97-F-14

The Stein Phenomenon in Simultaneous Estimation: A Review

Tatsuya Kubokawa
University of Tokyo

April 1997

The Stein Phenomenon in Simultaneous Estimation: A Review

Tatsuya KUBOKAWA
Department of Economics, University of Tokyo

In the simultaneus estimation of a mean of a multivariate normal distribution, Charles Stein discovered the surprising decision-theoretic result that the usual maximum likelihood estimator is *inadmissible* with respect to quadratic loss in three or more dimensions. Since then, the researches on this Stein phenomenon have received considerable attention. This paper surveys the theoretical study of the Stein phenomenon. The minimaxity of the James-Stein estimator and its improvements are demonstrated instructively, and various extensions and developments in Bayesian frameworks and non-normal distributions are reviewed. The paper shortly refers to the Stein phenomenon in confidence sets and a series of decision-theoretic results in estimation of a covariance matrix.

Key words and phrases: Simultaneous estimation, multivariate normal distribution, statistical decision theory, admissibility, minimaxity, Stein problem, shrinkage estimation, Bayes estimation, continuous and discrete exponential families, confidence sets, covariance matrix.

1 Introduction and Summary

The theory of the statistical parametric estimation has been remarkably developed for these twenty years in two main topics: the efficiency of the maximum likelihood estimator (MLE) in a large sample and the Stein problem for the inadmissibility of the MLE in a small sample. The large sample theory clarified the structure of the parametric estimation theory and provided the higher order asymptotic efficiency of the MLE and its differential-geometric interpretation. On the other hand, we have lots of practical situations where a large sample can not be expected in biometrics, engineering and others. In these cases, available informations, for instance, guessed ***a priori*** or contained in correlated data, may be used to modify usual estimators such as maximum likelihood or unbiased estimators for increasing their accuracies. The theoretical researches of the modification of the estimators in the small sample have been studied in a decision-theoretic framework since Wald(1950). For fundamental or general theories, see Zacks(1971) and Lehmann(1983). In particular, the researches on the Stein problem have received considerable attention since 1970, and a remarkably large amount of theoretical results have been produced.

When the simultaneous estimation of mean vector $\theta = (\theta_1, \dots, \theta_p)'$ based on random vector $X = (X_1, \dots, X_p)'$ having p-variate normal distribution $\mathcal{N}_p(\theta, I_p)$ is considered relative to a quadratic loss function, Stein(1956) presented the surprising, important and seminal result that the MLE X is inadmissible for $p \geq 3$ while it is admissible for $p = 1$ and 2. James and Stein(1961) succeeded in giving an explicit form of an estimator improving on X. This means that the MLE X of θ is inadmissible in the framework of the simultaneous estimation of several parameters, ***although the components X_i's of the estimator are separately admissible to estimate the corresponding one-dimensional parameters θ_i's***, which is called the ***Stein phenomenon*** or ***Stein problem.*** Stein(1973) developed an integration-by-parts approach, called the

Stein identity, which is very powerful and useful for deriving improved estimators. Since then, the researches on this Stein problem have been developed remarkably. For the good accounts, see Judge and Bock(1978), Berger(1985), Brandwein and Strawderman(1990), Hoffmann(1992), Mikhail and Vassily(1994), Robert(1994) and Rukhin(1995).

In this paper, we survey the theoretical results of the Stein problem from various aspects. Sections 2 gives an explanation of the Stein problem with a motivation from an empirical Bayes approach and a simple proof through the Stein identity. The improvements on the James-Stein estimator is treated in Section 3 through a new technique for improving on equivariant estimators. Several extensions and developments in the normal or non-normal distributions are provided in Sections 4 and 5. The Stein problem in a confidence set is surveyed in Section 6 and the related problem of estimating a covariance matrix is treated in Section 7.

Although our interest is limited to developments of theoretical results for a restriction of the volume of the paper, the concept of the Stein (or shrinkage) estimation has been applied to some practical problems. Since Efron and Morris(1972) indicated the empirical Bayesness of the James-Stein estimator, especially, the shrinkage estimation based on the empirical Bayes approach has been effectively used by Efron and Morris(1975) for estimation of batting averages of baseball players and for estimation of epidemic rates, by Fay and Herriot(1979), Battese *et al.*(1988), Prasad and Rao(1990), Ghosh and Rao(1994) and others for the small-area problem, by Tsutakawa *et al.*(1985), Clayton and Kaldor(1987) and others for estimation of mortality rates and indices and by Wahba(1985), Li(1985), Li and Hwang(1984) and Ansley *et al.*(1993) for smoothing data by a spline function (see also Copas(1983), Morris(1983) and Casella(1985)). Thus the shrinkage estimation originated by Stein has been evaluated as an effective procedure in a small sample from a practical point of view while the theoretical progress has been made markedly.

2 The Stein Phenomenon

Let $X = (X_1, \ldots, X_p)'$ be a random vector having p-variate normal distribution $\mathcal{N}_p(\theta, I_p)$ and consider the problem of estimating mean vector $\theta = (\theta_1, \ldots, \theta_p)'$ by estimator $\delta(X)$ based on X. Every estimator is evaluated in terms of the risk function relative to the quadratic loss function $||\delta(X) - \theta||^2$.

A natural estimator of θ is X and it is a maximum likelihood, uniformly minimum variance unbiased and minimax estimator. Also this estimation problem is invariant under the transformation $\Gamma X + d$, $\Gamma\theta + d$ for orthogonal matrix Γ, vector d when the estimator $\delta(X)$ satisfies the equivariance $\delta(\Gamma X + d) = \Gamma\delta(X) + d$, which implies $\delta(X) = X + d$ for vector d, and X is the best among this class of equivariant estimators.

For the admissibility of X, Stein(1956) presented the surprising, important and seminal result that X is inadmissible for $p \geq 3$ while it is admissible for $p = 1, 2$. This means that a usual estimator is inadmissible in the framework of the simultaneous estimation of several parameters, *although the components of the estimator are separately admissible to estimate the corresponding one-dimensional parameters*, and we call it the *Stein Phenomenon.* Every estimator equivariant with respect to the transformation ΓX and $\Gamma\theta$ is written by

$$\delta_\phi = \left\{1 - \frac{\phi(||X||^2)}{||X||^2}\right\} X \tag{1}$$

and Stein(1956) proved that there exists an estimator improving on X among the class (1).

James and Stein(1961) found an explicit form of an estimator better than X as

$$\delta^{JS} = \left\{1 - \frac{p-2}{||X||^2}\right\} X, \tag{2}$$

which is called the ***James-Stein Estimator***. Since it shrinks X towards the origin, such an estimator is generally called a ***Shrinkage Estimator***. Intuitive explanations of the Stein phenomenon are given in Stigler(1990), Brandwein and Strawderman(1990) and others.

Although one has an impression that the James-Stein estimator is artificial and strange from its form, it can be derived as a natural and empirical Bayes estimator from the Bayesian aspect (Efron-Morris(1972a), Robbins (1983)). Let us suppose that the parameter θ is a random variable whose prior distribution is $\mathcal{N}_p(\theta_0, a^{-1}I_p)$ where a is an unknown parameter and θ_0 is a known vector to be chosen beforehand. Then the posterior distribution of θ given X is $\mathcal{N}_p(\theta_0 + (1-\tau)(X-\theta_0), (1-\tau)I_p)$ for $\tau = a/(1+a)$, and the Bayes estimator of θ is thus given by

$$\hat{\theta}_B(\theta_0) = \theta_0 + (1-\tau)(X-\theta_0).$$

Since the hyperparameter τ is unknown, it is needed to be estimated from the marginal distribution of X, $\mathcal{N}_p(\theta_0, \tau^{-1}I_p)$. The marginal distribution of $||X-\theta_0||^2$ has $\tau^{-1}\chi_p^2$, and an unbiased estimator of τ is

$$\hat{\tau} = \frac{p-2}{||X-\theta_0||^2}.$$

By substituting $\hat{\tau}$ for τ in the Bayes estimator, we get the empirical Bays estimator

$$\hat{\theta}_{EB}(\theta_0) = \theta_0 + (1-\hat{\tau})(X-\theta_0) = \theta_0 + \left\{1 - \frac{p-2}{||X-\theta_0||^2}\right\}(X-\theta_0),$$

and $\hat{\theta}_{EB}(0)$ is identical to δ^{JS}. The value of θ_0 is given based on a prior information and $\hat{\theta}_{EB}(\theta_0)$ has a large risk-reduction for θ near θ_0, so that in the case where one can guess or take a prior information about θ, $\hat{\theta}_{EB}(\theta_0)$ brings a good estimate. Even if one can not suppose any exact prior information, the risk of $\hat{\theta}_{EB}(\theta_0)$ is always less than that of X and it does not yield any actual harm from a frequentist point of view, that is, $\hat{\theta}_{EB}(\theta_0)$ is robust for the prior information. The Bayes estimator depends on the prior knowledge completely while the knowledge is neglected in the maximum likelihood estimator. The empirical Bayes estimator is thus interpreted as an intermediate of the Bayes and maximum likelihood ones such that the drawbacks of both estimators are made up for.

The above motivation of δ^{JS} from the empirical Bayes aspect was presented by Efron and Morris(1972a), and they also proved the Stein phenomenon through the empirical Bayes approach. For the proof of the Stein phenomenon, we have two other approaches: one is an original method of James and Stein(1961), which utilizes the fact that a non-central chi square distribution is represented by a Poisson mixture of a central chi square distribution; the other is a method of Stein(1973), which uses an integration by parts. The latter is very simple, quite useful and powerful and so we introduce it here.

More generally we begin with obtaining sufficient conditions on the function ϕ for the estimator δ_ϕ improving on X. For absolutely continuous function $h(x)$ and its differential derivative $h'(x)$, an integration by parts gives the equality

$$E[(X_i - \theta_i)h(X_i)] = E[h'(X_i)], \tag{3}$$

which is called *the Stein identity*(Stein(1973,81)). Using this identity, we can write the risk function of δ_ϕ as

$$R(\theta,\delta_\phi) = E\left[p+\frac{\phi^2}{||X||^2}-2\sum_{i=1}^{p}(X_i-\theta_i)X_i\frac{\phi}{||X||^2}\right]$$
$$= E\left[p+\frac{\phi}{||X||^2}\{\phi-2(p-2)\}-4\phi'\right]. \quad (4)$$

The unknown parameters θ_i's thus disappear in the interior of the expectation $E[\cdot]$, which turns out to be an unbiased estimator of the risk function of δ_ϕ. Since $R(\theta,X)=p$, the conditions on $\phi(w)$ for δ_ϕ improving on X are given by solutions of the following differential inequality:

$$\phi(w)\{\phi(w)-2(p-2)\}/w-4\phi'(w)\le 0,$$

which is, for instance, satisfied by

(i) $\phi(w)$ is nondecreasing,

(ii) $0<\phi(w)\le 2(p-2)$.

Thus a class of the estimators δ_ϕ better than X is constructed. Since the conditions (i) and (ii) are satisfied by $\phi(w)=p-2$, the James-Stein estimator δ^{JS} is included in this class and the risk function is given by

$$R(\theta,\delta^{JS})=p-(p-2)^2E[||X||^{-2}],$$

which shows that the minimum risk is given when the non-centrality parameter $||\theta||^2$ or the mean vector θ is the origin. The usefulness of the Stein identity produced remarkable developments in this field.

When $||X||^2<p-2$, the James-Stein estimator yields an over-shrinkage and changes the sign of each X_i. For eliminating this drawback, the positive-part Stein estimator $\delta_+^{JS}=\max\{0,1-(p-2)/||X||^2\}X$ is considered and it is shown to be better than δ^{JS}. From the general theory that admissible estimators are analytic, it follows that δ_+^{JS} itself is inadmissible. However it was a big open problem to find an explicit estimator dominating δ_+^{JS}. Recently Shao and Strawderman(1994) successfully obtained the estimator

$$\delta_g^{SS}(a)=\delta_+^{JS}-\frac{a\,g(||X||^2)}{||X||^2}XI_{[p-2\le||X||^2\le p]}, \quad (5)$$

dominating δ_+^{JS}, where $g(t)$ is a function symmetric at $t=p-1$ satisfying $g(p-2)=g(p)=0$ and

$$g(t)=\begin{cases} t-p, & \text{if } p^*\le t\le p,\\ 2p^*-p-t, & \text{if } p-1\le t<p^*,\end{cases}$$

for suitable constants p^* and a. Since $\delta_g^{SS}(a)$ is not smooth, it is inadmissible still and the problem is not resolved completely yet. Sugiura and Takagi(1996) extends this result to the case where the covariance matrix is fully unknown.

An admissible estimator improving X (or minimax) was developed by Strawderman(1971) as

$$\delta_{GB}=\delta_{\phi_{GB}}=\{1-\phi_{GB}(||X||^2)/||X||^2\}X,$$

where

$$\phi_{GB}(w)=p-2-\frac{2}{\int_0^1 z^{p/2-2}e^{(1-z)w/2}dz}, \quad (6)$$

which is a generalized Bayes estimator against the prior distribution

$$\theta|\lambda \sim \mathcal{N}_p(0, \frac{1-\lambda}{\lambda} I_p), \quad \frac{1}{\lambda^2} I_{(0,1)}(\lambda) d\lambda.$$

In fact, the minimaxity of δ_{GB} can be easily shown by checking the conditions (i) and (ii). The admissibility can be verified from the results of Brown(1971) as pointed out in Berger(1980) and Brown and Hwang(1982). Also another type of admissible and minimax estimators was given by Alam(1973).

Letting $\nabla = (\partial/\partial x_1, \ldots, \partial/\partial x_p)'$ and $\nabla^2 = \sum \partial^2/\partial x_i^2$, Stein(1973,81) showed that if $f(x)$ satisfies the super-harmonic condition $\nabla^2 f(x) \leq 0$, then the estimator $\delta^{ST}(f) = X + \nabla \log f(X)$ dominates X, which suggests deep relations between the potential theory and the Stein phenomenon. The generalized Bayes estimators are represented by the form $\delta^{ST}(f)$ for the marginal density $f(x)$ and it was shown by Stein(1981) and Haff(1991) that if $f(x)$ satisfies the super-harmonic condition, then $\delta^{ST}(f)$ dominates X. Berger and Srinivasan(1978) characterized admissible estimators of θ through the generalized Bayesness. For other interesting discussions about admissibility, the characterization of admissibility of generalized Bayes estimators were given by Brown(1971), Srinivasan(1981), Brown and Hwang(1982) and Berger(1985), and the diffusion characterization of admissibility was given by Brown(1971), Johnstone(1984) and Eaton(1992).

3 Improvements on the James-Stein Estimator

The inadmissibility of the James-Stein estimator is stated in the previous section. We shall construct a broad class of estimators improving on the James-Stein estimator. From the expression (4) of the risk function based on the Stein identity, we get the condition for δ_ϕ to dominate δ^{JS} as

$$\{\phi(w) - (p-2)\}^2 - 4w\phi'(w) \leq 0.$$

From this inequality, however, we cannot find any meaningful or general solutions on $\phi(w)$ as stated in Rukhin(1995), which demonstrates one of limitations of the characteristics through the Stein identity. Kubokawa(1994a) and Takeuchi(1991) proposed a new approach to improving on equivariant estimators. Their idea is to express a difference of risk functions through an integral, and we shall call it the IERD (*Integral Expression of Risk Difference*) method. Letting $\lim_{w\to\infty}\phi(w) = p-2$ and using (4), we apply the IERD method to have

$$\begin{aligned}
&R(\theta, \delta^{JS}) - R(\theta, \delta_\phi) \\
&= E\left[\frac{\phi(\infty)}{||X||^2}\{\phi(\infty) - 2(p-2)\} - \frac{\phi(||X||^2)}{||X||^2}\{\phi(||X||^2) - 2(p-2)\}\right] + 4E[\phi'(||X||^2)] \\
&= E\left[\int_1^\infty \frac{d}{dt}\left\{\frac{\phi(t||X||^2)}{||X||^2}\{\phi(t||X||^2) - 2(p-2)\}\right\} dt\right] + 4E[\phi'(||X||^2)] \\
&= 2E[\int_1^\infty \{\phi(t||X||^2) - (p-2)\}\phi'(t||X||^2) dt] + 4E[\phi'(||X||^2)] \\
&= 2\int_0^\infty \int_1^\infty \{\phi(tx) - (p-2)\}\phi'(tx) f_p(x;\lambda) dt dx + 4E[\phi'(||X||^2)],
\end{aligned} \tag{7}$$

where $\lambda = ||\theta||^2$ and $f_p(x;\lambda)$ denotes a density of a non-central chi square distribution with p degrees of freedom and non-centrality parameter λ. Making the transformations $w = tx$ and $y = w/t$, we rewrite the first term in the r.h.s. of the extreme equation of (7) as

$$2\int_0^\infty \int_1^\infty \{\phi(w) - (p-2)\}\phi'(w) f_p(\frac{w}{t};\lambda)\frac{1}{t} dt dw$$

$$= 2\int_0^\infty \int_0^w \{\phi(w) - (p-2)\}\phi'(w) f_p(y;\lambda) y^{-1} dy dw,$$

so that

$$R(\theta, \delta^{JS}) - R(\theta, \delta_\phi)$$
$$= 2\int_0^\infty \phi'(w) \left\{ \{\phi(w) - (p-2)\} \int_0^w y^{-1} f_p(y;\lambda) dy + 2 f_p(w;\lambda) \right\} dw. \quad (8)$$

From (8) and the inequality

$$f_p(w;\lambda) / \int_0^w y^{-1} f_p(y;\lambda) dy \geq f_p(w) / \int_0^w y^{-1} f_p(y) dy$$

for $f_p(y) = f_p(y; 0)$, we see that δ_ϕ is better than δ^{JS} if the following conditions (a) and (b) hold:

(a) $\phi(w)$ is nondecreasing in w and $\lim_{w\to\infty}\phi(w) = p-2$,
(b) $\phi(w) \geq \phi_0(w)$, where

$$\begin{aligned}\phi_0(w) &= p - 2 - 2 f_p(w) / \int_0^w y^{-1} f_p(y) dy \\ &= \int_0^w f_p(y) dy / \int_0^w y^{-1} f_p(y) dy.\end{aligned}$$

It can be easily checked that the functions $\phi^+(w) = \min(w, p-2)$ and $\phi_0(w)$ satisfy the conditions (a) and (b). $\phi^+(w)$ yields the positive-part Stein estimator. Noting that $\phi_0(w) = \phi_{GB}(w)$, the admissible (hierarchical Bayes) estimator δ_{GB} given by Strawderman(1971) is seen to be better than the James-Stein estimator. Kubokawa(1991) applied the method of Brewster and Zidek(1974) in estimation of a variance and showed that δ_{GB} is derived as a limit when the number of partitions tends to infinity. It is interesting to note that the improvement on the James-Stein estimator is strongly related to the problems of estimating a variance of a normal distribution with an unknown mean and of estimating a positive normal mean as suggested by Rukhin(1992b), which established that their three problems are equivalent asymptotically.

From a practical sense, it is important to discuss the case where a variance of the underlying normal distribution is unknown. For instance, a canonical form of a multiple regression model is given by

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

where random variables X and S are independent and χ_n^2 denotes a chi square distribution with degrees of freedom n. When we consider the simultaneous estimation of the mean vector θ under the loss function $||\hat{\theta} - \theta||^2/\sigma^2$, James and Stein(1961) showed that $\delta^{JS} = \{1 - (p-2)(n+2)^{-1} W^{-1}\} X$ for $W = ||X||^2/S$ is better than X for $p \geq 3$. More generally we consider the shrinkage estimator

$$\delta_\phi = \left\{ 1 - \frac{\phi(W)}{W} \right\} X$$

for absolutely continuous function $\phi(\cdot)$. Using the Stein identity and the chi square identity given by

$$E[(S - n\sigma^2) h(S)] = 2\sigma^2 E[S h'(S)]$$

for absolutely continuous function $h(\cdot)$, Efron and Morris(1976a) gave an unbiased estimator of the risk function of δ_ϕ as

$$p - \left\{ \frac{\phi(W)}{W} \{2(p-2) - (n+2)\phi(W)\} + 4\phi'(W)\{1 + \phi(W)\} \right\},$$

which implies that δ_ϕ improves on X if $\phi(w)$ is nondecreasing and if $0 < \phi(w) \leq 2(p-2)/(n+2)$. By the arguments based on the IERD method, Kubokawa(1994a) showed that δ_ϕ dominates δ^{JS} if

(a) $\phi(w)$ is nondecreasing in w and $\lim_{w\to\infty}\phi(w) = (p-2)/(n+2)$,
(b) $\phi(w) \geq \phi_0(w)$, where for $h(w) = \int_0^\infty v f_n(v) f_p(vw)dv$,

$$\phi_0(w) = \frac{(p-2)\int_0^w s^{-1}h(s)ds - 2h(w)}{(n+2)\int_0^w s^{-1}h(s)ds + 2h(w)}.$$

The estimator δ_{ϕ_0} is a generalized Bayes estimator(Lin and Tsai(1973)), but the proof of the admissibility is so difficult.

It is interesting to note that the James-Stein estimator δ^{JS} is represented by

$$\delta^{JS} = \left\{1 - \hat{\sigma}_0^2 \frac{p-2}{||X||^2}\right\} X$$

where $\hat{\sigma}_0^2 = (n+2)^{-1}S$ is the best affine equivariant estimator of σ^2 relative to the quadratic loss. The problem of improving on $\hat{\sigma}_0^2$ by using the information contained in X has been studied by Stein(1964), Brown(1968), Brewster and Zidek(1974) and others. For a good review of this field, see Maatta and Casella(1990). The improved estimator of Stein(1964) is given by

$$\hat{\sigma}_{ST}^2 = \min\left\{\frac{S}{n+2}, \frac{S+||X||^2}{n+p+2}\right\}.$$

George(1990) conjectured that δ^{JS} is improved on by the estimator $(1 - \hat{\sigma}_{ST}^2(p-2)/||X||^2)X$ given by replacing $\hat{\sigma}_0^2$ with $\hat{\sigma}_{ST}^2$, and this conjecture was affirmatively verified by Kubokawa *et al.*(1993) and Berry(1994).

4 Extensions and Developments in Normal Distributions

Various extensions and developments for the Stein phenomenon have been studied. We here survey them for normal distributions.

The risk of the James-Stein estimator δ^{JS} is given by $p - (p-2)^2 E[1/\sum_{i=1}^p X_i^2]$, which implies that the risk-gain is quite small if one of X_i^2's is very large. For this drawback, some modifications of the James-Stein estimator were proposed by Efron and Morris(1972b) and Stein(1981). The truncated estimator proposed by Stein(1981) is given by, componentwise

$$\delta_i^{(\ell)} = \left(1 - \frac{(\ell-2)\min\{1, Z_{(\ell)}/|X_i|\}}{\sum_{j=1}^p X_j^2 \wedge Z_{(\ell)}^2}\right) X_i,$$

where ℓ is a suitable constant, $a \wedge b = \min(a,b)$ and

$$Z_{(1)} < Z_{(2)} < \cdots < Z_{(p)}$$

designate the order statistics of $Z_1, Z_2, \ldots, Z_p$ for $Z_i = ||X_i||$. Then the risk function of $\delta^{(\ell)}$ is represented by

$$R(\theta, \delta^{(\ell)}) = p - (\ell-2)^2 E[1/\sum_{j=1}^p X_j^2 \wedge Z_{(\ell)}^2],$$

which shows that some large values X_j^2's do not affect the risk of $\delta^{(\ell)}$. The discussions of a choice of ℓ and a robust estimation were given by Dey and Berger(1983), Berger and Dey(1985).

The multinormal model with unequal variances, $X = (X_1, \ldots, X_p)' \sim \mathcal{N}_p(\theta, D)$, for $D = \text{diag}(d_1, \ldots, d_p)$ is practically important, and the shrinkage procedures have been studied by Efron and Morris(1975), Fay and Herriot(1979) and Morris(1983) in some applications. A usual minimax shrinkage estimator is given by Berger(1976) as

$$\delta_i^{MS} = \left\{1 - \frac{p-2}{\Sigma_i X_i^2/d_i} \frac{1}{d_i}\right\} X_i, \quad i = 1, \ldots, p$$

but the risk-reduction is quite small when one variance of d_i's is much smaller than others. On the other hand, empirical Bayes estimators are much shrunken and they give practically reasonable estimates while their minimaxity is not guaranteed from a frequentist point of view. Using the implicit function theorem, Shinozaki and Chang(1994) developed a minimax empirical Bayes estimator as

$$\delta_i^{SC} = \left\{1 - \frac{\Sigma d_i^2 - 2d_{max}^2}{(p-2)d_{max}^2} \frac{d_i}{\hat{a} + d_i}\right\} X_i, \quad i = 1, \ldots, p$$

where $\hat{a}$ is a solution of the equation

$$\sum_{i=1}^{p} \frac{X_i^2}{\hat{a} + d_i} = p - 2,$$

and provided numerical comparison of the estimators.

When θ is guessed to be in subspace V based on a prior information, it is reasonable to consider the Stein estimator $\delta^{JS}(V)$ shrunken towards V, and a large risk-reduction is expected for θ in/near V. The prior information, however, may be vague. When several subspaces $V_1, \ldots, V_k$ up to θ are guessed, George(1986a,b) proposes an adaptive, random weighted combined estimator of the Stein estimator $\delta^{JS}(V_i)$ shrunken towards V_i, given by

$$\sum_{i=1}^{k} \rho_i(X)\delta^{JS}(V_i),$$

where $\rho_i(X)$ is an adaptively weighting function which has a high weight when a risk-reduction of $\delta^{JS}(V_i)$ is expected to be large.

The researches on the robust Bayes estimation were developed by Berger(1980b). In the model $X \sim \mathcal{N}_p(\theta, \Sigma)$ for known matrix Σ, the Bayes estimator of θ against prior distribution $\pi_0 : \ \theta \sim \mathcal{N}_p(\mu, A)$ is given by

$$\delta^{\pi_0} = X - \Sigma(\Sigma + A)^{-1}(X - \mu),$$

where the hyperparameters μ and A are determined subjectively. Let $A - \Sigma \geq 0$ and let $B(\lambda) = \lambda^{-1}A - \Sigma$ for $0 < \lambda < 1$. For $k > 0$, consider the hierarchical prior distribution

$$g_k(\theta) = \int_0^1 \frac{\lambda^{k-1-(p/2)}}{|B(\lambda)|^{1/2}} e^{-(\theta-\mu)'B(\lambda)^{-1}(\theta-\mu)/2} d\lambda,$$

which is an extension of Strawderman(1971). Then the generalized Bayes estimator is given by

$$\begin{aligned} \delta^{RB} &= X - \frac{r_k((X-\mu)'(\Sigma+A)^{-1}(X-\mu))}{(X-\mu)'(\Sigma+A)^{-1}(X-\mu)} \Sigma(\Sigma+A)^{-1}(X-\mu) \\ r_k(v) &= \frac{v \int_0^1 \lambda^k \exp\{-\lambda v/2\} d\lambda}{\int_0^1 \lambda^{k-1} \exp\{-\lambda v/2\} d\lambda}, \end{aligned}$$

which is approximated to $r_k(v) = \min\{p-2, v\}$. If

$$(X-\mu)'(\Sigma + A)^{-1}(X-\mu) \le p-2,$$

this is the subjective Bayes estimator δ^{π_0} and otherwise a Stein type shrinkage estimator. In other words, when the prior information is exact, the subjective Bayes estimator is selected, otherwise the Stein type shrinkage estimator or the usual estimator X yields and so δ^{RB} has a robust property for the subjective prior information.

The robust Bayes estimator δ^{RB} is not always minimax for any Σ and A. On the basis of the idea of Bhattacharya(1966), Berger(1982) developed a minimax estimator incorporating the subjective hyperparameters μ and A. For simplicity, let $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, $A = (a_1, \ldots, a_p)$ and $q_1 \ge q_2 \ge \ldots \ge q_p > 0 \equiv q_{p+1}$ for $q_i = \sigma_i^4/(\sigma_i^2 + a_i)$. Then the minimax and robust Bayes estimator is given by, componentwise

$$\begin{aligned}\delta_i^{MB} &= X_i - \frac{\sigma_i^2}{\sigma_i^2 + a_i}(X_i - \mu_i) \\ &\times \left[\frac{1}{q_i}\sum_{j=i}^{p}(q_j - q_{j+1})\min\left\{1, \frac{2(j-2)^+}{\sum_{k=1}^{j}(X_k - \mu_k)^2/(\sigma_k^2 + a_k)}\right\}\right].\end{aligned}$$

The robust Bayes estimation against an ε-contaminated prior distribution $\pi = (1-\varepsilon)\pi_0 + \varepsilon q$ for subjective prior π_0 has been studied by Berger(1985), Berger and Berliner(1986). Letting $\hat{q}$ be the distribution q maximizing the marginal distribution of x, $m(x|q)$, for the prior distribution π, they called $\hat{\pi} = (1-\varepsilon)\pi_0 + \varepsilon\hat{q}$ the *ML-II prior distribution* and obtained the Bayes estimator against the prior as

$$\delta^{\hat{\pi}}(x) = \hat{\lambda}_\varepsilon(x)\delta^{\pi_0}(x) + (1 - \hat{\lambda}_\varepsilon(x))\delta^{\hat{q}}(x),$$

where δ^{π_0} designates the subjective Bayes estimator against the prior π_0 and $\delta^{\hat{q}}$ is the empirical Bayes estimator against the prior $\hat{q}$. $\delta^{\hat{\pi}}$ is an intermediate between δ^{π_0} and $\delta^{\hat{q}}$ where the weight depends on the rate of the contamination ε. For a series of theoretical results concerning the robust Bayes estimation, see Berger and Berliner(1984). Berliner(1985) indicated that the problem of finding an estimator minimizing the Bayes risk uniformly against a class of prior distributions Γ comes down to that of the simultaneous estimation of scale parameters of independent Gamma distributions, and he showed that δ^{JS} is a robust Bayes estimator against the class Γ.

In connection with recent developments of Bayesian computations like the Gibbs sampling, the hierarchical Bayes estimation has been studied by Berger(1985), Ghosh and Sinha(1988), Berger and Robert(1990). In the model $X \sim \mathcal{N}_p(\theta, \sigma^2 I)$ for known σ^2, for instance, suppose the first step prior distribution $\theta \sim \mathcal{N}_p(\beta\mathbf{1}, \sigma_\pi^2 I)$ for $\mathbf{1} = (1, \ldots, 1)'$ and for the hyperparameters β and σ_π^2, suppose the second step prior distributions $\beta \sim \mathcal{N}(\beta^0, A)$ and $\sigma_\pi^2 \sim \pi_2^2(\sigma_\pi^2)$. Then the Bayes estimator can be obtained against the hierarchical prior distributions. If the second step is the non-informative prior distribution, that is, $A \to \infty$ and $\pi_2^2(\sigma_\pi^2) = 1$, then the hierarchical Bayes estimator is given by

$$\begin{aligned}\delta^{HB} &= X - E^{\pi_2^2(\sigma_\pi^2|X)}\left[\frac{\sigma^2}{\sigma^2 + \sigma_\pi^2}\right](X - \bar{x}\mathbf{1}) \\ \pi_2^2(\sigma_\pi^2|X) &\propto (\sigma^2 + \sigma_\pi^2)^{-(p-1)/2}\exp\left\{-\frac{||X - \bar{x}\mathbf{1}||^2}{2(\sigma^2 + \sigma_\pi^2)}\right\}, \quad \bar{x} = p^{-1}\sum_{i=1}^{p} X_i.\end{aligned} \tag{9}$$

It is seen that the drawback of the over-shrinkage in the empirical Bayes estimation does not arise in the hierarchical Bayes estimation. The hierarchical Bayes estimators in more general models, the derivations of unbiased estimators of their risk functions and general conditions for their minimaxity were discussed in Berger and Robert(1990).

In the case where the covariance matrix is fully unknown, it is sufficient to substitute an estimator for the unknown covariance matrix with a modification of a constant as long as we consider a loss function invariant under affine transformations (James and Stein(1961), Baranchik(1970), Lin and Tsai(1973), Bock(1975), Alam(1975), Efron and Morris(1976a)). For a non-invariant loss function, however, the uniform domination was recognized to be a difficult issue (Berger *et al.*(1977)). For this open problem, Gleser(1986) succeeded in derivation of a class of improved estimators, and Honda(1991) and Tan(1991) extended it to the multivariate regression models and growth curve models, respectively.

It is interesting to note that the estimation of a matrix mean is related to that of a covariance matrix. When $p \times r$ random matrix X has $\mathcal{N}_{p\times r}(\Theta, I_p \otimes I_r)$, Efron and Morris(1972a) showed that the MLE X is improved on by the empirical Bayes estimator

$$\hat{\Theta}_0^{EM} = \{I_p - (r-p-1)(XX')^{-1}\}X.$$

Stein(1973) discussed the further domination of $\hat{\Theta}_0^{EM}$ and through the empirical Bayes arguments Efron and Morris(1976b) showed that the improvement on $\hat{\Theta}_0^{EM}$ is reduced to the estimation of the inverse of a covariance matrix and developed the improved estimator

$$\hat{\Theta}_1^{EM} = \hat{\Theta}_0^{EM} - (p^2+p-2)(\mathrm{tr}XX')^{-1}X.$$

Zheng(1988) extended the result of Stein(1981) to the case of the matrix mean. A canonical form of a multivariate regression model with an unknown error covariance matrix is represented by

$$X \sim \mathcal{N}_{p\times r}(\Theta, \Sigma \otimes I_r), \qquad S \sim \mathcal{W}_p(n, \Sigma),$$

and various types of shrinkage estimators for Θ have been proposed by Bilodeau and Kariya(1989), Konno (1991) and Shieh(1993). Some extensions to a growth curve model have been done by Kubokawa *et al.* (1992) and Kariya *et al.*(1994, 96), which developed an interesting domination result about a double shrinkage estimation.

The Stein phenomenon has been studied in various other situations: by Chang(1982) and Sengupta and Sen(1991) for ordered restrictions of parameters $\theta_1, \ldots, \theta_p$, by Takada(1984) and Ghosh *et al.*(1987) for a sequential analysis and by Stein(1960), Baranchick(1973), Takada(1979), Zidek(1978) for a multiple regression problem with random ***independent*** variables correlated to ***dependent*** variables. Brown(1990) discussed the latest issue in two situations of the conditional and unconditional inference given an ancillary statistic and indicated the interesting result, called ***Ancillarity Paradox*** that the decision-theoretic conclusions are different in the respective cases. Brown's paper includes lots of interesting discussions and comments. The Stein phenomenon has been investigated relative to other criteria for comparing estimators: by Brown(1973), Shinozaki(1980) and Hwang(1985) for whether the Stein domination holds uniformly in a class of loss functions and by Sen *et al.*(1989) for the Pitman closeness criterion.

George(1991), Krishnamoorthy(1992) and Sarkar(1994) considered the problem of estimating a common mean vector of two multivariate normal distributions with possibly different unknown variances and established an innovative result that a linear estimator is dominated by a shrinkage estimator even if no statistics for estimating variances are available.

The problem of estimating the loss functions of the MLE and the James-Stein estimator was treated by Johnstone(1987) and the inadmissibility of usual unbiased estimators was shown.

Lu and Berger(1989a) studied the estimation of the loss functions of the positive-part Stein and the generalized Bayes estimators and Lele(1992) discussed similar issues for an exponential family.

5 The Stein Phenomenon in Non-Normal Distributions

5.1 Spherical distributions

The Stein phenomenon has been shown for non-normal distributions. Brandwein and Strawderman(1990) presented a good survey for spherically symmetric and elliptical distributions. For a distribution of η, $G(\eta)$, the $G(\eta)$-scale mixture of a normal distribution

$$f(||x-\theta||) = \int \left(\frac{\eta}{2\pi}\right)^{p/2} e^{-(\eta/2)||x-\theta||^2} G(d\eta)$$

has been studied by Strawderman(1974), Berger(1975), Srivastava and Bilodeau(1989) and Chou and Strawderman(1990), and the Stein phenomenon has been shown for $p \geq 3$. Bravo and MacGibbon(1988a) gave a domination result in a scale mixture of a normal distribution with an unknown variance.

For the general spherically symmetric distributions without any restrictions to scale mixtures of normal distributions, Brandwein and Strawderman(1978, 91a,b) and Bock(1985) showed the Stein phenomenon for $p \geq 4$, and Ralescu *et al.*(1992) established it for $p = 3$ in a uniform distribution on a compact support. In an elliptically contoured distribution, Cellier *et al.*(1989) proved that the condition for the James-Stein estimator dominating the least squares estimator does not depend on the form of the distribution, that is, the domination is robust. By using an elementary stochastic analysis, Evans and Stark(1996) recently generalized the Stein identity to a large class of distributions including spherically symmetric ones, and proved that shrinkage estimators dominate X for $p \geq 3$.

5.2 A continuous exponential family

Hudson(1978) extended the Stein identity for the normal distribution to the continuous exponential family and derived an improved estimator corresponding to the James-Stein estimator. Especially much attention has been paid to the simultaneous estimation of scale parameters of gamma distributions. Let $X_1, \dots, X_p$ be p independent random variables, X_i having the gamma distribution

$$f(x_i, \theta_i) = \exp(-\theta_i x_i) x_i^{\alpha_i - 1} \theta_i^{\alpha_i} / \Gamma(\alpha_i), \quad x_i > 0$$

and consider the simultaneous estimation of $\theta = (\theta_1, \dots, \theta_p)'$ and $\theta^{-1} = (\theta_1^{-1}, \dots, \theta_p^{-1})'$. Employing the loss function

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

for the estimation of θ^{-1}, Berger(1980a) obtained a differential inequality for the domination and developed solutions for $p \geq 2$ when $m = -2, -1, 1$ and for $p \geq 3$ when $m = 0$. It is the surprising result that the Stein phenomenon (especially called *Berger Phenomenon*) holds even for $p = 2$ while the dimension is at least three for the normal distribution. Ghosh and Parsian(1980) extended the class of the solutions. For the general loss function

$$L(\delta, \theta) = \sum_{i=1}^{p} c_i \theta_i^{m_i} (1 - \delta_i \theta_i)^2, \quad c_i > 0$$

DasGupta(1986) showed that the best equivariant estimator $\delta_{0,i} = X_i/(\alpha_i+1), i = 1, \ldots, p$ is dominated for $p \geq 2$ by the shrinkage estimator

$$\begin{aligned} \delta_i^{DG} &= \frac{X_i}{\alpha_i+1}(1+\phi_i(X)), \quad i = 1, \ldots, p, \\ \phi_i(X) &= -c(\text{sgn} m_i) x_i^{m_i/2} \prod_{j=1}^{p} x_j^{-m_j/2p}, \quad c > 0. \end{aligned} \tag{10}$$

Also he proved the inadmissibility of the best equivariant estimator of θ for $p \geq 2$ under the loss

$$L(\delta, \theta) = \sum_{i=1}^{p} c_i \theta_i^{m_i} (\delta_i/\theta_i - 1)^2, \quad c_i > 0.$$

DasGupta(1989) established a general theory concerning the domination in the simultaneous estimation of positive parameters such as scale parameters of general distributions, eigenvalues of a covariance matrix and other examples.

Other topics have been studied by Chou(1988) for an extension of Hudson's identity to a multi-dimensional exponential family, by Dey *et al.*(1987) for the Stein phenomenon for $p \geq 3$ under the Kullback-Leibler loss, by Haff and Johnson(1986) for super-harmonic conditions in the exponential family, by Ki and Tsui(1990) for a multiple shrinkage estimation, by Dey(1990) for estimation of scale parameters of a mixture distribution and by Bilodeau(1988) for estimation of $\theta^f = (\theta_1^{f_1}, \ldots, \theta_p^{f_p})'$, ($f_i = 1$ or $f_i = -1$). Shinozaki(1984) showed that the Stein phenomenon arises by considering the simultaneous estimation for uniform, double exponential and t-distributions. He also presented the interesting result that even if the underlying distribution is unknown, the Stein effect appears when the second and fourth moments of the usual estimator are known. The result in an inverse Gaussian distribution was given by Bravo and MacGibbon(1988b).

5.3 A discrete exponential family

The Stein phenomenon is known for the discrete exponential family including Poisson and negative binomial distributions. Let $X_1, \ldots, X_p$ be p independent random variables, X_i having $\mathcal{P}o(\theta_i)$. For the estimation of $\theta = (\theta_1, \ldots, \theta_p)'$, two types of loss functions $L_0(\delta, \theta)$ and $L_1(\delta, \theta)$ have been treated in the literature where

$$L_m(\delta, \theta) = \sum_{i=1}^{p} (\delta_i - \theta_i)^2/\theta_i^m, \quad m = 0, 1.$$

For the L_1 loss, Clevenson and Zidek(1975) obtained the innovative result that

$$\delta^{CZ} = \left[1 - \frac{\beta+p-1}{\sum_{i=1}^{p} X_i + \beta + p - 1}\right] X_i, \quad i = 1, \ldots, p$$

improves on $X = (X_1, \ldots, X_p)'$ for $p \geq 2$ and $0 \leq \beta \leq p-1$ and gave admissible and generalized Bayes estimators dominating X. Ghosh and Parsian(1981) constructed a class of generalized Bayes estimators, Tsui and Press(1982) provided various classes of improved estimators, and Tsui(1984) proved the superiority of the Clevenson-Zidek estimator in a negative binomial distribution.

For the loss L_0, on the other hand, Hudson(1978) derived an identity in the discrete exponential family to give a difference inequality for the domination, which was resolved for $p \geq 3$. A series of results was unified and summarized by Hwang(1982), Ghosh *et al.*(1983) and Chou(1991) in a general discrete exponential family under the general loss function $\sum(\delta_i - \theta_i)^2/\theta_i^{m_i}$. For comparison of several proposed estimators for the L_0 and L_1 losses, see Jun(1993).

Other topics have been studied by Ghosh and Yang(1988) for the Stein phenomenon for $p \geq 3$ under the Kullback-Leibler loss, by Dey and Chung(1992) for a discrete mixture distribution, by Johnson(1987) for a domination result in a binomial distribution, by Brown(1981), Johnstone(1984) and Brown and Farrell(1985) for admissibility in discrete distributions, by Gupta *et al.*(1989) and Albert(1987) for contingency tables, by Lwin and Maritz(1989) and Gupta and Saleh(1996) for a multinomial distribution and by Kuo(1986) for a Dirichlet distribution.

5.4 Asymptotic theories

In nonparametric models, the asymptotic improvements on L-, M-, R- estimators by the Stein effect in terms of a distributional risk criterion have been developed by Sen and Saleh(1985, 87), Saleh and Sen(1985) and Shiraishi(1991) and others. The Stein phenomenon in time series models was shown by Chaturvedi *et al.*(1993), Nickerson and Basawa(1992) and Koul and Saleh(1993), and the decision-theoretic results in a Gaussian process were given by Spruill(1982), Majumdar(1994) and Mandelbaum and Shepp(1987).

Yanagimoto(1994) noticed that the relation that

$$E[||X - \theta||^2] = E[||X - \delta^{JS}||^2] + E[||\delta^{JS} - \theta||^2]$$

holds for the James-Stein estimator δ^{JS} and called it the *Mean Pythagorean Relation.* Recently Eguchi and Yanagimoto(1994) showed that the mean Pythagorean relation holds asymptotically for every regular distributions. Let $(g^{ij}(\theta))_{ij}$ be the inverse of the Fisher information for density $p(x;\theta)$, $\theta \in \mathbf{R}^p$, and denote

$$(\mathrm{grad} f)^i = \sum_{j=1}^{p} g^{ij}(\theta) \frac{\partial}{\partial \theta_j} f(\theta).$$

When we consider the estimator

$$\hat{\theta}^* = \hat{\theta} + \frac{1}{n} \mathrm{grad}\ u(\hat{\theta})$$

for the MLE $\hat{\theta}$, the mean Pythagorean relation

$$R_n(\hat{\theta}, \theta) = R_n(\hat{\theta}, \hat{\theta}^*) + R_n(\hat{\theta}^*, \theta) + O(n^{-1})$$

holds asymptotically if and only if $\exp(u)$ satisfies the super-harmonic condition given by Stein (1981), where $R_n(\hat{\theta}, \theta)$ designates the risk function with respect to the Kullback-Leibler loss. In other words, $\hat{\theta}$ is asymptotically improved on by $\hat{\theta}^*$ if $\exp(u)$ satisfies the super-harmonic condition. Also Komaki(1996) provided an differential-geometric interpretation for the asymptotic domination of a shrinkage estimator.

6 Confidence Sets

One of marked developments of the Stein problem in the 1980s is the construction of improved confidence sets. A usual confidence set in the model $X \sim \mathcal{N}_p(\theta, I_p)$ is given by $C_0(X) = \{\theta; ||\theta - X||^2 \leq c\}$ where c is a constant satisfying $P(\chi_p^2 \leq c^2) = \gamma$ for confidence coefficient $1 - \alpha = \gamma$.

It is said that a confidence set $C(X)$ improves on $C_0(X)$ if the following two criteria are satisfied:

(I) $P_\theta\{\theta \in C(X)\} \geq P_\theta\{\theta \in C_0(X)\}$ for every θ,

(II) (the volume of $C(X)$)$\leq$(the volume of $C_0(X)$) for almost all X.

Since Brown(1966) and Joshi(1967) proved the inadmissibility of $C_0(X)$ for $p \geq 3$, various improved confidence sets have been proposed. Berger(1980b) obtained a confidence set based on the generalized Bayes estimator, and indicated that it has a marked improvement in both of the coverage probability and the volume while it has a bit computational troublesomeness and no uniform improvement. Hwang and Casella(1982) gave the first work of developing a confidence set improving on $C_0(X)$ in an explicit form, which is given by

$$C(a,X) = \{\theta; ||\theta - \delta^+(a,X)||^2 \leq c\}$$

for the positive-part Stein estimator $\delta^+(a,X) = \max\{0, 1 - a/||X||^2\}X$. In fact $C_0(X)$ can be proved to be improved by $C(a,X)$ in terms of (I) if $p \geq 4$ and a satisfies $0 < a \leq a_c$ where a_c is a solution of the equation

$$\{\sqrt{c} + \sqrt{c + a_c}\}^{p-3} = (a_c)^{(p-3)/2} e^{\sqrt{a_c c}}.$$

The condition on a was further extended in Hwang and Casella(1984) so that the domination holds for $p = 3$. Thus the simple and useful confidence set which guarantees the uniform improvement is presented. Extensions to the spherical symmetric distribution including uniform, double exponential and multivariate t-distributions and construction of improved confidence sets were studied by Hwang and Chen(1986) and Ki and Tsui(1985), and an improved confidence set shrinking towards a linear subspace was given by Casella and Hwang(1987).

In the models with unknown scale parameter(variance), Chen and Hwang(1988) and Hwang and Ullah(1994) demonstrated that a usual confidence set based on F-statistic is asymptotically or numerically dominated by a shrinkage confidence set. Robert and Casella(1990) succeeded in the derivation of an exact dominance result for the spherically symmetric distributions including a multivariate t-distribution, but the normal distribution is not contained and it remains still open as an interesting problem under the normality.

The above improvements are done in terms of (I) while the same volume holds. From a natural sense of a confidence set, however, the improvement in terms of (II) may be desirable. By shrinking the sphere $C_0(X)$ towards the origin, Shinozaki(1989) succeeded in the derivation of improved confidence sets in the sense of minimizing the volumes while the same confidence coefficient holds.

The empirical Bayes confidence sets have been studied by Morris(1983) and Casella and Hwang (1983). Let $I_C(\theta) = 1$ if $\theta \in C$ and $= 0$ if $\theta \notin C$. Then $C_0(X)$ is minimax relative to the loss

$$L(\theta, C) = k_0(\text{the volume of } C) - I_C(\theta),$$

for $k_0 = \exp(-c^2/2)/(2\pi)^{p/2}$. Casella and Hwang(1983) constructed an empirical Bayes confidence set with respect to the loss $L(\theta, C)$. The Bayes confidence set against the prior distribution $\theta \sim \mathcal{N}_p(0, \tau^2 I)$ is given by

$$C_B(X) = \left\{\theta; ||\theta - BX||^2 \leq B[c^2 - p\log B]\right\}$$

for $B = \tau^2/(\tau^2+1)$. Estimating τ or B from the marginal distribution, we get the empirical Bayes confidence set

$$C_{EB}(X) = \left\{\theta; ||\theta - \delta^+(p-2, X)||^2 \leq v_E(||X||)\right\},$$

where

$$v_E^2(||X||) = \begin{cases} \left(1-\frac{p-2}{c^2}\right)\left[c^2 - p\log\left(1-\frac{p-2}{c^2}\right)\right], & \text{if } ||X|| \le c, \\ \left(1-\frac{p-2}{||X||^2}\right)\left[c^2 - p\log\left(1-\frac{p-2}{||X||^2}\right)\right], & \text{if } ||X|| \ge c. \end{cases}$$

While $C_{EB}(X)$ has a smaller volume than $C_0(X)$, no analytical result is given for the coverage probability. It is numerically demonstrated that $C_{EB}(X)$ satisfies requested confidence coefficients for small p ($p \ge 5$).

As a problem related to the estimation of the loss function stated in Section 4, the estimation of the accuracy of the usual confidence set $I_{C_0(X)}(\theta) = I(\theta \in C_0(X))$ is discussed. Decision-theoretic results about the admissibility of the unbiased estimator γ have been given by Lu and Berger(1989b), Hwang and Brown(1991), Robert and Casella(1994), George and Casella(1994), and Casella *et al.*(1994). A similar estimation problem is also considered in the testing hypothesis, and decision-theoretic results were given(Hwang *et al.*(1992)).

7 Estimation of a Covariance Matrix

In this section, we survey the estimation of the covariance matrix, which is related to the simultaneous estimation of the matrix mean and of the scale parameters.

Let $p \times p$ random matrix S have Wishart distribution $\mathcal{W}_p(n, \Sigma)$ with mean $n\Sigma$ and consider the estimation of Σ by $\hat{\Sigma}$ relative to the Kullback-Leibler loss $\mathrm{tr}\hat{\Sigma}\Sigma^{-1} - \log|\hat{\Sigma}\Sigma^{-1}| - p$. It is known that unbiased estimator $\hat{\Sigma}_0 = n^{-1}S$ has a drawback that eigenvalues of $\hat{\Sigma}_0$ spread out more than those of Σ, and for modifying $\hat{\Sigma}_0$, it is necessary to shrink eigenvalues of $\hat{\Sigma}_0$ towards a middle value. Works along this direction can be found in Stein(1977), Efron and Morris(1976), Haff(1980), Sugiura and Fujimoto(1982) and others. By use of an integration by parts, Haff(1979) derived a useful formula, called the *Haff identity* or *Wishart identity*, in the Wishart distribution. For $p \times p$ matrix $V = (v_{ij}(S))$, define $V_{(1/2)} = (v'_{ij})$ where $v'_{ij} = v_{ij}$ for $i = j$ and $= 2^{-1}v_{ij}$ for $i \ne j$, and denote $D = (\partial/\partial s_{ij})_{(1/2)}$. Then the Haff identity is given by

$$\begin{aligned} E[h(S)\mathrm{tr}V\Sigma^{-1}] &= 2E[h(S)\mathrm{tr}(DV)] + 2E\left[\mathrm{tr}\left\{\frac{\partial h(S)}{\partial S}\cdot V_{(1/2)}\right\}\right] \\ &\quad +(n-p-1)E[h(S)\mathrm{tr}S^{-1}V], \end{aligned} \tag{11}$$

for absolutely continuous real-valued function $h(S)$. It is noted that the identity can be also obtained by using the Stein identity (Stein(1977) and Takemura(1991)). This identity is very powerful for the derivation of improved estimators of the covariance matrix.

Since the general linear (GL) group does not satisfy Kiefer's conditions for the minimaxity, the best equivariant estimator $\hat{\Sigma}_0$ is not minimax for $p \ge 2$. Letting G_T^+ be a set of lower triangular matrices, which is a subgroup of GL, James and Stein(1961) indicated that the best G_T^+-equivariant estimator is minimax and is given by

$$\hat{\Sigma}^{JS} = TDT', \quad D = \mathrm{diag}(d_1, \ldots, d_p), \quad d_i = (n+p+1-2i)^{-1},$$

where $T \in G_T^+$ such that $S = TT'$. However $\hat{\Sigma}^{JS}$ depends on a co-ordinate system, and it is desirable to construct orthogonally invariant minimax estimators.

Two approaches to the derivation of orthogonally invariant minimax estimators are known. One is the method of Stein(1977) and Dey and Srinivasan(1985). Let R be an orthogonal matrix and denote $L = \mathrm{diag}(\ell_1, \ldots, \ell_p)$, a diagonal matrix such that $S = RLR'$. Then $\hat{\Sigma}^{JS}$ can be dominated by the orthogonally invariant estimator

$$\hat{\Sigma}^{ST} = R\mathrm{diag}(\ell_1 d_1, \ldots, \ell_p d_p)R'.$$

Dey and Srinivasan(1985) developed an estimator improving on $\hat{\Sigma}^{ST}$ further for $p \geq 3$. Sheena and Takemura(1992) showed the inadmissibility of $\hat{\Sigma}^{ST}$ for $p \geq 2$ by considering a truncation rule. Haff(1991) developed a general theory of VFBE(*Variational Form of Bayes Estimator*), and demonstrated through simulation experiments that VFBE of Σ is better than $\hat{\Sigma}^{ST}$.

The other is an approach of Takemura(1984), which considered the orthogonally invariant estimator

$$\hat{\Sigma}^{TK} = \int_{O(p)} \Gamma T_\Gamma D T_\Gamma' \Gamma' d\mu(\Gamma),$$

dominating $\hat{\Sigma}^{JS}$, where μ designates the uniform distribution on the orthogonal group $O(p)$ and $T_\Gamma T_\Gamma' = \Gamma' S \Gamma$ for $\Gamma \in O(p)$ and $T_\Gamma \in G_T^+$. The explicit expressions of $\hat{\Sigma}^{TK}$ are given for $p \leq 3$, but it is too difficult to give them for $p \geq 4$(Takemura(1984)). The difficulty arises from a computation of an expectation of a ratio of random variables. Perron(1992) obtained explicitly an approximated solution by replacing it for a ratio of expectations, and showed it is an orthogonally invariant minimax estimator.

Other topics have been studied by Yang and Berger(1994) for the Bayes estimation of Σ, by Krishnamoorthy and Gupta(1989), Dey *et al.*(1990) for estimation of the inverse of the covariance matrix Σ^{-1}, by Konno(1995) for an extension to a growth curve model, by Dey(1988), Dey and Gelfand(1989), Jin(1993) and DasGupta(1989) for the simultaneous estimation of the eigenvalues of Σ, by Eaton and Olkin(1987) for the estimation of a Cholesky decomposition and by Loh(1991a,b) for the simultaneous estimation of two covariance matrices Σ_1 and Σ_2.

Related to the covariance matrix, the estimation of the ratio of two covariance matrices has been studied by DasGupta(1989), Konno(1992), Bilodeau and Srivastava(1992). Especially Bilodeau and Srivastava derived the Kullback-Leibler loss for estimation of the ratio, and showed that the quite similar results to the case of the covariance matrix hold in the estimation of the ratio. Related to the estimation of the covariance matrix and the mean vector, Kuriki(1993) considered the problem of estimating a skew-symmetric normal mean matrix with applications to paired comparisons models and derived unbiased estimators of risks of orthogonally invariant estimators, which provided a class of minimax estimators.

Acknowledgements

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. 08780216 and by a grant from the Research Institute for the Japanese Economy, University of Tokyo.

REFERENCES

Alam, K. (1973). A family of admissible minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* , **1**, 517-525.

Alam, K. (1975). Minimax and admissible minimax estimators of the mean of a multivariate normal distribution for unknown covariance matrix. *Ann. Statist.* , **5**, 83-95.

Albert, J.H. (1987). Empirical Bayes estimation in contingency tables. *Comm. Statist.-Theory Methods*, **16**, 2459-2485.

Ansley, C.F., Kohn, R. and Wong, C.(1993). Nonparametric spline regression with prior information. *Biometrika*, **80**, 75-88.

Baranchick, A.J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* , **41**, 642-645.

Baranchick, A. (1973). Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables. *Ann. Statist.*, **1**, 312-321.

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.

Berger, J. (1975). Minimax estimation of location vectors for a wide class of densities. *Ann. Statist.*, **3**, 1318-1328.

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

Berger, J.O. (1980a). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma parameters. *Ann. Statist.*, **8**, 545-571.

Berger, J.O. (1980b). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.*, **8**, 716-761.

Berger, J.O. (1982). Selecting a minimax estimator of a multivariate normal mean. *Ann. Statist.*, **10**, 81-92.

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

Berger, J.O. and Berliner, L.M. (1984). Bayesian input in Stein estimation and a new minimax empirical Bayes estimator. *J. Econometrics*, **25**, 87-108.

Berger, J.O. and Berliner, L.M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.*, **14**, 461-486.

Berger, J., Bock, M. E., Brown, L.D., Casella, G., and Gleser, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.*, **5**, 763-771.

Berger, J.O. and Dey, D.K. (1985). Truncation of shrinkage estimators of normal means in the nonsymmetric case. In *Multivariate Analysis* VI (P.R. Krishnaiah, ed.), 43-56. North-Holland, Amsterdam.

Berger, J.O. and Robert, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: On the frequentist interface. *Ann. Statist.*, **18**, 617-651.

Berger, J.O. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.*, **6**, 783-801.

Berger, J. and Wolpert, R. (1983). Estimating the mean function of a Gaussian process and the Stein effect. *J. Multivariate Anal.*, **13**, 401-424.

Berliner, L.M. (1985). A decision theoretic structure for robust Bayesian analysis with applications to the estimation of a multivariate normal mean. In *Bayesian Statistics* 2 (J.M. Bernardo, et al., eds.), 619-628. North-Holland, Amsterdam.

Berry, J.C. (1994). Improving the James-Stein estimator using the Stein variance estimator. *Statist. and Prob. Letters*, **20**, 241-245.

Bhattacharya, P.K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. *Ann. Math. Statist.*, **37**, 1819-1824.

Bilodeau, M. (1988). On the simultaneous estimation of scale parameters. *Canad. J. Statist.*, **16**, 169-174.

Bilodeau, M. and Kariya, T. (1989). Minimax estimators in the normal MANOVA model. *J. Multivariate Anal.*, **28**, 260-270.

Bilodeau, M. and Srivastava, M.S. (1992). Estimation of the eigenvalues of $\Sigma_1\Sigma_2^{-1}$. *J. Multivariate Anal.*, **41**, 1-13.

Bock, M.E. (1975). Minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.*, **3**, 209-218.

Bock, M.E. (1985). Minimax estimators that shift towards a hypersphere for location vectors of spherically symmetric distributions. *J. Multivariate Anal.*, **17**, 127-147.

Brandwein, A.C. and Strawderman, W.E. (1978). Minimax estimation of location parameters for spherical symmetric unimodal distributions. *Ann. Statist.*, **6**, 377-416.

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

Brandwein, A.C. and Strawderman, W.E. (1991a). Generalizations of James-Stein estimators under spherical symmetry. *Ann. Statist.*, **19**, 1639-1650.

Brandwein, A.C. and Strawderman, W.E. (1991b). Improved estimates of location in the presence of an unknown scale. *J. Multivariate Anal.*, **39**, 305-314.

Bravo, G. and MacGibbon, B. (1988a). Improved shrinkage estimators for the mean vector of a scale mixture of normals with unknown variance. *Canad. J. Statist.*, **16**, 237-245.

Bravo, G. and MacGibbon, B. (1988b). Improved estimation for the parameters of an inverse Gaussian distribution. *Comm. Statist.-Theory Methods*, **17**, 4285-4299.

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

Brown, L.D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.*, **37**, 1087-1136.

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. (1971). Admissible estimators, recurrent diffusions, and insolvable boundary value problems. *Ann. Math. Statist.*, **42**, 855-904.

Brown, L.D. (1973). Estimation with incompletely specified loss functions. *J. Amer. Statist. Assoc.*, **70**, 471-427.

Brown, L.D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.*, **9**, 1289-1300.

Brown, L.D. (1990). An ancillarity paradox which appears in multiple linear regression. *Ann. Statist.*, **18**, 471-538.

Brown, L.D. and Farrell, R.H. (1985). Complete class theorems for estimation of multivariate Poisson means and related problems. *Ann. Statist.*, **13**, 706-726.

Brown, L.D. and Hwang, J.T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics* III (S.S. Gupta, J. Berger, eds.), 205-230. Academic Press, New York.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *Amer. Statist.*, **39**, 83-87.

Casella, G. and Hwang, J.T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.*, **78**, 688-698.

Casella, G. and Hwang, J.T. (1987). Employing vague prior information in the construction of confidence sets. *J. Multivariate Anal.*, **21**, 79-104.

Casella, G., Hwang, J.T.G. and Robert, C.P. (1994). Loss functions for set estimation. In *Statistical Decision Theory and Related Topics* V (S.S. Gupta, J.O. Berger, eds.), 237-251. Springer-Verlag, New York.

Cellier, D., Fourdrinier, D. and Robert, C. (1989). Robust shrinkage estimators of the location parameter for elliptically symmetric distributions. *J. Multivariate Anal.*, **29**, 39-52.

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

Chaturvedi, A., Hoa, T.V. and Shukla, G. (1993). Performance of the Stein-rule estimators when the disturbances are misspecified as spherical. *Economic Studies Quarterly*, **44**, 97-107.

Chen, J. and Hwang, J.T. (1988). Improved set estimators for the coefficients of a linear model when the error distribution is spherically symmetric with unknown variances. *Canad. J. Statist.*, **16**, 293-299.

Chou, J.-P. (1988). An identity for multidimensional continuous exponential families and its applications. *J. Multivariate Anal.*, **24**, 129-142.

Chou, J.-P. (1991). Simultaneous estimation in discrete multivariate exponential families. *Ann. Statist.*, **19**, 314-328.

Chou, J.-P. and Strawderman, W.E.(1990). Minimax estimation of means of multivariate normal mixtures. *J. Multivariate Anal.*, **35**, 141-150.

Clayton, D. and Kaldor, J.(1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671-681.

Clevenson, M.L. and Zidek, J.V. (1975). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc.*, **70**, 698-705.

Copas, J.B. (1983). Regression, prediction, shrinkage. *J. Roy. Statist. Soc.* (Ser. B), **45**, 311-354.

DasGupta, A. (1986). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses. *Ann. Statist.*, **14**, 206-219.

DasGupta, A. (1989). A general theorem on decision theory for nonnegative functionals: with applications. *Ann. Statist.*, **17**, 1360-1374.

Dey, D. (1988). Simultaneous estimation of eigenvalues. *Ann. Inst. Statist. Math.*, **40**, 137-147.

Dey, D. (1990). Estimation of scale parameters in mixture distributions. *Canad. J. Statist.*, **18**, 171-178.

Dey, D. and Berger, J.O. (1983). On truncation of shrinkage estimators in simultaneous estimation of normal means. *J. Amer. Statist. Assoc.*, **78**, 865-869.

Dey, D. and Chung, Y. (1992). Compound Poisson distributions: Properties and estimation. *Comm. Statist.-Theory Methods*, **21**, 3097-3121.

Dey, D. and Gelfand, A.E. (1989). Improved estimation of a patterned covariance matrix. *J. Multivariate Anal.*, **31**, 107-116.

Dey, D., Ghosh, M. and Srinivasan, C. (1987). Simultaneous estimation of parameters under entropy loss. *J. Statist. Plann. Inference*, **15**, 347-363.

Dey, D., Ghosh, M. and Srinivasan, C. (1990). A new class of improved estimators of a multinormal precision matrix. *Statist. Decisions*, **8**, 141-151.

Dey, D.K. and Srinivasan, C. (1985). Estimation of covariance matrix under Stein's loss. *Ann. Statist.*, **13**, 1581-1591.

Eaton, M.L. (1992). A statistical diptych: admissible inferences - recurrence of symmetric Markov chains. *Ann. Statist.*, **20**, 1147-1179.

Eaton, M.L. and Olkin, I. (1987). Best equivariant estimators of a Cholesky decomposition. *Ann. Statist.*, **15**, 1639-1650.

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.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.*, **70**, 311-319.

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

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

Eguchi, S. and Yanagimoto, T. (1984). Second-order structure of relative entropy risk by Laplace-Beltrami operator. Unpublished manuscript.

Evans, S.N. and Stark, P.B. (1996). Shrinkage estimators, Skorokhod's problem and stochastic integration by parts. *Ann. Statist.*, **24**, 809-815.

Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.

George, E.I. (1986a). Minimax multiple shrinkage estimation. *Ann. Statist.*, **14**, 188-205.

George, E.I. (1986b). Combining minimax shrinkage estimators. *J. Amer. Statist. Assoc.*, **81**, 437-445.

George, E.I. (1990). Comment on paper by J.M. Maatta and G. Casella (1990). *Statist. Science*, **5**, 90-120.

George, E.I. (1991). Shrinkage domination in a multivariate common mean problem. *Ann. Statist.*, **19**, 952-960.

George, E.I. and Casella, G. (1994). An empirical Bayes confidence report. *Statistica Sinica*, **4**, 617-638.

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

Ghosh, M., Nickerson, D.M. and Sen, P.K. (1987). Sequential shrinkage estimation. *Ann. Statist.*, **15**, 817-829.

Ghosh, M. and Parsian, A. (1980). Admissible and minimax multiparameter estimation in exponential families. *J. Multivariate Anal.*, **10**, 551-564.

Ghosh, M. and Parsian, A. (1981). Bayes minimax estimation of multiple Poisson parameters. *J. Multivariate Anal.*, **11**, 280-288.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statist. Science*, **9**, 55-93.

Ghosh, M. and Sinha, B.K. (1988). Empirical and hierarchical Bayes competitors of preliminary test estimators in two sample problems. *J. Multivariate Anal.*, **27**, 206-227.

Ghosh, M. and Yang, M.-C. (1988) Simultaneous estimation of Poisson means under entropy loss. *Ann. Statist.*, **16**, 278-291.

Gleser, L.J. (1986). Minimax estimators of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.*, **14**, 1625-1633.

Gupta, A.K. and Saleh, A.K.Md.E. (1996). Estimation of multinomial probabilities under a model constraint. *J. Multivariate Anal.*, **58**, 151-161.

Gupta, A.K., Saleh, A.K.Md.E. and Sen, P.K. (1989). Improved estimation in a contingency tables: Independence structure. *J. Amer. Statist. Assoc.*, **84**, 525-532.

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

Haff, L.R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.*, **8**, 586-597.

Haff, L.R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.*, **19**, 1163-1190.

Haff, L.R. and Johnson, R.W. (1986). The superharmonic condition for simultaneous estimation of means in exponential families. *Canad. J. Statist.*, **14**, 43-54.

Hoffmann, K. (1992). *Improved Estimation of Distribution Parameters: Stein-Type Estimators.* B.G. Teubner Verlagsgesellschaft, Stuttgart.

Honda, T. (1991). Minimax estimators in the MANOVA model for arbitrary quadratic loss and unknown covariance matrix. *J. Multivariate Anal.*, **36**, 113-120.

Hudson, H.M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.*, **6**, 478-484.

Hwang, J.T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. *Ann. Statist.*, **10**, 857-867.

Hwang, J.T. (1985). Universal domination and stochastic domination: Estimation simultaneously under a broad class of loss functions. *Ann. Statist.*, **13**, 295-314.

Hwang, J.T. and Brown, L.D. (1991). Estimated confidence under the validity constraint. *Ann. Statist.*, **19**, 1964-1977.

Hwang, J.T. and Casella, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.*, **10**, 868-881.

Hwang, J.T. and Casella, G. (1984). Improved set estimators for a multivariate normal mean. *Statist. Decisions*, Supplement Issue **1**, 3-16.

Hwang, J.T., Casella, G., Robert, C., Wells, M.T. and Farrell, R.H. (1992). Estimation of accuracy in testing. *Ann. Statist.*, **20**, 490-509.

Hwang, J.T. and Chen, J. (1986). Improved confidence sets for the coefficients of a linear model with spherically symmetric errors. *Ann. Statist.*, **14**, 444-460.

Hwang, J.T. and Ullah, A. (1994). Confidence sets centered at James-Stein estimators. *J. Econometrics*, **60**, 145-156.

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.

Jin, C. (1993). A note on simultaneous estimation of eigenvalues of a multivariate normal covariance matrix. *Statist. Probab. Lett.*, **16**, 197-203.

Johnson, R.W. (1987). Simultaneous estimation of binomial N's. *Sankhya* (Ser. A), **49**, 264-267.

Johnstone, I. (1984). Admissibility, difference equations and recurrence in estimating a Poisson mean. *Ann. Statist.*, **12**, 1173-1198.

Johnstone, I. (1987). On inadmissibility of some unbiased estimates of loss. In *Statistical Decision Theory and Related Topics* IV (S.S. Gupta, J.O. Berger, eds.). Springer-Verlag, New York.

Joshi, V.M. (1967). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.*, **40**, 1042-1067.

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

Jun, C.-H. (1993). Heuristic shrinkage estimators of Poisson means and a Monte Carlo comparison. *J. Statist. Comput. Simul.*, **47**, 181-193.

Kariya, T., Konno, Y. and Strawderman, W.E. (1996). Double shrinkage estimators in the GMANOVA model. *J. Multivariate Anal.*, **56**, 245-258.

Kariya, T., Konno, Y. and Strawderman, W.E. (1994). Construction of improved estimators for the regression coefficient matrix in GMANOVA model. Discussion Paper Series A, Inst. Econ. Res., Hitotsubashi University.

Ki, F. and Tsui, K.-W. (1985). Improved confidence set estimators of a multivariate normal mean and generalizations. *Ann. Inst. Statist. Math.*, **37**, 487-498.

Ki, F. and Tsui, K.-W. (1990). Multiple-shrinkage estimators of means in exponential families. *Canad. J. Statist.*, **18**, 31-46.

Komaki, F. (1996). On asymptotic properties of predictive distributions. *Biometrika*, **83**, 299-313.

Konno, Y. (1991). On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivariate Anal.*, **36**, 44-55.

Konno, Y. (1992). On estimating eigenvalues of the scale matrix of the multivariate F distribution. *Sankhya* (Ser. A), **54**, 241-251.

Konno, Y. (1995). Estimation of a normal covariance matrix with incomplete data under Stein's loss. *J. Multivariate Anal.*, **52**, 308-324.

Koul, H.L. and Saleh, A.K.Md.E. (1993). R-estimation of the parameters of autoregressive $[AR(p)]$ models. *Ann. Statist.*, **21**, 534-551.

Krishnamoorthy, K. (1992). On a shrinkage estimator of a normal common mean vector. *J. Multivariate Anal.*, **40**, 109-114.

Krishnamoorthy, K. and Gupta, A.K. (1989). Improved minimax estimation of a normal precision matrix *Canad. J. Statist.*, **17**, 91-102.

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., Honda, T., Morita, K. and Saleh, A.K.Md.E. (1993). 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. (1993). Estimation of the variance and its applications. *J. Statist. Plann. Inference*, **35**, 319-333.

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., Saleh, A.K.Md.E. and Morita, K. (1992). Improving on MLE of coefficient matrix in a growth curve model. *J. Statist. Plann. Inference*, **31**, 169-177.

Kuo, L. (1986). A note on Bayes empirical Bayes estimation by means of Dirichlet processes. *Statist. Probab. Lett.*, **4**, 145-150.

Kuriki, T. (1993). Orthogonally invariant estimation of the skew-symmetric normal mean matrix. *Ann. Inst. Statist. Math.*, **45**, 731-739.

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

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

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

Li, K.-C. and Hwang, J.T. (1985). The data-smoothing aspect of Stein estimates. *Ann. Statist.*, **12**, 887-897.

Lin, P.-E. and Tsai, H.-L. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *Ann. Statist.*, **1**, 142-145.

Loh, W.-L. (1991a). Estimating covariance matrices. *Ann. Statist.*, **19**, 283-296.

Loh, W.-L. (1991b). Estimating covariance matrices II. *J. Multivariate Anal.*, **36**, 163-174.

Lu, K.L. and Berger, J.O. (1989a). Estimation of normal means: Frequentist estimation of loss. *Ann. Statist.*, **17**, 890-906.

Lu, K.L. and Berger, J.O. (1989b). Estimated confidence procedures for multivariate normal means. *J. Statist. Plann. Inference*, **23**, 1-19.

Lwin, T. and Maritz, J.S. (1989). Empirical Bayes approach to multiparameter estimation: With special reference to multinomial distribution. *Ann. Inst. Statist. Math.*, **41**, 81-99.

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

Majumdar, S. (1994). Bayes compound and empirical Bayes estimation of the mean of a Gaussian distribution on a Hilbert space. *J. Multivariate Anal.*, **48**, 87-106.

Mandelbaum, A. and Shepp, L.A. (1987). Admissibility as a touchstone. *Ann. Statist.*, **15**, 252-268.

Morris, C.N. (1983). Parametric empirical Bayes inference: Theory and application. *J. Amer. Statist. Assoc.*, **78**, 47-65.

Nickerson, D.M. and Basawa, I.V. (1992). Shrinkage estimation for linear models with correlated errors. *Sankhya* (Ser. A), **54**, 411-424.

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

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.*, **85**, 163-171.

Ralescu, S. Brandwein, A.C. and Strawderman, W.E. (1992). Stein estimation for non-normal spherically symmetric location families in three dimensions. *J. Multivariate Anal.*, **42**, 35-50.

Robbins, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Statist.*, **11**, 713-723.

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

Robert, C. and Casella, G. (1990). Improved confidence sets for spherically symmetric distributions. *J. Multivariate Anal.*, **32**, 84-94.

Robert, C.P. and Casella, G. (1994). Improved confidence statements for the usual multivariate normal confidence set. In *Statistical Decision Theory and Related Topics* V (S.S. Gupta, J. Berger, eds.), 351-368. Springer-Verlag, New York.

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

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

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.

Saleh, A.K.Md.E. and Sen, P.K. (1985). Nonparametric shrinkage estimation in a parallelism problem. *Sankhya* (Ser. A), **47**, 156-165.

Sarkar, S.K. (1994). Shrinkage domination of some usual estimators of the common mean of several multivariate normal populations. *J. Statist. Plann. Inference*, **39**, 43-55.

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

Sen, P.K. and Saleh, A.K.Md.E. (1985). On some shrinkage estimators of multivariate location. *Ann. Statist.*, **13**, 272-281.

Sen, P.K. and Saleh, A.K.Md.E.(1987). On preliminary test and shrinkage M-estimation in linear models. *Ann. Statist.*, **15**, 1580-1592.

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

Shao, P.Y.-S. and Strawderman. (1994). Improving on the James-Stein positive-part estimator. *Ann. Statist.*, **22**, 1517-1538.

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

Shieh, G. (1993). Empirical Bayes minimax estimators of matrix normal means for arbitrary quadratic loss and unknown covariance matrix. *Statist. Decisions*, **11**, 317-341.

Shinozaki, N. (1980). Estimation of a multivariate normal mean with a class of quadratic loss functions. *J. Amer. Statist. Assoc.*, **75**, 973-976.

Shinozaki, N. (1984). Simultaneous estimation of location parameters under quadratic loss. *Ann. Statist.*, **12**, 322-335.

Shinozaki, N. (1989). Improved confidence sets for the mean of a multivariate normal distribution. *Ann. Inst. Statist. Math.*, **41**, 331-346.

Shinozaki, N. and Chang, Y.-T. (1994). Minimaxity of empirical Bayes estimators of the means of independent normal variables with unequal variances. *Comm. Statist. - Theory Methods*, **22**, 2147-2169.

Shiraishi, T. (1991). On positive-part shrinkage R- and M-estimation in one-way ANOVA. *J. Japan Statist. Soc.*, **21**, 61-72.

Spruill, C. (1982). Admissibility of the natural estimator of the mean of a Gaussian process. *J. Multivariate Anal.*, **12**, 568-574.

Srinivasan, C. (1981). Admissible generalized Bayes estimators and exterior boundary value problems. *Sankhya* (Ser. A), **43**, 1-25.

Srivastava, M.S. and Bilodeau, M. (1989). Stein estimation under elliptical distributions. *J. Multivariate Anal.*, **28**, 247-259.

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. (1960). Multiple regression. In *Contributions to Probability and Statistics Essays in Honor of Harold Hotelling* (I. Olkin, et al., eds.), 424-443. Stanford University Press.

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

Stein, C. (1977). Lectures on multivariate estimation theory. (In Russian.) In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad.(In Russian)

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

Stigler, S.M. (1990). A Galtonian perspective on shrinkage estimators. *Statist. Science*, **5**, 147-155.

Strawderman. W.E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.*, **42**, 385-388.

Strawderman. W.E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *J. Multivariate Anal.*, **4**, 255-264.

Sugiura, N. and Fujimoto, M. (1982). Asymptotic risk comparison of improved estimators for normal covariance matrix. *Tsukuba J. Math.*, **6**, 103-126.

Sugiura, N. and Takagi, Y. (1996). Dominating James-Stein positive-part estimator for normal mean with unknown covariance matrix. *Commun. Statist.- Theory Methods*, **25**, 2875-2900.

Takada, Y. (1979). A family of minimax estimators in some multiple regression problems. *Ann. Statist.*, **7**, 1144-1147.

Takada, Y. (1984). Inadmissibility of a sequential estimation rule of the mean of a multivariate normal distribution. *Sequential Anal.*, **3**, 267-271.

Takemura, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba J. Math.*, **8**, 367-376.

Takemura, A. (1991). *The Foundation of the Multivariate Statistical Inference.* Kyoritsu Press, Tokyo (in Japanese).

Takeuchi, K. (1991). Personal communications.

Tan, M. (1991). Improved estimators for the GMANOVA problem with application to Monte Carlo simulation. *J. Multivariate Anal.*, **38**, 262-274.

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.

Tsui, K.-W. (1984). Robustness of Clevenson-Zidek-type estimators. *J. Amer. Statist. Assoc.*, **79**, 152-157.

Tsui, K.-W. and Press, S.J. (1982). Simultaneous estimation of several Poisson parameters under K-normalized squared error loss. *Ann. Statist.*, **10**, 932-100.

Tsutakawa, R.K., Shoop, G.L. and Marienfeld, C.J. (1985). Empirical Bayes estimation of cancer mortality rates. *Statist. Medicine*, **4**, 201-212.

Wahba, G. (1985). Comparison of GCV and GML for choosing the smoothing parameter in the generalized smoothing problem. *Ann. Statist.* , **13**, 1378-1402.

Wald, A. (1950). *Statistical Decision Functions.* Wiley, New York.

Yanagimoto, T. (1994). The Kullback-Leibler risk of the Stein estimator and the conditional MLE. *Ann. Inst. Statist. Math.*, **46**, 29-41.

Yang, R. and Berger, J.O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* , **22**, 1195-1211.

Zacks, S. (1971). *The Theory of Statistical Inference*, Wiley, New York.

Zheng, Z. (1986). On estimation of matrix of normal mean. *J. Multivariate Anal.*, **18**, 70-82.

Zidek, J. (1978). Deriving unbiased risk estimators of multinormal mean and regression coefficient estimators using zonal polynomials. *Ann. Statist.*, **6**, 769-782.