The Minimax Risk for Estimating a Bounded Normal Mean

by

Israel Feldman
Hebrew University

and

Lawrence D. Brown*
Hebrew University and Cornell University

Short Title: Minimax Risk Bounds

Key Words and Phrases: Minimax risk, Quadratic risk, Normal distribution, Restricted parameter space.

* Research supported in part by NSF DMS 8809015
The Minimax Risk for Estimating a Bounded Normal Mean

Introduction

Donoho and Liu (1988) produced a lower bound for the asymptotic minimax risk, applicable to a variety of non-parametric regression and density estimation problems. Their lower bound involves a constant, first mentioned in Ibragimov and Hasminskii (1987), which depends on the minimax values referred to in the title. For further explanation of their results see also Brown and Low (1988).

They conjectured that the value of this constant was 1.2465 (based on the results of Casella and Strawderman (1988)), but were only able to prove that the constant was less than 1.34. The numerical results to be given below prove that their constant is less than 1.250 and so make it certain that the number is very close to 1.2465.

This paper contains two main results. Section 3 reports a table of values of the minimax risk for estimating a bounded normal mean, for various values of the bound. This table extends the results given by Casella and Strawderman (1981). It also establishes that the Ibragimov-Hasminskii, Donoho-Liu constant is not greater than 1.294. Section 4 describes a numerical study designed to provide a more precise bound on this constant; and which results in the bound 1.250 mentioned above.

Definitions and Notations

Let $X/\theta \sim N(\theta, 1)$. For an observation $x$, let $\delta(x)$ be an estimate of $\theta$. 

The risk is defined by:

\[ R(\theta, \delta) = E((\theta - \delta(x))^2 / \theta) = \int (\theta - \delta(x))^2 \rho(x - \theta) dx \]

where \( \rho(t) = \frac{1}{(2\pi)^{1/2}} e^{-t^2/2} \)

Let \( g \) be the prior distribution function of \( \theta \); then

\[ r(g, \delta) = E_g R(\theta, \delta) \text{ is the Bayes risk.} \]

Consider the case that the mean is restricted in a given interval: \( \delta \in [-m, m] \). Let \( r_m \) denote the minimax risk for this case. It is well known that in this case the least favorable distribution exists and puts mass on a finite number of points (Ghosh (1964)).

Any prior distribution which puts mass on a finite number of points can be specified with two finite vectors \( (z_1, \ldots, z_n) \) and \( (y_1, \ldots, y_n) \) where \( z_i \) are the \( n \) points in the interval \([-m, m]\) and \( y_i \) is the mass that the least favorable distribution puts on \( z_i \).

**Lemma 1**

Let \( g \) be the distribution that puts masses \( (y_1, \ldots, y_n) \) on \( (z_1, \ldots, z_n) \) respectively, then the Bayes risk is given by:

\[
\gamma(z, y) = \sum_{i=1}^{n} z_i y_i - \int_{-\infty}^{\infty} \frac{\prod_{i=1}^{n} y_i \phi(u - z_i)}{\prod_{i=1}^{n} y_i \phi(u - z_i)} du
\]

**Proof:** For any prior distribution \( g \) the Bayes risk, \( B(g) \), satisfies

\[ B(g) = 1 - \int_{-\infty}^{\infty} \frac{(h'(u))^2}{h(u)} du \] where \( h(u) \) is the marginal density function.

See Brown (1971) or Bickel (1981). In our case
h(u) = \sum_{i=1}^{n} y_i \phi(u - z_i)

Since

h'(u) = \sum_{i=1}^{n} (z_i - u) y_i \phi(u - z_i) = \sum_{i=1}^{n} z_i y_i \phi(u - x_i) - uh(u)

we get that

\left( \frac{h'(u)}{h(u)} \right)^2 = \frac{\left[ \sum_{i=1}^{n} z_i y_i \phi(u - z_i) \right]^2}{\sum_{i=1}^{n} z_i y_i \phi(u - z_i)} - 2u \sum_{i=1}^{n} z_i y_i \phi(u - z_i) + u^2 \sum_{i=1}^{n} y_i \phi(u - z_i)

And finally by using the equality for B(g) the Lemma is proven.

Let \( r_m \) be the minimax risk for all distribution functions that are restricted to the interval \([-m, m]\). For every \( m, n \) we can find an approximation to the minimax risk by solving a numerical problem as follows:

Among all \((z_1, \ldots, z_n)\) and \((y_1, \ldots, y_n)\) such that:

\[-m \leq z_i \leq m, \quad i = 1, \ldots, n\]
\[y_i \geq 0, \quad i = 1, \ldots, n\]
\[\sum_{i=1}^{n} y_i = 1\]

find \((z_1, \ldots, z_n)\) and \((y_1, \ldots, y_n)\) such that:

1. \( \gamma(z^*, y^*) = \max_{z, y} \gamma(z, y) \).

Let \( \gamma_m = \gamma(z^*, y^*) \), as numerically determined above.
\( \gamma_m \) is an excellent estimate of \( r_m \) since, as already noted,
\[
\gamma_m = \sup \{ B(g) : g([-m,m]) = 1 \} = \gamma(z^*, y^*)
\]
for some choice of \( n \). Furthermore in any case \( \gamma_m \leq r_m \) since \( r_m \) is the minimax risk and \( \gamma_m \) is the Bayes risk for some prior.

The Ibragimov-Hasminskii, Donoho-Liu constant is defined by
\[
\lambda = \sup \frac{m^2}{(m^2 + 1)r_m}
\]
The Table 1 shows the values of \( \gamma_m \) and of
\[
\lambda_m^* = \frac{m^2}{(m^2 + 1)\gamma_m}
\]
for selected values of \( m \). It also shows the value of \( n \) which produced the corresponding \( \gamma_m \). Clearly, \( \max \lambda_m^* \) is an estimate of \( \lambda \). Note also that
\[
\lambda_m = m^2/[(m^2 + 1)r_m] \leq \lambda_m^*.
\]

The focus of the above discussion has been on the constant \( \lambda \), but it should be noted that the entries \( \gamma_m \) of Table 1 can be of independent interest. It is also true that they can be directly exploited in asymptotic theory as part of Donoho and Liu's hardest-linear family method. See Donoho and Liu (1988) and also Section 3 of Brown and Low (1988).
Table 1: The Values of \( n, \gamma_m \) and \( \lambda_m^* \) for Selected Values of \( m \)

<table>
<thead>
<tr>
<th>( m )</th>
<th>1</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8</th>
<th>1.9</th>
<th>2</th>
<th>2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>( \gamma_m )</td>
<td>.45</td>
<td>.472</td>
<td>.492</td>
<td>.513</td>
<td>.535</td>
<td>.556</td>
<td>.577</td>
<td>.596</td>
<td>.615</td>
<td>.631</td>
<td>.645</td>
<td>.669</td>
</tr>
<tr>
<td>( \lambda_m^* )</td>
<td>1.111</td>
<td>1.160</td>
<td>1.200</td>
<td>1.225</td>
<td>1.238</td>
<td>1.245</td>
<td>1.2465</td>
<td>1.243</td>
<td>1.243</td>
<td>1.241</td>
<td>1.240</td>
<td>1.239</td>
</tr>
<tr>
<td>( m )</td>
<td>2.4</td>
<td>2.7</td>
<td>3.0</td>
<td>3.3</td>
<td>3.5</td>
<td>4.0</td>
<td>4.2</td>
<td>4.4</td>
<td>4.5</td>
<td>4.6</td>
<td>4.8</td>
<td>6.0</td>
</tr>
<tr>
<td>( n )</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>( \gamma_m )</td>
<td>.693</td>
<td>.725</td>
<td>.751</td>
<td>.773</td>
<td>.787</td>
<td>.815</td>
<td>.824</td>
<td>.834</td>
<td>.838</td>
<td>.842</td>
<td>.850</td>
<td>.886</td>
</tr>
<tr>
<td>( \lambda_m^* )</td>
<td>1.23</td>
<td>1.213</td>
<td>1.198</td>
<td>1.185</td>
<td>1.175</td>
<td>1.155</td>
<td>1.148</td>
<td>1.140</td>
<td>1.137</td>
<td>1.134</td>
<td>1.128</td>
<td>1.098</td>
</tr>
</tbody>
</table>

It can be deduced directly from these numbers and the monotonicity of \( r_m \) that \( \lambda \leq 1.294 \). However, the following theorem establishes a better bound for \( \lambda \).

**Theorem:** \( \lambda_m \leq 1.25 \) for \( 0 \leq m < \infty \).

**Proof:** The main idea of the proof is based on the following Lemma.

**Lemma 2:** Let \( B_{m_0} \) be the Bayes risk for some distribution restricted to the interval \([-m_0, m_0]\). If \( \langle m_0 / [(m_0 + 1)B_{m_0}] \rangle < 1.25 \) and \( B_{m_0} < 0.8 \) then there exists an \( m_1 \), defined by

\[
(1.1) \quad m_1 = [1.25B_{m_0} / (1 - 1.25B_{m_0})]^{\frac{1}{2}} > m_0 ,
\]

such that

\[
\lambda_m \leq 1.25 \quad \text{for every} \quad m_0 \leq m \leq m_1 .
\]
Proof: Since $r_{m_0}$ is the minimax risk on $[-m_0, m_0]$, $r_{m_0} \geq B_{m_0}$ for every Bayes risk with respect to a distribution restricted to $[-m_0, m_0]$. Thus

$$\lambda_{m_0} = \frac{2}{[(m_0^2 + 1)]} r_{m_0} \leq \frac{2}{[(m_0^2 + 1)B_{m_0}]} < 1.25$$

Let $m_1$ be the solution of the following equation:

$$\frac{2}{m_1/[(m_1^2 + 1)B_{m_0}]} = 1.25$$

The solution exists for every $B_{m_0} < 0.8$ and is given by (1.1).

Since $r_m$ is non-decreasing in $m$, $\lambda_m \leq \frac{m^2}{[(m^2 + 1)r_m]} \leq \frac{2}{[(m_1^2 + 1)B_{m_0}]}$ for every $m \geq m_0$. By using the fact that $m^2/(m^2 + 1)$ is increasing with $m$ we finally get:

$$\lambda_m \leq \frac{m^2}{[(m^2 + 1)B_{m_0}]} \leq \frac{2}{[m_1^2/(m_1^2 + 1)/B_{m_0}]} = 1.25$$

for every $m$; $m_0 \leq m \leq m_1$.

Remark: For $B_m \geq 0.8$

$$\lambda_m \leq \frac{m^2}{[(m^2 + 1)r_m]} \leq \frac{m^2}{[(m_2^2 + 1)B_m]} \leq 1.25$$

Thus for $m \geq 4$, $\lambda_m < 1.25$ (see Table 1).

We shall use the idea of Lemma 2 for all $1.05 \leq m < 4$.

For $0 \leq m < 1.05$ we use the following argument:

Let $r_m$ denote the two point prior putting equal mass on $\pm m$, and let $\delta^0_m(x)$ denote the Bayes rule against $r_m$. It is straightforward to check that

$$\delta^0_m(x) = m \tanh(mx)$$

and that the Bayes risk is given by
\[ r_m = m^2 E_m [1 - \tanh(mx)]^2. \]

From Casella and Strawderman [2] it follows that this distribution is the least favorable for \(0 \leq m \leq 1.05\).

Let \(g(m) = E_m [1 - \tanh(mx)]^2\). From the proof of Lemma 3.2 in [2] it follows that \(dg(m)/dm < 0\), thus \(g(m)\) is decreasing in \(m\). Since

\[ \lambda_m \leq m^2 / [(m^2 + 1) r_m] = 1 / [(m^2 + 1) g(m)] \]

it follows that if \(m_1 < m_2\) and

\[ 1 / [(m_1^2 + 1) g(m_2)] \leq 1.25 \]

then

\[ \lambda_m \leq 1 / [(m_1^2 + 1) g(m_2)] \quad \text{for } m_1 \leq m \leq m_2 \]

since \(1 / (m_1^2 + 1) \leq 1 / (m^2 + 1)\).
The following table represents the values of the bounds $1/[(m_1^2 + 1)g(m_2)]$ for different intervals $[m_1, m_2]$ that are covering the interval $[0, 1.05]$.

<table>
<thead>
<tr>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$[1/(m_1^2+1)/g(m_2)]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4</td>
<td>1.1594</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
<td>1.1845</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7</td>
<td>1.1224</td>
</tr>
<tr>
<td>0.7</td>
<td>0.8</td>
<td>1.1485</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>1.1844</td>
</tr>
<tr>
<td>0.9</td>
<td>1.0</td>
<td>1.2277</td>
</tr>
<tr>
<td>1.0</td>
<td>1.05</td>
<td>1.1958</td>
</tr>
</tbody>
</table>

Thus $\lambda_m \leq 1.25$ for $0 \leq m \leq 1.05$.

For further computations we shall use Lemma 2. For example:

$m_0 = 1.05$, $B_{m_0} = .461$

\[
\lambda_{m_0} \leq \frac{m_0^2}{(m_0^2 + 1)B_{m_0}} = 1.1275 < 1.25
\]

\[
m_1 = \left[1.25 \cdot .461/(1 - 1.25 \cdot .461)\right]^{1/2} = 1.166
\]

Thus $\lambda_m \leq 1.25$ for $1.05 \leq m \leq 1.166$. After this stage we then choose a prior distribution on $[-m,m]$ and compute a lower bound for its Bayes risk.

Let $B_m$ denote this lower bound. If $\lambda_{m_1} = m_1^2/[(1 + m_1^2)/B_{m_1}] \leq 1.25$ we can proceed via the same procedure by taking $m_1$ as $m_0$. For $1.1 \leq m \leq 2.2$ we use the three point prior putting mass $\alpha$ at 0 and $\frac{1}{2}(1 - \alpha)$ at $\pm m$.

For $\alpha$ we use the values from Table 2 of Casella and Strawderman [2] and a linear interpolation when $1.1 \leq m \leq 2$. When $2 \leq m \leq 2.2$ we use $\alpha = 0.42$. 

8
When $2.2 \leq m \leq 2.7$ we use the four point prior with equal masses on $z_1, \ldots, z_4$: $z_1 = -m$, $z_2 = -m + D$, $z_3 = -z_2$, $z_4 = -z_1$ and $D$ was chosen such that for $m = 2.25$, $z_2 = -0.533$ as was suggested by Casella and Strawderman. When $2.7 < m \leq 4.0$ we use the five point prior with equal masses and equal distances between $z_1, \ldots, z_5$.

The following table represents the first few values of the sequence $(m_i, B'_m, \lambda'_m)$.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_i$</td>
<td>1.1</td>
<td>1.1996</td>
<td>1.2637</td>
<td>1.3103</td>
</tr>
<tr>
<td>$B'_m$</td>
<td>0.472</td>
<td>0.4919</td>
<td>0.5055</td>
<td>0.5155</td>
</tr>
<tr>
<td>$\lambda'_m$</td>
<td>1.160</td>
<td>1.1993</td>
<td>1.2164</td>
<td>1.2258</td>
</tr>
</tbody>
</table>

Table 3: Values of $(m_i, B'_m, \lambda'_m)$

The complete set of computed values is available from the authors.

Our procedure covered all possible values of $m$ with $\lambda'_m \leq 1.25$ and thus $\lambda'_m \leq 1.25$ for every $m$, $0 \leq m \leq \infty$.  

9
References


