Minimax Estimation of a Bounded Squared Mean

Jianqing Fan
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260

Irène Gijbels [1]
Department of Mathematics
Limburgs Universitair Centrum
B-3590 Diepenbeek, Belgium

Abstract

Consider a normal model with unknown mean bounded by a known constant. This paper deals with minimax estimation of the squared mean. We establish an expression for the asymptotic minimax risk. This result is applied in nonparametric estimation of quadratic functionals.

1 Introduction

Consider a random sample $X_1, \cdots, X_n$ from $N(\theta, \sigma^2)$. In this paper, we deal with estimation of θ^2 having the prior knowledge that $|\theta| \leq \tau$, where τ is known. In other words, the parameter space of θ is restricted to a bounded interval. By sufficiency considerations, the problem is equivalent to estimating θ^2 based on the random variable $\bar{X} \sim N(\theta, \sigma^2/n)$. Then, the minimax risk of estimating θ^2 is

$$\rho(\tau, \sigma^2/n) = \inf_T \sup_{|\theta| \leq \tau} E_\theta \left(T(\bar{X}) - \theta^2\right)^2, \tag{1.1}$$

where $T(\bar{X})$ denotes a statistic. By rescaling it is easy to see that

$$\rho(\tau, \sigma^2/n) = \frac{\sigma^4}{n^2} \rho(\sqrt{n}\tau/\sigma, 1). \tag{1.2}$$

[1]Completed while visiting Department of Statistics, University of North Carolina, Chapel Hill.

AMS 1980 subject classification. Primary 62F10. Secondary 62G05.

Key words and phrases. Bayes risk; Gaussian white noise model; hardest 1-dimensional subproblems; minimax risk; normal mean; quadratic functionals.

The asymptotic property of the latter risk is studied via the behavior of $\rho(a,1)$ as $a \to \infty$. In Section 2, we evaluate this behavior and establish that

$$\rho(a,1) = 4a^2(1 + o(1)).$$

This result can be used in nonparametric estimation of quadratic functionals by considering the hardest 1-dimensional subproblem. For discussions on that subproblem see Stein (1956) and Donoho and Liu (1991). This approach allows us to establish a lower bound for the minimax risk in this estimation problem. A discussion on this application forms the core of Section 3.

A related topic is the estimation of a bounded normal mean, i.e. estimation of θ knowing that $|\theta| \leq \tau$. Here, the appropriate minimax risk is

$$\inf_T \sup_{|\theta|\leq\tau} E_\theta \left(T(\bar{X}) - \theta\right)^2 .$$

For detailed studies on this estimation problem see Levit (1980), Casella and Strawderman (1981), Bickel (1981), Ibragimov and Has'minskii (1984), Gatsonis, MacGibbon and Strawderman (1987), and Donoho, Liu and MacGibbon (1990). An important application of this problem consists of obtaining nearly sharp minimax lower bounds for estimating linear functionals in an infinite-dimensional space. These estimation problems include, for example, nonparametric density estimation and signal recovery. See respectively Donoho and Liu (1991) and Ibragimov and Has'minskii (1984).

2 Main result

The main result of this paper is contained in the following theorem. It establishes the asymptotic behavior of the minimax risk $\rho(a,1)$.

Theorem 1. *As $a \to \infty$,*

$$\rho(a,1) = 4a^2(1 + o(1)).$$

Moreover, for a large enough, the least favorable prior density is the limit (as $m \to \infty$) of the density

$$\pi_{m,a}(\theta) = \frac{(2m+1)\theta^{2m}}{2a^{2m+1}} 1_{(|\theta| \le a)}, \tag{2.1}$$

i.e.

$$\lim_{m\to\infty} \lim_{a\to\infty} B_m(a)/\rho(a,1) = 1,$$

where $B_m(a)$ is the Bayes risk of estimating θ^2 with prior density (2.1).

Proof. Let $Y \sim N(\theta, 1)$ with $|\theta| \le a$. Assign the prior density $\pi_{m,a}(\theta)$, defined in (2.1), to θ. The Bayes estimate for θ^2 under quadratic loss is given by

$$E(\theta^2|Y) = \frac{\int_{-a}^{+a} \theta^{2m+2} \exp(-(Y-\theta)^2/2) d\theta}{\int_{-a}^{+a} \theta^{2m} \exp(-(Y-\theta)^2/2) d\theta} \equiv Y^2 + \delta_a(Y).$$

Write $Y = \theta + W$, where $W \sim N(0,1)$. Then, the risk of the Bayes estimate is

$$E(Y^2 + \delta_a(Y) - \theta^2)^2 = 4\theta^2 + 3 + E\delta_a^2(W+\theta) + 4\theta E(W\delta_a(W+\theta)) + 2E(W^2\delta_a(W+\theta)).$$

Taking expectation with respect to θ results in the Bayes risk

$$\frac{4(2m+1)}{2m+3} a^2 + 3 + E\delta_a^2(W+\vartheta_a) + 4E(W\vartheta_a\delta_a(W+\vartheta_a)) + 2E(W^2\delta_a(W+\vartheta_a)), \tag{2.2}$$

where $\vartheta_a \sim \pi_{m,a}(\cdot)$, and is independent of W.

We will show that

$$E\delta_a^2(W+\vartheta_a) = o(a^2), \tag{2.3}$$

and from this, we then proceed as follows. By the Cauchy-Schwarz inequality, we obtain

$$|E(W\vartheta_a\delta_a(W+\vartheta_a))| \le \sqrt{EW^2 E\vartheta_a^2}\sqrt{E\delta_a^2(W+\vartheta_a)} = o(a^2),$$

and

$$\left|E\left(W^2\delta_a(W+\vartheta_a)\right)\right| \le \sqrt{EW^4}\sqrt{E\delta_a^2(W+\vartheta_a)} = o(a)$$

Hence, (2.2) yields

$$\frac{4(2m+1)}{2m+3} a^2(1+o(1)),$$

and this tends to $4a^2(1+o(1))$ as $m \to \infty$. Now, since the minimax risk is greater than or equal to the Bayes risk, i.e.

$$\rho(a,1) \geq 4a^2(1+o(1)),$$

and on the other hand

$$\sup_{|\theta|\leq a} E_\theta(Y^2-\theta^2)^2 = 4a^2+3,$$

we conclude that

$$\rho(a,1) = 4a^2(1+o(1)).$$

Hence it remains to prove (2.3).

First of all, by integration by parts, we write

$$\begin{aligned} \delta_a(Y) &= \frac{\int_{-a}^{+a} \theta^{2m}(\theta^2-Y^2)\exp(-(Y-\theta)^2/2)d\theta}{\int_{-a}^{+a} \theta^{2m}\exp(-(Y-\theta)^2/2)d\theta} \\ &= -\delta_1(Y)+\delta_2(Y)+(2m+1)+2m\delta_3(Y), \end{aligned}$$

where

$$\delta_1(Y) = a^{2m}(a+Y)\exp(-(Y-a)^2/2)/I,$$

$$\delta_2(Y) = a^{2m}(-a+Y)\exp(-(Y+a)^2/2)/I,$$

$$\delta_3(Y) = Y\int_{-a}^{+a} \theta^{2m-1}\exp(-(Y-\theta)^2/2)d\theta/I, \tag{2.4}$$

with

$$I = \int_{-a}^{+a} \theta^{2m}\exp(-(Y-\theta)^2/2)d\theta. \tag{2.5}$$

Hence, in order to prove (2.3) we need to show that

$$E\delta_j^2(Y_a) = o(a^2), \qquad j = 1,2,3,$$

where $Y_a \stackrel{d}{=} \vartheta_a + W$, and this will be established in the following three lemmas. □

Lemma 1. *Under the above notations,*

$$E\delta_1^2(Y_a) = o(a^2).$$

Proof. We start from the decomposition

$$E\delta_1^2(Y_a) = E\left(\delta_1^2(Y_a)1_{(Y_a<a-1)}\right) + E\left(\delta_1^2(Y_a)1_{(a-1\le Y_a\le a)}\right) + E\left(\delta_1^2(Y_a)1_{(Y_a>a)}\right), \qquad (2.6)$$

and deal with each of the three terms separately. First of all, on the set $\{Y_a < a-1\}$, and for a large enough, we have

$$I \ge (a-1)^{2m}\int_{a-1}^{a} \exp(-(Y_a-\theta)^2/2)d\theta \ge \frac{1}{2}a^{2m}\exp(-(Y_a-a)^2/2),$$

and hence

$$\delta_1^2(Y_a) \le 4(a+Y_a)^2. \qquad (2.7)$$

Further, since EY_a^4/a^4 is bounded, the sequence of random variables $\{Y_a^2/a^2\}$ is uniformly integrable which implies the uniform integrability of $\{\delta_1^2(Y_a)1_{(Y_a<a-1)}/a^2\}$. Thus, showing that

$$\delta_1^2(Y_a)1_{(Y_a<a-1)}/a^2 \xrightarrow{P} 0, \quad \text{as } a\to\infty$$

leads to

$$E\left(\delta_1^2(Y_a)1_{(Y_a<a-1)}\right) = o(a^2). \qquad (2.8)$$

In order to prove the above probability statement, first note that on the set $\{2 \le |Y_a| < a-1\}$

$$I \ge \int_{Y_a}^{Y_a+1} \exp(-(Y_a-\theta)^2/2)d\theta \ge e^{-0.5}, \qquad (2.9)$$

and therefore

$$\begin{aligned}\frac{\delta_1^2(Y_a)}{a^2}1_{(Y_a<a-1)} &= \frac{\delta_1^2(Y_a)}{a^2}1_{(2\le|Y_a|<a-1)} + \frac{\delta_1^2(Y_a)}{a^2}\left(1_{(|Y_a|<2)} + 1_{(Y_a\le -a+1)}\right)\\ &\le ea^{4m}(1+Y_a/a)^2\exp(-(Y_a-a)^2/2)\\ &\quad +4(1+Y_a/a)^2\left(1_{(|Y_a|<2)} + 1_{(Y_a\le -a+1)}\right),\end{aligned}$$

where also (2.7) was used. By the inequality $(c-d)^2 \le 2(c^2+d^2)$ we have that $(Y_a-a)^2 \ge (a-\theta_a)^2/2 - W^2$. Using this fact it can be shown that the above term tends to zero in probability (we omit details here).

Next, on the set $\{a-1 \le Y_a \le a\}$ we have, for large a,

$$I \ge e^{-0.5}(a-1)^{2m} \ge 0.5e^{-0.5}a^{2m}\exp(-(Y_a-a)^2/2),$$

and thus,

$$E\left(\delta_1^2(Y_a)1_{(a-1\le Y_a\le a)}\right) \le 16ea^2P\{a-1\le Y_a\le a\} = o(a^2). \tag{2.10}$$

Finally, on the set $\{Y_a > a\}$,

$$\begin{aligned} I &\ge (a-1)^{2m}\int_{a-1}^{a}\exp(-(Y_a-\theta)^2/2)d\theta \\ &\ge \frac{(a-1)^{2m}}{Y_a-a+1}\left(\exp(-(Y_a-a)^2/2)-\exp(-(Y_a-a+1)^2/2)\right) \\ &\ge \frac{a^{2m}}{2(Y_a-a+1)}(1-e^{-0.5})\exp(-(Y_a-a)^2/2), \end{aligned}$$

and hence,

$$E\left(\delta_1^2(Y_a)1_{(Y_a>a)}\right) \le 4(1-e^{-0.5})^{-2}E\left((Y_a-a+1)^2(a+Y_a)^2 1_{(Y_a>a)}\right).$$

In order to deal with the last expectation, note that

$$P\{Y_a > a\} \le P\{\vartheta_a > a-a^{0.25}\} + P\{W > a^{0.25}\} = O\left(a^{-0.75}\right), \tag{2.11}$$

$$P\{Y_a > a+a^{0.25}\} \le P\{W > a^{0.25}\} = O\left(a^{-0.25}\exp(-a^{0.5}/2)\right), \tag{2.12}$$

and

$$E|Y_a|^l = O\left(a^l\right), \quad \text{for any } l > 0. \tag{2.13}$$

The Cauchy-Schwarz inequality, together with (2.11) — (2.13), assures that

$$\begin{aligned} E\left(\delta_1^2(Y_a)1_{(Y_a>a)}\right) &\le 4(1-e^{-0.5})^{-2}\left[E\left((Y_a-a+1)^2(a+Y_a)^2 1_{(a^{0.25}+a\ge Y_a>a)}\right)\right. \\ &\qquad \left. +E\left((Y_a-a+1)^2\,(a+Y_a)^2 1_{(Y_a>a^{0.25}+a)}\right)\right] \\ &\le O(a^{1.75}). \end{aligned} \tag{2.14}$$

Combination of (2.6), (2.8), (2.10) and (2.14) leads to the desired result. □

Lemma 2. *Under the above notations,*

$$E\delta_2^2(Y_a) = o(a^2).$$

Proof. The proof is similar to that of Lemma 1. □

Lemma 3. *Under the above notations,*

$$E\delta_3^2(Y_a) = o(a^2).$$

Proof. From (2.4) and (2.5) we obtain, for $a \geq 2$, that

$$\delta_3^2(Y_a) \leq 2Y_a^2 \left(1 + g^2(Y_a)\right), \tag{2.15}$$

where

$$g(Y_a) = \frac{\int_{-1}^{1} \theta^{2m-1} \exp(-(Y_a - \theta)^2/2) d\theta}{\int_{-2}^{2} \theta^{2m} \exp(-(Y_a - \theta)^2/2) d\theta}.$$

When $|Y_a| \leq 2$, the continuity of g assures that $|g(Y_a)| \leq c$, a finite constant. When $Y_a > 2$, we have

$$\int_{-1}^{1} \theta^{2m-1} \exp(-(Y_a - \theta)^2/2) d\theta \leq 2 \exp(-(Y_a - 1)^2/2),$$

and

$$\int_{-2}^{2} \theta^{2m} \exp(-(Y_a - \theta)^2/2) d\theta \geq \exp(-(Y_a - 1)^2/2).$$

Combining these two inequalities we obtain that $|g(Y_a)| \leq 2$, whenever $Y_a > 2$. Similarly, when $Y_a < -2$, $|g(Y_a)| \leq 2$. Therefore (2.15) leads to

$$\delta_3^2(Y_a) \leq 2 \max(c^2 + 1, 5) Y_a^2.$$

Note that the uniform integrability of $\{Y_a^2/a^2\}$ implies that of $\{\delta_3^2(Y_a)/a^2\}$. So, it suffices to show that

$$\delta_3^2(Y_a)/a^2 \xrightarrow{P} 0 \quad \text{as } a \to \infty, \tag{2.16}$$

in order to obtain that

$$E\frac{\delta_3^2(Y_a)}{a^2} \to 0 \quad \text{as } a \to \infty.$$

We now prove statement (2.16). First of all, by (2.9),

$$P(I \geq e^{-0.5}) \geq P(2 \leq |Y_a| < a-1) \to 1,$$

and hence from (2.4) we find

$$\begin{aligned}\frac{\delta_3^2(Y_a)}{a^2} &= \frac{Y_a^2}{a^2}\left[\int_{-M}^{M} \theta^{2m-1}\exp(-(Y_a-\theta)^2/2)d\theta/I + \frac{1}{M}\right]^2 \\ &\leq O_P(1)\left[2M^{2m}\exp(-Y_a^2/4 + M^2/2)e^{0.5} + \frac{1}{M}\right]^2.\end{aligned}$$

Letting $a \to \infty$ and then $M \to \infty$, statement (2.16) follows. This finishes the proof. □

3 Application

In this section, we will use the main result to obtain a lower bound for the minimax risk in estimating quadratic functionals. More precisely, we focus on estimation of

$$Q(\mathbf{x}) = \sum_{j=1}^{\infty} \lambda_j x_j^2, \tag{3.1}$$

with λ_j, $j = 1, 2, \cdots$, known constants, based on observations Y_j's from

$$Y_j = x_j + \sigma Z_j, \qquad Z_j \sim_{\text{i.i.d}} N(0,1), \quad j = 1, 2, \cdots, \tag{3.2}$$

where $\mathbf{x} \in \Sigma$, a convex subset in $\Re^\infty$. For a motivation of and a discussion on this estimation problem, see Donoho and Nussbaum (1990) and Fan (1991).

We deal with the above nonparametric estimation problem by using the hardest 1-dimensional heuristic. Let $\mathbf{c}$ be a fixed point in Σ, and denote by $\mathbf{c}_\theta = \theta\mathbf{c}, \quad \theta \in [-1, 1]$, a line segment through the origin. Estimating $Q(\mathbf{x})$ in the parametric subfamily $\{\mathbf{c}_\theta\}$ is equivalent to estimating $Q(\mathbf{c}_\theta) = \theta^2 Q(\mathbf{c})$ from observations (c.f. (3.2))

$$\mathbf{Y} = \theta\mathbf{c} + \sigma\mathbf{Z},$$

where θ is unknown and has to be estimated. Note that the inner product $(\mathbf{c}, \mathbf{Y})$ is a sufficient statistic for θ. Hence, by the Blackwell-Rao theorem, it suffices to consider statistics based on $(\mathbf{c}, \mathbf{Y})$ or equivalently on

$$\frac{\sqrt{Q(\mathbf{c})}}{\|\mathbf{c}\|^2}(\mathbf{c}, \mathbf{Y}) \sim N(\theta\sqrt{Q(\mathbf{c})}, \frac{Q(\mathbf{c})}{\|\mathbf{c}\|^2}\sigma^2)$$

for estimating θ. From definition (1.1), the minimax risk for $\theta^2 Q(\mathbf{c})$ is given by

$$\rho\left(\sqrt{Q(\mathbf{c})}, \frac{Q(\mathbf{c})}{\|\mathbf{c}\|^2}\sigma^2\right). \tag{3.3}$$

In other words, this is the minimax risk for estimating the quadratic functional $Q(\mathbf{x})$ in the subfamily $\{\mathbf{c}_\theta\}$. Now, the problem of estimating $Q(\mathbf{x})$ on Σ is at least as difficult as that of estimating $Q(\mathbf{x})$ on the subfamily $\{\mathbf{c}_\theta\}$. Indeed, for any estimator $T(\mathbf{Y})$,

$$\begin{aligned}
\inf_T \sup_{\mathbf{x}\in\Sigma} E_{\mathbf{x}}(T(\mathbf{Y}) - Q(\mathbf{x}))^2 &= \inf_T \sup_{\mathbf{c}\in\Sigma} \sup_{|\theta|\le 1} E_{\mathbf{c}_\theta}\left(T(\mathbf{Y}) - \theta^2 Q(\mathbf{c})\right)^2 \\
&\ge \sup_{\mathbf{c}\in\Sigma} \inf_T \sup_{|\theta|\le 1} E_{\mathbf{c}_\theta}\left(T(\mathbf{Y}) - \theta^2 Q(\mathbf{c})\right)^2 \\
&= \sup_{\mathbf{c}\in\Sigma} \rho\left(\sqrt{Q(\mathbf{c})}, \frac{Q(\mathbf{c})}{\|\mathbf{c}\|^2}\sigma^2\right) \qquad \text{(by (3.3))} \\
&= \sup_{\mathbf{c}\in\Sigma}\left(\frac{Q^2(\mathbf{c})\sigma^4}{\|\mathbf{c}\|^4}\right)\rho(\|\mathbf{c}\|/\sigma, 1) \qquad \text{(by (1.2))}.
\end{aligned}$$

This implies that a lower bound for the minimax risk for nonparametric estimation of quadratic functionals (3.1) can be obtained via studying $\rho(\cdot, 1)$. Theorem 1 provides that

$$\inf_T \sup_{\mathbf{x}\in\Sigma} E_{\mathbf{x}}(T(\mathbf{Y}) - Q(\mathbf{x}))^2 \ge \sup_{\mathbf{x}\in\Sigma, \mathbf{x}\ne 0} \frac{4Q^2(\mathbf{x})}{\|\mathbf{x}\|^2}\sigma^2(1 + o(1)), \tag{3.4}$$

as $\sigma \to 0$. This kind of asymptotics becomes natural when the white noise model (3.2) is connected with a density estimation model. Here σ plays a similar role as $\sqrt{n}$ in the density estimation setting. See Donoho, Liu and MacGibbon (1990) and Fan (1991) for further discussions on this connection. We also remark that the lower bound (3.4) is attainable for certain quadratic functionals (see Fan (1991)).

References

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

[2] Casella, G. and Strawderman, W.E. (1981). Estimating a bounded normal mean. *Ann. Statist.*, **9**, 870-878.

[3] Donoho, D.L. and Liu, R.C. (1991). Geometrizing rate of convergence I. *Ann. Statist.*, **19**, to appear.

[4] Donoho, D.L., Liu, R.C. and MacGibbon, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.*, **18**, 1416-1437.

[5] Donoho, D.L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. of Complexity*, **6**, 290-323.

[6] Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.*, **19**, to appear.

[7] Gatsonis, C., MacGibbon, B. and Strawderman, W. (1987). On the estimation of a restricted normal mean. *Statist. Probab. Lett.*, **6**, 21-30.

[8] Ibragimov, I.A. and Has'minskii, R.Z. (1984). Nonparametric estimation of the value of linear functional in Gaussian white noise. *Theory Probab. Appl.*, **29**, 1-32.

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

[10] Stein, C. (1956). Efficient nonparametric estimation and testing. *Proc. Third Berkeley Symp.*, **1**, 187-195.