EFFICIENCY OF LINEAR RULES FOR ESTIMATING
A BOUNDED NORMAL MEAN*

by

Brani Vidakovic and Anirban DasGupta

Purdue University

Technical Report #92-23

Department of Statistics
Purdue University

May 1992

*Research supported by NSF grant DMS 8923071 at Purdue University.

Efficiency of linear rules for estimating a bounded normal mean *

BRANI VIDAKOVIC and ANIRBAN DASGUPTA
Purdue University

May 13, 1992

Abstract

Inference for restricted parameters is often considerably harder than it is for the unrestricted case. A simple example is the estimation of a univariate normal mean, when the mean is known to lie in a fixed bounded interval. Levit (1980), Casella and Strawderman (1981), Bickel (1981), DasGupta (1985) describe a minimax estimator in this case. The general problem remains unsolved. An alternative to deriving the exact optimal rule is to use an easily computable rule with very good or near optimal performance. Donoho, Liu and MacGibbon (1990) demonstrate that the linear minimax rule is at most (about) 25% worse than the exact minimax rule uniformly over all bounded intervals. We consider a Bayesian version of this problem where the unknown mean is assigned a prior belonging to an appropriate family Γ of prior distributions. Using the criterion of usual minimaxity amounts to allowing all possible prior distributions. We show that if Γ is the class of all symmetric and unimodal priors, the linear minimax rule is at most 7.4% worse than the exact minimax rule, again uniformly over all bounded intervals. We also consider the high dimensional problem when the unknown mean is known to lie inside a sphere, and has a spherically symmetric and unimodal distribution.
Key Words: linear rule, Gamma minimax, linear Gamma minimax, Bayes risk, spherically symmetric, unimodal.
AMS 1985 subject classification: 62C20

1 Introduction

It is of a general statistical interest that estimators of unknown parameters be easily calculable and simple. As a rule, optimal estimators are sometimes not so. For example, in Bayesian estimation, in case of non-conjugate families of prior distributions, Bayes estimators can be difficult to calculate and investigate. We may instead use non-optimal but easy estimators, like linear ones, whenever the loss of efficiency, measured by some reasonable criteria, is small. Donoho, Liu and MacGibbon (1990) prove that in minimax estimation of a bounded normal mean, use of linear rules will produce a loss of efficiency of at most 25%, irrespective of the size of the compact bound on the mean. Their efficiency measure is the ratio of the minimax risks of the linear and plain minimax rules.

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

We consider the Bayesian version of the same problem. Restricting the class of all priors on the mean to unimodal and symmetric distributions, we prove the fact that we may lose at most 7.4% by using nonoptimal linear rules. As a criteria of optimality we compare corresponding Γ-minimax risks.

In the Γ-minimax approach, the statistician specifies a *class* Γ of prior distributions on the parameter space and for each decision rule from an appropriate class, calculates the supremum of the Bayes risk over the family Γ and then selects the decision rule that minimizes the above supremum.

This is attractive as a middle ground between the subjective Bayes setup, which is Γ-minimax for a family of prior distributions containing a single prior, and the full minimax setup which, as another extreme, is Γ-minimax for the family of *all* prior distributions. Though many Bayesians object to the use of Bayes risk as a measure of performance (since it contains averaging with respect to the sample space), it is a fact that frequentist measures do play a role in Bayesian analysis, especially for checking robustness of procedures.

For a nice discussion on the Γ-minimax approach and its relation to robust Bayesian estimation, see Berger (1984, 1985).

Let us consider the problem of estimating a bounded normal mean under squared error loss, and let X be a single observation from the $\mathcal{N}(\theta, 1)$ distribution where $\theta \in [-m, m]$. No generality is lost by the assumptions that the variance is 1, and that the interval is symmetric. Restricting the parameter space seems to be reasonable since in real life there hardly exists an example in which the mean is truly unbounded.

In the case where no more prior knowledge is imposed on θ, the problem has been studied by Miyasawa (1953), Levit (1980), Casella and Strawderman (1981), Bickel (1981), DasGupta (1985), and Gatsonis, MacGibbon and Strawderman (1987), among others.

Suppose that, in addition to boundedness, we believe that θ, viewed as a random variable, has a symmetric and unimodal distribution. The properties of symmetry and unimodality are very natural in describing our prior knowledge (or prior ignorance) about θ. Therefore, let us pose the following model:

$$\left\{ \begin{array}{c} X|\theta \sim \mathcal{N}(\theta, 1), \\ \theta \sim \pi \in \Gamma, \Gamma \text{ is the set of all symmetric and unimodal priors on } [-m, m], \\ \text{Squared error loss.} \end{array} \right. \tag{1}$$

Olman and Shmundak (1985), Eichenauer (1990) and Eichenauer, Ickstadt and Weiss (1991) have studied the problem of estimating θ under model (1) for small values of m.

In addition, let $\mathcal{D}$ be the set of all decision rules and $\mathcal{D}_L$ be the set of all linear decision rules. The rule $\delta^\star \in \mathcal{D}$ is Γ-minimax if

$$\inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{\pi \in \Gamma} r(\pi, \delta^\star) = r_\Gamma. \tag{2}$$

The quantity r_Γ is the corresponding Γ-minimax risk. Similarly, the rule $\delta_L^\star \in \mathcal{D}_L$ is linear Γ-minimax if

$$\inf_{\delta \in \mathcal{D}_L} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{\pi \in \Gamma} r(\pi, \delta_L^\star) = r_L, \tag{3}$$

and the quantity r_L is the corresponding linear Γ-minimax risk.

For the family of priors Γ as defined above, the ratio $\rho = \rho(m) = \frac{r_L}{r_\Gamma}$ measures how good the linear Γ-minimax rule is, as opposed to unconstrained Γ-minimax rule, in terms of appropriate Γ-minimax risks. If we believe in the Γ-minimax approach, then it seems reasonable to use easily calculable linear rules whenever ρ is close to 1.

Donoho, Liu and MacGibbon (1990) calculate an upper bound on $\mu^\star = \sup_m \rho(m)$ for the class Γ of *all* disrtibutions on bounded intervals. They prove that $\mu^\star$ (Ibragimov-Has'minskii constant, as they refer to it) is less than 1.25 and cite the work of Feldman (at Hebrew University of Jerusalem) who obtained $1.246 < \mu^\star < 1.247$.

By analytical and numerical considerations we establish a similar uniform upper bound on ρ when Γ is the class of all symmetric and unimodal distributions. It turns out that linear Γ-minimax rules in this case are more attractive since $\mu^\star < 1.074$. One can do even better. The bound for $\mu^\star$ of (approximately) 1.045 is obtained by replacing linear rules with truncated linear rules, i.e. with rules of the form:

$$\delta_{a,m}(x) = \begin{cases} -m, & x \le -\frac{m}{a} \\ ax, & -\frac{m}{a} \le x \le \frac{m}{a} \\ m, & x \ge \frac{m}{a} \end{cases} .$$

These necessarily take values inside the parameter space, unlike linear rules.

Any linear rule in this problem has the disturbing property that it is bound to assume values outside of the parameter space. Our calculations in the paper show that this can happen with the probability as large as 50% in the worst cases. With this in mind, we also consider the problem of deriving the linear Γ-minimax rule under the additional constraint that the estimate belongs to the parameter space with a prescribed probability of $1-\alpha$, for all θ in $[-m, m]$, where $0 \le \alpha < 1$ is any fixed number. Clearly, this leads to a larger loss of efficiency. But we prove that this is not much. Thus linear estimates can be chosen that they are inside of $[-m, m]$ with a large probability, these are still easily calculable, and yet the loss of efficiency is still typically small. This is attractive.

Motivated by the calculations in Ghosh (1964), we next consider the problem of deriving the Γ-minimax rule in the class of polynomials of a specified degree n, n any finite positive integer. The general problem is formulated, and the cubic case, in particular, is solved explicitly. In the process, we make some novel applications of canonical moments. This is of some independent theoretical interest.

The multivariate case, naturally, is even more important and interesting. Here, we let the unknown mean belong to a sphere of radius m, and assume that the prior is spherically symmetric and unimodal about the center of the sphere. The linear Γ-minimax estimate is derived in closed form, and analogous bounds on the loss of efficiency are derived. The argument here involves interesting uses of the multivariate Brown identity and calculations with the multivariate Bickel prior and induced marginal distributions. As in the univariate case, we again address the issue that linear estimates can go outside of the sphere in which the mean is assumed to lie and find new linear estimates which stay inside the sphere with prescribed probabilities and still keep the loss of efficiency within reasonable to very good bounds.

Sections 2 and 3 deal with the derivation of the linear Γ-minimax rule together with its efficiency calculations. Revised linear rules under probability constraints are derived in

Section 4. Section 5 discusses a number of modifications in the lines of Donoho et. al. In Section 6, polynomial rules and the corresponding theory are presented. Sections 7 and 8 address the multidimensional case. We would like to point out here that even though the results in the multivariate case are similar in spirit to those in the univariate case, due to the simple fact that one dimensional problems are easier to solve, we do have more explicit results in the univariate case. This is the central reason the univariate case is isolated for better focus and understanding.

The main message of this article is that easily computable rules can be verifiably near optimal according to standard criteria, in which case it is reasonable to use these instead of pursuing computationally hard exact optimal rules.

2 Linear Γ-minimax rule

If we restrict the decision rules to the class of linear rules of the form $\delta(x) = ax + b$, then the linear Γ-minimax rule and its risk is given by the following theorem.

Theorem 2.1 *In the model (1), the linear* Γ*-minimax rule is*

$$\delta_L^*(X) = \frac{m^2}{m^2+3}X, \tag{4}$$

with the corresponding Γ*-minimax risk*

$$r_L = \frac{m^2}{m^2+3}. \tag{5}$$

Proof: First, in a rule $\delta(x) = ax + b$ the constant b can be dropped, since the symmetry of the prior distribution of θ implies

$$E^{\theta}R(\theta, aX+b) \geq E^{\theta}R(\theta, aX).$$

We will use the fact that any symmetric and unimodal random variable θ has the representation

$$\theta = UZ, \tag{6}$$

where U is an uniform $\mathcal{U}[-1,1]$ random variable and Z is a nonnegative random variable, defined on $[0, m]$, and independent of U.

The frequentist risk of the linear rule $\delta(x) = ax$ is

$$R(\theta, aX) = E^{X|\theta}(\theta - aX)^2 = (a-1)^2\theta^2 + a^2,$$

and its Bayes risk with respect to the prior $\pi(\theta) \in \Gamma$ is

$$r(\pi, a) = a^2 + (a-1)^2 E^{\pi}\theta^2.$$

Taking into account that $EU^2 = \frac{1}{3}$, $EZ^2 \leq m^2$, and the independence of U and Z, we obtain

$$\begin{aligned}\sup_{\pi} r(\pi, a) &= a^2 + (a-1)^2 \sup_{\pi} E\theta^2 \\ &= a^2 + (a-1)^2 \frac{1}{3} \sup EZ^2 \\ &= a^2 + (a-1)^2 \frac{m^2}{3}, \end{aligned} \tag{7}$$

and the infimum of (7) over a is achieved at $a = \frac{m^2}{m^2+3}$. Therefore, the linear Γ-minimax rule is $\delta_L^*(x) = \frac{m^2}{m^2+3} x$, and the corresponding Γ-minimax risk is

$$r_L(m) = \inf_a \sup_{\pi} r(\pi, a) = \frac{m^4}{(m^2+3)^2} + \frac{3}{(m^2+3)^2} m^2 = \frac{m^2}{m^2+3}. \quad \Box \tag{8}$$

Remark 2.1 *The following secondary fact is interesting: the normal* $\mathcal{N}(0, \frac{m^2}{3})$ *prior on the whole real line yields the same Bayes rule, and uniformly in* m*, gives the probability of* 0.9164 *to the interval* $[-m, m]$*. This prior with unbounded support can be viewed as an "approximation" to the whole class* Γ.

3 Γ-minimax risk

Let us define $\sup_{\pi\in\Gamma} \inf_{\delta\in\mathcal{D}} r(\pi, \delta) = \sup_{\pi\in\Gamma} r(\pi) \quad (= \underline{r})$. In general, $r_\Gamma \geq \underline{r}$ holds, and one of the results of this paper is the proof of the fact $r_\Gamma = \underline{r}$, i.e. that the corresponding statistical game has a value. The motivation for the above interchange of *inf* and *sup* is that it is the one of the principal ways to obtain Γ-minimax solutions. When the decision problem, viewed as a statistical game, has a value, then the Bayes solution with respect to the least favorable prior and the Γ-minimax solution coincide, and we can use the more powerful mathematical machinery of the dual problem.

Ghosh (1964), using a Hilbert space approach, derives a sequence u_n of estimators that approximate the Γ-minimax estimator when Γ is the class of all distributions on $[-m, m]$. The risks of estimators $u_n(x) = \sum_{i=1}^{n} \alpha_{ni} x^{2i-1}$ tend to the risk of the Γ-minimax estimator uniformly on $[-m, m]$. For small values of n, Ghosh derives the values of α_{ni} explicitly.

Kempthorne (1987) gives conditions under which a more general statistical game has a value, proving that the support of the least favorable distribution is discrete and finite. He also gives a sequential algorithm for obtaining an approximation to the least favorable distribution of the underlying statistical game.

The following theorem gives an exact form of Γ-minimax rules in our setup.

Theorem 3.1 *Under the statistical model (1), the corresponding statistical game has a value*

$$\inf_{\delta\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\pi, \delta) = \sup_{\pi\in\Gamma} \inf_{\delta\in\mathcal{D}} r(\pi, \delta) = r(\pi_0, \delta_0).$$

The least favorable prior π_0 *is a finite linear combination of uniform distributions and the point mass at zero,*

$$\pi_0(\theta) = \alpha_0 \mathbf{1}(\theta = 0) + \sum_{i=1}^{n} \frac{\alpha_i}{2m_i} \mathbf{1}(-m_i \le \theta \le m_i),$$
$$0 < m_1 < \ldots < m_n = m, \quad \alpha_i \ge 0, \sum_{i=0}^{n} \alpha_i = 1; \tag{9}$$

the corresponding marginal density of X *is*

$$m(x) = \alpha_0 \phi(x) + \sum_{i=1}^{n} \frac{\alpha_i}{2m_i} (\Phi(x+m_i) - \Phi(x-m_i)), \tag{10}$$

and the Bayes rule $\delta_0(x) = x + \frac{m'(x)}{m(x)}$ *has the form*

$$\delta_0(x) = x - \frac{\alpha_0 x \phi(x) - \sum_{i=1}^{n} \frac{\alpha_i}{2m_i} (\phi(x+m_i) - \phi(x-m_i))}{\alpha_0 \phi(x) + \sum_{i=1}^{n} \frac{\alpha_i}{2m_i} (\Phi(x+m_i) - \Phi(x-m_i))}$$
$$(= \delta(x; 0, \alpha_0, m_1, \alpha_1, \ldots, m_n, \alpha_n)); \tag{11}$$

Furthermore, δ_0 *is* Γ*-minimax.*

The proof of Theorem 3.1 is given in the Appendix.

Remark 3.1 *Differing from the linear case, the explicit* Γ*-minimax solutions for general* m *seem to be intractable. Regardless of the fact that we know the form of the solution, the explicit expressions of the parameters* α_i *and* m_i *as functions of* m*, are unknown. The number* n *of uniforms in the least favorable distribution clearly increases with increase of* m*, but an explicit relationship is also unknown. Numerical evidence suggests that the point mass at zero alternatively comes in and out of the least favorable distribution with increase of* m*.*

In estabilishing an upper bound on $\rho(m)$, we consider small, moderate and large values of m separately.

3.1 Small m

For small m, the Γ-minimax rule is Bayes with respect to the uniform $\mathcal{U}[-m, m]$ prior. This is a special case of a more general result proved in DasGupta and Delampady (1990). Namely, when Γ is the class of all spherically symmetric and unimodal distributions on the p-dimensional ball $\Sigma_p(\underset{\sim}{0}, m)$ of radius m, then δ_u, the Bayes rule with respect to the uniform distribution on the entire ball $\Sigma_p(\underset{\sim}{0}, m)$, minimizes $\sup_{\pi \in \Gamma} r(\pi, \delta)$, provided the risk function $R(\underset{\sim}{\theta}, \delta_u)$ is subharmonic. Subharmonicity (convexity in one dimension) is sufficient for the uniform $\mathcal{U}[-m, m]$ distribution to be least favorable. To search for the value of m (m_0, say) until which a least favorable distribution remains uniform, we need to inspect the shape of the function

$$\bar{R}(z,\delta_0) = \frac{1}{2z}\int_{-z}^{z} R(\theta,\delta_0)d\theta \tag{12}$$

where $R(\theta,\delta_0)$ is the frequentist risk of $\delta_0 = \delta(x;0,0,m,1)$, the Bayes rule with respect to the $\mathcal{U}[-m,m]$ prior. As long as $\bar{R}(m,\delta_0) > \bar{R}(z,\delta_0)$, holds for $z \in [0,m)$, the least favorable distribution remains uniform on $[-m,m]$ (See DasGupta and Delampady (1990)). Numerically we establish that the above inequality holds if $m \leq m_0 = 2.532258$.

The Brown identity (See Appendix) in the univariate case has the form

$$r(\pi) = 1 - \int_{-\infty}^{\infty} \frac{(m'(x))^2}{m(x)}dx, \tag{13}$$

where $m(x)$ is the marginal density of X.

Theorem 3.1 gives that $r_\Gamma = r(\pi_0)$, and that the marginal density corresponding to π_0 is

$$m(x) = \frac{\Phi(x+m) - \Phi(x-m)}{2m},$$

if π_0 is uniform on $[-m,m]$. Combining these facts we establish the following

Theorem 3.2 *For $m \leq m_0 = 2.532258$,*

$$r_\Gamma = 1 - \frac{1}{m}\int_0^{\infty} \frac{(\phi(x+m)-\phi(x-m))^2}{\Phi(x+m)-\Phi(x-m)}dx. \tag{14}$$

The expression (14) is much more convenient for numerical work than the integral

$$r_\Gamma = \frac{1}{2m}\int_{-m}^{m}\int_{-\infty}^{\infty}(\theta - \delta(x;0,0,m,1))^2\phi(x-\theta)dxd\theta.$$

Table 1 gives values of r_Γ,r_L, and ρ, for some values of $m \leq m_0$. The values of $\rho(m)$ require numerical evaluation of the integral (14). The following theorem allows us to place a completly analytic upper bound on $\rho(m)$.

Theorem 3.3 *(Chentsov (1967)) Let $\underset{\sim}{X}|\underset{\sim}{\theta} \sim \mathcal{N}_p(\underset{\sim}{\theta},I)$ and let $L(\underset{\sim}{\theta},\delta) = ||\underset{\sim}{\theta} - \delta||^2$. If π is the uniform prior on the hypercube $[-m,m]^p$, then*

$$r(\pi) \geq p\,(1 - \frac{\tanh(m)}{m}). \tag{15}$$

Corollary 3.1 *For any $m > 0$,*

$$r_\Gamma \geq 1 - \frac{\tanh(m)}{m}, \tag{16}$$

and

$$\rho(m) \leq \frac{m^2}{(3+m^2)(1-\frac{\tanh(m)}{m})}. \tag{17}$$

Though the lower bound on the Γ-minimax risk (15) holds for any m, it is sharp when m is small. For example, when $m = 1$, (15) gives $r_\Gamma \geq 0.23841$, while the exact value is 0.24919.

Table 1: Values of r_Γ, r_L, and ρ, for small m

m	r_Γ	r_L	ρ
0.1	0.0033226	0.0033226	1
0.2	0.0131579	0.0131579	1
0.3	0.0291260	0.0291262	$1+7\ 10^{-6}$
0.5	0.0769152	0.0769231	1.0001
1	0.24919	0.25	1.0032
1.5	0.42184	0.42857	1.0160
2	0.55121	0.57143	1.03668
2.5	0.63895	0.67568	1.05748
2.53226	0.64351	0.68127	1.05870

3.2 Moderate m

For $m > m_0$, the uniform $\mathcal{U}[-m, m]$ prior is no longer the least favorable and the rule $\delta_0 = \delta(x; 0, 0, m, 1)$ is not Γ-minimax. Numerical work shows that for $2.5323 \le m \le 3.2962$, the prior $\pi_0(\theta) = \alpha \mathbf{1}(\theta = 0) + \frac{1-\alpha}{2m}\mathbf{1}(-m \le \theta \le m)$ is the least favorable for appropriate α, and the rule $\delta_0 = \delta(x; 0, \alpha, m, 1-\alpha)$ is Γ-minimax. In Table 2 we give values of α,r_Γ,r_L, and ρ for some chosen values of m.

Table 2: Values of α, r_Γ, r_L, and ρ for moderate m

m	α	r_Γ	r_L	ρ
2.6	0.01280	0.65284	0.69262	1.06093
2.7	0.02910	0.66610	0.70845	1.06358
2.8	0.04274	0.67870	0.72325	1.06564
2.9	0.05404	0.69064	0.73707	1.06723
3.0	0.06335	0.70190	0.75000	1.06853
3.2	0.07665	0.72246	0.77341	1.07052

For $3.2962 \le m \le m^\star$, the least favorable prior has the form $\pi_0(\theta) = \frac{\alpha}{2m_1}\mathbf{1}(-m_1 \le \theta \le m_1) + \frac{1-\alpha}{2m}\mathbf{1}(-m \le \theta \le m)$. The corresponding Γ-minimax rule is $\delta_0 = \delta(x; 0, 0, m_1, \alpha, m, 1-\alpha)$. The value of $m^\star$ is between 4.5 and 5. Table 3 gives values of $\alpha, m_1, r_\Gamma, r_L$, and ρ, for chosen values of m. The numbers in the column marked by r_Γ in Table 3.3 for $m \ge 5$ are only lower bounds on r_Γ since the prior we used there, $\pi_0(\theta) = \frac{\alpha}{2m_1}\mathbf{1}(-m_1 \le \theta \le m_1) + \frac{1-\alpha}{2m}\mathbf{1}(-m \le \theta \le m)$, is not least favorable. Better bounds can be obtained by using the prior which is "next" in line, namely a linear combination of two uniform distributions and a point mass at zero. Specifying this prior is an extremely intensive calculational problem.

Table 3: Values of $\alpha, m_1, r_\Gamma, r_L$, and ρ for moderate m

m	m_1	α	r_Γ	r_L	ρ
3.2962	0	0.08137	0.73144	0.78363	1.07135
3.5	0.8819	0.10753	0.74894	0.80328	1.07256
3.75	1.2675	0.13642	0.76819	0.82418	1.07299
4	1.5897	0.17431	0.78535	0.84211	1.07227
4.25	1.8324	0.20498	0.80068	0.85757	1.07105
4.5	2.0525	0.23280	0.81440	0.87097	1.06946
m	m_1	α	$r_\Gamma >$	r_L	$\rho <$
5	2.4548	0.28175	0.83771	0.89286	1.06583
5.5	2.8523	0.32348	0.85653	0.90977	1.06216
6	3.2615	0.35778	0.87187	0.92308	1.05874
7	4.0980	0.41225	0.89509	0.94230	1.05247
8	4.9660	0.45400	0.91163	0.95522	1.04782

3.3 Large m

Another analytic lower bound on r_Γ can be given which is more useful for large values of m. Levit (1980) and Bickel (1981) have independently shown that under the model (1), when Γ is the family of *all* distributions on $[-m, m]$, the weak limit of the least favorable priors, rescaled to the interval $[-1, 1]$ is

$$g_1(\theta) = \cos^2(\frac{\pi\theta}{2})\, \mathbf{1}(|\theta| \le 1).$$

Fortunately, the prior

$$g_m(\theta) = \frac{1}{m} g_1(\frac{\theta}{m}) = \frac{1}{m}\cos^2(\frac{\pi\theta}{2m})\mathbf{1}(|\theta| \le m) \tag{18}$$

belongs to the family Γ of all symmetric and unimodal distributions on $[-m, m]$, and this fact is used to give estimates of r_Γ and $\rho(m)$ for large m, though the Bayes rule with respect to g_m is not asymptotically Γ-minimax. Denote by G_m the cdf of the density g_m.

To derive upper bounds on ρ for large values of m, we will use the following two facts:

Theorem 3.4 *(Borovkov-Sakhanienko (1980)). Let X have a density $p_\theta(x)$ and let $\theta \in \Theta$ have a prior density $\pi(\theta)$. Let, further, $\mathcal{I}(\theta) = E_\theta(\frac{\partial}{\partial\theta}\log p_\theta(X))^2$ be the Fisher information for p_θ. Suppose that $\mathrm{cl}\{\theta|\pi(\theta) > 0\} \subset \mathrm{int}\{\Theta\}$, where cl and int are the closure and interior of a set, respectively. Under the usual regularity conditions needed for the Cramér-Rao inequality to be valid, and under the squared error loss, the following lower bound on $r(\pi)$ holds:*

$$r(\pi) \ge \frac{C^2}{C + D}, \tag{19}$$

where $C = \int \frac{\pi(\theta)}{\mathcal{I}(\theta)} d\theta$ and $D = \int \frac{((\frac{\pi(\theta)}{\mathcal{I}(\theta)})')^2}{\pi(\theta)} d\theta$.

Theorem 3.5 *(Huber (1964). Among all absolutely continuous distributions F, with a density f, supported on the interval $[-m, m]$, the distribution G_m minimizes the Fisher information functional,*

$$\mathcal{I}(F) = \int \frac{(f'(x))^2}{f(x)} dx.$$

Furthermore, this minimal value is $\frac{\pi^2}{m^2}$.

Corollary 3.2

$$r_\Gamma \geq \underline{r} = \sup_\pi r(\pi) \geq \frac{m^2}{m^2 + \pi^2}. \tag{20}$$

Proof: When $X|\theta \sim \mathcal{N}(\theta, 1)$, and $\pi(\theta) = g_m(\theta)$, then $C = 1$ and $D = \mathcal{I}(G_m(\theta)) = \int \frac{(g'_m(\theta))^2}{g_m(\theta)} d\theta$. The bound (19), in this particular case, gives (20). □

Corollary 3.3 *The quantity $\rho - 1$ tends to 0, as $m \to \infty$, as fast as $\frac{C}{m^2}$. The constant C can be taken as $\pi^2 - 3$.*

Remark 3.2 *A different bound*

$$\sup_\pi r(\pi) \geq 1 - \frac{\pi^2 + \epsilon}{m^2}$$

was obtained by Bickel (1981) by using properties of the Fisher information functional.

A sharper lower bound on r_Γ, in fact an improvement of the Borovkov-Sakhanienko bound, can be obtained by using a result in Brown and Gajek (1990).

Theorem 3.6 *(Brown and Gajek (1990)). The bound (20) can be improved as follows:*

$$\sup_\pi r(\pi) = \underline{r} \geq r(G_m) \geq 1 - (1 + \frac{m^2}{\pi^2} + \beta)^{-1}, \tag{21}$$

where

$$\beta = \frac{4m^2}{\pi} \int_0^{\pi/2} \frac{(\sin(t) - 2t\cos(t))^2}{\pi^2 + 2m^2 t \cot(t)} dt.$$

For selected values of m, Table 4 gives the values of the bounds (20) and (21), as well as the corresponding bounds on ρ. The graphs of the bounds are given in Figure 1. The graph marked by 1 corresponds to the bound on ρ obtained by using the Borovkov-Sakhanienko inequality, while the graph marked by 2 corresponds to the bound gotten by using the Brown-Gajek improvement.

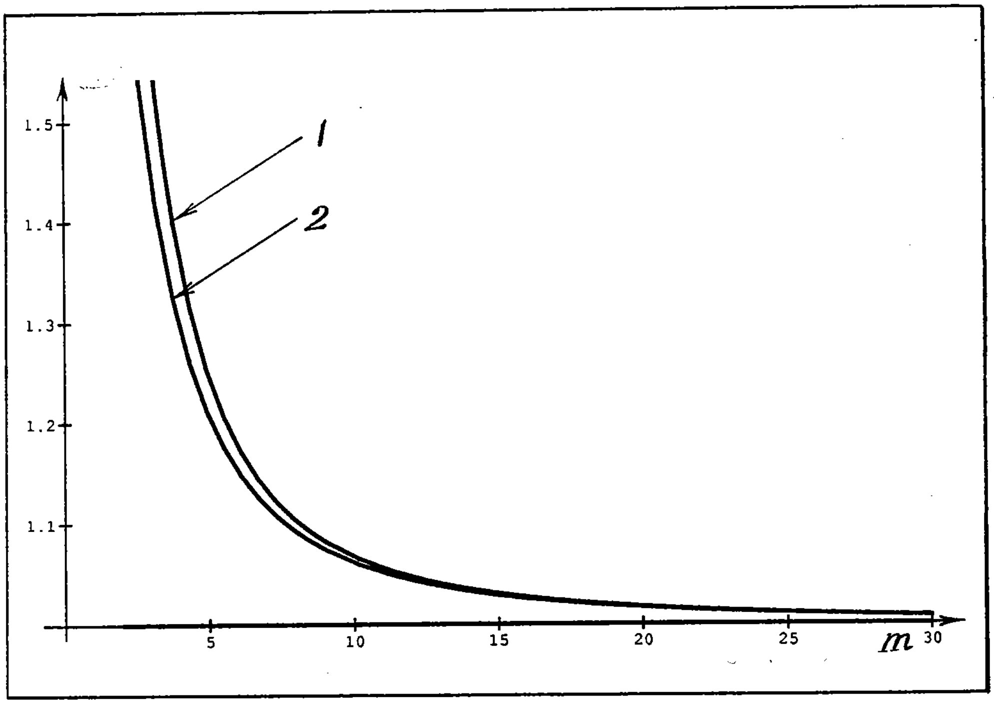

Figure 1: Bounds on ρ

Table 4: Values of the bounds by Borovkov-Sakhanienko and Brown-Gajek inequalities

m	Bound (20)	Bound (21)	Bound on ρ by (20)	Bound on ρ by (21)
6	0.78483	0.80137	1.17614	1.15188
7	0.83235	0.84427	1.13211	1.11612
8	0.86639	0.87511	1.10253	1.09154
10	0.91017	0.91509	1.06670	1.06096
12	0.93586	0.93882	1.04673	1.04343
15	0.95798	0.95951	1.03013	1.02849
20	0.97592	0.97654	1.01705	1.01640
30	0.98915	0.98931	1.00761	1.00744
50	0.99607	0.99610	1.00274	1.00272

Discussion: It is clear from the preceding results that we have used a variety of methods to bound the ratio ρ. For instance, we have used bounds implied by (i) a single uniform, (ii) linear combinations of many uniforms, and (iii) the Bickel prior. For the single uniform itself, we have sometimes used a fully analytic bound provided by Chentsov's inequality and at other times used the Brown identity followed by a numerical integration of the integral in (14). On the other hand, both the Borovkov-Sakhanienko inequality and the Brown-Gajek improvement were used to bound the Bayes risk with respect to the Bickel prior. Each of these methods can be used, in principle, for any m. However, using the uniform prior to bound r_Γ for large m is clearly not optimal, because use of the Bickel prior results in substantially sharper lower bounds in this case. Likewise, the use of the Bickel prior for small m is not advised. The Chentsov inequality has the aesthetic advantage of being fully analytic but a numerical evaluation of the exact Bayes risk with respect to the uniform prior on $[-m, m]$ produces sharper bounds on r_Γ. The ultimate goal of all of this analysis is to produce the sharpest possible uniform (in m) bound on ρ, while keeping the derivation as analytic as possible, using numerical work only as an unavoidable last resort. This calls for combining the bounds obtained above by various methods in just the optimal way so as to meet the aforementioned goal. This is done in the following single unifying table that gives the best possible bound on ρ for any given m. One can therefore regard Table 5 as the state of the art, so far as the loss of efficiency due to use of linear rules is concerned. Part 1 of Table 5 contains bounds on ρ obtained by Chentsov's inequality. Exact, but numerically obtained values of ρ are in part 2. Part 3 is made by using linear combinations of uniform priors. Finally, part 4 utilizes the Borovkov-Sakhanienko analytic bound on ρ.

The results obtained in the subsections 3.1, 3.2, and the current subsection 3.3 (small, moderate, and large m), prove the following theorem.

Theorem 3.7

$$\rho(m) \leq 1.074, \textit{ for all } m. \tag{22}$$

In other words, if we choose the family Γ of all symmetric and unimodal distribution to describe our prior knowledge about the parameter θ and use linear Γ-minimax rules, the loss of efficiency with respect to the calculationally hard, exact Γ-minimax rules is at most 7.4%. The general shape of the function $\rho = \rho(m)$ is given in Figure 2.

4 Probability constraints on linear rules

A somewhat disturbing property of linear Γ-minimax rules in our setup is that they might take values outside the set of possible values of the parameter they are estimating. In the most extreme case, the probability of $\delta_L(x)$ being outside $[-m, m]$ can be close to 0.5. The function

$$\xi(m|\theta) = P^{X|\theta}\left(\left|\frac{m^2}{m^2+3}X\right| \leq m\right) = \Phi\left(m + \frac{3}{m} - \theta\right) - \Phi\left(-m - \frac{3}{m} - \theta\right)$$

Table 5: Unifying table

	m	$\rho \leq$
1	0.1	1.00066
	0.2	1.00263
	0.3	1.00581
	0.5	1.01528
	0.8	1.03454
	1	1.04863
	1.2	1.06236
2	1.2	1.00702
	1.5	1.01595
	2	1.03668
	2.5	1.05748
3	2.8	1.06564
	3	1.06853
	3.2	1.07052
	3.5	1.07256
	3.75	1.07299
	4	1.07227
	4.5	1.06946
	5	1.06583
	6	1.06874
	7	1.05247
	8	1.04782
	10	1.04147
4	10	1.06690
	11	1.05540
	12	1.04673
	15	1.03013
	20	1.01705
	50	1.00274

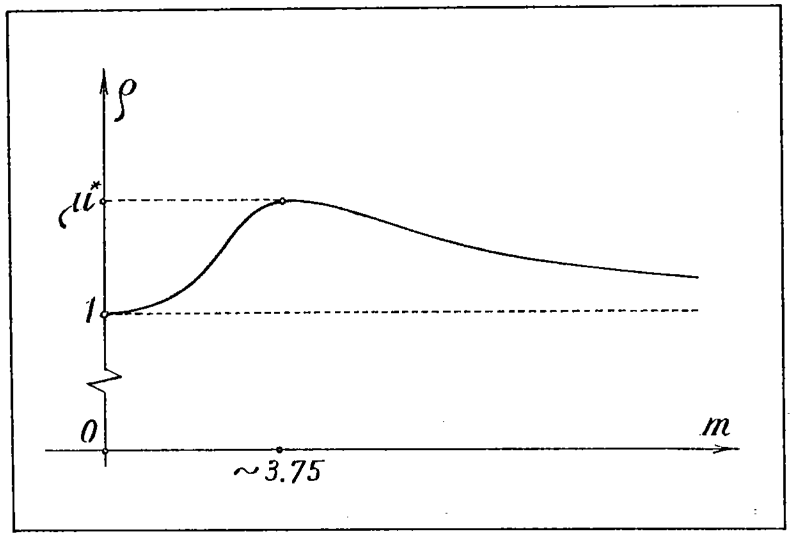

Figure 2: The function $\rho = \rho(m)$.

is decreasing in $|\theta|$, for fixed m, and hence

$$\max_{\theta} \xi(m|\theta) = \Phi(m + \frac{3}{m}) - \Phi(-m - \frac{3}{m}) = \xi_1(m) \text{ (say)}$$
$$\min_{\theta} \xi(m|\theta) = \Phi(\frac{3}{m}) - \Phi(-2m - \frac{3}{m}) = \xi_2(m) \text{ (say)}.$$

The function $\xi_2(m)$ has the limit 0.5, when $m \to \infty$.

A related natural problem is to derive an upper bound on the loss of efficiency by using a linear rule that belongs to the parameter space with a prescribed probability of at least $1 - \alpha$, uniformly in $|\theta| \leq m$. The following theorem is useful in this regard.

Theorem 4.1 *The rule $\delta_{L,\alpha}(x) = cx$, where*

$$c = \left(\frac{m^2}{m^2 + 3} \wedge \frac{m}{\sqrt{F^{-1}(1-\alpha)}} \right), \tag{23}$$

is Γ-minimax in the class of linear rules ax which satisfy $P^{X|\theta}(|aX| \leq m) \geq 1 - \alpha$, $\forall |\theta| \leq m$. Here $a \wedge b$ is the minimum of a and b, and $F^{-1}(1-\alpha)$ is the $(1-\alpha)$th quantile of the noncentral χ^2 distribution with one degree of freedom and parameter of noncentrality m^2. The corresponding Γ-minimax risk is

$$r_{L,\alpha} = c^2 + (1-c)^2 \frac{m^2}{3}. \tag{24}$$

Proof: Let us restrict our attention to only those linear rules for which

$$\min_{\theta} P^{X|\theta}(|aX| \leq m) \geq 1 - \alpha.$$

Due to the monotonicity of $P^{X|\theta}(|aX| \leq m)$ in $|\theta|$, for fixed m, the above constraint is equivalent to

$$P^{X|m}(|aX| \leq m) \geq 1 - \alpha \Leftrightarrow 0 \leq a \leq a_0, \tag{25}$$

where

$$a_0 = \frac{m}{\sqrt{F^{-1}(1-\alpha)}}. \tag{26}$$

The problem of determining the Γ-minimax linear rule in this subclass then amounts to maximizing $a^2 + (a-1)^2 \frac{m^2}{3}$ for $0 \leq a \leq a_0$. The assertion of the theorem now follows immediately.

Table 6 gives the slopes of the linear rule for $\alpha = 0.1, 0.05$ and 0.01, for different values of m. In the case when the slope c differs from $\frac{m^2}{m^2+3}$, the corresponding linear Γ-minimax risk as well as an upper bound on ρ are calculated. The next three theorems characterize the behavior of

$$\rho = \frac{r_{L,\alpha}}{r_\Gamma}.$$

Theorem 4.2 *$\rho \to 1$, as $m \to 0$.*

Proof: Easy.

Theorem 4.3 *$\rho \to 1 + \frac{z_\alpha^2}{3}$, as $m \to \infty$, where z_α is the $(1-\alpha)$th quantile of the standard normal distribution.*

Proof: It is well known (see, for example, Johnson and Kotz (1970), p.141) that

$$P(nc\chi_p^2(m^2) \leq x) = \Phi\left(\frac{x - p - m^2}{\sqrt{2(p + 2m^2)}}\right) + O\left(\frac{1}{m}\right), \tag{27}$$

uniformly in x, for any fixed p. This immediately implies that c converges to 1, as m goes to infinity. The same uniform approximation gives after a little work that $(1-c)m$ converges to z_α. Together, these imply that $r_{L,\alpha}$ converges to $1 + \frac{z_\alpha^2}{3}$. Since r_Γ converges to 1, as m goes to infinity, the theorem is proved. □

Finally, the following analog of Theorem 3.7 holds.

Theorem 4.4 *For any $0 < \alpha < 1$,*

$$\rho = \frac{r_{L,\alpha}}{r_\Gamma}$$

is bounded, uniformly in m.

Proof: Follows from Theorems 4.2 and 4.3, and the fact that ρ is a continuous function of m, for any α.

Remark 4.1 *An analog of Theorem 4.4 holds in the full minimax case (Donoho, Liu and MacGibbon setup) also. The proof is almost the same. Finding uniform upper bounds on ρ for different values of α may be a project of separate interest.*

Table 6: Slopes, risk, and ρ of restricted linear rules

m		$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
2	c		0.54872	0.46228
	$r_{L,\alpha}$		0.57263	0.59922
	ρ		1.03886	1.08710
3	c	0.70068	0.64588	0.56324
	$r_{L,\alpha}$	0.75973	0.79337	0.88952
	ρ	1.08239	1.13032	1.26730
4	c	0.75735	0.70861	0.63228
	$r_{L,\alpha}$	0.88760	0.95497	1.12095
	ρ	1.13020	1.21598	1.42733
5	c	0.79598	0.752462	0.68247
	$r_{L,\alpha}$	0.98045	1.07682	1.30598
	$\rho \leq$	1.17039	1.28543	1.55899
7	c	0.84525	0.80973	0.75056
	$r_{L,\alpha}$	1.10558	1.24697	1.57959
	$\rho \leq$	1.23516	1.39312	1.76473
10	c	0.88640	0.85875	0.81127
	$r_{L,\alpha}$	1.21585	1.40252	1.84545
	$\rho \leq$	1.33585	1.54094	2.02760
20	c	0.93978	0.92401	0.89503
	$r_{L,\alpha}$	1.36670	1.62378	2.27024
	$\rho \leq$	1.40042	1.66385	2.32626
50	c	0.97501	0.96815	0.95553
	$r_{L,\alpha}$	1.47106	1.78267	2.56102
	$\rho \leq$	1.47682	1.78965	2.57105
100	c	0.98735	0.98382	0.97726
	$r_{L,\alpha}$	1.50827	1.84054	2.67873
	$\rho \leq$	1.50975	1.84235	2.68136
250	c	0.99490	0.99346	0.99078
	$r_{L,\alpha}$	1.53170	1.87804	2.75265
	$\rho \leq$	1.53193	1.87832	2.75306
1000	c	0.99872	0.99836	0.99768
	$r_{L,\alpha}$	1.54348	1.89326	2.78950
	$\rho \leq$	1.54350	1.89328	2.78953

5 Truncated linear Γ-minimax rules

The fact that linear rules in a bounded parameter estimation problem necessarily take values outside the parameter space, as was elaborated in the previous section, leads us to consider modifications to linear rules.

Thus, for instance, take the truncated linear Γ-minimax rule:

$$\delta_m(x) = \begin{cases} -m, & x \leq -m - \frac{3}{m} \\ \frac{m^2}{m^2+3}x, & -m - \frac{3}{m} \leq x \leq m + \frac{3}{m} \\ m, & x \geq m + \frac{3}{m} \end{cases}$$

This estimate lies inside $[-m, m]$ with probability 1 for all θ. Calculations in the lines of Section 2 give that the loss of efficiency in comparison to the exact Γ-minimax rule is now at most 5.7%, uniformly in m. Of course, it is obvious that due to truncation the loss of efficiency will be smaller than in the untruncated case. This 5.7% bound is not hard to obtain.

One can do even better. Take any linear rule ax, and truncate it in the obvious way:

$$\delta_{a,m}(x) = \begin{cases} -m, & x \leq -\frac{m}{a} \\ ax, & -\frac{m}{a} \leq x \leq \frac{m}{a} \\ m, & x \geq \frac{m}{a} \end{cases} .$$

Then, for any given m, there exists an optimal choice of a in the sense of this article (it is not $\frac{m^2}{m^2+3}$!). This involves a moderate amount of calculational complexity. In any case, the efficiency calculations can be done again, and now it can be shown that the loss of efficiency is at most 4.5%, uniformly in m. Considering that it is a rather feeble improvement on the rule described above, it is perhaps a waste of time to try to use these rules as competitor to the exact Γ-minimax rule. We refrain from giving the technical details in these two cases because the calculations are largely similar to those in Section 2.

6 Polynomial Γ-minimax rules

The results in Section 3 show that the loss of efficiency due to the use of linear rules is at most 7.4%, uniformly over all compact intervals. This will naturally improve even more if one considers polynomial rules, of which linear rules are a special case. In this section, we formulate the general problem and work out the details for the cubic case. In the process, some interesting use of canonical moments is made.

Let $\mathcal{D}_n$ denote the class of all polynomial rules of the form

$$\delta_n(x) = \sum_{i=0}^{n} a_i x^i, n \in \mathbf{N}.$$

As an indirect consequence of the result of Ghosh (1964), exact Γ-minimax rules can be approximated arbitrarily close by Γ-minimax polynomial rules. The polynomial Γ-minimax rule (i.e. when the infimum in (2) is taken over $\mathcal{D}_n$) is skew-symmetric, i.e. $a_0 = a_2 = \ldots =$

$a_{[\frac{n}{2}]} = 0$. Define $\underset{\sim}{a} = (a_1, a_3, \ldots, a_{2k+1})'$ and $\underset{\sim}{y} = (x, x^3, x^5, \ldots, x^{2k+1})'$, where $k = [\frac{n-1}{2}]$. The frequentist risk of $\delta_n(x) = \underset{\sim}{a}'\underset{\sim}{y}$ is:

$$R(\theta, \delta_n(x)) = (\theta - \underset{\sim}{a}'E\underset{\sim}{y})^2 + \underset{\sim}{a}'\Sigma\underset{\sim}{a}, \tag{28}$$

where $\Sigma = Cov(\underset{\sim}{y}, \underset{\sim}{y})$. The quantities $E\underset{\sim}{y}$ and Σ can be expressed through Chebyshev-Hermite-like polynomials of θ. Let $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, and $D = \frac{d}{dx}$. The polynomials defined as

$$H_n(x) = \frac{(-D)^n\phi(x)}{\phi(x)}$$

are the usual Chebyshev-Hermite polynomials. Define now

$$Q_n(x) = \frac{1}{i^n}H_n(ix). \tag{29}$$

If $t_k = Q_k(\theta)$, then

$$E\underset{\sim}{y} = (t_1, t_3, \ldots, t_{2k+1})'$$

and

$$\Sigma = \begin{pmatrix} t_2 - t_1^2 & t_4 - t_1t_3 & \ldots & t_{2k+2} - t_1t_{2k+1} \\ t_4 - t_3t_1 & t_6 - t_3^2 & \ldots & t_{2k+4} - t_3t_{2k+1} \\ \vdots & & & \\ t_{2k+2} - t_{2k+1}t_1 & t_{2k+4} - t_{2k+1}t_3 & \ldots & t_{4k+2} - t_{2k+1}^2 \end{pmatrix}.$$

By simplifying (28) we get

Theorem 6.1

$$R(\theta, \underset{\sim}{a}'\underset{\sim}{y}) = \sum_{i\in\mathbf{O}_k} a_i^2t_{2i} + 2\sum\sum\nolimits_{i,j\in\mathbf{O}_k, i<j} a_ia_jt_{i+j} - 2\theta\sum_{i\in\mathbf{O}_k} a_it_i + \theta^2, \tag{30}$$

where $\mathbf{O}_k = \{1, 3, 5, \ldots, 2k+1\}$.

6.1 Cubic Γ-minimax rules

In finding the minimax solution we can interchange the *sup* and *inf*, since the corresponding statistical game can be formulated as a finite S-game and therefore has a value. Therefore to find the polynomial Γ-minimax rule we first minimize $E^\pi R(\theta, \underset{\sim}{a}'\underset{\sim}{y})$ with respect to $\underset{\sim}{a}$, for fixed π, and then maximize with respect to $\pi \in \Gamma$. For a fixed k, the least favorable distributions are linear combinations of at most $k+2$ uniform distributions, if we consider a point mass at zero as a degenerate uniform distribution (See Blackwell and Girshick (1954)).

We will elaborate the case $n = 3$ in detail. Larger values of n differ from the case $n = 3$ only by the amount of calculational complexity.

Theorem 6.2

$$\inf_{\delta\in\mathcal{D}_n}\sup_{\pi\in\Gamma} r(\pi,\delta) = \tag{31}$$
$$\sup_{0\le p_1,p_2,p_3\le 1} \frac{2ABCDE + AB^2E^2 - AC^2D^2 - A^2CE^2 + B^2CD^2 - 2B^3DE}{(AC-B^2)^2} + F,$$

where

$$\begin{aligned}
A &= \frac{1}{3}m^2\nu_1 + 1,\\
B &= \frac{1}{5}m^4\nu_2 + 2m^2\nu_1 + 3,\\
C &= \frac{1}{7}m^6\nu_3 + 3m^4\nu_2 + 15m^2\nu_1 + 15,\\
D &= \frac{1}{3}m^2\nu_1,\\
E &= \frac{1}{5}m^4\nu_2 + m^2\nu_1, \quad \textit{and}\\
F &= \frac{1}{3}m^2\nu_1,
\end{aligned}$$

and $\nu_i = \nu_i(p_1,p_2,p_3)$ are as in (63).

Proof: Let

$$\delta_3(x) = (a_1,a_3)\begin{pmatrix} x \\ x^3 \end{pmatrix} = a_1x + a_3x^3.$$

Then

$$\begin{aligned}
R(\theta,\delta_3(x)) &= (\theta - \underset{\sim}{a}'\begin{pmatrix} \theta \\ \theta^3+3\theta \end{pmatrix})^2 + \underset{\sim}{a}'\begin{pmatrix} 1 & 3\theta^2+3 \\ 3\theta^2+3 & 9\theta^4+36\theta^2+15 \end{pmatrix}\underset{\sim}{a}\\
&= a_1^2(\theta^2+1) + 2a_1a_3(\theta^4+6\theta^2+3) + a_3^2(\theta^6+15\theta^4+45\theta^2+15)\\
&\qquad -2a_1\theta^2 + -2a_3(\theta^4+3\theta^2) + \theta^2.
\end{aligned}$$

If we take the expectation of $R(\theta,\delta_3(x))$ with respect to θ, and use the representation (6) to replace $E\theta^n$ with $\frac{1}{n+1}EZ^n$, we get

$$r(\pi,\delta_3) = Aa_1^2 + 2Ba_1a_3 + Ca_3^2 - 2Da_1 - 2Ea_3 + F,$$

where ν_i is the i-th moment of $W = \frac{Z}{m} \in [0,1]$.

Minimizing $r(\pi,\delta_3)$ with respect to a_1 and a_3 first, we get, by standard calculus arguments, that the minimum

$$r(\pi,\delta_3^\star) = \frac{2ABCDE + AB^2E^2 - AC^2D^2 - A^2CE^2 + B^2CD^2 - 2B^3DE}{(AC-B^2)^2} + F,$$

is achieved for the rule $\delta_3^\star = a_1^\star x + a_3^\star x^3$, where

$$a_1^\star = \frac{DC-BE}{AC-B^2} \quad\text{and}\quad a_3^\star = \frac{AE-BD}{AC-B^2}.$$

To maximize $r(\pi, \delta_3^*)$ with respect to the moments ν_1, ν_2, and ν_3 of the random variable W, we will employ the *canonical moments* (see Appendix). Expressing ν_i's through canonical moments, as in (63), we transform the original extremal moment problem with complex boundary conditions, to an equivalent problem where boundary conditions on p_1, p_2, and p_3 are independent and simple. As a matter of fact, we then perform the maximization over the unit cube $[0,1] \times [0,1] \times [0,1]$. □

The numerical maximization (IMSL routine for constrained maximization DBCONF was used) gives that only two types of distributions can be least favorable in the cubic Γ-minimax problem. For $m < 2.7599$, the maximizing p_1 is equal to 1, which corresponds (regardless of p_2 and p_3) to the uniform $\mathcal{U}[-m, m]$ distribution of θ. Some values of $m < 2.7599$, and the corresponding values of $a_1^*, a_3^*, r_C, \frac{r_C}{r_\Gamma}$ (where r_C is the Γ-minimax risk of the cubic rule δ_3^*), are given below.

Table 7: Values of coefficients for the cubic rule and corresponding ρ

m	a_1^*	a_3^*	r_C	r_C/r_Γ
0.3	0.02962	-0.00016	0.029126	1
0.5	0.08023	-0.00102	0.076915	1
1	0.28032	-0.00777	0.24922	1.0001
1.5	0.50624	-0.01597	0.42241	1.0014
2	0.69311	-0.02	0.55315	1.0035
2.5	0.82672	-0.02	0.64193	1.0047
2.7	0.86704	0.01928	0.66862	1.0038

If $m \geq 2.7599$, then the maximizing p_1 is strictly less than 1, p_2 is equal to 1, and p_3 is arbitrary. This corresponds to the least favorable distribution of θ which is a linear combination of the uniform $\mathcal{U}[-m, m]$ and a point mass at zero, i.e. $\pi_0(\theta) = \alpha\delta(\{0\}) + (1-\alpha)\frac{1}{2m}\mathbf{1}(-m \leq \theta \leq m)$.

The cubic rules perform very well. Some bounds on r_C/r_Γ (for moderate m) in Table 8 are not sharp, and we think that the cubic rules are at most (about) 1% worse than the plain Γ-minimax rules. The drawback is the calculational complexity of such rules. Also, even though the loss of efficiency improves further from the linear case, cubic and other polynomial rules can badly suffer from taking values outside of the parameter space.

7 Multidimensional problem

Considerably harder and more numerically intensive but even more interesting is the multidimensional analogy of the problem. Let us consider the following model:

$$\left\{\begin{array}{c} \underset{\sim}{X} | \underset{\sim}{\theta} \sim \mathcal{MVN}_p(\underset{\sim}{\theta}, I), \\ \underset{\sim}{\theta} \sim \pi \in \Gamma - \text{the set of all spherically symmetric and unimodal priors on } ||\underset{\sim}{\theta}|| \leq m, \\ \text{Loss} \quad L(\underset{\sim}{\theta}, \underset{\sim}{\delta}) = ||\underset{\sim}{\theta} - \underset{\sim}{\delta}||^2. \end{array}\right. \quad (32)$$

Table 8: Continuation of the previous table with values of point mass at zero

m	$a_1^\star$	$a_3^\star$	$\alpha = 1 - p_1$	r_C	r_C/r_Γ
2.8	0.87906	-0.01845	0.01011	0.68054	1.0027
3	0.88499	-0.01581	0.05447	0.70300	1.0015
3.5	0.89928	-0.01080	0.13326	0.75217	1.0043
4	0.91228	-0.00748	0.18379	0.79215	1.0087
m	$a_1^\star$	$a_3^\star$	$\alpha = 1 - p_1$	r_C	$r_C/r_\Gamma <$
5	0.93347	-0.00379	0.24253	0.85037	1.0151
6	0.94885	-0.00207	0.27412	0.88861	1.0192
8	0.96793	-0.00075	0.30535	0.93256	1.0228
10	0.97835	-0.00033	0.31976	0.95526	1.0177
12	0.98451	-0.00016	0.32758	0.96831	1.0147
15	0.98984	-0.00007	0.33398	0.97938	1.0116
20	0.99417	-0.00002	0.33895	0.98825	1.0094
50	0.99905	-5.82 10^{-7}	0.34428	0.99809	1.0020
100	0.99976	-3.65 10^{-8}	0.34513	0.99952	~ 1

7.1 Affine Γ-minimax rules

Let us consider the set of affine rules

$$\mathcal{D} = \{A\underset{\sim}{X},\ A \text{ any } p \times p \text{ matrix}\}.$$

We prove the following theorem which is the multivariate analog of Theorem 2.1. In the following, $||A||^2$ denotes $trA'A$.

Theorem 7.1 *The affine* Γ*-minimax rule is*

$$\delta_L(\underset{\sim}{X}) = \frac{m^2}{p+2+m^2}\underset{\sim}{X} \tag{33}$$

with the risk

$$r_L = \frac{pm^2}{p+2+m^2}. \tag{34}$$

Proof: Any spherically symmetric and unimodal random variable $\underset{\sim}{\theta}$, restricted on the p-dimensional ball $\Sigma_p(\underset{\sim}{0}, m)$ of radius m, can be represented as a product UZ, where U is uniform on the unit ball $\Sigma_p(\underset{\sim}{0}, 1)$, and $Z \in [0, m]$ is an arbitrary univariate random variable independent of U. The density of U is

$$u(\underset{\sim}{\theta}) = \frac{\Gamma(\frac{p}{2}+1)}{\pi^{\frac{p}{2}}}\mathbf{1}(||\underset{\sim}{\theta}|| \le 1).$$

Straightforward calculation gives that

$$EUU' = \frac{1}{p+2}I.$$

Now, the frequentist risk of the affine rule AX is:

$$\|(A-I)\underset{\sim}{\theta}\|^2 + \|A\|^2$$

and hence,

$$\sup_{\pi} r(\pi, AX) = \frac{m^2}{p+2}\|A-I\|^2 + \|A\|^2,$$

for any A. By standard matrix calculus, one now has

$$\inf_{A}\left(\frac{m^2}{p+2}\|A-I\|^2 + \|A\|^2\right) = \frac{m^2}{p+2}\|A^\star - I\|^2 + \|A^\star\|^2,$$

where $A^\star = \frac{m^2}{p+2+m^2}\cdot I$.

The linear Γ-minimax risk is

$$r_L = \frac{m^2}{p+2}\, p\left(\frac{m^2}{p+2+m^2} - 1\right)^2 + p\left(\frac{m^2}{p+2+m^2}\right)^2 = \frac{pm^2}{p+2+m^2}. \quad \square \tag{35}$$

7.2 Γ-minimax risk

As in the case $p=1$ we will consider small, moderate and large values of m separately to derive a global upper bound on $\rho(m)$. However, it will be seen that the general technique adopted in the univariate case still works.

The following theorem is a multivariate generalization of Theorem 3.1. The proof is long but straightforward and is omitted.

Theorem 7.2 *Under statistical model (32), the corresponding statistical game has a value*

$$\inf_{\underset{\sim}{\delta}\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\pi, \underset{\sim}{\delta}) = \sup_{\pi\in\Gamma} \inf_{\underset{\sim}{\delta}\in\mathcal{D}} r(\pi, \underset{\sim}{\delta}) = r(\pi_0, \underset{\sim}{\delta}_0).$$

The least favorable prior π_0 is a finite linear combination of uniform distributions on p-dimensional subballs of radius m_i, and the point mass at zero,

$$\pi_0 = \alpha_0 \mathbf{1}(\underset{\sim}{\theta} = 0) + \sum_{i=1}^{n} \alpha_i \frac{\Gamma(\frac{p}{2}+1)}{\pi^{\frac{p}{2}} m_i^p} \mathbf{1}(\|\underset{\sim}{\theta}\| \le m_i),$$

$$0 < m_1 < m_2 < \ldots m_n = m, \quad \alpha_i \ge 0, \ \sum_{i=0}^{n} \alpha_i = 1; \tag{36}$$

the corresponding marginal density of $\underset{\sim}{X}$ is

$$m(\underset{\sim}{x}) = \alpha_0 \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}\|\underset{\sim}{x}\|^2} + \sum_{i=1}^{n} \alpha_i \frac{\Gamma(\frac{p}{2}+1)}{\pi^{p/2} m_i^p} F_{\frac{p}{2},\frac{1}{2},\|\underset{\sim}{x}\|}(m_i^2), \tag{37}$$

where F is the cdf of the Bessel distribution, with density given by

$$dF_{\nu,\alpha,\beta}(x) = \left(\frac{2}{\beta}\right)^{\nu-1} \alpha^\nu e^{-\frac{\beta^2}{4\alpha}} x^{\frac{1}{2}(\nu-1)} e^{-\alpha x} I_{\nu-1}(\beta\sqrt{x}) \mathbf{1}(x>0)dx. \tag{38}$$

The corresponding Bayes rule

$$\underset{\sim}{\delta}(\underset{\sim}{x}) = \underset{\sim}{x} + \frac{\nabla m(\underset{\sim}{x})}{m(\underset{\sim}{x})} \tag{39}$$

is, in addition, Γ-minimax.

7.3 Small values of m

For small values of m, Theorem 2.6 in DasGupta and Delampady (1990) implies that the prior

$$\pi(\underset{\sim}{\theta}) = \frac{\Gamma(\frac{p}{2}+1)}{\pi^{\frac{p}{2}} m^p} 1(\|\underset{\sim}{\theta}\| \le m), \tag{40}$$

is least favorable, since the frequentist risk function $R(\underset{\sim}{\theta}, \delta_\pi)$ is subharmonic. For arbitrary m the least favorable distributions are of the form given in Theorem 7.2. Numerically obtaining the Bayes risk $r(\pi_0)$ is extremely calculationally intensive. One elegant way to put lower bounds on r_Γ (for small values of m even exactly calculate r_Γ) is to employ the Brown identity with the uniform prior (40) (see Appendix for an exact statement of the Brown identity).

To transform the marginal distribution $M(\underset{\sim}{x})$, with the density

$$m(\underset{\sim}{x}) = \frac{\Gamma(\frac{p}{2}+1)}{\pi^{\frac{p}{2}} m^p} \int_{\|\underset{\sim}{\theta}\| \le m} \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}\|\underset{\sim}{x}-\underset{\sim}{\theta}\|^2} d\underset{\sim}{\theta} \tag{41}$$

to a form suitable for calculations, we will pretend, for a moment, that $\underset{\sim}{\theta}$ has the multivariate normal $\mathcal{MVN}_p(\underset{\sim}{x}, I)$ distribution. Then

$$\frac{1}{(2\pi)^{\frac{p}{2}}} \int_{\|\underset{\sim}{\theta}\|^2 \le m^2} e^{-\frac{1}{2}\|\underset{\sim}{\theta}-\underset{\sim}{x}\|^2} d\underset{\sim}{\theta} = P(\|\underset{\sim}{\theta}\|^2 \le m^2 | \underset{\sim}{\theta} \sim \mathcal{MVN}_p(\underset{\sim}{x}, I)). \tag{42}$$

In this case, $v = \|\underset{\sim}{\theta}\|^2$ has the noncentral χ^2 distribution, with p degrees of freedom and with the parameter of noncentrality $\|\underset{\sim}{x}\|^2$. The density function of v is given by

$$\varphi(v) = \frac{1}{2^{\frac{p}{2}}} e^{\frac{1}{2}(\|\underset{\sim}{x}\|^2+v)} v^{\frac{p}{2}-1} \sum_{m=0}^{\infty} (\frac{\|\underset{\sim}{x}\|^2}{4})^m \frac{v^m}{m!\Gamma(\frac{p}{2}+m)}. \tag{43}$$

After straightforward algebraic transformations, we get

$$\begin{aligned} \varphi(v) = & \ \frac{1}{2^{\frac{p}{2}}} e^{-\frac{1}{2}\|\underset{\sim}{x}\|^2} e^{-\frac{1}{2}v} v^{\frac{p}{2}-1} \textstyle\sum_{m=0}^{\infty} \frac{(\frac{\sqrt{v}\|\underset{\sim}{x}\|}{2})^{2m}}{m!\Gamma((\frac{p}{2}-1)+m+1)} \\ = & \ \frac{1}{2} \frac{e^{-\frac{1}{2}\|\underset{\sim}{x}\|^2}}{\|\underset{\sim}{x}\|^{\frac{p}{2}-1}} v^{\frac{1}{2}(\frac{p}{2}-1)} e^{-\frac{1}{2}v} \textstyle\sum_{m=0}^{\infty} \frac{(\frac{\sqrt{v}\|\underset{\sim}{x}\|}{2})^{2m+(\frac{p}{2}-1)}}{m!\Gamma((\frac{p}{2}-1)+m+1)}. \end{aligned}$$

The sum in the previous expression is exactly the definition of the modified Bessel function of the first kind, of order $\frac{p}{2}-1$ and argument $\sqrt{v}\|\underset{\sim}{x}\|$. The usual notation is $I_{\frac{p}{2}-1}(\sqrt{v}\|\underset{\sim}{x}\|)$. Using modified Bessel functions it is possible to define a family of Bessel distributions (see Laha (1954)), with densities given by (38), and in our case, $\varphi(v)dv = dF_{\frac{p}{2},\frac{1}{2},\|\underset{\sim}{x}\|}(v)$. Therefore,

$$P(\|\underset{\sim}{\theta}\|^2 \le m^2 | \underset{\sim}{\theta} \sim \mathcal{MVN}_p(\underset{\sim}{x}, I)) = \int_0^{m^2} dF_{\frac{p}{2},\frac{1}{2},\|\underset{\sim}{x}\|}(u) = F_{\frac{p}{2},\frac{1}{2},\|\underset{\sim}{x}\|}(m^2). \tag{44}$$

The marginal distribution $M(\underset{\sim}{x})$ therefore has the density

$$m(\|\underset{\sim}{x}\|) = \frac{\Gamma(\frac{p}{2}+1)}{\pi^{\frac{p}{2}} m^p} F_{\frac{p}{2},\frac{1}{2},\|\underset{\sim}{x}\|}(m^2). \tag{45}$$

Since the marginal density depends on $\underset{\sim}{x}$ only through $v = \|\underset{\sim}{x}\|$, then

$$\sum_{i=1}^{p}\left(\frac{\partial m(\underset{\sim}{x})}{\partial x_i}\right)^2 = \sum_{i=1}^{p}\left(m'(v)\frac{x_i}{v}\right)^2 = (m'(v))^2,$$

and

$$\mathcal{I}_p(M) = \frac{p\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2}+1)}\int_0^\infty v^{p-1}\frac{(m'(v))^2}{m(v)}dv.$$

The facts above and the Brown identity prove the following:

Theorem 7.3

$$r_\Gamma \geq p - \frac{p}{m^p}\int_0^\infty v^{p-1}\frac{(\frac{\partial}{\partial v}F_{\frac{p}{2},\frac{1}{2},v}(m^2))^2}{F_{\frac{p}{2},\frac{1}{2},v}(m^2)}dv. \tag{46}$$

The integral in (46) is now evaluated numerically to produce upper bounds on ρ, and the results in Table 9 come as a slight surprise. For small m the affine Γ-minimax rule does better with the increase of dimension.

Again, for small values of m, values in the ρ-columns in Table 9 are exact, while for larger values of m, the given numbers are only upper bounds on ρ. The analogy with the univariate case is complete. The replacement of the uniform prior (40) with the multivariate analogy of the least favorable prior, given by (36), yields the marginal distribution

$$m(\underset{\sim}{x}) = \alpha_0\frac{1}{(2\pi)^{p/2}}e^{-\frac{1}{2}\|\underset{\sim}{x}\|^2} + \sum_{i=1}^{n}\alpha_i\frac{\Gamma(\frac{p}{2}+1)}{\pi^{p/2}m_i^p}F_{\frac{p}{2},\frac{1}{2},\|\underset{\sim}{x}\|}(m_i^2), \tag{47}$$

$$\alpha_i \geq 0,\ \sum_{i=0}^{n}\alpha_i = 1,\ 0 < m_1 < m_2 < \ldots < m_n = m,$$

and then through the Brown identity yields the values of r_Γ and ρ.

Table 9: Bounds on ρ in multivariate case by the uniform prior

m	$p=3$ $\rho \leq$	$p=5$ $\rho \leq$	$p=7$ $\rho \leq$
1	1.0010	1.0002	1.00008
2	1.0144	1.0067	1.0035
3	1.0392	1.0225	1.0138
4	1.0571	1.0382	1.0263
5	1.0654	1.0483	1.0361
6	1.0678	1.0567	1.0423
7	1.0671	1.0551	1.0454
8	1.0650	1.0475	1.0359

7.4 Large values of m

For large values of m, as in the univariate case, we will use multivariate analogies to the Bickel prior, Borovkov-Sakhanienko inequality and Huber's result. This will give again an analytical bound on r_Γ.

Let J_t be the Bessel function of the first kind, order t, and let γ_t be its first positive zero. Let G_{1p} be the spherically symmetric distribution on $\Sigma_p(\underset{\sim}{0}, 1)$ given by the probability density function

$$g_{1p}(\|\underset{\sim}{\theta}\|) = C_p \|\underset{\sim}{\theta}\|^{-2t} J_t^2(\|\underset{\sim}{\theta}\|\gamma_t)\mathbf{1}(\|\underset{\sim}{\theta}\| \leq 1),$$

where

$$t = \begin{cases} \frac{p}{2} - 1 & \text{if } p \text{ is odd or divisible by 4} \\ -\frac{p}{2} + 1 & \text{if } p \text{ is even, not divisible by 4} \end{cases}$$

and C_p is the normalizing constant. For general $m > 0$, g_{mp} denotes $\frac{1}{m} g_{1p}(\frac{\theta}{m})$ and G_{mp} the corresponding cdf.

Shemyakin (1985) generalized the Borovkov-Sakhanienko bound (19) to the multivariate case.

Theorem 7.4 *(Shemyakin (1985)). Let $\underset{\sim}{X}$ be a single observation from a population with a distribution belonging to the family $\mathcal{P} = \{P_{\underset{\sim}{\theta}}, \underset{\sim}{\theta} \in \Theta\}$, where $\Theta \subset \mathbf{R}^p$ is compact. Suppose that the family $\mathcal{P}$ satisfies regularity conditions necessary for the Cramér-Rao inequality to hold. Let π be a prior on Θ, such that $\mathrm{cl}\{\underset{\sim}{\theta} | \pi(\underset{\sim}{\theta}) > 0\} \subset \mathrm{int}\{\Theta\}$, and let $H(\underset{\sim}{\theta}) = \pi(\underset{\sim}{\theta})\mathbf{I}(\underset{\sim}{\theta})$, where $\mathbf{I}(\underset{\sim}{\theta})$ is the Fisher information matrix for $P_{\underset{\sim}{\theta}}$.*

Then

$$r(\pi) \geq tr(E\mathbf{I}^{-1} - \Delta), \tag{48}$$

where $\Delta = E\frac{\nabla H(\nabla H)'}{\pi(\underset{\sim}{\theta})}$ and $(\nabla H)' = (\frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_p})H$.

The following result will also be useful.

Theorem 7.5 *(Bickel (1981)). G_{mp} uniquely minimizes $\mathcal{I}_p(F)$ among all spherically symmetric distributions F, concentrating on the ball $\Sigma_p(\underset{\sim}{0}, m)$, and furthermore,*

$$\mathcal{I}_p(G_{mp}) = \frac{4\gamma_t^2}{m^2}.$$

Corollary 7.1

$$r_\Gamma \geq r(G_{mp}) \geq p - \frac{4\gamma_t^2}{m^2}. \tag{49}$$

implying

$$\rho \leq \frac{pm^2}{(2 + p + m^2)(p - \frac{4\gamma_t^2}{m^2})}. \tag{50}$$

Proof: If $X|\theta \sim \mathcal{MVN}_p(\underset{\sim}{\theta}, I)$, and the prior distribution of $\underset{\sim}{\theta}$ is G_{mp}, then $\mathbf{I}(\underset{\sim}{\theta}) = I$, $tr\Delta = \mathcal{I}_p(G_{mp})$, and

$$r_\Gamma \geq r(g_{mp}) \geq p - \mathcal{I}_p(G_{mp}),$$

where $\mathcal{I}(F)$ is defined as in (61). The result now follows from Theorem 7.5. □

The first nonnegative zeros of J_t are given in the Table 10, and the implied bounds on ρ in Table 11.

Table 10: First nonnegative zeros of Bessel J function

p	t	γ_t
1	-0.5	$1.570796 = \frac{\pi}{2}$
2	0	2.404826
3	0.5	$3.141592 = \pi$
4	1	3.831706
5	1.5	4.493409
6	-2	5.135622
7	2.5	5.763459
8	3	6.380162
10	-4	7.588342
12	5	8.771484
15	6.5	10.512835

Table 11: Bounds on ρ by use of the Bickel prior

m \ p	3	5	7
8	1.1676	1.2057	1.2464
9	1.1246	1.1497	1.1755
10	1.0967	1.1146	1.1324
12	1.0636	1.0741	1.0841
15	1.0390	1.0448	1.0501
20	1.0213	1.0242	1.0267
30	1.0092	1.0104	1.0114
50	1.0033	1.0037	1.0040

As in the univariate case we give an unified table. The first part contains values and bounds on ρ implied by the uniform prior. In the second part of Table 12, bounds on ρ obtained by (50) are given. The uniform (in m) bounds on ρ are 1.1676, 1.2057, and 1.2464, for $p = 3, 5$, and 7, respectively. The given uniform bounds are probably very conservative. The work on sharpening these is still ongoing.

Table 12: Unified table

	m \ p	3	5	7
1	1	1.0010	1.0002	1.00008
	2	1.0144	1.0067	1.0035
	3	1.0392	1.0225	1.0138
	4	1.0571	1.0382	1.0263
	5	1.0654	1.0483	1.0361
	6	1.0678	1.0567	1.0423
	7	1.0671	1.0551	1.0454
	8	1.0650	1.0475	1.0359
2	8	1.1676	1.2057	1.2464
	9	1.1246	1.1497	1.1755
	10	1.0967	1.1146	1.1324
	12	1.0636	1.0741	1.0841
	15	1.0390	1.0448	1.0501
	20	1.0213	1.0242	1.0267
	50	1.0033	1.0037	1.0040

8 Probability constraints on affine rules

As in the univariate case, affine rules may take values outside the parameter space. Indeed, this may happen with an embarrassingly large probability.

Consider

$$P_m(\underset{\sim}{\theta}) = P_{\underset{\sim}{\theta}}\left(\|\frac{m^2}{m^2+p+2}X\| \leq m\right). \tag{51}$$

It is well known (see Anderson (1955)) that $P_m(\underset{\sim}{\theta})$ is monotone decreasing in $\|\underset{\sim}{\theta}\|$, and hence

$$\inf_{\|\underset{\sim}{\theta}\|\leq m} P_m(\underset{\sim}{\theta}) = P_m(m\underset{\sim}{e}), \tag{52}$$

where $\underset{\sim}{e} = (1, 0, \ldots, 0)'$. Table 13 gives values of this infimum for various m and p. As is apparent from this table, in the worst cases, the probability that the affine Γ-minimax rule takes values outside of the parameter space can approach 50%.

A much harder problem is to derive the rule which is Γ-minimax among all affine rules that have a prescribed probability of being in the p-ball of radius m. Mathematically, the problem is equivalent to:

$$\begin{aligned} \text{find} \quad & \inf_A(\sup_\pi r(\pi, A\underset{\sim}{X})), \text{ subject to:} \\ & P^{\underset{\sim}{\theta}}(\|A\underset{\sim}{X}\| \leq m) \geq 1-\alpha, \\ & \text{for all } \theta \text{ with } \|\underset{\sim}{\theta}\| \leq m. \end{aligned} \tag{53}$$

Our conjecture is that the minimizing A in this case is a multiple of the identity matrix, as in the unconstrained case. We were, however, unable to prove this.

Table 13: Minimal probabilities for different values p and m

$m \setminus p$	1	2	3	5
1	0.99865	0.99993	$1 - 2\ 10^{-6}$	1
2	0.93319	0.96587	0.98503	0.99824
3	0.84135	0.88370	0.91905	0.96688
5	0.72575	0.76013	0.79295	0.85217
10	0.61791	0.63717	0.65626	0.69371
50	0.52392	0.52791	0.53189	0.53985

A slightly more restricted class of affine rules allows us to carry the minimization process in this case.

Let A be a $p \times p$ matrix such that $A \leq \lambda I$ for fixed λ and let $\mathcal{A}$ be the class of all such matrices. Our goal is to find a matrix A in $\mathcal{A}$ for which $\sup_\pi r(\pi, A\underset{\sim}{X})$ achieves its minimum. By repeating the argument from Section 7.1 it follows that the minimizing A is

$$arg \min_{A \in \mathcal{A}} \frac{m^2}{p+2} tr(A-I)(A-I)' + trAA'. \tag{54}$$

If $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of the matrix A, the above problem is equivalent to finding

$$arg \min_{0 \leq \lambda_i \leq \lambda} \frac{m^2}{p+2} \sum (1-\lambda_i)^2 + \sum \lambda_i^2. \tag{55}$$

The solution is

$$\lambda_1 = \lambda_2 = \ldots = \lambda_p = \lambda \wedge \frac{m^2}{m^2+p+2}, \tag{56}$$

which gives the minimizing A as

$$A = (\lambda \wedge \frac{m^2}{m^2+p+2}) I. \tag{57}$$

For given m, p, and α we would like to find $\lambda_0 = \lambda_0(m, p, \alpha)$ such that the rule

$$(\lambda_0 \wedge \frac{m^2}{m^2+p+2}) \underset{\sim}{X} \tag{58}$$

is in the p-dimensional ball of radius m with the prescribed probability of at least $1-\alpha$.

Since

$$\{\underset{\sim}{X} |\ ||(\lambda \wedge \frac{m^2}{m^2+p+2}) \underset{\sim}{X}|| \leq m\} \supseteq \{\underset{\sim}{X} |\ ||\lambda \underset{\sim}{X}|| \leq m\},$$

one has

$$P(||(\lambda \wedge \frac{m^2}{m^2+p+2}) \underset{\sim}{X}|| \leq m) \geq P(\lambda^2 ||\underset{\sim}{X}||^2 \leq m^2).$$

Since $||\underset{\sim}{X}||^2$ is distributed as a noncentral χ^2 with p degrees of freedom and the parameter of noncentrality m^2, the choice

$$\lambda_0 = \frac{m}{\sqrt{F^{-1}(1-\alpha)}}, \tag{59}$$

where $F^{-1}(1-\alpha)$ is the $(1-\alpha)$th quantile of the above mentioned noncentral χ^2 distribution, guarantees that the prescribed probability requirement is satisfied.

Table 14 compares the values of $\frac{m^2}{m^2+p+2}$ and λ_0 for $\alpha = 0.05$ and $p = 5$ for chosen values of m.

Table 14: The values of λ_0 for affine rules ($\alpha = 0.05$ and $p = 5$)

m	$F^{-1}(1-\alpha)$	λ_0	$\frac{m^2}{m^2+p+2}$
1	13.170	0.27555	0.125
2	18.626	0.46342	0.36364
3	26.535	0.58239	0.5625
5	48.763	0.71602	0.78125
10	139.920	0.84540	0.93460
20	472.662	0.91993	0.98280
50	2671.250	0.96742	0.99721

As before, the affine rule (58) satisfies the following theorem:

Theorem 8.1 *Let $r_{L,\alpha}$ denote the Γ-minimax risk of the affine rule (58). Then, for any α, the ratio*

$$\rho = \frac{r_{L,\alpha}}{r_\Gamma}$$

is bounded, uniformly in m.

Proof: Similar to the proof of Theorem 4.4.

Next, we give Table 15, in which the constrained affine rules $c\underset{\sim}{X}$, the corresponding risks and bounds on ρ are calculated for $p = 3$ and $p = 5$.

9 Appendix

9.1 The Brown identity

Let $\underset{\sim}{X}|\underset{\sim}{\theta} \sim \mathcal{MVN}_p(\underset{\sim}{\theta}, I)$, and let Π be any prior distribution on $\underset{\sim}{\theta}$ such that $m(\underset{\sim}{x}) = \int \phi_p(\underset{\sim}{x} - \underset{\sim}{\theta}) d\Pi(\underset{\sim}{\theta}) < \infty$.

Let $r(\pi)$ denote the Bayes risk with respect to the prior π. Then,

$$r(\Pi) = p - \int_{\mathcal{R}^p} \frac{||\nabla m(\underset{\sim}{x})||^2}{m(\underset{\sim}{x})} d\underset{\sim}{x}. \tag{60}$$

Table 15: Restricted affine rules: Bounds on ρ

		$p=3$			$p=5$		
m		$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$
3	c		0.61156	0.53868			0.51724
	$r_{L,\alpha}$		1.93680	2.01974			2.83591
	$\rho \leq$		1.04364	1.08833			1.05989
5	c	0.77424	0.73352	0.66762	0.75423	0.71602	0.65372
	$r_{L,\alpha}$	2.56286	2.67933	2.99430	3.92294	4.00351	4.27800
	$\rho \leq$	1.07233	1.12106	1.25285	1.05278	1.07440	1.14263
10	c	0.87909	0.85199	0.80542	0.87196	0.84540	0.79968
	$r_{L,\alpha}$	3.19555	3.49208	4.21779	4.97259	5.28073	6.06373
	$\rho \leq$	1.22660	1.34042	1.61898	1.18610	1.25960	1.44637
20	c	0.93765	0.92196	0.89391	0.93553	0.91993	0.89203
	$r_{L,\alpha}$	3.57057	4.01169	5.09845	5.56362	6.06313	7.30931
	$\rho \leq$	1.23068	1.38272	1.75730	1.15955	1.26365	1.52338
50	c	0.97463	0.96778	0.95518	0.97426	0.96741	0.95483
	$r_{L,\alpha}$	3.81517	4.36699	5.75036	5.92903	6.57603	8.20195
	$\rho \leq$	1.27845	1.46337	1.92693	1.19352	1.32376	1.65106
100	c	0.98725	0.98372	0.97717	0.98715	0.98363	0.97708
	$r_{L,\alpha}$	3.89936	4.49335	5.99184	6.05177	6.75176	8.52567
	$\rho \leq$	1.30150	1.49976	1.99991	1.21231	1.35254	1.70789
250	c	0.99488	0.99345	0.99076	0.99487	0.99343	0.99075
	$r_{L,\alpha}$	3.95240	4.56967	6.14648	6.12369	6.86152	8.72768
	$\rho \leq$	1.31774	1.52364	2.04926	1.22505	1.37266	1.74599
1000	c	0.99872	0.99836	0.99768	0.99872	0.99836	0.99768
	$r_{L,\alpha}$	3.97536	4.60393	6.21554	6.15749	6.90476	8.82140
	$\rho \leq$	1.32514	1.53466	2.07187	1.23152	1.38097	1.76431

The Brown identity has been variously used by many authors, for example, Bickel (1981) and Donoho, Liu and MacGibbon (1990).

Notice, the function $m(\underset{\sim}{x})$ is exactly the marginal density of $\underset{\sim}{X}$ and the integral in (60) is the Fisher information integral $\mathcal{I}_p(\Phi_p * \Pi)$, defined as

$$\mathcal{I}_p(F) = \begin{cases} \int_{\mathcal{R}^p} \frac{\sum_{i=0}^{p} (\frac{\partial f(\underset{\sim}{x})}{\partial x_i})^2}{f(\underset{\sim}{x})} d\underset{\sim}{x}, & \text{if } F \text{ has a cont. differentiable density } f \\ \infty, & \text{otherwise} \end{cases} \tag{61}$$

and where the symbol $*$ denotes convolution. In other words, the Bayes risk $r(\Pi)$ is the difference of the dimension of the parameter space and the Fisher information of the marginal distribution.

9.2 Canonical moments

Let $\mathcal{P}$ denote the class of all probability measures on $[0,1]$. As in Skibinsky (1967), we define

$$M_n = \{(c_1, c_2, \ldots, c_n) | \; c_i = \int_0^1 x^i dP(x), \; P \in \mathcal{P}\}.$$

Let

$$\begin{aligned} \mu_n^- &= \min\{c \mid (c_1, \ldots, c_{n-1}, c) \in M_n\}, \\ \mu_n &= c_n, \\ \mu_n^+ &= \max\{c \mid (c_1, \ldots, c_{n-1}, c) \in M_n\}. \end{aligned}$$

We can speak in terms of *min* and *max* since M_n is a convex, bounded and closed set in $\mathbf{R}^n$. The canonical moments are defined as

$$p_n = \frac{\mu_n - \mu_n^-}{\mu_n^+ - \mu_n^-},$$

whenever $\mu_n^+ > \mu_n^-$.

Calculationaly, it is convenient to express μ_i's through canonical moments as follows (Skibinsky (1968)):

$$\mu_i = S_{i,i}$$

where

$$S_{0,j} = 1, \;\; j = 0, 1, 2, \ldots;$$

$$S_{i,j} = \sum_{k=i}^{j} \eta_{k-i+1} S_{i-1,k}; \;\; i, j = 1, 2, \ldots; i \leq j,$$

and

$$\eta_0 = q_0 = 1, \;\; \eta_j = q_{j-1} p_j, \;\; j = 1, 2, \ldots; \;\; p_i + q_i = 1.$$

Explicitly, the first three moments can be expressed as:

$$\begin{aligned} \mu_1 &= p_1, \\ \mu_2 &= p_1(p_1 + q_1 p_2), \\ \mu_3 &= p_1(p_1(p_1 + q_1 p_2) + q_1 p_2(p_1 + q_1 p_2 + q_2 p_3)). \end{aligned} \tag{62}$$

For nice examples of applications of canonical moments we refer the reader to Studden (1980), Lau and Studden (1985), DasGupta, Mukhopadhyay and Studden (1991), Skibinsky (1986), Dette (1990), among others.

9.3 Proof of Theorem 3.1

Let Π be an arbitrary distribution from the family Γ. Then Π is absolutely continuous, except possibly at the point $\theta = 0$.

(i) The density of Π on $[-m, m]\backslash\{0\}$ has the representation

$$\pi(\theta) = \int_{|\theta|}^{m} \frac{1}{2z} dF(z), \tag{63}$$

where $F(z)$ is the corresponding mixing distribution on $[0, m]$.

(ii) For any $z \in [-m, m]$, define "a new" risk function

$$\begin{aligned} \bar{R}(z, \delta) &= \frac{1}{2z} \int_{-z}^{z} R(\theta, \delta) d\theta, \\ \bar{R}(0, \delta) &= R(0, \delta). \end{aligned} \tag{64}$$

The Bayes risk of δ under the new risk and with respect to the "z-prior" F is

$$r'(F, \delta) = E^F \bar{R}(z, \delta).$$

Then

$$\begin{aligned} r'(F, \delta) &= \int_0^m \frac{1}{2z} \int_{-z}^{z} R(\theta, \delta) d\theta dF(z) \\ &= \int_{-m}^{m} \int_{|\theta|}^{m} \frac{1}{2z} dF(z) R(\theta, \delta) d\theta \\ &= \int_{-m}^{m} R(\theta, \delta) d\Pi(\theta) \\ &= r(\pi, \delta), \end{aligned}$$

and

$$\begin{aligned} \inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} r(\pi, \delta) &= \inf_{\delta \in \mathcal{D}} \sup_{F \text{ on } [0,m]} r'(F, \delta) \\ &= \sup_{F \text{ on } [0,m]} \inf_{\delta \in \mathcal{D}} r'(F, \delta) \\ &\quad \text{(from standard theory, since } F \text{ is arbitrary)} \\ &= \sup_{\pi \in \Gamma} \inf_{\delta \in \mathcal{D}} r(\pi, \delta). \end{aligned}$$

The least favorable prior F_0 in the new game with payoff $r'(F,\delta)$ is discrete. Moreover, for fixed m, the prior F_0 has a finite number of points of support (see Kempthorne (1987)). Because of the representation (6), the discrete distribution

$$F_0(z) = \alpha_0 \mathbf{1}(z=0) + \sum_{i=1}^{n} \alpha_i \mathbf{1}(z \geq m_i),$$
$$0 < m_1 < \ldots < m_n = m, \ \alpha_i \geq 0, \ \sum \alpha_i = 1,$$

corresponds to the prior distribution (9) from Γ, and straightforward calculation yields the rule (11).

We, however, need to check that the new statistical game and the risk $\bar{R}(z,\delta)$ satisfy conditions $(A1)$ to $(A5)$, as in Kempthorne (1987).

$(A1)$ *For every distribution F on $[0,m]$ the Bayes procedure δ_F is unique a.e. for all z.*

The uniqueness of the Bayes procedure δ_π in the original problem and the fact that Bayes solutions in both games are the same, i.e. $\delta_F \equiv \delta_\pi$, gives $(A1)$.

$(A2)$ *If $\{F_i, i=1,2,\ldots\}$ is any sequence of distributions on $[0,m]$ that converges weakly to F, then the risk functions $\{\bar{R}(z,\delta_{F_i}), i=1,2,\ldots\}$ of the corresponding Bayes procedures converge uniformly on compacts to the risk function $\bar{R}(z,\delta_F)$ of the Bayes procedure with respect to F.*

This property follows from the fact

$$|\bar{R}(z,\delta_{F_n}) - \bar{R}(z,\delta_F)| \leq \frac{1}{2z}\int_{-z}^{z} |R(\theta,\delta_{\pi_n}) - R(\theta,\delta_\pi)| d\theta, \tag{65}$$

where F_n and F, are the mixing distributions for the densities π_n and π, respectively. The fact that parameter space is bounded is crucial. Namely,

(i) By a characteristics function argument, it follows that $F_n \to F$ weakly implies that $\Pi_n \to \Pi$, weakly.

(ii) The weak convergence of Π_n to Π implies that

$$\delta_{\pi_n}(x) \longrightarrow \delta_\pi(x)$$

uniformly in x (see Rao (1962)).

(iii) Finally, the consequence of (ii) is that

$$R(\theta,\delta_{\pi_n}) \longrightarrow R(\theta,\delta_\pi)$$

uniformly in θ, which together with (65) gives $(A2)$.

$(A3)$ *The interval $[0,m]$ is a compact, separable metric space.*

$(A4)$ *$\bar{R}(z,\delta)$ is upper semicontinuous in θ for any measurable rule δ.*

$(A5)$ *$\bar{R}(z,\delta)$ is an analytic function of z, for any measurable decision rule δ.*

Both, $(A4)$ and $(A5)$, follow from the fact that $R(\theta,\delta)$ is well behaved, i.e. it is analytic for any measurable rule δ. $\square$

Acknowledgement. We are grateful to Larry Brown and Bill Strawderman for some helpful discussion and suggestions concerning the problem. Jayanta Ghosh, Bill Studden and Saurabh Mukhopadhyay pointed to the right references and made useful comments.

References

[1] ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** *170-176.*

[2] BERGER, J. (1984). The robust Bayesian viewpoint. In *Robustness of Bayesian Analyses* edited by J. Kadane. Elsevier Science Publishers *63-124.*

[3] BERGER, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, Springer Series in Statistics.

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

[5] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of games and statistical decisions.* John Wiley and Sons, New York.

[6] BOROVKOV, A. and SAKHANIENKO, A. (1980). On estimates of the expected quadratic risk. *Probab. Math. Statist.* **1** *185-195.*

[7] BROWN, L. (1971). Admissible estimators, recurrent diffusions and insoluble boundary-value problems. *Ann. Math. Statist.* **37** *1087-1136.*

[8] BROWN, L. and GAJEK, L. (1990). Information inequalities for the Bayes risk. *Ann. Statist.* **18** *1578-1594.*

[9] CASELLA, G. and STRAWDERMAN, W. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** *870-878.*

[10] CHENTSOV, N. (1967). On estimating an unknown mean of multivariate normal distribution. *Theory Probab. Appl.* **12** *560-574.*

[11] DASGUPTA, A. (1985). Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. *Sankhyā (A) 326-332.*

[12] DASGUPTA, A. and DELAMPADY, M. (1990). Γ-minimax estimation of vector parameters in restricted parameter spaces. *Technical Report* **90-42** *Department of Statistics, Purdue University.*

[13] DASGUPTA, A., MUKHOPADHYAY, S., and STUDDEN, W. (1991). Compromise designs in heteroscedastic linear models. *Technical Report* **91-48** *Department of Statistics, Purdue University.*

[14] DETTE, H. (1990). A generalization of D and D_1-optimal designs in polynomial regression. *Ann. Statist.* **18** *1784-1804.*

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

[16] EICHENAUER-HERRMANN, J. (1990). A gamma minimax result for the class of symmetric and unimodal priors. *Statistical Papers* **31** *301-304.*

[17] EICHENAUER-HERRMANN, J., ICKSTADT, K. and WEISS, E. (1991). Gamma-minimax results for the class of unimodal priors. *Technical report* **1388**, *Technische Hochschule Darmstadt.*

[18] GATSONIS, C., MACGIBBON, B., and STRAWDERMAN, W. (1987). The estimation of restricted normal mean. *Statist. Probab. Letters* **6** *21-30.*

[19] GHOSH, M. (1964). Uniform approximation of minimax point estimates. *Ann. Math. Statist.* **35** *1031-1047.*

[20] HUBER, P. (1964). Robust estimation of location parameter. *Ann. Math. Statist.* **35** *73-101.*

[21] JOHNSON, N. L. and KOTZ, S. (1970) *Continuous univariate distributions - 2.* John Wiley and Sons, New York.

[22] KEMPTHORNE, P. (1987). Numerical specification of discrete least favorable prior distributions. *J. Sci. Stat. Comput.* **8** *171-184.*

[23] LAHA, R. (1954). On some properties of the Bessel function distribution. *Bull. Calcutta Math. Soc.* **46** *59-72.*

[24] LAU TAI-SHING and STUDDEN, W. J. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments. *Ann. Statist.* **13** *383-394.*

[25] LEVIT, B. (1980). On asymptotic minimax estimates of second order. *Theory Probab. Appl.* **25** *552-568.*

[26] MIYASAWA, K. (1953). On the minimax point estimations. *Bull. Math. Statist.* **5** *1-17.*

[27] OLMAN, V. and SHMUNDAK, A. (1985). Minimax Bayes estimation of the mean of normal law for the class of unimodal priors. (In Russian) *Proc. Acad. Sci. Estonian Physics Math.* **34** *148-153.*

[28] RAO, R. (1972). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** *659-681.*

[29] SKIBINSKY, M. (1967). The range of the $(n+1)th$ moment for distributions on $[0,1]$. *J. Appl. Prob.* **4** *543-552.*

[30] SKIBINSKY, M. (1968). Extreme nth moments for distributions on $[0,1]$ and the inverse of a moment space map. *J. Appl. Prob.* **5** *693-701.*

[31] SKIBINSKY, M. (1986). Principal representations and canonical moment sequences for distributions on an interval. *J. Math. Anal. Appl.* **120** *95-117.*

[32] STUDDEN, W (1980). D_S-optimal designs for polynomial regression using continued fractions. *Ann. Statist.* **8** *1132-1141.*

[33] SHEMYAKIN, A. (1985). Rao-Cramér type integral inequalities for estimates of a vector parameter. *Theory Probab. Appl.* **32** *426-434*.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

brani@l.cc.purdue.edu
dasgupta@pop.stat.purdue.edu