1994) and others.

The estimation of a common mean of several normal distributions with possibly different variances is interpreted as one of the above problems with equality constraints. For instance, the two sample case has the canonical form

$$\begin{aligned} \overline{X} &\sim \mathcal{N}(\mu, \sigma_1^2/n_1), \qquad & S_1 &\sim \sigma_1^2 \chi^2_{n_1-1} \\ \overline{Y} &\sim \mathcal{N}(\mu, \sigma_2^2/n_2), \qquad & S_2 &\sim \sigma_2^2 \chi^2_{n_2-1}. \end{aligned}$$

The MLE of the common mean μ is not written in an explicit form, and so a combined procedure of $\overline{X}$ and $\overline{Y}$ of the form

$$\hat{\mu}^{GD} = \frac{cS_1^{-1}}{cS_1^{-1} + S_2^{-1}} \overline{X} + \frac{S_2^{-1}}{cS_1^{-1} + S_2^{-1}} \overline{Y}$$

may be taken as a natural choice. This two-stage GLS procedure is especially referred to as the Graybill-Deal(1959) estimator. The conditions for $\hat{\mu}^{GD}$ improving on $\overline{X}$ and $\overline{Y}$ and their extensions have been studied by Brown and Cohen(1974), Cohen and Sackrowitz(1974), Bhattacharya(1984) and others. This is related to the classical problem of recovery of interblock information in incomplete block designs with random effects (Yates(1940), Seshadri(1963), Shah(1964), Stein(1966)). Although Kubokawa(1987) developed one of admissible minimax estimators, the admissibility of $\hat{\mu}^{GD}$ remained as an open problem. Recently it was partly resolved by Kubokawa(1997b), who established the inadmissibility in the linear regression model by finding an unbiased improved estimator when the dimension is at least three.

3.5 Other problems

Besides the problems stated above, there are lots of estimation problems where the shrinkage or modification is required.

[1] Estimation after a selection. The player having taken the first place in the preliminary elimination round sometimes cannot get a good record in the final. This phenomenon is related to a so-called estimation problem after a selection. For simplicity, let us consider a k-sample problem, the i-th population having $\mathcal{N}_p(\theta_i, 1)$ and the θ_i being estimated by random variable X_i. We here call a population with the smallest mean θ_i 'the best population', which is usually selected as the i^*-th population such that

$$X_{i^*} = \max_{1 \le i \le k} X_i.$$

Then we want to estimate the mean θ_{i^*} of the selected population, being written by

$$\theta_{i^*} = \sum_{i=1}^{k} \theta_i I(X_i = \max_j X_j)$$

and X_{i*} is a natural estimator. In the case where $\theta_1 = \cdots = \theta_k = \theta_0$, or $\theta_{i*} = \theta_0$, there is some $c_k > 0$ such that

$$E[X_{i*}] = \theta_0 + c_k,$$

so that X_{i*} has a bias above. Hence X_{i*} is needed to be shrunken below. Some shrinkage procedures were proposed by Dahiya(1974), Cohen and Sackrowitz(1982) and others. Of these, Venter(1988) wrote the bias of X_{i*} as

$$E[X_{i*} - \theta_{i*}] = \sum_{i=1}^{k} \int z\phi(z) \prod_{j \neq i} \Phi(z + \theta_i - \theta_j)dz$$

for standard normal distribution function Φ and its density ϕ, and estimated this bias with

$$b(\boldsymbol{X}, \lambda) = \lambda \sum_{i=1}^{k} \int z\phi(z) \prod_{j \neq i} \Phi(z + \lambda(X_i - X_j)dz$$

for $\boldsymbol{X} = (X_1, \ldots, X_k)'$ and constant λ, which yielded the shrinkage estimator

$$\delta_{i*}^{VT} = X_{i*} - b(\boldsymbol{X}, \lambda),$$

and its bias and risk performances were investigated. However, the derivation of an unbiased estimator of θ_{i*} has not been resolved yet.

Hwang(1993) treated this issue in a framework of the mixed linear model $\theta_i = \mu + a_i$ where $a_i \sim \mathcal{N}(0, \sigma_A^2)$, and showed that the empirical Bayes estimator

$$\delta_{i*}^{EB} = \overline{X} + \left[1 - \frac{k-3}{\Sigma(X_j - \overline{X})^2}\right]^+ (X_{i*} - \overline{X}), \quad \overline{X} = p^{-1}\sum_{i=1}^{k} X_i$$

improves on X_{i*}.

For a similar problem in an exponential distribution, the UMVU and minimax estimators are given in explicit forms (Sackrowitz and SamuelCahn(1984)). Vellaisamy(1992) provided decision-theoretic results in a gamma distribution.

[2] **Multicollinearity**. It is well known that the least squared estimator (LSE) is unstable when explanatory variables have multicollinearity in the linear regression model $\boldsymbol{y} = \boldsymbol{Z\beta} + \boldsymbol{\varepsilon}$. For stabilizing the estimator, Hoerl and Kennard(1970) proposed the ridge regression estimator $\hat{\boldsymbol{\beta}}^R(k) = (\boldsymbol{Z}'\boldsymbol{Z} + k\boldsymbol{I})^{-1}\boldsymbol{Z}'\boldsymbol{y}$, $k > 0$. Although $\hat{\boldsymbol{\beta}}^R(k)$ stabilizes for some k, one is still faced with the problems: the arbitrariness of k and no uniform improvement on the LSE. Substituting estimator $\hat{k}$ for k, one can get the estimator $\hat{\boldsymbol{\beta}}^R(\hat{k})$ dominating the LSE. As indicated by Casella(1980), however, $\hat{\boldsymbol{\beta}}^R(\hat{k})$ falls unstable again by the influence of multicollinearity. This means that one cannot get an estimator possessing both properties of the stability and the uniform domination. Thus it is desirable to set up a criterion of balancing the two properties and to derive a new estimator under the criterion. Hill and Judge(1990) proposed the method for eliminating some of explanatory variables based on an unbiased estimator of a risk function of a shrinkage estimator.

Casella(1985) derived the condition on the design matrix $\boldsymbol{Z}$ for guaranteeing both of the stability and the minimaxity. For $p \times p$ orthogonal matrix $\boldsymbol{P}$ such that

$$\boldsymbol{P}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{P} = \boldsymbol{D}_\lambda = \text{diag}(\lambda_1, \ldots, \lambda_p), \quad \lambda_1 \geq \ldots \geq \lambda_p,$$

the generalized ridge regression estimator

$$\hat{\boldsymbol{\beta}}^{GR}(\boldsymbol{K}) = \boldsymbol{P}(\boldsymbol{D}_\lambda + \boldsymbol{K})^{-1}\boldsymbol{P}'\boldsymbol{Z}'\boldsymbol{y}, \quad \boldsymbol{K} = \text{diag}(k_1, \ldots, k_p)$$

is dealt with, and the *Condition Number*

$$\kappa[\hat{\boldsymbol{\beta}}(\boldsymbol{K})] = \max_{1\le i\le p}(\lambda_i + k_i)/\min_{1\le i\le p}(\lambda_i + k_i)$$

is defined as a criterion of measuring the stability. It is said that $\hat{\boldsymbol{\beta}}(\boldsymbol{K})$ improves on the LSE in terms of the condition number if $\kappa[\hat{\boldsymbol{\beta}}(\boldsymbol{K})] < \lambda_1/\lambda_p$. Let $\hat{k}_i = a_i\hat{\sigma}^2/\hat{\boldsymbol{\beta}}'\boldsymbol{Z}'\boldsymbol{Z}\hat{\boldsymbol{\beta}}$ for the unbiased estimator $\hat{\sigma}^2$ of σ^2 and constant a_i. Then Casella(1985) proved that $\hat{\beta}^{GR}(\hat{K})$ is minimax and the improvement in the condition number holds, namely, $\kappa[\hat{\boldsymbol{\beta}}(\hat{K})] \le \lambda_1/\lambda_p$ if and only if

$$\sum_{i=1}^{p-1}\lambda_i^{-2} > (\lambda_1\lambda_p)^{-1}.$$

When the design matrix $\boldsymbol{Z}$ satisfies the above inequality, one thus gets a ridge regression estimator possessing the stability and the minimaxity.

[3] Shrinkage toward a null hypothesis. Let X be a random variable having $\mathcal{N}(\mu, \sigma^2)$ for known variance σ^2, and consider the situation where the mean μ is expected to be near the null hypothesis $H_0 : \mu = \mu_0$. In this case, X should be shrunken toward μ_0 as

$$\hat{\mu}(k) = k(T)X + \{1 - k(T)\}\mu_0, \quad T = (X - \mu_0)/\sigma.$$

Here several methods about how to determine the shrinkage function $k(T)$ have been proposed. Hirano(1977) treated it in a framework of a preliminary test estimation, set $k_1(t) = I(|t| \ge z_{\alpha/2})$ and proposed to determine the constant α such that the Akaike Information Criterion(AIC) is minimized, where $z_{\alpha/2}$ denotes the $100(1-\alpha/2)\%$ point of the standard normal distribution. Since the MSE of the estimator $cX + (1-c)\mu_0$ is minimized at $c = \{(\mu-\mu_0)^2/\sigma^2\}/\{1 + (\mu - \mu_0)^2/\sigma^2\}$, Thompson(1968) considered to substitute an estimator for μ and proposed to choose $k_2(t) = t^2/(1+t^2)$. Combining both ideas, Inada(1984) considered

$$k_3(t) = d^*I(|t| < z_{\alpha/2}) + I(|t| \ge z_{\alpha/2})$$

and proposed to determine (α, d^*) through the minimax regret criterion. Another method was given by Hawkins and Han(1989), and the case of the scale parameter was discussed by Kambo *et al.*(1990). Theoretically X is admissible, and one cannot get any estimator $\hat{\mu}(k)$ superior to X. In some applications, however, it is expected that μ is near μ_0, or more generally that a model is in the neighborhood of the supposed hypothesis. In such a situation, the above type of shrinkage procedures may be effective and useful.

When the hypothesis $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0}$ is suspected in the linear regression model $\boldsymbol{y} = \boldsymbol{Z}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, Ghosh *et al.*(1989) gave the empirical Bayes estimator of $\boldsymbol{\beta}_1$ given by

$$\hat{\boldsymbol{\beta}}_1^{EB} = \hat{\boldsymbol{\beta}}_1 + \left(1 - \frac{c}{F}\right)(\tilde{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_1)$$

and derived the conditions on c for $\hat{\boldsymbol{\beta}}_1^{EB}$ improving on the minimax estimator $\tilde{\boldsymbol{\beta}}_1$, where $\hat{\boldsymbol{\beta}}_1$ is the restricted LSE under H_0, $(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2)$ is unrestricted LSE of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and

$$F = \frac{\tilde{\boldsymbol{\beta}}_2'\boldsymbol{Z}_2'(\boldsymbol{I} - \boldsymbol{Z}_1(\boldsymbol{Z}_1'\boldsymbol{Z}_1)^{-1}\boldsymbol{Z}_1')\boldsymbol{Z}_2\tilde{\boldsymbol{\beta}}_2}{||\boldsymbol{y} - \boldsymbol{Z}_1\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{Z}_2\tilde{\boldsymbol{\beta}}_2||^2},$$

which is a likelihood ratio test statistic for $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0}$ vs. $H_A : \boldsymbol{\beta}_2 \ne \boldsymbol{0}$. The empirical Bayes estimator is an intermediate between $\tilde{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_1$, and if H_0 is suspected, it approaches $\hat{\boldsymbol{\beta}}_1$; otherwise, it does $\tilde{\boldsymbol{\beta}}_1$.

[4] Estimation of the reliability. The problem of estimating the one-sided normal probability $\theta = P[X_i > c]$ based on random sample $X_1, \ldots, X_n$ from $\mathcal{N}(\mu, \sigma^2)$ is motivated from quality control. For instance, in the case where measurement X_i of a product is requested to be less than or equal to upper bound c of a standard, the one-sided propability θ means the inferior rate of the product and it is required to be estimated.

Brown and Rutemiller(1973) numerically compared performances of UMVUE and MLE, and Fujino(1987) extended the issue to a linear regression model and revealed the efficiency of a bias-corrected estimator derived through an asymptotic expansion of the MLE type estimator. Peszek and Rukhin(1993) gave a generalized Bayes rule of θ and demonstrated the admissibility. The estimation of $P[Y < X]$ for mutually independent random variables X and Y were discussed by Enis and Geisser(1971), Yu and Govindarajulu(1995), Guttman and Papandonatos(1997) and others.

In the exponential distribution $\sigma^{-1}\exp(-x/\sigma)$, the estimation of the reliability $P[X > t] = \exp(-t/\sigma)$ has been discussed by Zacks and Even(1966), Chiou(1993) and others. Pierce(1973), Varde(1969), Rukhin and Ananda(1989) studied Bayes estimators and their admissibility for reliability in an exponential distribution with unknown location-scale parameters.

[5] Estimation of a quantile. The quantile of the normal distribution $\mathcal{N}(\mu, \sigma^2)$ is written by a linear combination of μ and σ such as $\theta = \mu + a\sigma$ for constant a. Zidek(1969, 71) proved that a usual minimax estimator of θ is improved on by a truncated procedure. Some decision-theoretical results in an exponential distribution with unknown location-scale parameters have been given by Rukhin and Strawderman(1982), Rukhin and Zidek(1985), Rukhin(1986) and Sharma and Kumar(1994).

[6] Discriminant Analysis. The aim in discriminant analysis is to assign an individual to one of two or more distinct groups by means of an allocation rule. The allocation rule is constructed based on training sample for which the group membership of each observation is known. When the allocation rule in a multivariate discriminant analysis is performed on the basis of the linear or quadratic discriminant function, it incorporates an estimator of the inverse of the covariance matrix. This implies that it is quite sensitive to the smallest eigenvalue of the estimator of the covariance matrix and is likely to be more unstable as the dimension gets larger, resulting in a rise in the error rate in discriminant analysis (Peck and Van Ness(1982), Friedman(1989) and Matsuda *et al.*(1990)). It is, for instance, supposed that an individual with observation on a p-dimensional random vector is to be allocated into one of two p-variate normal distributions with a common covariance:

$$\Pi_i : \boldsymbol{X}_{ij} \sim \mathcal{N}_p(\boldsymbol{\theta}_i, \Sigma), \quad j = 1, \ldots, n_i, \; i = 1, 2.$$

The linear discriminant function is given by

$$h(\boldsymbol{x}; \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{S}) = \{\boldsymbol{x} - 2^{-1}(\overline{\boldsymbol{X}}_1 + \overline{\boldsymbol{X}}_2)\}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$$

where $\overline{\boldsymbol{X}}_i$ is a sample mean of the i-th population and $\boldsymbol{S}$ is an unbiased estimator of Σ. When $h(\boldsymbol{x}; \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{S}) > 0$ (resp. ≤ 0), $\boldsymbol{x}$ is allocated into Π_1 (resp. Π_2). If the dimension p is large in this function, the inverse matrix $\boldsymbol{S}^{-1}$ is likely to be unstable. When $p > n_1 + n_2 - 2$, $\boldsymbol{S}^{-1}$ falls into the ill condition. Haff(1986), Dey and Srinivasan(1991) considered the estimation of the coefficient vector $\eta = \Sigma^{-1}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$ in this linear discriminant function by use of $(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{S})$, and applied theoretical results in estimation of Σ^{-1} to show that the shrinkage procedure $c(\boldsymbol{S} + u\phi(u)\boldsymbol{I})^{-1}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$ for $u = 1/\mathrm{tr}\boldsymbol{S}^{-1}$ improves on $c\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$ under some conditions. Rukhin(1992a) gave a condition on c for admissibility of a usual procedure.

[7] Estimation of error rates in discriminant analysis. In discriminant analysis, it is important to estimate the error rates in allocating a randomly selected future observation.

The apparent error rates and MLE type estimators are known to be inclined to under-estimate the actual error rates. In the notations stated above, for instance, the actual error rate of misallocating $\boldsymbol{x}$ from Π_1 into Π_2 is given by

$$\begin{aligned}\lambda(\boldsymbol{\theta}_1, \Sigma, \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{S}) &= P[h(\boldsymbol{x}; \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{S}) \leq 0 | \boldsymbol{x} \sim \mathcal{N}_p(\boldsymbol{\theta}_1, \Sigma)] \\ &= \Phi\left\{-\frac{\{\boldsymbol{\theta}_1 - 2^{-1}(\overline{\boldsymbol{X}}_1 + \overline{\boldsymbol{X}}_2)\}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)}{\sqrt{(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1}\Sigma\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)}}\right\},\end{aligned}$$

where Φ designates the standard normal distribution function. The issue is to estimate this actual error rate based on $\overline{\boldsymbol{X}}_1$, $\overline{\boldsymbol{X}}_2$ and $\boldsymbol{S}$. Substituting $\hat{\boldsymbol{\theta}}_1 = \overline{\boldsymbol{X}}_1$, $\hat{\Sigma} = \boldsymbol{S}$ for $\boldsymbol{\theta}_1$, Σ in $\lambda(\boldsymbol{\theta}_1, \Sigma, \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{S})$ yields the MLE type estimator

$$\hat{\lambda}^{ML} = \Phi(-D/2), \quad D^2 = (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2),$$

which, however, has an undesirable property of under-estimating the actual rate. For eliminating it, McLachlan(1974) proposed a bias-corrected estimator on the basis of an asymptotic expansion of $\lambda(\boldsymbol{\theta}_1, \Sigma, \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{S})$ derived by Okamoto(1963). Efron(1979) introduced the bootstrap method to provide a bias-corrected apparent error rate in a nonparametric model. Theoretical and numerical comparisons of the proposed bias-corrected procedures involving the cross-validation method were studied by Konishi and Honda(1990, 92).

[8] Coping with improper solutions in factor analysis. Consider the factor analysis model with random common factor vector $\boldsymbol{f}$, given by

$$\boldsymbol{x}_i = \boldsymbol{\mu} + \Lambda \boldsymbol{f}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n$$

where $\boldsymbol{x}_i \in \boldsymbol{R}^p$ is an observed vector, $\boldsymbol{\mu} \in \boldsymbol{R}^p$ and Λ $(p \times k)$ are unknown parameters, $\boldsymbol{f}_i \in \boldsymbol{R}^k$ and $\boldsymbol{e}_i \in \boldsymbol{R}^p$ are mutually independent random variables having $\boldsymbol{f}_i \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{e}_i \sim \mathcal{N}_p(\boldsymbol{0}, \Psi)$ for unknown $\Psi = \text{diag}(\psi_1, \ldots, \psi_p)$. Letting

$$\begin{aligned}\boldsymbol{S} &= n^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})', \\ \Sigma &= \mathbf{Cov}(\boldsymbol{x}_i) = \Lambda\Lambda' + \Psi,\end{aligned}$$

the MLE of Ψ and Λ can be derived by minimizing

$$q(\Psi, \Lambda) = \text{tr}(\boldsymbol{S}\Sigma^{-1}) - \log|\boldsymbol{S}\Sigma^{-1}|.$$

Then it is known that improper (non-positive) solutions are sometimes yielded for Ψ. This is related to the problem of estimating parameters that are subject to a set of inequality and equality functional constraints in covariance structure analysis. For coping with this drawback, Lee(1980, 81) proposed the penalty function and the Lagrange multiplier methods. Takeuchi(1986) showed the existence of the ML solutions, clarified the structure of the MLE and gave in-depth studies of the case where $\hat{\psi}_i^{ML} = 0$. Akaike(1987) treated the above factor model in a Bayesian framework with a similar prior distribution as in Martin and McDonald(1975), and based on the AIC, proposed to obtain solutions minimizing the equation

$$q^*(\Psi, \Lambda) = \text{tr}(\boldsymbol{S}(\Lambda\Lambda' + \Psi)^{-1}) - \log|\boldsymbol{S}(\Lambda\Lambda' + \Psi)^{-1}| + \delta\ \text{tr}\Lambda'\Psi^{-1}\Lambda.$$

The second term in the r.h.s. functions as a penalty so that improper solutions are not yielded, since it gets large as one of ψ_i's approaches zero. It is thus noted that the proposed procedures are modified to keep estimators far from zero rather than to shrink them towards somewhere.

Since δ is a hyper-parameter adjusting the extent of the penalty, it may be needed to be estimated from data through empirical Bayes rules. Further theoretical developments will be desired.

Acknowledgements. The author is grateful to the referee for his valuable comments which improved on the presentation of the paper. This research was supported in part by a grant from the Shirakawa Institute of Animal Genetics and the Japan Livestock Technology Association, by the Ministry of Education, Japan, Grant No. 09780214 and by a grant from the Research Institute for the Japanese Economy, University of Tokyo. I am also grateful to Mr. M. Ushijima for his help in the simulation experiments.

REFERENCES

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317-332.

Amari, S. (1982). Differential geometry of curved exponential families - Curvature and information loss. *Ann. Statist.* , **10**, 357-385.

Amemiya, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite. *Amer. Statist.*, **39**, 112-117.

Arnold, B.C. (1970). Inadmissibility of the usual scale estimate for a shifted exponential distribution. *J. Amer. Statist. Assoc.*, **65**, 1260-1264.

Barlow, R.F., Bartholomew, D.D., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

Berger, J.O.(1976a). Inadmissibility results for generalized Bayes estimators of coordinates of a location vector. *Ann. Statist.*, **4**, 302-333.

Berger, J.O.(1976b). Admissibility results for generalized Bayes estimators of coordinates of a location vector. *Ann. Statist.*, **4**, 334-356.

Berger, J.O.(1985). *Statistical Decision Theory and Bayesian Analysis.* 2nd Ed., Springer-Verlag, New York.

Berger, J.O., Berliner, L.M. and Zaman, A. (1982). General admissibility results for estimation in a control problem. *Ann. Statist.*, **10**, 838-856.

Berliner, L.M. (1983). Improving on inadmissible estimators in control problems. *Ann. Statist.*, **11**, 814-826.

Bhattacharya, C.G. (1984). Two inequalities with an application. *Ann. Inst. Statist. Math.*, **36**, 129-134.

Bickel, P.J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.*, **9**, 1301-1309.

Blumenthal, S. and Cohen, A. (1968a). Estimation of the larger translation parameter. *Ann. Math. Statist.*, **39**, 502-516.

Blumenthal, S. and Cohen, A. (1968b). Estimation of two ordered translation parameters. *Ann. Math. Statist.*, **39**, 517-530.

Brandwein, A.C. and Strawderman, W.E. (1990). Stein estimation: The spherically symmetric case. *Statist. Science*, **5**, 356-369.

Brewster, J.F. (1974). Alternative estimators for the scale parameter of the exponential distribution with unknown location. *Ann. Statist.*, **2**, 553-557.

Brewster, J.F. and Zidek, J.V. (1974). Improving on equivariant estimators. *Ann. Statist.*, **2**, 21-38.

Brown, G.G. and Rutemiller, H.C. (1973). The efficiencies of maximum likelihood and minimum variance unbiased estimators of fraction defective in the normal case. *Technometrics*, **15**, 849-855.

Brown, L.D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.*, **39**, 29-48.

Brown, L.D. and Cohen, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.*, **2**, 963-976.

Brown, P.J. (1982). Multivariate calibration. *J. Roy. Statist. Soc.* (Ser. B), **44**, 287-321.

Calvin, J.A. and Dykstra, R.L. (1991). Maximum likelihood estimation of a set of covariance matrices under Lowner order restrictions with applications to balanced multivariate variance components models. *Ann. Statist.*, **19**, 850-869.

Casella, G. (1980). Minimax ridge regression estimation. *Ann. Statist.*, **8**, 1036-1056.

Casella, G. (1985). Condition numbers and minimax ridge regression estimators. *J. Amer. Statist. Assoc.*, **80**, 753-758.

Casella, G. and Strawderman, W.E. (1981). Estimating a bounded mean. *Ann. Statist.*, **9**, 870-879.

Chang, Y.-T. (1982). Stein-type estimators for parameters in truncated spaces. *Keio Sci. Tech. Rep.*, **35**, 185-193.

Chiou, P. (1993). Empirical Bayes shrinkage estimation of reliability in the exponential distribution. *Comm. Statist.-Theory Methods*, **22**, 1483-1494.

Chow, M.S. (1987). A complete class theorem for estimating a noncentrality parameter. *Ann. Statist.*, **15**, 800-804.

Cohen, A. (1972). Improved confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.*, **67**, 382-387.

Cohen, A. and Sackrowitz, H.B. (1970). Estimation of the last mean of a monotone sequence. *Ann. Math. Statist.*, **41**, 2021-2034.

Cohen, A. and Sackrowitz, H.B. (1974). On estimating the common mean of two normal distributions. *Ann. Statist.*, **2**, 1274-1282.

Cohen, A. and Sackrowitz, H.B. (1982). Estimating the mean of the selected population. In *Statistical Decision Theory and Related Topics* III (S.S. Gupta, J.O. Berger, eds.), **1**, 243-270. Academic Press, New York.

Dahiya, R.C. (1974). Estimation of the mean of the selected population. *J. Amer. Statist. Assoc.*, **69**, 226-230.

Dey, D.K. and Srinivasan, C. (1991). On estimation of discriminant coefficients. *Statist. Probab. Lett.*, **11**, 189-193.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, **7**, 1-26.

Efron, B. and Morris, C. (1972a). Empirical Bayes on vector observations: An extension of Stein's method. *Biometrika*, **59**, 335-347.

Efron, B. and Morris, C. (1972b). Limiting the risk of Bayes and empirical Bayes estimators - Part II: The empirical Bayes case. *J. Amer. Statist. Assoc.*, **67**, 130-139.

Enis, P. and Geisser, S. (1971). Estimation of the probability that $Y < X$. *J. Amer. Statist. Assoc.*, **66**, 162-168.

Friedman, J.H. (1989). Regularized discriminant analysis. *J. Amer. Statist. Assoc.*, **84**, 165-175.

Fujikoshi, Y. and Nishii, R. (1986). Selection of variables in a multivariate inverse regression problem. *Hiroshima Math. J.*, **16**, 267-277.

Fujino, Y. (1987). Estimation of one-sided normal probability. *Japanese J. Applied Statist.*, **16**, 119-130 (in Japanese).

Gelfand, A.E. and Dey, D.K. (1988). On the estimation of a variance ratio. *J. Statist. Plann. Inference*, **19**, 121-131.

Ghosh, M. (1994). On some Bayesian solutions of the Neyman-Scott problem. *Statistical*

Decision Theory and Related Topics V (eds. S.S. Gupta and J.O. Berger), 267-276, Springer-Verlag, New York.

Ghosh, M., Saleh, A.K.Md.E. and Sen, P.K. (1989). Empirical Bayes subset estimation in regression models. *Statist. Decisions*, **7**, 15-35.

Goutis, C. and Casella, G. (1991). Improved invariant confidence intervals for a normal variance. *Ann. Statist.*, **19**, 2015-2031.

Graybill, F.A. and Deal, R.B. (1959). Combining unbiased estimators. *Biometrics*, **15**, 543-550.

Guttman, I. and Papandonatos, G.D. (1997). A Bayesian approach to a reliability problem: theory, analysis and interesting numerics. *Canadian J. Statist.*, **25**, 143-158.

Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531-542.

Haff, L.R. (1986). On linear log-odds and estimation of discriminant coefficients. *Comm. Statist.-Theory Methods*, **15**, 2131-2144.

Hawkins, D.L. and Han, C.-P. (1989). A minimum average risk approach to shrinkage estimators of the normal mean. *Ann. Inst. Statist. Math.*, **41**, 347-363.

Hill, R.C. and Judge, G.G. (1990). Improved estimation under collinearity and squared error loss. *J. Multivariate Anal.*, **32**, 296-312.

Hirano, K. (1977). Estimation procedures based on preliminary test, shrinkage technique and information criterion. *Ann. Inst. Statist. Math.*, **29**, 21-34.

Hoadley, B. (1970). A Bayesian look at inverse linear regression. *J. Amer. Statist. Assoc.*, **65**, 356-369.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55-68.

Hwang, J.T. (1993). Empirical Bayes estimation for the means of the selected populations. *Sankhya* (Ser. A), **55**, 285-311.

Hwang, J.T. and Peddada, S.D. (1994). Confidence interval estimation subject to order restrictions. *Ann. Statist.*, **22**, 67-93.

Inada, K. (1984). A minimax regret estimator of a normal mean after preliminary test. *Ann. Inst. Statist. Math.*, **36**, 207-215.

Iwasa, M. and Moritani, Y. (1996). A note on admissibility of the maximum likelihood estimator for a bounded normal mean. *Statist. Probab. Letters*, **32**, 99-105.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, **1**, 361-379. University of California Press, Berkeley.

Johnstone, I.M. and MacGibbon, K.B. (1992). Minimax estimation of a constrained Poisson vector. *Ann. Statist.*, **20**, 807-831.

Judge, G. and Bock, M.E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics.* North-Holland, Amsterdam.

Kambo, N.S., Handa, B.R. and Al-Hemyari, Z.A. (1990). On shrunken estimators for exponential scale parameter. *J. Statist. Plann. Inference*, **24**, 87-94.

Kariya, T. (1989). Equivariant estimation in a model with an ancillary statistic. *Ann. Statist.*, **17**, 920-928.

Kariya, T., Giri, N.C. and Perron, F. (1988). Equivariant estimation of a mean vector $\boldsymbol{\mu}$ of $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu} = 1$ or $\Sigma^{-1/2}\boldsymbol{\mu} = \boldsymbol{c}$ or $\Sigma = \sigma^2\boldsymbol{\mu}'\boldsymbol{\mu}\boldsymbol{I}$. *J. Multivariate Anal.*, **27**, 270-283.

Kleffe, J. and Rao, J.N.K. (1986). The existence of asymptotically unbiased non-negative quadratic estimates of variance components in ANOVA models. *J. Amer. Statist. Assoc.*, **81**, 692-698.

Konishi, S. and Honda, M. (1990). Comparison of procedures for estimation of error rates in discriminant analysis under nonnormal populations. *J. Statist. Comput. Simul.*, **36**, 105-115.

Konishi, S. and Honda, M. (1992). Bootstrap methods for error rate estimation in discriminant analysis. *Japanese J. Applied Statist.*, **21**, 67-100 (in Japanese).

Kourouklis, S. (1997). A new property of the inverse Gaussian distribution with applications. *Statist. Probab. Letters*, **32**, 161-166.

Krutchkoff, R.G. (1967). Classical and inverse regression methods of calibration. *Technometrics*, **9**, 429-439.

Kubokawa, T. (1987). Admissible minimax estimation of a common mean of two normal populations. *Ann. Statist.*, **15**, 1245-1256.

Kubokawa, T. (1991). An approach to improving the James-Stein estimator. *J. Multivariate Anal.*, **36**, 121-126.

Kubokawa, T. (1994a). A unified approach to improving equivariant estimators. *Ann. Statist.*, **22**, 290-299.

Kubokawa, T. (1994b). Double shrinkage estimation of ratio of scale parameters. *Ann. Inst. Statist. Math.*, **46**, 95-116.

Kubokawa, T. (1995). Estimation of variance components in mixed linear models. *J. Multivariate Anal.*, **53**, 210-236.

Kubokawa, T. (1998a). The Stein phenomenon in simultaneous estimation: A review. *Applied Statistical Science III* (eds. S.E. Ahmed, M. Ahsanullah and B.K. Sinha), NOVA Science Publishers, Inc., New York, to appear.

Kubokawa, T. (1998b). Double shrinkage estimation of common coefficients in two regression equations with heteroscedasticity. *Discussion Paper Series 97-F-20, Faculty of Economics, University of Tokyo.*

Kubokawa, T., Honda, T., Morita, K. and Saleh, A.K.Md.E. (1993a). Estimating a covariance matrix of a normal distribution with unknown mean. *J. Japan Statist. Soc.*, **23**, 131-144.

Kubokawa, T., Morita, K., Makita, S. and Nagakura, K. (1993b). Estimation of the variance and its applications. *J. Statist. Plann. Inference*, **35**, 319-333.

Kubokawa, T. and Robert, C.P. (1994). New perspectives on linear calibration. *J. Multivariate Anal.*, **51**, 178-200.

Kubokawa, T., Robert, C. and Saleh, A.K.Md.E. (1992). Empirical Bayes estimation of the variance parameter of a normal distribution with unknown mean under an entropy loss. *Sankhya* (Ser. A), **54**, 402-410.

Kubokawa, T., Robert, C.P. and Saleh, A.K.Md.E. (1993c). Estimation of noncentrality parameters. *Canad. J. Statist.*, **21**, 45-57.

Kubokawa, T. and Saleh, A.K.Md.E. (1994). Estimation of location and scale parameters under order restrictions. *J. Statist. Research*, **28**, 41-51.

Kubokawa, T., Saleh, A.K.Md.E., Konno, Y. and Ushijima, M. (1996). Estimation of variance components under Kullback-Leibler loss with applications. *Discussion Paper Series 96-F-12, Faculty of Economics, University of Tokyo.*

Kubokawa, T., Saleh, A.K.Md.E. and Makita, S. (1993d). On improved positive estimators of variance components. *Statist. Decisions*, Supplement Issue **3**, 1-16.

Kubokawa, T. and Srivastava, M.S. (1996). Double shrinkage estimators of ratio of variances. In *Multidimensional Statistical Analysis and Theory of Random Matrices* (Proceedings of the Sixth Eugene Lukacs Symposium, Bowling Green, Editors: A.K. Gupta and V.L. Girko), 139-154, VSP, Netherlands.

LaMotte, L.R. (1973). On non-negative quadratic unbiased estimation of variance components. *J. Amer. Statist. Assoc.*, **68**, 728-730.

Lee, C.C. (1981). The quadratic loss of isotonic regression under normality. *Ann. Statist.*, **9**, 686-688.

Lee, S.-Y. (1980). Estimation of covariance structure models with parameters subject to functional restraints. *Psychometrika*, **45**, 309-324.

Lee, S.-Y. (1981). The multiplier method in constrained estimation of covariance structure models. *J. Statist. Comput. Simul.*, **12**, 247-257.

Lele, C. (1992). Inadmissibility of loss estimators. *Statist. Decisions*, **10**, 309-322.

Leung, P.L. and Muirhead, R.J. (1987). Estimation of parameter matrices and eigenvalues in MANOVA and canonical correlation analysis. *Ann. Statist.*, **15**, 1651-1666.

Li, K.-C. (1985). From Stein's unbiased risk estimates to the method of generalized cross validation. *Ann. Statist.*, **13**, 1352-1377.

Maatta, J.M. and Casella, G. (1990). Developments in decision-theoretic variance estimation. *Statist. Science*, **5**, 90-120.

MacGibbon, B. and Shorrock, G. (1994). Estimation of the lamda parameter of an inverse Gaussian distribution. *Statist. Probab. Lett.*, to appear.

Marchand, E. (1994). On the estimation of the mean of a $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ population with $\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$ known. *Statist. Probab. Lett.*, **21**, 69-75.

Martin, J.K. and McDonald, R.P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, **40**, 505-516.

Maruyama, Y. (1996). Estimation of the mean vector and the variance of a multivariate normal distribution. Unpubilished manuscript.

Mathew, T., Sinha, B.K. and Sutradhar, B.C. (1992). Nonnegative estimation of variance components in unbalanced mixed models with two variance components. *J. Multivariate Anal.*, **42**, 77-101.

Mathew, T., Niyogi, A. and Sinha, B.K. (1994). Improved nonnegative estimation of variance components in balanced multivariate mixed models. *J. Multivariate Anal.*, **51**, 83-101.

Matsuda, S., Fujimoto, T. and Yoshimura, I. (1990). A robust quadratic discriminant function using a shrinkage estimator of variance matrix. *Japanese J. Applied Statist.*, **19**, 33-51 (in Japanese).

McLachlan, G.J. (1974). An asymptotic unbiased technique for estimating the error rates in discriminant analysis. *Biometrics*, **30**, 239-249.

Miwa, T. (1979). On the inference of the SN ratio. *Proc. of the 9th annual meeting of the Japanese Quality Control Society*, 37-40 (in Japanese).

Miwa, T. (1985). Comparison among point estimators in linear calibration in terms of means squared error. *Japanese J. Applied Statist.*, **14**, 83-93 (in Japanese).

Moors, J.J.A. (1981). Inadmissibility of linearly invariant estimators in truncated parameter spaces. *J. Amer. Statist. Assoc.*, **76**, 910-915.

Muirhead, R.J. (1985). Estimating a particular function of the multiple-correlation coefficient. *J. Amer. Statist. Assoc.*, **80**, 923-925.

Nagata, Y. (1989). Improvements of interval estimations for the variance and the ratio of two variances. *J. Japan Statist. Soc.*, **19**, 151-161.

Nagata, Y. (1991). Improved interval estimation for the scale parameter of exponential distribution with an unknown location-scale parameter. *Quality*, **21**, 5-10 (in Japanese).

Neff, N. and Strawderman, W.E. (1976). Further remarks on estimating the parameter of a noncentral chi-squared distribution. *Comm. Statist.* , **A5**, 65-76.

Nishii, R. and Krishnaiah, P.R. (1988). On the moments of classical estimates of explanatory variables under a multivariate calibration model. *Sankhya* (Ser. A), **50**, 137-148.

Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.*, **34**, 1286-1301.

Osborne, C. (1991). Statistical calibration: A review. *Inter. Statist. Review*, **59**, 309-336.

Pal, N. and Sinha, B.K. (1989). Improved estimators of dispersion of an inverse Gaussian distribution. In *Statistical Data Analysis and Inference* (Y. Dodge, ed.), 215-222. Elsevier Science Publishers.

Peck, R. and Van Ness, J. (1982). The use of shrinkage estimators in linear discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intelligence*, **4**, 530-537.

Perlman, M.D. and Rasmussen, V.A. (1975). Some remarks on estimating a noncentrality parameter. *Comm. Statist.*, **4**, 455-468.

Perron, F. (1992). Minimax estimators of a covariance matrix. *J. Multivariate Anal.*, **43**, 16-28.

Peszek, I. and Rukhin, A.L. (1993). Estimating normal distribution function and normal density. *Statist. Decisions*, **11**, 391-406.

Pierce, D.A. (1973). Fiducial, frequency, and Bayesian inference on reliability for the two parameter negative exponential distribution. *Technometrics*, **15**, 249-253.

Portnoy, S. (1971). Formal Bayes estimation with application to a random effect model. *Ann. Math. Statist.*, **42**, 1379-1402

Proskin, H.M. (1985). An admissibility theorem with applications to the estimation of the variance of the normal distribution. Ph.D. dissertation, Dept. Statist., Rutgers University.

Remadi, S. and Amemiya, Y. (1994). Asymptotic properties of the estimators for multivariate components of variance. *J. Multivariate Anal.*, **49**, 110-131.

Robert, C.P. (1994). *The Bayesian Choice : A Decision-Theoretic Motivation.* Springer-Verlag, New York.

Robertson, T., Wright, F.T. and Dykstra, R. (1988). *Order Restricted Statistical Inference.* North-Holland, New York.

Rukhin, A.L. (1986). Admissibility and minimaxity results in the estimation problem of exponential quantiles. *Ann. Statist.*, **14**, 220-237.

Rukhin, A.L. (1987). How much better are better estimators of a normal variance ? *J. Amer. Statist. Assoc.*, **82**, 925-928.

Rukhin, A.L. (1991). Asymptotic variance estimation in multivariate distributions. *J. Multivariate Anal.*, **38**, 366-384.

Rukhin, A.L. (1992a). Generalized Bayes estimators of a normal discriminant function. *J. Multivariate Anal.*, **41**, 154-162.

Rukhin, A.L. (1992b). Asymptotic risk behavior of mean vector and variance estimators and the problem of positive normal mean. *Ann. Inst. Statist. Math.*, **44**, 299-311.

Rukhin, A.L. (1995). Admissibility: Survey of a concept in progress. *Inter. Statist. Review*, **63**, 95-115.

Rukhin, A.L. and Ananda, M.M.A. (1989). Estimating exponential reliability function and exponential density. *Statist. Decisions*, **7**, 277-296.

Rukhin, A.L., Kuo, L. and Dey, D.K. (1990). A class of minimax estimators of the scale parameter of the uniform distribution. *Statist. Probab. Lett.*, **9**, 317-321.

Rukhin, A.L. and Sinha, B.K. (1991). Decision-theoretic estimation of the product of Gamma scales and generalized variance. *Calcutta Statist. Assoc. Bull.*, **40**, 257-265.

Rukhin, A.L. and Strawderman, W.E. (1982). Estimating a quantile of an exponential distribution. *J. Amer. Statist. Assoc.*, **77**, 159-162.

Rukhin, A.L. and Zidek, J.V. (1985). Estimation of linear parametric functions for several exponential samples. *Statist. Decisions*, **3**, 225-238.

Sackrowitz, H. and Samuel-Cahn, E. (1984). Estimation of the mean of a selected negative exponential population. *J. Roy. Statist. Soc.*, **46**, 242-249.

Sarkar, S.K. (1989). On improving the shortest length confidence interval for the generalized variance. *J. Multivariate Anal.*, **31**, 136-147.

Saxena, K.M.L. and Alam, K. (1982). Estimation of the noncentrality parameter of a chi-squared distribution. *Ann. Statist.*, **10**, 1012-1016.

Sengupta, D. and Sen, P.K. (1991). Shrinkage estimation in a restricted parameter space. *Sankhya* (Ser. A), **53**, 389-411.

Seshadri, V. (1963). Combining unbiased estimators. *Biometrics*, **19**, 163-170.

Shah, K.R. (1964). Use of inter-block information to obtain uniformly better estimators. *Ann. Math. Statist.*, **35**, 1064-1078.

Shao, P.Y.-S. and Strawderman. (1995). Improving on the positive part of the UMVUE of a noncentrality parameter of a noncentral chi-square distribution. *J. Multivariate Anal.*, **53**, 52-66.

Sharma, D. and Kumar, S. (1994). Estimating quantiles of exponential populations. *Statist. Decisions*, **12**, 343-352.

Shinozaki, N. (1995). Some modifications of improving estimators of a normal variance. *Ann. Inst. Statist. Math.*, to appear.

Shorrock, G. (1990). Improved confidence intervals for a normal variance. *Ann. Statist.*, **18**, 972-980.

Shorrock, R.B. and Zidek, J.V. (1976). An improved estimator of the generalized variance. *Ann. Statist.*, **4**, 629-638.

Sinha, B.K. (1976). On improved estimators of the generalized variance. *J. Multivariate Anal.*, **6**, 617-626.

Sinha, B.K. and Ghosh, M. (1987). Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss. *Statist. Decisions*, **5**, 201-227.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symp. Math. Statist. Probab.*, **1**, 197-206. University of California University, Berkeley.

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

Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Research Papers in Statistics* (Neyman Festschrift, F. N. David, ed.), 351-366. Wiley, New York.

Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. In *Proc. Prague Symp. Asymptotic Statist.*, 345-381.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135-1151.

Strawderman. W.E. (1974). Minimax estimation of powers of the variance of a normal population under squared error loss. *Ann. Statist.*, **2**, 190-198.

Sugiura, N. and Konno, Y. (1988). Entropy loss and risk of improved estimators for the generalized variance and precision. *Ann. Inst. Statist. Math.*, **40**, 329-341.

Takeuchi, K. (1968). On the problem of fixing the level of independent variables in a linear regression function. IMM 367, Courant Institute of Math. Sciences, New York University.

Takeuchi, K. (1986). On the structure of the maximum likelihood estimator in factor analysis. *Japanese J. Applied Statist.*, **15**, 29-45 (in Japanese).

Takeuchi, K. (1991). Personal communications.

Tate, R.F. and Klett, G.W. (1959). Optimal confidence intervals for the variance of a normal distribution. *J. Amer. Statist. Assoc.*, **54**, 674-682.

Thompson, J.R. (1968). Some shrinkage techniques for estimating the mean. *J. Amer. Statist. Assoc.*, **63**, 113-122.

Varde, S.D. (1969). Life testing and reliability estimation for the two parameter exponential distribution. *J. Amer. Statist. Assoc.*, **64**, 621-631.

Vellaisamy, P. (1992). Inadmissibilty results for the selected scale parameters. *Ann. Statist.*, **20**, 2183-2191.

Venter, J.H. (1988). Estimation of the mean of the selected population. *Comm. Statist.-Theory Methods*, **17**, 791-805.

Yates, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugenics*, **10**, 317-325.

Yu, Q. and Govindarajulu, Z. (1966). Admissibility and minimaxity of the UMVU estimator of $P\{X < Y\}$. *Ann. Statist.*, **23**, 598-607.

Zacks, S. and Even, M. (1966). The efficiency in small samples of the maximum likelihood and best unbiased estimators of reliability functions. *J. Amer. Statist. Assoc.*, **61**, 1033-1051.

Zaman, A. (1981). A complete class theorem for the control problem, and further results on admissibility and inadmissibility. *Ann. Statist.*, **9**, 812-821.

Zidek, J. (1969). Inadmissibility of the best invariant estimator of extreme quantiles of the normal law under squared error loss. *Ann. Math. Statist.*, **40**, 1801-1808.

Zidek, J. (1971). Inadmissibility of a class of estimators of a normal quantile. *Ann. Math. Statist.*, **42**, 1444-1447.

Zidek, J. (1973). Estimating the scale parameter of the exponential distribution with unknown location. *Ann. Statist.*, **1**, 264-278.