LINEAR VERSUS NONLINEAR RULES FOR MIXTURE NORMAL PRIORS*

by

Brani Vidakovic and Anirban DasGupta
Purdue University

Technical Report #91-46

Department of Statistics
Purdue University

August 1991

* This research was supported by NSF Grant DMS-8923071.

◇

Linear versus nonlinear rules for mixture normal priors *

Brani Vidakovic and Anirban DasGupta
Purdue University

August 26, 1991

Abstract

The problem under consideration is Γ-minimax estimation of a multivariate normal mean with squared error loss. The family Γ of priors is induced by mixing zero mean multivariate normals with a covariance matrix τI by nonnegative random variables τ, whose distributions belong to a suitable family $\mathcal{G}$. For fixed family $\mathcal{G}$, the linear Γ-minimax rule is compared with the usual Γ-minimax rule in terms of appropriate Γ-minimax risks. It is customary to measure the performance of linear Γ-minimax rule by the ratio (Γ-minimax risk of linear rule)/(Γ-minimax risk), (cf. Donoho, Liu and MacGibbon (1990)). It is shown that the linear rule is "good"(i.e. the ratio of the risks is close to 1) whenever $\sup_{\mathcal{G}} \frac{E\tau}{1+E\tau}$ is "close" to $\sup_{\mathcal{G}} E\frac{\tau}{1+\tau}$, regardless of the dimension of the model. Several examples illustrate a different behavior of the linear Γ-minimax rule.

Key Words: Γ-minimax, scale mixture of normals, Brown's identity, linear Γ-minimax rule.
AMS 1985 subject classification: 62C20.

1 Introduction

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

On the other hand, if we have a lot of prior information, then the class Γ is not rich. An extreme case is a class Γ that contains only one prior. In this case, the Γ-minimax framework becomes the minimum Bayes risk principle framework. The spirit of Γ-minimaxity can be 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 distribution that is "true" in an absolute sense. There are only more or less relevant priors, and

*This research was supported by NSF Grant DMS-8923071 at Purdue University.

> 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 the belief in Γ-minimax may produce "demonstrable incoherence", since there are examples when the Γ-minimax rule is not Bayes (Watson (1974)). But in most cases 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:

$$\begin{aligned} \underset{\sim}{X}|\underset{\sim}{\theta} &\sim \mathcal{MVN}_p(\underset{\sim}{\theta}, I), \\ \underset{\sim}{\theta}|\tau &\sim \mathcal{MVN}_p(\underset{\sim}{0}, \tau I), \\ \tau &\sim G(t), (\tau \geq 0). \end{aligned} \tag{1}$$

Let $\mathcal{G}$ be the family of distribution functions G. Suppose that random variables τ have uniformly bounded expectations, and that $E\tau = 0$ not for all τ, i.e.

$$0 < \sup_{G \in \mathcal{G}} \int_T t dG(t) < \infty \tag{2}$$

where T is the union of supports of τ's. Class $\mathcal{G}$ determines the family of priors Γ as

$$\Gamma = \left\{ \int_T \phi_{p,t}(\underset{\sim}{\theta}) dG(t), G \in \mathcal{G} \right\} \tag{3}$$

where

$$\phi_{p,t}(\underset{\sim}{\theta}) = \frac{1}{(2\pi)^{p/2} t^{p/2}} e^{\frac{-\|\theta\|^2}{2t}} \tag{4}$$

is the density of $\mathcal{MVN}_p(\underset{\sim}{0}, tI)$ distribution. The class (3) is an example of so called "scale mixture of normals" or "normal scale mixtures" family. An interesting fact is that some of the well known families of distributions such as: t (particularly Cauchy), logistic, double-exponential, $cosh^{-1}$, etc., can be obtained as appropriate normal scale mixtures. These classes are attractive because they are easy to work with (Monte Carlo studies, Gibbs sampler, Bayesian calculations), or they possess some desirable properties (Robust Bayesian inference). For some accounts of significance of scale mixture of normals in statistics, we refer the reader to Efron and Olshen (1978), Robert (1990), and DasGupta, Ghosh, and Zen (1990). Having the family of priors, it seems natural to employ the Γ-minimax setup for estimating the unknown parameter $\underset{\sim}{\theta}$. Let $\mathcal{D}$ be the set of all decision rules. The estimator $\delta^* \in \mathcal{D}$ that minimizes $\sup_{\pi \in \Gamma} r(\pi, \delta)$, i.e.

$$\inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{\pi \in \Gamma} r(\pi, \delta^*), \tag{5}$$

is the Γ-minimax 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

$$\inf_{\delta\in\mathcal{D}_L} \sup_{\pi\in\Gamma} r(\pi, \delta) = \sup_{\pi\in\Gamma} r(\pi, \delta_L^\star), \tag{6}$$

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 performance of linear Γ-minimax rules compared to plain Γ-minimax rules, and 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.

The calculation and estimation of ρ (or $\rho^\star$) seems to be the problem of interest. 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, Liu, and MacGibbon (1990) have derived very sharp upper bound $\rho \leq 1.25$ that holds uniformly in m. When Γ is a family of all unimodal and symmetric distributions on $[-m, m]$, Vidakovic and DasGupta (1991) have shown that $\rho \leq 1.074$.

In the multivariate case, we think about a linear rule as an affine transformation $\delta(\underset{\sim}{x}) = A\underset{\sim}{x} + B$, for some matrices A and B. Solomon (1972) has shown that $B \neq \mathbf{0}$ is an inadmissible choice. Motivation for using linear rules is apparent: they are easily calculable and simple to use.

2 Preliminaries

2.1 An Information Integral Inequality

First, we will prove an inequality involving the Fisher information integral. The lemma that follows can be stated and proven in much more generality. The integral

$$\mathcal{I}(f(\underset{\sim}{x})) = \int_{\Re^p} \frac{\sum_{i=1}^p (\frac{\partial}{\partial x_i} f(\underset{\sim}{x}))^2}{f(\underset{\sim}{x})} d\underset{\sim}{x} \tag{7}$$

is the trace of the Fisher information matrix $\mathbf{I}(\underset{\sim}{\theta})$ (for a location parameter $\underset{\sim}{\theta}$ of a family $f(\underset{\sim}{x} - \underset{\sim}{\theta})$). Let G be an arbitrary distribution function.

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

$$\mathcal{I}(\int_T \phi_{p,t+1}(\underset{\sim}{x}) dG(t)) \leq \int_T \mathcal{I}(\phi_{p,t+1}(\underset{\sim}{x})) dG(t) \tag{8}$$

Proof:

$$\begin{aligned}
&(\int_T \frac{\partial}{\partial x_i} \phi_{p,t+1}(\underset{\sim}{x}) dG(t))^2 \\
= \; &(\int_T \frac{\frac{\partial}{\partial x_i}\phi_{p,t+1}(\underset{\sim}{x})}{\sqrt{\phi_{p,t+1}(\underset{\sim}{x})}} \sqrt{\phi_{p,t+1}(\underset{\sim}{x})} dG(t))^2 \\
\leq \; &\int_T \frac{(\frac{\partial}{\partial x_i}\phi_{p,t+1}(\underset{\sim}{x}))^2}{\phi_{p,t+1}(\underset{\sim}{x})} dG(t) \int_T \phi_{p,t+1}(\underset{\sim}{x}) dG(t).
\end{aligned}$$

Therefore:

$$\frac{(\int_T \frac{\partial}{\partial x_i}\phi_{p,t+1}(\underset{\sim}{x})dG(t))^2}{\int_T \phi_{p,t+1}(\underset{\sim}{x})dG(t)} \leq \int_T \frac{(\frac{\partial}{\partial x_i}\phi_{p,t+1}(\underset{\sim}{x}))^2}{\phi_{p,t+1}(\underset{\sim}{x})}dG(t). \tag{9}$$

If we take the derivative out of the integral sign and make a summation with respect to i in (9), we get

$$\frac{\sum_{i=1}^p(\frac{\partial}{\partial x_i}\int_T \phi_{p,t+1}(\underset{\sim}{x})dG(t))^2}{\int_T \phi_{p,t+1}(\underset{\sim}{x})dG(t)} \leq \int_T \frac{\sum_{i=1}^p(\frac{\partial}{\partial x_i}\phi_{p,t+1}(\underset{\sim}{x})dG(t))^2}{\phi_{p,t+1}(\underset{\sim}{x})}dG(t).$$

Finally,

$$\begin{aligned}\int_{\Re^p}\frac{\sum_{i=1}^p(\frac{\partial}{\partial x_i}\int_T \phi_{p,t+1}(\underset{\sim}{x})dG(t))^2}{\int_T \phi_{p,t+1}(\underset{\sim}{x})dG(t)}d\underset{\sim}{x} &\leq \int_{\Re^p}\int_T \frac{\sum_{i=1}^p(\frac{\partial}{\partial x_i}\phi_{p,t+1}(\underset{\sim}{x}))^2}{\phi_{p,t+1}(\underset{\sim}{x})}dG(t)d\underset{\sim}{x}\\ &= \int_T\int_{\Re^p} \frac{\sum_{i=1}^p(\frac{\partial}{\partial x_i}\phi_{p,t+1}(\underset{\sim}{x}))^2}{\phi_{p,t+1}(\underset{\sim}{x})}d\underset{\sim}{x}dG(t).\end{aligned}$$

The positivity of the integrand allows the interchange of the order of integration. □

2.2 Brown's Identity

When the model is $X|\theta \sim \mathcal{N}(\theta, 1)$ 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

$$r(\pi) = 1 - \mathcal{I}(\phi_1 * \pi(x)). \tag{10}$$

Under such a model, the convolution $\phi_1 * \pi(x)$ can be interpreted as the marginal distribution for X. For derivation and some applications of Brown's identity see Bickel (1981), also Brown (1987). In the p-variate case Brown's identity has the form

$$r(\pi) = p - \mathcal{I}(\phi_{p,1} * \pi(\underset{\sim}{x})). \tag{11}$$

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

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

$$r_\Gamma \geq \sup_\pi (p - \mathcal{I}(\phi_{p,1} * \pi(\underset{\sim}{x}))\ ), \tag{12}$$

but equality holds in most regular cases. For rigorous discussion on Γ-minimax theorems see Stein (1982).

3 Main Result

Assume model (1) and squared error loss. Let $\mathcal{G}$ be any family of distributions that satisfies (2), and let Γ be the induced family (3).

Theorem 3.1 :

$$(13) \qquad \rho = \frac{r_L}{r_\Gamma} \leq \frac{\sup_{G\in\mathcal{G}} \frac{E\tau}{1+E\tau}}{\sup_{G\in\mathcal{G}} E\frac{\tau}{1+\tau}}.$$

Proof: (a) Bound on r_Γ

If the prior density is

$$\pi(\underset{\sim}{\theta}) = \int_T \phi_{p,t}(\underset{\sim}{\theta})dG(t),$$

then under the model (1), elementary calculations give the marginal density for $\underset{\sim}{X}$:

$$(14) \qquad m_\pi(\underset{\sim}{x}) = \int_T \phi_{p,t+1}(\underset{\sim}{x})dG(t).$$

This fact is apparent after the algebraic transformation

$$\begin{aligned}
\pi(\underset{\sim}{\theta}|t)\phi_{p,1}(\underset{\sim}{x}-\underset{\sim}{\theta}) &= \frac{1}{(2\pi)^{p/2}t^{p/2}}e^{-\frac{\|\theta\|^2}{2t}}\frac{1}{(2\pi)^{p/2}}e^{-\frac{\|x-\theta\|^2}{2}} \\
&= \frac{1}{(2\pi)^{p/2}(\frac{t}{t+1})^{p/2}}e^{-\frac{1}{2}\frac{\|\underset{\sim}{\theta}-\frac{t}{t+1}\underset{\sim}{x}\|^2}{t/(t+1)}}\frac{1}{(2\pi)^{p/2}(t+1)^{p/2}}e^{-\frac{\|x\|^2}{2(t+1)}}.
\end{aligned}$$

Integrating out $\underset{\sim}{\theta}$ we get (14). Now

$$\begin{aligned}
r_\Gamma &\geq \sup_{\pi\in\Gamma}(p - \mathcal{I}(m_\pi(\underset{\sim}{x}))\ ) \\
&= \sup_{G\in\mathcal{G}}(p - \mathcal{I}(\int_T \phi_{p,t+1}(\underset{\sim}{x})dG(t))\ ) \\
&\geq \sup_{G\in\mathcal{G}}(p - \int_T \mathcal{I}(\phi_{p,t+1}(\underset{\sim}{x}))dG(t)\ ).
\end{aligned}$$

Using the well known fact that

$$\mathcal{I}(\phi_{p,t+1}(\underset{\sim}{x})) = \frac{p}{t+1},$$

we obtain a lower bound on r_Γ :

$$(15) \qquad r_\Gamma \geq \sup_{G\in\mathcal{G}}(p - pE\frac{1}{1+\tau}) = p\sup_{G\in\mathcal{G}} E\frac{\tau}{1+\tau}.$$

(b) Calculation of r_L

Notice that $E\underset{\sim}{\theta} = E(E\underset{\sim}{\theta}|\tau) = 0$, and $E\underset{\sim}{\theta}\underset{\sim}{\theta}' = E(E(\underset{\sim}{\theta}\underset{\sim}{\theta}'|\tau)) = E(\tau I) = (E\tau)I$. If our estimator is constrained to be linear, i.e. of the form

$$\delta_L(\underset{\sim}{x}) = A\underset{\sim}{x} \in \mathcal{D}_L,$$

where A is a $p \times p$ matrix, then we need to find a matrix $A^\star$ such that $\delta_L^\star(\underset{\sim}{x}) = A^\star\underset{\sim}{x}$ satisfies

$$(16) \qquad r_L = \inf_{\delta_L\in\mathcal{D}_L}\sup_{\pi\in\Gamma} r(\pi,\delta_L) = \sup_{\pi\in\Gamma} r(\pi,\delta_L^\star).$$

Under squared error loss we have

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

Therefore

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

and

$$\sup_{\pi\in\Gamma} r(\pi, \delta_L) = (\sup_{G\in\mathcal{G}} E\tau)\|A - I\|^2 + \|A\|^2. \tag{17}$$

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 (17) with respect to A, we get

$$\begin{aligned} d(t_0\|A - I\|^2 + \|A\|^2) &= t_0\,2tr(A - I)'dA + 2trA'dA \\ &= 2tr((1 + t_0)A - t_0I)'dA = 0. \end{aligned}$$

It follows that

$$(1 - t_0)A - t_0I = 0,$$

and

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

(since $\frac{x}{1+x}$ is monotonically increasing in x). The fact that $A^\star$ minimizes (17) follows from

$$d^2(t_0\|A - I\|^2 + \|A\|^2) = 2d(tr(1 + t_0)A - t_0I)dA = 2tr(1 + t_0)dA'dA \geq 0.$$

Therefore, the linear Γ-minimax risk is

$$r_L = t_0\|(\frac{t_0}{1 + t_0} - 1)I\|^2 + \|\frac{t_0}{1 + t_0}I\|^2 = p\frac{t_0}{1 + t_0} = p\sup_{G\in\mathcal{G}} \frac{E\tau}{1 + E\tau}. \tag{18}$$

From (17) and (18) we get

$$\rho = \frac{r_L}{r_\Gamma} \leq \frac{\sup_{G\in\mathcal{G}} \frac{E\tau}{1+E\tau}}{\sup_{G\in\mathcal{G}} E\frac{\tau}{1+\tau}}(= \rho^\star). \quad \square \tag{19}$$

Remark 3.1 :

Bound $\rho^\star$ does not depend on the dimension p of the model.

Remark 3.2 :

Chebyshev inequality[1] gives

$$
\begin{aligned}
\rho^\star &= \frac{\sup_{G\in\mathcal{G}} \frac{E\tau}{1+E\tau}}{\sup_{G\in\mathcal{G}} E\frac{\tau}{1+\tau}} \\
&\le \sup_{G\in\mathcal{G}} \frac{\frac{E\tau}{1+E\tau}}{E\frac{\tau}{1+\tau}} \\
&\le \frac{1}{\inf_{G\in\mathcal{G}} P(\tau \ge E\tau)}. \qquad (20)
\end{aligned}
$$

If $\mathcal{G}$ is *any* family of point mass distributions satisfying (2), then the upper bound of (20) is achieved and is equal to 1. This is not a surprise. As we will see in Remark 3.4, the bound $\rho^\star = 1$ is valid for many general classes $\mathcal{G}$.

Remark 3.3 :

Since for $x \ge 0$, the function $f(x) = \frac{x}{1+x}$ is concave, and Jensen's inequality gives the expected relation

$$\rho^\star \ge 1.$$

Remark 3.4 :

a. It is easy to show (as an elementary moment problem) that in the case when $\mathcal{G}$ is a set of *all* distributions satisfying condition (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 yields that the Bayes linear rule is Γ-minimax and that $\rho = \rho^\star = 1$.

Even if we restrict a class $\mathcal{G}$ to be the class of (i) all symmetric or (ii) all symmetric unimodal distributions on an interval $[a, b]$, then the Γ-minimax rule is linear, the corresponding hyperprior G puts all mass at the middle point $\frac{a+b}{2}$, and $\rho = \rho^\star = 1$. Let us show that this is the case when $\mathcal{G}$ is as (ii).

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

$$\tau \sim c + U \cdot Z, \qquad (21)$$

where U is an uniform on $[-1, 1]$ and Z is the corresponding random variable defined on $[0, \frac{b-a}{2}]$, and independent of U.

Since $E\tau \equiv c$, then

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

[1] If $f(x)$ is a nonnegative and nondecreasing function and X is a nonnegative random variable,then for any c: $P(X \ge c) \le \frac{Ef(X)}{f(c)}$. In our case $f(x) = \frac{x}{1+x}$ and $c = EX$.

and

$$\begin{aligned}
\sup_{\mathcal{G}} E\frac{\tau}{1+\tau} &= \sup_{Z} E\frac{c+UZ}{1+c+UZ} \\
&= \sup_{Z} E^{Z}(E^{U|Z}\frac{c+UZ}{1+c+UZ}) \\
&= \sup_{Z} E(1-\frac{ln\frac{1+c+Z}{1+c-Z}}{2Z}).
\end{aligned}$$

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 $0 \le Z \le \frac{b-a}{2}$. That choice of Z gives $\tau \equiv c$, and the maximal expectation is $1-\frac{1}{1+c}$. So, in this case the corresponding linear Bayes rule is Γ-minimax. The standard way to check for 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

$$\begin{aligned}
E^{\underset{\sim}{\theta}}E^{\underset{\sim}{X}|\underset{\sim}{\theta}}||\underset{\sim}{\theta}-\frac{c}{1+c}\underset{\sim}{X}||^2 &= E^{\underset{\sim}{\theta}}\frac{1}{1+c}||\underset{\sim}{\theta}||^2+\frac{pc^2}{(1+c)^2} \\
&= \frac{pc}{1+c},
\end{aligned}$$

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

b. Let now $\mathcal{G}$ be the class of all distributions on $[0,\infty)$, unimodal about $c>0$, such that $0<\sup_{\mathcal{G}} E\tau = t_0 < \infty$. We will show that in this case $\rho^\star > 1$. Any random variable τ, unimodal about c, can be written as

$$\tau = c + UZ, \tag{22}$$

where U is uniform $\mathcal{U}[0,1]$, and Z is fixed mixing random variable, independent of U. 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 $Z \ne 0$,

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

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

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

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

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

Therefore,

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

and

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

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

	$c=0$	$c=1$	$c=3$	$c=4$	$c=10$
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

Table 1.

4 Examples

Example 4.1 :

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

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

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

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

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

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

and

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

Therefore, we can not claim that the linear rule is good 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 $\underset{\sim}{\theta}$ 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 (2)), we assume that $m > 2$.

So, 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 let S be 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

$$E\frac{\tau}{1+\tau} = (\frac{m\sigma^2}{2})^{\frac{m}{2}} e^{\frac{m\sigma^2}{2}} \Gamma(1 - \frac{m}{2}, \frac{m\sigma^2}{2}), \tag{23}$$

where $\Gamma(a, b) = \int_b^\infty t^{a-1} e^{-t} dt$ is the incomplete Gamma function. When m is fixed, the expression (23) is increasing in σ, and

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

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

$$\rho^\star = \frac{\frac{mS^2}{mS^2+m-2}}{(\frac{mS^2}{2})^{\frac{m}{2}} e^{\frac{mS^2}{2}} \Gamma(1 - \frac{m}{2}, \frac{mS^2}{2})}. \tag{24}$$

The following table gives the worst choice of S(the ratio $\rho^\star$ maximal), 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

Table 2.

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 $\rho^\star$. The fact $\rho^\star_{max} = 1.16415$, 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 shows the function $\rho^\star = \rho^\star(S)$, for $m = 3, 5$, and 10 degrees of freedom.

Example 4.3 :

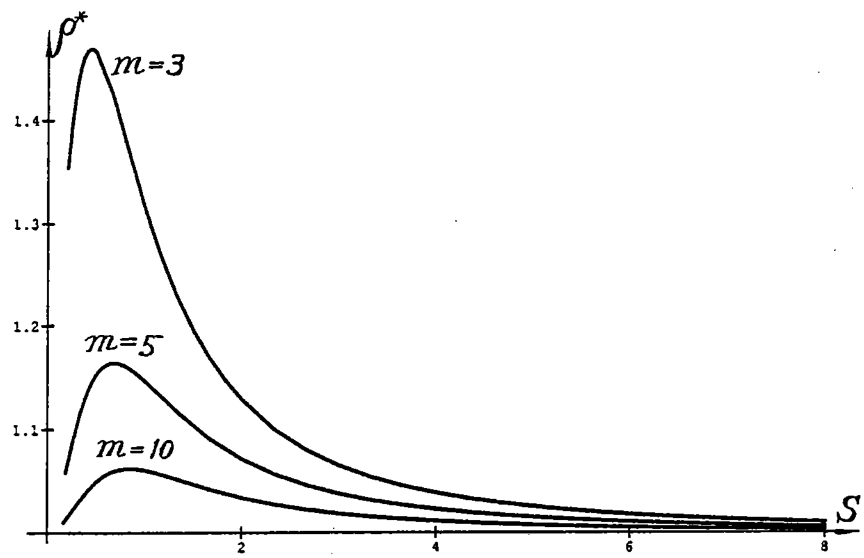

Figure 1: $\rho^\star$ for $m = 3, 5$, and 10

A very interesting example is when the family $\mathcal{G}$ is $\{\mathcal{U}[0, m],\ 0 \le m \le M\}$. Here

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

As a function of M, $\rho^\star \longrightarrow 1$ when $M \longrightarrow 0$ or ∞. The least favorable choice of M is 3.4764 for which the linear Γ-minimax rule is 11.6% worse than the plain one. This can be viewed as special case of (22) 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 on $r_\Gamma.(r_\Gamma^\star$, say). Instead of evaluating the integral $\int_T \mathcal{I}(\phi_{p,t+1}(\underset{\sim}{x}))dG(t)$ we numerically evaluate $\mathcal{I}(\int_T \phi_{p,t+1}(\underset{\sim}{x})dG(t))$. For the sake of being explicit, fix $\epsilon = 0.1$. In this case, $r_\Gamma^\star$ has the maximum 0.60378 at $T = 25.888745$. This gives $\rho \le 1/0.60378 = 1.65623$ uniformly in T. This means that the linear Γ-minimax rule is "worse" than the plain Γ- minimax rule by (about) 66% in the least favorable case. At the same time, Theorem 3.1 gives

$$\rho^\star(T) = \frac{\frac{1-\epsilon+\epsilon T}{2-\epsilon+\epsilon T}}{\frac{1}{2} + \frac{\epsilon}{2} - \frac{\epsilon}{1+T}}. \tag{25}$$

The function $\rho^\star(T)$ is increasing in T and

$$\lim_{T\to\infty} \rho^\star(T) = \frac{2}{1+\epsilon}.$$

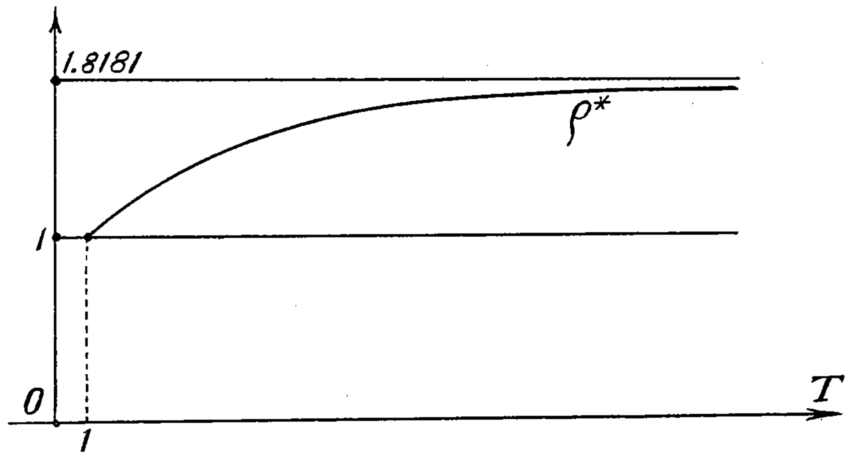

Figure 2: Plot of $\rho^*(T)$

The choice $\epsilon = 0.1$ gives an upper bound on ρ of 1.81818. (See Figure 2). Table 3 gives some numerical results for $\epsilon = 0.1$ and the selected values of T.

T	r_Γ^*	lower bnd on r_Γ (15)	r_L	$\frac{r_L}{r_\Gamma^*}$	ρ^* (13)
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

Table 3.

Because of the discussion in Remark 3.4, we can easily generalize this example. Let $\mathcal{G}$ be the class $\{(1-\epsilon)\mathbf{1}(\tau = 1) + \epsilon G(t)\}$, where $G(t)$ is (i) an arbitrary, (ii) symmetric, or (iii) symmetric unimodal distribution on $[1, T]$. In the case (i), the bound ρ^* is the same as in (25), while in cases (ii) and (iii) we have

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

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

5 Generalization

A very natural question is what happens to $\rho^\star$ when covariance matrices in the model and the prior are not identity matrices. So, let us assume the following setup:

$$
\begin{aligned}
\underset{\sim}{X}|\underset{\sim}{\theta} &\sim \mathcal{MVN}_p(\underset{\sim}{\theta}, \Sigma) \\
\underset{\sim}{\theta}|\tau &\sim \mathcal{MVN}_p(\underset{\sim}{0}, \tau\Psi) \qquad (26)\\
\tau &\sim G(t), (\tau \geq 0).
\end{aligned}
$$

where Σ and Ψ are arbitrary positive definite matrices. Let G belongs to the family $\mathcal{G}$ for which the condition (2) is satisfied. Taking the nonidentity matrix Σ is unessential, since we can rescale the model by multiplying $\underset{\sim}{X}$ with $\Sigma^{-\frac{1}{2}}$. Consequently, let $\Sigma \equiv I$.

Denote with $\phi_{p,\Sigma}(\underset{\sim}{x} - \underset{\sim}{\mu})$ the density of $\mathcal{MVN}_p(\underset{\sim}{\mu}, \Sigma)$ distribution. The marginal distribution of $\underset{\sim}{X}$ in the model (26) can be expressed as

$$m(\underset{\sim}{x}) = \int_T \phi_{p,I+t\Psi}(\underset{\sim}{x}) dG(t).$$

Mimicking the calculation in the Proof of Theorem 3.1, we get

$$r_\Gamma \leq \sup_{G \in \mathcal{G}} E(p - tr(I + \tau\Psi)^{-1}). \qquad (27)$$

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

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

and the corresponding linear Γ-minimax risk is

$$p - tr(I + (\sup_{G \in \mathcal{G}} E\tau)\Psi)^{-1}. \qquad (28)$$

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

$$\rho^\star = \frac{p - tr(I + (\sup_{G \in \mathcal{G}} E\tau)\Psi)^{-1}}{\sup_{G \in \mathcal{G}} (p - trE(I + \tau\Psi)^{-1})}. \qquad (29)$$

As we can see, the bound (29) depends on the dimension p of the model.

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

Example 5.1 :

If a random variable τ is as in (21), i.e. belongs to the class of all symmetric and unimodal distributions on $[a, b]$, then the relation $\rho^\star = 1$ continues to be true. Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of the matrix Ψ. Then

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

and

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

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 5.2 :

The bound (24) can also be generalized. Let, again, $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of the matrix Ψ.

$$r_L = p - tr(I + \frac{mS^2}{m-2}\Psi)^{-1} = p - \sum_{i=1}^{p} \frac{m-2}{m-2+mS^2\lambda_i}.$$

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

Therefore,

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

References

[1] ALBERT, J. H. (1984). Estimation of a normal mean using mixtures of normal priors. *Technical Report, Bowling Green State University.*

[2] BERGER, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer, Berlin-Heildeberg-New York.

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

[4] BROWN, L. (1986). *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory.* Institute of Mathematical Statistics, Lecture Notes vol. **9**, Series Editor: Shanti S. Gupta.

[5] DASGUPTA, A., GHOSH, J.K., AND ZEN, M.M. (1990). Bayesian analysis of multivariate normal mean with flat tailed priors. *Technical Report* **90-22**, *Purdue University.*

[6] DONOHO, D., LIU, R.C., AND MCGIBBON, B. (1990). Minimax risk over hyperrectangles and implications. *Ann. Statist.* **18** *1416-1437.*

[7] 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.*

[8] EFRON, B. AND OLSHEN, R. (1978). How broad is the class of normal scale mixtures? *Ann. Statist.* **5** *1159-1164.*

[9] 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.*

[10] KUDŌ, H. (1967). On partial prior information and the property of parametric sufficiency. *Proc. Fifth Berkeley Symp. Math. Stat. 251-265.*

[11] ROBERT, C. (1990). Hidden mixtures and Bayesian sampling. *Rapport technique* **115**, *LSTA Université Paris 6.*

[12] 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.*

[13] SOLOMON, D.L. (1972). Λ-minimax estimation of multivariate location parameter. *J. Amer. Statist. Assoc.* **67** *641-646.*

[14] STEIN, C. (1983). *Lecture Notes in Statistics 336 (Decision Theory)*, Stanford, Fall 1982. (Notes prepared by Persi Diaconis)

[15] VIDAKOVIC, B., AND DASGUPTA, A. (1991). Linear versus nonlinear rules for estimating a bounded normal mean. *Contributed paper on: 54-th Annual Meeting of the Institute of Mathematical Statistics, 19-22. August 1991, Atlanta, Georgia.*

[16] WATSON, S.R. (1974). On Bayesian inference with incompletely specified prior distributions. *Biometrika* **61** *193-196.*