Ann. Inst. Statist. Math.
Vol. 51, No. 1, 111–124 (1999)

LINEAR VERSUS NONLINEAR RULES FOR MIXTURE NORMAL PRIORS*

BRANI VIDAKOVIC

Institute of Statistics and Decision Sciences, Duke University, Box 90251, Durham, NC 27708-0251, U.S.A.

(Received March 10, 1997; revised September 29, 1997)

Abstract. The problem under consideration is the Γ-minimax estimation, under L_2 loss, of a multivariate normal mean when the covariance matrix is known. The family Γ of priors is induced by mixing zero mean multivariate normals with covariance matrix τI by nonnegative random variables τ, whose distributions belong to a suitable family $\mathcal{G}$. For a fixed family $\mathcal{G}$, the linear (affine) Γ-minimax rule is compared with the usual Γ-minimax rule in terms of corresponding Γ-minimax risks. It is shown that the linear rule is "good", i.e., the ratio of the risks is close to 1, irrespective of the dimension of the model. We also generalize the above model to the case of nonidentity covariance matrices and show that independence of the dimensionality is lost in this case. Several examples illustrate the behavior of the linear Γ-minimax rule.

Key words and phrases: Affine rules, efficiency, Bayes risk, Brown's identity, Γ-minimax rules.

1. Introduction

Partial prior information can be well formalized and leads naturally to the description of a class of priors Γ that forms the basis for the Γ-minimax approach (Skibinsky and Cote (1962), Kudō (1967)). If prior information is scarce, the class Γ of priors under consideration is large, and we are close to the usual minimax principle. The extreme case is when no information is available, in which case the Γ-minimax setup is the minimax setup.

On the other hand, if we have a lot of prior information, then the class Γ is sparse. An extreme case is a class Γ that contains only one prior. In this case, the Γ-minimax framework becomes the minimum Bayes risk framework. The spirit of Γ-minimaxity is vividly expressed by the often quoted sentence of Efron and Morris (1971):

> ... We have referred to the "true prior distribution" ... but in realistic situations there is seldom any one population or corresponding prior

* The research was supported in part by NSF Grant DMS-9626159 at Duke University.

distribution that is "true" in an absolute sense. There are only more or less relevant priors, and Bayesian statistician chooses among those as best he can, compromising between his limited knowledge of subpopulation distributions and what is usually an embarrassingly large number of identifying labels attached to the particular problem.

Some Bayesians object that working within the context of Γ-minimaxity may produce "demonstrable incoherence," since there are examples in which the Γ-minimax rule is not Bayes (Watson (1974)). But in most cases of interest, the Γ-minimax rule is the Bayes rule with respect to some prior from the family Γ.

For a nice discussion on Γ-minimaxity in the context of Bayesian robustness, we refer the reader to Berger (1985).

Consider the following model:

$$\begin{aligned} &\boldsymbol{X} \mid \boldsymbol{\theta} \sim \mathcal{MVN}_p(\boldsymbol{\theta}, I), \\ &\boldsymbol{\theta} \mid \tau \sim \mathcal{MVN}_p(\mathbf{0}, \tau I), \\ &\tau \sim G(t), (\tau \geq 0). \end{aligned} \tag{1.1}$$

Let $\mathcal{G}$ be the family of distribution functions G. Suppose that the random variables τ have uniformly bounded expectations, and that $\exists \tau$ s.t. $E\tau > 0$, i.e.

$$0 < \sup_{G \in \mathcal{G}} \int t dG(t) < \infty. \tag{1.2}$$

The class $\mathcal{G}$ determines the family of priors Γ as

$$\Gamma = \left\{ \int \phi_{p,t}(\boldsymbol{\theta}) dG(t), G \in \mathcal{G} \right\} \tag{1.3}$$

where

$$\phi_{p,t}(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{p/2} t^{p/2}} e^{-\|\boldsymbol{\theta}\|^2/2t}$$

is the density of the $\mathcal{MVN}_p(\mathbf{0}, tI)$ distribution.

The class (1.3) is an example of a "scale mixture of normals" or a "normal scale mixtures" family. Some of the well-known families of distributions, such as: the t (particularly the Cauchy), logistic, double-exponential, $\cosh^{-1}$, etc., can be expressed as appropriate scale mixtures of normals. These classes are attractive both because they are easy to work with (studies in which the Gibbs sampler, Monte Carlo method, and Bayesian calculations are used), and because they possess other desirable properties (e.g. in robust Bayesian inference). For accounts of the significance of scale mixture of normals in statistics, we refer the reader to Efron and Olshen (1978), West (1987), Robert (1990), DasGupta *et al.* (1990), and DasGupta (1992). Model (1.1) was also considered by Faith (1978) in the context of James-Stein estimation of a multivariate normal mean.

Having a family of priors, it seems natural to employ Γ-minimax to estimate the unknown parameter $\boldsymbol{\theta}$. Let $\mathcal{D}$ be the set of all measurable decision rules. The estimator $\delta^\star \in \mathcal{D}$ that minimizes $\sup_{\pi \in \Gamma} r(\pi, \delta)$, i.e.

$$\inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{\pi \in \Gamma} r(\pi, \delta^\star),$$

is the Γ-minimax rule, and $r_\Gamma = \sup_{\pi \in \Gamma} r(\pi, \delta^\star)$ is the corresponding Γ-minimax risk.

If we consider the set of linear decision rules $\mathcal{D}_L$, then the rule $\delta_L^\star \in \mathcal{D}_L$ for which

$$\inf_{\delta \in \mathcal{D}_L} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{\pi \in \Gamma} r(\pi, \delta_L^\star),$$

is called the Γ-minimax linear rule, and $r_L = \sup_{\pi \in \Gamma} r(\pi, \delta_L^\star)$ is the linear Γ-minimax risk.

We are interested in the performance of linear Γ-minimax rules compared to general Γ-minimax rules. Performance will be measured through the ratio $\rho = \frac{r_L}{r_\Gamma}$. More precisely, for a prespecified class $\mathcal{G}$ of hyperprior distributions G, we have an induced class Γ. For the class Γ, we want to calculate ρ or at least an upper bound on ρ, say $\rho^\star$. Values of ρ close to 1 suggest good performance of the linear Γ-minimax rule.

When the model is $X \sim \mathcal{N}(\theta, 1)$ and Γ is the family of *all* distributions on $[-m, m]$ (the bounded normal mean), Ibragimov and Has'minskii (1984) argued that ρ is finite. Donoho *et al.* (1990) have derived an upper bound $\rho \leq 1.25$ that holds uniformly in m. When Γ is the family of all unimodal and symmetric distributions on $[-m, m]$, Vidakovic and DasGupta (1996) have shown that $\rho \leq 1.074$.

In the multivariate case, we think about a linear rule as an affine transformation $\delta(\boldsymbol{x}) = A\boldsymbol{x} + B$, for some matrices A and B. Solomon (1972) has shown that $B \neq \mathbf{0}$ is an inadmissible choice. The motivation for using linear rules is apparent: they are easily calculable and simple to use. The method used in calculating $\rho^\star$ will not generalize to non-normal models. This is because Brown's identity employed in evaluating $\rho^\star$ (Subsection 2.1) is valid only in normal-location setup under the sqaured-error loss.

2. Preliminaries

2.1 *An information integral inequality*

First, we will prove an inequality involving the Fisher information integral,

$$\mathcal{I}(f(\boldsymbol{x})) = \int_{\Re^p} \frac{\sum_{i=1}^p \left(\frac{\partial}{\partial x_i} f(\boldsymbol{x}) \right)^2}{f(\boldsymbol{x})} d\boldsymbol{x}, \tag{2.1}$$

which will be helpful in calculating a bound on r_Γ.

The integral in (2.1) is the trace of the Fisher information matrix $\boldsymbol{I}(\boldsymbol{\theta})$ for a location parameter $\boldsymbol{\theta}$ in $f(\boldsymbol{x} - \boldsymbol{\theta})$. The lemma that follows can be stated and proven in much more generality. Let G be an arbitrary distribution function.

LEMMA 2.1. *Let $\phi_{p,t}(\boldsymbol{x})$ be the density of the $\mathcal{MVN}_p(\mathbf{0}, tI)$. Then*

$$\mathcal{I}\left(\int \phi_{p,t+1}(\boldsymbol{x}) dG(t) \right) \leq \int \mathcal{I}(\phi_{p,t+1}(\boldsymbol{x})) dG(t).$$

PROOF. By the Cauchy-Schwartz inequality

$$\frac{\left(\int \frac{\partial}{\partial x_i}\phi_{p,t+1}(\boldsymbol{x})dG(t)\right)^2}{\int \phi_{p,t+1}(\boldsymbol{x})dG(t)} \le \int \frac{\left(\frac{\partial}{\partial x_i}\phi_{p,t+1}(\boldsymbol{x})\right)^2}{\phi_{p,t+1}(\boldsymbol{x})}dG(t). \tag{2.2}$$

By taking the derivative outside the integral sign and summing with respect to i in (2.2), we obtain

$$\frac{\sum_{i=1}^p \left(\frac{\partial}{\partial x_i}\int \phi_{p,t+1}(\boldsymbol{x})dG(t)\right)^2}{\int \phi_{p,t+1}(\boldsymbol{x})dG(t)} \le \int \frac{\sum_{i=1}^p \left(\frac{\partial}{\partial x_i}\phi_{p,t+1}(\boldsymbol{x})dG(t)\right)^2}{\phi_{p,t+1}(\boldsymbol{x})}dG(t).$$

Finally, after interchanging the order of integration we get

$$\int_{\Re^p} \frac{\sum_{i=1}^p \left(\frac{\partial}{\partial x_i}\int \phi_{p,t+1}(\boldsymbol{x})dG(t)\right)^2}{\int \phi_{p,t+1}(\boldsymbol{x})dG(t)}d\boldsymbol{x}$$
$$\le \iint_{\Re^p} \frac{\sum_{i=1}^p \left(\frac{\partial}{\partial x_i}\phi_{p,t+1}(\boldsymbol{x})\right)^2}{\phi_{p,t+1}(\boldsymbol{x})}d\boldsymbol{x}dG(t). \qquad \square$$

2.2 *Brown's identity*

When the model is $X \mid \theta \sim \mathcal{N}(\theta, 1)$, and the loss is squared-error, the following identity (attributed to L. Brown) holds. For *any* prior distribution π, the Bayes risk $r(\pi)$ satisfies

$$r(\pi) = 1 - \mathcal{I}(\phi_1 * \pi(x)).$$

Under such a model, the convolution $\phi_1 * \pi(x)$ can be interpreted as the marginal distribution for X. For the derivation and some applications of Brown's identity, see Brown (1971, 1986), and Bickel (1981). In the p-variate case, Brown's identity takes the form

$$r(\pi) = p - \mathcal{I}(\phi_{p,1} * \pi(\boldsymbol{x})).$$

The function $\phi_{p,1} * \pi(\boldsymbol{x})$ has an interpretation as the marginal density for $\boldsymbol{X}$ under the model $\boldsymbol{X} \mid \boldsymbol{\theta} \sim \mathcal{MVN}_p(\boldsymbol{\theta}, I)$.

Since, in general, $r_\Gamma \ge \sup_{\pi\in\Gamma} r(\pi)$, Brown's identity gives only a lower bound on Γ-minimax risk. Therefore,

$$r_\Gamma \ge \sup_\pi (p - \mathcal{I}(\phi_{p,1} * \pi(\boldsymbol{x}))).$$

However the equality holds in most regular cases.

3. Main result

Here, we determine an upper bound for the ratio of risks by calculating r_L and finding an lower bound for r_Γ.

The model (1.1) and L_2 loss are assumed. Let $\mathcal{G}$ be any family of distributions that satisfies (1.2), and let Γ be the induced family (1.3).

THEOREM 3.1.

$$\rho = \frac{r_L}{r_\Gamma} \le \frac{\sup_{G\in\mathcal{G}} \dfrac{E\tau}{1+E\tau}}{\sup_{G\in\mathcal{G}} E\dfrac{\tau}{1+\tau}}.$$

PROOF. (a) Bound on r_Γ: If the prior density is $\pi(\boldsymbol{\theta}) = \int \phi_{p,t}(\boldsymbol{\theta})dG(t)$, then under the model (1.1), elementary calculations give the marginal density for $\boldsymbol{X}$:

$$m_\pi(\boldsymbol{x}) = \int \phi_{p,t+1}(\boldsymbol{x})dG(t). \tag{3.1}$$

This fact is apparent after the algebraic transformation

$$\begin{aligned}\pi(\boldsymbol{\theta} \mid t)\phi_{p,1}(\boldsymbol{x}-\boldsymbol{\theta}) &= \frac{1}{(2\pi)^{p/2}t^{p/2}}e^{-\|\boldsymbol{\theta}\|^2/2t}\frac{1}{(2\pi)^{p/2}}e^{-\|\boldsymbol{x}-\boldsymbol{\theta}\|^2/2}\\ &= \frac{1}{(2\pi)^{p/2}\left(\dfrac{t}{t+1}\right)^{p/2}}e^{-(1/2)(\|\boldsymbol{\theta}-(t/t+1)\boldsymbol{x}\|^2/(t/(t+1)))}\\ &\quad\cdot\frac{1}{(2\pi)^{p/2}(t+1)^{p/2}}e^{-\|\boldsymbol{x}\|^2/2(t+1)}.\end{aligned}$$

Equation (3.1) is obtained by integrating out $\boldsymbol{\theta}$. Now

$$r_\Gamma \ge \sup_{\pi\in\Gamma}(p - \mathcal{I}(m_\pi(\boldsymbol{x}))) \ge \sup_{G\in\mathcal{G}}\left(p - \int \mathcal{I}(\phi_{p,t+1}(\boldsymbol{x}))dG(t)\right).$$

After the substitution $\mathcal{I}(\phi_{p,t+1}(\boldsymbol{x})) = \frac{p}{t+1}$, we obtain a lower bound on r_Γ:

$$r_\Gamma \ge \sup_{G\in\mathcal{G}}\left(p - pE\frac{1}{1+\tau}\right) = p\sup_{G\in\mathcal{G}} E\frac{\tau}{1+\tau}.$$

(b) Calculation of r_L: Notice that $E\boldsymbol{\theta} = E(E\boldsymbol{\theta} \mid \tau) = 0$, and $E\boldsymbol{\theta}\boldsymbol{\theta}' = E(E(\boldsymbol{\theta}\boldsymbol{\theta}' \mid \tau)) = E(\tau I) = (E\tau)I$. Our estimator is constrained to be affine, i.e. of the form $\delta_L(\boldsymbol{x}) = A\boldsymbol{x} \in \mathcal{D}_L$, where A is a $p \times p$ matrix. We need to find a matrix $A^\star$ such that $\delta_L^\star(\boldsymbol{x}) = A^\star\boldsymbol{x}$ satisfies

$$r_L = \inf_{\delta_L\in\mathcal{D}_L}\sup_{\pi\in\Gamma} r(\pi,\delta_L) = \sup_{\pi\in\Gamma} r(\pi,\delta_L^\star).$$

Under L_2 loss we have

$$\begin{aligned} R(\boldsymbol{\theta}, A\boldsymbol{X}) &= E^{\boldsymbol{X}|\boldsymbol{\theta}}\|A\boldsymbol{X} - \boldsymbol{\theta}\|^2 \\ &= E\boldsymbol{X}'A'A\boldsymbol{X} - E\boldsymbol{X}'A'\boldsymbol{\theta} - E\boldsymbol{\theta}'A\boldsymbol{X} + \boldsymbol{\theta}'\boldsymbol{\theta} \\ &= \operatorname{tr} A'A(I + \boldsymbol{\theta}\boldsymbol{\theta}') - \boldsymbol{\theta}'A'\boldsymbol{\theta} - \boldsymbol{\theta}'A\boldsymbol{\theta} + \boldsymbol{\theta}'\boldsymbol{\theta} \\ &= \boldsymbol{\theta}'(A - I)'(A - I)\boldsymbol{\theta} + \operatorname{tr} A'A \\ &= \|(A - I)\boldsymbol{\theta}\|^2 + \|A\|^2. \end{aligned}$$

Therefore

$$r(\pi, \delta_L) = E^{\theta}\|(A - I)\boldsymbol{\theta}\|^2 + \|A\|^2 = E\tau\|A - I\|^2 + \|A\|^2,$$

and

$$\sup_{\pi \in \Gamma} r(\pi, \delta_L) = \left(\sup_{G \in \mathcal{G}} E\tau\right) \|A - I\|^2 + \|A\|^2. \tag{3.2}$$

Let $t_0 = \sup_{G \in \mathcal{G}} E\tau (< \infty)$. The next step is to find $\inf_A t_0\|A - I\|^2 + \|A\|^2$. If we differentiate (3.2) with respect to A, and set the derivative equal to 0,

$$\begin{aligned} d(t_0\|A - I\|^2 + \|A\|^2) &= t_0 2 \operatorname{tr}(A - I)'dA + 2 \operatorname{tr} A'dA \\ &= 2 \operatorname{tr}((1 + t_0)A - t_0 I)'dA = 0, \end{aligned}$$

we obtain

$$(1 - t_0)A - t_0 I = 0,$$

and

$$A^\star = \frac{t_0}{1 + t_0} I = \sup_{G \in \mathcal{G}} \frac{E\tau}{1 + E\tau} I,$$

since the function $\frac{x}{1+x}$ is increasing in x. That $A^\star$ minimizes (3.2) follows from the standard argument

$$d^2(t_0\|A - I\|^2 + \|A\|^2) = 2d(\operatorname{tr}(1 + t_0)A - t_0 I)dA = 2 \operatorname{tr}(1 + t_0)dA'dA \geq 0.$$

Therefore, the linear Γ-minimax risk is

$$r_L = t_0 \left\| \left(\frac{t_0}{1 + t_0} - 1\right) I \right\|^2 + \left\| \frac{t_0}{1 + t_0} I \right\|^2 = p \frac{t_0}{1 + t_0} = p \sup_{G \in \mathcal{G}} \frac{E\tau}{1 + E\tau}. \tag{3.3}$$

From (3.2) and (3.3) we get

$$\rho = \frac{r_L}{r_\Gamma} \leq \frac{\sup_{G \in \mathcal{G}} \dfrac{E\tau}{1 + E\tau}}{\sup_{G \in \mathcal{G}} E\dfrac{\tau}{1 + \tau}} (= \rho^\star). \qquad \square$$

Notice that the upper bound $\rho^\star$ does not depend on the dimension p of the model.

Remark 3.1. The Chebyshev inequality applied to (3.1) gives

$$\rho^\star \le \frac{1}{\inf_{G\in\mathcal{G}} P(\tau \ge E\tau)}. \tag{3.4}$$

If $\mathcal{G}$ is *any* family of point mass distributions satisfying (1.2), then the upper bound in (3.4) is achieved and is equal to 1. This is not surprising: we will see in Theorem 3.2, that the bound $\rho^\star = 1$ remains valid for variety of classes $\mathcal{G}$.

It is straightforward to show (as an elementary moment problem) that in the case when $\mathcal{G}$ is the set of *all* distributions satisfying condition (1.2) with $\sup_{\mathcal{G}} E\tau = t_0$, that $\sup_{\mathcal{G}} E\frac{\tau}{1+\tau} = \frac{t_0}{1+t_0}$. This means that the Bayes linear rule is Γ-minimax and that $\rho = \rho^\star = 1$. The next two theorems assess $\rho^\star$ for different classes $\mathcal{G}$.

THEOREM 3.2. *If we restrict the class $\mathcal{G}$ to be the class of* (i) *all symmetric distributions or* (ii) *all symmetric unimodal distributions on the interval* $[a, b]$, *then the Γ-minimax rule is linear, the corresponding hyperprior G puts all its mass at the middle point $\frac{a+b}{2}$, and $\rho = \rho^\star = 1$.*

PROOF. Let us prove the theorem in case (ii). We fix two nonnegative numbers a and b, such that $a < b$, and consider the class $\mathcal{G}$ of all unimodal and symmetric (about $c = \frac{a+b}{2}$) distributions on the interval $[a, b]$. Now, any random variable τ with distribution in the class $\mathcal{G}$, can be represented as

$$\tau \stackrel{d}{=} c + U \cdot Z, \tag{3.5}$$

where U is an uniform on $[-1, 1]$ and Z is the corresponding random variable defined on $[0, \frac{b-a}{2}]$, and is independent of U. Since $E\tau \equiv c$, then $\sup_{\mathcal{G}} \frac{E\tau}{1+E\tau} = \frac{c}{1+c}$, and

$$\sup_{\mathcal{G}} E\frac{\tau}{1+\tau} = \sup_Z E^Z \left(E^{U|Z} \frac{c+UZ}{1+c+UZ}\right) = \sup_Z E\left(1 - \frac{\ln \dfrac{1+c+Z}{1+c-Z}}{2Z}\right).$$

We maximize the above expectation by taking Z to be identically 0, since the function under the expectation sign is monotone decreasing in Z, when $0 \le Z \le \frac{b-a}{2}$. This choice of Z gives $\tau \equiv c$, and the maximal expectation is $1 - \frac{1}{1+c}$. We now show that the corresponding linear Bayes rule is Γ-minimax.

The standard way to check for the Γ-minimaxity of a rule δ_0 that is Bayes with respect to the prior $\pi_0 \in \Gamma$, is to prove that for any other prior $\pi \in \Gamma$

$$r(\pi, \delta_0) \le r(\pi_0, \delta_0).$$

The right-hand side is $\frac{pc}{1+c}$. The left-hand side is

$$E^{\boldsymbol\theta} E^{X|\boldsymbol\theta} \left\|\boldsymbol\theta - \frac{c}{1+c}\boldsymbol X\right\|^2 = E^{\boldsymbol\theta}\frac{1}{1+c}\|\boldsymbol\theta\|^2 + \frac{pc^2}{(1+c)^2} = \frac{pc}{1+c},$$

because $E^{\boldsymbol{\theta}}\|\boldsymbol{\theta}\|^2 \equiv pc$ for any prior in the class Γ. Therefore, the linear rule is Γ-minimax, the prior $\pi(\boldsymbol{\theta}) = \phi_{p,c}(\boldsymbol{\theta})$ is the least favorable, and $\rho^\star = 1$. □

THEOREM 3.3. *Let $\mathcal{G}$ be the class of all distributions on $[0, \infty)$ that are unimodal about $c > 0$, such that $0 < \sup_{\mathcal{G}} E\tau = t_0 < \infty$. In this case $\rho^\star > 1$.*

PROOF. Any random variable τ, unimodal about c, can be written as

$$\tau \stackrel{d}{=} c + UZ, \tag{3.6}$$

where U is uniform $\mathcal{U}[0,1]$, and Z is a fixed mixing random variable, independent of U. The condition $0 \le E\tau \le t_0$ is equivalent to $-2c \le EZ \le 2(t_0 - c)$. Let us assume that $t_0 \ge c$. The case $t_0 \le c$ is analogous. First, $\sup_\tau \frac{E\tau}{1+E\tau} = \frac{t_0}{1+t_0}$. If $Z \neq 0$,

$$E\frac{\tau}{1+\tau} = E\frac{c+UZ}{1+c+UZ} = E^Z\left(1 - \frac{1}{Z}\log\frac{1+c+Z}{1+c}\right),$$

while for $Z = 0$, $E\frac{\tau}{1+\tau} = \frac{c}{1+c}$. The function $1 - \frac{1}{Z}\log\frac{1+c+Z}{1+c}$ is increasing in Z, and the solution of the moment problem

$$\begin{cases} \sup_Z E\left(1 - \dfrac{1}{Z}\log\dfrac{1+c+Z}{1+c}\right) \\ \text{subject to } EZ \le 2(t_0 - c) \end{cases}$$

is the random variable Z_0, degenerate at $2(t_0 - c)$. This corresponds to

$$\tau \sim \mathcal{U}[c, 2t_0 - c].$$

Therefore,

$$E\frac{\tau}{1+\tau} \le \frac{1}{2(t_0-c)}\int_c^{2t_0-c}\frac{t}{1+t}dt = 1 + \frac{\log\dfrac{1+c}{1+2t_0-c}}{2(t_0-c)},$$

and

$$\rho^\star = \frac{\dfrac{t_0}{1+t_0}}{1 + \dfrac{\log\dfrac{1+c}{1+2t_0-c}}{2(t_0-c)}}. \qquad \square$$

Table 1 gives maximal values for $\rho^\star$, as a function of $t_0 (> c)$, for different choices of c.

Table 1. Maximal values of $\rho^\star$ as a function of t_0.

	$c = 0$	$c = 1$	$c = 3$	$c = 4$	$c = 10$
$\max \rho^\star$	1.11593	1.04439	1.02024	1.01593	1.00699
t_0	1.7382	6.6368	15.8827	20.4786	48.0086

4. Examples

In this section we illustrate the performance of linear Γ-minimax rules in four different situations.

Example 4.1. Let $\mathcal{G}$ be the Poisson family $\{\mathcal{P}(\lambda), 0 < \lambda < \Lambda\}$. In this case

$$r_L = \sup_{\mathcal{G}} \frac{E\tau}{1+E\tau} = \frac{\Lambda}{1+\Lambda},$$

$$r_\Gamma \geq \sup_{\mathcal{G}} E\frac{\tau}{1+\tau} = \frac{\Lambda - 1 + e^{-\Lambda}}{\Lambda}, \quad \text{and}$$

$$\rho \leq \rho^\star = \frac{\Lambda^2}{(\Lambda+1)(\Lambda - 1 + e^{-\Lambda})}.$$

The limiting behavior of $\rho^\star$ is as follows:

$$\rho^\star \to 1, \Lambda \to \infty,$$

and

$$\rho^\star = \frac{\Lambda^2}{(1+\Lambda)(\Lambda^2/2 + o(\Lambda^2))} \to 2, \Lambda \to 0.$$

Apparently the linear rule does not perform well for small values of Λ.

Example 4.2. If τ has the Inverse Gamma $\mathcal{IG}(\frac{m}{2}, \frac{m\sigma^2}{2})$ distribution, then $\boldsymbol{\theta}$ has the multivariate $\mathcal{T}_p(m, 0, \sigma^2 I)$ distribution. Since $E\tau = \frac{m\sigma^2}{m-2}$ should be finite (condition (1.2)), we assume that $m > 2$.

Let Γ be the family

$$\{\mathcal{T}_p(m, 0, \sigma^2 I), 0 \leq \sigma \leq S\},$$

where $m > 2$ is the number of degrees of freedom, and S is a nonnegative real number. We are interested in calculating $\rho^\star$ for the above family. For $\tau \sim \mathcal{IG}(\frac{m}{2}, \frac{m\sigma^2}{2})$, we have

$$E\frac{\tau}{1+\tau} = \left(\frac{m\sigma^2}{2}\right)^{m/2} e^{m\sigma^2/2}\Gamma\left(1 - \frac{m}{2}, \frac{m\sigma^2}{2}\right), \tag{4.1}$$

where $\Gamma(a, b) = \int_b^\infty t^{a-1}e^{-t}dt$ is the incomplete Gamma function. When m is fixed, the expression (4.1) is increasing in σ, and

$$\sup_{0<\sigma<S} E\frac{\tau}{1+\tau} = \left(\frac{mS^2}{2}\right)^{m/2} e^{mS^2/2}\Gamma\left(1 - \frac{m}{2}, \frac{mS^2}{2}\right).$$

Since $\sup_{0\leq\sigma\leq S} \frac{E\tau}{1+E\tau} = \frac{mS^2}{mS^2+m-2}$, we have

$$\rho^\star = \frac{\dfrac{mS^2}{mS^2+m-2}}{\left(\dfrac{mS^2}{2}\right)^{m/2} e^{mS^2/2}\Gamma\left(1 - \dfrac{m}{2}, \dfrac{mS^2}{2}\right)}. \tag{4.2}$$

Table 2. Worst choice of S for the selected values of m.

	$m=3$	$m=4$	$m=5$	$m=7$	$m=10$	$m=20$
$\max\rho^\star$	1.46969	1.24333	1.16415	1.09932	1.06231	1.02776
S	0.44908	0.59973	0.68457	0.77822	0.84642	0.92399

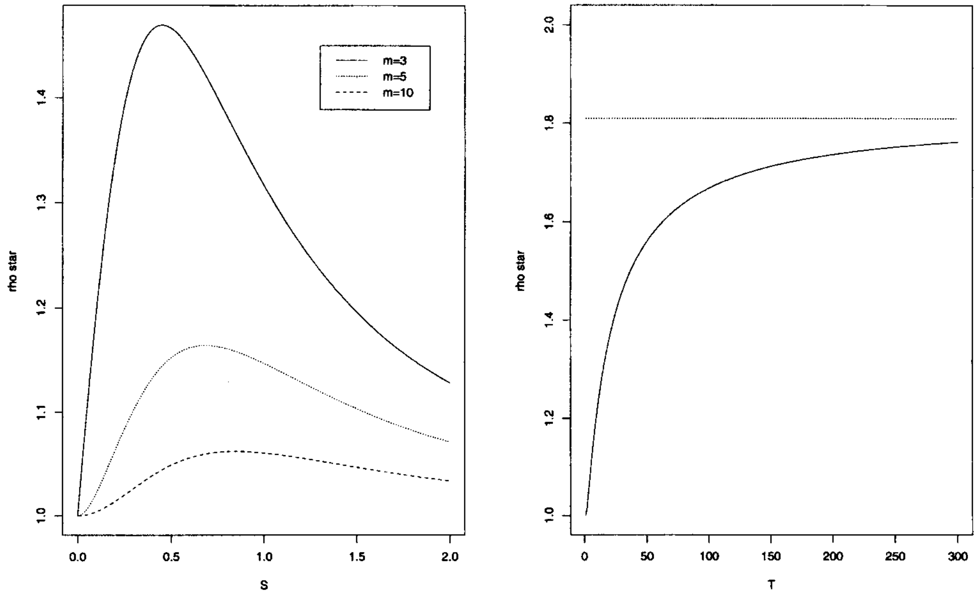

Fig. 1. Left: $\rho^\star$ for $m = 3$, 5, and 10; Right: Plot of $\rho^\star(T)$.

Table 2 gives the worst choice of S (the choice that maximizes $\rho^\star$), for selected values of m.

For example, for $m = 5$ degrees of freedom, the family $\{\mathcal{T}_p(5, 0, \sigma^2 I), 0 \leq \sigma \leq 0.68457\}$ maximizes $\rho^\star$. $\rho^\star_{\max} = 1.16415$ means that the loss (in terms of Γ-minimax risks) incurred by using the linear Γ-minimax rule instead of the unrestricted one is less than 16.5%. Figure 1, Left, shows the function $\rho^\star = \rho^\star(S)$, for $m = 3$, 5, and 10 degrees of freedom.

Example 4.3. An interesting example is when the family $\mathcal{G}$ is $\{\mathcal{U}[0, m], 0 \leq m \leq M\}$. Here

$$\rho^\star = \frac{\dfrac{M}{2+M}}{1 - \dfrac{\ln(1+M)}{M}}.$$

As a function of M, $\rho^\star \to 1$ when $M \to 0$ or ∞. The least favorable choice of M is 3.4764 for which the linear Γ-minimax rule is 11.6% worse than the general one. This can be viewed as a special case of (3.6) with $c = 0$ and $Z = M$.

Example 4.4. Let $\mathcal{G} = \{(1-\epsilon)\mathbf{1}(\tau = 1) + \epsilon\mathbf{1}(\tau = t), 1 \le t \le T\}$. This class of hyperpriors was considered by Albert (1984) in a different context. The induced class of priors Γ is the normal $\mathcal{N}(0,1)$ distribution ϵ-contaminated by the normal $\mathcal{N}(0,t)$ distribution, $1 \le t \le T$. This case is interesting since, using numerical methods, we can give sharper lower bounds, $r_\Gamma^\star$, on r_Γ. For $\epsilon = 0.1$, $r_\Gamma^\star$ has the maximum 0.60378 at $T = 25.888745$. This gives $\rho \le 1.65623$ uniformly in T. The linear Γ-minimax rule is "worse" than the general Γ-minimax rule by (about) 66% in the least favorable case. On the other hand, Theorem 3.1 gives

$$\rho^\star(T) = \frac{\dfrac{1-\epsilon+\epsilon T}{2-\epsilon+\epsilon T}}{\dfrac{1}{2}+\dfrac{\epsilon}{2}-\dfrac{\epsilon}{1+T}}. \tag{4.3}$$

The function $\rho^\star(T)$ is increasing in T and $\lim_{T\to\infty} \rho^\star(T) = \frac{2}{1+\epsilon}$. The choice $\epsilon = 0.1$ gives an upper bound on ρ of 1.81818. (See Fig. 1, Right.) Table 3 gives some numerical results for $\epsilon = 0.1$ and the selected values of T.

Table 3. Comparison of bounds.

T	$r_\Gamma^\star$	lower bound on r_Γ	r_L	$\frac{r_L}{r_\Gamma^\star}$	$\rho^\star$
1	0.5	0.5	0.5	1	1
1.5	0.51217	0.51	0.51220	1.00004	1.00430
2	0.52353	0.51667	0.52381	1.00053	1.01388
3	0.54277	0.525	0.54545	1.00494	1.03896
5	0.56799	0.53333	0.58333	1.02702	1.09375
10	0.59307	0.54090	0.65517	1.10472	1.21124
20	0.60314	0.54524	0.74359	1.23288	1.36388
50	0.60378	0.54804	0.85507	1.41620	1.56024

By using Theorem 3.2 Example 4.4 can be readily generalized. Let $\mathcal{G}$ be the class $\{(1-\epsilon)\mathbf{1}(\tau = 1) + \epsilon G(t)\}$, where $G(t)$ is (i) an arbitrary, (ii) a symmetric, or (iii) a symmetric unimodal distribution on $[1, T]$. In case (i), the bound $\rho^\star$ is the same as in (4.3), while in cases (ii) and (iii) we have

$$\rho^\star = \frac{\dfrac{1-\epsilon+\epsilon c}{2-\epsilon+\epsilon c}}{\dfrac{1}{2}+\dfrac{\epsilon}{2}-\dfrac{\epsilon}{1+c}},$$

where $c = \frac{T+1}{2}$.

5. Generalization

It is natural to wonder what happens to $\rho^\star$ when the covariance matrices in the model and the prior are not identity matrices. Let us assume the following setup:

$$\begin{aligned} \boldsymbol{X} \mid \boldsymbol{\theta} &\sim \mathcal{MVN}_p(\boldsymbol{\theta}, \Sigma) \\ \boldsymbol{\theta} \mid \tau &\sim \mathcal{MVN}_p(\boldsymbol{0}, \tau\Psi) \\ \tau &\sim G(t), \qquad (\tau \geq 0), \end{aligned} \tag{5.1}$$

where Σ and Ψ are arbitrary positive definite matrices. Let G belong to the family $\mathcal{G}$ for which the condition (1.2) is satisfied. Without loss of generality, we can take $\Sigma = I$, since we can always rescale the model by multiplying $\boldsymbol{X}$ by $\Sigma^{1/2}$. Denote with $\phi_{p,\Sigma}(\boldsymbol{x} - \boldsymbol{\mu})$ the density of the $\mathcal{MVN}_p(\boldsymbol{\mu}, \Sigma)$ distribution. The marginal distribution of X in the model (5.1) can be expressed as

$$m(\boldsymbol{x}) = \int \phi_{p,I+t\Psi}(\boldsymbol{x})dG(t).$$

Mimicking the calculation in the proof of Theorem 3.1 we get

$$r_\Gamma \leq \sup_{G \in \mathcal{G}} E(p - \text{tr}(I + \tau\Psi)^{-1}).$$

On the other hand, matrix differentiation and basic matrix algebra show that the linear Γ-minimax estimate is

$$\left(\left(I + \left(\sup_{G \in \mathcal{G}} E\tau \right) \Psi \right)^{-1} \right)^{-1} \boldsymbol{X},$$

and the corresponding linear Γ-minimax risk is

$$p - \text{tr} \left(I + \left(\sup_{G \in \mathcal{G}} E\tau \right) \Psi \right)^{-1}.$$

Therefore, the bound on ρ (which generalizes that in Theorem 3.1) is

$$\rho^\star = \frac{p - \text{tr}(I + (\sup_{G \in \mathcal{G}} E\tau)\Psi)^{-1}}{\sup_{G \in \mathcal{G}}(p - \text{tr}\, E(I + \tau\Psi)^{-1})}. \tag{5.2}$$

As is apparent from this equation, the bound (5.2) depends on the dimension p of the model.

We give two examples of the calculation of $\rho^\star$ in the general case.

THEOREM 5.1. *If a random variable τ is as in (3.5), i.e. it belongs to the class of all symmetric and unimodal distributions on $[a, b]$, then the relation $\rho^\star = 1$ remains valid.*

PROOF. Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of the matrix Ψ. Then

$$r_L = p - \operatorname{tr}(1 + c\Psi)^{-1} = p - \sum_{i=1}^{p} \frac{1}{1 + c\lambda_i},$$

and

$$\begin{aligned} r_\Gamma &\geq \sup_Z \left(p - E^Z E^{U|Z} \sum_{i=1}^{p} \frac{1}{1 + (c + UZ)\lambda_i} \right) \\ &= \sup_Z \left(p - E^Z \sum_{i=1}^{p} \frac{1}{2Z\lambda_i} \log \frac{1 + c\lambda_i + Z\lambda_i}{1 + c\lambda_i - Z\lambda_i} \right). \end{aligned}$$

For the choice $Z = 0$ (among all random variables on $[0, \frac{b-a}{2}]$), the previous supremum is achieved and has the value

$$p - \sum_{i=1}^{p} \frac{1}{1 + c\lambda_i}.$$

Therefore, $\rho^\star = 1$. □

Example 5.1. The bound (4.2) can be generalized as well. Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of the matrix Ψ.

$$\begin{aligned} r_L &= p - \operatorname{tr}\left(I + \frac{mS^2}{m-2}\Psi\right)^{-1} = p - \sum_{i=1}^{p} \frac{m-2}{m - 2 + mS^2\lambda_i}. \\ r_\Gamma &\geq \sup_{0 \leq \sigma \leq S} (p - E \operatorname{tr}(I + \tau\Psi)^{-1}) \\ &= \sup_{0 \leq \sigma \leq S} p - \frac{\left(\frac{m\sigma^2}{2}\right)^2}{\Gamma\left(\frac{m}{2}\right)} \int_0^\infty \sum_{i=1}^{p} \frac{t^{-1-m/2}}{1 + t\lambda_i} e^{-m\sigma^2/2t} dt \\ &= p - \sum_{i=1}^{p} \left(\frac{mS^2\lambda_i}{2}\right)^{m/2} \frac{m}{2} e^{mS^2\lambda_i/2} \Gamma\left(-\frac{m}{2}, \frac{mS^2\lambda_i}{2}\right). \end{aligned}$$

Therefore,

$$\rho^\star = \frac{p - \sum_{i=1}^{p} \frac{m-2}{m - 2 + mS^2\lambda_i}}{p - \sum_{i=1}^{p} \left(\frac{mS^2\lambda_i}{2}\right)^{m/2} \frac{m}{2} e^{mS^2\lambda_i/2} \Gamma\left(-\frac{m}{2}, \frac{mS^2\lambda_i}{2}\right)}.$$

Acknowledgements

I would like to thank Anirban DasGupta for many hours of fruitful discussions. Comments of the anonymous referee and the Associate Editor improved the clarity of presentation.

REFERENCES

Albert, J. H. (1984). Estimation of a normal mean using mixtures of normal priors, Tech. Report, Bowling Green State University, Ohio.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer, Berlin.

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

Brown, L. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems, *Ann. Math. Statist.*, **37**, 1087–1136.

Brown, L. (1986). *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory*, Lecture Notes vol. 9, Institute of Mathematical Statistics.

DasGupta, A. (1992). Distributions which are Gaussian convolutions, Tech. Report, No. 92-34, Department of Statistics, Purdue University, West Lafayette, Indiana.

DasGupta, A., Ghosh, J. K. and Zen, M. M. (1990). Bayesian analysis of multivariate normal mean with flat tailed priors, Tech. Report, No. 90-22, Department of Statistics, Purdue University, West Lafayette, Indiana.

Donoho, D., Liu, R. C. and McGibbon, B. (1990). Minimax risk over hyperrectangles and implications, *Ann. Statist.*, **18**, 1416–1437.

Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators—part I: the Bayes case, *J. Amer. Statist. Assoc.*, **66**, 807–815.

Efron, B. and Olshen, R. (1978). How broad is the class of normal scale mixtures?, *Ann. Statist.*, **5**, 1159–1164.

Faith, E. R. (1978). Minimax Bayes estimators of multivariate normal mean, *J. Multivariate Anal.*, **8**, 372–379.

Ibragimov, I. A. and Has'minskii, R. Z. (1984). Nonparametric estimation of the values of a linear functional in Gaussian white noise, *Theory Probab. Appl.*, **29**, 1–32.

Kudō, H. (1967). On partial prior information and the property of parametric sufficiency, *Proc. Fifth Berkeley Symp. on Math. Statist. Prob.*, 251–265, University of California Press, Berkeley.

Robert, C. (1990). Hidden mixtures and Bayesian sampling, Rapport technique 115, LSTA Université Paris 6.

Skibinsky, M. and Cote, L. (1962). On the inadmissibility of some standard estimates in the presence of prior information, *Ann. Math. Statist.*, **34**, 539–548.

Solomon, D. L. (1972). Λ-minimax estimation of multivariate location parameter, *J. Amer. Statist. Assoc.*, **67**, 641–646.

Vidakovic, B. and DasGupta, A. (1996). Efficiency of linear rules for estimating a bounded normal mean, *Sankhyā Ser. A*, **58**, 81–100.

Watson, S. R. (1974). On Bayesian inference with incompletely specified prior distributions, *Biometrika*, **61**, 193–196.

West, M. (1987). On scale mixtures of normal distributions, *Biometrika*, **74**, 646–648.