Minimax Admissible Estimation of a Multivariate Normal Mean and Improvement upon the James-Stein Estimator

By
Yuzo Maruyama

Faculty of Mathematics,
Kyushu University

A dissertation submitted to the
Graduate School of Economics
of University of Tokyo
in partial fulfillment of the requirements
for the degree of Doctor of Economics

June, 2000
(Revised December, 2000)

Abstract

This thesis mainly focuses on the estimation of a multivariate normal mean, from the decision-theoretic point of view. We concentrate our energies on investigating a class of estimators satisfying two optimalities: minimaxity and admissibility. We show that all classes of minimax admissible generalized Bayes estimators in the hitherto researches are divided in two, on the basis of whether a certain function of the prior distribution is bounded or not. Furthermore by using Stein's idea, we propose a new class of minimax admissible estimators. The relations among minimaxity, a prior distribution and a shrinkage factor of the genaralized Bayes estimator, are also investigated.

For the problem of finding admissible estimators which dominate the inadmissible James-Stein estimator, we have some results.

Moreover, in non-normal case, we consider the problem of proposing admissible minimax estimators. For a certain class of spherically symmetric distributions which includes at least multivariate-t distributions, admissible minimax estimators are derived.

Acknowledgements

I am indebted to my adviser, Professor Tatsuya Kubokawa for his guidance and encouragement during the research for and preparation of this thesis.

I wish to take opportunity to express my deep gratitude to Professors of Graduate School of Economics, University of Tokyo: Akimichi Takemura, Naoto Kunitomo, Yoshihiro Yajima, Kazumitsu Nawata and Nozomu Matsubara, and to Professors of Faculty of Mathematics, Kyushu University: Takashi Yanagawa, Sadanori Konishi, Hiroto Hyakutake, Yoshihiko Maesono, Masayuki Uchida and Kaoru Fueda, for their encouragement and suggestions.

Finally, special thanks to my wife, Hiromi, for her support, encouragement, tenderness and patience.

Contents

Chapter 1

Introduction

Statistical decision theory, which was made by Wald(1950), characterizes the mode of modern statistical thinking. The contrast between the simplicity of its basic notions and the broad area of applicability is astonishing. The minimax principle and least favorable distributions are central notions of statistical decision theory and are close parallel to the corresponding concepts of the game theory. Admissibility is another basic concept of statistical decision theory and is related to the concept of Pareto optimality in economics. During the last thirty years, the study of statistical decision theory, in particular of admissibility, led to unexpected, amazing connections to many branches of mathematics, for example, partial differential equations of mathematical physics, differential inequalities and recurrence properties of stochastic process. For the good accounts, see Brown(1971), Eaton(1992) and Rukhin(1995). It is noted that these connections were discovered and have developed in the process of investigating the estimation problem of a multivariate normal mean, a typical topic of what is called the *Stein phenomenon*, which is introduced as follows.

Many statisticians would like to believe that a natural estimator as maximum likelihood estimator(MLE) is admissible because admissibility is a weak criterion for optimality. For the estimation problem of a mean vector $\theta = (\theta_1, \dots, \theta_p)'$ based on random vector $X = (X_1, \dots, X_p)'$ having p-variate normal distribution $N(\theta, I_p)$ relative to a quadratic loss function, Stein(1956) showed that the MLE X, which is also minimax, is inadmissible for $p \geq 3$ while it is admissible for $p = 1, 2$. James and Stein(1961) succeeded in giving an explicit form of an estimator improving on X as

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

which is called the *James-Stein estimator*. This result surprises us because 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*.

Generally if δ dominates a minimax estimator, then δ is also minimax and, thus, definitely preferred. It is, therefore, important from decision-theoretic point of view, to ascertain whether a proposed minimax estimator δ is admissible. Moreover if the estimator δ in the above is also inadmissible, it is interesting to derive admissible estimators improving upon δ. In this thesis, we mainly consider the estimation problem in Stein(1956)'s setting and concentrate our energies on investigating a class of estimators satisfying two optimalities: minimaxity and admissibility. We also consider the problem of deriving admissible estimators improving upon the James-Stein estimator which is known to be inadmissible. The outline of this thesis is as follows.

In Chapter 2, we review basic notions of statistical decision theory, particular in estimation problem. The concepts of "admissibility", "minimaxity", "Bayes" and "permissibility" are introduced and results concerning close connections among these concepts are presented with proof.

Chapter 3 contains a variety of results concerning an estimation of a multivariate normal mean. In Section 3.1, the minimaxity and the inadmissibility for $p \geq 3$ of the natural estimator X are proved. In Section 3.2, results of Brown(1971,1988) and Srinivasan(1981) concerning admissibility and permissibility in this estimation problem are presented. In Section 3.3, we consider a generalized Bayes estimator δ_h with respect to scale mixtures of multivariate normal distribution whose density is proportional to

$$\int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left(-\frac{\lambda}{2(1-\lambda)}\|\theta\|^2\right) \lambda^{-a} h(\lambda) d\lambda, \tag{1.1}$$

where $h(\lambda)$ is a measurable positive function on $(0, 1)$. By Brown(1971)'s result, it is shown that δ_h is admissible if and only if $a \leq 2$. Next we consider the minimaxity of δ_h. As seen easily, all classes of minimax generalized Bayes estimators given in the hitherto researches, that is, Strawderman(1971), Alam(1973), Berger(1976b), Faith(1978), Maruyama(1998) and Fourdrinier *et al.*(1998) are expressed as δ_h although relations among classes above have not been clear. Hence we succeed in treating uniformly classes above. We point out that Fourdrinier *et al.*(1998)'s result is based on the boundedness of $h(\lambda)$ although their result is powerful and includes results of Strawderman(1971), Berger(1976b) and Faith(1978). Indeed, the generalized Bayes estimators proposed by Alam(1973) and Maruyama(1998), are expressed as δ_h in the case of unbounded $h(\lambda)$. Thus we can see that the hitherto researches are divided in two, as Table 3.1 in Section 3.3.3, on the basis of whether $h(\lambda)$ is bounded or not.

Moreover we pay attention to the following Stein(1973)'s simple idea and provide it with the warrant for statistical decision theory. Stein(1973) suggested that the generalized Bayes estimator with respect to the prior distribution: the weighted sum of that determined by the density $\|\theta\|^{2-p}$, which corresponds to (1.1) with $a = 2$ and $h \equiv 1$, and a measure concentrated at the origin, may dominate the James-Stein estimator. Since Kubokawa(1991) afterward showed that the generalized Bayes estimator with respect to the density $\|\theta\|^{2-p}$ dominates the James-Stein estimator, Stein's suggestion is to the point in the case where the ratio of a measure concentrated at the origin is very small. Efron and Morris(1976) expressed Stein(1973)'s estimator explicitly. As for the decision-theoretic properties of the estimator, admissibility is easily able to be checked by using Brown(1971)'s theorem. On the other hand, even minimaxity, to say nothing of dominance over the James-Stein estimator has not been proved yet. Extending Stein's prior distribution, we consider a prior distribution: the weighted sum of a measure concentrated at the origin and the prior distribution whose density is proportional to (1.1). This prior distribution corresponds to one which has a density function (1.1) with $h(\lambda) = \beta h(\lambda) + \delta(\lambda - 1)$, where $\beta > 0$ and $\delta(\cdot)$ is the Dirac delta function. We show that for $h(\lambda)$ such that δ_h is shown to be minimax, there exists a constant β^* such that for $\beta \geq \beta^*$, the generalized Bayes estimator with respect to the above distribution is minimax.

Furthermore the relations among minimaxity, the prior distribution, and the behavior of $\phi(w)$, a shrinkage factor of the generalized Bayes estimator, are investigated. Especially, in the case where $\phi(w)$ is not monotone, the researches on the relations are fragmentary and have not been arranged yet. We characterize the subclass of the prior distributions which lead minimax estimators with not-monotone $\phi(w)$. We believe that the investigation of estimators with not-monotone $\phi(w)$ is more and more important because such estimators are considered as candidates for the solution of the most difficult problem in this field, that is, the problem of finding admissible estimators dominating the James-Stein positive-part rule.

In Section 3.4, the problem of finding admissible estimators which dominate the James-Stein estimator is treated. We show that an estimator which is proposed in Maruyama(1996) and improves upon the James-Stein estimator, is permissible. We unfortunately conjecture that it is inadmissible, which implies that it is quite difficult to find a class of admissible estimators which improve upon the James-Stein estimator.

The Stein phenomenon is not limited in the case of a multivariate normal distribution. Since Stein(1956), considerable effort has been given to improving upon the best equivariant estimator δ_0, which is minimax under some mild conditions. The theoretical questions were answered quite thoroughly by Brown(1966). He showed that in 3 or more dimensions δ_0 is inadmissible for an extremely wide variety of distributions and loss functions. It is however very difficult to give explicitly improved minimax estimators without restriction of distributions. For the explicit construction of improved minimax estimators in non-normal case, the estimation of a mean vector of spherically symmetric distributions under the quadratic loss function, has been an important topic in this field. As a special case of spherically symmetric distributions, Strawderman(1974) introduced scale mixtures of multivariate normal distributions, which have a probability density function

$$f(\|x-\theta\|^2) = \int_0^\infty (2\pi)^{-p/2} v^{p/2} \exp\left(-\frac{\|x-\theta\|^2 v}{2}\right) G(dv),$$

where G is a known distribution function. Strawderman(1974) proposed a class of improved minimax estimators of the mean vector. Generally Berger(1975), Bock(1985) and Brandwein and Strawderman(1978,1991) found and discussed classes of improved minimax estimators for spherically symmetric distributions. It is however noted that, in non-normal case, an estimator satisfying both minimaxity and admissibility, has not been derived yet. In Chapter 4, the problem of estimating the mean vector θ of scale mixtures of multivariate normal distributions, in the case where G has a probability density function, with the quadratic loss function is considered. For a certain subclass of these distributions, which includes at least multivariate-t distributions, admissible minimax estimators are derived for $p \geq 5$.

In Chapter 5 and 6, we review my published papers Maruyama(1997,1999a,1999b).

Chapter 2

Basic notions of statistical decision theory

In this chapter, referring to Takemura(1991) and Lehmann and Casella(1999), we review statistical decision theory, particular in estimation problem. Let X be a random variable taking on values in a sample space $\mathcal{X}$ according to a distribution P_θ, which is known to belong to a family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. The aim of an estimation problem is to determine an *estimator*, a plausible real-valued function δ defined over the sample space, for an *estimand*, $g(\theta)$, a real-valued function of the parameter θ. Needless to say, we hope that $\delta(X)$ will tend to be close to the unknown $g(\theta)$. Since $\delta(X)$ is a random variable, we shall interpret this to mean that it will be close on the average. To make this requirement precise, it is necessary to specify a measure of the average closeness of an estimator to $g(\theta)$. One example of such measures is mean square error $E(\delta(X) - g(\theta))^2$. Quite generally, we introduce a *loss function* $L(\theta, d)$, a nonnegative function of θ and d. It is usually assumed that $L(\theta, d) \geq 0$ for all θ, d and $L(\theta, g(\theta)) = 0$ for all θ, so that the loss is zero when the correct value is estimated. The accuracy of an estimator δ is then measured by the *risk function*

$$R(\theta, \delta) = E_\theta\{L(\theta, \delta(X))\}. \tag{2.1}$$

If estimators are given, we are able to compare the estimators using the risk functions calculated as (2.1).

Definition 2.1. For two estimators δ_1 and δ_2, δ_1 *dominates* δ_2 if

$$R(\theta, \delta_1) \le R(\theta, \delta_2)$$

for all θ and with strict inequality for some θ. An estimator δ is said to be *inadmissible* if there exists another estimator δ' which dominates it and *admissible* if no such estimator δ' exists.

Since admissibility is a desirable property, it is of interest to determine the totality of admissible estimators.

Definition 2.2. A class of $\mathcal{C}$ of estimators is *complete* if for any δ not in $\mathcal{C}$ there exists an estimator δ' in $\mathcal{C}$ such that δ' dominates δ.

It follows from this definition that any estimator outside a complete class is inadmissible. It is therefore reasonable, in the search for an optimal estimator, to restrict attention to a complete class. We note however that admissibility is a weak criterion of optimality and many admissible estimators exist. For example, an unreasonable constant estimator $\delta_0 = \theta_0$, which does not utilize the information of X, is typically admissible. Therefore, only on the basis of admissibility, we cannot solve the problem "Which estimator should be chosen?" As one more criterion for comparing estimators, minimax criterion, which aims at minimizing the risk in the least favorable case, is often utilized. Needless to say, a constant estimator is not minimax.

Definition 2.3. An estimator δ^M, which minimizes the maximum risk, that is, which satisfies

$$\inf_{\delta} \sup_{\theta} R(\theta, \delta) = \sup_{\theta} R(\theta, \delta^M)$$

is called a *minimax* estimator.

There exist close connections between admissibility and minimaxity as follows.

Theorem 2.4. *If an estimator is unique minimax, it is admissible.*

Proof. If it were inadmissible, another estimator would dominate it in risk and, hence, would be minimax. □

For an estimator with constant risk, we have the following.

Theorem 2.5. *If an estimator has constant risk and is admissible, it is minimax.*

Proof. If it were not, another estimator would have smaller maximum risk and, hence, uniformly smaller risk. □

If there is a unique minimax estimator δ, by Theorem 2.4, δ is admissible and we should use it as the best estimator. If the uniqueness is removed from the above assumption, there can exist other minimax estimators which improve upon δ. In particular, if the constant risk minimax estimator is inadmissible, this is the worst minimax estimator in the sense that every other minimax estimator has a uniformly small risk. For example, suppose that δ is minimax and has constant risk r. For any other minimax estimator δ', which implies that $\sup_\theta R(\theta, \delta')$ equals r, the case where $R(\theta, \delta') < r$ for any θ and $\lim_{\|\theta\|\to\infty} R(\theta, \delta') = r$ may happen. Indeed, on the most important theme of this thesis, the Stein phenomenon, this happens. In this case, the one and only best estimator on the basis of minimaxity and admissibility, is not able to be determined since there can exist many admissible minimax estimators. It is however noted that a problem of characterizing a class of estimators satisfying both admissibility and minimaxity appears.

To prove admissibility or minimaxity of a certain estimator δ, we often have to express it as *Bayes estimator* or variant of Bayes estimator which are defined in the following.

Definition 2.6. Let Θ have distribution π, that is, $\int \pi(d\theta) = 1$. An estimator δ minimizing a *Bayes risk*

$$r(\pi, \delta) = \int R(\theta, \delta)\pi(d\theta)$$

is called a *Bayes estimator* with respect to π.

This definition immediately leads the following principal method for proving admissibility.

Theorem 2.7. *Any unique Bayes estimator is admissible.*

Proof. If δ is unique Bayes with respect to the distribution π and is dominated by δ', then

$$\int R(\theta, \delta')\pi(d\theta) \leq \int R(\theta, \delta)\pi(d\theta),$$

which contradicts uniqueness. □

The following result replaces the uniqueness assumption by the continuity of all risk functions.

Theorem 2.8. *For a possibly vector-valued parameter θ, suppose that δ^ν is a Bayes estimator having finite Bayes risk with respect to a distribution which has a density function ν, that ν is positive for all θ, and that the risk function of every estimator δ is continuous function of θ. Then δ^ν is admissible.*

Proof. If δ^ν is not admissible, there exists an estimator δ such that $R(\theta, \delta) \leq R(\theta, \delta^\nu)$ for all θ with strict inequality for some θ. It then follows from the continuity of risk functions that $R(\theta, \delta) < R(\theta, \delta^\nu)$ for all θ in some open subset Θ_0 of the parameter space and hence that

$$\int R(\theta, \delta)\nu(\theta)d\theta < \int R(\theta, \delta^\nu)\nu(\theta)d\theta,$$

which contradicts the definition of δ^ν. □

In the Bayesian analysis, π and ν in the above are called a *prior distribution* and a *prior density function*, respectively. Namely θ is viewed as a realization of a random vector $\boldsymbol{\theta}$ whose prior distribution is π. A sample X is drawn from $P_\theta = P_{x|\theta}$, which is viewed as the conditional distribution of X given $\boldsymbol{\theta} = \theta$. The sample $X = x$ is then used to obtain an updated prior distribution, which is called *posterior distribution.* Noting that the joint distribution of X and $\boldsymbol{\theta}$ is a probability measure on $\mathcal{X} \times \Theta$ determined by

$$P(A \times B) = \int_B P_{x|\theta}(A)\pi(d\theta), \qquad A \in \mathcal{B}_{\mathcal{X}}, \quad B \in \mathcal{B}_{\Theta},$$

where $\mathcal{B}_{\mathcal{X}}$ and $\mathcal{B}_{\Theta}$ are the Borel σ-field on $\mathcal{X}$, Θ respectively, we see that the posterior distribution of $\boldsymbol{\theta}$, given by $X = x$, is expressed as the conditional distribution $P_{\theta|x}$. If $P_{x|\theta}$ and π have probability density functions $f(x|\theta)$ and $\nu(\theta)$ respectively, the posterior distribution $P_{\theta|x}$ has a posterior density function

$$\nu(\theta|x) = f(x|\theta)\nu(\theta)\left(\int f(x|\theta)\nu(\theta)d\theta\right)^{-1}.$$

By using the concept of *posterior*, the determination of a Bayes estimator is, in principle, quite simple as follows.

Theorem 2.9. *Suppose the following assumptions hold*

1. *There exists an estimator δ_0 with finite risk,*
2. *for almost all x, there exists a value $\delta^\pi(x)$ minimizing*

$$E\left\{L(\boldsymbol{\theta}, \delta(x))|X = x\right\}, \tag{2.2}$$

where the expectation is taken with respect to the posterior distribution $P_{\theta|x}$.

Then δ^π is a Bayes estimator.

The expectation (2.2) is called *posterior expected loss.* Therefore a Bayes estimator is interpreted as one which minimizes the posterior expected loss.

Proof. Let δ be any estimator with finite risk. Then (2.2) is finite since L is nonnegative. Hence,

$$E\left\{L(\boldsymbol{\theta}, \delta(x))|X = x\right\} \geq E\left\{L(\boldsymbol{\theta}, \delta^\pi(x))|X = x\right\} \qquad \text{a.}e.,$$

and the result follows by taking the expectation of both sides. □

Corollary 2.9.1. *Suppose the assumptions of Theorem 2.9 hold. If* $L(\theta, d) = (d - g(\theta))^2$*, then* $\delta^\pi(x) = E(g(\boldsymbol{\theta})|x)$*, which is the mean of the posterior distribution.*

Proof. By Theorem 2.9, the Bayes estimator is obtained by minimizing

$$E\left\{(g(\boldsymbol{\theta}) - \delta(x))^2|X = x\right\}. \tag{2.3}$$

Since (2.3) is expanded as

$$(\delta(x) - E(g(\boldsymbol{\theta})|x))^2 - (E(g(\boldsymbol{\theta})|x))^2 + E(g(\boldsymbol{\theta})^2|x),$$

it is minimized by $\delta(x) = E(g(\boldsymbol{\theta})|x)$. □

Next we extend the above definition of the Bayes estimator to the case where π is a measure satisfying $\int \pi(d\theta) = \infty$, a so-called *improper prior distribution.* In the case where (2.2) is finite for each x, the Bayes estimator can formally be defined.

Definition 2.10. An estimator δ^π is a *generalized Bayes estimator* with respect to a measure $\pi(\theta)$(even if it is not a proper probability distribution) if the posterior expected loss, $E\{L(\boldsymbol{\theta}, \delta(x))|X = x\}$, is minimized at $\delta = \delta^\pi$ for all x.

Therefore a class of generalized Bayes estimators coincides with direct sum of a class of proper Bayes estimators and a class of improper Bayes estimators. There is another useful variant of a Bayes estimator, a *limit of Bayes estimators.*

Definition 2.11. An estimator δ is a *limit of Bayes estimators* if there exists a sequence of proper prior distributions π_n and Bayes estimators δ^{π_n} such that $\delta^{\pi_n} \to \delta$ a.e. as $n \to \infty$.

By using the property of a limit of Bayes estimators, a sufficient condition for minimaxity of a certain estimator δ is given as follows.

Lemma 2.12. *Suppose that* $\{\pi_n\}$ *is a sequence of proper prior distributions and* δ_n *is the Bayes estimator with respect to* π_n*. If*

$$\sup_\theta R(\theta, \delta) = \lim_{n\to\infty} r(\pi_n, \delta_n),$$

then δ *is minimax.*

Proof. Suppose that δ' is any other estimator. Then we have

$$\sup_\theta R(\theta, \delta') \geq \int R(\theta, \delta') d\pi_n(\theta) \geq r(\pi_n, \delta_n),$$

and this holds for every n. Hence,

$$\sup_\theta R(\theta, \delta') \geq \sup_\theta R(\theta, \delta)$$

and δ is minimax. □

In the case where at least one minimax estimator is derived, a following idea of *unbiased estimator of risk*, often enables us to show that the other estimators are minimax.

Definition 2.13. For a given estimator $\delta(X)$ and its risk $R(\theta, \delta)$, if there exists a statistic $\hat{R}(\delta(X))$, which is not depend on θ and satisfies

$$R(\theta, \delta) = E_\theta(\hat{R}(\delta(X))),$$

$\hat{R}(\delta(X))$ is called the *unbiased estimator of risk.*

Lemma 2.14. *Suppose that $\delta'(X)$ with the unbiased estimator of risk, $\hat{R}(\delta'(X))$, is minimax and that $\delta(X)$ with the unbiased estimator of risk, $\hat{R}(\delta(X))$, is any other estimator.*

1. *If $\hat{R}(\delta(x)) \leq \hat{R}(\delta'(x))$ for any x, $\delta(X)$ is minimax.*

2. *If $\hat{R}(\delta(x)) < \hat{R}(\delta'(x))$ for any x, $\delta'(X)$ is inadmissible.*

Proof. To prove part *1*, by taking the expectation of both sides of $\hat{R}(\delta(x)) \leq \hat{R}(\delta'(x))$, we have $R(\theta, \delta) \leq R(\theta, \delta')$ for any θ. Since δ' is minimax, δ is also minimax. Part *2* follows immediately from the definition of inadmissibility. □

The concept of unbiased estimator of risk leads to a criterion of optimality which is weaker than admissibility.

Definition 2.15. A estimator $\delta(X)$ with the unbiased estimator of risk, $\hat{R}(\delta(X))$, is said to be *permissible* if the inequality $\hat{R}(\delta(x)) \geq \hat{R}(\delta'(x))$ is valid for some δ' for all values of x, implies the equality for all x.

If the estimator δ is not permissible, then δ is inadmissible in the original problem. A permissible estimator is not always admissible. The notion of permissibility is not dealt with in standard texts of mathematical statistics. See Brown(1988) and Rukhin(1995) for details and more discussion.

Chapter 3

Estimation of a multivariate normal mean

3.1 Inadmissibility of the usual estimator

Let X be a random variable having p-variate normal distribution $N_p(\theta, I_p)$ and $p_\theta(x)$ denote its density function. Then we consider the problem of estimating the mean vector θ by $\delta(X)$ relative to the quadratic loss function $\|\delta(X) - \theta\|^2$. Therefore every estimator is evaluated on the risk function

$$R(\theta, \delta) = E_\theta \left[\|\delta(X) - \theta\|^2\right] = \int_{R^p} \|\delta(x) - \theta\|^2 p_\theta(x) dx.$$

The usual estimator of θ is X, which is a uniformly minimum variance unbiased and maximal likelihood estimator. This estimation problem is invariant under the transformation $\Gamma X + d$, $\Gamma\theta + d$ for any orthogonal matrix Γ and any 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. Hence we easily see that X is the best among this class of equivariant estimators. Moreover the risk of X is clearly constant, that is, $R(\theta, X) = p$. Following results concerning the minimaxity and admissibility of X are well-known and fundamental.

Theorem 3.1. *1. The estimator X is minimax.*

2. For $p \leq 2$, X is admissible.

3. For $p \geq 3$, X is inadmissible.

In this section, we prove part *1* and *3*. We will prove part *2* in the next section. Part *1* and *3* of this theorem imply that for $p \geq 3$, if δ dominates X, then δ is also minimax and, thus, definitely preferred.

proof of part1. For the prior distribution of θ, we select p-variate normal distribution $N_p(0, mI_p)$. The joint density function of θ and X is then given by

$$\begin{aligned} f(x,\theta) &\propto \exp\left(-\frac{\|x-\theta\|^2}{2}\right)\exp\left(-\frac{\|\theta\|^2}{2m}\right) \\ &= \exp\left(-\frac{m+1}{m}\frac{\|\theta - mx/(m+1)\|^2}{2}\right)\exp\left(-\frac{\|x\|^2}{2(m+1)}\right). \end{aligned}$$

Since the posterior density is the joint density divided by the marginal density of X, we see that posterior distribution of $\theta|X = x$ is $N_p(mx/(m+1), m(m+1)^{-1}I_p)$. Under the quadratic loss function, the Bayes estimator is $\delta_m(X) = mX/(m+1)$, by Corollary 2.9.1. The risk of $\delta_m(X)$ is

$$\begin{aligned} R(\theta, \delta_m) &= E\left[\|\delta_m(X) - \theta\|^2\right] \\ &= E\left[\|m(m+1)^{-1}(X-\theta) - \theta/(m+1)\|^2\right] \\ &= (m/(m+1))^2 p + (m+1)^{-2}\|\theta\|^2, \end{aligned}$$

so that the Bayes risk is

$$r(\pi_m, \delta_m) = (m/(m+1))^2 p + (m+1)^{-2} pm,$$

which approaches the value p or the risk of X, when m tends to infinity. Lemma 2.12 guarantees the minimaxity of X. □

Next by proposing a class of estimators dominating X, we prove part *3*. The inadmissibility of X for $p \geq 3$ was first presented by Stein(1956) and surprised many statisticians because 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.* Although Stein(1956)'s proof of inadmissibility was non-constructive, James and Stein(1961) succeeded in giving an explicit form of an estimator improving on X as

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

which is called the *James-Stein estimator*. Their method of proof of the Stein phenomenon utilizes the fact that a non-central chi square distribution is represented by a Poisson mixture of a central chi square distribution. In this thesis, a more simple and powerful method derived by Stein(1973) is introduced. The following relation which is called the *Stein identity* is essential.

Lemma 3.2 (Stein(1973)). *If $Y \sim N(\mu, 1)$ and g is any differentiable function such that $E[|g'(Y)|] < \infty$, then*

$$E[g(Y)(Y-\mu)] = E[g'(Y)]. \tag{3.1}$$

proof. We write $p_0(y)$ for the standard normal density, with derivative $p_0'(y) = -yp_0(y)$. The right-hand side of the equation (3.1) is expressed as

$$E[g'(Y)] = \int_{-\infty}^{0} g'(y)p_0(y-\mu)dy + \int_{0}^{\infty} g'(y)p_0(y-\mu)dy. \tag{3.2}$$

By using the relation $p_0(y-\mu) = \int_y^\infty (z-\mu)p_0(z-\mu)dz$, the second term in the right-hand side of the equation (3.2) is written as

$$\begin{aligned}\int_0^\infty g'(y)p_0(y-\mu)dy &= \int_0^\infty g'(y)\left(\int_y^\infty (z-\mu)p_0(z-\mu)dz\right)dy \\ &= \int_0^\infty (z-\mu)p_0(z-\mu)\left(\int_0^\infty g'(y)I_{[y\le z]}dy\right)dz \\ &= \int_0^\infty (z-\mu)p_0(z-\mu)(g(z)-g(0))dz, \qquad (3.3)\end{aligned}$$

where $I_{[y \le z]}$ is a indicator function, that is, 1 if $y \le z$, 0 otherwise. Fubini's Theorem ensures the second equality in (3.3). Similarly we have

$$\int_{-\infty}^{0} g'(y)p_0(y-\mu)dy = \int_{-\infty}^{0} (z-\mu)p_0(z-\mu)(g(z)-g(0))dz. \tag{3.4}$$

Combining (3.3) and (3.4), we have

$$\begin{aligned} E[g'(Y)] &= \int_{-\infty}^{\infty} (z-\mu)p_0(z-\mu)(g(z)-g(0))dz \\ &= E[(Y-\mu)g(Y)]. \end{aligned}$$

□

Using this identity (3.1), we can write the risk function of the estimator of the form $\delta(X) = X + g(X)$ as

$$\begin{aligned} R(\theta,\delta) &= E_\theta\left[\|\delta(X)-\theta\|^2\right] \\ &= E_\theta\left[\|X-\theta\|^2\right] + E_\theta\left[\|g(X)\|^2\right] + 2\sum_{i=1}^{p} E_\theta\left[(X-\theta)'g(X)\right] \\ &= p + E_\theta\left[\|g(X)\|^2\right] + 2\sum_{i=1}^{p} E_\theta\left[\frac{\partial}{\partial x_i} g_i(X)\right], \end{aligned}$$

where we assume that $g(x) = \{g_i(x)\}$ is differentiable and $E|(\partial/\partial x_i)g_i(X)| < \infty$ for $i = 1, \ldots, p$. In this case, the unbiased estimator of risk is

$$p + \|g(X)\|^2 + 2\sum_{i=1}^{p} \frac{\partial}{\partial x_i} g_i(X) \tag{3.5}$$

and by Theorem 2.14 we have a sufficient condition for minimaxity.

Theorem 3.3 (Stein(1973)). *For the estimator of the form $X + g(X)$, assume that $g(x)$ is given by solutions of the following partial differential inequality*

$$\|g(x)\|^2 + 2\sum_{i=1}^{p} \frac{\partial}{\partial x_i} g_i(x) \le 0. \tag{3.6}$$

Then $X + g(X)$ is minimax.

To make the structure of a minimax estimator comprehensible, we consider an estimator of the form of

$$\delta_\phi(X) = (1 - \phi(\|X\|^2)/\|X\|^2)X, \tag{3.7}$$

which is equivariant with respect to the orthogonal transformation ΓX and $\Gamma\theta$ for orthogonal matrix Γ. Assigning $g(x) = -\phi(\|x\|^2)x/\|x\|^2$ in (3.6), we have the following result.

Corollary 3.3.1 (Efron and Morris(1976)). *For the estimator of the form* (3.7), *assume that* $\phi(w)$ *is given by solutions of the following differential inequality*

$$\phi(w)\left(2(p-2) - \phi(w)\right)/w + 4\phi'(w) \geq 0, \tag{3.8}$$

which is, for example, satisfied by

(A1) $\phi(w)$ *is monotone nondecreasing.*

(A2) $0 \leq \phi(w) \leq 2(p-2)$ *for every* $w \geq 0$.

Then δ_ϕ *is minimax.*

Thus a class of the estimators better than X is constructed. Since the inequality (3.8) are satisfied by $\phi(w) = p - 2$, the James-Stein estimator δ^{JS} is included in this class. It is noted that the sufficient condition for minimaxity, (A1) and (A2), had been derived by Baranchik(1970) before Stein(1973).

Remark 3.4. Although the monotonity of ϕ which is required by (A1) is very comprehensible and is easily checked by differentiating $\phi(w)$, the differential inequality (3.8) allows $\phi(w)$ to decrease. Indeed (3.8) is equivalent to that

$$\frac{w^{p/2-1}\phi(w)}{2(p-2) - \phi(w)} \tag{3.9}$$

is nondecreasing, as shown in Efron and Morris(1976). Hence considering the case where (3.9) equals a nonnegative constant c, we have an estimator which has constant risk p, $\delta_d(X) = (1 - \phi_d(\|X\|^2)/\|X\|^2)X$ where

$$\phi_d(w) = \frac{2c(p-2)}{c + w^{p/2-1}}.$$

Notice that $\phi_d(w)$ is strictly monotone decreasing in w and that $c = 0$ corresponds to the usual estimator X while $c = \infty$ corresponds to the James-Stein estimator $(1 - 2(p-2)/\|X\|^2)X$. These facts on $\phi_d(w)$ were first pointed out by DasGupta and Strawderman(1997).

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 James-Stein estimator

$$\delta_+^{JS} = \max\left(0, 1 - \frac{p-2}{\|X\|^2}\right) X$$

is considered and it dominates the James-Stein estimator as shown in Baranchik(1964).

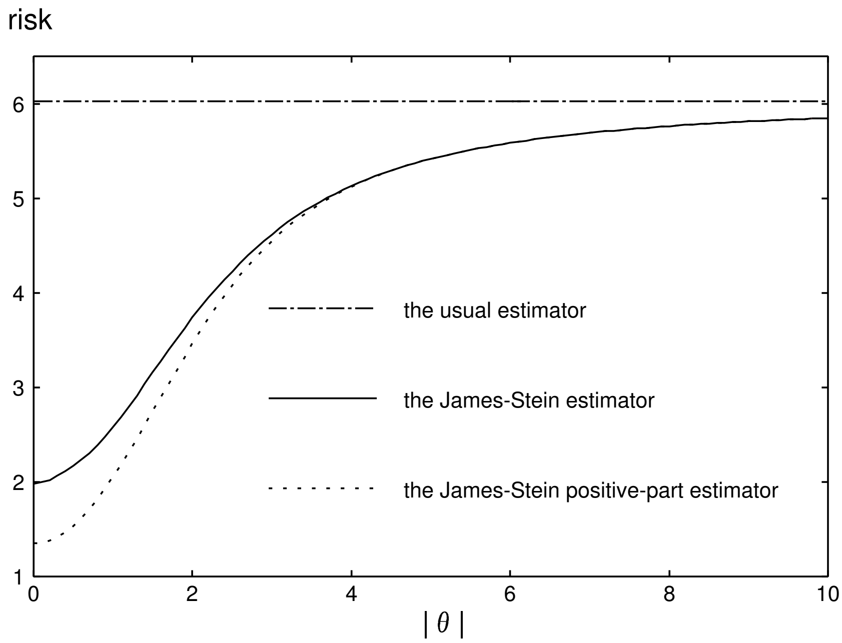

Figure 3.1 : Comparison of the risks of X, δ^{JS} and δ_+^{JS}

Figure 3.1 gives a comparison of the respective risks of X, δ^{JS} and δ_+^{JS}, for $p = 6$. Furthermore the complete class theorem, which will be introduced in the next section, implies that the positive-part James-Stein estimator is not analytic and thus is inadmissible. Since we are interested in proposing a class of admissible minimax estimators, we will review, in the next section, some results of complete class theorem, admissibility and permissibility in this problem.

3.2 Admissibility and permissibility

3.2.1 Preparations

First we briefly review a generalized Bayes estimator since it plays an important role in this section. Let F be a nonnegative σ-finite Borel measure on R^p such that $f(x) = \int p_\theta(x)F(d\theta) < \infty$ for almost all x in R^p. Then the generalized Bayes estimator with respect to F exists and is denoted by δ_F. The estimator δ_F is given by

$$\delta_F(X) = \frac{\int \theta p_\theta(X)F(d\theta)}{\int p_\theta(X)F(d\theta)}.$$

It is easy to see, by differentiating under integral sign that we can write $\delta_F(X) = \nabla f(X)/f(X) + X$.

The basic tool for our study is the necessary and sufficient condition for admissibility due to Farrell(1968). Although his original result includes many regularity conditions, the result for our estimation problem, can be simplified as follows.

Theorem 3.5 (Farrell(1968)). *δ is admissible if and only if there exists a sequence of finite measures $\{G_n\}$ satisfying*

1. *$\{G_n\}$ has compact support,*
2. *$G_n(\{\theta : \|\theta\| \leq 1\}) \geq 1$ for all n and*

$$\lim_{n\to\infty} \int \left(R(\theta, \delta) - R(\theta, \delta_{G_n})\right) G_n(d\theta) = 0. \tag{3.10}$$

proof. Since the necessity of a method of proof is quite deep, we omit it in this thesis. We shall prove the sufficiency below. Suppose δ is not admissible. Then there is an estimator δ' such that $R(\theta, \delta') \leq R(\theta, \delta)$ and

$$\int \|\delta'(x) - \delta(x)\| dx > 0. \tag{3.11}$$

We define δ'' by $\delta'' = (\delta + \delta')/2$. Then, using Jensen's inequality and (3.11), we have

$$\begin{aligned} R(\theta, \delta'') &= \int \|\delta''(x) - \theta\|^2 p_\theta(x) dx \\ &< \frac{1}{2}\left(\int \|\delta(x) - \theta\|^2 p_\theta(x) dx + \int \|\delta'(x) - \theta\|^2 p_\theta(x) dx\right) \\ &= \left(R(\theta, \delta) + R(\theta, \delta')\right)/2 \\ &\leq R(\theta, \delta). \end{aligned}$$

$R(\theta, \delta'')$ and $R(\theta, \delta)$ are both continuous functions. Hence the inequality above yields the existence of an $\epsilon > 0$ such that $R(\theta, \delta'') < R(\theta, \delta) - \epsilon$ for $\|\theta\| \leq 1$. Hence if G_n satisfies $G_n(\{\theta : \|\theta\| \leq 1\}) \geq 1$, we have

$$\begin{aligned} \int \left(R(\theta, \delta) - R(\theta, \delta_{G_n})\right) G_n(d\theta) &\geq \int \left(R(\theta, \delta) - R(\theta, \delta'')\right) G_n(d\theta) \\ &\geq \int_{\|\theta\| \leq 1} \left(R(\theta, \delta) - R(\theta, \delta'')\right) G_n(d\theta) \\ &\geq \epsilon. \end{aligned}$$

This contradicts (3.10). It follows that if (3.10) is satisfied, δ is admissible. □

If $u : R^p \to R$, we shall say that u is *piecewise differentiable* if there exists a countable collection of disjoint open set $\{O_i\}$ such that $R^p = \cup_{i=1}^{\infty} \bar{O}_i$ ($\bar{O}_i$ is the closure of O_i) and u is continuously differentiable in each O_i.

3.2.2 The necessary condition for the admissibility

Interchanging the order of integration we have, as in James and Stein(1961), for any estimator δ

$$\begin{aligned}\int (R(\theta,\delta) - R(\theta,\delta_{G_n}))G_n(d\theta) &= \int\int \left(\|\delta-\theta\|^2 - \|\delta_{G_n}-\theta\|^2\right) p_\theta(x)G_n(d\theta) \\ &= \int \Bigg(\|\delta(x)\|^2 g_n(x) - 2\delta(x)\cdot\int \theta p_\theta(x)G_n(d\theta) \\ &\quad -\|\delta_{G_n}(x)\|^2 g_n(x) + 2\delta_{G_n}(x)\cdot\int \theta p_\theta(x)G_n(d\theta)\Bigg)\,dx \\ &= \int \|\delta(x)-\delta_{G_n}(x)\|^2 g_n(x)dx. \end{aligned} \tag{3.12}$$

Using a continuity theorem for Laplace transforms shown in Brown(1971) and (3.12), we have the following result, what is called, Brown's complete class theorem.

Theorem 3.6 (Brown(1971)). *If δ is admissible, there is a nonnegative measure F such that $f(x) < \infty$ for all x and such that $\delta(X) = \delta_F(X)$ a.e. (dx).*

proof. If $\{G_n\}$ is a sequence satisfying part *2* of Theorem 3.5 then we have

$$\begin{aligned} g_n(x) &\geq (2\pi)^{-p/2}\int_{\|\theta\|\leq 1} \exp(-\|x\|^2/2 + \theta\cdot x - \|\theta\|^2/2)G_n(d\theta) \\ &\geq (2\pi)^{-p/2}\exp(-\|x\|^2/2 - \|x\|)\int_{\|\theta\|\leq 1}\exp(-\|\theta\|^2/2)d\theta \\ &= C\exp(-\|x\|^2/2 - \|x\|), \end{aligned} \tag{3.13}$$

for some constant $C(>0)$. Hence (3.12) and (3.13) imply that $\nabla g_n(x)/g_n(x)$ converges to $\delta(x) - x$ in measure (dx) on each compact set in R^p. Thus there is a subsequence $\{n'\}$ such that

$$\|\nabla g_{n'}(x)/g_{n'}(x)\| \to \|\delta(x) - x\| < \infty \quad \text{a.e.} \quad (dx).$$

Defining $\mu_n(d\theta) = \exp(-\|\theta\|^2/2)G_n(d\theta)$ and the multivariate Laplace transform $\tilde{\mu}_i$ by

$$\tilde{\mu}_i(x) = \int \exp(x\cdot\theta)\mu_i(d\theta)$$

for $x \in R^p$, the above implies that $\nabla\tilde{\mu}_{n'}(x)/\tilde{\mu}_{n'}(x)$ converges to $\delta(x)$ a.e. (dx). Thus, using a continuity theorem for multivariate Laplace transform shown in Brown(1971), there is a measure $\mu_0(d\theta)$ and a subsequence $\{n''\}$ of $\{n'\}$ such that $\nabla\tilde{\mu}_{n''}(x)/\tilde{\mu}_{n''}(x) \to \nabla\tilde{\mu}_0(x)/\tilde{\mu}_0(x)$ for all $x \in R^p$. It follows that $\nabla\tilde{\mu}_0(x)/\tilde{\mu}_0(x) = \delta(x)$ a.e. (dx). Defining $F(d\theta) = \exp(\|\theta\|^2/2)\mu_0(d\theta)$, we have $f(x) = \int p_\theta(x)F(d\theta) < \infty$ and $x + \nabla f(x)/f(x) = \nabla\tilde{\mu}_0(x)/\tilde{\mu}_0(x)$ and hence $\delta(x) = \delta_F(x)$ a.e. (dx). This completes the proof. □

3.2.3 The sufficient condition for admissibility

As shown in Theorem 3.6, admissible estimators are generalized Bayes. Therefore, our study of admissibility of estimators can be confined to generalized Bayes estimators. The main aim of this subsection is to obtain the sufficient conditions on $f(x)$ for δ_F to be admissible. Throughout the rest of this subsection we assume that F is a fixed nonnegative σ-finite Borel measure with unbounded support. (If the support of F is bounded then F is a finite measure and therefore δ_F is proper Bayes and hence admissible.) We also assume that the closed convex hull of the support is equal to R^p. For example, this assumption is satisfied if F is spherically symmetric.

Moreover we restrict estimators considered in this subsection ones which have bounded risk, because we are interested in minimaxity, that is, risk bounded less than p. A necessary and sufficient condition for bounded risk is the following.

Theorem 3.7 (Brown(1971)). *There is a constant r such that $R(\theta, \delta_F) < r$ for all $\theta \in R^p$ if and only if there is a constant B such that*

$$\|\nabla f(x)/f(x)\| < B \qquad \text{for all} \quad x \in R^p. \tag{3.14}$$

proof. The "only if" part is omitted. See Theorem 3.3.1 of Brown(1971). We shall prove the sufficiency below. Suppose that (3.14) is satisfied. When we have

$$\begin{aligned} R(\theta, \delta_F) &= \int \|x + \nabla f(x)/f(x) - \theta\|^2 p_\theta(x)dx \\ &\le 2\int \left(\|x-\theta\|^2 + \|\nabla f(x)/f(x)\|^2\right) p_\theta(x)dx \\ &\le 2p + B^2. \end{aligned}$$

This completes the proof. □

Using the fact that $\delta_F(x) = x + \nabla f(x)/f(x)$ and interchanging the order of integration in (3.12), we have, as in Brown(1971),

$$\begin{aligned}
\int (R(\theta, \delta_F) - R(\theta, \delta_{G_n}))G_n(d\theta) &= \int \|\delta_F(x) - \delta_{G_n}(x)\|^2 g_n(x)dx \\
&= \int \left\| \frac{\nabla f(x)}{f(x)} - \frac{\nabla g_n(x)}{g_n(x)} \right\|^2 g_n(x)dx \\
&= \int \left\| \frac{f(x)\nabla g_n(x) - g_n(x)\nabla f(x)}{f(x)^2} \right\|^2 \frac{f(x)^2}{g_n(x)}dx \\
&= 4\int \frac{\|\nabla h_n(x)\|^2}{h_n(x)} f(x)dx \\
&= \int \|\nabla h_n(x)^{1/2}\|^2 f(x)dx, \qquad (3.15)
\end{aligned}$$

where $h_n(x) = g_n(x)/f(x)$. This identity plays a crucial role in the rest of this subsection.

We note that by (3.13), $g_n(x)$ is greater than $\exp(-3/2)$ for $\|x\| \leq 1$. Multiplying F by a positive constant does not affect the value of $R(\theta, \delta_F) - R(\theta, \delta_{G_n})$. Without loss of generality, we may thus assume F has been normalized on the unit sphere so that $h_n(x) \geq 1$ for $\|x\| \leq 1$ for all n. Moreover we have the following result.

Lemma 3.8 (Birnbaum(1955)). *If a sequence of* $\{G_n\}$ *has a compact support,*

$$\lim_{r\to\infty} \sup_{\|x\|=r} h_n(x) = 0 \qquad (3.16)$$

for all n.

Now let J be the class of all nonnegative piecewise differentiable real valued functions j defined R^p satisfying

i) $j(x) \geq 1$ for $\|x\| \leq 1$,

ii) $\lim_{r\to\infty} \sup_{\|x\|=r} j(x) = 0$ for all n.

By (3.15) it is easy to see that

$$\int (R(\theta, \delta_F) - R(\theta, \delta_{G_n}))G_n(d\theta) \geq \inf_{j\in J} \int \|\nabla j(x)\|^2 f(x)dx \tag{3.17}$$

for all n. By Theorem 3.5, if δ_F is admissible

$$\inf_{j\in J} \int \|\nabla j(x)\|^2 f(x)dx \tag{3.18}$$

equals to zero. In particular, if (3.18) is positive then δ_F is inadmissible. The converse, that (3.18) is zero implies δ_F is admissible, is our goal.

Theorem 3.9 (Brown(1971) and Srinivasan(1981)). *The bounded risk generalized Bayes estimator δ_F is admissible if and only if*

$$\inf_{j\in J} \int \|\nabla j(x)\|^2 f(x)dx = 0.$$

proof. The "only if" part is trivial. We shall the sufficiency below. The proof involves constructing a sequence of finite measures $\{G_n\}$ as in Theorem 3.5 and showing that the sequence satisfies (3.10).

Let $j_r(x)$ be a sequence of nonnegative functions satisfying

$$j_r(x) = \begin{cases} 1 & \text{for} \quad \|x\| \leq 1 \\ \exp(-2Br) & \text{for} \quad r \leq \|x\| \leq 2r \\ 0 & \text{for} \quad \|x\| > 2r \end{cases} \tag{3.19}$$

and $L_f j_r(x) = 0$ for $1 < \|x\| < r$. Such a sequence exists by Theorem 3.11 in the next subsection. Let $G_r(d\theta) = j_r^2(\theta)F(d\theta)$, $g_r(x) = \int p_\theta(x)G_r(d\theta)$ and $\psi_r(x) = \int j_r(\theta)p_\theta(x)F(d\theta)/f(x)$. We shall now prove that

$$\int (R(\theta, \delta_F) - R(\theta, \delta_{G_r}))\, G_r(d\theta) \tag{3.20}$$

converges to 0 as $r \to \infty$. We can write (3.20) as

$$\int_{\|x\|<2r} \left\| \int \left(j_r^2(\theta) - \psi_r^2(x)\right)(\theta - \delta_F(x)) \frac{p_\theta(x)F(d\theta)}{g_r(x)} \right\|^2 g_r(x)dx$$

$$\int_{\|x\|\geq 2r} \left\| \int j_r^2(\theta)\,(\theta - \delta_F(x)) \frac{p_\theta(x)F(d\theta)}{g_r(x)} \right\|^2 g_r(x)dx. \tag{3.21}$$

We will call the two terms in (3.21) T_1 and T_2 respectively. We first consider the second term T_2. Since $j_r(\theta) = 0$ for $\|\theta\| > 2r$, using Schwartz inequality and (3.14), we have

$$T_2 \leq C \int_{\|x\|>2r} \|x\|^2 \int_{\|\theta\|<2r} j_r^2(\theta) p_\theta(x) F(d\theta) dx, \tag{3.22}$$

for some constant C. Now let $D = \{\theta : \|\theta\| \leq r\}$ and $E = \{x : \|x\| > 3r\}$. It is easy to see that (3.14) implies $\int \exp(-B\|\theta\|) F(d\theta) = B_1 < \infty$. Hence using maximum modulus principle on $j_r(\theta)$ shown in Srinivasan(1981), we have

$$\int_{\|x\|>2r} \|x\|^2 \int_D j_r^2(\theta) p_\theta(x) F(d\theta) dx \leq B_1 \exp(Br) \int_{\|x\|>2r} \|x\|^2 \exp\left(-(\|x\|^2 - r)^2/2\right) dx. \tag{3.23}$$

The right-hand side of (3.23) is easily seen to converge to zero as $r \to \infty$. By a similar argument we can show that

$$\lim_{r\to\infty} \int_E \|x\|^2 \int_{\|\theta\|\leq 2r} j_r^2(\theta) p_\theta(x) F(d\theta) = 0. \tag{3.24}$$

Moreover from the fact that $j_r(\theta) = \exp(-2Br)$ for $r < \|\theta\| < 2r$, we have

$$\lim_{r\to\infty} \int_{2r<\|x\|<3r} \|x\|^2 \int_{r<\|\theta\|<2r} j_r^2(\theta) p_\theta(x) F(d\theta) = 0.$$

Hence we have proved that $T_2 \to 0$ as $r \to \infty$. We will now complete the proof of the theorem by showing T_1 goes to zero as $r \to \infty$. Expressing $(j_r^2(\theta) - \psi_r^2(x))$ as $(j_r(\theta) - \psi_r(x))(j_r(\theta) + \psi_r(x))$ and using Schwartz inequality we have

$$\begin{aligned} T_1 \leq 2 \int_{\|x\|<2r} &\left(\int (j_r(\theta) - \psi_r(x))^2 \|\theta - \delta_F(x)\|^2 p_\theta(x) F(d\theta) \right) \\ &\times \left(\int \left(j_r^2(\theta) + \psi_r^2(x)\right) p_\theta(x) F(d\theta) \right) g_r^{-1}(x) dx. \end{aligned} \tag{3.25}$$

Since the inequality $\psi_r^2(x) f(x) \leq g_r(x)$ is satisfied by Schwartz inequality, T_1 is evaluated as

$$
\begin{aligned}
T_1 &\le 4\int_{\|x\|<2r}\int (j_r(\theta)-\psi_r(x))^2\|\theta-\delta_F(x)\|^2 p_\theta(x)F(d\theta)dx \\
&\le 4\int_{\|x\|<2r}\int (j_r(\theta)-j_r(x))^2\|\theta-\delta_F(x)\|^2 p_\theta(x)F(d\theta)dx \\
&\quad +4\int_{\|x\|<2r}\int (j_r(x)-\psi_r(x))^2\|\theta-\delta_F(x)\|^2 p_\theta(x)F(d\theta)dx. \qquad (3.26)
\end{aligned}
$$

We call the two terms in (3.26) T_3 and T_4 respectively. By Lemmas A.3, A.4 and (3.14), we have

$$
\begin{aligned}
T_3 &\le C\left[\int_{\|x\|<2r}\int (j_r(\theta)-j_r(x))^2\|\theta-x\|^2 p_\theta(x)F(d\theta)dx \right.\\
&\quad \left. +\int_{\|x\|<2r}\int (j_r(\theta)-j_r(x))^2 p_\theta(x)F(d\theta)dx\right] \\
&\le C_1\int \|\nabla j_r(x)\|^2 f(x)dx \qquad (3.27)
\end{aligned}
$$

where C and C_1 are some positive constants. The choice of j_r's imply that (3.27) goes to zero and hence T_3 goes to zero as $r\to\infty$. Finally we consider the term T_4. Using Schwartz inequality, Lemma A.4 and (3.14),

$$
\begin{aligned}
T_4 &\le \int_{\|x\|<2r}\int \left(\int (j_r(x)-j_r(\eta))^2 p_\eta(x)F(d\eta)\right)\frac{\|\theta-\delta_F(x)\|^2}{f(x)} p_\theta(x)F(d\theta)dx \\
&\le C_2\int_{\|x\|<2r}\int (j_r(x)-j_r(\eta))^2 p_\eta(x)F(d\eta)dx,
\end{aligned}
$$

for some constant C_2. Appealing to Lemma A.3, it is easy to see that

$$
T_4 \le C_3\int \|\nabla j_r(x)\|^2 f(x)dx \to 0 \quad \text{as} \quad r\to\infty. \qquad (3.28)
$$

Hence T_1 goes to zero as $r\to\infty$ and this completes the proof. □

Now we restrict the class of the estimators considered in this subsection, to the class of generalized Bayes estimators with respect to spherically symmetric generalized prior distributions. Namely, for any measure set $A\in R^p$ and any orthogonal transformation

Γ, $F(A) = F(\Gamma A)$ is satisfied. In this case, the marginal density function $f(x)$ satisfies $f(\Gamma x) = f(x)$ for any Γ, because we have

$$\begin{aligned} f(\Gamma x) &= \int_{R^p} \exp(-\|\theta - \Gamma x\|^2/2) F(d\theta) \\ &= \exp(-\|x\|^2/2) \int_{R^p} \exp(x'\Gamma'\theta) \exp(-\|\theta\|^2/2) F(d\theta) \\ &= \exp(-\|x\|^2/2) \int_{R^p} \exp(x'\eta) \exp(-\|\eta\|^2/2) F_\Gamma(d\eta) \\ &= \exp(-\|x\|^2/2) \int_{R^p} \exp(x'\eta) \exp(-\|\eta\|^2/2) F(d\eta) \\ &= f(x), \end{aligned}$$

where F_Γ is a measure such that $F(\Gamma A) = F_\Gamma(A)$ for any measure set $A \in R^p$. Therefore $f(x)$ is a function only of $\|x\|^2$ and we put $m(\|x\|^2) = f(x)$. We have the following result as the corollary of Theorem 3.9.

Corollary 3.9.1. *Suppose that F is spherically symmetric. The bounded risk generalized Bayes estimator δ_F is admissible if and only if*

$$\int_1^\infty t^{-p/2} m(t)^{-1} dt = \infty.$$

proof. The necessity part will be proved in Theorem 3.12, where we show that if $\int_1^\infty t^{-p/2} m(t)^{-1} dt < \infty$ then δ_F is not permissible and is thus inadmissible. We shall prove the sufficiency below. We define

$$j_i(x) = \begin{cases} 1 & \text{for} \quad \|x\|^2 \le 1 \\ (\int_1^i t^{-p/2} m(t)^{-1} dt)^{-1} \int_{\|x\|^2}^i t^{-p/2} m(t)^{-1} dt & \text{for} \quad 1 < \|x\|^2 \le i \\ 0 & \text{for} \quad \|x\|^2 > i. \end{cases}$$

A direct computation yields

$$\begin{aligned}\int \|\nabla j_i(x)\|^2 f(x)dx &= 4\left(\int_1^i t^{-p/2}m(t)^{-1}dt\right)^{-2}\int_{\{x:1\le\|x\|^2\le i\}}\|x\|^{2-p}m(\|x\|^2)^{-2}f(x)dx\\ &= 4\left(\int_1^i t^{-p/2}m(t)^{-1}dt\right)^{-2}\int_{\{x:1\le\|x\|^2\le i\}}\|x\|^{2-p}m(\|x\|^2)^{-1}dx\\ &= C\left(\int_1^i t^{-p/2}m(t)^{-1}dt\right)^{-1}\end{aligned}$$

for some positive constant C. Therefore we have

$$\lim_{i\to\infty}\int \|\nabla j_i(x)\|^2 f(x)dx = \lim_{i\to\infty}\left(\int_1^i t^{-p/2}m(t)^{-1}dt\right)^{-1} = 0$$

and this completes the proof. □

By using this theorem, we can prove part *2* of Theorem 3.1.

proof of part 2 of Theorem 3.1. Since the estimator X is the generalized Bayes estimator with respect to the Lebesque measure on R^p, $m(t)$ is constant. Clearly $\int_1^\infty t^{-p/2}dt$ diverges for $p = 1, 2$. This completes the proof. □

By Theorem 3.6, it is very important to determine whether a proposed estimator is generalized Bayes or not. We note that every estimator of the form (3.7) is able to be written as

$$X + \nabla \log m_\phi(\|X\|^2), \tag{3.29}$$

where

$$m_\phi(w) = \exp\left(-\frac{1}{2}\int_\epsilon^w \frac{\phi(w)}{w}dw\right), \tag{3.30}$$

for every $\epsilon > 0$. Because (3.29) takes the same form of the generalized Bayes estimator, Bock(1988) called it a *pseudo-Bayes estimator* and investigated its properties. If there exists a measure π which satisfies

$$m_\phi(\|x\|^2) = \int_{R^p}\exp(-\|\theta - x\|^2/2)\pi(d\theta), \tag{3.31}$$

the estimator of the form (3.29) is truly generalized Bayes. Generally, it is however quite difficult to find such a p-dimensional measure π. Strawderman and Cohen(1971) derived one-dimensional necessary condition for an estimator of the form (3.29) to be generalized Bayes.

Theorem 3.10 (Strawderman and Cohen(1971)). *A necessary condition for an estimator of the form* (3.29) *to be generalized Bayes, is that there exists a one-dimensional measure ν which satisfies*

$$m_\phi(y^2) = \exp(-y^2/2)\int_{-\infty}^{\infty} \exp(y\eta)\nu(d\eta). \tag{3.32}$$

proof. Since an estimator of the form (3.29) is generalized Bayes, there exists a measure π which satisfies (3.31). Noting that the left-hand side of (3.31) is spherically symmetric, for any orthogonal transformation Γ, we have

$$m_\phi(\|x\|^2) = \int_{R^p} \exp(-\|\theta - x\|^2/2)\pi(d\theta) = \int_{R^p} \exp(-\|\theta - \Gamma x\|^2/2)\pi(d\theta). \tag{3.33}$$

Choosing the orthogonal transformation Γ_x which carries x into $(\|x\|, 0, \cdots, 0)$, we have

$$m_\phi(\|x\|^2) = \exp(-\|x\|^2/2)\int_{-\infty}^{\infty} \exp(\theta_1\|x\|)\nu(d\theta_1), \tag{3.34}$$

where

$$\nu(d\theta_1) = \int_{\theta_2} \cdots \int_{\theta_p} \exp(-\|\theta\|^2/2)\pi(d\theta). \tag{3.35}$$

By letting $y = \|x\|$ and $\eta = \theta_1$, this completes the proof. □

3.2.4 Exterior boundary problem and recurrent diffusions

We observed in the previous subsection, that the following calculus of variation problem

$$\inf_{j\in J} \int \|\nabla j(x)\|^2 f(x)dx$$

is crucial to our study of admissibility of δ_F. Hence finding conditions under which this infimum is not zero is important. We introduce, below an exterior boundary value problem which describes when this infimum is zero.

Let L_f denote the elliptic differential operator given by

$$L_f u(x) = \Delta u(x) + \nabla \log f(x) \cdot \nabla u(x) \tag{3.36}$$

where u is a twice continuously differential function. We say that Exterior Boundary Problem for L_f(BP for L_f) is solvable if there exists a unique bounded solution u_0 for the equation $L_f u(x) = 0$ in the region $\{x : \|x\| > 1\}$ satisfying the condition $u_0(x) = 1$ on $\|x\| = 1$. Note that the unique bounded solution u_0 is identically equal to 1. The following result is well-known.

Theorem 3.11. *A necessary and sufficient condition for*

$$\inf_{j \in J} \int \|\nabla j(x)\|^2 f(x) dx = 0$$

is BP for L_f is solvable.

proof. See Meyers and Serrin(1960) and Srinivasan(1981). □

The operator $L_f u(x)$ can be considered as the generator of the diffusion with the local covariance $2I$ and the local mean $\nabla \log f(x)$. There exist some bounded solutions of $L_f u(x) = 0$ if and only if the diffusion is transient in which case $u(x)$ can be taken to the probability of the reaching the unit ball. Thus the estimator is admissible if and only if the diffusion process determined by $L_f u(x)$ is recurrent.

For the usual estimator X, the corresponding diffusion process is Brownian motion, which is known to be recurrent if and only if the dimension p does not exceed two. Equation $L_f u(x) = 0$ now means that

$$\Delta u(x) = 0 \qquad \text{for} \quad \|x\| \geq 1,$$

i.e. u is a bounded harmonic function outside the unit ball whose determination is a classical Dirichlet boundary problem. Putting $u(x) = v(\|x\|^2)$, one obtains

$$pv'(y) + 2yv''(y) = 0, \qquad \text{for} \quad y \geq 1.$$

All solutions of this equation are $v(y) = 1$ for all p and

$$v(y) = \begin{cases} y^{1-p/2}, & \text{for} \quad p \neq 2 \\ \log y + 1 & \text{for} \quad p = 2. \end{cases}$$

Hence only for $p = 1, 2$, the exterior boundary problem for L_f is solvable.

3.2.5 Permissibility

As shown in Chapter 2, permissibility is a weaker criterion of optimality than admissibility. It is however noted that if the estimator δ is permissible then it is impossible to find an estimator δ' whose unbiased estimate of risk is always less than that of δ. In short, if δ is inadmissible, the customary method for finding an estimator improving upon δ cannot succeed.

Suppose that a proposed estimator δ_ϕ is the form of (3.29) and that $\delta_{\phi,h}(X) = X + \nabla \log\{m_\phi(\|X\|^2)h(\|X\|^2)\}$ is a proposed competing estimator. In the following of this subsection, let $m = m_\phi$ for simplicity. Then substituting $g(x) = \nabla \log m(\|x\|^2)$ and $g(x) = \nabla \log\{m(\|x\|^2)h(\|X\|^2)\}$ in (3.5), we have the unbiased estimators of risk as $\hat{R}(\delta_\phi) = p + 4\mathcal{R}(\delta_\phi)$ where

$$\mathcal{R}(\delta_\phi) = -w \left(\frac{m'(w)}{m(w)} \right)^2 + p\frac{m'(w)}{m(w)} + 2w\frac{m''(w)}{m(w)} \tag{3.37}$$

and $w = \|x\|^2$ and $\hat{R}(\delta_{\phi,h}) = p + 4\mathcal{R}(\delta_{\phi,h})$ where

$$\begin{aligned} \mathcal{R}(\delta_{\phi,h}) &= -w \left(\frac{(m(w)h(w))'}{m(w)h(w)} \right)^2 + p\frac{m'(w)}{m(w)} + p\frac{h'(w)}{h(w)} + 2w\frac{(m(w)h(w))''}{m(w)h(w)} \\ &= \mathcal{R}(\delta_\phi) - w\left(\frac{h'(w)}{h(w)} \right)^2 + 2w\frac{h''(w)}{h(w)} + p\frac{h'(w)}{h(w)} + 2w\frac{h'(w)}{h(w)}\frac{m'(w)}{m(w)}. \end{aligned}$$

The estimator is then permissible if $\mathcal{R}(\delta_{\phi,h}) \leq \mathcal{R}(\delta_\phi)$ implies $h(w) \equiv 1$. Now we assume that $\int_0^1 w^{-p/2}m^{-1}(w)dw = \infty$, which is satisfied by reasonable estimators, for example, the generalized Bayes estimator δ_F such that $f(x) < \infty$ and the James-Stein estimator. The following result is a special case of Brown(1988)'s result.

Theorem 3.12 (Brown(1988)). *The estimator of the form* (3.29) *is permissible if and only if*

$$\int_1^\infty w^{-p/2} m(w)^{-1} dw = \infty.$$

proof. We first consider the "only if" part. We assume that $\int_1^\infty w^{-p/2} m(w)^{-1} dw < \infty$. By letting $M(w) = w^{p/2} m(w)$, the difference $\mathcal{R}(\delta_{\phi,h}) - \mathcal{R}(\delta_\phi)$ is written as

$$2\frac{M'(w)}{M(w)} h'(w) + 2h''(w) - h(w) \left(\frac{h'(w)}{h(w)}\right)^2. \tag{3.38}$$

Note that the sum of the first and second terms is interpreted as the elliptic differential operator $L_M h(w)$. Let $h_1(w)$ be a function satisfying

$$h_1(w) = \begin{cases} 1 & \text{for} \quad w \le 1 \\ \left(\int_1^\infty M^{-1}(t) dt\right)^{-1} \int_w^\infty M^{-1}(t) dt & \text{for} \quad w > 1. \end{cases}$$

Then we have $\mathcal{R}(\delta_{\phi,h_1}) - \mathcal{R}(\delta_\phi) = 0$ for $w \le 1$ and

$$\mathcal{R}(\delta_{\phi,h_1}) - \mathcal{R}(\delta_\phi) = -h_1(w) \left(\frac{h_1'(w)}{h_1(w)}\right)^2 \le 0$$

for $w > 1$ with strict inequality for some w which implies that δ_ϕ is not permissible.

Next we shall prove the sufficiency. We assume that $\int_1^\infty w^{-p/2} m(w)^{-1} dw = \infty$. By letting $g(w) = h'(w)M(w)/h(w)$, the inequality $\mathcal{R}(\delta_{\phi,h}) - \mathcal{R}(\delta_\phi) \le 0$ is written as

$$h(w) \left(2\frac{g'(w)}{M(w)} + \left(\frac{g(w)}{M(w)}\right)^2\right) \le 0. \tag{3.39}$$

We shall now show that the inequality (3.39) implies that $g(w) \ge 0$ for all w. Suppose to the contrary that $g(w) < 0$ for some w_0. Then $g(w) < 0$ for all $w \ge w_0$ since $g'(w) < 0$ and for all $w > w_0$, we can write (3.39) as

$$\frac{d}{dw}\left(\frac{1}{g(w)}\right) \ge \frac{1}{2M(w)}.$$

Integrating both sides from w_0 to w^* leads to

$$\frac{1}{g(w^*)} - \frac{1}{g(w_0)} \geq \frac{1}{2}\int_{w_0}^{w^*} M^{-1}(t)dt.$$

As $w^* \to \infty$, the right-hand side of above inequality tends to infinity, and this provides a contradiction since the left-hand side is less than $-1/g(w_0)$.

Similarly we have $g(w) \leq 0$ for all w. It follows that $g(w)$ is zero for all w, which implies that $h(w) \equiv 1$ for all w. This completes the proof. □

By Corollary 3.9.1 and Theorem 3.12, we have the following result.

Proposition 3.13. *A bounded risk permissible estimator is admissible if it is generalized Bayes and inadmissible if not.*

It is difficult to find estimators dominating an estimator which is both permissible and inadmissible, because we have to overcome the limitation of the characteristics through the Stein identity. For example, the corresponding $m(\|x\|^2)$ for the James-Stein estimator is $\|x\|^{2-p}$ and the James-Stein estimator is permissible since $\int_1^\infty (t^{p/2}t^{1-p/2})^{-1}dt = \infty$. Finding estimators dominating the James-Stein estimator except for the positive-part estimator is therefore difficult. In Section 3.4, we deal with the problem of improving upon the James-Stein estimator.

3.3 The class of admissible minimax estimators

3.3.1 Generalized Bayes estimators and their admissibility

In this subsection, we construct a certain class of generalized Bayes estimators and consider their admissibility. All but Lemma 3.17 is due to the hitherto researches, in particular Fourdrinier *et al.*(1998). We deal with a convenient subclass of spherically symmetric prior distributions, a class of variance mixtures of multivariate normal distributions, whose density is proportional to

$$\int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left(-\frac{\lambda}{2(1-\lambda)}\|\theta\|^2\right) \lambda^{-a} h(\lambda) d\lambda, \tag{3.40}$$

where $h(\lambda)$ is a measurable positive function on $(0,1)$ and we assume that $\lim_{\lambda\to 0} h(\lambda) = c_1 > 0$. This prior distribution is interpreted as the following two-stage prior distribution. For a fixed value of λ, let the θ be according to $N(0, \lambda^{-1}(1-\lambda)I_p)$. In addition, suppose that λ itself is a random variable, Λ, with distribution $\Lambda \sim \lambda^{-a}h(\lambda)$. Hence this prior distribution is proper provided $\int_0^1 \lambda^{-a}h(\lambda)d\lambda$ is finite. By using Fubini's theorem for positive functions, the marginal density function of X is calculated as

$$
\begin{aligned}
f_h(x) &\propto \int_0^1 \int_{R^p} \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \lambda^{-a}h(\lambda) \\
&\quad \times \exp\left(-\frac{\|x-\theta\|^2}{2} - \frac{\lambda}{2(1-\lambda)}\|\theta\|^2\right) d\theta d\lambda \\
&= \int_0^1 \int_{R^p} \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left(-\frac{\lambda}{2}\|x\|^2\right) \\
&\quad \times \exp\left(-\frac{\|\theta-(1-\lambda)x\|^2}{2(1-\lambda)}\right) \lambda^{-a}h(\lambda) d\theta d\lambda \\
&\propto \int_0^1 \exp\left(-\frac{\lambda}{2}\|x\|^2\right) \lambda^{p/2-a}h(\lambda)d\lambda. \qquad (3.41)
\end{aligned}
$$

Lebesque's dominated convergence theorem ensures that differentiating under the integral sign is valid. Hence we obtain

$$
\nabla f_h(x) = -\left(\int_0^1 \lambda^{p/2+1-a} \exp(-\|x\|^2\lambda/2)h(\lambda)d\lambda\right) x \qquad (3.42)
$$

and

$$
\begin{aligned}
\Delta f_h(x) &= \left(\int_0^1 \lambda^{p/2+2-a}h(\lambda)\exp(-\|x\|^2\lambda/2)d\lambda\right)\|x\|^2 \\
&\quad -p\left(\int_0^1 \lambda^{p/2+1-a}h(\lambda)\exp(-\|x\|^2\lambda/2)d\lambda\right). \qquad (3.43)
\end{aligned}
$$

Noting that the generalized Bayes estimator is written as $X + \nabla \log f_h(X)$, we see that $f_h(x)$ exists for all x and the generalized Bayes estimator is well defined if and only if

$$
\int_0^1 \lambda^{p/2-a}h(\lambda)d\lambda < \infty, \qquad (3.44)
$$

which requires that $a < p/2+1$. Therefore we have the generalized Bayes estimator

$$\delta_h(X) = \left(1 - \frac{\int_0^1 \lambda^{p/2+1-a} h(\lambda)\exp(-\|X\|^2\lambda/2)d\lambda}{\int_0^1 \lambda^{p/2-a} h(\lambda)\exp(-\|X\|^2\lambda/2)d\lambda}\right) X. \tag{3.45}$$

It is noted that by putting

$$\phi_h(w) = w\frac{\int_0^1 \lambda^{p/2+1-a} h(\lambda)\exp(-w\lambda/2)d\lambda}{\int_0^1 \lambda^{p/2-a} h(\lambda)\exp(-w\lambda/2)d\lambda}, \tag{3.46}$$

$\delta_h(X)$ is represented as $(1-\phi_h(\|X\|^2)/\|X\|^2)X$.

Concerning finiteness of the risk of $\delta_h(X)$, we have the following result.

Lemma 3.14 (Fourdrinier *et al.***(1998)).** *The generalized Bayes estimator $\delta_h(X)$ has finite risk.*

proof. As the risk of the estimator X is finite (constant risk p), by a straight application of Schwarz's inequality the risk of the estimator is evaluated as

$$\begin{aligned} R(\theta,\delta_h) &= E[\|X+\nabla\log f_h(X)-\theta\|^2] \\ &= E[\|X-\theta\|^2] + E[\|\nabla\log f_h(X)\|^2] \\ &\quad +2E[(X-\theta)'\log\nabla f_h(X)] \\ &\le p + E[\|\nabla\log f_h(X)\|^2] \\ &\quad +2\left(E[\|X-\theta\|^2]E[\|\nabla\log f_h(X)\|^2]\right)^{1/2}. \end{aligned}$$

Hence the risk is finite if and only if $E[\|\nabla\log f_h(X)\|^2] < \infty$. For any x, we have

$$\begin{aligned} \|\nabla\log f_h(x)\|^2 &= \frac{\|\nabla f_h(x)\|^2}{f_h^2(x)} \\ &= \|x\|^2\left(\frac{\int_0^1 \lambda^{p/2+1-a} h(\lambda)\exp(-\|x\|^2\lambda/2)d\lambda}{\int_0^1 \lambda^{p/2-a} h(\lambda)\exp(-\|x\|^2\lambda/2)d\lambda}\right)^2 \\ &\le \|x\|^2, \end{aligned}$$

so that $E[\|\nabla\log f_h(X)\|^2] \le E[\|X\|^2] \le p+\|\theta\|^2 < \infty$. This completes the proof. □

Now the admissibility of $\delta_h(X)$ is considered. By Theorem 3.9.1, we have only to investigate the behavior of $f_h(x)$ in the case where $\|x\|^2$ tends to infinity. A following Tauberian theorem gives a nice technique for relating the tail behavior of a function and its Laplace transform.

Theorem 3.15 (Tauberian Theorem). *For the Laplace transform*

$$f(s) = \int_0^\infty \exp(-st)g(t)dt,$$

if we have $g(t) \sim t^\gamma$ *as* $t \to +0$*, then* $f(s) \sim s^{-\gamma-1}\Gamma(\gamma+1)$ *as* $s \to \infty$*.*

Using this theorem, we can evaluate $f_h(x)$ as

$$\begin{aligned} f_h(x) &= \int_0^1 \exp\left(-\lambda\|x\|^2/2\right) \lambda^{p/2-a} h(\lambda) d\lambda \\ &= 2^{p/2-a+1} \int_0^\infty \exp\left(-\|x\|^2 t\right) t^{p/2-a} h(2t) I_{(0,1/2)}(t) dt, \\ &\sim c_1 2^{p/2-a+1}\Gamma(p/2-a+1)\|x\|^{-2(p/2-a+1)}. \end{aligned} \tag{3.47}$$

Putting $m_h(\|x\|^2) = f_h(x)$, we have

$$t^{-p/2} m_h(t)^{-1} \sim C^{-1} t^{-a+1}$$

as $t \to \infty$, where $C = c_1 2^{p/2-a+1}\Gamma(p/2-a+1)$. Therefore the integral

$$\int_1^\infty t^{-p/2} m_h(t)^{-1} dt$$

diverges if and only if $a \leq 2$. Hence we have the following result.

Theorem 3.16 (Fourdrinier *et al.*(1998)). $\delta_h(X)$ *is admissible if and only if* $a \leq 2$*.*

Moreover using Theorem 3.15, we can investigate the limitation of $\phi_h(w)$ in the case $w \to \infty$. Similarly to (3.47), we can evaluate as

$$\int_0^1 \exp\left(-\lambda\|x\|^2/2\right) \lambda^{p/2-a+1} h(\lambda) d\lambda \sim c_1 2^{p/2-a+2}\Gamma\left(p/2-a+2\right)\|x\|^{-2(p/2-a+2)},$$

which implies that when w tends to infinity, we have

$$\begin{aligned}\phi_h(w) &\sim \frac{c_1 2^{p/2-a+2}\Gamma(p/2-a+2)}{c_1 2^{p/2-a+1}\Gamma(p/2-a+1)} w^{1-(p/2-a+2)+(p/2-a+1)} \\ &= 2(p/2-a+1).\end{aligned}$$

Hence we have the following result.

Lemma 3.17. $\lim_{w\to\infty}\phi_h(w) = p - 2a + 2$, *which is not depend upon* $h(\lambda)$.

3.3.2 The construction of minimax Bayes estimator

In this subsection and next two subsections, we consider a class of minimax generalized Bayes estimators. If we are interested in deriving admissible minimax estimators, by Theorem 3.16, we have only to restrict the following results concerning minimax generalized Bayes estimators to the case of $a \le 2$.

In this subsection, we assume that

$$\lim_{\lambda\to 1} h(\lambda) = c_2(\ge 0), \tag{3.48}$$

which implies that $h(\lambda)$ is bounded. Before considering the minimaxity of δ_h, we investigate the properties of the behavior of $\phi_h(w)$ with bounded $h(\lambda)$.

Theorem 3.18.

1. *Assume that* $\lambda h'(\lambda)/h(\lambda)$ *with bounded* $h(\lambda)$ *is monotone nonincreasing. Then* $\phi_h(w)$ *is monotone increasing.*
2. *Assume that* $\lambda h'(\lambda)/h(\lambda)$ *with bounded* $h(\lambda)$ *is monotone increasing. Then* $\phi_h(w)$ *is increasing from the origin to a certain point and is decreasing from the point.*

It is noted that the assumption of part *1* implies $h'(\lambda) \le 0$ and the assumption of part *2* implies $h'(\lambda) \ge 0$ because the value $\lambda h'(\lambda)/h(\lambda)$ on $\lambda = 0$ is 0.

proof. Applying an integration by parts to the numerator on right-hand side of (3.46), we have

$$\int_0^1 \lambda^{p/2+1-a} h(\lambda) \exp(-w\lambda/2) d\lambda = -\frac{2}{w} \left[\lambda^{p/2+1-a} h(\lambda) \exp(-w\lambda/2)\right]_0^1 + \frac{p+2-2a}{w} \int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-w\lambda/2) d\lambda + \frac{2}{w} \int_0^1 \lambda^{p/2+1-a} h'(\lambda) \exp(-w\lambda/2) d\lambda. \quad (3.49)$$

It is noted that the assumption (3.48) guarantees this integration by parts. By (3.49), $\phi_h(w)$ is written as

$$\phi_h(w) = 2(p/2-a+1) + 2\frac{\int_0^1 \lambda^{p/2-a+1} h'(\lambda) \exp(-w\lambda/2) d\lambda}{\int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-w\lambda/2) d\lambda} - \frac{2c_2 \exp(-w/2)}{\int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-w\lambda/2) d\lambda}.$$

Putting

$$\varphi_1(w) = \exp(w/2) \left(\int_0^1 \lambda^{p/2-a+1} h'(\lambda) \exp(-w\lambda/2) d\lambda - c_2 \exp(-w/2) \right),$$

$$\varphi_2(w) = \exp(w/2) \left(\int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-w\lambda/2) d\lambda \right),$$

$\varphi(w) = \varphi_1(w)/\varphi_2(w)$ and

$$\psi(w) = \frac{\varphi_1'(w)}{\varphi_2'(w)} = \frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda) h'(\lambda) \exp(w(1-\lambda)/2) d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda) h(\lambda) \exp(w(1-\lambda)/2) d\lambda},$$

we have $\phi_h(w) = p - 2a + 2 + 2\varphi(w)$ and

$$\varphi'(w) = \varphi_2'(w)(\varphi_2(w))^{-1} \left(\psi(w) - \varphi(w)\right). \quad (3.50)$$

Note that $\varphi(0) = -p/2 + a - 1$ and

$$\psi(0) = \frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda) h'(\lambda) d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda) h(\lambda) d\lambda} = -p/2 + a - 1 + \frac{\int_0^1 \lambda^{p/2-a+1} h(\lambda) d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda) h(\lambda) d\lambda},$$

which implies that $\varphi(0) < \psi(0)$. Moreover we have

$$\lim_{w\to\infty} \varphi(w) = \lim_{w\to\infty} \psi(w) = 0, \tag{3.51}$$

which is shown by Tauberian's theorem.

First we prove part *1.* By Lemma 3.19 in the below, $\psi(w)$ is strictly increasing in w from $\psi(0)(< 0)$ to $\psi(\infty) = 0$. If we had the point w_0 which satisfies $\varphi(w_0) = \psi(w_0)$, it is necessary to satisfy $\varphi'(w_0) \geq \psi'(w_0) > 0$. This however contradicts the equation (3.50). Therefore for every $w \geq 0$, we have $\varphi(w) < \psi(w)$, that is, $\varphi(w)$ is increasing.

Next we prove part *2.* By Lemma 3.19, $\psi(w)$ is strictly decreasing in w from $\psi(0)(> 0)$ to $\psi(\infty) = 0$. Note that for sufficiently large w, $\varphi_1(w)$ is positive. If the inequality $\varphi(w) \leq \psi(w)$ were satisfied for every $w \geq 0$, by (3.51) we have $\varphi(w) \leq 0$ for every $w \geq 0$. This contradicts the positivity of $\varphi_1(w)$ for large w. Hence there exists w_1 which satisfies $\varphi(w_1) > \psi(w_1)$. By decreasingness of $\psi(w)$, we have the only one point w_2 which is strictly less than w_1 and satisfies $\varphi(w_2) = \psi(w_2)$. Clearly for the point w which satisfies $0 < w < w_2$, we have $\varphi(w) < \psi(w)$. For the point w which satisfies $w > w_2$, similarly to the proof of part *1*, we easily show that $\varphi(w) > \psi(w)$. This completes the proof. □

A following lemma is known as a *correlation inequality*.

Lemma 3.19 (correlation inequality). *Let Y be a random variable, and $g(y)$ and $h(y)$ any functions for which $E[g(Y)]$, $E[h(Y)]$ and $E[g(Y)h(Y)]$ exist. Then:*

1. *If one of the functions $g(\cdot)$ and $h(\cdot)$ is nonincreasing and the other is nondecreasing, then*

$$E[g(Y)]E[h(Y)] \geq E[g(Y)h(Y)].$$

2. *If both functions are either nondecreasing or nonincreasing, then*

$$E[g(Y)]E[h(Y)] \leq E[g(Y)h(Y)].$$

In the case of both part 1 and 2, the equation is satisfied if and only if either $g(\cdot)$ or $h(\cdot)$ is constant or Y has a degenerating distribution.

Combining (A1) and (A2) of Corollary 3.3.1, Lemma 3.17 and part *1* of Theorem 3.18, we have the following result which had been derived by Faith(1978).

Proposition 3.20 (Faith(1978)). *Assume that $h(\lambda)$ is a nonnegative bounded function such that $\lambda h'(\lambda)/h(\lambda)$ is monotone nonincreasing. Then the generalized Bayes estimator δ_h with $3-2/p \le a < 1+p/2$ is minimax and $\phi_h(w)$ is monotone increasing.*

The minimax condition (A1) and (A2) of Corollary 3.3.1 is however included in the sufficient condition for minimaxity represented by the differential inequality (3.8), which allows $\phi(w)$ to decrease with increasing w. Therefore by using (3.8), which is equivalent to (3.6), a larger class of minimax generalized Bayes estimators is able to be constructed as follows.

Theorem 3.21. *Let $h(\lambda)$ be a bounded nonnegative function such that $\lambda h'(\lambda)/h(\lambda)$ can be decomposed as $h_1(\lambda) + h_2(\lambda)$ where $h_1(\lambda) \le H_1$ and is nondecreasing while $0 \le h_2(\lambda) \le H_2$ with $H_1 + 2H_2 \le p/2 + a - 3$. Then the generalized Bayes estimator δ_h is minimax.*

Needless to say, this theorem includes Proposition 3.20. We note that this theorem is equivalent to the result by Fourdrinier *et al.*(1998). For the proof of this theorem, the integration by parts (3.49) is essential. Although Fourdrinier *et al.*(1998) admitted (3.49) unconditionally, the boundedness of $h(\lambda)$ is indispensable for it.

proof. By substituting $g(x) = \nabla \log f_h(x)$ in (3.6), the sufficient condition for minimaxity is equivalent to

$$\frac{\Delta f_h(x)}{\|\nabla f_h(x)\|} - \frac{1}{2}\frac{\|\nabla f_h(x)\|}{f_h(x)} \le 0.$$

By (3.42) and (3.43), we rewrite the above inequality as

$$p - \|x\|^2 \frac{\int_0^1 \lambda^{p/2+2-a} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda}{\int_0^1 \lambda^{p/2+1-a} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda} \\ + \frac{\|x\|^2}{2} \frac{\int_0^1 \lambda^{p/2+1-a} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda}{\int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda} \geq 0. \quad (3.52)$$

We put $s = \|x\|^2/2$ and apply an integration by parts in the numerators of second and third term on the left-hand side of (3.52), similarly to (3.49). Therefore (3.52) becomes

$$a + p/2 - 3 + \left[\frac{2c_2 \exp(-s)}{\int_0^1 \lambda^{p/2+1-a} h(\lambda) \exp(-s\lambda) d\lambda} - \frac{c_2 \exp(-s)}{\int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-s\lambda) d\lambda} \right] \\ - 2 \frac{\int_0^1 \lambda^{p/2+2-a} h'(\lambda) \exp(-s\lambda) d\lambda}{\int_0^1 \lambda^{p/2+1-a} h(\lambda) \exp(-s\lambda) d\lambda} + \frac{\int_0^1 \lambda^{p/2+1-a} h'(\lambda) \exp(-s\lambda) d\lambda}{\int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-s\lambda) d\lambda} \geq 0. \quad (3.53)$$

Clearly the term in bracket in (3.53) is nonnegative since the denominator of the second term is larger than that of the first. Putting, for fixed s,

$$f_k(\lambda) = \frac{\lambda^k g(\lambda) \exp(-s\lambda)}{\int_0^1 \lambda^k g(\lambda) \exp(-s\lambda) d\lambda},$$

we have a sufficient condition for minimaxity of the Bayes estimator δ_h as

$$a + p/2 - 3 - 2E_{p/2+1-a}\left(\frac{\lambda h'(\lambda)}{h(\lambda)}\right) + E_{p/2}\left(\frac{\lambda h'(\lambda)}{h(\lambda)}\right) \geq 0, \quad (3.54)$$

where E_k denotes expectation with respect to the density $f_k(\lambda)$. Let $\lambda h'(\lambda)/h(\lambda) = h_1(\lambda) + h_2(\lambda)$ where $h_1 \leq H_1$ and is nonincreasing and $0 \leq h_2 \leq H_2$. Noting that h_1 is nonincreasing and bounded by H_1, by the correlation inequality we have

$$2E_{p/2+1-a}(h_1) - E_{p/2-a}(h_1) \leq E_{p/2-a}(h_1) \leq H_1.$$

Moreover we have $2E_{p/2+1-a}(h_2) - E_{p/2-a}(h_2) \leq 2H_2$. Hence the left-hand side of (3.54) is bounded below by $a + p/2 - 3 - H_1 - 2H_2$, which in turn is nonnegative by the hypothesis of the theorem. This completes the proof. □

In the case of part *2* of Theorem 3.18, that is, $\lambda h'(\lambda)/h(\lambda)$ is monotone nondecreasing, we have the following by putting $h_1(\lambda) = 0$ on Theorem 3.21.

Corollary 3.21.1. *Assume that $h(\lambda)$ is a nonnegative bounded function such that $\lambda h'(\lambda)/h(\lambda)$ is monotone nondecreasing and that $3 - p/2 < a < 1 + p/2$ and that $h'(1) \le c_2(a + p/2 - 3)/2$. Then the generalized Bayes estimator δ_h is minimax and $\phi_h(w)$ has one extremum.*

Example 3.22. We consider the case of

$$h(\lambda) = (1-\lambda)^b \exp\left(c\frac{\lambda}{1-\lambda} + d\lambda\right).$$

Fourdrinier *et al.*(1998) considered the case of $d = 0$ in the above. Following results are the correct and extended version of Section 4.3 in Fourdrinier *et al.*(1998).

We easily calculate as

$$\begin{aligned} \lambda\frac{h'(\lambda)}{h(\lambda)} &= -b\frac{\lambda}{1-\lambda} + c\frac{\lambda}{(1-\lambda)^2} + d\lambda \\ &= -(b-c)\frac{\lambda}{1-\lambda} + c\left(\frac{\lambda}{1-\lambda}\right)^2 + d\lambda. \end{aligned} \tag{3.55}$$

If $3 - p/2 \le a < p/2 + 1$, $c \le 0$, $b \ge c$ and $d \le 0$, the generalized Bayes estimator is minimax by Proposition 3.20. If $3 - p/2 < a < p/2 + 1$, $b = c = 0$ and $0 < d \le (p/2 - 3 + a)/2$, the generalized Bayes estimator is minimax by Corollary 3.21.1. Finally we consider the case $3 - p/2 < a < p/2 + 1$, $c < 0$, $b < c$ and $d = 0$. Putting

$$h_1(\lambda) = \left(-(b-c)\frac{\lambda}{1-\lambda} + c\left(\frac{\lambda}{1-\lambda}\right)^2\right) I_{[\lambda/(1-\lambda) \ge (b-c)/(2c)]}(\lambda)$$

and

$$h_2(\lambda) = \left(-(b-c)\frac{\lambda}{1-\lambda} + c\left(\frac{\lambda}{1-\lambda}\right)^2\right) I_{[\lambda/(1-\lambda) < (b-c)/(2c)]}(\lambda),$$

we easily see that $h_1(\lambda)$ is monotone decreasing and $h_2(\lambda)$ is monotone increasing and that $h_1(\lambda)$ and $h_2(\lambda)$ are less than $-(b-c)^2/(4c)$ for any λ. Therefore by Theorem 3.21, if $a + p/2 - 3 + 3(b-c)^2/(4c) \ge 0$, the generalized Bayes estimator is minimax.

Example 3.23. We consider the case of $h(\lambda) = \sum_{i=0}^{n} a_i \lambda^{b_i}$, where $a_i > 0$ for every i, $b_0 = 0$, $b_i < b_j$ if $i < j$ and n is either finite or infinite. If the termwise differentiation is allowed and $\sum_{i=0}^{n} a_i b_i$ exists (Needless to say, these assumptions are satisfied unconditionally in the case of finite n), we have

$$\lambda h'(\lambda)/h(\lambda) = \frac{\sum_{i=0}^{n} a_i b_i \lambda^{b_i}}{\sum_{i=0}^{n} a_i \lambda^{b_i}},$$

which is monotone increasing in λ by Lemma 3.24 in the below. Therefore the generalized Bayes estimator is minimax if $3 - p/2 < a < p/2 + 1$ and $\sum_{i=0}^{n} a_i b_i / \sum_{i=0}^{n} a_i \leq (a + p/2 - 3)/2$.

A following lemma is a generalization of Lehmann(1986)'s.

Lemma 3.24. *Let* $g(y) = \left(\sum_{i=0}^{n} c_i y^{b_i}\right) / \left(\sum_{i=0}^{n} a_i y^{b_i}\right)$ *where* a_i *and* c_i *are positive for every* i, $b_0 = 0$, $b_i < b_j$ *if* $i < j$ *and* n *is either finite or infinite. We assume that* $\sum_{i=0}^{n} c_i y^{b_i}$ *and* $\sum_{i=0}^{n} a_i y^{b_i}$ *converge for all* $y > 0$ *and that the termwise differentiation is allowed. If the sequence* c_i/a_i *is monotone increasing(decreasing), then* $g(y)$ *is monotone increasing(decreasing) in* y.

proof. We have

$$\begin{aligned} g'(y) &= \left(\sum_{i=0}^{n} a_i y^{b_i}\right)^{-2} y^{-1} \left(\sum_{i=0}^{n} a_i y^{b_i} \sum_{i=0}^{n} b_i c_i y^{b_i} - \sum_{i=0}^{n} c_i y^{b_i} \sum_{i=0}^{n} b_i a_i y^{b_i}\right) \\ &= \left(\sum_{i=0}^{n} a_i y^{b_i}\right)^{-1} y^{-1} \Big(E(\{c_i/a_i\} b_i) - E(c_i/a_i) E(b_i) \Big), \end{aligned}$$

where E denotes expectation with respect to the probability function $a_i y^{b_i} / \sum_{i=0}^{n} a_i y^{b_i}$, $i = 0, \ldots, n$. The correlation inequality guarantees the result. □

3.3.3 Alam-Maruyama's type generalized Bayes estimator

Although Theorem 3.21 is very powerful and easily enables the construction of minimax Bayes estimators, there exist two famous classes of generalized Bayes estimators to which Theorem 3.21 is not able to be applied. One is the generalized Bayes estimator δ_h for $h(\lambda) = (1-\lambda)^b$ with $-1 < b < 0$, which is obviously not bounded and is considered in this subsection. The other is Stein(1973)'s Bayes estimator, which is considered in the next subsection.

The generalized Bayes estimator in the case of $h(\lambda) = (1-\lambda)^b$ with $-1 < b < 0$ was originally investigated by Alam(1973) and was developed by Maruyama(1998). As a matter of fact, Alam(1973) considered the prior distribution whose density is proportional to $\|\theta\|^{2\alpha}$ with $\alpha < 1 - p/2$. It is noted that scale mixtures of multivariate normal distributions which have a density (3.40) for $h(\lambda) = (1-\lambda)^b$ with $a = b+2$ is equivalent to $\|\theta\|^{2(b+1-p/2)}$ since we can calculate as

$$
\begin{aligned}
&\int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left(-\frac{\lambda}{2(1-\lambda)}\|\theta\|^2\right) \lambda^{-a}(1-\lambda)^b d\lambda \\
&= \int_0^\infty t^{p/2-a}(t+1)^{a-b-2} \exp(-t\|\theta\|^2/2) dt \\
&\propto \|\theta\|^{2(b+1-p/2)}.
\end{aligned}
$$

Results in this subsection are essentially equivalent to Maruyama(1998,2000a)'s although these are more simplificated than Maruyama(1998)'s . The generalized Bayes estimator is written as $\delta_b(X) = (1 - \phi_b(\|X\|^2)/\|X\|^2)X$, where

$$
\phi_b(w) = w \frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^b \exp(-w\lambda/2) d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)^b \exp(-w\lambda/2) d\lambda}.
$$

To investigate the properties of the behavior of $\phi_b(w)$ and the minimaxity of $\delta_b(X)$, the representation through the confluent hypergeometric function

$$
M(a, b, x) = 1 + ax/b + \cdots + (a)_n x^n/(b)_n n! + \cdots ,
$$

where $(a)_n = a \cdot (a+1) \cdots (a+n-1)$ for $n \geq 1$ and $(a)_0 = 1$ is quite useful. From the formulas (13.1.27), (13.2.1), (13.4.3), (13.4.4) and (13.4.8) given in Abramowitz and Stegun(1964), we have the following relations which will be used in the sequel,

$$
\Gamma(b-a)\Gamma(a)(\Gamma(b))^{-1} M(a, b, x) = \int_0^1 e^{xt} t^{a-1}(1-t)^{b-a-1} dt, \quad \text{for} \quad b > a > 0, \quad (3.56)
$$

$$M(a,b,x) = e^x M(b-a,b,-x), \tag{3.57}$$

$$bM(a,b,x) - bM(a-1,b,x) - xM(a,b+1,x) = 0, \tag{3.58}$$

$$(1+a-b)M(a,b,x) - aM(a+1,b,x) + (b-1)M(a,b-1,x) = 0 \tag{3.59}$$

and

$$\frac{d}{dx}M(a,b,x) = \frac{a}{b}M(a+1,b+1,x). \tag{3.60}$$

From (3.56), (3.57) and (3.58), we obtain

$$\begin{aligned} \phi_b(w) &= w\frac{p/2-a+1}{p/2-a+b+2}\frac{M(b+1,p/2-a+b+3,w/2)}{M(b+1,p/2-a+b+2,w/2)} \\ &= (p-2a+2)(1+\varphi(w)), \end{aligned}$$

where $\varphi(w) = -M(b,c,w/2)/M(b+1,c,w/2)$ and $c = p/2-a+b+2$. Noting that

$$\psi(w) = \frac{-\frac{d}{dw}M(b,c,w/2)}{\frac{d}{dw}M(b+1,c,w/2)} = -\frac{b}{b+1}\frac{M(b+1,c+1,w/2)}{M(b+2,c+1,w/2)},$$

is strictly decreasing in w from $\psi(0) = -b/(b+1)$ to $\psi(\infty) = 0$ by Lemma 3.24, we can prove the following result of the properties of the behavior of $\phi_b(w)$, similarly to part *2* of Theorem 3.18.

Lemma 3.25 (Maruyama(1998)). *$\phi_b(w)$ with $-1 < b < 0$ is increasing from the origin to a certain point and is decreasing from the point.*

The result of minimaxity of $\delta_b(X)$ is following.

Theorem 3.26 (Maruyama(1998,2000a)). *The generalized Bayes estimator $\delta_b(X)$ with $3-p/2 < a < 1+p/2$ and $b \geq -(a+p/2-3)/(3p/2+1-a)$ is minimax.*

proof. Similarly to the proof of Theorem 3.21, the sufficient condition for minimaxity is written as

$$p - \|x\|^2 \frac{\int_0^1 \lambda^{p/2-a+2}(1-\lambda)^b \exp(-\|x\|^2\lambda/2)d\lambda}{\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^b \exp(-\|x\|^2\lambda/2)d\lambda} + \frac{1}{2}\|x\|^2 \frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^b \exp(-\|x\|^2\lambda/2)d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)^b \exp(-\|x\|^2\lambda/2)d\lambda} \geq 0. \quad (3.61)$$

Putting $s = \|x\|^2/2$ and using the relations (3.58) and (3.59), we rewrite the inequality (3.61) as

$$p/2 + a - 3 - 2b + 2c\frac{M(b,c,s)}{M(b+1,c+1,s)} - (p/2+1-a)\frac{M(b,c,s)}{M(b+1,c,s)} \geq 0. \quad (3.62)$$

For s which satisfies $M(b,c,s) \geq 0$, the left-hand side of (3.62) is bounded below by $p/2 + a - 3 - 2b$ since clearly we have $c \geq p/2 + 1 - a$ and $M(b+1,c+1,s)^{-1} \geq M(b+1,c,s)^{-1}$. For s which satisfies $M(b,c,s) < 0$, the left-hand side of (3.62) is bounded below by

$$p/2 + a - 3 - 2b + 2b\frac{c+1}{b+1} = \frac{p/2 + a - 3 + (3p/2+1-a)b}{b+1}, \quad (3.63)$$

since the inequality

$$(b+1)cM(b,c,s) - b(c+1)M(b+1,c+1,s) \geq 0$$

is satisfied for $-1 < b < 0$. The above inequality is shown by verifying that the coefficients of s^n for every $n \geq 0$ are nonnegative. By the hypothesis of the theorem, the left hand side of (3.62) is nonnegative and this completes the proof. □

Remark 3.27. To investigate the properties of the behavior of $\phi_b(w)$ and the minimaxity of $\delta_b(X)$, the relation (3.58) substitutes for the integration by parts (3.49) in Section 3.3.2 where $h(\lambda)$ is bounded.

Theorem 3.26 has been already proved by Maruyama(1998), in which he however made a mistake in evaluation of a certain inequality. We provide the correct version of the proof in the following.

proof. Putting

$$S(a,b,w) = \left(4\frac{d}{dw}\phi_b(w) + \phi_b(w)\left(2(p-2) - \phi_b(w)\right)/w \right),$$

we can write the sufficient condition for minimaxity as $S(a,b,w) \geq 0$ for every $w \geq 0$. Putting $c = p/2 - a + b + 2$ and $s = w/2$ and using the relation (3.60), we obtain

$$\begin{aligned} S(a,b,w) &= \frac{2(c-b-1)}{cM^2(b+1,c,s)} \Big((p-c-b-1)M(b+1,c+1,s)M(b+1,c,s) \\ &\quad +(c-b-1)M(b+1,c+1,s)M(b,c,s) \\ &\quad +2(b+1)M(b+2,c+1,s)M(b,c,s) \Big). \end{aligned} \tag{3.64}$$

The following calculation is mainly based on the method of Alam(1973). Using the series expansion for the confluent hypergeometric function, we have

$$\begin{aligned} &M(b+1,c+1,s)M(b+1,c,s) \\ &= \sum_{n=0}^{\infty} \frac{s^n}{n!} \sum_{\gamma=0}^{n} \binom{n}{\gamma} \frac{(b+1)_\gamma (b+1)_{n-\gamma}}{(c+1)_\gamma (c)_{n-\gamma}} \\ &= \sum_{n=0}^{\infty} \frac{s^n}{n!} \sum_{\gamma=0}^{n} \binom{n}{\gamma} \frac{(b+1)_\gamma (b+1)_{n-\gamma} c}{(c)_\gamma (c)_{n-\gamma} (c+n-\gamma)} \\ &= \sum_{n=0}^{\infty} \frac{s^n}{n!} \sum_{\gamma=0}^{[n/2]} \binom{n}{\gamma} \frac{(b+1)_\gamma (b+1)_{n-\gamma}}{(c)_\gamma (c)_{n-\gamma}} c q_\gamma \left[\frac{1}{c+\gamma} + \frac{1}{c+n-\gamma} \right], \end{aligned} \tag{3.65}$$

where $[n/2]$ denotes the largest integer less than or equal to $n/2$, $q_\gamma = 1$ for $\gamma < n/2$ and $q_\gamma = 1/2$ for $\gamma = n/2$. Similarly to (3.65), we have

$$\begin{aligned} &M(b+1,c+1,s)M(b,c,s) \\ &= \sum_{n=0}^{\infty} \frac{s^n}{n!} \sum_{\gamma=0}^{[n/2]} \binom{n}{\gamma} \frac{(b+1)_\gamma (b+1)_{n-\gamma}}{(c)_\gamma (c)_{n-\gamma}} c q_\gamma \\ &\qquad \times b \left[\frac{1}{(c+\gamma)(b+n-\gamma)} + \frac{1}{(c+n-\gamma)(b+\gamma)} \right] \end{aligned} \tag{3.66}$$

and

$$M(b+2,c+1,s)M(b,c,s)$$
$$= \sum_{n=0}^{\infty} \frac{s^n}{n!} \sum_{\gamma=0}^{[n/2]} \binom{n}{\gamma} \frac{(b+1)_\gamma (b+1)_{n-\gamma}}{(c)_\gamma (c)_{n-\gamma}} c q_\gamma$$
$$\times \frac{b}{b+1} \left[\frac{b+1+\gamma}{(c+\gamma)(b+n-\gamma)} + \frac{b+1+n-\gamma}{(c+n-\gamma)(b+\gamma)} \right]. \quad (3.67)$$

Combining (3.64), (3.65), (3.66) and (3.67), we can see that $S(a,b,w)$ is nonnegative if

$$\begin{aligned} T(c,b,n,\gamma) &= (c-b-1)b \left[\frac{1}{(c+\gamma)(b+n-\gamma)} + \frac{1}{(c+n-\gamma)(b+\gamma)} \right] \\ &\quad +2(b+1)\frac{b}{b+1} \left[\frac{b+1+\gamma}{(c+\gamma)(b+n-\gamma)} + \frac{b+1+n-\gamma}{(c+n-\gamma)(b+\gamma)} \right] \\ &\quad +(p-c-b-1) \left[\frac{1}{c+\gamma} + \frac{1}{c+n-\gamma} \right] \end{aligned} \quad (3.68)$$

is nonnegative for each $\gamma = 0, 1, \cdots, [n/2]$ and for each $n = 0, 1, \cdots$. Clearly $n = 0$ implies $\gamma = 0$ and we have $T(c,b,0,0) = 2p/c > 0$. Thus we deal with the case of $n \geq 1$. Note that $n \geq 1$ implies $n - \gamma \geq 1$. We arrange the right-hand side of (3.68) as

$$\begin{aligned} T(c,b,n,\gamma) &= (p-c-b-1)\frac{2c+n}{(c+\gamma)(c+n-\gamma)} \\ &\quad +b \left[\frac{c+b+1+2\gamma}{(c+\gamma)(b+n-\gamma)} + \frac{c+b+1+2n-2\gamma}{(c+n-\gamma)(b+\gamma)} \right]. \end{aligned} \quad (3.69)$$

The quantity inside the braces on the right-hand side of (3.69) can be evaluated as

$$\begin{aligned} &\left[\frac{c+b+1+2\gamma}{(c+\gamma)(b+n-\gamma)} + \frac{c+b+1+2n-2\gamma}{(c+n-\gamma)(b+\gamma)} \right] \\ &= \frac{1}{(c+\gamma)(c+n-\gamma)} \left[(c+b+1+2\gamma)\frac{c+n-\gamma}{b+n-\gamma} + (c+b+1+2n-2\gamma)\frac{c+\gamma}{b+\gamma} \right] \\ &\leq \frac{1}{(c+\gamma)(c+n-\gamma)} \left[(c+b+1+2\gamma)\frac{c+1}{b+1} + (c+b+1+2n-2\gamma)\frac{c+1}{b+1} \right] \\ &= 2\frac{c+1}{b+1}\frac{2c+n}{(c+\gamma)(c+n-\gamma)}\frac{c+b+1+n}{2c+n} \\ &\leq 2\frac{c+1}{b+1}\frac{2c+n}{(c+\gamma)(c+n-\gamma)}. \end{aligned} \quad (3.70)$$

Hence we have

$$T(c,b,n,\gamma) \geq \frac{2c+n}{(c+\gamma)(c+n-\gamma)}\left[p-c-b-1+2\frac{b(c+1)}{b+1}\right],$$

which is nonnegative if $p-c-b-1+2b(c+1)/(b+1) \geq 0$, that is, $(3p/2+1-a)b+p/2-3+a \geq 0$. In this case, $S(a,b,w)$ is also nonnegative and this completes the proof. □

Remark 3.28. In Section 3.3.2 and 3.3.3, we see that all classes of minimax generalized Bayes estimators given in the hitherto researches, that is, Strawderman(1971), Alam(1973), Berger(1976b), Faith(1978), Maruyama(1998,2000a) and Fourdrinier *et al.*(1998) are expressed as δ_h, although relations among classes above have not been clear. Hence we succeed in treating uniformly classes above as the following table.

Table 3.1: All classes of minimax Bayes estimators given in the hitherto researches

Author	The sufficient condition for minimaxity on $h(\lambda)$ and a	$\phi_h(w)$
Strawderman-Berger	$h \equiv 1$ $3-p/2 \leq a < p/2+1$	$\nearrow$
Faith	$h(\lambda)$ is bounded $\lambda h'(\lambda)/h(\lambda)$ is monotone nondecreasing $3-p/2 \leq a < p/2+1$	$\nearrow$
Fourdrinier *et al.*	$h(\lambda)$ is bounded $\lambda h'(\lambda)/h(\lambda)$ can be decomposed as $h_1(\lambda)+h_2(\lambda)$ $h_1(\lambda) \leq H_1$ and is monotone decreasing $0 \leq h_2(\lambda) \leq H_2$ $H_1+2H_2 \leq p/2+a-3$	$\nearrow$ or $\nearrow\searrow$ or $\cdots$
Alam	$h(\lambda)=(1-\lambda)^b$ with $b<0$ $a=b+2$	$\nearrow\searrow$
Maruyama	$h(\lambda)=(1-\lambda)^b$ with $b<0$ $3-p/2 < a < p/2+1$ $-(a+p/2-3)/(3p/2+1-a) \leq b < 0$	$\nearrow\searrow$

3.3.4 Generalization of Stein's Bayes estimator

In this subsection, we pay attention to the following Stein(1973)'s simple idea and provide it with the warrant for statistical decision theory. Stein(1973) suggested that the generalized Bayes estimator with respect to the prior distribution: the weighted sum of that determined by the density $\|\theta\|^{2-p}$, which corresponds to (3.40) with $a = 2$ and $h(\lambda) \equiv 1$, and a measure concentrated at the origin may dominate the James-Stein estimator. Since Kubokawa(1991) afterward showed that the generalized Bayes estimator with respect to the density $\|\theta\|^{2-p}$ dominates the James-Stein estimator, Stein's suggestion is to the point in the case where the ratio of a measure concentrated at the origin is very small. Efron and Morris(1976) expressed Stein(1973)'s estimator explicitly. As for the decision-theoretic properties of the estimator, admissibility is easily able to be checked by using Brown(1971)'s theorem. On the other hand, even minimaxity, to say nothing of dominance over the James-Stein estimator has not been proved yet. (Hara(1997) tried to prove the minimaxity of Stein's estimator. Unfortunately his proof is incorrect.)

In this subsection, we investigate minimaxity and admissibility of the generalized Bayes estimator with respect to the prior distribution which includes Stein's.

We consider the following prior distribution: scale mixtures of multivariate normal distribution whose density is proportional to

$$\int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left(-\frac{\lambda}{2(1-\lambda)}\|\theta\|^2\right) \lambda^{-a} h(\lambda) d\lambda,$$

which is equivalent to the density function considered in Section 3.3.1, is chosen with probability $\beta/(1+\beta)$ and a measure concentrated at the origin is chosen with probability $1/(1+\beta)$ for $\beta > 0$. This prior distribution is interpreted as the one whose density is (3.40) with

$$h(\lambda) = \beta h(\lambda) + \delta(\lambda - 1), \tag{3.71}$$

where $\delta(\cdot)$ is the Dirac delta function. Similarly to (3.41), the marginal density function of X is calculated as

$$f_\beta(x) = \beta \int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda + \exp(-\|x\|^2/2).$$

Since the generalized Bayes estimator is written as $X + \nabla \log f_\beta(X)$, $f_\beta(x)$ exists for all x and the generalized Bayes estimator is well defined if and only if the integral $\int_0^1 \lambda^{p/2-a} h(\lambda) d\lambda$ is finite. Noting that

$$\nabla f_\beta(x) = -\left(\int_0^1 \lambda^{p/2-a+1} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda + \beta^{-1} \exp(-\|x\|^2/2) \right) x, \qquad (3.72)$$

we have the generalized Bayes estimator

$$\delta_\beta(X) = (1 - \phi_\beta(\|X\|^2)/\|X\|^2)X,$$

where

$$\phi_\beta(w) = w \frac{\beta \int_0^1 \lambda^{p/2-a+1} h(\lambda) \exp(-w\lambda/2) d\lambda + \exp(-w/2)}{\beta \int_0^1 \lambda^{p/2-a} h(\lambda) \exp(-w\lambda/2) d\lambda + \exp(-w/2)}. \qquad (3.73)$$

Results in Section 3.3.1 are clearly succeeded regardless of β.

Theorem 3.29. *$\delta_\beta(X)$ is admissible if and only if $a \leq 2$.*

Lemma 3.30. $\lim_{w\to\infty} \phi_\beta(w) = p - 2a + 2$, *which is not depend upon $h(\lambda)$ and β.*

Before considering the minimaxity of $\delta_\beta(X)$, we furthermore investigate the properties of the behavior of $\phi_\beta(w)$.

Theorem 3.31.

1. *Assume that $\lambda h'(\lambda)/h(\lambda)$ with bounded $h(\lambda)$ is monotone nondecreasing or that $h(\lambda) = (1-\lambda)^b$ with $-1 < b < 0$. Then $\phi_\beta(w)$ for every $\beta > 0$ is increasing from the origin to a certain point and is decreasing from the point.*

2. *Assume that $\lambda h'(\lambda)/h(\lambda)$ with bounded $h(\lambda)$ is monotone nonincreasing and that $h(\lambda)$ is not constant. Then there exists a constant β_1 such that the following are satisfied.*

   (a) *If $\beta \geq \beta_1$, $\phi_\beta(w)$ is monotone nondecreasing.*

   (b) *If $\beta < \beta_1$, the function $\phi_\beta(w)$ is increasing from the origin to a certain point w_1 and is decreasing from the point w_1 to the other point w_2 and is increasing from the point w_2. Namely, $\phi_\beta(w)$ has two extreme points.*

We cannot derive the value β_1 analytically although we see that $\beta_1 < \beta_2$ where

$$\begin{aligned}\beta_2 &= \left(\int_0^1 \lambda^{p/2-a}(1-\lambda)h(\lambda)d\lambda\right)^{-1} \\ &\quad \cdot \left(\frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^2 h'(\lambda)d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)^2 h(\lambda)d\lambda} - \frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda)h'(\lambda)d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)h(\lambda)d\lambda}\right)^{-1}.\end{aligned}$$

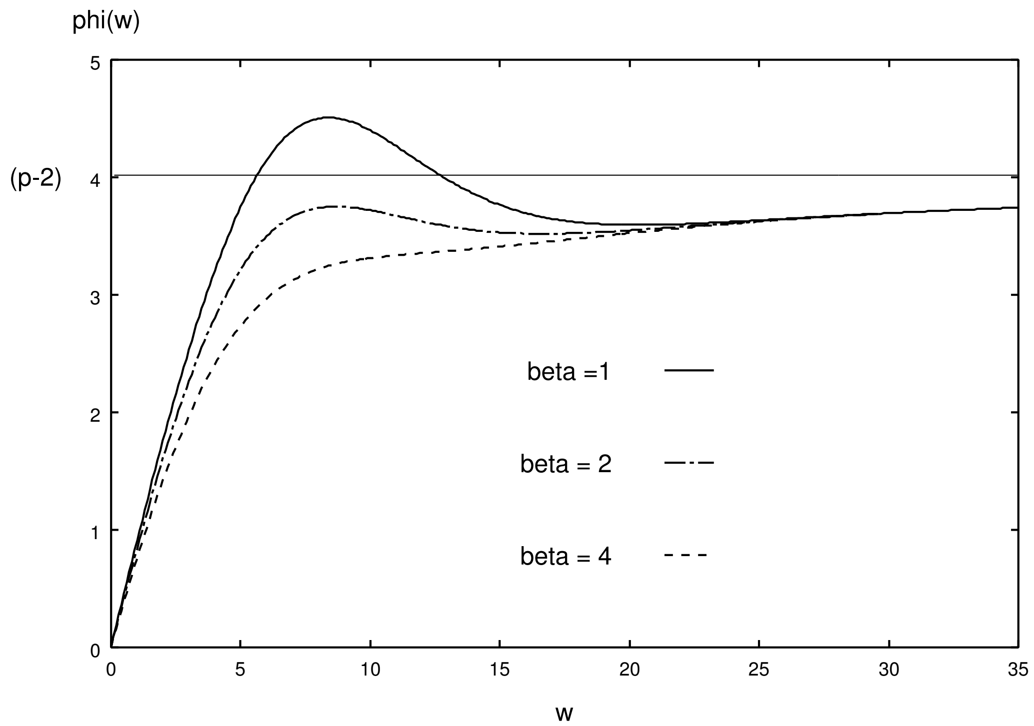

Figure 3.2 : The behavior of $\phi_\beta(w)$ for $p = 6$, $a = 2$ and $h(\lambda) = 1 - \lambda$ when $\beta = 1,\ 2,\ 4$.

Figure 3.2 gives the behavior of $\phi_\beta(w)$ with $a = 2$ and $h(\lambda) = 1 - \lambda$, in the case of $\beta = 1,\ 2,\ 4$, for $p = 6$. We can guess that there exists β_1 in this case, between 2 and 4.

proof. First we consider the case of $h(\lambda) = (1-\lambda)^b$ for $-1 < b < 0$. Using the relation (3.58), we rewrite $\phi_\beta(w)$ as

$$\phi_\beta(w) = p - 2a + 2 + \frac{w - (p-2a+2) - \beta d(p-2a+2)M(b,c,w/2)}{1 + \beta d M(b+1,c,w/2)},$$

where $c = p/2 - a + b + 2$ and $d = \Gamma(p/2 + 1 - a)\Gamma(b + 1)/\Gamma(c)$. Similarly to the proof of Theorem 3.18, we have the result.

Next we consider the case of bounded $h(\lambda)$. Applying an integration by parts to the numerator of the right-hand side of (3.73), we rewrite $\phi_\beta(w)$ as

$$\begin{aligned}\phi_\beta(w) &= p - 2a + 2 + \frac{2\beta \int_0^1 \lambda^{p/2-a+1} h'(\lambda) \exp((1-\lambda)w/2) d\lambda + w}{1 + \beta \int_0^1 \lambda^{p/2-a} h(\lambda) \exp((1-\lambda)w/2) d\lambda} \\ &\quad - \frac{2\beta c_2 + (p - 2a + 2)}{1 + \beta \int_0^1 \lambda^{p/2-a} h(\lambda) \exp((1-\lambda)w/2) d\lambda}.\end{aligned}$$

Putting

$$\varphi_1(w) = \beta \int_0^1 \lambda^{p/2-a+1} h'(\lambda) \exp((1-\lambda)w/2) d\lambda + w/2 - \beta c_2 - (p/2 - a + 1),$$

$$\varphi_2(w) = 1 + \beta \int_0^1 \lambda^{p/2-a} h(\lambda) \exp((1-\lambda)w/2) d\lambda,$$

$\varphi(w) = \varphi_1(w)/\varphi_2(w)$, and

$$\psi(w) = \frac{\varphi_1'(w)}{\varphi_2'(w)} = \frac{\beta \int_0^1 \lambda^{p/2-a+1}(1-\lambda) h'(\lambda) \exp((1-\lambda)w/2) d\lambda + 1}{\beta \int_0^1 \lambda^{p/2-a}(1-\lambda) h(\lambda) \exp((1-\lambda)w/2) d\lambda},$$

we have $\phi_\beta(w) = p - 2a + 2 + 2\varphi(w)$ and $\varphi'(w) = \varphi_2'(w)(\varphi_2(w))^{-1} (\psi(w) - \varphi(w))$. Moreover putting

$$\rho(w) = \frac{\varphi_1''(w)}{\varphi_2''(w)} = \frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^2 h'(\lambda) \exp((1-\lambda)w/2) d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)^2 h(\lambda) \exp((1-\lambda)w/2) d\lambda},$$

we have $\psi'(w) = \varphi_2''(w)(\varphi_2'(w))^{-1} (\rho(w) - \psi(w))$. Furthermore it is noted that $\varphi(0) = -p/2 + a - 1$,

$$\begin{aligned}\psi(0) &= \frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda) h'(\lambda) d\lambda + \beta^{-1}}{\int_0^1 \lambda^{p/2-a}(1-\lambda) h(\lambda) d\lambda} \\ &= -p/2 + a - 1 + \frac{\int_0^1 \lambda^{p/2-a+1} h(\lambda) d\lambda + \beta^{-1}}{\int_0^1 \lambda^{p/2-a}(1-\lambda) h(\lambda) d\lambda},\end{aligned}$$

$$\rho(0) = \frac{\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^2 h'(\lambda) d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)^2 h(\lambda) d\lambda} \tag{3.74}$$

and

$$\lim_{w\to\infty} \varphi(w) = \lim_{w\to\infty} \psi(w) = \lim_{w\to\infty} \rho(w) = 0. \tag{3.75}$$

If $\lambda h'(\lambda)/h(\lambda)$ is monotone nondecreasing, by Lemma 3.19, $\psi(w)$ is monotone decreasing. Hence similarly to the proof of Theorem 3.18, we have the result.

Fillaly we deal with the case where $\lambda h'(\lambda)/h(\lambda)$ is monotone nonincreasing. By Lemma 3.19, $\rho(w)$ is monotone increasing. Therefore if $\rho(0) \le \psi(0)$, that is, $\beta \ge \beta_2$, similarly to the proof of Theorem 3.18, we see that $\psi(w)$ is monotone increasing, which implies that $\varphi(w)$ is monotone increasing. For $\beta < \beta_2$, $\psi(w)$ is decreasing from the origin to a certain point and is increasing from the point. For fixed $w(> p-2a+2)$, as β is smaller and smaller, there exists β such that $\phi_\beta(w)$ exceeds $p-2a+2$, that is, $\varphi(w)$ exceeds 0, which implies that $\varphi(w)$ is not monotone by (3.75). By continuity of $\phi_\beta(w)$ in β, there exists a constant β_1 such that for $\beta \ge \beta_1$, $\psi(w) \ge \varphi(w)$ for every w and for $\beta < \beta_1$, we have the point w_0 which satisfies that $\psi(w_0) < \varphi(w_0)$. Similarly to the proof of Theorem 3.18, we see that for $\beta < \beta_1$, there exist two points $w_1(< w_0 <)w_2$ which satisfies that $\psi(w) = \varphi(w)$. Moreover we see that for $0 \le w < w_1$, $\psi(w) > \varphi(w)$ is satisfied and for $w_1 < w < w_2$, $\psi(w) < \varphi(w)$ is satisfied and for $w > w_2$, $\psi(w) > \varphi(w)$ is satisfied. This completes the proof. □

The result of minimaxity of $\delta_\beta(X)$ is following.

Theorem 3.32. *$\delta_\beta(X)$ is minimax if $\beta \ge \beta^*$ where*

$$\begin{aligned} \beta^* &= \left(D\int_0^1 \lambda^{p/2-a+1}(1-\lambda)h(\lambda)d\lambda\right)^{-1} \\ &\quad \times \left(1+\sqrt{1+D\left(\int_0^1 \lambda^{p/2-a}h(\lambda)d\lambda\right)^{-1}\int_0^1 \lambda^{p/2-a+1}(1-\lambda)h(\lambda)d\lambda}\right) \end{aligned}$$

for $D > 0$ where $D = a + p/2 - 3 - H_1 - 2H_2$ under the assumption of Theorem 3.21: $(a+p/2-3+b(3p/2+1-a))(b+1)^{-1}$ in the case of $h(\lambda) = (1-\lambda)^b$ with $-1 < b < 0$.

By this theorem, we see that, for $h(\lambda)$ such that δ_h is shown to be minimax in Section 3.3.2 and 3.3.3, there exists a constant β^* such that for $\beta \geq \beta^*$, δ_β is also minimax.

proof. Similarly to the proof of Theorem 3.21, the sufficient condition for minimaxity is written as

$$p - \|x\|^2 \frac{\beta \int_0^1 \lambda^{p/2+2-a} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda + \exp(-\|x\|^2/2)}{\beta \int_0^1 \lambda^{p/2+1-a} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda + \exp(-\|x\|^2/2)} + \frac{\|x\|^2}{2} \frac{\beta \int_0^1 \lambda^{p/2+1-a} h(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda + \exp(-\|x\|^2/2)}{\beta \int_0^1 \lambda^{p/2-a} g(\lambda) \exp(-\|x\|^2\lambda/2) d\lambda + \exp(-\|x\|^2/2)} \geq 0.$$

Putting $s = \|x\|^2/2$ and

$$Z_k(s) = \exp(s) \int_0^1 \lambda^k h(\lambda) \exp(-s\lambda) d\lambda,$$

we can rewrite the above inequality as

$$\left(p - 2s \frac{Z_{p/2-a+2}(s)}{Z_{p/2-a+1}(s)} + s \frac{Z_{p/2-a+1}(s)}{Z_{p/2-a}(s)}\right) \beta^2 + \left(\frac{p}{Z_{p/2-a+1}(s)} + \frac{p}{Z_{p/2-a}(s)} + 2s \frac{Z_{p/2-a+1}(s) - Z_{p/2-a+2}(s)}{Z_{p/2-a}(s) Z_{p/2-a+1}(s)} - \frac{2s}{Z_{p/2-a+1}(s)}\right) \beta + \frac{p-s}{Z_{p/2-a}(s) Z_{p/2-a+1}(s)} \geq 0. \quad (3.76)$$

Noting that $Z_{p/2-a+1}(s) - Z_{p/2-a+2}(s) \geq 0$, the value β which satisfies the inequality

$$\left(p - 2s \frac{Z_{p/2-a+2}(s)}{Z_{p/2-a+1}(s)} + s \frac{Z_{p/2-a+1}(s)}{Z_{p/2-a}(s)}\right) \beta^2 - \frac{2s}{Z_{p/2-a+1}(s)} \beta - \frac{s}{Z_{p/2-a}(s) Z_{p/2-a+1}(s)} \geq 0 \quad (3.77)$$

also satisfies the inequality (3.76). Let

$$A(s) = p - 2s Z_{p/2-a+2}(s) Z_{p/2-a+1}(s)^{-1} + s Z_{p/2-a+1}(s) Z_{p/2-a}(s)^{-1},$$

$B(s) = s/Z_{p/2-a+1}(s)$ and $C(s) = s/(Z_{p/2-a+1}(s)Z_{p/2-a}(s))$. For the quadratic equation $A(s)\beta^2 - 2B(s)\beta - C(s) = 0$, we would like to bound the larger solution

$$\beta^*(s) = B(s)/A(s) + \sqrt{(B(s)/A(s))^2 + C(s)/A(s)}$$

above. For this, it is sufficient to bound $A(s)$ below and $B(s)$, $C(s)$ above.

First we deal with $A(s)$. Under the assumption of Theorem 3.21, $A(s)$ is bounded below by $a + p/2 - 3 - H_1 - 2H_2$. In the case of $h(\lambda) = (1-\lambda)^b$ with $-1 < b < 0$, by (3.63), $A(s)$ is bounded below by $(a + p/2 - 3 + b(3p/2 + 1 - a))(b+1)^{-1}$.

Next we can evaluate $B(s)$ as

$$\begin{aligned} B(s) = sZ_{p/2-a+1}(s)^{-1} &= s\left(\sum_{n=0}^{\infty}\frac{s^n}{n!}\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^n h(\lambda)d\lambda\right)^{-1} \\ &\leq \left(\int_0^1 \lambda^{p/2-a+1}(1-\lambda)h(\lambda)d\lambda\right)^{-1}. \end{aligned} \tag{3.78}$$

Moreover noting that for every $s \geq 0$, the inequality

$$\int_0^1 \lambda^{p/2-a}h(\lambda)d\lambda \leq \int_0^1 \lambda^{p/2-a}h(\lambda)\exp((1-\lambda)s)d\lambda$$

is satisfied, we have

$$C(s) \leq \left(\int_0^1 \lambda^{p/2-a+1}(1-\lambda)h(\lambda)d\lambda \int_0^1 \lambda^{p/2-a}h(\lambda)d\lambda\right)^{-1}. \tag{3.79}$$

Therefore we see that $\beta^*(s)$ is bounded above by β^* where

$$\begin{aligned} \beta^* &= \left(D\int_0^1 \lambda^{p/2-a+1}(1-\lambda)h(\lambda)d\lambda\right)^{-1} \\ &\quad\times\left(1 + \sqrt{1 + D\left(\int_0^1 \lambda^{p/2-a}h(\lambda)d\lambda\right)^{-1}\int_0^1 \lambda^{p/2-a+1}(1-\lambda)h(\lambda)d\lambda}\right) \end{aligned}$$

for $D > 0$ where $D = a + p/2 - 3 - H_1 - 2H_2$ under the assumption of Theorem 3.21: $(a + p/2 - 3 + b(3p/2 + 1 - a))(b+1)^{-1}$ in the case of $h(\lambda) = (1-\lambda)^b$ with $-1 < b < 0$. This completes the proof. □

Example 3.33. We consider the sufficient condition for minimaxity of Stein(1973)'s Bayes estimator. Since the prior density $\|\theta\|^{2-p}$ corresponds to the density (3.40) with $a = 2$ and $h(\lambda) \equiv 1$, we have

$$\beta^* = \frac{p(p+2)}{2(p-2)} \left(1 + \sqrt{1 + \frac{(p-2)^2}{p(p+2)}} \right).$$

Moreover we consider Stein(1973)'s Bayes estimator. As introduced in the next section, Kubokawa(1991) showed that the generalized Bayes estimator with respect to the density $\|\theta\|^{2-p}$ is admissible and dominates the James-Stein estimator. As far as we know, the estimator is the only one which has above two optimalities, that is, domination over the James-Stein estimator and admissibility. The sole drawback is that Kubokawa's estimator has the same risk of the James-Stein estimator at $\|\theta\| = 0$. We note that shrinkage estimators are derived by using the vague prior information that $\|\theta\|$ is close to 0. It goes without saying that we would like to get significant improvement of risk when the prior information is accurate. Therefore Stein's estimator with a sufficiently large β, has quite possibility of dominating the James-Stein estimator significantly at the origin, although we cannot have an analytic result about it.

3.4 Improvement upon the James-Stein estimator

The inadmissibility of the James-Stein estimator is stated in Section 3.1. Since the James-Stein estimator is permissible as shown in Section 3.2, for the construction of a class of estimators improving on the James-Stein estimator, it is necessary to overcome the limitation of the characteristics through the Stein identity. Kubokawa(1991, 1994) succeeded in overcoming it. Kubokawa(1991) applied the method of Brewster and Zidek(1974) in estimation of a normal variance to this estimation problem and showed that the estimator $\delta_K(X) = (1 - \phi_K(\|X\|^2)/\|X\|^2)X$ where

$$\phi_K(w) = w\frac{\int_0^1 \lambda^{p/2-1}\exp(-w\lambda/2)d\lambda}{\int_0^1 \lambda^{p/2-2}\exp(-w\lambda/2)d\lambda},$$

dominates the James-Stein estimator and is admissible. Moreover Kubokawa(1994) proposed a new approach to improving on the James-Stein estimator, by the idea to express a difference of risk functions through an integral. The main theorem of Kubokawa(1994) is stated in the following.

Theorem 3.34 (Kubokawa(1994)). *Assume that*

(B1) $\phi(w)$ *is nondecreasing and* $\lim_{w\to\infty}\phi(w) = p-2$

(B2) $\phi(w) \geq \phi_K(w)$.

Then $\delta_\phi(X)$ *of form* (3.7) *dominates the James-Stein estimator.*

proof. By the Stein identity, we have

$$R(\theta, \delta_\phi) = p + E\left[\frac{\phi^2(\|X\|^2) - 2(p-2)\phi(\|X\|^2)}{\|X\|^2} - 4\phi'(\|X\|^2)\right]$$

and

$$R(\theta, \delta^{JS}) = p - (p-2)^2 E\left[\frac{1}{\|X\|^2}\right].$$

By letting $\lim_{w\to\infty}\phi(w)=p-2$, the difference of risk functions can be written as

$$\begin{aligned}
&R(\theta,\delta^{JS})-R(\theta,\delta_\phi)\\
&=E\left[\frac{\phi(\infty)}{\|X\|^2}\left(\phi(\infty)-2(p-2)\right)-\frac{\phi(\|X\|^2)}{\|X\|^2}\left(\phi(\|X\|^2)-2(p-2)\right)\right]\\
&\quad+4E\left[\phi'(\|X\|^2)\right]\\
&=E\left[\int_1^\infty\frac{d}{dt}\left(\frac{\phi(t\|X\|^2)}{\|X\|^2}\left(\phi(t\|X\|^2)-2(p-2)\right)\right)dt\right]+4E\left[\phi'(\|X\|^2)\right]\\
&=2E\left[\int_1^\infty\left(\phi(t\|X\|^2)-(p-2)\right)\phi'(t\|X\|^2)dt\right]+4E\left[\phi'(\|X\|^2)\right]\\
&=2\int_0^\infty\int_1^\infty\left(\phi(tx)-(p-2)\right)\phi'(tx)f_p(x;\lambda)dtdx+4E\left[\phi'(\|X\|^2)\right], \qquad (3.80)
\end{aligned}$$

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 right-hand side of the equation of (3.80) as

$$\begin{aligned}
&\int_0^\infty\int_1^\infty\left(\phi(w)-(p-2)\right)\phi'(w)f_p(w/t;\lambda)t^{-1}dtdw\\
&=\int_0^\infty\int_0^w\left(\phi(w)-(p-2)\right)\phi'(w)f_p(y;\lambda)y^{-1}dydw.
\end{aligned}$$

Moreover by the inequality

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

where $f_p(y)=f_p(y;0)$, which can be shown by the correlation inequality, we can evaluate the risk difference as

$$\begin{aligned}
&R(\theta,\delta^{JS})-R(\theta,\delta_\phi)\\
&=2\int_0^\infty\phi'(w)\left(\left(\phi(w)-(p-2)\right)\int_0^w y^{-1}f_p(y,\lambda)dy+2f(w,\lambda)\right)dw\\
&\geq 2\int_0^\infty\phi'(w)\left(\phi(w)-\phi_0(w)\right)\left(\int_0^w y^{-1}f_p(y,\lambda)dy\right)dw, \qquad (3.81)
\end{aligned}$$

where

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

The assumptions of the theorem guarantee the left-hand side of the inequality (3.81) is nonnegative. Using an integration by parts, we easily verify that $\phi_0 = \phi_K$. This completes the proof. □

$\delta_K(X)$ and $\delta_+^{JS}(X)$ are included in this class. We note that a shrinkage estimators such as the James-Stein estimator is derived by using the vague prior information that $\|\theta\|^2$ is close to 0. It goes without saying that we would like to get the significant improvement of risk when the prior information is accurate. Though $\delta_K(X)$ is an admissible generalized Bayes estimator and thus analytic, it does not improve upon $\delta^{JS}(X)$ at $\|\theta\|^2 = 0$, which is easily verified by (3.81). On the other hand, $\delta_+^{JS}(X)$ improves on $\delta^{JS}(X)$ significantly at $\|\theta\|^2 = 0$, but it is not analytic and is thus inadmissible by complete class theorem. Therefore it is desirable to get an analytic estimator (which implies that it has some possibilities of being admissible) improving on $\delta^{JS}(X)$ especially at $\|\theta\|^2 = 0$.

Maruyama(1996) proposed an estimator having such a property:

$$\delta_\alpha(X) = \left(1 - \phi_\alpha(\|X\|^2)/\|X\|^2\right) X \tag{3.82}$$

where

$$\begin{aligned} \phi_\alpha(w) &= w \frac{\int_0^1 \lambda^{\alpha(p/2-1)} \exp(-w\alpha\lambda/2) d\lambda}{\int_0^1 \lambda^{\alpha(p/2-1)-1} \exp(-w\alpha\lambda/2) d\lambda} \\ &= w \frac{M(1, \alpha(p/2-1)+2, \alpha w/2)}{M(1, \alpha(p/2-1)+1, \alpha w/2)} \\ &= (p-2)\left(1 - \frac{1}{M(1, \alpha(p/2-1)+1, \alpha w/2)}\right). \end{aligned} \tag{3.83}$$

The above second and third equations are from formula (3.56), (3.57) and (3.58). Clearly we have $\delta_1(X) = \delta_K(X)$. Using the representation of $\phi_\alpha(w)$ as (3.83), we have the following result.

Theorem 3.35 (Maruyama(1996)). *$\delta_\alpha(X)$ dominates $\delta^{JS}(X)$ for $\alpha \geq 1$.*

proof. We shall verify that $\phi_\alpha(w)$ for $\alpha \geq 1$ satisfies (B1) and (B2) of Theorem 3.34. As $M(1, \alpha(p/2-1)+1, \alpha w/2)$ is increasing in w, $\phi_\alpha(w)$ is increasing in w. As $\lim_{w\to\infty} M(1, \alpha(p/2-1)+1, \alpha w/2) = \infty$, it is clear that $\lim_{w\to\infty} \phi_\alpha(w) = p-2$. In order to show that $\phi_\alpha(w) \geq \phi_K(w) = \phi_1(w)$ for $\alpha \geq 1$, we have only to check that $\phi_\alpha(w)$ is increasing in α, which is easily verified because the coefficient of each term of the right-hand side of the equation

$$\begin{aligned} M(1, \frac{\alpha}{2}(p-2)+1, \frac{\alpha w}{2}) &= 1 + \frac{1}{p-2+2/\alpha} w \\ &\quad + \frac{1}{(p-2+2/\alpha)(p-2+4/\alpha)} w^2 + \cdots \end{aligned} \tag{3.84}$$

is increasing in α. We have thus proved the theorem. □

Proposition 3.36 (Maruyama(1996)). *$\delta_\alpha(X)$ approaches the positive-part James-Stein estimator when α tends to infinity, that is,*

$$\lim_{\alpha\to\infty} \delta_\alpha(X) = \delta_+^{JS}(X).$$

proof. By the proof of Theorem 3.35, $M(1, \alpha(p-2)/2+1, \alpha w/2)$ is increasing in α, so that it converges to $\sum_{i=0}^{\infty}(w/(p-2))^i$ by the monotone convergence theorem. Considering two cases: $w < (\geq) p-2$, we obtain $\lim_{\alpha\to\infty} \phi_\alpha(w) = w$ if $w < p-2$; $= p-2$ otherwise. This completes the proof. □

Needless to say, we are interested in determining whether $\delta_\alpha(X)$ for $\alpha > 1$ is admissible or not. Before considering admissibility, we consider the permissibility of $\delta_\alpha(X)$. We have

$$\begin{aligned} \exp\left(-\frac{1}{2}\int_\epsilon^{\|x\|^2} \frac{\phi_\alpha(w)}{w} dw\right) &\propto \left(\int_0^1 \lambda^{\alpha(p/2-1)-1} \exp(-\frac{\alpha\|x\|^2}{2}\lambda) d\lambda\right)^{1/\alpha} \\ &\sim c\|x\|^{2-p}, \end{aligned}$$

for some constant c. The evaluation in the above is from Tauberian's theorem. Hence $\delta_\alpha(X)$ is permissible by Theorem 3.12. Moreover by Proposition 3.13, $\delta_\alpha(X)$ with $\alpha > 1$ is admissible if it is generalized Bayes, that is, there exists a measure π which satisfies

$$\int_{R^p} \exp(-\|\theta - x\|^2/2)\pi(d\theta) = \left(\int_0^1 \lambda^{\alpha(p/2-1)-1} \exp(-\alpha\|X\|^2\lambda/2)d\lambda\right)^{1/\alpha}.$$

By Theorem 3.10, if $\delta_\alpha(X)$ with $\alpha > 1$ is admissible, then it is necessary to be written as

$$\left(\int_0^1 \lambda^{\alpha(p/2-1)-1} \exp(\alpha(1-\lambda)y^2/2)d\lambda\right)^{1/\alpha} = \int_{-\infty}^{\infty} \exp(yt)\nu(dt) \tag{3.85}$$

for a certain measure ν.

Remark 3.37. Recently, Maruyama and Iwasaki(2000) considered the inverse Laplace transform of the left-hand side of equation (3.85)

$$f_\alpha(t) \equiv \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F(y)^{1/\alpha} \exp(yt)dy, \tag{3.86}$$

for suitable constants c, where

$$F(y) = \int_0^1 \lambda^{\alpha(p/2-1)-1} \exp(\alpha(1-\lambda)y^2/2)d\lambda$$

and investigated the properties of asymptotic behavior of $F(y)$, although the investigation is still incomplete. As a result, Maruyama and Iwasaki(2000) conjecture that $f_\alpha(t)$ with $\alpha > 1$ takes negative values for a sufficiently large t, which implies that there does not exist a measure ν by unicity of inverse Laplace transform and hence $\delta_\alpha(X)$ with $\alpha > 1$ is inadmissible.

$\delta_\alpha(X)$ can be represented as follows.

Definition 3.38. We say that an estimator is *fake-Bayes* if it is a (generalized) Bayes estimator for a different problem and for a suitable prior distribution.

Proposition 3.39. *$\delta_\alpha(X)$ is a fake-Bayes estimator for the following problem: X is a random variable having $N_p(\theta, \alpha^{-1}I_p)$, θ is estimated relative to the loss $\|\delta - \theta\|^2$ and the prior distribution which has a density function*

$$\int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left(-\frac{\alpha\lambda}{2(1-\lambda)}\|\theta\|^2\right) \lambda^{(\alpha-1)(p/2-1)-2} d\lambda$$

is taken.

proof. The marginal density function of X is written as

$$\begin{aligned} f(x) \;\propto\; & \int_0^1 \int_{R^p} \exp\left(-\alpha\frac{\|x-\theta\|^2}{2} - \frac{\alpha\lambda}{2(1-\lambda)}\|\theta\|^2\right) \\ & \times \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \lambda^{(\alpha-1)(p/2-1)-2} d\theta d\lambda \\ \propto\; & \int_0^1 \exp\left(-\frac{\alpha\lambda}{2}\|x\|^2\right) \lambda^{\alpha(p/2-1)-1} d\lambda. \end{aligned}$$

Noting that the generalized Bayes estimator is expressed as $X + \alpha^{-1}\nabla \log f(X)$ in this case, we have the estimator (3.82). □

As a conclusion of this section, we consider a necessary condition for dominance over the James-Stein estimator and the James-Stein positive-part estimator. For the risk of the estimator of the form (3.7),

$$R(\theta, \delta_\phi) = p + E\left[\frac{\phi(\|X\|^2)(\phi(\|X\|^2) - 2(p-2))}{\|X\|^2} - 4\phi'(\|X\|^2)\right], \tag{3.87}$$

by using a Laplace approximation on the expectation in (3.87), we have

$$\begin{aligned} & E\left[\frac{\phi(\|X\|^2)(\phi(\|X\|^2) - 2(p-2))}{\|X\|^2} - 4\phi'(\|X\|^2)\right] \\ & \approx \frac{\phi(\eta)(\phi(\eta) - 2(p-2))}{\eta} - 4\phi'(\eta), \end{aligned}$$

where $\eta = \|\theta\|^2$. By carefully working with the error terms in the Laplace approximation, Berger(1976a) showed that the error of approximation is $o(\eta^{-1})$, that is,

$$R(\theta, \delta_\phi) = p + \frac{\phi(\eta)(\phi(\eta) - 2(p-2))}{\eta} - 4\phi'(\eta) + o(\eta^{-1}). \tag{3.88}$$

Here it is noted that if δ_ϕ is minimax, it is necessary that $\lim_{\eta\to\infty} \eta\phi'(\eta) = 0$ is satisfied because $\lim_{\eta\to\infty} \eta\phi'(\eta) \neq 0$ implies that $\phi(\eta)$ is not bounded, so that we have

$$\begin{aligned}\lim_{\eta\to\infty} \eta(R(\theta,\delta_\phi) - p) &= \lim_{\eta\to\infty} (\phi(\eta)(\phi(\eta) - 2(p-2)) - 4\eta\phi'(\eta)) \\ &= +\infty.\end{aligned}$$

By the approximation $R(\theta, \delta^{JS}) = p - (p-2)^2/\eta + o(\eta^{-1})$, we have

$$R(\theta, \delta^{JS}) - R(\theta, \delta_\phi) = -\frac{(\phi(\eta) - (p-2))^2}{\eta} + o(\eta^{-1}), \tag{3.89}$$

for a minimax estimator δ_ϕ. Clearly we have the following result.

Theorem 3.40. *A necessary condition for an estimator δ_ϕ to dominate the James-Stein estimator is that $\lim_{\eta\to\infty} \phi(\eta) = p - 2$ and $\lim_{\eta\to\infty} \eta\phi'(\eta) = 0$.*

Next we consider a necessary condition for an admissible estimator δ_ϕ to dominate the James-Stein positive-part estimator. If an estimator δ_ϕ is admissible, $\phi(w) < w$ is clearly satisfied. Moreover note that $\delta_+^{JS} = \max(0, 1 - (p-2)/\|X\|^2)X$ is written as $(1 - \phi_+^{JS}(\|X\|^2)/\|X\|^2)X$, where $\phi_+^{JS}(w) = \min(w, p-2)$. If the inequality $\phi_+^{JS}(w) > \phi(w)$ for every $w \geq 0$ is satisfied, for the risks of these two estimators at $\|\theta\| = 0$, we have $R(0, \delta_+^{JS}(X)) > R(0, \delta_\phi(X))$, which implies that δ_ϕ does not dominate $\delta_+^{JS}(X)$. Therefore there exists $w_0(> p-2)$ such that $\phi(w_0)$ exceeds $p-2$. By noting Theorem 3.40, we have the following result.

Theorem 3.41. *A necessary condition for an admissible estimator δ_ϕ to dominate the James-Stein positive-part estimator is that $\lim_{\eta\to\infty} \phi(\eta) = p-2$ and $\lim_{\eta\to\infty} \eta\phi'(\eta) = 0$ and that there exists $w_0(> p-2)$ such that $\phi(w_0)$ exceeds $p-2$, which implies that $\phi(w)$ is not monotone.*

This theorem implies that an admissible estimator satisfying the sufficient condition for dominance over the James-Stein estimator, (B1) and (B2), could never improve the James-Stein positive-part estimator.

Chapter 4

Estimation of a mean vector of scale mixtures of multivariate normal distributions

4.1 Introduction and summary

Since Stein(1956), considerable effort has been given to improving upon the best equivariant estimator δ_0, which is minimax under the quadratic loss function as shown in Kiefer(1957). The theoretical questions were answered quite thoroughly by Brown(1966). He showed that in 3 or more dimensions δ_0 is inadmissible for an extremely wide variety of distributions and loss functions. It is however very difficult to explicitly give improved minimax estimators without restriction of distributions and of loss functions. For the explicit construction of improved minimax estimators in non-normal case, the estimation of a mean vector of spherically symmetric distributions under the quadratic loss function has been an important topic in this field. As a special case of spherically symmetric distributions, Strawderman(1974) introduced scale mixtures of multivariate normal distributions, which have a probability density function

$$f(\|x-\theta\|^2) = \int_0^\infty (2\pi)^{-p/2} v^{p/2} \exp(-\frac{\|x-\theta\|^2 v}{2}) G(dv),$$

where G is a known distribution function. Strawderman(1974) proposed a class of improved minimax estimators of the mean vector. Generally Berger(1975) and Brandwein and Strawderman(1978,1991) found and discussed a class of minimax estimators for spherically symmetric distributions. In view of statistical decision theory, however, we are interested in proposing a class of estimators satisfying not only minimaxity but also admissibility. So far as I know, explicit results of constructing admissible minimax estimators of a mean vector were restricted to the case of the multivariate normal distribution. For the normal case, see Strawderman(1971), Alam(1973), Berger(1976b), Fourdrinier *et al.*(1998), Maruyama(1998) and Chapter 3.

In this chapter, the problem of estimating the mean vector θ of scale mixtures of multivariate normal distributions, in the case that G has a probability density function, under the quadratic loss function is considered. For certain class of these distributions which includes at least multivariate-t distributions, admissible minimax estimators are derived.

4.2 The construction of the generalized Bayes estimators

In this section, we derive an generalized Bayes estimator with respect to the following prior distribution. Let the conditional distribution of θ given t, $0 < t < 1$, be normal with mean 0 and covariance matrix $t^{-1}(1-t)I_p$ and a density function of t is proportional to $t^{-a}(1-t)^b I_{(0,1)}(t)$ with $b > -1$. Therefore the density function of θ, $h_{a,b}(\theta)$ is proportional to

$$h_{a,b}(\theta) = \int_0^1 \left(\frac{t}{1-t}\right)^{p/2} \exp\left(-\frac{t}{2(1-t)}\|\theta\|^2\right) t^{-a}(1-t)^b dt.$$

It is noted that this prior distribution is proper for $a < 1$ and is improper for $a \geq 1$. Under the quadratic loss function, the generalized Bayes estimator is expressed as

$$\delta_{a,b}^{g}(x) = \frac{\int \theta f(\|x-\theta\|^2) h_{a,b}(\theta) d\theta}{\int f(\|x-\theta\|^2) h_{a,b}(\theta) d\theta}. \tag{4.1}$$

To make the following calculations easy, by change of variables formula, we rewrite $h_{a,b}(\theta)$ as, for every $v > 0$,

$$\begin{aligned} h_{a,b}(\theta) &= \int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \lambda^{-a}(1-\lambda)^b \\ &\quad \times \exp\left(-\frac{v\lambda\|\theta\|^2}{2(1-\lambda)}\right) v^{p/2-a+1}(1-\lambda+\lambda v)^{a-b-2} d\lambda. \end{aligned}$$

The integrals with respect to θ of both the denominator and numerator of right-hand side of (4.1) are calculated as follows:

$$\begin{aligned} \text{numerator} &= \int_{R^p} \theta \exp\left(-\frac{v\|x-\theta\|^2}{2} - \frac{v\lambda}{2(1-\lambda)}\|\theta\|^2\right) d\theta \\ &= \int_{R^p} \theta \exp\left(-\frac{\lambda v}{2}\|x\|^2\right) \exp\left(-\frac{v\|\theta-(1-\lambda)x\|^2}{2(1-\lambda)}\right) d\theta \\ &= C_1 \exp\left(-\frac{\lambda v}{2}\|x\|^2\right) \left(\frac{1-\lambda}{v}\right)^{p/2} (1-\lambda)x \end{aligned}$$

and

$$\begin{aligned} \text{denominator} &= \int_{R^p} \exp\left(-\frac{v\|x-\theta\|^2}{2} - \frac{v\lambda}{2(1-\lambda)}\|\theta\|^2\right) d\theta \\ &= C_1 \exp\left(-\frac{\lambda v}{2}\|x\|^2\right) \left(\frac{1-\lambda}{v}\right)^{p/2}, \end{aligned}$$

for a constant C_1. Hence we have the generalized Bayes estimator $\delta^g_{a,b}(X) = (1 - \phi^g_{a,b}(\|X\|^2)/\|X\|^2)X$, where

$$\phi^g_{a,b}(w) = w\frac{\int_0^1\int_0^\infty \lambda^{p/2-a+1}v^{p/2-a+1}(1-\lambda)^b(1-\lambda+v\lambda)^{a-b-2}\exp(-wv\lambda/2)g(v)dvd\lambda}{\int_0^1\int_0^\infty \lambda^{p/2-a}v^{p/2-a+1}(1-\lambda)^b(1-\lambda+v\lambda)^{a-b-2}\exp(-wv\lambda/2)g(v)dvd\lambda}.$$

Next we investigate some properties of the behavior of $\phi^g_{a,b}(w)$.

Theorem 4.1. *1. Assume that $g(s_1t_1)g(s_2t_2) \le g(s_1t_2)g(s_2t_1)$ for $s_1 \le s_2$ and $t_1 \le t_2$. Then $\phi^g_{a,b}(w)$ for $b-a+2>0$ and $b \ge 0$ is monotone nondecreasing.*

2. $\phi^g_{a,b}(w)$ for $b-a+2=0$ and $b\ge 0$ is monotone nondecreasing.

proof. Putting

$$\Phi_1(w) = \int_0^1 \int_0^\infty \lambda^{p/2-a+1} v^{p/2-a+1} (1-\lambda)^b (1-\lambda+v\lambda)^{a-b-2} \exp(-wv\lambda/2) g(v) dv d\lambda$$

and

$$\Phi_2(w) = \int_0^1 \int_0^\infty \lambda^{p/2-a} v^{p/2-a+1} (1-\lambda)^b (1-\lambda+v\lambda)^{a-b-2} \exp(-wv\lambda/2) g(v) dv d\lambda,$$

we have $\phi_{a,b}^{\,g}(w) = w\Phi_1(w)/\Phi_2(w)$. The derivative of $\phi_{a,b}^{\,g}(w)$ is written as

$$\begin{aligned}
\frac{d}{dw}\phi_{a,b}^{\,g}(w) &= \Phi_1(w)/\Phi_2(w) \\
&\quad -\frac{w}{2\Phi_2(w)} \int_0^1 \int_0^\infty \lambda^{p/2-a+2} v^{p/2-a+2} (1-\lambda)^b (1-\lambda+v\lambda)^{a-b-2} \\
&\quad \times \exp\left(-\frac{wv\lambda}{2}\right) g(v) dv d\lambda \\
&\quad +\frac{w\Phi_1(w)}{2\Phi_2(w)^2} \int_0^1 \int_0^\infty \lambda^{p/2-a+1} v^{p/2-a+2} (1-\lambda)^b (1-\lambda+v\lambda)^{a-b-2} \\
&\quad \times \exp\left(-\frac{wv\lambda}{2}\right) g(v) dv d\lambda.
\end{aligned}$$

For $b > 0$, applying an integration by parts gives

$$\begin{aligned}
&\int_0^1 \lambda^{p/2-a+2} (1-\lambda)^b (1-\lambda+v\lambda)^{a-b-2} \exp(-wv\lambda/2) d\lambda \\
&= \frac{2}{wv} \Big(\int_0^1 (p/2-a+2) \lambda^{p/2-a+1} (1-\lambda)^b (1-\lambda+v\lambda)^{a-b-2} \exp(-wv\lambda/2) d\lambda \\
&\quad - \int_0^1 b\lambda^{p/2-a+2} (1-\lambda)^{b-1} (1-\lambda+v\lambda)^{a-b-2} \exp(-wv\lambda/2) d\lambda \\
&\quad + \int_0^1 (a-b-2)(v-1) \lambda^{p/2-a+2} (1-\lambda)^b (1-\lambda+v\lambda)^{a-b-3} \exp(-wv\lambda/2) d\lambda \Big) \\
&= \frac{p}{wv} \int_0^1 \lambda^{p/2-a+1} (1-\lambda)^b (1-\lambda+v\lambda)^{a-b-2} \exp(-wv\lambda/2) d\lambda \\
&\quad - \frac{2}{wv} \int_0^1 ((a-2)(1-\lambda) + bv\lambda) \lambda^{p/2-a+1} (1-\lambda)^{b-1} \\
&\quad \times (1-\lambda+v\lambda)^{a-b-3} \exp(-wv\lambda/2) d\lambda. \qquad (4.2)
\end{aligned}$$

Similarly, we get

$$
\begin{aligned}
&\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^b(1-\lambda+v\lambda)^{a-b-2}\exp(-wv\lambda/2)d\lambda \\
&= \frac{p-2}{wv}\int_0^1 \lambda^{p/2-a}(1-\lambda)^b(1-\lambda+v\lambda)^{a-b-2}\exp(-wv\lambda/2)d\lambda \\
&\quad -\frac{2}{wv}\int_0^1 ((a-2)(1-\lambda)+bv\lambda)\lambda^{p/2-a}(1-\lambda)^{b-1} \\
&\quad\quad \times(1-\lambda+v\lambda)^{a-b-3}\exp(-wv\lambda/2)d\lambda.
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
\frac{d}{dw}\phi^{\,g}_{a,b}(w) \;=\; & \frac{1}{\Phi_2(w)^2}\Bigg(\Phi_2(w)\Big(\int_0^1\int_0^\infty ((a-2)(1-\lambda)+bv\lambda)\lambda^{p/2-a+1}(1-\lambda)^{b-1} \\
& \times(1-\lambda+v\lambda)^{a-b-3}\exp(-wv\lambda/2)v^{p/2-a+1}g(v)dvd\lambda\Big) \\
& -\Phi_1(w)\Big(\int_0^1\int_0^\infty ((a-2)(1-\lambda)+bv\lambda)\lambda^{p/2-a}(1-\lambda)^{b-1} \\
& \times(1-\lambda+v\lambda)^{a-b-3}\exp(-wv\lambda/2)v^{p/2-a+1}g(v)dvd\lambda\Big)\Bigg).
\end{aligned}
$$

Making the transformation $t=(1-\lambda)^{-1}\lambda v$ gives

$$
\begin{aligned}
\frac{d}{dw}\phi^{\,g}_{a,b}(w) \;=\; & \frac{1}{\Phi_2(w)^2}\Bigg(\int_0^1\int_0^\infty \frac{(a-2)+bt}{1+t}\frac{\lambda}{1-\lambda}f_w(t,\lambda)dtd\lambda \\
& \times\int_0^1\int_0^\infty f_w(t,\lambda)dtd\lambda \\
& -\int_0^1\int_0^\infty \frac{(a-2)+bt}{1+t}\frac{1}{1-\lambda}f_w(t,\lambda)dtd\lambda \\
& \times\int_0^1\int_0^\infty \lambda f_w(t,\lambda)dtd\lambda,
\end{aligned}
$$

where

$$
f_w(t,\lambda)=t^{p/2-a+1}(1+t)^{a-b-2}\lambda^{-2}(1-\lambda)^{p/2}\exp\left(-\frac{wt}{2}(1-\lambda)\right)g\left(\frac{1-\lambda}{\lambda}t\right).
$$

Here FKG inequality due to Fortuin *et al.*(1971) is useful.

Lemma 4.2 (FKG inequality). *Let ξ denote a probability density function with respect to ν for a q-variate random vector. For two points $y = (y_1, \cdots, y_q)$ and $z = (z_1, \cdots, z_q)$, define*

$$y \wedge z = (y_1 \wedge z_1, \cdots, y_q \wedge z_q),$$

$$y \vee z = (y_1 \vee z_1, \cdots, y_q \vee z_q)$$

where $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$. Suppose that ξ satisfies

$$\xi(y)\xi(z) \leq \xi(y \vee z)\xi(y \wedge z) \tag{4.3}$$

and that $\alpha(y)$, $\beta(y)$ are nondecreasing in each argument and that α, β, $\alpha\beta$ are integrable with respect to ξ. then

$$\int \alpha\beta\xi d\nu \geq \int \alpha\xi d\nu \int \beta\xi d\nu.$$

Here we note that $f_w(t, \lambda)$ satisfies (4.3) by assumption of the theorem. In the case of $b - a + 2 \geq 0$, as the function $(a - 2 + bt)(1 + t)^{-1}(1 - \lambda)^{-1}$ is nondecreasing in t and λ, the derivative of $\phi^{\,g}_{a,b}(w)$ is nonnegative. For $b - a + 2 = 0$, it is noted that the condition with respect to $g(v)$ is not needed.

In the case of $b = 0$, while the difference of the calculation on (4.2) give rise to slight difference of the form of the derivative of $\phi^{\,g}_{a,b}(w)$, we are able to show that $\phi^{\,g}_{a,b}(w)$ is nondecreasing similarly. This completes the proof. □

Theorem 4.3. *$\phi^{\,g}_{a,b}(w)/w$ is nonincreasing if $a - b - 2 \geq 0$ and $vg'(v)/g(v)$ is nonincreasing.*

proof. By change of variables formula, we write $\phi^{\,g}_{a,b}(w)/w$ as

$$\begin{aligned} \frac{\phi^{\,g}_{a,b}(w)}{w} &= \frac{\int_0^1 \int_0^\infty \lambda^{-1} t^{p/2-a+1}(1-\lambda)^b (1-\lambda+t)^{a-b-2} \exp(-wt/2) g(t/\lambda) dt d\lambda}{\int_0^1 \int_0^\infty \lambda^{-2} t^{p/2-a+1}(1-\lambda)^b (1-\lambda+t)^{a-b-2} \exp(-wt/2) g(t/\lambda) dt d\lambda} \\ &= \frac{\int_0^\infty \psi_1(t) t^{p/2-a+1} \exp(-wt/2) dt}{\int_0^\infty \psi_2(t) t^{p/2-a+1} \exp(-wt/2) dt}, \end{aligned}$$

where

$$\psi_1(t) = \int_0^1 \lambda^{-1}(1-\lambda)^b(1-\lambda+t)^{a-b-2} g(t/\lambda) d\lambda$$

and

$$\psi_2(t) = \int_0^1 \lambda^{-2}(1-\lambda)^b(1-\lambda+t)^{a-b-2} g(t/\lambda) d\lambda.$$

The derivative of $\phi_{a,b}^{\,g}(w)/w$ is

$$\begin{aligned}\frac{d}{dw}\left(\frac{\phi_{a,b}^{\,g}(w)}{w}\right) &= 2^{-1}\left(\int_0^1 k_w(t)dt\right)^{-2}\left(\int_0^1 \psi(t)k_w(t)dt\int_0^1 tk_w(t)dt\right. \\ &\quad \left. -\int_0^1 k_w(t)dt\int_0^1 t\psi(t)k_w(t)dt\right),\end{aligned}$$

where $k_w(t) = \psi_2(t)t^{p/2-a+1}\exp(-wt/2)$ and $\psi(t) = \psi_1(t)/\psi_2(t)$. By the correlation inequality, $\phi_{a,b}^{\,g}(w)/w$ is nonincreasing if $\psi(t)$ is nondecreasing in t. We can calculate the derivative of $\psi(t)$ as

$$\begin{aligned}\psi'(t) &= \psi_2(t)^{-1}\int_0^1 \big((a-b-2)(1-\lambda+t)^{a-b-3}g(t/\lambda) \\ &\quad +(1-\lambda+t)^{a-b-2}\lambda^{-1}g'(t/\lambda)\ \big)\ \lambda^{-1}(1-\lambda)^b d\lambda \\ &\quad -\psi_2(t)^{-2}\psi_1(t)\int_0^1 \lambda^{-2}(1-\lambda)^b\big((a-b-2)(1-\lambda+t)^{a-b-3}g(t/\lambda) \\ &\quad +(1-\lambda+t)^{a-b-2}\lambda^{-1}g'(t/\lambda)\ \big)\ d\lambda.\end{aligned}$$

By the assumption of the theorem, $(a-b-2)(1-\lambda+t)^{-1}+\lambda^{-1}g'(t/\lambda)/g(t/\lambda)$ is increasing in λ, which implies that $\psi(t)$ is nondecreasing in t by the correlation inequality. This completes the proof. □

Theorem 4.4. *For* $\max(a-2,0) \le b < p/2-1$,

$$\lim_{w\to\infty}\phi_{a,b}^{\,g}(w) = 2(p/2-a+1)\int_0^\infty v^{-1}g(v)dv.$$

proof. $\phi_{a,b}^{\,g}(w)$ can be expressed as

$$\phi_{a,b}^{\,g}(w) = \int_0^\infty \Psi_1(v,w)v^{-1}g(v)dv \Big/ \int_0^\infty \Psi_2(v,w)g(v)dv$$

where

$$\Psi_1(v,w) = (wv)^{p/2-a+2}\int_0^1 \lambda^{p/2-a+1}(1-\lambda)^b(1-\lambda+v\lambda)^{a-b-2}\exp(-wv\lambda/2)d\lambda$$

and

$$\Psi_2(v,w) = (wv)^{p/2-a+1}\int_0^1 \lambda^{p/2-a}(1-\lambda)^b(1-\lambda+v\lambda)^{a-b-2}\exp(-wv\lambda/2)d\lambda.$$

Making a transformation gives

$$\Psi_1(v,w) = (2w)^{p/2-a+2}\int_0^\infty q(s,v)\exp(-ws)ds$$

where $g(s,v) = s^{p/2-a+1}(1-2s/v)^b(1-2s/v+2s)^{a-b-2}I_{(0,v/2)}(s)$. Since $q(s,v) \sim s^{p/2-a+1}$ as $s \to 0+$, from Tauberian's theorem, we have

$$\int_0^\infty q(s,v)\exp(-ws)ds \sim \Gamma(p/2-a+2)w^{-(p/2-a+2)}$$

as $w \to \infty$, which implies that $\lim_{w\to\infty}\Psi_1(v,w) = 2^{p/2-a+2}\Gamma(p/2-a+2)$. Next we will show $\Psi_1(v,w)$ is bounded to the above. For $\max(a-2,0) \le b < p/2$, $\Psi_1(v,w)$ is evaluated as

$$\begin{aligned}\Psi_1(v,w) &\le w^{p/2-a+2}v^{p/2-b}\int_0^1 \lambda^{p/2-b-1}\exp(-wv\lambda/2)d\lambda \\ &\le w^{-a+b+2}2^{p/2-b}\int_0^\infty t^{p/2-b-1}\exp(-t)dt.\end{aligned}$$

Therefore $\Psi_1(v,w)$ for $\max(a-2,0) \le b < p/2$ is bounded by a constant value which is not depend upon v and w. Similarly we can show that $\lim_{w\to\infty}\Psi_2(v,w) = 2^{p/2-a+1}\Gamma(p/2-a+1)$ and that $\Psi_2(v,w)$ for $\max(a-2,0) \le b < p/2-1$ is bounded to the above. By Lebesque's dominated convergence theorem, we have

$$\begin{aligned}\lim_{w\to\infty}\phi^g_{a,b}(w) &= \lim_{w\to\infty}\int_0^\infty \Psi_1(v,w)v^{-1}g(v)dv \Big/ \lim_{w\to\infty}\int_0^\infty \Psi_2(v,w)g(v)dv \\ &= \int_0^\infty \lim_{w\to\infty}\Psi_1(v,w)v^{-1}g(v)dv \Big/ \int_0^\infty \lim_{w\to\infty}\Psi_2(v,w)g(v)dv \\ &= 2(p/2-a+1)\int_0^\infty v^{-1}g(v)dv.\end{aligned}$$

□

4.3 Admissible Minimax estimators of the mean vector

For the shrinkage estimator $\delta_\phi(X) = (1-\phi(\|X\|^2)/\|X\|^2)X$, the known minimax condition in the case of scale mixtures of multivariate normal distributions is following.

Theorem 4.5 (Strawderman(1974) and Berger(1975)). *Assume the following conditions:*

(C1) $\phi(w)$ *is nondecreasing.*

(C2) $\phi(w)/w$ *is nonincreasing.*

(C3) $0 \le \phi(w) \le 2(p-2)/\int_0^\infty vg(v)dv.$

Then $\delta_\phi(X)$*is minimax.*

By Theorem 4.1 and 4.3, except for the case of $b-a+2=0$, we cannot get the range of values (a,b) where both assumption (C1) and (C2) of Theorem 4.5 are satisfied. Taking account of admissibility, we have to restrict the range of $a<1$ and $b>-1$ which implies that $b-a+2>0$. Therefore we would like to derive the sufficient condition for minimaxity which does not require the decrease of $\phi^g_{a,b}(w)/w$. The below theorem is a special case of Berger(1975)'s Theorem 3, which provides the sufficient condition for minimaxity of general spherically symmetric distributions.

Theorem 4.6 (Berger(1975)). *Assume the following conditions:*

(D1) $\phi(w)$ *is nondecreasing.*

(D2) $0 \le \phi(w) \le 2(p-2)\int_0^\infty v^{p/2-1}g(v)dv / \int_0^\infty v^{p/2}g(v)dv.$

Then $\delta_\phi(X)$*is minimax.*

proof. By using the Stein identity, Lemma 3.2, the risk difference can be written as

$$\begin{aligned} R(\theta, X) - R(\theta, \delta_\phi) &= E\left(\|X-\theta\|^2\right) - E\left(\|\left(1-\phi(\|X\|^2)/\|X\|^2\right)X - \theta\|^2\right) \\ &= \int_{R^p}\int_0^\infty \frac{2(p-2)\phi(\|x\|^2)}{\|x\|^2}\frac{v^{p/2-1}}{(2\pi)^{p/2}}\exp(-\frac{\|x-\theta\|^2 v}{2})g(v)dvdx \\ &\quad + \int_{R^p}\int_0^\infty 4\phi'(\|x\|^2)\frac{v^{p/2-1}}{(2\pi)^{p/2}}\exp(-\frac{\|x-\theta\|^2 v}{2})g(v)dvdx \\ &\quad - \int_{R^p}\int_0^\infty \frac{\phi^2(\|x\|^2)}{\|x\|^2}\frac{v^{p/2}}{(2\pi)^{p/2}}\exp(-\frac{\|x-\theta\|^2 v}{2})g(v)dvdx. \end{aligned}$$

Using the inequality

$$\frac{\int_0^\infty v^{p/2-1}\exp(-v\|x-\theta\|^2/2)g(v)dv}{\int_0^\infty v^{p/2}\exp(-v\|x-\theta\|^2/2)g(v)dv} \ge \frac{\int_0^\infty v^{p/2-1}g(v)dv}{\int_0^\infty v^{p/2}g(v)dv},$$

which is shown by the correlation inequality, we evaluate the risk difference as

$$\begin{aligned} R(\theta, X) - R(\theta, \delta_\phi) &\ge \int_{R^p}\int_0^\infty \left(\left(2(p-2) - \frac{\int_0^\infty v^{p/2}g(v)dv}{\int_0^\infty v^{p/2-1}g(v)dv}\phi(\|x\|^2)\right)\frac{\phi(\|x\|^2)}{\|x\|^2}\right. \\ &\quad \left. +4\phi'(\|x\|^2)\right)(2\pi)^{-p/2}v^{p/2-1}\exp(-\frac{\|x-\theta\|^2 v}{2})g(v)dvdx. \end{aligned}$$

By the conditions of (D1) and (D2), we see that $R(\theta, X) - R(\theta, \delta_\phi)$ is nonnegative for every θ. This completes the proof. □

Combining Theorem 4.1, 4.3, 4.4, 4.5 and 4.6, we have the theorem about minimaxity and admissibility of $\delta_{a,b}^g(X)$.

Theorem 4.7. *1. If $a = b + 2$ and $vg'(v)/g(v)$ is nonincreasing, $\delta^g_{a,b}(X)$ for $\max(b_g, 0) \le b < p/2 - 1$ is minimax, where*

$$b_g = (p/2-1)(1 - 2\left(\int_0^\infty vg(v)dv \int_0^\infty v^{-1}g(v)dv\right)^{-1}).$$

2. If $g(v)$ satisfies the assumption of part 1 of Theorem 4.1, $\delta^g_{a,b}(X)$ for $a_g \le a < p/2+1$ and $\max(a-2, 0) \le b < p/2 - 1$ is minimax, where

$$a_g = p/2 + 1 - (p-2)\int_0^\infty v^{p/2-1}g(v)dv\left(\int_0^\infty v^{-1}g(v)dv \int_0^\infty v^{p/2}g(v)dv\right)^{-1}.$$

3. If $g(v)$ satisfies the assumption of part 1 of Theorem 4.1 and $a_g < 1$, $\delta^g_{a,b}(X)$ for $a_g \le a < 1$ and $0 \le b < p/2 - 1$ is admissible and minimax.

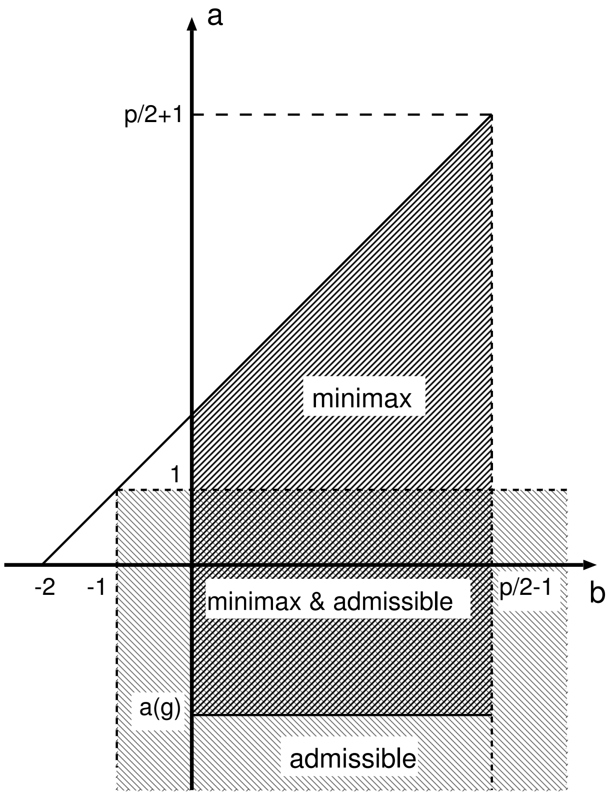

Figure 4.1 : Ranges of values (a, b) for minimaxity and admissibility (Theorem 4.7).

The result of part *1* of Theorem 4.7 is a revised version of Strawderman(1974)'s one about minimax generalized Bayes estimators. In fact, $h_{a,b}(\theta)$ with $a = b + 2$ is proportional to $\|\theta\|^{2-p+2b}$, which corresponds to the generalized prior density considered in Strawderman(1974).

Finally we consider a class of $g(v)$ satisfying the assumption of part *1* of Theorem 4.1. We easily check that $g_{\alpha,\beta}(v) = Cv^{\alpha-1}\exp(-\beta v)$ for $\alpha > 0$ and $\beta > 0$ satisfies this condition. In particular, $g_{\alpha,\beta}(v)$ for $\alpha = \beta = m/2$ corresponds to the case of multivariate-t distribution whose degrees of freedom is m. Since we calculate $a_g = p/2 + 1 - (p-2)(m-2)/(p+m-2)$ in this case, minimax admissible estimators can be proposed if $p \geq 5$ and $m > (p-2)(p+4)/(p-4)$.

Chapter 5

Estimation problems on a multiple regression model

In Chapter 3, we considered the estimation problem of a multivariate normal mean with known covariance matrix. From a practical sense, it is however 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

$$X \sim N_p(\theta, \sigma^2 I_p), \qquad S/\sigma^2 \sim \chi^2_n,$$

where random variables X and S are independent and χ^2_n denotes a chi square distribution with degrees of freedom n. In this chapter, we deal with the problem of estimating the mean vector θ and the variance σ^2.

5.1 Minimax estimators of a normal variance

5.1.1 Introduction

In this section, we review my published paper Maruyama(1999a), where the problem of estimating the unknown variance σ^2 by an estimator δ relative to the quadratic loss function $L_1(\sigma^2, \delta) = (\delta/\sigma^2 - 1)^2$ is considered. Stein(1964) showed that the best affine equivariant estimator is $\delta_0 = (n+2)^{-1}S$ and it can be improved by considering a class of scale equivariant estimators $\delta_\phi = \phi(Z)S$, for $Z = \|X\|^2/S$. He really found an improved estimator $\delta^{ST} = \phi^{ST}(Z)S$, where $\phi^{ST}(z) = \min((n+2)^{-1}, (n+p+2)^{-1}(1+z))$. Brewster and Zidek(1974) derived an improved generalized Bayes estimator $\delta^{BZ} = \phi^{BZ}(Z)S$, where

$$\phi^{BZ}(z) = \frac{1}{n+p+2} \frac{\int_0^1 \lambda^{p/2-1}(1+\lambda z)^{-(n+p)/2-1} d\lambda}{\int_0^1 \lambda^{p/2-1}(1+\lambda z)^{-(n+p)/2-2} d\lambda}.$$

We note that shrinkage estimators such as Stein's procedure and Brewster-Zidek's procedure are derived by using the vague prior information that $\lambda = \|\theta\|^2/\sigma^2$ is close to 0. It goes without saying that we would like to get significant improvement of risk when the prior information is accurate. Though δ^{ST} improves on δ_0 at $\lambda = 0$, it is not analytic and thus inadmissible. On the other hand, Brewster-Zidek's estimator does not improve on δ_0 at $\lambda = 0$ though it is admissible as shown in Proskin(1985). Therefore it is desirable to get analytic improved estimators dominating δ_0 especially at $\lambda = 0$. In this thesis, we give such a class of improved estimators $\delta_\alpha^V = \phi_\alpha^V(Z)S$, where

$$\phi_\alpha^V(z) = \frac{1}{n+p+2} \frac{\int_0^1 \lambda^{\alpha p/2-1}(1+\lambda z)^{-\alpha((n+p)/2+1)} d\lambda}{\int_0^1 \lambda^{\alpha p/2-1}(1+\lambda z)^{-\alpha((n+p)/2+1)-1} d\lambda},$$

for $\alpha > 1$. It is noted that δ_1^V coincides with Brewster-Zidek's estimator. Further we demonstrate that δ_α^V with $\alpha > 1$ improves on δ_0 especially at $\lambda = 0$ and that δ_α^V approaches Stein's estimator when α tends to infinity.

5.1.2 Main results

We first derive the estimator δ_α^V. The problem of estimating α times the variance $\alpha\sigma^2$ relative to the loss function $L_\alpha(\sigma^2, d) = (d/(\alpha\sigma^2) - 1)^2$ is considered, which is slight different from that of the variance σ^2. Among many generalized Bayes estimators for this problem, by selecting a suitable prior distribution, we can propose the estimator δ_α^V with $\alpha > 1$ which is not suitable for minimax estimator of $\alpha\sigma^2$ but that of σ^2. So far we have not determine whether or not δ_α^V with $\alpha > 1$ is the generalize Bayes estimator of σ^2.

Calculation for deriving the estimator δ_α^V is following. For $\eta = 1/(\alpha\sigma^2)$, let the conditional distribution of θ given λ, $0 < \lambda < 1$, be normal with mean 0 and covariance matrix $\lambda^{-1}(1-\lambda)\alpha^{-1}\eta^{-1}I_p$ and density functions of λ and η are proportional to $\lambda^{(\alpha-1)p/2-1}I_{(0,1)}(\lambda)$ and $\eta^{(\alpha-1)((p+n)/2+1)-1}I_{(0,\infty)}(\eta)$, respectively . Then the joint density function $g(\eta, x, s)$ of η, X, S is given by

$$
\begin{aligned}
g(\eta,x,s) &\propto \int \eta^{p/2}\exp\left(-\frac{\alpha\eta}{2}\|x-\theta\|^2\right)\left(\eta\frac{\lambda}{1-\lambda}\right)^{p/2}\exp\left(-\frac{\lambda}{1-\lambda}\frac{\alpha\eta}{2}\|\theta\|^2\right) \\
&\quad \cdot\eta^{(\alpha-1)((p+n)/2+1)-1}\lambda^{(\alpha-1)p/2-1}\eta^{n/2}\exp(-\alpha\eta s/2)d\theta d\lambda \\
&\propto \int \eta^{p/2}\left(\eta\frac{\lambda}{1-\lambda}\right)^{p/2}\exp\left(-\alpha\eta\frac{\|\theta-(1-\lambda)x\|^2}{2(1-\lambda)}-\frac{\alpha\eta\|x\|^2\lambda}{2}\right) \\
&\quad \cdot\eta^{(\alpha-1)((p+n)/2+1)-1}\lambda^{(\alpha-1)p/2-1}\eta^{n/2}\exp(-\alpha\eta s/2)d\theta d\lambda \\
&\propto \eta^{\alpha((p+n)/2+1)-2}\int_0^1 \lambda^{\alpha p/2-1}\exp\left(-\alpha\eta\frac{\|x\|^2\lambda+s}{2}\right)d\lambda.
\end{aligned}
$$

As a generalized Bayes estimator for the loss function $L_\alpha(d, \alpha\sigma^2)$ is $E(\eta \mid X, S)/E(\eta^2 \mid X, S)$, we have the generalized Bayes estimator

$$
\begin{aligned}
\delta_\alpha^V &= \frac{\int_0^1 \lambda^{\alpha p/2-1}\int_0^\infty \eta^{\alpha((p+n)/2+1)-1}\exp\left(-\alpha\eta(\|X\|^2\lambda+S)/2\right)d\lambda}{\int_0^1 \lambda^{\alpha p/2-1}\int_0^\infty \eta^{\alpha((p+n)/2+1)}\exp\left(-\alpha\eta(\|X\|^2\lambda+S)/2\right)d\lambda} \\
&= \frac{1}{n+p+2}\frac{\int_0^1 \lambda^{\alpha p/2-1}(1+\lambda Z)^{-\alpha((n+p)/2+1)}d\lambda}{\int_0^1 \lambda^{\alpha p/2-1}(1+\lambda Z)^{-\alpha((n+p)/2+1)-1}d\lambda}S.
\end{aligned}
$$

In the following, $\phi_\alpha^V(z)$ is represented through the hypergeometric function

$$
F(a,b,c,x) = 1+\sum_{n=1}^\infty \frac{(a)_n(b)_n}{(c)_n}\frac{x^n}{n!} \qquad \text{for} \qquad (a)_n = a\cdot(a+1)\cdots(a+n-1).
$$

The following facts about $F(a, b, c, x)$, from Abramowitz and Stegun(1964), are needed;

$$
\int_0^x t^{a-1}(1-t)^{b-1}dt = \frac{x^a}{a}F(a,1-b,a+1,x) \quad \text{for} \quad a,b>1, \tag{5.1}
$$

$$
F(a,b,c,x) = (1-x)^{c-a-b}F(c-a,c-b,c,x), \tag{5.2}
$$

$$(c-a-b)F(a,b,c,x)-(c-a)F(a-1,b,c,x)+b(1-x)F(a,b+1,c,x)=0, \quad (5.3)$$

$$(b-a)(1-x)F(a,b,c,x)-(c-a)F(a-1,b,c,x)+(c-b)F(a,b-1,c,x)=0, \quad (5.4)$$

$$F(a,b,c,1)=\infty \qquad \text{when } \ c-a-b\le -1. \quad (5.5)$$

Making a transformation and using (5.1) and (5.2), we have

$$\begin{aligned}
&\int_0^1 \lambda^{\alpha p/2-1}(1+\lambda z)^{-\alpha((n+p)/2+1)}d\lambda \Big/ \int_0^1 \lambda^{\alpha p/2-1}(1+\lambda z)^{-\alpha((n+p)/2+1)-1}d\lambda \\
&= \int_0^{\frac{z}{z+1}} t^{\alpha p/2-1}(1-t)^{\alpha(n/2+1)-1}dt \Big/ \int_0^{\frac{z}{z+1}} t^{\alpha p/2-1}(1-t)^{\alpha(n/2+1)}dt \\
&= (z+1)\frac{F\left(1,\alpha(n+p+2)/2,\alpha p/2+1,z/(z+1)\right)}{F\left(1,\alpha(n+p+2)/2+1,\alpha p/2+1,z/(z+1)\right)}.
\end{aligned}$$

Moreover by (5.3) and (5.4), $\phi_\alpha^V(z)$ is expressed as

$$\begin{aligned}
\phi_\alpha^V(z) &= \frac{1}{n+2}\left[1-\frac{p}{p+n+2}\frac{z+1}{F(1,\alpha(p+n+2)/2+1,\alpha p/2+1,z/(z+1))}\right] \\
&= \frac{1}{n+2}\left[1-\frac{p}{(n+2)F(1,\alpha(p+n+2)/2,\alpha p/2+1,z/(z+1))+p}\right]. \quad (5.6)
\end{aligned}$$

The following sufficient condition for minimaxity is very useful.

Theorem 5.1 (Brewster and Zidek(1974)). *Assume that*

(E1) $\phi(z)$ *is nondecreasing and* $\lim_{z\to\infty}\phi(z)=1/(n+2)$

(E2) $\phi(z)\ge\phi^{BZ}(z)$.

Then $\delta(X,S)=\phi(Z)S$ *is minimax.*

proof. See Brewster and Zidek(1974) and Kubokawa(1994). □

Making use of this theorem, we have the following.

Theorem 5.2. *The estimator δ^V_α with $\alpha \geq 1$ is minimax.*

proof. We shall verify that $\phi^V_\alpha(z)$ with $\alpha \geq 1$ satisfies the conditions (E1) and (E2). Since $F(1, \alpha(p+n+2)/2, \alpha p/2+1, z/(z+1))$ is increasing in z, $\phi^V_\alpha(z)$ is increasing in z. By (5.5), it is clear that $\lim_{z\to\infty} \phi^V_\alpha(z) = 1/(n+2)$. In order to show that $\phi^V_\alpha(z) \geq \phi^V_1(z)$ for $\alpha \geq 1$, we have only to check that $F(1, \alpha(p+n+2)/2, \alpha p/2+1, z/(z+1))$ is increasing in α, which is easily verified because the coefficient of each term of the right-hand side of the equation

$$\begin{aligned} &F(1, \alpha(p+n+2)/2, \alpha p/2+1, z/(z+1)) \\ &= \ 1 + \frac{p+n+2}{p+2/\alpha}\frac{z}{1+z} + \frac{(p+n+2)(p+n+2+2/\alpha)}{(p+2/\alpha)(p+4/\alpha)}\left(\frac{z}{1+z}\right)^2 + \cdots \end{aligned} \tag{5.7}$$

is increasing in α. We have thus proved the theorem. □

Now we investigate the nature of the risk of δ^V_α with $\alpha > 1$. By using Kubokawa(1994)'s method, the risk difference between δ_0 and δ^V_α at $\lambda = 0$ is written as

$$R(0, \delta_0) - R(0, \delta^V_\alpha) = 2\int_0^\infty \frac{d}{dz}\phi^V_\alpha(z)\left(\phi^V_\alpha(z) - \phi^V_1(z)\right)\int_0^\infty t^2 F_p(zt) f_n(t)\,dt\,dz,$$

where $f_k(t)$ and $F_k(t)$ designate the density and the distribution functions of χ^2_k. Therefore we see that Brewster-Zidek's estimator (δ^V_α with $\alpha = 1$) does not improve upon the best equivariant estimator at $\lambda = 0$. See also Rukhin(1992). On the other hand, since $\phi^V_\alpha(z)$ is strictly increasing in α, δ^V_α with $\alpha > 1$ improves on the best equivariant estimator especially at $\lambda = 0$. Figure 5.1 gives a comparison of the respective risks of the best equivariant estimator, Stein's estimator, Brewster-Zidek's estimator, δ^V_α with $\alpha = 2, 4$ and 10 for $p = 10$ and $n = 4$. This figure reveals that the risk behavior of δ^V_α with $\alpha = 10$ is similar to that of Stein's truncated estimator. In fact, we have the following result.

Proposition 5.3. *δ^V_α approaches Stein's estimator when α tends to infinity.*

proof. Since $F(1, \alpha(p+n+2)/2, \alpha p/2+1, z/(z+1))$ is increasing in α, by the monotone convergence theorem this function converges to $\sum_{i=0}^\infty \{(p(z+1))^{-1}(n+p+2)z\}^i$ when

α tends to infinity. Considering two cases: $(n+p+2)z < (\geq) p(z+1)$, we obtain $\lim_{\alpha\to\infty} \phi^V_\alpha(z) = (1+z)/(n+p+2)$ if $z < p/(n+2)$; $= 1/(n+2)$ otherwise. This completes the proof. □

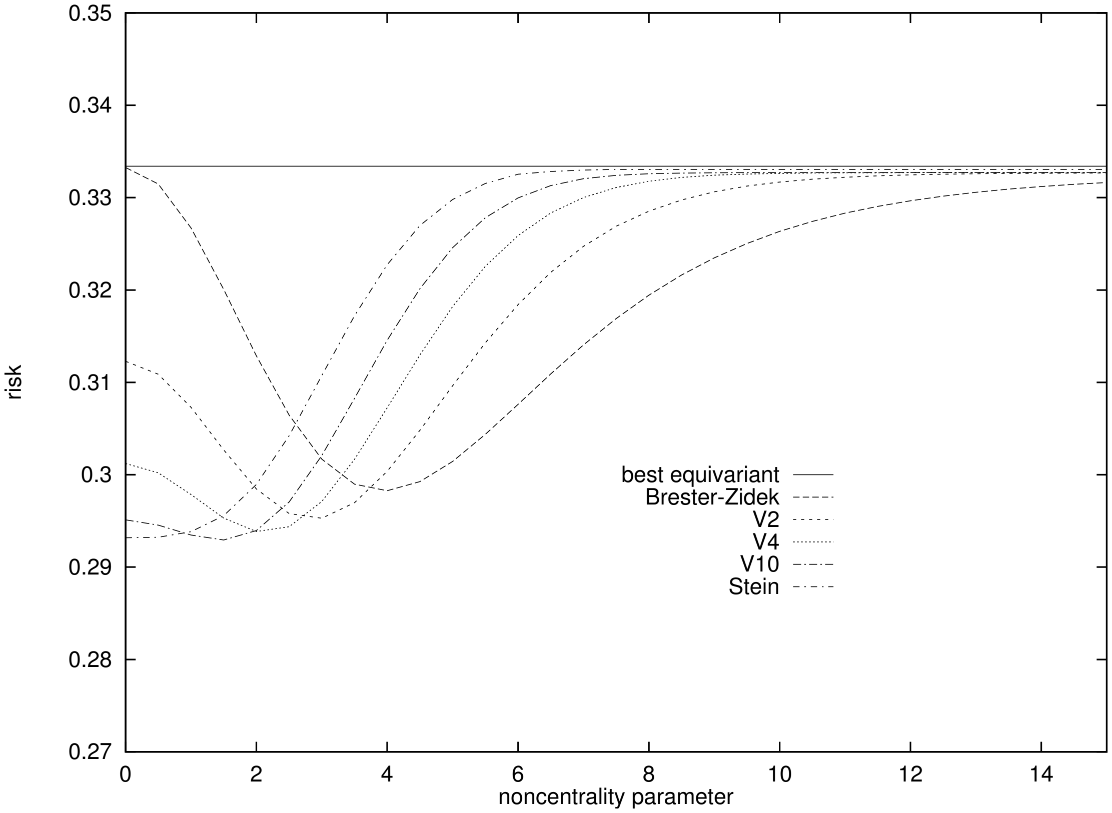

Figure 5.1 : Comparison of the risks of the estimators δ_0, δ^{ST}, δ^{BZ}, δ^V_α with $\alpha = 2, 4, 10$. 'noncentrality parameter' denotes $\|\theta\|/\sigma$. Vα denotes the risk of δ^V_α.

Remark 5.4. Ghosh(1994) proposed the minimax generalized Bayes estimator $\delta^{GH}_k = \phi^{GH}_k(Z)S$, where

$$\phi^{GH}_k(z) = \frac{1}{n+p+2(k+2)} \frac{\int_0^1 \lambda^{p/2+k}(1+\lambda z)^{-(n+p)/2-(k+2)} d\lambda}{\int_0^1 \lambda^{p/2+k}(1+\lambda z)^{-(n+p)/2-(k+3)} d\lambda},$$

for $-1-p/2 < k \le -1$. Clearly δ_k^{GH} with $k = -1$ coincides with Brewster-Zidek's estimator and we can see that δ_k^{GH} with $-1-p/2 < k < -1$ improves on the best equivariant estimator especially at $\lambda = 0$. It is noted that without the troublesome calculation in Ghosh(1994), the minimaxity of Ghosh's estimators is easily proved in the same way as Theorem 5.2 because ϕ_k^{GH} with $-1-p/2 < k \le -1$ satisfies conditions (E1) and (E2).

Remark 5.5. For the entropy loss function $L_0(\sigma^2, d) = d/\sigma^2 - \log(d/\sigma^2) - 1$, the discussions in this section are directly applied. In this case, the best equivariant estimator is the unbiased estimator $\delta_0 = S/n$ and Stein's truncated estimator is $\delta^{ST} = \phi^{ST}(Z)S$, where $\phi^{ST}(z) = \min(1/n, (n+p)^{-1}(1+z))$. Moreover Brewster-Zidek's estimator is $\delta^{BZ} = \phi^{BZ}(Z)S$, where

$$\phi^{BZ}(z) = \frac{1}{n+p} \frac{\int_0^1 \lambda^{p/2-1}(1+\lambda z)^{-(n+p)/2} d\lambda}{\int_0^1 \lambda^{p/2-1}(1+\lambda z)^{-(n+p)/2-1} d\lambda}.$$

Then the proposed minimax estimator is $\delta_\alpha^V = \phi_\alpha^V(Z)S$, where

$$\phi_\alpha^V(z) = \frac{1}{n+p} \frac{\int_0^1 \lambda^{\alpha p/2-1}(1+\lambda z)^{-\alpha(n+p)/2} d\lambda}{\int_0^1 \lambda^{\alpha p/2-1}(1+\lambda z)^{-\alpha(n+p)/2-1} d\lambda},$$

for $\alpha > 1$.

5.2 Estimation of a multivariate normal mean with unknown variance

5.2.1 Introduction

In this section, we review results of Maruyama(1999b) where the problem of estimating the mean vector θ relative to the quadratic loss function $\|\delta - \theta\|^2/\sigma^2$ is considered. The usual minimax estimator X is inadmissible for $p \ge 3$. James and Stein(1961) constructed the improved estimator $\delta^{JS} = [1 - ((p-2)/(n+2))S/\|X\|^2]\, X$, which is also dominated by the James-Stein positive-part estimator $\delta_+^{JS} = \max(0, \delta^{JS})$ as shown in

Baranchik(1964). We note that shrinkage estimators such as James-Stein's procedure are derived by using the vague prior information that $\lambda = \|\theta\|^2/\sigma^2$ is close to 0. It goes without saying that we would like to get significant improvement of risk when the prior information is accurate. Though δ_+^{JS} improves on δ^{JS} at $\lambda = 0$, it is not analytic and thus inadmissible. Kubokawa(1991) showed that one of minimax generalized Bayes estimators derived by Lin and Tsai(1973) dominates δ^{JS}. This estimator, however, does not improve on δ^{JS} at $\lambda = 0$. Therefore it is desirable to get analytic improved estimators dominating δ^{JS} especially at $\lambda = 0$. In this thesis, we show that the estimators in the subclass of Lin-Tsai's minimax generalized Bayes estimators $\delta_\alpha^M = (1 - \psi_\alpha^M(Z)/Z)X$, where $Z = \|X\|^2/S$ and

$$\psi_\alpha^M(z) = z\frac{\int_0^1 \lambda^{\alpha(p-2)/2}(1+\lambda z)^{-\alpha(n+p)/2-1}d\lambda}{\int_0^1 \lambda^{\alpha(p-2)/2-1}(1+\lambda z)^{-\alpha(n+p)/2-1}d\lambda}$$

with $\alpha \geq 1$, dominate δ^{JS}. It is noted that δ_α^M with $\alpha = 1$ coincides with the estimator derived by Kubokawa(1991). Further we demonstrate that δ_α^M with $\alpha > 1$ improves on δ^{JS} especially at $\lambda = 0$ and that δ_α^M approaches the James-Stein positive-part estimator when α tends to infinity.

5.2.2 Main results

We first represent $\psi_\alpha^M(z)$ through the hypergeometric function $F(a, b, c, x)$. The formulas from (5.1) to (5.5) and

$$c(1-x)F(a,b,c,x) - cF(a,b-1,c,x) + (c-a)xF(a,b,c+1,x) = 0, \qquad (5.8)$$

are needed. Making a transformation and using (5.1) and (5.2), we have

$$\begin{aligned}
&\int_0^1 \lambda^{\alpha(p/2-1)}(1+\lambda z)^{-\alpha(n+p)/2-1}d\lambda \Big/ \int_0^1 \lambda^{\alpha(p/2-1)-1}(1+\lambda z)^{-\alpha(n+p)/2-1}d\lambda \\
&= \frac{\alpha(p/2-1)}{\alpha(p/2-1)+1}\frac{F\left(1, \alpha(n+p)/2+1, \alpha(p/2-1)+2, z/(z+1)\right)}{F\left(1, \alpha(n+p)/2+1, \alpha(p/2-1)+1, z/(z+1)\right)}.
\end{aligned}$$

Moreover by (5.3) and (5.8), we obtain

$$\psi_\alpha^M(z) = \frac{p-2}{n+2}\left[1 - \frac{n+p}{(n+2)F(1,\alpha(n+p)/2,\alpha(p-2)/2+1,z/(z+1))+p-2}\right]. \tag{5.9}$$

Making use of (5.9), we can easily prove the theorem.

Theorem 5.6. *The estimator δ_α^M with $\alpha \geq 1$ dominates δ^{JS}.*

proof. We shall verify that $\psi_\alpha^M(z)$ with $\alpha \geq 1$ satisfies the condition for dominating the James-Stein estimator derived by Kubokawa(1994): for $\delta_\psi = (1 - \psi(Z)/Z)X$, $\psi(z)$ is nondecreasing, $\lim_{z\to\infty}\psi(z) = (p-2)/(n+2)$, and $\psi(z) \geq \psi_1^M(z)$.

Since $F(1,\alpha(n+p)/2,\alpha(p-2)/2+1,z/(z+1))$ is increasing in z, $\psi_\alpha^M(z)$ is increasing in z. By (5.5), it is clear that $\lim_{z\to\infty}\psi_\alpha^M(z) = (p-2)/(n+2)$. In order to show that $\psi_\alpha^M(z) \geq \psi_1^M(z)$ for $\alpha \geq 1$, we have only to check that $F(1,\alpha(n+p)/2,\alpha(p-2)/2+1,z/(z+1))$ is increasing in α, which is easily verified because the coefficient of each term of the right-hand side of the equation

$$\begin{aligned} &F(1,\alpha(n+p)/2,\alpha(p-2)/2+1,z/(z+1)) \\ &= 1 + \frac{p+n}{p-2+2/\alpha}\frac{z}{1+z} + \frac{(p+n)(p+n+2/\alpha)}{(p-2+2/\alpha)(p-2+4/\alpha)}\left(\frac{z}{1+z}\right)^2 + \cdots \end{aligned} \tag{5.10}$$

is increasing in α. We have thus proved the theorem. □

Now we investigate the nature of the risk of δ_α^M with $\alpha \geq 1$. By using Kubokawa(1994)'s method, the risk difference between the James-Stein estimator and δ_α^M at $\lambda = 0$ is written as

$$R(0,\delta^{JS}) - R(0,\delta_\alpha^M) = 2(n+p)\int_0^\infty \frac{d}{dz}\psi_\alpha^M(z)\left(\frac{\psi_\alpha^M(z)-\psi_1^M(z)}{1+\psi_1^M(z)}\right)\int_0^z s^{-1}h(s)ds dz,$$

where $h(s) = \int_0^\infty v f_n(v) f_p(vs) dv$ and $f_p(t)$ designates a density of χ_p^2. Therefore we see that Kubokawa(1991)'s estimator (δ_α^M with $\alpha = 1$) does not improve upon the James-Stein estimator at $\lambda = 0$. On the other hand, since $\psi_\alpha^M(z)$ is strictly increasing in α, δ_α^M with $\alpha > 1$ improves on the James-Stein estimator especially at $\lambda = 0$ and the risk

$R(\lambda, \delta_\alpha^M)$ at $\lambda = 0$ is decreasing in α. Figure 5.2 gives a comparison of the respective risks of the James-Stein estimator, its positive-part rule and δ_α^M with $\alpha = 1, 2$ and 10 for $p = 8$ and $n = 4$.

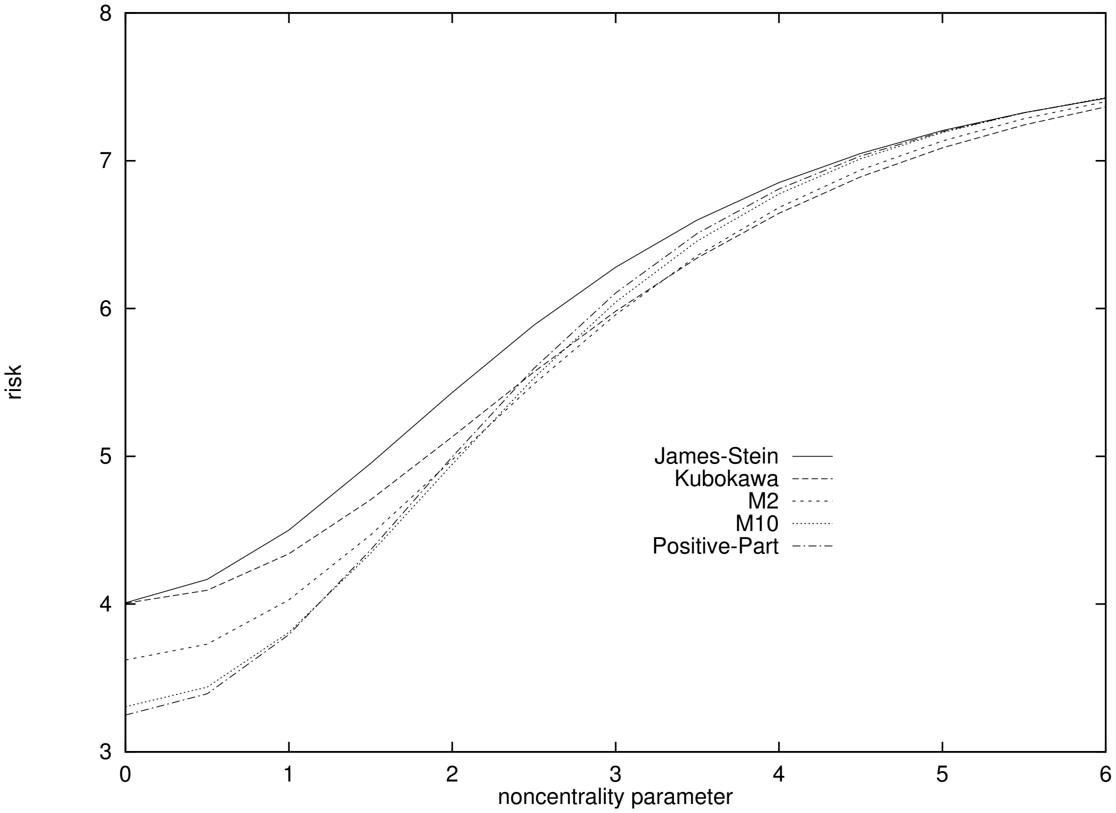

Figure 5.2 : Comparison of the risks of the estimators δ^{JS}, δ_+^{JS}, δ_α^M with $\alpha = 1, 2, 10$. 'noncentrality parameter' denotes $\|\theta\|/\sigma$. Mα denotes the risk of δ_α^M.

This figure reveals that the risk behavior of δ_α^M with $\alpha = 10$ is similar to that of the James-Stein positive-part estimator. In fact, we have the following result.

Proposition 5.7. *δ_α^M approaches the James-Stein positive-part estimator when α tends to infinity, that is,* $\lim_{\alpha\to\infty} \delta_\alpha^M = \delta_+^{JS}$.

proof. Since $F(1, \alpha(n+p)/2, \alpha(p-2)/2+1, z/(z+1))$ is increasing in α, by the monotone convergence theorem this function converges to $\sum_{i=0}^{\infty}\{((p-2)(z+1))^{-1}(n+p)z\}^i$.

Considering two cases: $(n+p)z < (\geq)(p-2)(z+1)$, we obtain $\lim_{\alpha\to\infty}\psi_\alpha^M(z) = z$ if $z < (p-2)/(n+2)$; $= (p-2)/(n+2)$ otherwise. This completes the proof. □

We, here, consider the connection between the estimation of the mean vector and the variance. It is noted that the James-Stein estimator is expressed as $(1-((p-2)/\|X\|^2)\hat{\sigma}_0^2)X$, for $\hat{\sigma}_0^2 = (n+2)^{-1}S$ and that $\hat{\sigma}_0^2$ is the best affine equivariant estimator of the variance under the quadratic loss function $L_1(\sigma^2, d) = (d/\sigma^2 - 1)^2$, as seen in Section 5.1. Stein(1964) showed that $\hat{\sigma}_0^2$ is inadmissible. George(1990) suggested that it might be possible to use an improved variance estimator to improve on the James-Stein estimator. In the same way as Theorem 5.6, we have the following result, which includes Berry(1994)'s and Kubokawa *et al.*(1993)'s.

Theorem 5.8. *Assume that $\hat{\sigma}^2$ is an improved estimator of variance, which satisfies the conditions (E1) and (E2) of Theorem 5.1. Then $\delta = (1-(p-2)\hat{\sigma}^2/\|X\|^2)X$ dominates δ^{JS}.*

It should be, however, noticed that the estimator derived from Theorem 5.8 is obviously inadmissible, since $(1-(p-2)\hat{\sigma}^2/\|X\|^2)$ sometimes takes a negative value.

Chapter 6

A New Positive Estimator of Loss Function

6.1 Introduction

In this chapter, we review my published paper Maruyama(1997). Let X be a random variable having a p-variate normal distribution $N_p(\theta, I_p)$. We assume that for the problem of estimating θ relative to the quadratic loss function, the usual minimax estimator X is used. Now it is of great interest in this loss $\|X - \theta\|^2$, which depends on θ and is unobservable. Therefore we wish to estimate it from the data using rule $\delta(X)$. To study how well $\delta(X)$ estimates the loss, we use the squared error loss function. Thus the expected distance or risk incurred by $\delta(X)$ is $R(\delta, \theta) = E[\{\delta(X) - \|X - \theta\|^2\}^2]$. Johnstone(1987) showed that the unbiased estimator $\delta_U(X) = p$ of the loss is admissible when $p \leq 4$, but is dominated by $\delta^{JO}(X) = p - 2(p-4)/\|X\|^2$ when $p \geq 5$. While $\|X - \theta\|^2$ is non-negative, $\delta^{JO}(X)$ takes negative values when $\|X\|^2 < 2(p-4)/p$. We need to modify $\delta^{JO}(X)$ in order to eliminate this undesirable property. One possible modification is, as proposed in Johnstone(1987), to truncate the rule at zero, that is $\delta_0^{JO}(X) = \max(0, p - 2(p-4)/\|X\|^2)$. However, this modification is imperfect because the loss is positive with probability 1. In this thesis, We propose a class of positive estimators improving on $\delta^{JO}(X)$:

$$\delta_a^{JO}(X) = \max\left[a, p - \frac{2(p-4)}{\|X\|^2}\right] \quad \text{for} \quad 0 < a \leq 2\frac{p}{p-2}.$$

Through our simulation results, we see that positive estimators should be chosen if one has no information about $\|\theta\|^2$. The case where the covariance matrix of X is $\sigma^2 I_p$ for unknown σ^2 is also discussed.

6.2 A positive estimator improving on Johnstone's estimator

We consider a class of orthogonally invariant estimators $\delta_\phi(X) = p - \phi(\|X\|^2)/\|X\|^2$. Using Stein(1973)'s identity, Lemma 3.2, Johnstone(1987) derived the unbiased estimator of the risk, which is written as

$$\begin{aligned} R(\delta_\phi(X), \theta) &= E\left[\left(p - \frac{\phi(\|X\|^2)}{\|X\|^2} - \|X-\theta\|^2\right)^2\right] \\ &= 2p + E\left[4(p-4)\left(\frac{\phi'(\|X\|^2)}{\|X\|^2} - \frac{\phi(\|X\|^2)}{\|X\|^4}\right)\right] \\ &\quad + E\left[8\phi''(\|X\|^2) + \frac{\phi^2(\|X\|^2)}{\|X\|^4}\right]. \end{aligned} \tag{6.1}$$

Then by the above equation and Kubokawa(1994)'s argument, the condition on $\phi(w)$ for the improvement on $\delta^{JO}(X)$ is obtained.

Theorem 6.1. *Assume the following conditions;*

1. $\phi(w)$ *is nondecreasing and* $\lim_{w\to\infty}\phi(w) = 2(p-4)$,

2. $\phi(w) \geq \min(2(p-4), \phi_0(w))$ *for every* w*, where,*

$$\phi_0(w) = 2(p-4) + (2 - 4/w)\frac{w^{p/2-1}e^{-w/2}}{\int_0^w y^{p/2-3}e^{-y/2}dy}.$$

Then $\delta_\phi(X)$ improves on $\delta^{JO}(X)$.

proof. We want to get the condition of $\phi(w)$ satisfying $R(\delta^{JO},\theta) - R(\delta_\phi,\theta) \geq 0$ for every θ. By using the equation (6.1) and noting $\phi(\infty) = 2(p-4)$, the risk difference is written as

$$\begin{aligned}
&R(\delta^{JO},\theta) - R(\delta_\phi,\theta)\\
&= E\left[-\frac{4(p-4)^2}{\|X\|^4} - 4(p-4)(\frac{\phi'(\|X\|^2)}{\|X\|^2} - \frac{\phi(\|X\|^2)}{\|X\|^4}) - 8\phi''(\|X\|^2) - \frac{\phi^2(\|X\|^2)}{\|X\|^4}\right]\\
&= E\left[\frac{\phi(\infty)(\phi(\infty)-4(p-4))}{\|X\|^4} - \frac{\phi^2(\|X\|^2)}{\|X\|^4} + 4(p-4)\frac{\phi(\|X\|^2)}{\|X\|^4}\right]\\
&\quad +E\left[-4(p-4)\frac{\phi'(\|X\|^2)}{\|X\|^2} - 8\phi''(\|X\|^2)\right]. \qquad (6.2)
\end{aligned}$$

The first term of the right-hand side of the extreme equation in (6.2) is rewritten as

$$\begin{aligned}
&E\left[\int_1^\infty \frac{d}{dt}\left[\frac{\phi(t\|X\|^2)}{\|X\|^4}\left(\phi(t\|X\|^2) - 4(p-4)\right)\right]dt\right]\\
&= E\left[\int_1^\infty \left(2\frac{\phi(t\|X\|^2)\phi'(t\|X\|^2)}{\|X\|^2} - \frac{4(p-4)\phi'(t\|X\|^2)}{\|X\|^2}\right)dt\right]\\
&= 2\int_0^\infty\int_1^\infty \frac{\phi'(tx)}{x}\left(\phi(tx) - 2(p-4)\right)f_p(x,\lambda)dt\,dx\\
&= 2\int_0^\infty\int_1^\infty \frac{\phi'(w)}{w}\left(\phi(w) - 2(p-4)\right)f_p(w/t,\lambda)dt\,dw\\
&= 2\int_0^\infty\int_0^w \phi'(w)\frac{1}{y^2}\left(\phi(w) - 2(p-4)\right)f_p(y,\lambda)dy\,dw.
\end{aligned}$$

It is noted that $f_p(t,\lambda)$ and $f_p(t)$ designate a density of $\chi^2_p(\lambda)$ with non-centrality $\lambda = \|\theta\|^2/2$ and a density of χ^2_p, respectively. For the second term of the right-hand side of the extreme equation in (6.2), applying an integration by parts gives

$$\int_0^\infty \phi''(w)f_p(w,\lambda)dw = -\int_0^\infty \phi'(w)\frac{d}{dw}f_p(w,\lambda)dw.$$

Hence the second term is expressed by

$$\int_0^\infty \phi'(w)\left[8\frac{d}{dw}f_p(w,\lambda) - \frac{4(p-4)}{w}f_p(w,\lambda)\right]dw.$$

Hence we get

$$R(\delta^{JO},\theta) - R(\delta_\phi,\theta) = 2\int_0^\infty \phi'(w)\left[(\phi(w)-2(p-4))\int_0^w y^{-2} f_p(y,\lambda)dy \right. \\ \left. -2(p-4)f_p(w,\lambda)/w + 4\frac{d}{dw}f_p(w,\lambda)\right] dw. \tag{6.3}$$

It is here noted that the following inequalities hold:

$$f_p(w,\lambda)/\int_0^w \frac{f_p(y,\lambda)}{y^2}dy \geq f_p(w)/\int_0^w \frac{f_p(y)}{y^2}dy \tag{6.4}$$

and

$$\frac{d}{dw}f_p(y,\lambda)/f_p(y,\lambda) \geq \frac{d}{dw}f_p(y)/f_p(y) = (p/2-1)/w - 1/2. \tag{6.5}$$

Combining (6.3), (6.4) and (6.5), we see that $R(\delta^{JO},\theta) - R(\delta_\phi,\theta) \geq 0$ for every θ, if

1. $\phi(w)$ is nondecreasing and $\lim_{z\to\infty}\phi(z) = 2(p-4)$,

2. for $\phi(w) - 2(p-4) < 0$,

$$(\phi(w)-2(p-4))\int_0^w 1/y^2 f_p(y)dy/f_p(y) - 2(p-4)/w + (2p-4)/w - 2 \geq 0,$$

which are guaranteed by the conditions of this theorem and the proof is complete. □

The following result is a simple and useful consequence of Theorem 6.1.

Corollary 6.1.1. *The estimator $\delta_\phi(X)$ is superior to $\delta^{JO}(X)$ if*

1. *$\phi(w)$ is nondecreasing and $\lim_{z\to\infty}\phi(z) = 2(p-4)$,*

2. *$\phi(w) \geq \min(2(p-4), p(p-4)/(p-2)w)$.*

The class of improved estimators given by Corollary 6.1.1 include nonnegative-part Johnstone's estimator $\delta_0^{JO}(X) = \max(0, p - 2(p-4)/\|X\|^2)$ and positive estimators

$$\delta_a^{JO}(X) = \max\left[a, p - \frac{2(p-4)}{\|X\|^2}\right] \quad \text{for} \quad 0 \leq a \leq 2\frac{p}{p-2}.$$

proof. Our task is only to show that $\phi_0(w) \le p(p-4)/(p-2)w$. By applying an integration by parts, $\phi_0(w)$ is rewritten as

$$\begin{aligned}
\phi_0(w) &= 2(p-4) + (2-4/w)\frac{w^{p/2-1}e^{-w/2}}{\int_0^w y^{p/2-3}e^{-y/2}dy} \\
&= 2\frac{(p-4)\int_0^w y^{p/2-3}e^{-y/2}dy + (1-2/w)w^{p/2-1}e^{-w/2}}{\int_0^w y^{p/2-3}e^{-y/2}dy} \\
&= 2\frac{\int_0^w y^{p/2-2}e^{-y/2}dy}{\int_0^w y^{p/2-3}e^{-y/2}dy} + 2\frac{w^{p/2-1}e^{-w/2}}{\int_0^w y^{p/2-3}e^{-y/2}dy}.
\end{aligned}$$

Noting the following two inequalities:

$$\int_0^w y^{p/2-3}e^{-y/2}dy \ge e^{-w/2}\int_0^w y^{p/2-3}dy = \frac{2}{p-4}w^{p/2-2}e^{-w/2}$$

and

$$\frac{1}{w}\frac{\int_0^w y^{p/2-2}e^{-y/2}dy}{\int_0^w y^{p/2-3}e^{-y/2}dy} = \frac{\int_0^1 t^{p/2-2}e^{-wt/2}dt}{\int_0^1 t^{p/2-3}e^{-wt/2}dt} \le \frac{\int_0^1 t^{p/2-2}dt}{\int_0^1 t^{p/2-3}dt} = \frac{p-4}{p-2},$$

we get the required inequality

$$\phi_0(w) \le 2\frac{p-4}{p-2}w + (p-4)w = p\frac{p-4}{p-2}w.$$

□

Table 6.1 reports the results of simulation for the "relative risk improvement" which is defined by

$$RRI(\delta) = 100 \times \{R(\|\theta\|^2/p,\ \delta^{JO}) - R(\|\theta\|^2/p,\ \delta)\}/R(\|\theta\|^2/p,\ \delta^{JO}),$$

where $\delta = \delta_0^{JO},\ \delta_{p/(p-2)}^{JO},\ \delta_{2p/(p-2)}^{JO}$. We see that if one has the prior knowledge that $\|\theta\|^2$ nearly equals zero, then δ_0^{JO} may be desirable. When one has no information about $\|\theta\|^2$, positive estimators, for example $\delta_{p/(p-2)}^{JO}$ or $\delta_{2p/(p-2)}^{JO}$, should be chosen. Moreover, as seen in the case of $p = 10$, improvements in risk over δ^{JO} are negligible for large values of p.

Table 6.1 : The relative risk improvements in estimation of the loss function for $p = 5, 6, 10$

p	$\|\theta\|^2/p$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
5	δ_0^{JO}	4.92	2.25	1.08	1.06	0.93	0.36	0.24	0.18	0.12	0.10	0.02
	$\delta_{p/(p-2)}^{JO}$	4.89	2.31	1.16	1.14	1.00	0.41	0.27	0.20	0.14	0.12	0.03
	$\delta_{2p/(p-2)}^{JO}$	3.50	1.84	1.11	1.27	1.18	0.59	0.42	0.32	0.23	0.19	0.08
6	δ_0^{JO}	2.06	0.99	1.13	0.47	0.43	0.15	0.09	0.04	0.03	0.01	0.00
	$\delta_{p/(p-2)}^{JO}$	2.06	1.05	1.21	0.53	0.48	0.18	0.12	0.06	0.05	0.01	0.01
	$\delta_{2p/(p-2)}^{JO}$	1.14	0.85	1.29	0.68	0.62	0.29	0.22	0.13	0.10	0.05	0.02
10	δ_0^{JO}	0.07	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$\delta_{p/(p-2)}^{JO}$	0.06	0.05	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$\delta_{2p/(p-2)}^{JO}$	0.06	0.05	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00

Now we extend the above results to the case where the covariance matrix of X is $\sigma^2 I_p$ for unknown σ^2. Let X and S be independent random variables where X has a p-variate normal distribution $N_p(\theta, \sigma^2 I_p)$ and S/σ^2 has a chi square distribution χ_n^2 with n degrees of freedom. When X is used to estimate θ under the loss $L(d, \theta, \sigma^2) = \|d - \theta\|^2/\sigma^2$, it is of great interest in estimating the loss $\|X - \theta\|^2/\sigma^2$ from the data using rule $\delta(X, S)$. For evaluating the loss estimator $\delta(X, S)$, the quadratic loss $(\delta(X, S) - \|X - \theta\|^2/\sigma^2)^2$ is exploited.

We consider a class of estimators $\delta_\phi(X, S) = p - \phi(Z)/Z$ for $Z = \|X\|^2/S$. By the Stein and the chi-square identities, we get the unbiased estimator of risk $R(\delta_\phi, \theta, \sigma^2)$, which is written as

$$\begin{aligned} R(\delta_\phi, \theta, \sigma^2) &= E\left[\left(p - \frac{\phi(Z)}{Z} - \frac{\|X - \theta\|^2}{\sigma^2}\right)^2\right] \\ &= 2p + E\left[\frac{\phi^2(Z)}{Z^2} - \frac{4(p-4)}{n+2}\frac{\phi(Z)}{Z^2}\right. \\ &\qquad \left. + \frac{4n(p+n-2)}{n+2}\frac{\sigma^2}{\|X\|^2}\phi'(Z) - 4\frac{\phi'(Z)}{Z}\right]. \end{aligned}$$

Therefore we can see that, in the class of $\delta_c(X, S) = p - c/Z$, $\delta_c(X, S)$ with $0 < c < 4(p-4)/(n+2)$ is superior to the unbiased estimator $\delta_U(X, S) = p$ for $p \geq 5$ and

$\delta_c^*(X,S) = \delta_c(X,S)$ for $c^* = 2(p-4)/(n+2)$ is optimal in this class. Moreover, the estimator $\delta_c^*(X,S)$ has the drawback that it takes negative values with a positive probability. Here, in the same way as Theorem 6.1, modified estimators improving on $\delta_c^*(X,S)$ are obtained.

Theorem 6.2. *The estimator $\delta_\phi(X,S)$ improves on $\delta_c^*(X,S)$ if*

1. *$\phi(z)$ is nondecreasing and $\lim_{z\to\infty}\phi(z) = 2(p-4)/(n+2)$,*
2. *$\phi(z) \geq \min(2(p-4)/(n+2), \phi_0(z))$ for every z, where,*

$$\phi_0(z) = 2(p-4)/(n+2) + 2\frac{nz-2}{z(n+2)}\frac{z^{p/2-1}(z+1)^{-(p+n)/2}}{\int_0^z t^{p/2-3}(t+1)^{-(p+n)/2}dt}.$$

proof. Using Kubokawa's method gives

$$\begin{aligned} R(\delta_c^*,\theta,\sigma^2) - R(\delta_\phi,\theta,\sigma^2) &= 2\int_0^\infty \phi'(z)\left[\left(\phi(z) - 2\frac{p-4}{n+2}\right)\int_0^z t^{-2}h(t,\lambda)dt \right. \\ &\left. + \frac{2}{z}h(z,\lambda) - 2n\frac{p+n-2}{n+2}\frac{1}{z}\int_0^\infty f_p(vz,\lambda)f_n(v)dv\right]dw, \end{aligned}$$

where $h(t,\lambda) = \int_0^\infty v f_p(vt,\lambda)f_n(v)dv$. The following inequalities:

$$\int_0^z t^{-2}h(t,\lambda)dt/h(z,\lambda) \leq \int_0^z t^{-2}h(t)dt/h(z),$$

$$\int_0^\infty f_p(vz,\lambda)f_n(v)dv/h(z,\lambda) \leq \int_0^\infty f_p(vz)f_n(v)dv/h(z) = \frac{z+1}{p+n-2}$$

and some calculation yield the condition of the theorem. □

Furthermore, in the same way as the proof of Collorary 6.1.1, the inequality $\phi_0(z) \leq p(p-4)z/(p-2)$ is derived, which implies that

$$\delta_a^{TR}(X,S) = \max\left[a, p - \frac{2(p-4)}{n+2}\frac{S}{\|X\|^2}\right] \quad \text{for} \quad 0 \leq a \leq 2\frac{p}{p-2}$$

improves on $\delta_c^*(X,S)$.

Through the simulation experiment, we can see that the relative risk improvement of $\delta_0^{TR}(X,S)$ and $\delta_a^{TR}(X,S)$ for $a = p/(p-2),\ 2p/(p-2)$ have similar performance to Table 6.1, regardless of values of n, although the details are omitted. In the sequel, positive estimators should be chosen as good procedures when when one has no information about $\|\theta\|^2/\sigma^2$.

Chapter 7

Conclusion

In this chapter, we emphasize our original results in this thesis.

We mainly consider the estimation of a multivariate normal mean and try to find a broader class of admissible minimax estimators. In Section 3.3.2 and 3.3.3, we see that all classes of minimax admissible estimators given in the hitherto researches, that is, Strawderman(1971), Alam(1973), Berger(1976b), Faith(1978), Maruyama(1998) and Fourdrinier *et al.*(1998) are expressed as δ_h although relations among classes above have not been clear. See Table 3.1. We not only summarize the hitherto researches but also show that Fourdrinier *et al.*'s class includes estimators whose shrinkage factor $\phi_h(w)$ is not monotone and has one extremum as shown in Corollary 3.21.1. Moreover we present more simple proof of Maruyama(1998)'s result, Theorem 3.26.

In Section 3.3.4, we pay attention to Stein(1973)'s simple idea and provide it with the warrant for statistical decision theory. Stein(1973) suggested that the generalized Bayes estimator with respect to the prior distribution: the weighted sum of that determined by the density $\|\theta\|^{2-p}$ and a measure concentrated at the origin, may dominate the James-Stein estimator. The generalized Bayes estimator has, however, not-monotone $\phi(w)$ as shown in Efron and Morris(1976), so that even minimaxity, to say nothing of dominance over the James-Stein estimator has not been proved yet. Extending Stein's prior distribution, we consider a prior distribution: the weighted sum of a measure concentrated at the origin and the prior distribution considered in Section 3.3.2 and

3.3.3. In Theorem 3.32, We derive analytically the sufficient condition for minimaxity on the ratio of a measure concentrated at the origin. Moreover, in Theorem 3.31 we show that this class of generalized Bayes estimators includes variety of not-monotone $\phi(w)$, such that

- $\phi(w)$ is increasing from the origin to a certain point and is decreasing from the point. Namely, $\phi(w)$ has one extreme point.
- the function $\phi(w)$ is increasing from the origin to a certain point w_1 and is decreasing from the point w_1 to the other point w_2 and is increasing from the point w_2. Namely, $\phi(w)$ has two extreme points.

Hence the relations among minimaxity, the prior distribution, and the behavior of $\phi(w)$, are quite clear although the researches on the relations, especially in the case where $\phi(w)$ is not monotone, have been fragmentary and have not been arranged yet. We believe that the investigation of estimators with not-monotone $\phi(w)$ is more and more important because such estimators are considered as candidates for the solution of the most difficult problem in this field, that is, the problem of finding admissible estimators dominating the James-Stein positive-part rule.

In Section 3.4, the problem of finding admissible estimators which dominate the James-Stein estimator is treated. We show that an estimator which is proposed in Maruyama(1996) and improves upon the James-Stein estimator, is permissible. We unfortunately conjecture that it is inadmissible, which implies that it is quite difficult to find a class of admissible estimators which improve upon the James-Stein estimator.

In Chapter 4, the problem of estimating the mean vector of scale mixtures of multivariate normal distributions is considered. For a certain subclass of these distributions, which includes at least multivariate-t distributions, we derive minimax admissible estimators although, in non-normal case, an estimator satisfying both minimaxity and admissibility, has not been derived yet.

Appendix A

Technical results

The proof of Theorem 3.9 requires certain technical results which we present in this chapter. All results in this appendix are due to Brown(1971) and Srinivasan(1981). Let u be a bounded piecewise differential functions defined on R^p. The following result is a rather standard one.

Lemma A.1. *There exists a constant K_0 such that*

$$\begin{aligned}\int (u(\theta) - u(x))^2 p_\theta(x)dx &\leq K_0 \left[\int \|\nabla u(x)\|^2 \|x - \theta\|^{-p+1} p_\theta(x)dx \right. \\ &\quad \left. + \int \|\nabla u(x)\|^2 p_\theta(x)dx \right].\end{aligned}$$

proof. Let $r = \|x - \theta\|$ and let φ denote the usual orthogonal angular coordinates in R^p around the point θ. (φ is a $(p-1)$ vector.) In short $(r, \varphi) = (r(x), \varphi(x))$ are spherical coordinates around the point θ. For covenience we normalize φ so that $\int_{\|x\|<1} dx = \int r^{p-1} dr d\varphi$. Moreover for convenience, let u_s denote u expressed in terms of these coordinates; i.e. $u(x) = u_s(r(x), \varphi(x))$. By Schwartz inequality, we have

$$\begin{aligned}(u(\theta) - u(x))^2 &= \left(\int_0^{r(x)} \|\nabla u_s(s, \varphi(x))\| ds \right)^2 \\ &\leq r(x) \int_0^{r(x)} \|\nabla u_s(s, \varphi(x))\|^2 ds.\end{aligned}$$

Therefore we have

$$\begin{aligned}\int (u(\theta)-u(x))^2 p_\theta(x)dx &\le \int\int r\left(\int_0^r \|\nabla u_s(s,\varphi)\|^2 ds\right)\exp(-r^2/2)r^{p-1}drd\varphi \\ &= \int\int_0^\infty \|\nabla u_s(s,\varphi)\|^2 \int_s^\infty r^p\exp(-r^2/2)drdsd\varphi\end{aligned}$$

Integrating by parts gives

$$\int_s^\infty r^p\exp(-r^2/2)dr \le K_0(s^{p-1}+1)\exp(-s^2/2)$$

for some constant $K_0>0$. Hence we have

$$\begin{aligned}\int (u(\theta)-u(x))^2 p_\theta(x)dx &\le K_0\left[\int_0^\infty \|\nabla u_s(s,\varphi)\|^2 s^{p-1}\exp(-s^2/2)dsd\varphi\right. \\ &\quad \left.+\int_0^\infty \|\nabla u_s(s,\varphi)\|^2\exp(-s^2/2)dsd\varphi\right] \\ &= K_0\left[\int \|\nabla u(x)\|^2\|x-\theta\|^{-p+1}p_\theta(x)dx\right. \\ &\quad \left.+\int \|\nabla u(x)\|^2 p_\theta(x)dx\right].\end{aligned}$$

We have proved the theorem. □

Lemma A.2. *Let ρ be a constant such that $0<\rho<1/2$. Suppose that $u(x)$ is a bounded piecewise differentiable function satisfying $L_f u(x)=0$ for $1<\|x\|<r$, $u(x)\equiv 1$ for $\|x\|\le 1$ and $u(x)\equiv c$, $0<c<1$ for $\|x\|>r$ for some $r>2+\rho$. Then there exists a constant $K_1>0$ such that*

$$\int_{\|x-\theta\|\le\rho}(u(x)-u(\theta))^2dx \le K_1\int_{\|x-\theta\|\le K_1}\|\nabla u(x)\|^2dx.$$

proof. Let $u(x)$ be as given in the lemma. Then it follows from Brown(1971) that $u(x)$ has mean value property. More precisely for any given θ, there exists $q_\theta(\eta)>0$ such that $u(\theta)=\int u(\eta)q_\theta(\eta)d\eta$. Moreover $q_\theta(\eta)=0$ if $\|\theta-\eta\|>1$, $\int q_\theta(\eta)d\eta=1$ and

$q_\theta(\eta) < K_2$ for some constant K_2. Then we have

$$\begin{aligned}
\int_{\|x-\theta\|\le\rho} (u(x)-u(\theta))^2 dx &= \int_{\|x-\theta\|\le\rho} \left(u(x) - \int u(\eta) q_\theta(\eta) d\eta\right)^2 dx \\
&\le \int \left(\int_{\|x-\theta\|\le\rho} (u(x)-u(\eta))^2 dx \right) q_\theta(\eta) d\eta \\
&\le \int \left(\int_{\|x-\eta\|\le 1+\rho} (u(x)-u(\eta))^2 dx \right) q_\theta(\eta) d\eta. \quad \text{(A.1)}
\end{aligned}$$

The last line (A.1) follows from the fact $\{x : \|x-\theta\| \le \rho\} \subset \{x : \|x-\eta\| \le 1+\rho\}$ for any η such that $\|\eta-\theta\| \le 1$. In the similar way as the proof of Lemma A.1, we have

$$\begin{aligned}
\int_{\|x-\eta\|<1+\rho} (u(x)-u(\eta))^2 dx &\le \int_0^{1+\rho} v^p \left(\int_0^v \|\nabla u_s(s,\varphi)\|^2 ds \right) dv d\varphi \\
&= \int_0^{1+\rho} \|\nabla u_s(s,\varphi)\|^2 \left(\int_s^{1+\rho} v^p dv \right) ds d\varphi \\
&\le (1+\rho)^{m+1} \int_0^{1+\rho} \|\nabla u_s(s,\varphi)\|^2 s^{-p+1} s^{p-1} ds d\varphi \\
&= (1+\rho)^{m+1} \int_E \|\nabla u(x)\|^2 \|x-\eta\|^{-p+1} dx, \qquad \text{(A.2)}
\end{aligned}$$

where $E = \{x : \|x-\eta\| \le 1+\rho\}$. Therefore, bounding the second integral in (A.1) using (A.2) and the upper bound of $q_\theta(\eta)$, we have

$$\begin{aligned}
\int_{\|x-\theta\|<\rho} (u(x)-u(\theta))^2 dx &\le (1+\rho)^{m+1} K_2 \int_{\|\theta-\eta\|\le 1} \\
&\quad \left(\int_{\|x-\theta\|\le 3+\rho} \|\nabla u(x)\|^2 \|x-\eta\|^{-p+1} dx \right) d\eta \\
&\le cK_2(1+\rho)^{m+1} \int_{\|x-\theta\|\le 3+\rho} \|\nabla u(x)\|^2 dx,
\end{aligned}$$

where

$$c = \int_{\|z\|\le 1} \|z\|^{1-p} dz \ge \int_{\|\theta-\eta\|\le 1} \|x-\eta\|^{-p+1} d\eta.$$

Letting $K_1 = cK_2(1+\rho)^{m+1} + 4$ the lemma follows. □

Lemma A.3. *Let $u(x)$ satisfy the condition of Lemma A.2. Then there exists a constant $K_3 > 0$ such that*

$$\int (u(x) - u(\theta))^2 p_\theta(x) dx \leq K_3 \int \|\nabla u(x)\|^2 p_\theta(x) dx. \tag{A.3}$$

proof. By Lemma A.1, we have

$$\begin{aligned}
\int_{\|x-\theta\|>\rho} (u(x) - u(\theta))^2 p_\theta(x) dx &\leq C_1 \int_{\|x-\theta\|>\rho} \|\nabla u(x)\|^2 \|x-\theta\|^{-p+1} p_\theta(x) dx \\
&\quad + C_1 \int \|\nabla u(x)\|^2 p_\theta(x) dx \\
&\leq C_2 \int \|\nabla u(x)\|^2 p_\theta(x) dx
\end{aligned} \tag{A.4}$$

where $C_2 = 2C_1\rho^{-p+1}$. For $0 < \rho < 1/2$, we have by Lemma A.2

$$\begin{aligned}
\int_{\|x-\theta\|\leq\rho} (u(x) - u(\theta))^2 p_\theta(x) dx &\leq (2\pi)^{-p/2} \int_{\|x-\theta\|\leq\rho} (u(x) - u(\theta))^2 dx \\
&\leq (2\pi)^{-p/2} K_1 \int_{\|x-\theta\|\leq K_2} \|\nabla u(x)\|^2 dx \\
&\leq K_1 \exp(K_1^2/2) \int_{\|x-\theta\|\leq K_2} \|\nabla u(x)\|^2 p_\theta(x) dx \\
&\leq K_1 \exp(K_1^2/2) \int \|\nabla u(x)\|^2 p_\theta(x) dx.
\end{aligned} \tag{A.5}$$

Hence letting $K = K_1 \exp(K_1^2/2) + C_2$, we have the lemma. □

The next lemma is an extension of Lemma A.3.

Lemma A.4. *Let $u(x)$ satisfy the condition of Lemma A.2. Then there exists a constant $K_4 > 0$ such that*

$$\int (u(x) - u(\theta))^2 \|x-\theta\|^2 p_\theta(x) dx \leq K_4 \int \|\nabla u(x)\|^2 \left(\int_{\|\xi\|<K_4} p_\theta(x+\xi) d\xi \right) dx$$

proof. We can show this result by using an argument similar to Lemmas A.1, A.2 and A.3 and noting the following technical result shown in Brown(1971).

Lemma A.5. *For a positive constant α, there exists a constant K_5 depending only upon K_5 and p such that*

$$\exp(\alpha\|x-\theta\|)p_\theta(x) \leq K_5 \int_{\|\xi\|<K_5} p_\theta(x+\xi)d\xi.$$

□

Bibliography

[1] Abramowitz, M and Stegun, I.A. (1964). *Handbook of Mathematical Functions*, Dover Publications, New York.

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

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

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

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

[6] Berger, J. (1976a). Tail minimaxity in location vector problems and its applications. *Ann. Statist.*, **4**, 33-50.

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

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

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

[10] Birnbaum, A. (1955). Characterization of complete classes of tests of some multiparametric hypotheses with applications to likelihood ratio tests. *Ann. Math. Statist.*, **22**, 22-42.

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

[12] Bock, M.E. (1988). Shrinkage estimators: pseudo-Bayes rules for normal mean vectors. In *Statistical Decision Theory and Related Topics IV* (S.S. Gupta, J.O. Berger, eds.), 281-297, Springer-Verlag, Newyork.

[13] Brandwein, A. and Strawderman. W.E. (1978). Minimax estimation of location parameters for spherically symmetric unimodal distributions under quadratic loss. *Ann. Statist.* , **6**, 377-416.

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

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

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

[17] Brown, L.D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.*, **42**, 855-903.

[18] Brown, L.D. (1988). The differential inequality of a statistical estimation problem. In *Statistical Decision Theory and Related Topics IV* (S.S. Gupta, J.O. Berger, eds.), 299-324, Springer-Verlag, Newyork.

[19] DasGupta, A. and Strawderman. W.E. (1997). All estimates with a given risk, Riccaci differential equations and a new proof of a theorem of Brown. *Ann. Statist.*, **25**, 1208-1221.

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

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

[22] Faith, R.E. (1978). Minimax Bayes estimators of a multivariate normal mean. *J. Multivariate Anal.*, **8**, 372-379.

[23] Farrell, R. (1968). On a necessary and sufficient condition for admissibility of estimators when strictly convex loss is used. *Ann. Math. Statist.*, **38**, 23-28.

[24] Fortuin, C.M., Kasteleyn, P.W. and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22**, 89-103.

[25] Fourdrinier, D., Strawderman. W.E. and Wells, T. (1998). On the construction of Bayes minimax estimators. *Ann. Statist.*, **26**, 660-671.

[26] George, E.I. (1990). Comment on "Developments in decision theoretic variance estimation", by Maatta, J.M. and Casella, G. *Statist. Sci.*, **5**, 107-109.

[27] Ghosh, M.(1994). On some Bayesian solutions of the Neyman-Scott problem. In *Statistical Decision Theory and Related Topics V*, 267-276.

[28] Hara, H. (1997). unpublished manuscript.

[29] James, W. and Stein, C. (1961). Estimation of quadratic loss. *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, Vol.1, 361-379, Univ. of California Press, Berkeley.

[30] Kiefer, J.(1957). Invariance, minimax sequential estimation and continuous time processes. *Ann. Math. Statist.*, **28**, 573-601.

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

[32] Kubokawa, T.(1994). An unified approach to improving equivariant estimators. *Ann. Statist.*, **22**, 290-299.

[33] Kubokawa, T.(1998). The Stein phenomenon in simultaneous estimation: a review. *Applied Statistical Science III* (eds. S.E. Ahmed, M. Ahsanullah and B.K. Sinha), 143-173, NOVA Science Publishers, Inc., New York.

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

[35] Lehmann,E.L. (1986). *Testing Statistical Hypotheses. 2nd Ed.*, Wiley, New York.

[36] Lehmann, E.L. and Casella, G.(1999). *Theory of Point Estimation. 2nd Ed.*, Springer, New York.

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

[38] Maruyama, Y. (1996). Estimation of the mean vector and the variance of a multivariate normal distribution. Master thesis of graduate school of economics, University of Tokyo.

[39] Maruyama, Y. (1997). A New Positive Estimator of Loss Function. *Statistics & Probability Letters*, **36**, 269-274.

[40] Maruyama, Y. (1998). A unified and broadened class of admissible minimax estimators of a multivariate normal mean. *J. Multivariate Anal.*, **64**, 196-205.

[41] Maruyama, Y. (1999a). Minimax estimators of a normal variance. *Metrika*, **48**, 209-214.

[42] Maruyama, Y. (1999b). Improving on the James-Stein Estimator. *Statistics & Decisions*, **17**, 137-140.

[43] Maruyama, Y. (2000a). Correction to "A unified and broadened class of admissible minimax estimators of a multivariate normal mean". to appear in *J. Multivariate Anal.*

[44] Maruyama, Y. (2000b). Admissible minimax estimators of a mean vector of scale mixtures of multivariate normal distributions. submitted.

[45] Maruyama, Y. and Iwasaki, K. (2000). unpublished manuscript.

[46] Meyers, N. and Serrin, J. (1960). The exterior problem for second order elliptic equations. *J. Math. and Mech.*, **9**, 513-539.

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

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

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

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

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

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

[53] 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.

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

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

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

[57] Strawderman, W.E. and Cohen, A. (1971). Admissibility of estimators of the mean vector of multivariate normal distribution with quadratic loss. *Ann. Math. Statist.*, **42**, 270-296.

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

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