SOME EFFICIENT SIMPLE RULES IN
\( \Gamma \)-MINIMAX ESTIMATION

Brani Vidakovic  
Institute of Statistics & Decision Sciences, Duke University  
&  
David Rios Insua  
Departamento de Inteligencia Artificial  
Facultad de Información  
Universidad Politécnica de Madrid

DP #93-A10
SOME EFFICIENT SIMPLE RULES IN $\Gamma$-MINIMAX ESTIMATION

BRANI VIDAKOVIC

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

DAVID RIOS INSUA

Departamento de Inteligencia Artificial
Facultad de Informatica
Universidad Politecnica de Madrid
28660 Madrid, Spain

Key Words and Phrases: Normal means, robust Bayesian analysis, $\Gamma$-minimax, linear rule, Fisher information, almost $\Gamma$-minimax.

Abstract. We discuss the $\Gamma$-minimax estimation of normal means under squared-error loss. Since this problem is computationally intensive, we study linear $\Gamma$-minimax estimation. We identify contexts in which linear rules are adequately efficient, in terms of the ratio of their risk to the optimal risk. We also provide strategies that may work in other contexts.
1 Introduction

We study the problem of normal mean estimation under squared-error loss within the robust Bayesian approach: we consider the case in which we elicit only a class \( \Gamma \) of priors. Since the computations typically undertaken in this approach are very involved (e.g., computation of \( \Gamma \)-nondominated estimators, bounds for the posterior expected loss, etc.), we adopt the \( \Gamma \)-minimax criteria. This provides us with a sensible way of choosing a rule in the case where no additional information about \( \Gamma \) is available, and we are interested in procedural robustness (see Berger, 1984).

The optimization problems are still very demanding. We find contexts in which we may limit our search to the class of linear rules. We express the efficiency of the linear \( \Gamma \)-minimax (LGM) rule in terms of the ratio between the risks of the \( \Gamma \)-minimax (GM) rule and the LGM rule. This requires the use of information theoretic inequalities together with procedures to compute distributions with minimum Fisher information. We illustrate these ideas with some popular classes in the robust Bayesian literature: \( \epsilon \)-Kolmogorov neighborhood classes, \( \epsilon \)-contamination classes, and quantile classes.

Since the resulting rules may not be sensible in certain contexts, e.g., when the parameter space is bounded, we suggest another type of efficient simple rules, which we call almost \( \Gamma \)-minimax.

2 The problem

Let \( X = (X_1, X_2, \ldots, X_n) \) be a sample from \( N(\theta, 1) \). We want to estimate \( \theta \) under squared-error loss. In the robust Bayesian spirit (Berger, 1993), we consider the case in which we elicited only a class \( \Gamma \) of priors. In this context, we would like to compute the class \( \Gamma^* \) of posteriors and the set of \( \Gamma^* \)-nondominated estimators. This is computationally involved. Additionally, the set of \( \Gamma^* \)-nondominated estimators is not typically a singleton. Therefore, it is convenient to adopt a \( \Gamma \)-minimax approach as a way to avoid computational problems and to provide protection against heavy losses due to unfavorable priors.

Let \( \mathcal{D} \) be the class of decision rules under consideration. Given a rule \( \delta \in \mathcal{D} \) and
a prior $\pi \in \Gamma$, define the Bayes risk

$$r(\pi, \delta) = E^\pi(E^{X|\theta}L(\theta, \delta)).$$

Given the class $\Gamma$ of priors, the $\Gamma$-minimax rule $\delta_\Gamma$ satisfies

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

Note that when $\Gamma$ is the set of all priors, this coincides with the minimax approach; when $\Gamma$ is a singleton, this is the Bayes risk criterion. Under suitable conditions - when the corresponding statistical game has a value - $\delta_\Gamma$ is Bayes with respect to $\pi_\Gamma$, the least favorable prior.

In general, the computation of GM rules is still too involved. As a consequence, we use the following strategy. Consider a subclass $\mathcal{D}_L \subset \mathcal{D}$, for which it is easy to compute

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

Clearly, $r_\Gamma \leq r_L$. The relative efficiency of the restricted rule $\delta_L$ may be measured by

$$e = \frac{r_\Gamma}{r_L}.$$ 

If we could estimate this ratio and show that it is close to 1, we would not lose too much using $\delta_L$ instead of $\delta_\Gamma$. Implementations of this idea in other statistical problems may be found in Donoho, Liu and MacGibbon (1990), Johnstone and MacGibbon (1992) and Vidakovic and DasGupta (1992). We consider robust Bayesian normal mean estimation, using the restricted class of linear rules and information theoretic results to estimate $e$.

Since the resulting rules are not sensible in certain contexts, we also study other simple rules which we call almost $\Gamma$-minimax rules, since they are close in a Bayes risk sense, to $\Gamma$-minimax rules.

## 3 LGM rules

Let $\mathcal{D}_L$ be the class of linear rules, i.e., rules $\delta$ such that

$$\delta(X) = a'X + b$$
with \( a \in \mathbb{R}^n, b \in \mathbb{R} \). We first compute the LGM rule, i.e., the \( \Gamma \)-minimax rule within \( \mathcal{D}_L \), under the assumption that the class consists of priors with bounded second moment.

**Proposition 1** Let \( X_1, \ldots, X_n \sim N(\theta, 1) \) i.d., \( L(\theta, \delta) = (\theta - \delta)^2 \), \( \mathcal{D}_L \) be the class of linear rules, and \( \Gamma = \{ \pi : E \theta^2 \leq M \} \). The LGM rule is

\[
\delta_L(X) = \frac{nM}{1 + nM} \bar{X},
\]

with LGM risk

\[
\tau_L = \frac{M}{nM + 1}.
\]

**Proof:** Let \( \delta(X) = a^T X + b \) with \( a \in \mathbb{R}^n, b \in \mathbb{R} \). We have

\[
r(\pi, \delta) = \Sigma a_i^2 + b^2 + 2b(\Sigma a_i - 1)E \theta + (\Sigma a_i - 1)^2 E \theta^2.
\]

Let \( \Psi(a, b) = \sup_{\pi \in \Gamma} r(\pi, \delta) \). Note that we may maximize \( r(\pi, \delta) \) by changing \( E \theta^2 \) and \( E \theta \) independently, as long as \( |E \theta| \leq \sqrt{M} \). Therefore,

\[
\Psi(a, b) = \begin{cases}
\Sigma a_i^2 + b^2, & \text{if } \Sigma a_i = 1 \\
\Sigma a_i^2 + (\Sigma a_i - 1)^2 M, & \text{if } b = 0 \\
\Sigma a_i^2 + b^2 + (\Sigma a_i - 1)^2 M + 2b(\Sigma a_i - 1)\sqrt{M}, & \text{if } b(\Sigma a_i - 1) > 0 \\
\Sigma a_i^2 + b^2 + (\Sigma a_i - 1)^2 M - 2b(\Sigma a_i - 1)\sqrt{M}, & \text{if } b(\Sigma a_i - 1) < 0
\end{cases}
\]

If \( \Sigma a_i = 1 \), we find, with the aid of Lagrange multipliers that

\[
\inf_{a, b} \Psi(a, b) = \Psi(a^1, b_1)
\]

with \( a_1 = 1/n, b_1 = 0 \), and risk \( 1/n \).

If \( b = 0 \),

\[
\inf_a \Psi(a, 0) = \Psi(a^2, 0)
\]

with \( a_2^2 = M/(nM + 1) \), and risk \( M/(nM + 1) \).

If \( b(\Sigma a_i - 1) > 0 \), the stationary point of \( \Psi(a, b) \) is outside the region, so the minima are on the border, as computed above. Similar results hold if \( b(\Sigma a_i - 1) < 0 \). \( \square \)

Note that \( \delta_L \) coincides with the linear Bayes rule for a prior \( \pi \) such that \( E^\pi \theta = 0 \),
and $E^x\theta^2 = M$ (Goldstein, 1974). This is a consequence of the fact that supremum and infimum can be interchanged in the statistical game involving the class of linear rules.

Note also that a result similar to Proposition 1 holds for the class $\Gamma = \{\pi : |E\theta| \leq F, Var(\theta) \leq V\}$, with $(V + m^2)$ in place of $M$ in the definitions of the linear $\Gamma$-minimax rule and risk, and $m = \min(F, \sqrt{V})$. Therefore, the forthcoming results may be applied to this class. Note that it may be easier to elicit bounds on the mean and variance than on the second moment. However, we choose the first class for economy.

To carry out our program, we need to compute $r_{\Gamma}$, the $\Gamma$-minimax risk. This task is complicated unless we know the underlying statistical game has a value. However, we may appeal to several bounds on $r_{\Gamma}$ to estimate the efficiency of linear rules. Specifically, we shall use one based on Schutzenberger's (1957) inequality. Since this result is not well-known, we review it for completeness. Given a density $g$, $I(g) = \int_{\Theta} \frac{\partial^2}{\partial \theta^2} \log g(\theta) \, d\theta$ will designate the Fisher information of the corresponding distribution. Given a prior $\pi$, $r(\pi)$ will designate its Bayes risk.

**Proposition 2** Let $X = (X_1, \ldots, X_n)$ be a sample from a population with density $f_\theta(x)$, and $\theta$ be a random variable with prior density $\pi(\theta)$. Suppose that Cramér-Rao regularity conditions hold for the corresponding posterior $\pi_x(\theta)$. Let $I = E^X(I(\pi_X))$, with $E^X$ the expectation taken w.r.t. the marginal distribution of $X$. Then,

$$r(\pi) \geq \frac{1}{I}.$$

**Proof:** Let $\rho(\pi_x, \delta_\pi)$ be the posterior expected loss of the Bayes rule $\delta_\pi$. Then

$$r(\pi) = E^X \rho(\pi_X, \delta_\pi) \geq E^X \frac{1}{I_X} \geq \frac{1}{E^X I_X} = \frac{1}{I}.$$

This inequality has a simpler form in our context.

**Corollary 1** When $f_\theta(x) = \Pi_{i=1}^n \phi(x_i - \theta)$, with $\phi$ the standard normal density,

$$r(\pi) \geq \frac{1}{n + I(\pi)}.$$

**Proof:** The identity

$$-\frac{\partial^2}{\partial \theta^2} \ln \pi_x(\theta) = -\frac{\partial^2}{\partial \theta^2} \ln \pi(\theta) - \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \phi(x_i - \theta),$$

5
implies
\[ I = I(\pi) + n. \]
\[ \square \]

As an immediate consequence, we have the bound on \( r_r \) that we shall use.

**Corollary 2**

\[ r_r \geq \frac{1}{n + \inf_{\pi \in \Pi} I(\pi)}. \]

Finally, we combine the above results to provide a bound on the efficiency of linear rules.

**Theorem 1** *Under the conditions of Proposition 1 and Corollary 1, let \( \pi_0 \) be such that \( I(\pi_0) = \inf_{\pi \in \Pi} I(\pi) \). Then*

\[ e \geq \frac{n + 1/M}{n + I(\pi_0)}. \]

Clearly, as \( n \to \infty \), the lower bound tends to 1. Thus, asymptotically, LGM rules are as good as GM rules. When \( n \) is small, this assessment of efficiency depends, essentially, on the relation between \( I(\pi_0) \) and \( 1/M \).

The problem is, therefore, the computation of \( I(\pi_0) \), which is a typical computational problem in robust Bayesian analysis (Berger, 1993). For several important cases, we may appeal to specific algorithms to compute \( I(\pi_0) \) and, therefore, study the efficiency of LGM rules, as we do in the next section.

## 4 Efficiency of LGM rules

We shall study the efficiency of LGM rules for three classes of priors widely used in Bayesian robustness: the \( \epsilon \)-Kolmogorov neighborhood of a standard normal distribution, the \( \epsilon \)-contamination of a standard normal distribution with arbitrary distributions as contamination, the class of distributions giving probability greater than \( \epsilon \) to the interval \([-1, 1]\).
Huber (1981) provides algorithms to compute minimum Fisher information distributions for these classes. We adapt these to our framework, imposing the additional constraint that the priors have second moments bounded by $M$. We assume that $M$ is not small, i.e., $M \geq M_0 = \int \theta^2 \pi_0(\theta) d\theta$, with $\pi_0$ designating the distribution minimizing Fisher information in the incumbent class. Therefore, the minimizing prior coincides in the initial, unconstrained class and the additionally constrained class. Implementations of the algorithms in Mathematica appear in the Appendix.

4.1 Kolmogorov neighborhoods of normal priors

Let

$$\Gamma_1 = \{ \Pi(\theta) \mid \sup_{\theta \in \Theta} |\Pi(\theta) - \Phi(\theta)| \leq \epsilon, \ E\theta^2 \leq M\},$$

where $\Phi$ is the standard normal distribution function.

If $\epsilon \leq 0.0303$, the prior $\pi_0$ for $\Gamma_1$ (Huber, 1981) is

$$\pi_0(\theta) = \pi_0(-\theta) = \begin{cases} \frac{\phi(\theta_0)}{\cos^2\left(\frac{\omega_0}{2}\right)} \cos^2\left(\frac{\omega_0}{2}\right), & 0 \leq \theta \leq \theta_0 \\ \phi(\theta), & \theta_0 \leq \theta \leq \theta_1 \\ \phi(\theta_1)e^{-\lambda(\theta-\theta_1)}, & \theta \geq \theta_1 \end{cases}$$

(2)

with $\theta_0, \theta_1, \omega$ given by:

$$
\begin{align*}
  u &= \omega \theta_0, \\
  \theta_0 &= \sqrt{u \tan \frac{u}{2}}, \\
  \epsilon &= \Phi(\theta_0) - \frac{1}{2} - \theta_0 \phi(\theta_0) \left(1 + \frac{\sin u}{1 + \cos u}\right), \\
  \epsilon &= \frac{\phi(\theta_1)}{\theta_1} - \Phi(-\theta_1).
\end{align*}
$$

If $\epsilon > 0.0303$, the prior $\pi_0$ for $\Gamma_1$ is

$$\pi_0(\theta) = \pi_0(-\theta) = \begin{cases} \frac{\phi(\theta_0)}{\cos^2\left(\frac{\omega_0}{2}\right)} \cos^2\left(\frac{\omega_0}{2}\right), & 0 \leq \theta \leq \theta_0 \\ \phi(\theta_0)e^{-\lambda(\theta-\theta_0)}, & \theta \geq \theta_0 \end{cases}$$

(3)

with

$$u = \omega \theta_0,$$
\[
\lambda \theta_0 = \omega \theta_0 \tan \frac{\omega \theta_0}{2}, \\
\theta_0 \phi(\theta_0) = \frac{\cos^2 \frac{\theta_0}{2}}{1 + \frac{\omega}{u \tan \frac{\theta_0}{2}}} \quad \text{and} \\
\epsilon = \frac{\phi(\theta_0) - \Phi(-\theta_0)}{\lambda}.
\]

The table below gives the lower bounds \( \epsilon \) of the efficiency of \( \Gamma_1 \)-minimax rules for selected values of \( \epsilon \) and \( n \), when \( M = 10 \).

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>( n = 5 )</th>
<th>( n = 10 )</th>
<th>( n = 50 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.85265</td>
<td>0.91974</td>
<td>0.98271</td>
</tr>
<tr>
<td>0.005</td>
<td>0.85997</td>
<td>0.92403</td>
<td>0.98369</td>
</tr>
<tr>
<td>0.01</td>
<td>0.86735</td>
<td>0.92831</td>
<td>0.98467</td>
</tr>
<tr>
<td>0.05</td>
<td>0.91006</td>
<td>0.95247</td>
<td>0.99004</td>
</tr>
<tr>
<td>0.1</td>
<td>0.94749</td>
<td>0.97278</td>
<td>0.99439</td>
</tr>
<tr>
<td>0.2</td>
<td>0.99159</td>
<td>0.99574</td>
<td>0.99914</td>
</tr>
</tbody>
</table>

Note that, the performance of linear rules improves with increasing sample size. For fixed \( n \), the risk of the linear \( \Gamma_1 \)-minimax rule is fixed for any \( \epsilon \) and depends only on \( M \). Since the \( \Gamma_1 \)-minimax risk increases when \( \epsilon \) increases, the relative efficiency of the linear \( \Gamma_1 \)-minimax rules increases.

### 4.2 \( \epsilon \)-contaminations

Let

\[
\Gamma_2 = \{ \Pi(\theta) : \Pi(\theta) = (1 - \epsilon) \Phi(\theta) + \epsilon H(\theta), E\theta^2 \leq M \},
\]

where \( H \) is an arbitrary probability distribution on the real line.

The prior \( \pi_0 \) for \( \Gamma_2 \) (Huber, 1981) is

\[
\pi_0(\theta) = \begin{cases} 
\frac{1 - \epsilon}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}}, & |\theta| \leq k \\
\frac{1 - \epsilon}{\sqrt{2\pi}} e^{-\frac{k^2}{2} - k|\theta|}, & |\theta| > k 
\end{cases}
\]

with

\[
\frac{2\phi(k)}{k} - 2\Phi(-k) = \frac{\epsilon}{1 - \epsilon},
\]

8
The table below gives the lower bounds $\epsilon$ of the efficiency of linear $\Gamma_2$-minimax rules for several values of $\epsilon$ and $n$, when $M = 10$.

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>$n = 5$</th>
<th>$n = 10$</th>
<th>$n = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.85134</td>
<td>0.91897</td>
<td>0.98254</td>
</tr>
<tr>
<td>0.005</td>
<td>0.85505</td>
<td>0.92115</td>
<td>0.98304</td>
</tr>
<tr>
<td>0.01</td>
<td>0.85877</td>
<td>0.92332</td>
<td>0.98353</td>
</tr>
<tr>
<td>0.05</td>
<td>0.87990</td>
<td>0.93552</td>
<td>0.98630</td>
</tr>
<tr>
<td>0.1</td>
<td>0.89928</td>
<td>0.94647</td>
<td>0.98873</td>
</tr>
<tr>
<td>0.2</td>
<td>0.92915</td>
<td>0.96294</td>
<td>0.99230</td>
</tr>
<tr>
<td>0.3</td>
<td>0.95248</td>
<td>0.97543</td>
<td>0.99495</td>
</tr>
</tbody>
</table>

Again note that the linear rules gain efficiency as the sample size increases and, that their performance improves as $\epsilon$ increases, due to fixed $r_L$, and increasing $r_R$.

4.3 Quantile class

Let

$$\Gamma_2 = \{\Pi(\theta)| \Pi([-1, 1]) \geq \epsilon, E\theta^2 \leq M\}.$$  \hspace{1cm} (7)

The prior $\pi_0$ for $\Gamma_3$ (Huber, 1981) is

$$\pi_0(\theta) = \begin{cases} 
\frac{C}{\cos^2 \frac{\psi}{2}} \cos^2 \frac{\omega \theta}{2}, & |\theta| \leq 1 \\
C e^{\lambda |\theta|}, & |\theta| > 1
\end{cases}$$ \hspace{1cm} (8)

with

$$\lambda = \omega \tan \frac{\omega}{2},$$

$$C = \frac{\cos^2 \frac{\psi}{2}}{1 + \omega \tan \frac{\omega}{2}}, \text{ and}$$

$$\epsilon = 1 - \frac{2C}{\lambda}.$$
The table below gives bounds for the efficiency of the linear $\Gamma_3$-minimax rule for several values of $\epsilon$ and $n$, when $M = 50$.

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>$n = 5$</th>
<th>$n = 10$</th>
<th>$n = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.99480</td>
<td>0.99739</td>
<td>0.99948</td>
</tr>
<tr>
<td>0.3</td>
<td>0.98178</td>
<td>0.99079</td>
<td>0.99814</td>
</tr>
<tr>
<td>0.4</td>
<td>0.96150</td>
<td>0.98034</td>
<td>0.99600</td>
</tr>
<tr>
<td>0.5</td>
<td>0.93236</td>
<td>0.96493</td>
<td>0.99277</td>
</tr>
<tr>
<td>0.6</td>
<td>0.89220</td>
<td>0.94292</td>
<td>0.98802</td>
</tr>
<tr>
<td>0.7</td>
<td>0.83787</td>
<td>0.91162</td>
<td>0.98095</td>
</tr>
<tr>
<td>0.8</td>
<td>0.76387</td>
<td>0.86590</td>
<td>0.96991</td>
</tr>
<tr>
<td>0.9</td>
<td>0.65682</td>
<td>0.70254</td>
<td>0.95018</td>
</tr>
</tbody>
</table>

Again, we see that these rules are more efficient as the sample size grows. They are less efficient, the more mass is put in the interval $[-1, 1]$. This is to be expected, since in this case we are closer to a case in which the space parameter is bounded. The use of linear rules is not very sensible in this context. See Vidakovic and DasGupta (1992) for a discussion, and Section 6 for related comments.

4.4 Other classes

As we have mentioned, the computation of the distribution minimizing Fisher information in a class may be seen as a general computational problem in Bayesian robustness. For other important classes, there are also specific results available, which may be used in a similar fashion. Collins and Wiens (1989) provide minimizing distributions for $(\epsilon, \delta)$-Levy neighborhoods. They also show that the minimizing distribution in an $\epsilon$-neighborhood with $\mathcal{L}^1$ distance is the minimizing distribution in an $\epsilon/4$-Kolmogorov neighborhood.

Bickel and Collins (1983) minimize Fisher information in classes of distributions of the form $\int F(., \gamma)\nu(d\gamma)$, where $F(., \gamma)$ is a parametric class of distributions, and $\nu$ is an arbitrary mixing measure. Unfortunately, not many closed form results are available for this family.
5 Bounding the maximum second moment

The results in Sections 4.1 and 4.2 suggest the relative inefficiency of LGM rules when \( n \) and \( \epsilon \) are small. This seems counterintuitive: in the limit, when \( \epsilon = 0 \), the optimal rule is linear. Therefore, we would expect LGM rules to perform well.

Note that, for fixed \( M \), the LGM rule protects us against a prior for which \( E\theta^2 = M \). This may be extremely conservative, especially when \( M \gg M_0 \). As \( \epsilon \to 0 \), \( M_0(\epsilon) \downarrow 1 \), which accentuates this. (\( M_0(\epsilon) \) is the second moment of the minimizing prior \( \pi_0 \), for the class associated with \( \epsilon \).) This suggests reducing \( M \) as \( \epsilon \) decreases. When we reduce \( \epsilon \), we imply more precision in the class of priors, thus we would expect that the priors in the class more closely resemble the standard normal distribution. Likewise, the moments of the priors resemble those of the standard normal. In our case, we take \( M = M_0 = \int \theta^2 \pi_0(\theta) d\theta \). This choice of \( M \) further constrains \( \Gamma \). This additional constraint may be unduly restrictive when \( \epsilon \) is very small; however, we may view this as an automatic procedure to assess an upper bound on the second moment.

The table summarizes the efficiency of LGM rules in the Kolmogorov neighborhood case, with this modification of \( M \).

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>( n = 5 )</th>
<th>( n = 10 )</th>
<th>( n = 50 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.99915</td>
<td>0.99954</td>
<td>0.99990</td>
</tr>
<tr>
<td>0.005</td>
<td>0.99604</td>
<td>0.99785</td>
<td>0.99954</td>
</tr>
<tr>
<td>0.01</td>
<td>0.99300</td>
<td>0.99622</td>
<td>0.99919</td>
</tr>
<tr>
<td>0.05</td>
<td>0.98515</td>
<td>0.99215</td>
<td>0.99863</td>
</tr>
<tr>
<td>0.1</td>
<td>0.98591</td>
<td>0.99270</td>
<td>0.99850</td>
</tr>
<tr>
<td>0.2</td>
<td>0.99193</td>
<td>0.99591</td>
<td>0.99917</td>
</tr>
</tbody>
</table>

Note that efficiency is higher in all cases. Specifically, note that the performance does not deteriorate when \( \epsilon \to 0 \), because, if \( \delta_\epsilon \) designates the LGM rule for a given \( \epsilon \), we have \( \delta_\epsilon \to \delta_0 \), when \( \epsilon \to 0 \). We believe that this is a desirable property.

Similar results hold for the \( \epsilon \)-contamination class.
6 Almost $\Gamma$-minimax rules

The bound modification strategy was suggested by the observed deterioration in performance of LGM rules when there is little imprecision about the prior and the bound on the second moment is large. Since the bounds given above may be unduly restrictive, we suggest another type of simple estimators, based on a result by Ibragimov and Khas'minski (1974).

Consider the problem of estimating a normal mean under squared-error loss, with a prior $\pi$. The Bayes rule $\delta_\pi$ is the expected value of the posterior distribution. Ibragimov and Khas'minski suggest replacing the marginal by the prior in the variational form of the Bayes rule in this context, and propose the rule

$$\delta^*(\bar{X}) = \bar{X} + \frac{1}{n} \frac{\tau'((\bar{X})}{\pi(\bar{X}).} \tag{9}$$

Under suitable conditions, they show that

$$|r(\pi, \delta^*) - r(\pi)| = o\left(\frac{1}{n^2}\right). \tag{10}$$

Therefore,

$$r(\pi, \delta^*) = \frac{1}{n} - \frac{1}{n^2} I(\pi) + o\left(\frac{1}{n^2}\right). \tag{11}$$

Consider a class of priors such that $\inf_\delta \sup_\pi r(\pi, \delta)$ is equal or close to $\sup_\pi \inf_\delta r(\pi, \delta)$. Note that this happens, trivially, when the prior is unique; we are closer to this case when $\Gamma$ is small. This motivates the type of rules we are suggesting. Then, the rule of form (9) maximizing (11) in the class $\Gamma$, is close to the $\Gamma$-minimax rule. We call it almost $\Gamma$-minimax.

The prior maximizing (11) is exactly the prior $\pi_0$ minimizing the Fisher information in $\Gamma$. We have

$$r_\Gamma \approx \sup_\Gamma r(\pi) \approx \sup_\Gamma \left(\frac{1}{n} - \frac{1}{n^2} I(\pi)\right) = \frac{1}{n} - \frac{1}{n^2} I(\pi_0).$$

Substituting $\pi_0$ in (9), we obtain the desired rule.

Again, we are led to the computation of $\pi_0$. For example, for $\Gamma_1$, when $\epsilon$ is small ($\epsilon < 0.0303$), the almost $\Gamma_1$-minimax rule is

$$\delta^*_1(\bar{X}) = \begin{cases}  
\bar{X} - \frac{\omega}{2} \tan \frac{\omega^2 \bar{X}}{2}, & 0 \leq \bar{X} \leq \theta_0 \\
\bar{X}(1 - \frac{1}{n}), & \theta_0 \leq \bar{X} \leq \theta_1 \\
\bar{X} - \frac{\theta_1}{n}, & \bar{X} \geq \theta_1 
\end{cases} \tag{12}$$
Figure 1: Almost $\Gamma_2$-minimax rule for $n = 10$ and $\epsilon = 0.02$.

with $\omega$, $\theta_0$ and $\theta_1$ as in Section 4.1. A similar rule is obtained for $\epsilon > 0.0303$. For the class $\Gamma_2$, the almost $\Gamma_2$-minimax rule is

$$
\delta^*(\bar{X}) = \bar{X} - \frac{1}{n} \max\{-k, \min\{k, \bar{X}\}\},
$$

where $k$ is defined by equation (6). $\delta^*_2(\bar{X})$ is a shrinkage rule. The graph for $n = 10$ and $\epsilon = 0.02$ is given in Figure 1.

In certain contexts, linear rules are not sensible. For example, when the parameter space is bounded, Vidakovic and DasGupta (1992) show that, for different classes of priors, the probability that the LGM estimator is outside the parameter space may be as large as 0.5. Ibragimov and Khas'minski (1974) specify almost Bayes estimators for the bounded parameter case, which we could use to derive almost $\Gamma$-minimax estimators much as we have derived them in the general case.

Wells (1992) suggests the use of Bayes rules with respect to priors minimizing Fisher information. His rationale is that the chosen priors are "noninformative" and the corresponding rules are reasonably robust for many families of priors. We support such rules from a $\Gamma$-minimax standpoint.

7 Conclusions

LGM rules are efficient enough when one is interested in $\Gamma$-minimax estimation, there is a bound on the second moment of the class and the sample size goes to
infinity. For small sample sizes, we have identified cases in which LGM are not that efficient. However, we have provided alternative strategies: almost GM rules and LGM rules with modified second moment.

Acknowledgements The work of David Rios Insua was supported by projects from DGICYT and UPM and a MEC grant to visit ISDS at Duke University. We are grateful to discussions with Anirban DasGupta.

References


8 Appendices

8.1 Appendix A

(*-----------------------------FishInf.m-----------------------------*)
(* This Mathematica program computes the distribution minimizing Fisher information in the Kolmogorov neighborhood of size \eps of the standard normal law. The functions FishInts (\eps<0.0303) and FishIntl (\eps>0.0303) compute the minimum information *)

BeginPackage["FisherInfo1"]

FishInf::usage = "FishInf[\epsilon] gives the minimum of Fisher Information in the epsilon Kolmogorov neighborhood of the standard normal law. The notation is as in Huber, P. 'Robust Statistics', J. Wiley, 1980."

(* ------standard normal pdf and cdf ----------- *)
fi[x_]:= 0.398942280401432677939946 Exp[-0.5 x] ;
Fi[x_]:=0.5 + 0.5 Erf[0.7071067811865475 x] ;

FishInf[\eps_]:=FishInf[\eps] /; \eps <= 0.0303 ;
FishInf[\eps_]:=FishIntl[\eps] /; \eps > 0.0303 ;

(* --------------------- small \eps (<0.0303) --------------------- *)
FishInf[\eps_]:= Module[{uu,om,thetanul,thetaone},
    thetaone = t /. FindRoot[\eps==fi[t]/t -Fi[-t],{t,1}] ;
    (* gives theta_1 *)
    uu = u /. FindRoot[\eps==Fi[Sqrt[u Tan[u/2]]]-0.5 -Sqrt[u \ Tan[u/2]] fi[Sqrt[u Tan[u/2]]] (1+Sin[u]/u) /(1+Cos[u]),
{u,1,1.1}] ; (* gives u *)
    om = Sqrt[uu /Tan[ uu/2]] ; (* gives omega *)
    thetanul = Sqrt[uu Tan[uu/2]] ; (* gives theta_0 *)
}
Print["\"theta_0\"=\", thetanul] ;
Print["\"theta_1=\"lambda\"\", thetaone] ;
Print["\"omega=\", om] ;
secmom=2 NIntegrate[ theta^2 fi[thetanul] Cos[om theta/2]^2 \
/ Cos[om thetanul/2]^2 , {theta,0,thetanul}] + 2 NIntegrate[
theta^2 fi[theta],{theta,thetanul,thetaone}] + 2 NIntegrate[
theta^2 fi[thetaone] Exp[ -thetaone (theta-thetaone)], \
{theta, thetaone, Infinity}] ;
Print["E(theta)^2=", secmom] ;

fis=2 NIntegrate[ (om Tan[om theta/2])^2 fi[thetanul] \\
(Cos[om theta/2]/Cos[om thetanul/2])^2, {theta,0,thetanul}] \\
+ 2 NIntegrate[ theta^2 fi[theta],{theta,thetanul,thetaone}] \\
+ 2 NIntegrate[ fi[thetaone] Exp[-thetaone (theta-thetaone)] \\
thetaone^2,{theta,thetaone,Infinity}] ;
Print["Minimum Fisher Information",fis] ;
]

(* ----------------------------large eps (>0.0303) ------------------- *)
FishInfl[eps_]:= Module[{li, thetai},
li = {theta,om} /. FindRoot[{ eps==fi[theta]/(om Tan[om \ntheta/2]) -Fi[-theta], theta fi[theta]==(Cos[om theta/2])^2 \\
1/(1+2/(om theta Tan[om theta/2]))}, {theta,1.4,1.5}, \\
{om,1.0,9}] ;

thetai= li[[2]] Tan[li[[1]]] li[[2]]/2 ; (* gives theta_1 *)
Print["\"theta_0\"=\",li[[1]]] ] ;
Print["\"theta_1=\"lambda\"\",thetai] ;
Print["\"omega=\",li[[2]] ] ;

secmom1=2 NIntegrate[ theta^2 fi[li[[1]]] Cos[ li[[2]] \\
theta/2]^2/Cos[ li[[2]] li[[1]]/2]^2,{theta,0,li[[1]]}] \\
+2 NIntegrate[ theta^2 fi[li[[1]]] Exp[ -thetai (theta \n- li[[1]])}, {theta,li[[1]],Infinity}] ;
Print["E(theta)^2=",secmom1] ;

fil=2 NIntegrate[(li[[2]] Tan[li[[2]] theta/2])^2 \
\n\n17
\[ f(l_1[l]) (\cos(l_1[l] \theta) \theta)^2 / (\cos(l_1[l]) l_1[l]/2)^2, \{\theta, 0, l_1[l]\} + 2 \theta^2 \]
\[ \text{NIntegrate}[f[l_1[l]] \text{Exp}[-\theta l_1[l]], \{\theta, l_1[l], \infty}\}; \]
\[ \text{Print} ["Minimum Fisher Information =", f[l]] ; \]
\]
EndPackage[ ]

8.2 Appendix B

(*------------------------FishInf2.m-----------------------------*)
(* This Mathematica program computes the distribution minimizing
Fisher information in the epsilon contaminated family of the
standard normal law *)

BeginPackage["FishInf2"]
FishInf::usage = "FishInf[epsilon] gives the minimum of
the Fisher information in the class of arbitrary epsilon
contaminations of the standard normal law."
(* ---------standard normal pdf and cdf  --------- *)
\[ f[i][x_] := 0.398942280401432677939946 \text{Exp}[-0.5 \times x] ; \]
\[ F[i][x_] := 0.5 + 0.5 \text{Erf}[0.7071067811865475 \times x] ; \]
(* --------------------------------------------------------*)

FishInf[epsilon_, k0_] := Module[{ka},
  ka = k/FindRoot[2 f[i][k]/k - 2 F[i][-k] = epsilon/(1-epsilon), \{k, k0}\];
  Print["epsilon =", epsilon];
  Print["k =", ka];
  sm = 2 \text{NIntegrate}[\theta^2 (1-epsilon) \text{Exp}[-\theta^2/2]/\text{Sqrt}[2 \pi], \{\theta, 0, ka\} + 2 \text{NIntegrate}[\theta^2 (1-epsilon) \text{Exp}[ka^2/2 - ka \theta]/\text{Sqrt}[2 \pi], \{\theta, ka, \infty}\];
  Print["E(\theta^2) =", sm];
  fi = 2 \text{N}[\text{epsilon}/\text{Sqrt}[2 \pi]] (\text{NIntegrate}\[\theta^2 \text{Exp}[-\theta^2/2], \{\theta, 0, ka\}] + \text{NIntegrate}[ka^2 \text{Exp}[ka^2/2 - ka \theta], \]
{theta, ka, Infinity}];
Print["Minimum Fisher Information=", fi];
]
EndPackage[

8.3 Appendix C

(* -----------------------FishInf3.m ----------------------- *)
(* This Mathematica package computes the distribution
minimizing
Fisher information in the class
of all distributions satisfying P([-1, 1])= prob;
The constants c (C is protected), lambda, omega are
the same as in Huber [Robust Statistics, J. Wiley, p. 84] *)

BeginPackage["FishInf3"]
FishInf::usage = "FishInf[prob] gives the minimum
of Fisher information in the class of all distributions
that satisfy P([-1, 1]) = prob."

FishInf[prob_] := Module[{c, omega, om, lambda, lam},
c = Simplify[Cos[omega/2]^2 / (1 + 2/(omega Tan[omega/2]))];
lambda = omega Tan[omega/2.];
om = omega /. FindRoot[ prob == 1 - 2 c/lambda, {omega, 1, 1.2}];
lam = om Tan[om/2.];
c = Cos[om/2]^2 / (1 + 2/(om Tan[om/2]));
Print["C=", c];
Print["omega=", om];
Print["lambda=", lam];
sm = 2 NIntegrate[ theta^2 c Cos[om theta/2]^2/Cos[om/2]^2, 
{theta, 0, 1}] + 2 NIntegrate[ theta^2 c Exp[lam] Exp[-lam 
theta], {theta, 1, 10, Infinity}];
Print["E(theta^2) =", sm];
\text{fi} = \text{om}^2 / (1 + 2/(\text{om} \times \text{Tan}[\text{om}/2]));

\text{Print} ["\text{Minimum Fisher Information} = ", \text{fi}];

\text{EndPackage}[ ]

### 8.4 Appendix D: Tables

The tables describe relevant features of the distribution minimizing Fisher information for the classes \( \Gamma_1, \Gamma_2 \) and \( \Gamma_3 \).

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>( \theta_0 )</th>
<th>( \theta_1 = \lambda )</th>
<th>( \omega )</th>
<th>( I(H_0) )</th>
<th>( E\theta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>( \infty )</td>
<td>1.4142</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.001</td>
<td>0.65338</td>
<td>2.43641</td>
<td>1.36582</td>
<td>0.981359</td>
<td>1.02439</td>
</tr>
<tr>
<td>0.002</td>
<td>0.75341</td>
<td>2.23172</td>
<td>1.35069</td>
<td>0.966873</td>
<td>1.04522</td>
</tr>
<tr>
<td>0.003</td>
<td>0.81943</td>
<td>2.10814</td>
<td>1.33976</td>
<td>0.953870</td>
<td>1.06493</td>
</tr>
<tr>
<td>0.005</td>
<td>0.91180</td>
<td>1.94830</td>
<td>1.32335</td>
<td>0.930420</td>
<td>1.10259</td>
</tr>
<tr>
<td>0.01</td>
<td>1.05637</td>
<td>1.72409</td>
<td>1.29529</td>
<td>0.879991</td>
<td>1.19217</td>
</tr>
<tr>
<td>0.015</td>
<td>1.15335</td>
<td>1.58923</td>
<td>1.27509</td>
<td>0.836169</td>
<td>1.27923</td>
</tr>
<tr>
<td>0.02</td>
<td>1.22880</td>
<td>1.49206</td>
<td>1.25873</td>
<td>0.796450</td>
<td>1.36568</td>
</tr>
<tr>
<td>0.03</td>
<td>1.34617</td>
<td>1.35336</td>
<td>1.23238</td>
<td>0.725191</td>
<td>1.53972</td>
</tr>
<tr>
<td>0.03033</td>
<td>1.34958</td>
<td>1.34958</td>
<td>1.23161</td>
<td>0.722977</td>
<td>1.54553</td>
</tr>
<tr>
<td>0.04</td>
<td>1.33041</td>
<td>1.25027</td>
<td>1.20710</td>
<td>0.661596</td>
<td>1.72198</td>
</tr>
<tr>
<td>0.05</td>
<td>1.32165</td>
<td>1.16369</td>
<td>1.17877</td>
<td>0.604027</td>
<td>1.92011</td>
</tr>
<tr>
<td>0.08</td>
<td>1.33093</td>
<td>0.95880</td>
<td>1.08652</td>
<td>0.459832</td>
<td>2.64019</td>
</tr>
<tr>
<td>0.1</td>
<td>1.35282</td>
<td>0.84960</td>
<td>1.02395</td>
<td>0.382639</td>
<td>3.25926</td>
</tr>
<tr>
<td>0.15</td>
<td>1.43374</td>
<td>0.63221</td>
<td>0.87384</td>
<td>0.238115</td>
<td>5.57266</td>
</tr>
<tr>
<td>0.2</td>
<td>1.53630</td>
<td>0.46742</td>
<td>0.73633</td>
<td>0.143242</td>
<td>9.82812</td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>( k )</td>
<td>( I(\Pi_0) )</td>
<td>( E\theta^2 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>2.63288</td>
<td>0.990542</td>
<td>1.01219</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.005</td>
<td>2.16253</td>
<td>0.964576</td>
<td>1.05057</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>1.94511</td>
<td>0.938756</td>
<td>1.09362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>1.71744</td>
<td>0.895819</td>
<td>1.17482</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>1.39838</td>
<td>0.796100</td>
<td>1.45124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>1.14017</td>
<td>0.671206</td>
<td>1.83463</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.86159</td>
<td>0.488870</td>
<td>2.88820</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.68448</td>
<td>0.354428</td>
<td>4.40144</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>( C )</th>
<th>( \omega )</th>
<th>( \lambda )</th>
<th>( I(\Pi_0) )</th>
<th>( E\theta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.04825</td>
<td>0.45899</td>
<td>0.10722</td>
<td>0.010750</td>
<td>174.283</td>
</tr>
<tr>
<td>0.2</td>
<td>0.09263</td>
<td>0.66769</td>
<td>0.23157</td>
<td>0.046261</td>
<td>37.6110</td>
</tr>
<tr>
<td>0.25</td>
<td>0.11316</td>
<td>0.75785</td>
<td>0.30175</td>
<td>0.075294</td>
<td>22.2747</td>
</tr>
<tr>
<td>0.3</td>
<td>0.13244</td>
<td>0.84345</td>
<td>0.37841</td>
<td>0.113188</td>
<td>14.2718</td>
</tr>
<tr>
<td>0.4</td>
<td>0.16678</td>
<td>1.00800</td>
<td>0.55592</td>
<td>0.220998</td>
<td>6.76568</td>
</tr>
<tr>
<td>0.5</td>
<td>0.19434</td>
<td>1.17160</td>
<td>0.77735</td>
<td>0.384190</td>
<td>3.59252</td>
</tr>
<tr>
<td>0.6</td>
<td>0.21322</td>
<td>1.34237</td>
<td>1.06610</td>
<td>0.626555</td>
<td>2.02984</td>
</tr>
<tr>
<td>0.7</td>
<td>0.22035</td>
<td>1.53008</td>
<td>1.46902</td>
<td>0.991399</td>
<td>1.18265</td>
</tr>
<tr>
<td>0.8</td>
<td>0.21008</td>
<td>1.75161</td>
<td>2.10085</td>
<td>1.571800</td>
<td>0.69190</td>
</tr>
<tr>
<td>0.9</td>
<td>0.16903</td>
<td>2.05097</td>
<td>3.38068</td>
<td>2.642930</td>
<td>0.39062</td>
</tr>
</tbody>
</table>