On Minimax Filtering over Ellipsoids

Eduard N. Belitser and Boris Y. Levit

Mathematical Institute, University of Utrecht
Budapestlaan 6, 3584 CD Utrecht, The Netherlands

The problem of estimating the mean θ of an observed Gaussian infinite-dimensional vector with independent components is studied, when vector θ is known to lie in a l_2-ellipsoid and the variances of the components need not be equal. Under some general assumptions on the ellipsoid we provide the second-order behaviour of the minimax risk.

Key words: Gaussian noise, ellipsoid, minimax linear risk, asymptotically minimax estimator, second-order asymptotics.

1 Introduction

Pinsker (1980) initiated the study of minimax estimation procedures for the filtration problem in Gaussian noise which can be described, in equivalent terms, as

$$Y_k = \theta_k + \epsilon\sigma_k\xi_k \quad k = 1, 2, \dots \; . \tag{1}$$

Here $\sigma_k \geq 0,\;\; k = 1, 2, \dots$, are given, ξ_i's are independent standart Gaussian random variables, $\epsilon > 0$ is a small parameter and $\theta = (\theta_1, \theta_2, \dots)$ is the unknown infinite-dimensional parameter of interest, $\theta \in \Theta$, where

$$\Theta = \Theta(Q) = \{\theta : \; \sum_{k=1}^{\infty} a_k^2\theta_k^2 \leq Q\}\,, \tag{2}$$

$(a_k,\;\; k = 1, 2, \dots)$ is a nonnegative sequence converging to infinity.

The observation model (1) arise as the limiting experiment in many other estimation problems. This model has been actively pursued recently, see [1], [2], [5], [6] and further references therein. These papers demonstrate amply the importance of the asymptotic minimax estimators and their practical relevance.

In [9] it was shown that the quadratic minimax risk over the elipsoids Θ coincides asymptotically with the minimax risk within the class of linear estimators. A procedure for obtaining the minimax linear estimators and evaluating their risks was described.

In this article, developing further the approach of [9], we describe the second-order behaviour of the minimax estimators and the quadratic minimax risk for the model (1)–(2). These results are illustrated by a number of examples.

The authors are grateful to G.K. Golubev for a number of comments resulting in the improvement of some results of the paper and their better presentation.

2 Minimax linear estimation

Let the model of observations be given by (1). For the sake of simplicity we assume that the sequence $(a_k,\ k=1,2,\dots)$ in (2) is positive and monotone.

In this Section we investigate the minimax linear risk which will be shown later to be asymptotically equal, under some conditions, to the minimax risk. Denote $x=(x_1,x_2,\dots)$ and introduce the class of linear estimators:

$$\hat\theta=\hat\theta(x)=(\hat\theta_1,\hat\theta_2,\dots),\quad \hat\theta_k=x_kY_k,\quad k=1,2,\dots\,. \tag{3}$$

Define the risk of a linear estimator

$$R_\epsilon(x,\theta)=\mathbf{E}_\theta\|\hat\theta(x)-\theta\|^2 \tag{4}$$

and the minimax linear risk

$$r_\epsilon^l=r_\epsilon^l(\Theta)=\inf_x\sup_{\theta\in\Theta}R_\epsilon(x,\theta), \tag{5}$$

where $\|\theta\|^2=\sum_{k=1}^\infty\theta_k^2$.

To formulate the result about the minimax linear risk, we introduce some notations. Let c_ϵ be a solution of the equation

$$\epsilon^2\sum_{k=1}^\infty\sigma_k^2a_k(1-ca_k)_+=cQ \tag{6}$$

and

$$d_\epsilon=d_\epsilon(\Theta)=\epsilon^2\sum_{k=1}^\infty\sigma_k^2(1-c_\epsilon a_k)_+\,. \tag{7}$$

Here b_+ denotes nonnegative part of b.

The following Theorem is due to Pinsker [9], but we give its elementary proof for the sake of completeness.

Theorem 1. *Let c_ϵ and d_ϵ are defined by (6) and (7). Then*

$$\inf_x\sup_{\theta\in\Theta}R_\epsilon(x,\theta)=\sup_{\theta\in\Theta}\inf_xR_\epsilon(x,\theta); \tag{8}$$

the saddle point $(\tilde x,\tilde\theta)$ for the problem (5) is given by

$$\tilde x_k=(1-c_\epsilon a_k)_+\,, \tag{9}$$

$$\tilde\theta_k^2=\epsilon^2\sigma_k^2(1-c_\epsilon a_k)_+/(c_\epsilon a_k) \tag{10}$$

and the linear minimax risk satisfies the following equations:

$$r_\epsilon^l = d_\epsilon = \sup_{\theta \in \Theta} \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 \theta_k^2 / (\theta_k^2 + \epsilon^2 \sigma_k^2) \,. \tag{11}$$

Proof. It follows immediately that the risk of a linear estimator has the form

$$R_\epsilon(x, \theta) = \sum_{k=1}^{\infty} \left(\epsilon^2 \sigma_k^2 x_k^2 + (1 - x_k)^2 \theta_k^2 \right) . \tag{12}$$

Since, according to (6) $Qc_\epsilon^2 = \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 c_\epsilon a_k (1 - c_\epsilon a_k)_+$,

$$\begin{aligned}
\inf_x \sup_{\theta \in \Theta} R_\epsilon(x, \theta) &\leq \sup_{\theta \in \Theta} R_\epsilon(\tilde{x}, \theta) \leq Q \sup_{k \geq 1} (1 - \tilde{x}_k)^2 / a_k^2 + \sum_{k=1}^{\infty} \epsilon^2 \sigma_k^2 \tilde{x}_k^2 \\
&\leq Qc_\epsilon^2 + \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 (1 - c_\epsilon a_k)_+^2 \\
&= \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 ((c_\epsilon a_k (1 - c_\epsilon a_k)_+ + (1 - c_\epsilon a_k)_+^2)) \\
&= \epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 (1 - c_\epsilon a_k)_+ = d_\epsilon \,. \qquad (13)
\end{aligned}$$

Note now that equation (6) can be also rewritten as

$$\sum_{k=1}^{\infty} a_k^2 \tilde{\theta}_k^2 = Q \,, \tag{14}$$

so that $\tilde{\theta} \in \Theta$. Taking into account (13), (14) and (12), we obtain

$$\begin{aligned}
d_\epsilon &\geq \inf_x \sup_{\theta \in \Theta} R_\epsilon(x, \theta) \geq \sup_{\theta \in \Theta} \inf_x R_\epsilon(x, \theta) \\
&= \sup_{\theta \in \Theta} \epsilon^2 \sum_{k=1}^{\infty} \frac{\sigma_k^2 \theta_k^2}{\theta_k^2 + \epsilon^2 \sigma_k^2} \geq \epsilon^2 \sum_{k=1}^{\infty} \frac{\sigma_k^2 \tilde{\theta}_k^2}{\tilde{\theta}_k^2 + \epsilon^2 \sigma_k^2} = d_\epsilon \,.
\end{aligned}$$

This completes the proof of Theorem 1. □

Remark 1. The equations (6) and (10) can be obtained by the Lagrange multiplier method for a problem of maximizing a functional $\epsilon^2 \sum_{k=1}^{\infty} \sigma_k^2 \theta_k^2 / (\theta_k^2 + \epsilon^2 \sigma_k^2)$ subject to the convex constraint (2).

Remark 2. Due to monotonocity of $(a_k,\ k = 1, 2, \ldots)$, $d_\epsilon = \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 (1 - c_\epsilon a_k)$, where

$$N = N_\epsilon(\Theta) = \max\{k\colon\ a_k \leq c_\epsilon^{-1}\} \,. \tag{15}$$

One can easily derive the explicit formulas for c_ϵ and N (cf. [3]):

$$c_\epsilon = c_\epsilon(\Theta) = \frac{\sum_{k=1}^{N} \sigma_k^2 a_k}{Q\epsilon^{-2} + \sum_{k=1}^{N} \sigma_k^2 a_k^2} \,,$$

$$N = \max\left\{ l : \ \epsilon^2 \sum_{k=1}^{l} \sigma_k^2 a_k (a_l - a_k) \le Q \right\}.$$

Note that (15) entails that

$$0 \le c_\epsilon a_k \le 1, \quad k = 1, 2, \ldots, N \ . \tag{16}$$

3 Asymptotically minimax estimation

In this section we investigate the asymtotic behaviour of the minimak risk with respect to all possible estimators.

We define the minimax risk:

$$r_\epsilon = r_\epsilon(\Theta) = \inf_{\hat\theta} \sup_{\theta\in\Theta} \mathbf{E}_\theta \|\hat\theta - \theta\|^2, \tag{17}$$

where $\hat\theta$ is an arbitrary estimator based on $Y = (Y_1, Y_2, \ldots)$.

In the proof of Proposition 1 we use the van Trees inequality [10, p. 72]. Now we describe the version of this inequality which we use below. Let $dP_\theta(y)$, $y = (y_1, y_2, ...)$, denote distribution of the vector of observations $Y = (Y_1, Y_2, ...)$ in (1) and $\varphi(y_k, \theta_k)$ be the marginal (gaussian) density of Y_k. Assume that a prior distribution $d\Lambda(\theta), \theta = (\theta_1, \theta_2, ...)$ is defined according to which θ_k are independent random variables, with corresponding densities $\nu_k(x)$. Let, for all k, $\nu_k(x)$ be absolutely continuous, with finite Fisher information

$$I(\nu_k) = \int \left(\frac{\partial \log \nu_k(x)}{\partial x} \right)^2 \nu_k(x)\, dx.$$

Assume also that $\nu_k(x)$ is positive inside a bounded interval of the real line and zero outside it.

We write $\mathbf{E}$ for the expectation with respect to the joint distribution of Y and θ. Then, according to the van Trees inequality (cf. [4] and [5]), the Bayes risk $\mathbf{E}(\hat\theta_k - \theta_k)^2$ admits a lower bound:

$$\mathbf{E}(\hat\theta_k - \theta_k)^2 \ge \frac{1}{I_k + I(\nu_k)}, \tag{18}$$

where $I_k = \epsilon^{-2}\sigma_k^{-2}$ is the Fisher information about θ_k contained in the observation Y_k and $\hat\theta_k = \hat\theta_k(Y)$.

Since our setup here is slightly different fron those of [4] and [5], below we sketch a short proof of (18). Let

$$A = \hat\theta_k - \theta_k, \quad B = \frac{\partial}{\partial\theta_k} \log\left(\varphi(Y_k, \theta_k)\nu_k(\theta_k)\right).$$

Denote $Y^{(k)} = (Y_1, ..., Y_{k-1}, Y_{k+1}, ...)$, $\theta^{(k)} = (\theta_1, ..., \theta_{k-1}, \theta_{k+1}, ...)$ and let $dP_\theta^{(k)}(y^{(k)})$ and $d\Lambda^{(k)}(\theta^{(k)})$ respectively be their distributions.

Use the Cauchy-Schwarz inequality $\mathbf{E}A^2 \geq (\mathbf{E}AB)^2/\mathbf{E}B^2$. One can assume, without loss of generality, that $\mathbf{E}A^2 < \infty$. Our assumptions permit integration by parts and interchanging the order of integration in the following integral yields

$$\begin{aligned}
\mathbf{E}AB &= \int (\hat{\theta}_k - \theta_k) \frac{\partial}{\partial \theta_k} \log \left(\varphi(y_k, \theta_k) \nu_k(\theta_k) \right) \, dP_\theta(y) \, d\Lambda(\theta) \\
&= \int (\hat{\theta}_k - \theta_k) \frac{\partial}{\partial \theta_k} \left(\varphi(y_k, \theta_k) \nu_k(\theta_k) \right) \, dy_k \, d\theta_k dP_\theta^{(k)}(y^{(k)}) \, d\Lambda^{(k)}(\theta^{(k)}) \\
&= \int dP_\theta(y) \, d\Lambda(\theta) = 1.
\end{aligned}$$

It remains to note that $\mathbf{E}B^2 = I_k + I(\nu_k)$.

Next theorem describes the lower bound for the minimax risk (17). The proof of this and the following results of this Section will be given in the Appendix.

Theorem 2. *Let $(m_k,\ k = 1, 2, \ldots)$ be a sequence such that, for some $\gamma > 0$,*

$$\sum_{k=1}^{\infty} a_k^2 m_k^2 + \left(8\gamma \left(\log \epsilon^{-1} \right) \sum_{k=1}^{\infty} a_k^4 m_k^4 \right)^{1/2} \leq Q \,. \tag{19}$$

Then the following lower bound holds:

$$r_\epsilon \geq \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2}{m_k^2 + \epsilon^2 \sigma_k^2} + O(\epsilon^\gamma) \,, \qquad \epsilon \to 0 \,. \tag{20}$$

To derive a good lower bound, one should, in principle, maximize the functional appearing in (20) under the restriction (19). However, the following Theorem shows that, under rather mild condition, this problem is asymptotically equivalent to the maximization problem (11) which has already been solved by Theorem 1. This implies, in particular, the asymptotic equivalence of the minimax risk and the minimax linear risk.

Theorem 3. *Let c_ϵ and N be defined by (6) and (15). If condition*

$$\log \epsilon^{-1} \frac{\sum_{k=1}^{\infty} a_k^2 \sigma_k^4 (1 - c_\epsilon a_k)_+^2}{\left(\sum_{k=1}^{\infty} a_k \sigma_k^2 (1 - c_\epsilon a_k)_+ \right)^2} = o(1) \,, \qquad \epsilon \to 0 \,,$$

holds, then

$$r_\epsilon = \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 \left(c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k \right) (1 + o(1)) \,, \qquad \epsilon \to 0 \,.$$

The next Proposition, although looking quite general, provides exact asymptotics of the minimax risk for a more restricted class of ellipsoids. In particular, it is convenient in applications where the sequence $(a_k,\ k = 1, 2, \ldots)$ is increasing faster than k^m for any $m > 0$. Since in such cases the limiting behaviour of the minimax linear risk d_ϵ

typically does not depend on Q (cf. Example 4–5 in Section 4 below), this Proposition leads also to the exact asymptotics of the minimax risk r_ϵ. In the context of curve estimation this corresponds to estimating "very smooth" function, with rapidly decreasing Fourier coefficients (cf. [5]).

Proposition 1. *Let d_ϵ be defined by (7). Then*

$$d_\epsilon(\Theta(Q/\pi^2)) \le r_\epsilon \le d_\epsilon(\Theta(Q)) . \tag{21}$$

Note that these lower and upper bounds for the minimax risk are nonasymptotic.

Corollary 1. *Let c_ϵ and N be defined by (6) and (15). If $\sum_{k=1}^\infty \sigma_k^2 = \tau < \infty$ and*

$$c_\epsilon \sum_{k=1}^N \sigma_k^2 a_k = o\left(\sum_{k=N+1}^\infty \sigma_k^2\right), \qquad \epsilon \to 0 ,$$

then the following asymptotic expansion holds:

$$r_\epsilon = \epsilon^2\tau - \epsilon^2 \left(\sum_{k=N+1}^\infty \sigma_k^2\right)(1+o(1)) , \qquad \epsilon \to 0 .$$

Remark 3. There are two terms in asymptotic expression of the minimax risk in Theorem 3. They can be either of the same order or the second term can be of smaller order than the first. In the latter case Theorem 3 provides at least two terms of asymptotic expansion of the minimax risk (cf. Example 1 below).

Remark 4. Recall that the sequence $(a_k,\ k = 1, 2, \ldots)$ was assumed positive. The results remain valid under the weaker assumption: $a_k \ge 0,\ k = 1, 2, \ldots$(cf. [2]).

4 Examples

The results presented below illustrate the assertions in the previous Section.

Example 1. Consider model (1)–(2) with $a_k = k^\alpha,\ \alpha > 0,\ \sigma_k^2 = k^{\delta-1},\ \alpha + \delta > 0$. In this case it is easy to prove that $c_\epsilon N^\alpha \to 1$ as $\epsilon \to 0$. Using this and (6), one can calculate

$$N = \Big((2\alpha+\delta)(\alpha+\delta)Q/(\alpha\epsilon^2)\Big)^{\frac{1}{2\alpha+\delta}} (1+o(1)) ,$$

$$c_\epsilon = \Big(\alpha\epsilon^2/((2\alpha+\delta)(\alpha+\delta)Q)\Big)^{\frac{\alpha}{2\alpha+\delta}} (1+o(1)) .$$

Here we make use of the asymptotic relation

$$\sum_{m=1}^M m^\kappa = \frac{M^{\kappa+1}}{(\kappa+1)}(1+o(1)) \quad \text{as} \quad M \to \infty , \quad \kappa > -1 . \tag{22}$$

Now one can easily verify the condition of Theorem 3. By applying Theorem 3, we derive the asymptotics of the minimax risk.

- Case $\delta > 0$. The asymptotics (22) and relations for N and c_ϵ yield (cf. [9] for $\delta = 1$)

$$r_\epsilon = \epsilon^{4\alpha/(2\alpha+\delta)}\delta^{-1}(Q(2\alpha+\delta))^{\delta/(2\alpha+\delta)}(\alpha/(\alpha+\delta))^{2\alpha/(2\alpha+\delta)}(1+o(1)).$$

  In this case Theorem 3 gives only the first-order term of the minimax risk. Note also that although $(\sigma_k^2,\ k = 1, 2, \ldots)$ can be increasing to infinity, the minimax risk still converges to zero.

- Case $\delta = 0$. By using (22) and the asymptotics

$$\sum_{k=1}^{M} k^{-1} = \log M + C_e + o(1) \quad \text{as} \quad M \to \infty\,, \tag{23}$$

  one obtains

$$r_\epsilon = \alpha^{-1}\epsilon^2 \log\epsilon^{-1} + \epsilon^2(C_e + (2\alpha)^{-1}\log(2\alpha Q) - \alpha^{-1})(1+o(1)),$$

  where $C_e = 0.5772156\ldots$ is the Euler constant ([7, 6.360.2]).

- Case $\delta < 0$. Using the asymptotic relation

$$\sum_{m=M}^{\infty} m^{-\kappa} = \frac{M^{1-\kappa}}{\kappa-1}(1+o(1)) \quad \text{as} \quad M \to \infty\,, \quad \kappa > 1\,,$$

  we calculate

$$r_\epsilon = \epsilon^2\tau + \epsilon^{4\alpha/(2\alpha+\delta)}\delta^{-1}(Q(2\alpha+\delta))^{\delta/(2\alpha+\delta)}(\alpha/(\alpha+\delta))^{2\alpha/(2\alpha+\delta)}(1+o(1)).$$

Example 2: $a_k = k^\alpha,\ \alpha > 0,\ \sigma_k^2 = k^{-(1+\alpha)}$. In this case the condition of Theorem 3 is again satisfied, and

$$N = (Q\alpha\epsilon^{-2}/\log\epsilon^{-2})^{1/\alpha}(1+o(1))\,, \qquad c_\epsilon = \epsilon^2\log\epsilon^{-2}(Q\alpha)^{-1}(1+o(1)).$$

Then by Theorem 3,

$$r_\epsilon = \epsilon^2\tau - \epsilon^4(\log\epsilon^{-2})^2\alpha^{-2}Q^{-1}(1+o(1)).$$

Example 3: $a_k = k^\alpha,\ \alpha > 0,\ \sigma_k^2 = k^{-(1+\delta)},\ \delta > \alpha$. One calculates

$$N = \Big((\delta-\alpha)Q\epsilon^{-2}\Big)^{1/\alpha}(1+o(1))\,, \qquad c_\epsilon = \epsilon^2((\delta-\alpha)Q)^{-1}(1+o(1)).$$

With these asymptotic relations, one can show that

$$d_\epsilon(\Theta(Q)) = \epsilon^2\tau - \epsilon^4 Q^{-1}(\delta-\alpha)^{-2}(1+o(1)).$$

By applying Proposition 1, we can obtain only the rate of the second-order term of the minimax risk:

$$r_\epsilon = \epsilon^2\tau - \epsilon^4 Q^{-1}(\delta-\alpha)^{-2}\phi_\epsilon\,,$$

where
$$\liminf_{\epsilon\to 0}\phi_\epsilon \geq 1\,, \qquad \limsup_{\epsilon\to 0}\phi_\epsilon \leq \pi^2\,.$$

Example 4: $a_k = e^{\beta k}$, $\beta > 0$, $\sigma_k^2 = k^{\delta-1}$. From (15) one can see that

$$e^{-\beta} \leq c_\epsilon e^{N\beta} \leq 1\,. \tag{24}$$

Using (6), (24) and the asimptotics
$$\sum_{m=1}^{M} m^\kappa e^{\beta m} = \frac{M^\kappa e^{\beta(M+1)}}{(e^\beta - 1)}(1+o(1)) \quad \text{as} \quad M\to\infty$$
gives
$$N = \beta^{-1}\log\epsilon^{-1} + (2\beta)^{-1}(1-\delta)\log\log\epsilon^{-1} + O(1)\,.$$
By the last two relations and (24), we have
$$c_\epsilon\sum_{k=1}^{N}\sigma_k^2 a_k = \frac{N^{\delta-1}c_\epsilon e^{\beta(N+1)}}{e^\beta - 1}(1+o(1)) \leq \frac{N^{\delta-1}e^\beta}{e^\beta-1}(1+o(1)) = O\left((\log\epsilon^{-1})^{\delta-1}\right).$$
We apply Proposition 1 to this Example.

- Case $\delta > 1$. Since, according to [7, 0.121],
  $$\sum_{m=1}^{M} m^\kappa = \frac{M^{\kappa+1}}{(\kappa+1)} + \frac{M^\kappa}{2}(1+o(1)) \quad \text{as} \quad M\to\infty\,, \quad \kappa>0\,,$$
  we calculate
  $$\sum_{k=1}^{N}\sigma_k^2 = \frac{N^\delta}{\delta} + \frac{N^{\delta-1}}{2}(1+o(1)) = \frac{(\log\epsilon^{-1})^\delta}{\delta\beta^\delta} + \frac{(1-\delta)(\log\epsilon^{-1})^{\delta-1}\log\log\epsilon^{-1}}{2\beta^\delta}(1+o(1))\,,$$
  and obtain that
  $$r_\epsilon = \frac{\epsilon^2(\log\epsilon^{-1})^\delta}{\delta\beta^\delta} + \frac{\epsilon^2(1-\delta)(\log\epsilon^{-1})^{\delta-1}\log\log\epsilon^{-1}}{2\beta^\delta}(1+o(1))\,.$$

- Case $\delta = 1$. In this case we have that $c_\epsilon\sum_{k=1}^{N}\sigma_k^2 a_k = O(1)$,
  $$\sum_{k=1}^{N}\sigma_k^2 = N = \beta^{-1}\log\epsilon^{-1} + O(1)\,,$$
  and therefore,
  $$r_\epsilon = \beta^{-1}\epsilon^2\log\epsilon^{-1} + O(\epsilon^2)\,.$$

- Case $0 < \delta < 1$. One can show that

$$\sum_{m=1}^{M} m^{-\kappa} = \frac{2^\kappa}{2-2^\kappa}\left(\sum_{m=1}^{M}(M+m)^{-\kappa} - \sum_{m=1}^{2M}(-1)^{m+1}m^{-\kappa}\right)$$
$$= \frac{M^{1-\kappa}}{(1-\kappa)} + \zeta(\kappa) + o(1) \quad \text{as} \quad M \to \infty\,, \quad 0 < \kappa < 1\,,$$

where

$$\zeta(\kappa) = \frac{2^\kappa}{2^\kappa - 2}\sum_{m=1}^{\infty}(-1)^{m+1}m^{-\kappa}$$

is the Riemann zeta function ([7, 7.422.2]). Using this asymptotics, we obtain

$$\sum_{k=1}^{N}\sigma_k^2 = \frac{(\log\epsilon^{-1})^\delta}{\delta\beta^\delta} + \zeta(1-\delta) + o(1)\,.$$

Consequently,

$$r_\epsilon = \epsilon^2(\log\epsilon^{-1})^\delta\delta^{-1}\beta^{-\delta} + \epsilon^2\zeta(1-\delta)(1+o(1))\,.$$

- Case $\delta = 0$. Since, by (23),

$$\sum_{k=1}^{N}\sigma_k^2 = \log N + C_e + o(1)\,,$$

we get

$$r_\epsilon = \epsilon^2 \log\log\epsilon^{-1} + \epsilon^2(C_e + \log\beta^{-1})(1+o(1))\,.$$

- Case $\delta < 0$. In this case one can verify that

$$\sum_{k=N+1}^{\infty}\sigma_k^2 = -\frac{N^\delta}{\delta}(1+o(1)) = -\frac{(\log\epsilon^{-1})^\delta}{\delta\beta^\delta}(1+o(1))\,.$$

Therefore, by Corollary 1 we have

$$r_\epsilon = \epsilon^2\tau + \epsilon^2(\log\epsilon^{-1})^\delta\beta^{-\delta}\delta^{-1}(1+o(1))\,.$$

Example 5: $a_k = e^{\beta k^r}$, $\beta > 0$, $0 < r < 1$, $\sigma_k^2 = k^{\delta-1}$. With the asymptotics

$$\sum_{m=1}^{M} m^\kappa e^{\beta m^r} = C_r e^{\beta M^r} M^{\kappa+(1-r)_+}(1+o(1)) \quad \text{as} \quad M \to \infty\,,$$

where

$$C_r = \begin{cases} (r\beta)^{-1}, & 0 < r < 1 \\ e^\beta/(e^\beta - 1), & r = 1 \\ 1, & r > 1\,, \end{cases}$$

one can obtain

$$N = \left(\beta^{-1}\log\epsilon^{-1}\right)^{1/r}(1+o(1)).$$

By definition of N, we evaluate

$$\begin{aligned} c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k &= C_r N^{\delta-1+(1-r)_+} c_\epsilon e^{\beta N^r}(1+o(1)) \le C_r N^{\delta-1+(1-r)_+}(1+o(1)) \\ &= O\left(\left(\log\epsilon^{-1}\right)^{(\delta-1+(1-r)_+)/r}\right)(1+o(1)). \end{aligned}$$

Now the asymptotics of the minimax risk may be obtained in the same way as in Example 4.

- Case $\delta > 0$.

$$r_\epsilon = \epsilon^2(\log\epsilon^{-1})^{\delta/r}\beta^{-\delta/r}\delta^{-1}(1+o(1)).$$

- Case $\delta = 0$.

$$r_\epsilon = r^{-1}\epsilon^2\log\log\epsilon^{-1} + \epsilon^2(C_e + r^{-1}\log\beta^{-1})(1+o(1)).$$

- Case $\delta < 0$.

$$r_\epsilon = \epsilon^2\tau + \epsilon^2(\log\epsilon^{-1})^{\delta/r}\beta^{-\delta/r}\delta^{-1}(1+o(1)).$$

Remark 5. Note that in most cases in Examples 4 and 5 both the first- and the second-order terms of the minimax risk do not depend on the "size" Q of ellipsoid $\Theta(Q)$.

Remark 6. Let $a_k = a_k(\beta)$, $k = 1, 2, \ldots$, be as in Example 4 or Example 5. Define the correspondent hyperrectangle in l_2-space:

$$\mathcal{H}_\beta = \mathcal{H}_\beta(Q) = \{\theta : \ |\theta_k| \le \sqrt{Q}a_k^{-1}(\beta), \ k = 1, 2, \ldots\}.$$

The assertions of Examples 4 and 5 concerning the first-order behaviour (also the second-order behaviour for the cases $\delta = 0$ and $\delta < 0$) of the minimax risk remain valid with $\Theta = \Theta_\beta$ replaced by $\mathcal{H}_\beta$. This follows immediately from the following easily verified relation:

$$\text{for any} \quad Q > 0, \ \beta > 0, \ 0 < \mu < \beta \quad \text{there exists} \quad Q_1 > 0 \quad \text{such that}$$

$$\Theta_\beta(Q) \subseteq \mathcal{H}_\beta(Q) \subseteq \Theta_{\beta-\mu}(Q_1).$$

Example 6: $a_k = k^\alpha$, $\sigma_k^2 = e^{\beta k^r}$, $\alpha, \beta, r > 0$. Let us establish first an upper bound for the minimax risk $r_\epsilon(\Theta)$ (see (17)). Such a bound is provided by the minimax linear risk which, according to Theorem 1, equals d_ϵ (see (6)–(7)). Using the asymptotic expansions ($M \to \infty$):

$$\begin{aligned} \sum_{m=1}^{M} m^\alpha e^{\beta m^r} &= M^\alpha e^{\beta M^r}\left(\frac{M^{1-r}}{\beta r} + \frac{r-1-\alpha}{(\beta r)^2}M^{1-2r}(1+o(1))\right), \qquad 0 < r < 1, \\ \sum_{m=1}^{M} m^\alpha e^{\beta m} &= M^\alpha e^{\beta M}\left(\frac{e^\beta}{e^\beta - 1} - \frac{\alpha e^\beta}{(e^\beta-1)^2}M^{-1}(1+o(1))\right), \\ \sum_{m=1}^{M} m^\alpha e^{\beta m^r} &= M^\alpha\left(e^{\beta M^r} + e^{\beta(M-1)^r}(1+o(1))\right), \qquad r > 1, \end{aligned}$$

one can solve (6)–(7), thus obtaining

$$c_\epsilon = \left(\beta^{-1}\log\epsilon^{-2}\right)^{-\alpha/r}(1+o(1)),$$

$$d_\epsilon = Qc_\epsilon^2(1+o(1)) = Q\left(\beta^{-1}\log\epsilon^{-2}\right)^{-2\alpha/r}(1+o(1)).$$

The last formula exhibits a distinctive feature of this example, as compared to all previous ones. Indeed, analyzing the proof of Theorem 1 (cf. inequality (13)), one realizes that the term Qc_ϵ^2, contributing to d_ϵ, arises solely as the squared bias term of the linear minimax estimator. Thus, only the bias of the estimator contributes to its maximal risk, up to the first order.

To show that d_ϵ coincides asymptotically with the minimax risk $r_\epsilon(\Theta)$, we choose a prior distribution Λ on Θ and use the obvious inequality $r_\epsilon(\Theta) \geq \mathcal{R}_\epsilon(\Lambda)$, where $\mathcal{R}_\epsilon(\Lambda)$ denotes the Bayes risk. Let Λ be a distribution on Θ such that

$$\theta_N = \pm\rho, \qquad \text{with probabilities } 1/2$$

and

$$\theta_i = 0, \qquad i \neq N, \qquad \Lambda-\text{almost surely},$$

where $\rho = (Q/a_N)^{1/2}$ and $N = [c_\epsilon^{-\alpha}]$. Clearly $\Lambda(\Theta) = 1$, $\rho^2 = Qa_N^{-2} = Qc_\epsilon^2(1+o(1)) = d_\epsilon(1+o(1))$ and $\sigma_N^2 = e^{\beta N^r} = \epsilon^{-2}e^{O(1)}$.

Due to sufficiency considerations, the Bayes risk $\mathcal{R}_\epsilon(\Lambda)$ in estimating θ is equal to the Bayes risk in estimating θ_N, based on the observation Y_N only. Since

$$\lim_{\epsilon\to 0}\frac{\rho^2}{\mathbf{Var}\,Y_N} = 0,$$

it follows (see [8], proof of Lemma 3.2) that

$$r_\epsilon(\Theta) \geq \mathcal{R}_\epsilon(\Lambda) = \rho^2(1+o(1)) = d_\epsilon(1+o(1)).$$

Thus

$$r_\epsilon(\Theta) = Q\left(\beta^{-1}\log\epsilon^{-2}\right)^{-2\alpha/r}(1+o(1)).$$

5 Appendix: Proofs

The proof of Theorem 2 is based on the following elementary result.

Proposition 2. *Let $\eta_1,\ldots,\eta_m$ be independent Gaussian random variables with $\mathbf{E}\eta_k = 0$, $\mathbf{Var}\,\eta_k = d_k^2$. Then*

$$\mathbf{P}\left\{\sum_{k=1}^m \eta_k^2 > P\right\} \leq \exp\left\{-\frac{P-\sum_{k=1}^m d_k^2}{4\sum_{k=1}^m d_k^4}\right\}.$$

Indeed, using Markov inequality and moment-generating function of χ^2_m, one obtains for any $\lambda > 0$

$$\begin{aligned} \mathbf{P}\left\{\sum_{k=1}^m \eta_k^2 > P\right\} &\le e^{-\lambda P}\mathbf{E}e^{\lambda\sum_{k=1}^m \eta_k^2} = \exp\left\{-\lambda P + \sum_{k=1}^m \log(1-2\lambda d_k^2)^{-1/2}\right\} \\ &\le \exp\left\{-\lambda(P - \sum_{k=1}^m d_k^2) + \lambda^2 \sum_{k=1}^m d_k^4\right\}. \end{aligned}$$

It remains to set

$$\lambda = \frac{P - \sum_{k=1}^m d_k^2}{2\sum_{k=1}^m d_k^4}.$$

Proof of Theorem 2. We select a prior measure $d\mu(\theta)$ such that $\theta_k,\ k = 1, 2, \ldots$, are distributed independently and normally with zero means and variances m_k^2, $k = 1, 2, \ldots$. Let $\mathbf{E}$ denote the expectation with respect to the joint distribution of $Y_1, Y_2, \ldots$ and $\theta_1, \theta_2, \ldots$.

Since Θ is closed and convex, $r_\epsilon = \inf_{\hat\theta\in\Theta}\sup_{\theta\in\Theta}\mathbf{E}_\theta\|\hat\theta - \theta\|^2$. We bound the minimax risk from below as follows:

$$\begin{aligned} r_\epsilon &= \inf_{\hat\theta\in\Theta}\sup_{\theta\in\Theta}\mathbf{E}_\theta\|\hat\theta-\theta\|^2 \ge \inf_{\hat\theta\in\Theta}\int_\Theta \mathbf{E}_\theta\|\hat\theta-\theta\|^2 d\mu(\theta)/\mu(\Theta) \\ &\ge \inf_{\hat\theta}\mathbf{E}\|\hat\theta-\theta\|^2 - \sup_{\hat\theta\in\Theta}\int_{\Theta^C}\mathbf{E}_\theta\|\hat\theta-\theta\|^2 d\mu(\theta)/\mu(\Theta) \\ &\ge \inf_{\hat\theta}\sum_{k=1}^\infty \mathbf{E}(\hat\theta_k-\theta_k)^2 - 2Qa_1^{-2}(1+\sqrt{3})\left(\mu(\Theta^C)\right)^{1/2}/\mu(\Theta)\,, \end{aligned} \tag{25}$$

because, by Cauchy-Shwarz inequality and (19),

$$\begin{aligned} \sup_{\hat\theta\in\Theta}\int_{\Theta^C}\mathbf{E}_\theta\|\hat\theta-\theta\|^2 d\mu(\theta) &\le 2\mu(\Theta^C)\sup_{\hat\theta\in\Theta}\mathbf{E}_\theta\|\hat\theta\|^2 + 2\sum_{k=1}^\infty\int_{\Theta^C}\theta_k^2 d\mu(\theta) \\ &\le 2Q\mu(\Theta^C)a_1^{-2} + 2\sum_{k=1}^\infty(\mu(\Theta^C))^{1/2}\left(\int\theta_k^4 d\mu(\theta)\right)^{1/2} \\ &= 2Q\mu(\Theta^C)a_1^{-2} + 2\sqrt{3}\left(\mu(\Theta^C)\right)^{1/2}\sum_{k=1}^\infty m_k^2 \\ &\le 2Qa_1^{-2}(1+\sqrt{3})\left(\mu(\Theta^C)\right)^{1/2}. \end{aligned}$$

Note that Proposition 2, together with (19), entails

$$\left(\mu(\Theta^C)\right)^{1/2} \le \epsilon^\gamma\,. \tag{26}$$

Now we recall the following known result. If ξ and ζ are independent Gaussian random variables with $\mathbf{E}\xi = 0$, $\mathbf{Var}\,\xi = \sigma_\xi^2$, $\mathbf{Var}\,\zeta = \sigma_\zeta^2$, and $\eta = \zeta + \xi$, then

$$\inf_{f(\eta)}\mathbf{E}(\zeta - f(\eta))^2 = \mathbf{E}\left(\zeta - \mathbf{E}(\zeta|\eta)\right)^2 = \sigma_\zeta^2\sigma_\xi^2/(\sigma_\zeta^2+\sigma_\xi^2)\,.$$

Using this, we estimate the first term of the right-hand side of (25)

$$\inf_{\hat\theta} \sum_{k=1}^{\infty} \mathbf{E}(\hat\theta_k - \theta_k)^2 \geq \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2}{m_k^2 + \epsilon^2 \sigma_k^2} .$$

From the last inequality, (25) and (26) we obtain finally

$$r_\epsilon \geq \sum_{k=1}^{\infty} \frac{\epsilon^2 \sigma_k^2 m_k^2}{m_k^2 + \epsilon^2 \sigma_k^2} + O(\epsilon^\gamma) .$$

Theorem 2 is proved. $\square$

Proof of Theorem 3. The following upper bound for the minimax risk follows immediately from Theorem 1:

$$r_\epsilon \leq r_\epsilon^l = d_\epsilon = \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k . \tag{27}$$

Introduce

$$\alpha_\epsilon = \alpha_\epsilon(\gamma) = \frac{\left(8\gamma \left(\log \epsilon^{-1}\right) \sum_{k=1}^{\infty} a_k^2 \sigma_k^4 (1 - c_\epsilon a_k)_+^2\right)^{1/2}}{\sum_{k=1}^{\infty} a_k \sigma_k^2 (1 - c_\epsilon a_k)_+} .$$

Note that $\alpha_\epsilon \geq 0$ and $\alpha_\epsilon = o(1)$ as $\epsilon \to 0$ for any $\gamma > 0$ because of condition of Theorem. Now we take the sequence $m_k^2 = \tilde\theta_k^2 (1 + \alpha_\epsilon)^{-1}$, $k = 1, 2, \ldots$, with $\tilde\theta_k^2$ defined by (10). Equation (19) is satisfied. Indeed, by virtue of (14) we have

$$\sum_{k=1}^{\infty} a_k^2 m_k^2 + \left(8\gamma \left(\log \epsilon^{-1}\right) \sum_{k=1}^{\infty} a_k^4 m_k^4\right)^{1/2} = \sum_{k=1}^{\infty} a_k^2 \tilde\theta_k^2 = Q .$$

Applying now Theorem 2 for the sequence $(m_k^2,\ k = 1, 2, \ldots)$, we calculate

$$\begin{aligned} r_\epsilon &\geq \frac{\epsilon^2}{(1+\alpha_\epsilon)} \sum_{k=1}^{\infty} \frac{\sigma_k^2 \tilde\theta_k^2}{\tilde\theta_k^2 (1+\alpha_\epsilon)^{-1} + \epsilon^2 \sigma_k^2} + O(\epsilon^\gamma) \\ &= \epsilon^2 \sum_{k=1}^{N} \sigma_k^2 - \epsilon^2 c_\epsilon \sum_{k=1}^{N} \sigma_k^2 a_k - \epsilon^2 c_\epsilon \alpha_\epsilon \sum_{k=1}^{N} \frac{\sigma_k^2 a_k (1 - c_\epsilon a_k)}{1 + \alpha_\epsilon c_\epsilon a_k} + O(\epsilon^\gamma) . \end{aligned}$$

From (6) it follows that c_ϵ can not be of smaller order than ϵ^2. Choosing now some $\gamma > 4$ and recalling that $\alpha_\epsilon > 0$, $\alpha_\epsilon = o(1)$ as $\epsilon \to 0$, and $0 \leq c_\epsilon a_k \leq 1$ for $k = 1, 2, \ldots, N$ (see (16)), we conclude that the last lower bound, together with upper bound (27), proves Theorem 3. $\square$

Proof of Proposition 1. Let m_k, $k = 1, 2, \ldots$ be some sequence of positive numbers such that

$$\sum_{k=1}^{\infty} a_k^2 m_k^2 \leq Q , \tag{28}$$

i.e. $m=(m_k,\ k=1,2,\dots)\in\Theta$. Introduce

$$\nu_k(x)=(1/m_k)\nu_0(x/m_k),\quad k=1,2,\dots,$$

where $\nu_0(x)=I\{|x|\le 1\}\cos^2(\pi x/2)$. These are the probability densities with supports $[-m_k,m_k]$ respectively. It is easy to calculate Fisher information of distribution defined by density $\nu_k(x)$:

$$I(\nu_k)=\mathbf{E}_{\nu_k}[(\log\nu_k(\theta_k))']^2=I(\nu_0)/m_k^2=\pi^2/m_k^2\,, \tag{29}$$

where $\mathbf{E}_{\nu_k}$ denotes the expectation with respect to density ν_k.

We select a prior measure $d\mu(\theta)$ such that $\theta_k,\ k=1,2,\dots,$ are distributed independently with densities $\nu_k(x),\ k=1,2,\dots$, respectively. Since (28) provides that $supp\,\mu\in\Theta$ we proceed estimating the minimax risk (17) from below as follows:

$$\begin{aligned} r_\epsilon=\inf_{\hat\theta}\sup_{\Theta}\sum_{k=1}^{\infty}\mathbf{E}_\theta(\hat\theta_k-\theta_k)^2 \ &\ge\ \inf_{\hat\theta}\int_\Theta\sum_{k=1}^{\infty}\mathbf{E}_\theta(\hat\theta_k-\theta_k)^2d\mu(\theta)\\ &=\ \inf_{\hat\theta}\sum_{k=1}^{\infty}\mathbf{E}(\hat\theta_k-\theta_k)^2\,. \end{aligned}\tag{30}$$

For this case (see (29)) the inequality (18) yields

$$\mathbf{E}(\hat\theta_k-\theta_k)^2\ge\frac{1}{\pi^2/m_k^2+\epsilon^{-2}\sigma_k^{-2}}\,.$$

From this inequality and (30) we get that for any $m\in\Theta$ the minimax risk r_ϵ satisfies

$$r_\epsilon\ge\epsilon^2\sum_{k=1}^{\infty}\frac{\sigma_k^2m_k^2/\pi^2}{m_k^2/\pi^2+\epsilon^2\sigma_k^2}\,. \tag{31}$$

From this one obtains the following lower bound:

$$r_\epsilon\ge\sup_{m\in\Theta(Q)}\ \epsilon^2\sum_{k=1}^{\infty}\frac{\sigma_k^2m_k^2/\pi^2}{m_k^2/\pi^2+\epsilon^2\sigma_k^2}=\sup_{m\in\Theta(Q/\pi^2)}\ \epsilon^2\sum_{k=1}^{\infty}\frac{\sigma_k^2m_k^2}{m_k^2+\epsilon^2\sigma_k^2}=d_\epsilon(\Theta(Q/\pi^2))$$

The last relation and Theorem 1 complete the proof. □

Proof of Corollary 1. The left-hand side of the inequality (31) does not depend on m. Therefore, we can take any $m\in\Theta$. Now we make use of the vector $(\tilde\theta_k,\ k=1,2,\dots)$ defined by (10). Relation (14) provides that $\tilde\theta\in\Theta$. Substituting $\tilde\theta_k$ in (31), $k=1,2,\dots$, one calculates

$$r_\epsilon\ge\epsilon^2\sum_{k=1}^{\infty}\frac{\sigma_k^2\tilde\theta_k^2/\pi^2}{\tilde\theta_k^2/\pi^2+\epsilon^2\sigma_k^2}=\epsilon^2\sum_{k=1}^{N}\sigma_k^2-\epsilon^2\sum_{k=1}^{N}\frac{\sigma_k^2c_\epsilon a_k}{(1-\pi^{-2})c_\epsilon a_k+\pi^{-2}}\,.$$

Using now this and (16), we obtain that

$$r_\epsilon\ge\epsilon^2\sum_{k=1}^{N}\sigma_k^2-\epsilon^2\pi^2\sum_{k=1}^{N}\sigma_k^2c_\epsilon a_k\,.$$

Combining last relation with condition of Corollary and the upper bound (27) completes the proof. □

References

[1] D.L. Donoho and I.M. Johnstone, *Minimax risk over l_p-balls*, Technical Report 204, Department of Statistics, University of California, Berkeley, 1989.

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

[3] S.Y. Efroimovich and M.S. Pinsker, *Estimation of square-integrable density of a random veriable*, Problems Inform. Transmission, 3 (1982), pp. 175-189.

[4] R.D. Gill and B.Y. Levit, *Applications of the van Trees inequality: A Bayesian Cramer-Rao bound*, Bernoulli, 1 (1995).

[5] G.K. Golubev and B.Y. Levit, *On the Second Order Minimax Esimation of Distribution Functions*, Preprint 881, Department of Mathematics, University of Utrecht, 1994.

[6] G.K. Golubev and M. Nussbaum, *A risk bound in Sobolev class regression*, Ann. Statist., 18 (1990), pp. 758-778.

[7] I.S. Gradshtein and I.M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, New York, 1980.

[8] I.A. Ibragimov, R.Z. Khasminskii, *On nonparametric estimation of the value of a linear functional in Gaussian white noise*, Theor. Prob. Appl., 29 (1984), 18-32.

[9] M.S. Pinsker, *Optimal filtering of square integrable signals in Gaussian white noise*, Problems Inform. Transmission, 2 (1980), pp. 120-133.

[10] H.L. van Trees, *Detection, Estimation and Modulation Theory, Part 1*, Wiley, New York, 1968.