Solutions to the exercises from "Introduction to Nonparametric Estimation" A.B.Tsybakov

Pavel Chigansky

Department of Statistics, The Hebrew University, Mount Scopus, Jerusalem 91905, Israel

E-mail address: pchiga@mscc.huji.ac.il

These solutions have been written during the course, based on A. Tsybakov's "Introduction to Nonparametric Estimation", I taught at the Department of Statistics of the Hebrew University during the spring semester of 2012. All the errors and mistypes are exclusively mine and I will be delighted to get any improvement suggestions, remarks or bug reports. I am grateful to Mr. Luong Minh Khoa for proofreading of these notes.

P.Ch, 2012

1. Nonparametric Estimators

EXERECISE 1.1.

(1) Argue that any symmetric kernel K is a kernel of order 1 whenever the function $u \mapsto uK(u)$ is integrable.

(2) Find the maximum order of the Silverman kernel:

$$K(u) = \frac{1}{2} \exp\big(- |u|/\sqrt{2}\big) \sin\big(|u|/\sqrt{2} + \pi/4\big).$$

Hint: Apply the Fourier transform and write the Silverman kernel as

$$K(u) = \int_{-\infty}^{\infty} \frac{\cos(2\pi t u)}{1 + (2\pi t)^4} dt.$$

Solution

Note that $u \mapsto |u|^j K(u)$ is integrable for any $j \geq 1$ and hence

$$\int u^m K(u) du = i^m \hat{K}^{(m)}(0), \quad m \geq 1,$$

where $\hat{K}(t) = \int e^{itu} K(u) du$ is the Fourier transform of K, given by

$$\hat{K}(\omega) = \frac{2\pi}{1 + \omega^4}.$$

Hence for $m = 1$,

$$\hat{K}'(0) = -\frac{4\omega^3}{(1+\omega^4)^2}\Big|_{\omega:=0} = 0$$

which is also obvious by symmetry. Further, a calculation reveals that $\hat{K}''(0) = \hat{K}'''(0) = 0$, while $\hat{K}^{(4)}(0) = -2\pi \cdot 24$. Hence the Silverman kernel is of order 3.

EXERECISE 1.2. Kernel estimator of the s-th derivative $p^{(s)}$ of a density $p \in \mathcal{P}(\beta, L)$, $s < \beta$, can be defined as follows:

$$\hat{p}_{n,s}(x) = \frac{1}{nh^{s+1}} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right).$$

Here $h > 0$ is a bandwidth and $K : \mathbb{R} \mapsto \mathbb{R}$ is a bounded kernel with support $[-1, 1]$ satisfying for $\ell = \lfloor \beta \rfloor$:

$$\int u^j K(u) du = 0, \quad j \in \{0, ..., \ell\} \setminus \{s\} \tag{1.1}$$

$$\int u^s K(u) du = s!. \tag{1.2}$$

(1) Prove that, uniformly over the class $\mathcal{P}(\beta, L)$, the bias of $\hat{p}_{n,s}(x_0)$ is bounded by $ch^{\beta - s}$ and the variance of $\hat{p}_{n,s}(x_0)$ is bounded by $\frac{c'}{nh^{2s+1}}$, where c and c' are appropriate constants and x_0 is a point in $\mathbb{R}$.

(2) Prove that the maximum of the MSE of $\hat{p}_{n,s}(x_0)$ over $\mathcal{P}(\beta, L)$ is of order $O\Big(n^{-\frac{2(\beta-s)}{2\beta+1}}\Big)$ as $n \to \infty$ if the bandwidth $h := h_n$ is chosen optimally.

(3) Let (φ_m) be the orthonormal Legendre basis on $[-1,1]$. Show that the kernel

$$K(u) = \sum_{m=0}^{\ell} \varphi_m^{(s)}(0)\varphi_m(u)\mathbf{1}_{\{|u|\le 1\}}$$

satisfies the conditions (1.1) and (1.2).

<u>Solution</u>

(1) The variance of the estimate $\hat{p}_{n,s}(x_0)$ is

$$\mathsf{E}_p\Big(\hat{p}_{n,s}(x_0) - \mathsf{E}_p\hat{p}_{n,s}(x_0)\Big)^2 = \frac{1}{nh^{2(s+1)}}\mathsf{E}_p\left(K\left(\frac{X_1-x_0}{h}\right) - \mathsf{E}_pK\left(\frac{X_1-x_0}{h}\right)\right)^2 \le$$
$$\frac{1}{nh^{2(s+1)}}\mathsf{E}_pK^2\left(\frac{X_1-x_0}{h}\right) = \frac{1}{nh^{2s+1}}\int K^2(v)p(x_0+vh)dv \le K_{\max}^2\frac{1}{nh^{2s+1}},$$

i.e. the claim holds with $c' := K_{\max}^2 = \max_{x\in[-1,1]}|K(x)|$.
The bias term satisfies

$$b(x_0) = \mathsf{E}_p\hat{p}_{n,s}(x_0) - p^{(s)}(x_0) = \frac{1}{h^{s+1}}\int K\left(\frac{u-x_0}{h}\right)p(u)du - p^{(s)}(x_0) =$$
$$\frac{1}{h^s}\int K(v)p(x_0+vh)dv - p^{(s)}(x_0) =$$
$$\frac{1}{h^s}\int K(v)\Big(p(x_0) + p^{(1)}(x_0)vh + \dots + \frac{1}{\ell!}p^{(\ell)}(x_0+\tau vh)(vh)^\ell\Big)dv - p^{(s)}(x_0).$$

By the conditions (1.1) and (1.2),

$$|b(x_0)| \le \frac{1}{\ell!}\frac{1}{h^s}\int\Big|K(v)\Big(p^{(\ell)}(x_0+\tau vh) - p^{(\ell)}(x_0)\Big)(vh)^\ell\Big|dv \le$$
$$\frac{1}{\ell!}\frac{1}{h^s}\int |K(v)|L|vh|^{\beta-\ell}|vh|^\ell dv \le \frac{L}{\ell!}h^{\beta-s}\int |K(v)||v|^\beta dv =: ch^{\beta-s}.$$

(2) For any $n \ge 1$ and $h > 0$,

$$\sup_{p\in\mathcal{P}(\beta,L)}\mathsf{E}_p\big(\hat{p}_{n,s}(x_0) - p(x_0)\big)^2 = b^2(x_0) + \sigma^2(x_0) \le ch^{2(\beta-s)} + \frac{c'}{nh^{2s+1}}.$$

The right hand side is minimized over h by

$$h_n^* := \left(\frac{c'(2s+1)}{2c(\beta-s)}\right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}$$

so that

$$\sup_{p\in\mathcal{P}(\beta,L)}\mathsf{E}_p\big(\hat{p}_{n,s}(x_0) - p(x_0)\big)^2 \le Cn^{-\frac{2(\beta-s)}{2\beta+1}}.$$

(3) Since φ_i is a complete orthonormal basis in $L([-1,1])$, the functions u^j satisfy

$$u^j = \sum_{i=1}^{j} b_{ji}\varphi_i(u),$$

with some constants b_{ji}. Hence for $j \in \{0, ..., \ell\}$,

$$\int u^j K(u)du = \sum_{m=0}^{\ell} \varphi_m^{(s)}(0) \sum_{i=1}^{j} b_{ji} \int_{-1}^{1} \varphi_i(u)\varphi_m(u)du =$$

$$\sum_{m=0}^{j} \varphi_m^{(s)}(0) b_{jm} = \frac{d^s}{du^s}\left(\sum_{m=0}^{j} \varphi_m(u) b_{jm}\right)_{u=0} = \left(\frac{d^s}{du^s} u^j\right)_{u=0}.$$

The latter term vanishes both for $j < s$ and $j > s$. For $j = s$, $\int u^j K(u)du = s!$ is obtained.

Exerecise 1.3. Consider the estimator of the two dimensional kernel density $p(x,y)$ from the i.i.d. sample $(X_1,Y_1), ..., (X_n,Y_n)$

$$\hat{p}_n(x,y) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)K\left(\frac{Y_i - y}{h}\right), \quad (x,y) \in \mathbb{R}^2.$$

Assume that the density p belongs to the class of all probability densities on $\mathbb{R}^2$ satisfying the Holder condition

$$\big|p(x,y) - p(x',y')\big| \le L\Big(|x - x'|^\beta + |y - y'|^\beta\Big), \quad (x,y), (x'y') \in \mathbb{R}^2,$$

with given constants $0 < \beta \le 1$ and $L > 0$. Let (x_0, y_0) be a fixed point in $\mathbb{R}^2$. Derive the upper bounds for the bias and the variance of $\hat{p}_n(x_0,y_0)$ and an upper bound for the mean squared risk at (x_0,y_0). Find the minimizer $h = h_n^*$ of the upper bound of the risk and the corresponding rate of convergence.

<u>Solution</u>

The variance is given by

$$\sigma^2(x_0,y_0) = \mathsf{E}_p\Big(\hat{p}_n(x_0,y_0) - \mathsf{E}_p\hat{p}(x_0,y_0)\Big)^2 =$$

$$\frac{1}{n^2h^4}\mathsf{E}_p\left(\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)K\left(\frac{Y_i - y}{h}\right) - \mathsf{E}_p K\left(\frac{X_i - x}{h}\right)K\left(\frac{Y_i - y}{h}\right)\right)^2 =$$

$$\le \frac{1}{n}\frac{1}{h^4}\int\int K^2\left(\frac{u - x_0}{h}\right)K^2\left(\frac{v - y_0}{h}\right)p(u,v)dudv =$$

$$\frac{1}{n}\frac{1}{h^2}\int\int K^2(u)K^2(v)p(x_0 + uh, y_0 + vh)dudv \le \frac{1}{n}\frac{1}{h^2}p_{\max}\left(\int K^2(u)du\right)^2 =: c'\frac{1}{n}\frac{1}{h^2}$$

where $p(x,y) \le p_{\max}$ is assumed (later we shall see that under the assumptions of the problem, such $p_{\max}$ indeed exists). The bias term can be bounded as follows

$$b(x_0,y_0) = \mathsf{E}_p\hat{p}_n(x_0,y_0) - p(x_0,y_0) = \frac{1}{h^2}\int\int K\Big(\frac{u-x_0}{h}\Big)K\Big(\frac{v-y_0}{h}\Big)p(u,v)dudv - p(x_0,y_0) =$$
$$\int\int K(u)K(v)\Big(p(x_0+uh, y_0+vh) - p(x_0,y_0)\Big)dudv.$$

Using the Holder property we get

$$|b(x_0,y_0)| \le \int\int \Big|K(u)K(v)\Big| L\Big(|uh|^\beta + |vh|^\beta\Big)dudv =$$
$$h^\beta \int\int \Big|K(u)K(v)\Big| L\Big(|u|^\beta + |v|^\beta\Big)dudv =: ch^\beta.$$

Note that if K is taken to be bounded, then $|\mathsf{E}_p\hat{p}_n(x_0,y_0)| \le K^2_{\max}$ and hence $|p(x_0,y_0)| \le K^2_{\max} + c =: p_{\max}$ for $h < 1$. Hence the MSE is given by

$$\mathrm{MSE}(x_0,y_0) = b^2(x_0,y_0) + \sigma^2(x_0,y_0) \le c^2h^{2\beta} + c'\frac{1}{n}\frac{1}{h^2},$$

which is optimized by

$$h_n^* := Cn^{-\frac{1}{2\beta+2}}$$

where C is a constant. This gives

$$\mathrm{MSE}(x_0,y_0) = O\Big(n^{-\frac{2\beta}{2\beta+2}}\Big), \quad n \to \infty.$$

Exerecise 1.4. Define the $LP(\ell)$ estimators of the derivatives $f^{(s)}(x)$, $s = 1, ..., \ell$ by

$$\hat{f}_{ns}(x) = \Big(U^{(s)}(0)\Big)^\top \hat{\theta}_n(x)h^{-s},$$

where $U^{(s)}(x)$ is the vector whose coordinates are the s-th derivatives of the corresponding coordinates of $U(x)$.

(1) Prove that if $B_n(x) > 0$, then the estimator $\hat{f}_{ns}(x)$ is linear and it reproduces the s-th derivative of polynomials of degree less than or equal ℓ, i.e. if $\hat{f}_{ns}(x)$ is applied to $Y_i := Q(X_i)$, $i = 1, ..., n$ where Q is a polynomial with $\deg(Q) \le \ell$, it yields $Q^{(s)}(x)$.

(2) Prove that, under the assumptions similar to the case $s = 0$, the maximum of the MSE of $\hat{f}_{ns}(x)$ over $\Sigma(\beta, L)$ is of order $O\Big(n^{-\frac{2(\beta-s)}{2\beta+1}}\Big)$ as $n \to \infty$ if the bandwidth is chosen optimally.

Solution

1. If $B_n(x) > 0$, then

$$\hat{\theta}_n(x) = \sum_{i=1}^n Y_i \frac{1}{nh} B_n^{-1}(x) U\left(\frac{X_i - x}{h}\right) K\left(\frac{X_i - x}{h}\right),$$

and hence

$$\hat{f}_{ns}(x) = \left(U^{(s)}(0)\right)^\top \hat{\theta}_n(x)h^{-s} = \sum_{i=1}^{n} Y_i W_{ni}^{s*}(x),$$

with

$$W_{ni}^{s*}(x) = \frac{1}{nh^{1+s}}\left(U^{(s)}(0)\right)^\top B_n^{-1}(x)U\left(\frac{X_i-x}{h}\right)K\left(\frac{X_i-x}{h}\right)$$

i.e. the estimator $\hat{f}_{ns}$ is linear.

If Q is a polynomial of order ℓ, then

$$Q(X_i) = Q(x) + Q'(x)(X_i-x) + ... + \frac{1}{\ell!}Q^{(\ell)}(x)(X_i-x)^\ell = q(x)^\top U\Big(\frac{X_i-x}{h}\Big),$$

where

$$q(x) = \big(Q(x), Q'(x)h, ..., Q^{(\ell)}h^\ell\big).$$

Consequently,

$$\hat{\theta}_n(x) = \text{argmin}_{\theta\in\mathbb{R}^{\ell+1}} \sum_{i=1}^{n}\Big(Q(X_i) - \theta^\top U\Big(\frac{X_i-x}{h}\Big)\Big)^2 K\Big(\frac{X_i-x}{h}\Big) =$$
$$\text{argmin}_{\theta\in\mathbb{R}^{\ell+1}} \sum_{i=1}^{n}\Big((q(x)-\theta)^\top U\Big(\frac{X_i-x}{h}\Big)\Big)^2 K\Big(\frac{X_i-x}{h}\Big) = q(x).$$

Note that

$$U_j^{(s)}(x) = \frac{d^s}{dx^s}\frac{1}{j!}x^j = \begin{cases} 0 & s>j \\ \frac{1}{(j-s)!}x^{j-s} & s\le j \end{cases}, \quad j=1,...,\ell$$

and hence the s-th entry of the vector $U_j^{(s)}(0)$ is 1 and all the rest are zeros.

Hence if $\hat{f}_{ns}(x)$ is applied to $Y_i = Q(X_i)$ it yields

$$\hat{f}_{ns}(x) = \left(U^{(s)}(0)\right)^\top \hat{\theta}_n(x)h^{-s} = \left(U^{(s)}(0)\right)^\top q(x)h^{-s} = Q^{(s)}(x).$$

In particular,

$$\sum_{i=1}^{n}\Big(\frac{X_i-x}{h}\Big)^k W_{ni}^{s*}(x) = \begin{cases} 0 & k\in\{0,...,\ell\}\setminus\{s\} \\ s! & k=s \end{cases} \tag{1.3}$$

2. The weights $W_{ni}^{s*}(x)$ satisfy the properties
   (i) $\max_{1\le i\le n}\left|W_{ni}^{s*}(x)\right| \le C_*\frac{1}{nh^{1+s}}$
   (ii) $\sum_{i=1}^{n}\left|W_{ni}^{s*}(x)\right| \le C_*\frac{1}{h^s}$
   (iii) $W_{ni}^{s*}(x) = 0$ whenever $|X_i-x|>h$

with a constant C_* and $h \geq \frac{1}{2n}$. Indeed,

$$\max_{1\leq i\leq n} \left|W_{ni}^{s*}(x)\right| =$$
$$\frac{1}{nh^{1+s}} \max_{1\leq i\leq n} \left|\left(U^{(s)}(0)\right)^\top B_n^{-1}(x) U\left(\frac{X_i - x}{h}\right) K\left(\frac{X_i - x}{h}\right)\right| \leq$$
$$\frac{1}{nh^{1+s}} \max_{1\leq i\leq n} \left\|B_n^{-1}(x) U\left(\frac{X_i - x}{h}\right) K\left(\frac{X_i - x}{h}\right)\right\| \mathbf{1}_{\{|X_i - x|\leq h\}} \leq$$
$$\frac{1}{nh^{1+s}} \frac{K_{\max}}{\lambda_0} \max_{1\leq i\leq n} \left\|U\left(\frac{X_i - x}{h}\right)\right\| \mathbf{1}_{\{|X_i - x|\leq h\}} \leq$$
$$\frac{1}{nh^{1+s}} \frac{K_{\max}}{\lambda_0} \max_{1\leq i\leq n} \left\|U\left(1\right)\right\| \leq \frac{2}{nh^{1+s}} \frac{K_{\max}}{\lambda_0}.$$

Similarly,

$$\sum_{i=1}^n \left|W_{ni}^{s*}(x)\right| =$$
$$\frac{1}{nh^{1+s}} \sum_{i=1}^n \left|\left(U^{(s)}(0)\right)^\top B_n^{-1}(x) U\left(\frac{X_i - x}{h}\right) K\left(\frac{X_i - x}{h}\right)\right| \mathbf{1}_{\{|X_i - x|\leq h\}} \leq$$
$$\frac{1}{h^{1+s}} \frac{K_{\max}}{\lambda_0} \left\|U\left(1\right)\right\| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{|X_i - x|\leq h\}} \leq$$
$$\frac{1}{h^{1+s}} \frac{K_{\max}}{\lambda_0} \left\|U\left(1\right)\right\| a_0 \max\left(2h, \frac{1}{n}\right) \leq \frac{1}{h^s} \frac{4K_{\max} a_0}{\lambda_0},$$

and (i) and (ii) hold with e.g. $C_* := \frac{2K_{\max}}{\lambda_0}(1 + 2a_0)$. The claim (iii) is obvious.

Further,

$$\mathsf{E}_f\left(\hat{f}_{ns}(x) - f^{(s)}(x)\right)^2 =$$
$$\mathsf{E}_f\left(\hat{f}_{ns}(x) - \mathsf{E}_f \hat{f}_{ns}(x)\right)^2 + \left(\mathsf{E}_f \hat{f}_{ns}(x) - f^{(s)}(x)\right)^2 = \sigma^2(x) + b^2(x).$$

For the variance term we have

$$\sigma^2(x) = \mathsf{E}_f\left(\sum_{i=1}^n \xi_i W_{ni}^{s*}(x)\right)^2 \leq \sigma_{\max}^2 \sum_{i=1}^n \left(W_{ni}^{s*}(x)\right)^2 \leq \sigma_{\max}^2 C_*^2 \frac{1}{nh^{1+2s}} := q_1 \frac{1}{nh^{1+2s}}.$$

The bias term can be bounded as follows:

$$b(x) = \mathsf{E}_f \hat{f}_{ns}(x) - f^{(s)}(x) = \sum_{i=1}^n f(X_i) W_{ni}^{s*}(x) - f^{(s)}(x) =$$
$$\sum_{i=1}^n \left(f(x) + f'(x)(X_i - x) + \ldots + \frac{1}{\ell!} f^\ell\big(x + \tau_i(X_i - x)\big)(X_i - x)^\ell\right) W_{ni}^{s*}(x) - f^{(s)}(x) =$$
$$\frac{1}{\ell!} \sum_{i=1}^n \left(f^\ell\big(x + \tau_i(X_i - x)\big) - f^\ell\big(x\big)\right)(X_i - x)^\ell W_{ni}^{s*}(x),$$

where $|\tau_i| < 1$ and we used (1.3). Hence

$$|b(x)| \le \frac{1}{\ell!} \sum_{i=1}^{n} \Big| f^{\ell}\big(x + \tau_i(X_i - x)\big) - f^{\ell}\big(x\big)\Big| |X_i - x|^{\ell} |W_{ni}^{s*}(x)| \le$$

$$\frac{L}{\ell!} \sum_{i=1}^{n} |x - X_i|^{\beta} |W_{ni}^{s*}(x)| \mathbf{1}_{\{|x-X_i| \le h\}} \le \frac{LC_*}{\ell!} h^{\beta - s} =: q_2 h^{\beta - s}.$$

Assembling all parts together we obtain

$$\sup_{x \in [0,1]} \mathsf{E}_f \Big(\hat{f}_{ns}(x) - f^{(s)}(x) \Big)^2 \le q_1 \frac{1}{nh^{1+2s}} + q_2^2 h^{2(\beta - s)}.$$

The optimal choice of the bandwidth is $h_n^* := cn^{-\frac{1}{2\beta+1}}$ for which we get

$$\sup_{f \in \Sigma(\beta, L)} \sup_{x \in [0,1]} \mathsf{E}_f \Big(\hat{f}_{ns}(x) - f^{(s)}(x) \Big)^2 \le Cn^{-\frac{2(\beta - s)}{2\beta+1}}$$

with a constant $C > 0$, for all sufficiently large n.

Exerecise 1.5. Show that the rectangular kernel

$$K(u) = \frac{1}{2} I(|u| \le 1)$$

and the biweight kernel

$$K(u) = \frac{15}{16}(1 - u^2)^2 \mathbf{1}_{\{|u| \le 1\}}$$

are inadmissible.

Solution

The Fourier transform of the rectangular kernel is

$$\hat{K}(\omega) = \frac{1}{2} \int_{-1}^{1} e^{ix\omega} dx = \frac{e^{i\omega} - e^{i\omega}}{2i\omega} = \frac{\sin(\omega)}{\omega}.$$

This is a continuous function (when extended to zero by continuity) and $\hat{K}(3\pi/2) = -2\pi/3$. Hence the kernel is inadmissible by Proposition 1.8.

For the biweight kernel

$$\hat{K}(\omega) = \frac{15}{\omega^5} \Big((3 - \omega^2) \sin \omega - 3\omega \cos \omega \Big).$$

The kernel is inadmissible by continuity, since $\hat{K}(2\pi) < 0$.

Exerecise 1.6. Let $K \in \mathbb{L}_2(\mathbb{R})$ be symmetric and such that $\hat{K} \in L_\infty(\mathbb{R})$. Show that

(1) the condition

$$\exists A < \infty : \quad \operatorname*{ess\,sup}_{t \in \mathbb{R} \setminus \{0\}} \frac{\big|1 - \hat{K}(t)\big|}{|t|^{\beta}} \le A, \tag{1.4}$$

is equivalent to

$$\exists t_0, A_0 < \infty : \quad \operatorname*{ess\,sup}_{0<|t|<|t_0|} \frac{|1-\hat{K}(t)|}{|t|^\beta} \le A_0. \tag{1.5}$$

(2) for integer β the condition (1.4) is satisfied if K is a kernel of order $\beta - 1$ and $\int |u|^\beta |K(u)| du < \infty$

Solution

1. Suppose (1.5) holds, then

$$\operatorname*{ess\,sup}_{t\in\mathbb{R}\setminus\{0\}} \frac{\left|1-\hat{K}(t)\right|}{|t|^\beta} \le \operatorname*{ess\,sup}_{0<|t|<|t_0|} \frac{|1-\hat{K}(t)|}{|t|^\beta} + \operatorname*{ess\,sup}_{|t|\ge|t_0|} \frac{\left|1-\hat{K}(t)\right|}{|t|^\beta} \le$$
$$A_0 + \frac{1+\|\hat{K}\|_\infty}{|t_0|^\beta},$$

which verifies (1.4). The other direction is obvious.

2. If $\int |u|^\beta K(u) du < \infty$, the Fourier transform is β times differentiable at zero and $\hat{K}^{(i)}(0) = 0$ for $i = 1, ..., \beta - 1$, since K is of order $\beta - 1$. Hence,

$$\hat{K}(t) = 1 + \frac{1}{\beta!}\hat{K}^{(\beta)}(\tau t)t^\beta,$$

where $\tau \in [-1, 1]$. Hence

$$\sup_{|t|\le 1} \frac{|1-\hat{K}(t)|}{|t|^\beta} \le \frac{1}{\beta!} \sup_{s\in[-1,1]} |\hat{K}^{(\beta)}(s)| < \infty,$$

where we used the fact that $\hat{K}(t)$ is bounded on $[-1, 1]$, being continuous on it. By the Riemann-Lebesgue lemma, $\lim_{t\to\infty} \hat{K}(t) = 0$ and by continuity $\hat{K}$ is bounded. The claim now follows from 1.

Exerecise 1.7. Let $\mathcal{P}$ be the class of all probability densities p on $\mathbb{R}$ such that

$$\int \exp(\alpha|\omega|^r)|\varphi(\omega)|^2 d\omega \le L^2,$$

where $\alpha > 0$, $r > 0$, $L > 0$ are given constants and φ is the Fourier transform of p. Show that for any $n \ge 1$ the kernel density estimator $\hat{p}$ with the sinc kernel and appropriately chosen bandwidth $h = h_n$ satisfies

$$\sup_{p\in\mathcal{P}} \mathsf{E}_p \int \left(\hat{p}_n(x) - p(x)\right)^2 dx \le C\frac{(\log n)^{1/r}}{n},$$

where $C > 0$ is a constant dependng only on r, L and α.

Solution

Recall that

$$2\pi\text{MISE} = 2\pi\mathsf{E}_p \int \big(\hat{p}_n(x) - p(x)\big)^2 dx = \int \big|1 - \hat{K}(h\omega)\big|^2 |\varphi(\omega)|^2 d\omega +$$
$$\frac{1}{n}\int |\hat{K}(h\omega)|^2 d\omega - \frac{1}{n}\int |\hat{K}(h\omega)|^2 |\varphi(\omega)|^2 d\omega.$$

For the sinc kernel $K(u) = \frac{\sin u}{\pi u}$ with $\hat{K}(\omega) = \mathbf{1}_{\{|\omega|\le 1\}}$, the latter gives

$$2\pi\text{MISE} = \int \mathbf{1}_{\{|\omega|>1/h\}} |\varphi(\omega)|^2 d\omega + \frac{1}{n}\int \mathbf{1}_{\{|\omega|\le 1/h\}} d\omega - \frac{1}{n}\int \mathbf{1}_{\{|\omega|\le 1/h\}} |\varphi(\omega)|^2 d\omega =$$
$$\int_{\mathbb{R}\setminus[-1/h,1/h]} |\varphi(\omega)|^2 d\omega + \frac{2}{nh} - \frac{1}{n}\int_{-1/h}^{1/h} |\varphi(\omega)|^2 d\omega \le$$
$$\int_{\mathbb{R}\setminus[-1/h,1/h]} e^{-\alpha|\omega|^r} e^{\alpha|\omega|^r} |\varphi(\omega)|^2 d\omega + \frac{2}{nh} \le$$
$$e^{-\alpha|1/h|^r} \int_{\mathbb{R}\setminus[-1/h,1/h]} e^{\alpha|\omega|^r} |\varphi(\omega)|^2 d\omega + \frac{2}{nh} \le e^{-\alpha|1/h|^r} L^2 + \frac{2}{nh}.$$

For $h_n := \dfrac{\alpha^{1/r}}{(\log n)^{1/r}}$ we get the bound

$$\text{MISE} \le \frac{1}{2\pi}\left(\frac{1}{n}L^2 + \frac{1}{\alpha^{1/r}}\frac{2(\log n)^{1/r}}{n}\right) \le C\frac{(\log n)^{1/r}}{n},$$

with an obvious constant C.

Exerecise 1.8. Let $\mathcal{P}_a$, where $a > 0$, be the class of all probability densities p on $\mathbb{R}$ such that the support of the characteristic function φ is included in a given interval $[-a, a]$. Show that for any $n \ge 1$, the kernel density estimator $\hat{p}_n$ with the sinc kernel and appropriately chosen bandwidth h satisfies

$$\sup_{p\in\mathcal{P}_a} \int \big(\hat{p}_n(x) - p(x)\big)^2 dx \le \frac{a}{\pi n}.$$

This example, due to Ibragimon and Hasminskii (1983), shows that it is possible to estimate the density with rate $\sqrt{n}$ on sufficiently small nonparametric classes of functions.

Solution

As in the previous problem,

$$2\pi\text{MISE} \le \int_{\mathbb{R}\setminus[-1/h,1/h]} |\varphi(\omega)|^2 d\omega + \frac{2}{nh}.$$

The claim follows with the choice $h := \frac{1}{a}$, since $\varphi(\omega) = 0$ for $|\omega| > a$.

Exerecise 1.9. Let $(X_1, ..., X_n)$ be an i.i.d. sample from a density $p \in L_2[0, 1]$.

(1) Show that $\hat{c}_j$ are unbiased estimators of the Fourier coefficients $c_j = \int_0^1 p(x)\varphi_j(x)dx$ and find the variance of $\hat{c}_j$.

(2) Express the MISE of the estimator $\hat{p}_{n,N}$ as a function of p and the sequence (φ_j). Denote it by MISE(N).

(3) Derive an unbiased risk estimation method. Show that

$$\mathsf{E}_p\big(\hat{J}(N)\big) = \text{MISE}(N) - \int p^2,$$

where

$$\hat{J}(N) = \frac{1}{n-1}\sum_{j=1}^{N}\left(\frac{2}{n}\sum_{i=1}^{n}\varphi_j^2(X_i) - (n+1)\hat{c}_j^2\right).$$

Propose the data driven selector of N.

(4) Suppose now that (φ_j) is the trigonometric basis. Show that the MISE of $\hat{p}_{n,N}$ is bounded by

$$\frac{N+1}{n} + \rho_N,$$

where $\rho_N = \sum_{j=N+1}^{\infty} c_j^2$. Use this bound to prove that uniformly over the class of all the densities p belonging to the Sobolev class of periodic functions $W^{per}(\beta, L)$, $\beta > 0$ and $L > 0$, the MISE of $\hat{p}_{n,N}$ is of the order $O\left(n^{-\frac{2\beta}{2\beta+1}}\right)$ for an appropriate choice of $N = N_n$.

Solution

1. Clearly, $\hat{c}_j$ are unbiased:

$$\mathsf{E}_p\hat{c}_j = \mathsf{E}_p\varphi_j(X_1) = c_j$$

and

$$\text{var}_p(\hat{c}_j) = \mathsf{E}_p(\hat{c}_j - c_j)^2 = \frac{1}{n}\mathsf{E}_p\big(\varphi_j(X_1) - c_j\big)^2 = \frac{1}{n}\Big(\mathsf{E}_p\varphi_j^2(X_1) - c_j^2\Big) = \frac{1}{n}\Big(\int \varphi_j^2 p - \Big(\int \varphi_j p\Big)^2\Big)$$

2. Note that

$$\mathsf{E}\hat{p}_{n,N}(x) = \sum_{j=1}^{N} c_j\varphi_j(x)$$

and

$$\text{var}_p\Big(\hat{p}_{n,N}(x)\Big) = \mathsf{E}_p\Big(\sum_{j=1}^{N}(\hat{c}_j - c_j)\varphi_j(x)\Big)^2.$$

Hence

$$\text{MISE}(N) = \mathsf{E}_p\int\Big(\sum_{j=1}^{N}(\hat{c}_j - c_j)\varphi_j\Big)^2 + \int\Big(\sum_{j=1}^{N}c_j\varphi_j - \sum_{j=1}^{\infty}c_j\varphi_j\Big)^2 =$$

$$\sum_{j=1}^{N}\mathsf{E}_p(\hat{c}_j - c_j)^2 + \sum_{j=N+1}^{\infty}c_j^2 = \frac{1}{n}\sum_{j=1}^{N}\Big(\mathsf{E}_p\varphi_j^2(X_1) - c_j^2\Big) + \sum_{j=N+1}^{\infty}c_j^2$$

3. Following the suggestion

$$\mathsf{E}_p\big(\hat{J}(N)\big) = \frac{1}{n-1}\sum_{j=1}^{N}\left(\frac{2}{n}\sum_{i=1}^{n}\mathsf{E}_p\varphi_j^2(X_i) - (n+1)\mathsf{E}_p\hat{c}_j^2\right) =$$

$$\frac{1}{n-1}\sum_{j=1}^{N}\left(2\mathsf{E}_p\varphi_j^2(X_1) - (n+1)\Big(\frac{1}{n}\Big(\mathsf{E}_p\varphi_j^2(X_1) - c_j^2\Big) + c_j^2\Big)\right) =$$

$$\frac{1}{n}\sum_{j=1}^{N}\Big(\mathsf{E}_p\varphi_j^2(X_1) - (n+1)c_j^2\Big)$$

and hence

$$\mathsf{E}_p\big(\hat{J}(N)\big) - \mathrm{MISE}(N) =$$

$$\frac{1}{n}\sum_{j=1}^{N}\Big(\mathsf{E}_p\varphi_j^2(X_1) - (n+1)c_j^2\Big) - \frac{1}{n}\sum_{j=1}^{N}\Big(\mathsf{E}_p\varphi_j^2(X_1) - c_j^2\Big) - \sum_{j=N+1}^{\infty}c_j^2 =$$

$$-\sum_{j=1}^{\infty}c_j^2 = -\int p^2,$$

where we used Parseval's identity.

Since the same N maximizes both $\mathrm{MISE}(N)$ and $\mathsf{E}_p\big(\hat{J}(N)\big)$, it makes sense to select

$$\hat{N} := \mathrm{argmin}_{N\ge 1} J(N)$$

and to plug it into $\hat{p}_{n,N}$.

4. Recall that for even j

$$\varphi_j(x) = \sqrt{2}\cos(\pi j x),$$

and hence

$$\varphi_j^2(x) = 2\frac{1+\cos(2\pi j x)}{2} = 1 + \frac{1}{\sqrt{2}}\varphi_{2j}(x).$$

Similarly, for odd j

$$\varphi_j(x) = \sqrt{2}\sin(\pi(j-1)x),$$

and

$$\varphi_j^2(x) = 2\frac{1-\cos(2\pi(j-1)x)}{2} = 1 - \frac{1}{\sqrt{2}}\varphi_{2(j-1)}(x).$$

Hence

$$\mathsf{E}_p\varphi_j^2(X_1) = 1 + \frac{1}{\sqrt{2}}\begin{cases} c_{2j} & j \text{ is even} \\ -c_{2(j-1)} & j \text{ is odd}\end{cases}$$

Consequently, e.g. for even N

$$\mathrm{MISE}(N) = \frac{1}{n}\sum_{j=1}^{N}\Big(\mathsf{E}_p\varphi_j^2(X_1) - c_j^2\Big) + \rho_N = \frac{N}{n} + \frac{1}{\sqrt{2}}c_{2N} - \frac{1}{n}\sum_{j=1}^{N}c_j^2 + \rho_N \le \frac{N+1}{n} + \rho_N,$$

The same bound holds for odd N's.

If $p \in W^{per}(\beta, L)$, then the sequence of Fourier coefficients (c_j) belong to the Sobolev ellipsoid $\Theta(\beta, Q)$ with $Q = L^2/\pi^{2\beta}$ and hence

$$\rho_N = \sum_{j=N+1}^{\infty} c_j^2 \le \frac{1}{a_{N+1}^2} \sum_{j=N+1}^{\infty} c_j^2 a_j^2 \le Q N^{-2\beta}.$$

Now the choice $N_n := \lfloor \alpha n^{\frac{1}{2\beta+1}} \rfloor$ with a constant $\alpha > 0$ yields the claimed upper bound.

EXERECISE 1.10. Consider the nonparametric regression model under Assumption (A) and suppose that f belongs to the Sobolev class of periodic functions $W^{per}(\beta, L)$ with $\beta \ge 2$. The aim of this exercise is to study the weighted projection estimator

$$\hat{f}_{n,\lambda}(x) = \sum_{j=1}^{n} \lambda_j \hat{\theta}_j \varphi_j(x),$$

where λ_j's are real constants (weights), $\hat{\theta}_j = \frac{1}{n}\sum_{i=1}^n Y_i \varphi_j(X_i)$ are the Fourier coefficients estimates.

(1) Prove that the risk MISE of $\hat{f}_{n,\lambda}$ is minimized with respect to λ_j's by [1]

$$\lambda_j^* := \frac{\theta_j(\theta_j + \alpha_j)}{\varepsilon^2 + (\theta_j + \alpha_j)^2}, \quad j = 1, ..., n$$

where $\varepsilon^2 = \sigma_\xi^2/n$ (λ_j^*'s are the weights corresponding to the weighted projection *oracle*).

(2) Check that the corresponding value of the risk is

$$\text{MISE}(\lambda^*) = \sum_{j=1}^{n} \frac{\varepsilon^2 \theta_j^2}{\varepsilon^2 + (\theta_j + \alpha_j)^2} + \rho_n,$$

where $\rho_n = \sum_{j=n+1}^{\infty} \theta_j^2$.

(3) Prove that

$$\sum_{j=1}^{n} \frac{\varepsilon^2 \theta_j^2}{\varepsilon^2 + (\theta_j + \alpha_j)^2} = \Big(1 + o(1)\Big) \sum_{j=1}^{n} \frac{\varepsilon^2 \theta_j^2}{\varepsilon^2 + \theta_j^2}.$$

(4) Prove that

$$\rho_n = \big(1 + o(1)\big) \sum_{j=n+1}^{\infty} \frac{\varepsilon^2 \theta_j^2}{\varepsilon^2 + \theta_j^2}.$$

(5) Deduce from the above results that

$$\text{MISE}(\lambda^*) = \mathcal{A}_n^* \big(1 + o(1)\big), \quad n \to \infty,$$

---

[1] recall that

$$\alpha_j = \frac{1}{n} \sum_{i=1}^{n} f(i/n)\varphi_j(i/n) - \int f \varphi_j.$$

where

$$\mathcal{A}_n^* = \sum_{j=1}^{\infty} \frac{\varepsilon^2 \theta_j^2}{\varepsilon^2 + \theta_j^2}.$$

(6) Check that[2]

$$\mathcal{A}_n^* < \min_{N \ge 1} \mathcal{A}_{n,N}.$$

<u>Solution</u>

1. Recall that for $1 \le j \le n-1$ (an in fact for $j = n$ as well, check!)

$$\mathsf{E}\hat{\theta}_j = \theta_j + \alpha_j, \quad \mathsf{E}\big(\theta_j - \hat{\theta}_j\big)^2 = \sigma_\xi^2/n + \alpha_j^2 = \varepsilon^2 + \alpha_j^2.$$

Hence

$$\begin{aligned} \text{MISE} =& \mathsf{E}\int \Big(\sum_{j=1}^{n} \lambda_j \hat{\theta}_j \varphi_j(x) - \sum_{j=1}^{\infty} \theta_j \varphi_j(x)\Big)^2 dx = \\ & \mathsf{E}\int \Big(\sum_{j=1}^{n} (\lambda_j \hat{\theta}_j - \theta_j)\varphi_j(x) - \sum_{j=n+1}^{\infty} \theta_j \varphi_j(x)\Big)^2 dx = \\ & \sum_{j=1}^{n} \mathsf{E}\big(\lambda_j \hat{\theta}_j - \theta_j\big)^2 + \sum_{j=n+1}^{\infty} \theta_j^2. \end{aligned}$$

Since

$$\begin{aligned} & \mathsf{E}\big(\lambda_j \hat{\theta}_j - \theta_j\big)^2 = \mathsf{E}\big(\lambda_j(\hat{\theta}_j - \theta_j) + (\lambda_j - 1)\theta_j\big)^2 = \\ & \lambda_j^2 \mathsf{E}(\hat{\theta}_j - \theta_j)^2 + 2\lambda_j \mathsf{E}(\hat{\theta}_j - \theta_j)(\lambda_j - 1)\theta_j + (\lambda_j - 1)^2\theta_j^2 = \\ & \lambda_j^2\big(\varepsilon^2 + \alpha_j^2\big) + 2\lambda_j \alpha_j (\lambda_j - 1)\theta_j + (\lambda_j - 1)^2 \theta_j^2 = \\ & \lambda_j^2 \varepsilon^2 + \Big(\lambda_j \alpha_j + (\lambda_j - 1)\theta_j\Big)^2, \end{aligned}$$

we obtain

$$\text{MISE}(\lambda) = \sum_{j=1}^{n} \bigg(\lambda_j^2 \varepsilon^2 + \Big(\lambda_j \alpha_j + (\lambda_j - 1)\theta_j\Big)^2\bigg) + \rho_n.$$

The summands in the first term are parabolas, minimized by

$$\lambda_j^* = \frac{\theta_j(\alpha_j + \theta_j)}{\varepsilon^2 + (\alpha_j + \theta_j)^2}.$$

2. Direct calculation shows that the corresponding MISE is given by

$$\text{MISE}(\lambda^*) = \sum_{j=1}^{n} \frac{\varepsilon^2 \theta_j^2}{\varepsilon^2 + (\alpha_j + \theta_j)^2} + \rho_n$$

[2] $\mathcal{A}_{n,N} = \frac{\sigma_\xi^2 N}{n} + \rho_N$ is the leading term (for $\beta > 1$ in the upper bound for the MISE of the simple projection estimator we studied in class.)

3. We have

$$\frac{\varepsilon^2\theta_j^2}{\varepsilon^2+(\alpha_j+\theta_j)^2} = \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2}\frac{\varepsilon^2+\theta_j^2}{\varepsilon^2+(\alpha_j+\theta_j)^2} =$$

$$\frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2}\left(1+\frac{\theta_j^2-(\alpha_j+\theta_j)^2}{\varepsilon^2+(\alpha_j+\theta_j)^2}\right) = \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2}\left(1-\frac{\alpha_j(\alpha_j+2\theta_j)}{\varepsilon^2+(\alpha_j+\theta_j)^2}\right).$$

Further,

$$\left|\frac{\alpha_j(\alpha_j+2\theta_j)}{\varepsilon^2+(\alpha_j+\theta_j)^2}\right| = \left|\frac{2\alpha_j(\alpha_j+\theta_j)}{\varepsilon^2+(\alpha_j+\theta_j)^2} - \frac{\alpha_j^2}{\varepsilon^2+(\alpha_j+\theta_j)^2}\right| \le |\alpha_j/\varepsilon| + |\alpha_j/\varepsilon|^2$$

where we used the elementary inequality $\frac{|x|}{\varepsilon^2+x^2} \le \frac{1}{2}\frac{1}{\varepsilon}$, $x \in \mathbb{R}$. Recall that for $\theta \in \Theta(\beta, Q)$ with $\beta > 1/2$

$$\max_{1\le j\le n-1} |\alpha_j| \le C_{\beta,Q} n^{-\beta+1/2},$$

and that $\varepsilon = \sigma_\xi^2/n$. Hence for $\beta \ge 2$,

$$|\alpha_j/\varepsilon| + |\alpha_j/\varepsilon|^2 \le C_1 n^{-\beta+1},$$

for all n, large enough. Consequently,

$$\left|\sum_{j=1}^n \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+(\alpha_j+\theta_j)^2} - \sum_{j=1}^n \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2}\right| \le C_1 n^{-\beta+1}\sum_{j=1}^n \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2},$$

which verifies the claim for $\beta > 1$.

4. Note that

$$\theta_j^2 = \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2} + \theta_j^2\frac{\theta_j^2}{\varepsilon^2+\theta_j^2}.$$

Since $\sum_{j=1}^\infty \theta_j^2 a_j^2 \le Q < \infty$, the sequence $r_j := \theta_j^2 a_j^2$ converges to zero as $j \to \infty$. Hence

$$\sum_{j=n+1}^\infty \theta_j^2\frac{\theta_j^2}{\varepsilon^2+\theta_j^2} = \frac{1}{\varepsilon^2}\sum_{j=n+1}^\infty \frac{r_j}{a_j^2}\frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2} \le$$

$$\frac{1}{\varepsilon^2}\frac{\sup_{j\ge n+1} r_j}{a_{n+1}^2}\sum_{j=n+1}^\infty \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2} \le C\frac{1}{n^{-1}}\frac{\sup_{j\ge n+1} r_j}{(n+1)^{2\beta}}\sum_{j=n+1}^\infty \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2}.$$

Hence for $\beta \ge 1$,

$$\left|\rho_n - \sum_{j=n+1}^\infty \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2}\right| = \left|\sum_{j=n+1}^\infty \theta_j^2 - \sum_{j=n+1}^\infty \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2}\right| =$$

$$\sum_{j=n+1}^\infty \theta_j^2\frac{\theta_j^2}{\varepsilon^2+\theta_j^2} \le o(1)\sum_{j=n+1}^\infty \frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2},$$

as claimed.

5. Obviously follows from (3) and (4).

6. Using the identity $y^2 = \frac{y^2x^2}{y^2+x^2} + \frac{y^4}{y^2+x^2}$, we get

$$\min_{N\geq 1} \mathcal{A}_{n,N} = \min_{N\geq 1}\Big(\frac{\sigma_\xi^2 N}{n} + \rho_N\Big) = \min_{N\geq 1}\left(\sum_{j=1}^{N}\varepsilon^2 + \sum_{j=N+1}^{\infty}\theta_j^2\right) =$$

$$\min_{N\geq 1}\left(\sum_{j=1}^{N}\Big(\frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2} + \frac{\varepsilon^4}{\varepsilon^2+\theta_j^2}\Big) + \sum_{j=N+1}^{\infty}\Big(\frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2} + \frac{\theta_j^4}{\varepsilon^2+\theta_j^2}\Big)\right) =$$

$$\sum_{j=1}^{\infty}\frac{\varepsilon^2\theta_j^2}{\varepsilon^2+\theta_j^2} + \min_{N\geq 1}\left(\sum_{j=1}^{N}\frac{\varepsilon^4}{\varepsilon^2+\theta_j^2} + \sum_{j=N+1}^{\infty}\frac{\theta_j^4}{\varepsilon^2+\theta_j^2}\right) > \mathcal{A}_n^*.$$

Exerecise 1.11. Consider the nonparametric regression model under the Assumption (A). The smoothing spline estimator $\hat{f}_n^{sp}(x)$ is defined as a solution of the following minimization problem

$$\hat{f}_n^{sp} = \operatorname{argmin}_{f\in W}\left(\frac{1}{n}\sum_{i=1}^{n}\big(Y_i - f(X_i)\big)^2 + \kappa\int_0^1 \big(f''(x)\big)^2 dx\right),$$

where κ is the smoothing parameter and W is one of the sets of functions defined below.

(1) First suppose that W is the set of all the functions : $[0,1]\mapsto\mathbb{R}$ such that f' is absolutely continuous. Prove that the estimator $\hat{f}_n^{sp}$ reproduces polynomials of degree ≤ 1 if $n\geq 2$

(2) Suppose next that W is the set of all the functions $f:[0,1]\mapsto\mathbb{R}$ such that (i) f' is absolutely continuous and (ii) the periodicity condition is satisfied: $f(0)=f(1)$, $f'(0)=f'(1)$. Prove that the minimization problem is equivalent to

$$\min_{(b_j)}\sum_{j=1}^{\infty}\Big(-2\hat{\theta}_j b_j + b_j^2(\kappa\pi^4 a_j^2+1)\big(1+O(n^{-1})\big)\Big),$$

where b_j are the Fourier coefficients of f, the term $O(n^{-1})$ is uniform in (b_j) and a_j are defined as for the Sobolev ellipsoid.

(3) Assume now that the term $O(n^{-1})$ in the latter minimization problem is negligible. Formally replacing it by zero, find the solution and conclude that the periodic spline estimator is approximately equal to a weighted projection estimator

$$\hat{f}_n^{sp}(x)\approx\sum_{j=1}^{\infty}\lambda_j^*\hat{\theta}_j\varphi_j(x),$$

with the weights λ_j^* written explicitly.

(4) Use (3) to show that for sufficiently small κ the spline estimator $\hat{f}_n^{sp}$ is approximated by the kernel estimator

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^{n} Y_i K\Big(\frac{X_i - x}{h}\Big),$$

where $h = \kappa^{1/4}$ and K is the Silverman kernel.

Solution

1. Let $\mathcal{Q}$ be the space of linear polynomials, then

$$\min_{f\in W}\left(\frac{1}{n}\sum_{i=1}^{n}\big(Y_i - f(X_i)\big)^2 + \kappa\int_0^1 \big(f''(x)\big)^2 dx\right) \le$$
$$\min_{f\in \mathcal{Q}}\left(\frac{1}{n}\sum_{i=1}^{n}\big(Y_i - f(X_i)\big)^2 + \kappa\int_0^1 \big(f''(x)\big)^2 dx\right) =$$
$$\min_{a,b\in\mathbb{R}}\frac{1}{n}\sum_{i=1}^{n}\big(Y_i - b - aX_i\big)^2.$$

If $Y_i = b' + a'X_i$ and $n \ge 2$, the latter expression vanishes and the minimizing constants are exactly a' and b', i.e. $\hat{f}_n^{sp}(x) = b' + a'x$ as claimed.

2. In this case, $f(x) = \sum_{j=1}^{\infty} b_j\varphi_j(x)$, where (φ_j) is the trigonometric Fourier basis. Note that for even j

$$\frac{d^2}{dx^2}\varphi_j(x) = \frac{d^2}{dx^2}\cos\Big(2\pi\frac{j}{2}x\Big) = -\pi^2 j^2\varphi_j(x),$$

and for odd j

$$\frac{d^2}{dx^2}\varphi_j(x) = \frac{d^2}{dx^2}\sin\Big(2\pi\frac{j-1}{2}x\Big) = -\pi^2(j-1)^2\varphi_j(x),$$

i.e. for any $j \ge 0$

$$\frac{d^2}{dx^2}\varphi_j(x) = -\pi^2 a_j\varphi_j(x),$$

where a_j's are as in the definition of the Sobolev ellipsoid, corresponding to $\beta = 2$.

Hence, if we assume that f'' is square integrable[3], by the Parseval identity

$$\int_0^1 (f''(x))^2 dx = \sum_{j=1}^{\infty}\big(\pi^2 a_j b_j\big)^2.$$

Further,

$$\frac{1}{n}\sum_{i=1}^{n}\big(Y_i - f(X_i)\big)^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \frac{2}{n}\sum_{i=1}^{n} Y_i f(X_i) + \frac{1}{n}\sum_{i=1}^{n} f^2(X_i).$$

For the design $X_i = i/n$,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i f(X_i) = \frac{1}{n}\sum_{i=1}^{n} Y_i\sum_{j=1}^{\infty} b_j\varphi_j(i/n) = \sum_{j=1}^{\infty} b_j\hat{\theta}_j.$$

[3] if f'' is not square integrable, then the original minimization problem is ill defined

Also

$$\frac{1}{n}\sum_{i=1}^{n} f^2(X_i) - \int_0^1 f^2 = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} b_j b_k \frac{1}{n}\sum_{i=1}^{n} \varphi_j(i/n)\varphi_k(i/n) - \sum_{j=1}^{\infty} b_j^2 =$$
$$\sum_{j=n}^{\infty}\sum_{k=n}^{\infty} b_j b_k \frac{1}{n}\sum_{i=1}^{n} \varphi_j(i/n)\varphi_k(i/n) - \sum_{j=n}^{\infty} b_j^2,$$

and applying the Cauchy Schwarz inequality, we get

$$\left|\frac{1}{n}\sum_{i=1}^{n} f^2(X_i) - \int_0^1 f^2\right| \le 2\frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j=n}^{\infty}|b_j|\Big)^2 + \sum_{j=n}^{\infty} b_j^2 \le$$
$$2\sum_{j=1}^{\infty} b_j^2 a_j^2 \sum_{j=n}^{\infty}\frac{1}{a_j^2} + \frac{1}{a_n^2}\sum_{j=1}^{\infty} b_j^2 a_j^2 = \sum_{j=1}^{\infty} b_j^2 a_j^2 \left(2\sum_{j=n}^{\infty}\frac{1}{a_j^2} + \frac{1}{a_n^2}\right) \le Cn^{-3}\sum_{j=1}^{\infty} b_j^2 a_j^2,$$

where C is an absolute constant.

Assembling all parts together, we get

$$\frac{1}{n}\sum_{i=1}^{n}\big(Y_i - f(X_i)\big)^2 + \kappa\int_0^1 \big(f''(x)\big)^2 dx =$$
$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - 2\sum_{j=1}^{\infty} b_j\hat{\theta}_j + \sum_{j=1}^{\infty} b_j^2 + O(n^{-3})\sum_{j=1}^{\infty} b_j^2 a_j^2 + \kappa\pi^4\sum_{j=1}^{\infty} b_j^2 a_j^2,$$

which verifies the claim.

3. Note that the expression $-2\hat{\theta}_j b_j + b_j^2(\kappa\pi^4 a_j^2 + 1)$ is minimized by

$$b_j^* = \frac{\hat{\theta}_j}{\kappa\pi^4 a_j^2 + 1}$$

and hence the approximate equality holds with

$$\lambda_j^* = \frac{1}{\kappa\pi^4 a_j^2 + 1}.$$

4. Suppose that $h := \frac{1}{2}\kappa^{1/4}$ is small enough to justify the approximation:

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right) = \frac{1}{nh}\sum_{i=1}^n Y_i \int_{\mathbb{R}} \frac{\cos\left(2\pi t \frac{X_i - x}{h}\right)}{1+(2\pi t)^4} dt \approx$$

$$\frac{1}{nh}\sum_{i=1}^n Y_i \sum_{j=-\infty, j\neq 0}^{\infty} \frac{\cos\left(2\pi h j \frac{X_i - x}{h}\right)}{1+(2\pi h j)^4} h = \frac{1}{n}\sum_{i=1}^n Y_i \sum_{j=-\infty, j\neq 0}^{\infty} \frac{\cos\left(2\pi j (X_i - x)\right)}{1+(2\pi h j)^4} =$$

$$\frac{1}{n}\sum_{i=1}^n Y_i \sum_{j=-\infty, j\neq 0}^{\infty} \frac{\cos(2\pi j X_i)\cos(2\pi j x) + \sin(2\pi j X_i)\sin(2\pi j x)}{1+(2\pi h j)^4} =$$

$$\frac{1}{n}\sum_{i=1}^n Y_i \sum_{j=1}^{\infty} \frac{2\cos(2\pi j X_i)\cos(2\pi j x) + 2\sin(2\pi j X_i)\sin(2\pi j x)}{1+(2\pi h j)^4} =$$

$$\frac{1}{n}\sum_{i=1}^n Y_i \sum_{j=1}^{\infty} \frac{\varphi_{2j}(X_i)\varphi_{2j}(x) + \varphi_{2j+1}(X_i)\varphi_{2j+1}(x)}{1+(2\pi h j)^4} =$$

$$\sum_{j=1}^{\infty} \frac{\hat\theta_{2j}\varphi_{2j}(x) + \hat\theta_{2j+1}\varphi_{2j+1}(x)}{1+(2\pi h j)^4} = \sum_{j=1}^{\infty} \frac{\hat\theta_j \varphi_j(x)}{1+\pi^4 \kappa a_j^2} = \hat f_n^{sp}(x).$$

2. Lower bounds on the minimax risk

Exerecise 2.1. Give an example of measures P_0 and P_1 such that $p_{e,1}$ is arbitrarily close to 1.

Hint: consider two discrete measures on $\{0,1\}$.

Solution

Let $X \sim \mathrm{Ber}(p)$ under P_0 and $X \sim \mathrm{Ber}(q)$ under P_1. Since $X \in \{0,1\}$, the only possible (non-randomized) tests are $\psi'(X) = X$ and $\psi''(X) = 1 - X$. For the test $\psi'(X) = X$, we have

$$\mathsf{P}_0(\psi'(X) = 1) = \mathsf{P}_0(X=1) = p, \quad \mathsf{P}_1(\psi'(X)=0) = 1-q$$

and

$$\mathsf{P}_0(\psi''(X) = 1) = \mathsf{P}_0(X=0) = 1-p, \quad \mathsf{P}_1(\psi''(X)=0) = \mathsf{P}_1(X=1) = q.$$

Hence

$$p_{e,1} = \inf_\psi \Big(\mathsf{P}_0(\psi=1) \vee \mathsf{P}_1(\psi=0)\Big) = \min\Big(p \vee (1-q), (1-p)\vee q\Big),$$

which approaches 1 as e.g. both p and q approach 1.

Exerecise 2.2. Let P and Q be two probability measures with densities p and q w.r.t. the Lebesgue measure on $[0,1]$ such that

$$0 < c_1 \le p(x), q(x) < c_2 < \infty, \quad \forall x \in [0,1].$$

Show that the Kullback divergence $K(P,Q)$ is equivalent to the squared L_2 distance between the two densities, i.e.

$$k_1 \int (p-q)^2 \le K(\mathsf{P},\mathsf{Q}) \le k_2 \int (p-q)^2$$

where k_1, k_2 are constants (independent of p and q). The same true for the χ^2 divergence.

Solution

Since both densities are bounded,

$$0 < c_1/c_2 \le q(x)/p(x) \le c_2/c_1 < \infty.$$

Since $1 \in [c_1/c_2, c_2/c_1]$, by the Taylor theorem

$$\log x = x - 1 - \frac{1}{2}\frac{1}{\xi^2}(x-1)^2, \quad x \in [c_1/c_2, c_2/c_1]$$

where ξ is a point, possibly dependent on x, such that $|\xi - 1| \le |x-1|$. Hence for all $x \in [c_1/c_2, c_2/c_1]$

$$x - 1 - \frac{1}{2}(c_2/c_1)^2(x-1)^2 \le \log x \le x - 1 - \frac{1}{2}(c_1/c_2)^2(x-1)^2.$$

Since $p(x) \ge c_1$ and $q(x) \ge c_1$, $\mathsf{P} \sim \mathsf{Q}$ and

$$K(\mathsf{P},\mathsf{Q}) = -\int p \log\frac{q}{p} \le -\int p\Big(\frac{q}{p} - 1 - \frac{1}{2}(c_2/c_1)^2(q/p-1)^2\Big) = \frac{1}{2}(c_2/c_1)^2 \int p(q/p-1)^2 \le \frac{1}{2}(c_2/c_1)^2\frac{1}{c_1}\int (q-p)^2.$$

Similarly, the lower bound is obtained.

Exerecise 2.3. Prove that if the probability measures P and Q are mutually absolutely continuous (i.e. equivalent), then

$$K(\mathsf{P},\mathsf{Q}) \le \chi^2(\mathsf{Q},\mathsf{P})/2.$$

Solution

The claim is apparently false, as the following counterexample shows. Let P and Q be $\mathrm{Ber}(p)$ and $\mathrm{Ber}(q)$ distributions. Then

$$K(\mathsf{P},\mathsf{Q}) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q},$$

and

$$\frac{1}{2}\chi^2(\mathsf{Q},\mathsf{P}) = \frac{1}{2}p\Big(\frac{q}{p}-1\Big)^2 + \frac{1}{2}(1-p)\Big(\frac{1-q}{1-p}-1\Big)^2.$$

For $p = 0.5$ and $q = 0.1$, the numerical calculation yields

$$K(\mathsf{P},\mathsf{Q}) = 0.5108...$$

and

$$\frac{1}{2}\chi^2(\mathsf{Q},\mathsf{P}) = 0.3200...$$

EXERECISE 2.4. Consider the nonparametric regression model

$$Y_i = f(i/n) + \xi_i, \quad i = 1, ..., n$$

where f is a function on $[0,1]$ with values in $\mathbb{R}$ and ξ_i are arbitrary random variables. Using the technique of two hypotheses show that

$$\varliminf_{n\to\infty} \inf_{T_n} \sup_{f\in C[0,1]} \mathsf{E}_f \|T_n - f\|_\infty = +\infty,$$

where $C[0,1]$ is the space of al continuous functions on $[0,1]$. In words, no rate of convergence can be attained uniformly on such a large functional class as $C[0,1]$.

Solution

Consider the hypotheses $f_{0n}(x) \equiv 0$ and

$$f_{1n}(x) = n\Big(1 - \big|2nx - 1\big|\Big)\mathbf{1}_{\{x\in[0,1/n]\}}, \quad x \in [0,1].$$

Clearly, $f_{in} \in C[0,1]$, $i = 0, 1$ and

$$d(f_{0n}, f_{1n}) = \|f_{1n}\|_\infty = n := 2s,$$

with $s := A\psi_n$, $\psi_n = n$ and $A = 1/2$. Further, since $\mathsf{P}_0 = \mathsf{P}_1$, $K(\mathsf{P}_0, \mathsf{P}_1) = 0$ and hence $p_{e,1} \geq 1/2$. Consequently,

$$\inf_{T_n} \sup_{f\in C[0,1]} \mathsf{E}_f n^{-1}\|T_n - f\|_\infty \geq c$$

with a constant $c > 0$, which verifies the claim.

EXERECISE 2.5. Suppose that Assumptions (B) and (LP2) hold and assume that the random variables ξ_i are Gaussian. Prove (2.38) using Theorem 2.1.

Solution

Take the same hypotheses as in the text, i.e. $f_{0n}(x) \equiv 0$ and $f_{1n}(x) = Lh_n^\beta K\Big(\frac{x-x_0}{h_n}\Big)$ where K is a nonnegative function in $C^\infty(\mathbb{R}) \cap \Sigma(\beta, 1/2)$ satisfying $K(u) > 0$ if and only if $|u| < 1/2$. Clearly, $\mathsf{P}_0 \sim \mathsf{P}_1$ and

$$\frac{d\mathsf{P}_0}{d\mathsf{P}_1} = \prod_{i=1}^n \frac{p_\xi(Y_i)}{p_\xi(Y_i - f_{1n}(X_i))} = \exp\left(-\sum_{i=1}^n Y_i f_{1n}(X_i) + \frac{1}{2}\sum_{i=1}^n f_{1n}^2(X_i)\right) =$$

$$\exp\left(-\sum_{i=1}^n \xi_i f_{1n}(X_i) - \frac{1}{2}\sum_{i=1}^n f_{1n}^2(X_i)\right) =: \exp\left(\sigma_n Z - \frac{1}{2}\sigma_n^2\right),$$

where $Z \sim N(0,1)$ under P_1 and $\sigma_n^2 = \sum_{i=1}^n f_{1n}^2(X_i)$. As in the text (page 94), under the assumption (LP2), $\sigma^2 := \sup_n \sigma_n^2 < \infty$ and hence

$$\mathsf{P}_1\Big(\frac{d\mathsf{P}_0}{d\mathsf{P}_1} \geq 1\Big) = \mathsf{P}_1\Big(Z \geq \frac{1}{2}\sigma_n\Big) \geq \mathsf{P}_1\Big(Z \geq \frac{1}{2}\sigma\Big) = \Phi(-\sigma/2),$$

and consequently,

$$p_{e,1} \geq \sup_{\tau>0}\left\{\frac{\tau}{1+\tau}\mathsf{P}_1\Big(\frac{d\mathsf{P}_0}{d\mathsf{P}_1} \geq \tau\Big)\right\} \geq \frac{1}{2}\mathsf{P}_1\Big(\frac{d\mathsf{P}_0}{d\mathsf{P}_1} \geq 1\Big) \geq \frac{1}{2}\Phi(-\sigma/2) > 0,$$

and the claim follows from the general reduction scheme.

EXERECISE 2.6. Improve the bound of Theorem 2.6 by computing the maximum on the right hand side of (2.48). Do we obtain that $p_{e,M}$ is arbitrarily close to 1 for $M \to \infty$ and $\alpha \to 0$, as in the Kullback case (cf. (2.53)) ?

Solution

The function

$$\Phi(\tau) = \frac{M\tau}{1+M\tau}\Big(1-\tau(\alpha_*+1)\Big)$$

is maximized at

$$\tau^* := \frac{1}{M}\left(\sqrt{1+M/(\alpha_*+1)}-1\right)$$

with

$$\Phi(\tau^*) = \left(1-\sqrt{\frac{\alpha_*+1}{\alpha_*+1+M}}\right)\left(1-\frac{\alpha_*+1}{M}\left(\sqrt{1+\frac{M}{\alpha_*+1}}-1\right)\right).$$

For the choice $\alpha_* := \alpha M$, we get

$$\begin{aligned}
\Phi(\tau^*) &= \left(1-\sqrt{\frac{\alpha M+1}{\alpha M+1+M}}\right)\left(1-\frac{\alpha M+1}{M}\left(\sqrt{1+\frac{M}{\alpha M+1}}-1\right)\right)\\
&\xrightarrow{M\to\infty} \left(1-\sqrt{\frac{\alpha}{\alpha+1}}\right)\left(1-\alpha\left(\sqrt{1+\frac{1}{\alpha}}-1\right)\right) = \\
&\left(1-\sqrt{\frac{\alpha}{\alpha+1}}\right)\left(1-\sqrt{\alpha}\left(\sqrt{\alpha+1}-\sqrt{\alpha}\right)\right) \xrightarrow{\alpha\to 0} 1
\end{aligned}$$

EXERECISE 2.7. Consider the regression model with random design:

$$Y_i = f(X_i)+\xi_i, \quad i=1,...,n$$

where X_i are i.i.d. random variables with density $\mu(\cdot)$ on $[0,1]$ such that $\mu(x) \le \mu_0 < \infty$ for all $x \in [0,1]$, the random variables ξ_i are i.i.d. with density p_ξ on $\mathbb{R}$, and the random vector $(X_1,...,X_n)$ is independent of $(\xi_1,...,\xi_n)$. Let $f \in \Sigma(\beta, L)$, $\beta > 0$, L> 0 and let $x_0 \in [0,1]$ be a fixed point.

(1) Suppose first that p_ξ satisfies

$$\int \left(\sqrt{p_\xi(y)}-\sqrt{p_\xi(y+t)}\right)^2 dy \le p_* t^2, \quad t \in \mathbb{R},$$

with a positive constant p_*. Prove the bound

$$\varliminf_{n\to\infty} \inf_{\hat f_n} \sup_{f\in\Sigma(\beta,L)} \mathsf{E}_f n^{\frac{2\beta}{2\beta+1}}\left(\hat f_n(x_0)-f(x_0)\right)^2 \ge c,$$

where $c > 0$ depends only on β, L, μ_0, p_*.

(2) Suppose now that the variables ξ_i are i.i.d. and uniformly distributed on $[-1,1]$. Prove the bound

$$\varliminf_{n\to\infty} \inf_{\hat f_n} \sup_{f\in\Sigma(\beta,L)} \mathsf{E}_f n^{\frac{2\beta}{\beta+1}} \left(\hat f_n(x_0) - f(x_0)\right)^2 \ge c',$$

where $c'>0$ depends only on β, L, μ_0. Note that the rate here is $n^{-\frac{\beta}{\beta+1}}$, which is faster than the usual rate $n^{-\frac{\beta}{2\beta+1}}$. Furthermore, it can be proved that $\psi_n = n^{-\frac{\beta}{\beta+1}}$ is the optimal rate of convergence in the model with uniformly distributed errors.

Solution

(1) We shall use the following property of the Hellinger distance. Define the convolution (the subscript ξ in p_ξ is omitted for brevity)

$$(p*\mu)(x) = \int_0^1 p\big(x - f(y)\big)\mu(y)dy, \quad x\in\mathbb{R}.$$

By the Jensen inequality

$$\sqrt{(p*\mu)(x)} = \sqrt{\int_0^1 p\big(x-f(y)\big)\mu(y)dy} \ge \int_0^1 \sqrt{p\big(x-f(y)\big)}\mu(y)dy,$$

and hence

$$\begin{aligned}
&H^2(p, p*\mu) = \int \left(\sqrt{p(x)} - \sqrt{(p*\mu)(x)}\right)^2 dx = \\
&2 - 2\int \sqrt{p(x)}\sqrt{(p*\mu)(x)}dx \le 2 - 2\int \sqrt{p(x)} \int_0^1 \sqrt{p\big(x-f(y)\big)}\mu(y)dydx = \\
&\int_0^1 \left(2 - 2\int \sqrt{p(x)}\sqrt{p(x-f(y))}dx\right)\mu(y)dy = \\
&\int_0^1 \int \left(\sqrt{p(x)} - \sqrt{p(x-f(y))}\right)^2 dx\mu(y)dy \le p_* \int_0^1 f^2(y)\mu(y)dy.
\end{aligned}$$

To prove the required bound, we shall use the two hypotheses as in the text (see eq. (2.32)-(2.33))

$$f_{0n}(x)\equiv 0, \quad \text{and} \quad f_{1n}(x) = Lh_n^\beta K\left(\frac{x-x_0}{h_n}\right), \quad x\in[0,1].$$

It is left to show that $p_{e,1}\ge c>0$ with a constant c independent of n. To this end, we shall use the bound

$$H^2(\mathsf{P}_0,\mathsf{P}_1)\le \alpha < 2 \quad \Longrightarrow \quad p_{e,1} \ge \frac{1}{2}\left(1 - \sqrt{\alpha(1-\alpha/4)}\right).$$

By the independence, P_0 corresponds to the density

$$\prod_{i=1}^n p(u_i), \quad u\in\mathbb{R}^n$$

and P_1 to the density

$$\prod_{i=1}^n (p*\mu)(u_i), \quad u\in\mathbb{R}^n.$$

Since

$$H^2(\mathsf{P}_0, \mathsf{P}_1) = 2\left(1 - \left(1 - \frac{H^2(p, p*\mu)}{2}\right)^n\right),$$

it is left to show that $\inf_n \left(1 - \frac{1}{2}H^2(p, p*\mu)\right)^n > 0$ or equivalently

$$\sup_n nH^2(p, p*\mu) < \infty.$$

For f_{1n} as above, we have (recall that $h_n = c_0 n^{-\frac{1}{2\beta+1}}$)

$$\begin{aligned}&nH^2(p, p*\mu) \le np_* \int_0^1 f_{1n}^2(y)\mu(y)dy = L^2 p_* n h_n^{2\beta} \int_0^1 K^2\left(\frac{y-x_0}{h_n}\right)\mu(y)dy \le \\ &L^2 p_* \|K\|_\infty^2 n h_n^{2\beta} \int_0^1 \mathbf{1}_{\{|y-x_0|\le h_n/2\}}\mu(y)dy \le L^2 p_* \|K\|_\infty^2 \mu_0 n h_n^{2\beta+1} = \\ &L^2 p_* \|K\|_\infty^2 \mu_0 c_0^{2\beta+1},\end{aligned} \tag{2.1}$$

as required.

(2) For the uniform density $p(x) = \frac{1}{2}\mathbf{1}_{\{|x|\le 1\}}$, we have

$$\int \left(\sqrt{p(y)} - \sqrt{p(y+t)}\right)^2 dy = \frac{1}{2}\int \left(\mathbf{1}_{\{|y|\le 1\}} - \mathbf{1}_{\{|y+t|\le 1\}}\right)^2 dy = \min(2, |t|) \le |t|,$$

and as in (1)

$$H^2(p, p*\mu) \le \int_0^1 |f(y)|\mu(y)dy.$$

Choosing $h_n = c_0 n^{-\frac{1}{\beta+1}}$ and proceeding as in (1), we get

$$\begin{aligned}&nH^2(p, p*\mu) \le np_* \int_0^1 f_{1n}(y)\mu(y)dy = Lnh_n^\beta \int_0^1 K\left(\frac{y-x_0}{h_n}\right)\mu(y)dy \le \\ &L\|K\|_\infty \mu_0 n h_n^{\beta+1} \le L\|K\|_\infty \mu_0 c_0^{\beta+1},\end{aligned}$$

which yields the claimed result.

Exerecise 2.8. Let $X_1, ..., X_n$ be i.i.d. random variables on $\mathbb{R}$ having density $p \in \mathcal{P}(\beta, L)$, $\beta > 0$, $L > 0$. Show that

$$\varliminf_{n\to\infty} \inf_{\hat{f}_n} \sup_{p\in\mathcal{P}(\beta,L)} \mathsf{E}_p n^{\frac{2\beta}{2\beta+1}} \Big(\hat{p}_n(x_0) - p(x_0)\Big)^2 \ge c,$$

for any $x_0 \in \mathbb{R}$, where $c > 0$ depends only on β and L.

Solution

Consider the hypotheses

$$p_0(x) = \varphi_\sigma(x), \quad p_{1n}(x) = \varphi_\sigma(x) + g_n(x),$$

where φ_σ is the $N(0, \sigma^2)$ density with some $\sigma > 0$ to be chosen later,

$$g_n(x) := Lh_n^\beta R\Big(\frac{x-x_0}{h_n}\Big)$$

and R is a bounded $C_\infty(\mathbb{R})\cap\Sigma_{\mathbb{R}}(\beta,1/2)$ function, supported on an interval $\Delta\subset\mathbb{R}$, $\int R(x)dx=0$ and $R(0)>0$. For example, one can take $R(x)=K(x)-K(x-1)$, where K is defined in (2.34) of [**1**].

Note that since R is bounded and $\int R(x)dx=0$, $p_{1n}(x)$ is a probability density for all n large enough. Following the reduction scheme presented in text, we shall check the conditions

(i) $p_{1n}\in\Sigma(\beta,L)$

(ii) $|p_0(x_0)-p_{1n}(x_0)|\ge 2s=2A\psi_n$, where $\psi_n=n^{-\frac{\beta}{2\beta+1}}$ and $A>0$

(iii) $K(\mathsf{P}_1,\mathsf{P}_0)\le\alpha<\infty$

The condition (i) obviously holds for σ large enough, since φ and its derivatives are bounded (similarly to eq. (2.35) in [**1**]). Further,

$$|p_0(x_0)-p_{1n}(x_0)|=|p_{1n}(x_0)|=Lh_n^{2\beta}|R(0)|=2A\psi_n,$$

i.e. (ii) holds with $A=L|R(0)|/2>0$.

Finally, since X_i's are i.i.d. both under P_0 and P_1, using the elementary inequality $\log(a+x)\le\log a+x/a$, we get

$$K(\mathsf{P}_1,\mathsf{P}_0)=nK(p_{1n},p_0)=n\int\Big(\varphi_\sigma(x)+g_n(x)\Big)\log\frac{\varphi_\sigma(x)+g_n(x)}{\varphi_\sigma(x)}dx=$$
$$\le n\int\Big(\varphi_\sigma(x)+g_n(x)\Big)\frac{g_n(x)}{\varphi_\sigma(x)}dx=n\int_0^1\frac{g_n^2(x)}{\varphi_\sigma(x)}dx\le$$
$$\frac{1}{\inf_{x\in\Delta}\psi_\sigma(x)}nL^2h_n^{2\beta+1}\|R\|_2^2=:\alpha,$$

which completes the proof.

Exerecise 2.9. Suppose that Assumptions (B) and (LP2) hold and let $x_0\in[0,1]$. Prove the bound (Stone, 1980)

$$\lim_{a\to0}\varliminf_{n\to\infty}\inf_{T_n}\sup_{f\in\Sigma(\beta,L)}\mathsf{P}_f\Big(n^{\frac{\beta}{2\beta+1}}\big|T_n(x_0)-f(x_0)\big|\ge a\Big)=1.$$

Hint: introduce the hypotheses

$$f_{0n}(x)\equiv0,\quad f_{jn}(x)=\theta_jLh_n^\beta K\left(\frac{x-x_0}{h_n}\right),$$

with $\theta_j=j/M$, $j=1,...,M$.

Solution

Recall that

$$\inf_{T_n}\sup_{f\in\Sigma(\beta,L)}\mathsf{P}_f\Big(\big|T_n(x_0)-f(x_0)\big|\ge an^{-\frac{\beta}{2\beta+1}}\Big)\ge p_{e,M}\ge$$
$$\frac{\sqrt{M}}{1+\sqrt{M}}\Big(1-2\alpha-\sqrt{\frac{2\alpha}{\log M}}\Big),\quad(2.2)$$

whenever

$$d(f_{jn}, f_{kn}) = |f_{jn}(x_0) - f_{kn}(x_0)| \ge 2A\psi_n, \quad j \ne k \tag{2.3}$$

with $A = a/2$ and $\psi_n = n^{-\frac{\beta}{2\beta+1}}$ and

$$\frac{1}{M}\sum_{j=1}^{M} K(\mathsf{P}_j, \mathsf{P}_0) \le \alpha \log M, \tag{2.4}$$

for some $\alpha > 0$.

For the hypotheses appearing in the hint with $h_n = c_0 n^{-\frac{1}{2\beta+1}}$,

$$|f_{jn}(x_0) - f_{kn}(x_0)| = \frac{|j-k|}{M} L h_n^{\beta} K(0) \ge \frac{1}{M} L K(0) c_0^{\beta} \psi_n,$$

and hence (2.3) holds, if we choose

$$c_0 := \left(\frac{aM}{LK(0)}\right)^{1/\beta}.$$

Further,

$$\begin{aligned}
&K(\mathsf{P}_j, \mathsf{P}_0) = \sum_{i=1}^{n} \int p_\xi(u) \log \frac{p_\xi(u)}{p_\xi(u - f_{jn}(X_i))} du \le p_* \sum_{i=1}^{n} f_{jn}^2(X_i) \le \\
&p_*(j/M)^2 L^2 h_n^{2\beta} \|K\|_\infty^2 \sum_{i=1}^{n} \mathbf{1}_{\{|X_i - x_0| \le h_n/2\}} \le \\
&p_*(j/M)^2 L^2 h_n^{2\beta} \|K\|_\infty^2 n a_0 \max(h_n, 1/n) \le p_* \|K\|_\infty^2 a_0 L^2 (j/M)^2 c_0^{2\beta+1} = \\
&p_* \|K\|_\infty^2 a_0 L^2 (j/M)^2 \left(\frac{aM}{LK(0)}\right)^{2+1/\beta} =: C j^2 M^{1/\beta} a^{2+1/\beta}
\end{aligned}$$

and hence

$$\frac{1}{M}\sum_{j=1}^{M} K(\mathsf{P}_j, \mathsf{P}_0) \le C M^{1/\beta - 1} a^{2+1/\beta} \sum_{j=1}^{M} j^2 \le C'(aM)^{2+1/\beta}.$$

Now choose $\alpha := \frac{1}{\log \frac{1}{a}}$ and $M := 1/a$, then for any $a \in (0,1)$,

$$\varliminf_{n\to\infty} \inf_{T_n} \sup_{f \in \Sigma(\beta, L)} \mathsf{P}_f\left(\left|T_n(x_0) - f(x_0)\right| \ge a n^{-\frac{\beta}{2\beta+1}}\right) \ge$$

$$\frac{1}{1+\sqrt{a}}\left(1 - 2\frac{1}{\log \frac{1}{a}} - \sqrt{\frac{2}{\left(\log \frac{1}{a}\right)^2}}\right) \xrightarrow{a \to 0} 1,$$

as claimed.

Exerecise 2.10. Let $X_1, ..., X_n$ be i.i.d. random variables on $\mathbb{R}$ with density $p \in \mathcal{P}(\beta, L)$ where $\beta > 0$ and $L > 0$. Prove the bound

$$\varliminf_{n\to\infty} \inf_{T_n} \sup_{p \in \mathcal{P}(\beta, L)} \mathsf{E}_p n^{\frac{2\beta}{2\beta+1}} \|T_n - p\|_2^2 \ge c$$

where $c > 0$ depends only on β and L.

Solution

Using the notations of Section 2.6.1 [**1**], define

$$g_k(x) = Lh_n^{\beta} K'\left(\frac{x - x_k}{h_n}\right), \quad k = 1, ..., m, \quad x \in \mathbb{R}$$

where K' is the derivative of the function, defined in (2.34) of [**1**]. Let

$$p_{jn}(x) = \frac{1}{\sigma}\varphi(x/\sigma) + \sum_{k=1}^{m} \omega_k^{(j)} g_k(x), \quad j = 0, ..., M, \quad x \in \mathbb{R}$$

where φ is the standard Gaussian density, $\sigma > 0$ is a constant to be chosen shortly and $\omega^{(j)}$'s are m-tuples in the Varshamov-Gilbert subset of $\Omega = \{0,1\}^m$ (see Lemma 2.9, [**1**]). Since $\int g_k(x)dx = 0$ and $\sup_{x\in\mathbb{R}} |g_k(x)| \le L\|K'\|_\infty h_n^{\beta}$, all p_{jn}'s are probability densities for sufficiently large n. Following the reduction scheme presented in the text, the claimed bound follows from the conditions (see Theorem 2.7)

(i) $p_{jn} \in \mathcal{P}(\beta, L)$, $j = 0, ..., M$

(ii) $\|p_{jn} - p_{in}\|_2 \ge 2s = 2A\psi_n$, where $\psi_n = n^{-\frac{\beta}{2\beta+1}}$ and $A > 0$

(iii) $\frac{1}{M}\sum_{i=1}^{M} K(\mathsf{P}_j, \mathsf{P}_0) \le \alpha \log M$ for some $\alpha \in (0, 1/8)$.

The condition (i) is obvious for sufficiently large σ, since φ and all its derivatives are bounded (see eq. (2.35) in [**1**]). Further,

$$\|p_{jn} - p_{in}\|_2 = \int \left(\sum_{k=1}^{m} \big(\omega_k^{(j)} - \omega_k^{(i)}\big) g_k(x)\right)^2 dx =$$

$$\|g_1\|_2^2 \sum_{k=1}^{m} \big(\omega_k^{(j)} - \omega_k^{(i)}\big)^2 = L^2 h_n^{2\beta+1} \|K'\|_2^2 \rho(\omega^{(i)}, \omega^{(j)}),$$

where $\rho(\cdot,\cdot)$ is the Hamming distance. By the Varshamov-Gilbert lemma $\rho(\omega^{(i)}, \omega^{(j)}) \ge m/8$ and hence (ii) holds if $m := 1/h_n$, as in the text.

Using the elementary inequality $\log(a+x) \le \log(a) + \frac{x}{a}$ (and setting $\varphi_\sigma(x) := \frac{1}{\sigma}\varphi(x/\sigma)$ for brevity):

$$K(p_{jn}, p_{j0}) = \int p_{jn}(x) \log \frac{p_{jn}(x)}{p_{0n}(x)} dx =$$

$$\int \left(\varphi_\sigma(x) + \sum_{k=1}^m \omega_k^{(j)} g_k(x) \right) \left(\log \left(\varphi_\sigma(x) + \sum_{k=1}^m \omega_k^{(j)} g_k(x) \right) - \log \varphi_\sigma(x) \right) dx \le$$

$$\int \left(\varphi_\sigma(x) + \sum_{k=1}^m \omega_k^{(j)} g_k(x) \right) \frac{\sum_{k=1}^m \omega_k^{(j)} g_k(x)}{\varphi_\sigma(x)} dx = \int_0^1 \frac{\left(\sum_{k=1}^m \omega_k^{(j)} g_k(x) \right)^2}{\varphi_\sigma(x)} dx \le$$

$$\varphi_\sigma^{-1}(1) \|g_1\|_2^2 \sum_{k=1}^m \omega_k^{(j)} \le \varphi_\sigma^{-1}(1) L^2 \|K'\|_2^2 h_n^{2\beta+1} m =: C h_n^{2\beta},$$

where we used $m = 1/h_n$. Note that $n = c_0^{2\beta+1} h_n^{-2\beta-1}$ and hence

$$K(\mathsf{P}_j, \mathsf{P}_0) \le C n h_n^{2\beta} = C' h_n^{-1} = C' m \le \alpha \log M,$$

where we used the V-G inequality $M \ge 2^{m/8}$. The design constant c_0 can be chosen so that $\alpha \in (0, 1/8)$ and hence (iii) holds. This completes the proof.

EXERECISE 2.11. Consider the nonparametric regression model

$$Y_i = f(i/n) + \xi_i, \quad i = 1, ..., n,$$

where the random variables ξ_i are i.i.d. with distribution $N(0,1)$ and where $f \in W^{per}(\beta, L)$, $L > 0$ and $\beta \in \{1, 2, ...\}$. Prove the bound

$$\varliminf_{n\to\infty} \inf_{T_n} \sup_{f \in W^{per}(\beta,L)} \left(\frac{n}{\log n} \right)^{\frac{2\beta-1}{2\beta}} \mathsf{E}_f \|T_n - f\|_\infty^2 \ge c,$$

where $c > 0$ depends only on β and L.

<u>Solution</u>

Consider the hypotheses

$$f_{0n}(x) \equiv 0, \quad f_{jn}(x) = L h_n^{\beta-1/2} K \left(\frac{x - x_j}{M} \right), \quad j = 1, ..., M$$

where $h_n = c_0 \left(\frac{n}{\log n} \right)^{-\frac{1}{2\beta}}$, $M = \lceil 1/h_n \rceil$ and $x_j = \frac{j-1/2}{M}$ and K is the function defined in eq. (2.33) [**1**]. To prove the claimed bound, we shall check

(i) $f_{jn} \in W^{per}(\beta, L)$, $j = 1, ..., M$

(ii) $\|f_{jn} - f_{in}\|_\infty \ge 2s = 2A\psi_n$ with $\psi_n = \left(\frac{n}{\log n} \right)^{-\frac{\beta-1/2}{2\beta}}$

(iii) $\frac{1}{M} \sum_{j=1}^M K(\mathsf{P}_j, \mathsf{P}_0) \le \alpha \log M$ with $\alpha \in (0, 1/8)$

Note that by construction, f_{jn}'s and all their derivatives vanish at $x = 0$ and $x = 1$ and

$$\int_0^1 \left(f_{jn}^{(\beta)}(x)\right)^2 dx = L^2 h_n^{2(\beta-\frac{1}{2})} h_n^{-2\beta} \int_0^1 \left(K^{(\ell)}\left(\frac{x-x_j}{h_n}\right)\right)^2 dx = L^2 \|K^{(\ell)}\|_2^2,$$

and (i) holds, if we adjust K appropriately (i.e. choose a small enough in eq. (2.34), [**1**]).

Next we have

$$\|f_{jn} - f_{in}\|_\infty = L h_n^{\beta-1/2} \sup_{x\in[0,1]} \left| K\left(\frac{x-x_i}{M}\right) - K\left(\frac{x-x_j}{M}\right) \right| = L h_n^{\beta-1/2} K(0) = 2A\psi_n,$$

with $A = \frac{1}{2} L \|K(0)\|_\infty c_0^{\beta-1/2}$, and thus (ii) is satisfied.

Finally, since ξ_i's are i.i.d $N(0,1)$,

$$K(\mathsf{P}_j, \mathsf{P}_0) = \frac{1}{2}\sum_{i=1}^n f_{jn}^2(X_i) = \frac{1}{2} L^2 h_n^{2\beta-1} \sum_{i=1}^n K^2\left(\frac{X_i - x_j}{M}\right) \le$$

$$\frac{1}{2} L^2 h_n^{2\beta-1} \|K\|_\infty^2 \sum_{i=1}^n \mathbf{1}_{\{|X_i - x_j| \le M/2\}} = \frac{1}{2} L^2 h_n^{2\beta-1} \|K\|_\infty^2 \mathrm{Card}\big\{X_i \in \mathrm{supp}(f_{jn})\big\}$$

and hence

$$\frac{1}{M}\sum_{j=1}^M K(\mathsf{P}_j, \mathsf{P}_0) = \frac{1}{2} L^2 h_n^{2\beta-1} \|K\|_\infty^2 \frac{1}{M} \sum_{j=1}^M \mathrm{Card}\big\{X_i \in \mathrm{supp}(f_{jn})\big\} =$$

$$\frac{1}{2} L^2 h_n^{2\beta-1} \|K\|_\infty^2 n/M = \frac{1}{2} L^2 \|K\|_\infty^2 h_n^{2\beta} n = \frac{1}{2} L^2 \|K\|_\infty^2 c_0^{2\beta} \log n.$$

For sufficiently large n,

$$\log M \ge \log h_n^{-1} = \log c_0 + \frac{1}{2\beta}(\log n - \log\log n) \ge \frac{1}{4\beta}\log n$$

and (iii) follows, if c_0 is chosen small enough.

3. Asymptotic efficiency and adaptation

Exerecise 3.1. Consider an exponential ellipsoid

$$\Theta = \left\{ \theta \in \mathbb{R}^\infty : \sum_{j=1}^\infty e^{2\alpha j}\theta_j^2 \le Q \right\}$$

where $\alpha > 0$ and $Q > 0$.

(1) Give an asymptotic expression, as $\varepsilon \to 0$, for the minimax linear risk on Θ.

(2) Prove that the simple projection estimator defined by

$$\hat{\theta}_k = y_k \mathbf{1}_{\{k \le N^*\}}, \quad k = 1, 2, ...$$

with an appropriately chosen integer $N^* = N^*(\varepsilon)$, is an asymptotically minimax linear estimator on the ellipsoid Θ. Therefore it shares this property with the Pinsker estimator for the same ellipsoid.

Solution

1. We shall first find the asymptotic of κ applying Lemma 3.1 [**1**] to $a_j = e^{\alpha j}$:

$$Q = \frac{\varepsilon^2}{\kappa}\sum_{j=1}^{\infty} a_j(1-\kappa a_j)_+ = \frac{\varepsilon^2}{\kappa}\sum_{j=1}^{\infty} e^{\alpha j}(1-\kappa e^{\alpha j})_+ = \frac{\varepsilon^2}{\kappa}\sum_{j=1}^{M} e^{\alpha j}(1-\kappa e^{\alpha j}),$$

where $M = \lfloor \frac{1}{\alpha}\log\frac{1}{\kappa}\rfloor$. Using the geometric series summation formula

$$\sum_{j=1}^{M} a^j = a\frac{a^M-1}{a-1} = \frac{a^{M+1}}{a-1}(1+o(1)), \quad a>1, \quad M\to\infty$$

we get

$$\frac{\varepsilon^2}{\kappa}\sum_{j=1}^{M} e^{\alpha j}(1-\kappa e^{\alpha j}) = \frac{\varepsilon^2}{\kappa}\frac{e^{\alpha(M+1)}}{e^{\alpha}-1}\big(1+o(1)\big) - \varepsilon^2\frac{e^{2\alpha(M+1)}}{e^{2\alpha}-1}\big(1+o(1)\big) =$$
$$\frac{\varepsilon^2}{\kappa^2}\frac{e^{\alpha}}{e^{\alpha}-1}\big(1+o(1)\big) - \frac{\varepsilon^2}{\kappa^2}\frac{e^{2\alpha}}{e^{2\alpha}-1}\big(1+o(1)\big) = \frac{\varepsilon^2}{\kappa^2}\frac{e^{\alpha}}{e^{2\alpha}-1}\big(1+o(1)\big)$$

and consequently

$$\kappa = \frac{\varepsilon}{Q^{1/2}}\frac{e^{\alpha/2}}{(e^{2\alpha}-1)^{1/2}}\big(1+o(1)\big) =: \kappa^*\big(1+o(1)\big)$$

Next we shall apply Lemma 3.2 [**1**] to calculate the optimal risk:

$$\begin{aligned}D^* =&\varepsilon^2\sum_{j=1}^{\infty}\big(1-\kappa a_j\big)_+ = \varepsilon^2\sum_{j=1}^{M}\big(1-\kappa e^{\alpha j}\big) = \\ &\varepsilon^2\left(M-\kappa\frac{e^{\alpha(M+1)}}{e^{\alpha}-1}\big(1+o(1)\big)\right) = \varepsilon^2\left(\frac{1}{\alpha}\log\frac{1}{\kappa} - \frac{e^{\alpha}}{e^{\alpha}-1}\big(1+o(1)\big)\right) = \\ &\frac{1}{\alpha}\varepsilon^2\log\frac{1}{\varepsilon}\big(1+o(1)\big), \quad \varepsilon\to 0\end{aligned}$$

2. The risk of the suggested estimator (i.e. $\lambda_j = \mathbf{1}_{\{j\le N^*\}}$) is given by

$$R(\lambda,\theta) = \sum_{j=1}^{\infty}(1-\lambda_j)^2\theta_j^2 + \varepsilon^2\lambda_j^2 = \varepsilon^2 N^* + \sum_{j=N^*+1}^{\infty}\theta_j^2 =$$
$$\varepsilon^2 N^* + \sum_{j=N^*+1}^{\infty} e^{-2\alpha j}e^{2\alpha j}\theta_j^2 = \varepsilon^2 N^* + e^{-2\alpha N^*}\sum_{j=N^*+1}^{\infty} e^{-2\alpha(j-N^*)}e^{2\alpha j}\theta_j^2 \le$$
$$\varepsilon^2 N^* + e^{-2\alpha N^*}\sum_{j=1}^{\infty} e^{2\alpha j}\theta_j^2 \le \varepsilon^2 N^* + e^{-2\alpha N^*}Q.$$

If we choose $N^*_\varepsilon = \frac{1}{\alpha}\log\frac{1}{\varepsilon}$, we obtain the upper bound

$$R(\lambda,\theta) \le \frac{1}{\alpha}\varepsilon^2\log\frac{1}{\varepsilon} + \varepsilon^2 Q = \frac{1}{\alpha}\varepsilon^2\log\frac{1}{\varepsilon}\big(1+o(1)\big),$$

which coincides with the lower bound obtained in (1) and hence the simple estimate is asymptotically minimax among the linear estimators.

Exerecise 3.2. Suppose that we observe

$$y_j = \theta_j + \xi_j, \quad j = 1, ..., d$$

where the random variables ξ_j are i.i.d. with distribution $N(0,1)$. Consider the estimation of parameter $\theta = (\theta_1, ..., \theta_d)$. Take $\Theta(Q) = \big\{\theta \in \mathbb{R}^d : \|\theta\|^2 \le Qd\big\}$ with some $Q > 0$, where $\|\cdot\|$ denotes the Euclidian norm on $\mathbb{R}^d$. Define the minimax risk

$$R_d^*\big(\Theta(Q)\big) = \inf_{\hat\theta} \sup_{\theta\in\Theta(Q)} \mathsf{E}_\theta \frac{1}{d}\|\hat\theta - \theta\|^2,$$

where E_θ is the expectation with respect to the joint distribution of $(y_1, ..., y_d)$. Prove that

$$\lim_{d\to\infty} R_d^*\big(\Theta(Q)\big) = \frac{Q}{Q+1}.$$

Hint: to obtain the lower bound on the minimax risk, take $0 < \delta < 1$ and apply the scheme of Section 3.3.2 with the prior distribution $N(0, \delta Q)$ on each of the coordinates of θ.

Solution

We shall derive the upper bound in two ways: by means of the James-Stein estimator and by explicit calculation of the risk of the linear minimax estimator.

The upper bound I.

Recall that the James-Stein estimator $\hat\theta_{JS} = \Big(1 - \frac{d-2}{\|y\|^2}\Big) y$ has the risk

$$\frac{1}{d}\mathsf{E}_\theta\|\hat\theta_{JS} - \theta\|^2 = 1 - \frac{1}{d}\mathsf{E}_\theta\frac{(d-2)^2}{\|y\|^2} = 1 - \frac{(d-2)^2}{d^2}\mathsf{E}_\theta\frac{1}{\frac{1}{d}\|y\|^2}.$$

Denote by P the probability induced by the vector ξ, then

$$R_d^*\big(\Theta(Q)\big) \le \sup_{\theta\in\Theta(Q)} \frac{1}{d}\mathsf{E}_\theta\|\hat\theta_{JS} - \theta\|^2 = \sup_{\theta\in\Theta(Q)}\left(1 - \frac{(d-2)^2}{d^2}\mathsf{E}_\theta\frac{1}{\frac{1}{d}\|y\|^2}\right) =$$
$$1 - \frac{(d-2)^2}{d^2}\inf_{\theta\in\Theta(Q)}\mathsf{E}\frac{1}{\frac{1}{d}\|\theta+\xi\|^2}.$$

Calculations similar to those in Lemma 3.7 [**1**] and the dominated convergence theorem imply continuity of the function $h(\theta) := \mathsf{E}\frac{1}{\frac{1}{d}\|\theta+\xi\|^2}$ on $\mathbb{R}^d$ for d large enough. Since $\Theta(Q)$ is compact, the infimum in the latter expression is attained at a point $\theta_d^* \in \Theta(Q)$. Hence

$$\varlimsup_{d\to\infty} R_d^*\big(\Theta(Q)\big) \le 1 - \varliminf_{d\to\infty}\mathsf{E}\frac{1}{\frac{1}{d}\|\theta_d^*+\xi\|^2} \overset{\dagger}{\le}$$
$$1 - \mathsf{E}\frac{1}{\varlimsup_{d\to\infty}\frac{1}{d}\Big(\|\theta_d^*\|^2 + 2\langle\theta_d^*,\xi\rangle + \|\xi\|^2\Big)} \overset{\ddagger}{\le} 1 - \mathsf{E}\frac{1}{Q + 1 + \varlimsup_{d\to\infty}\frac{2}{d}\langle\theta_d^*,\xi\rangle}$$

where the inequality † holds by the Fatou lemma and the bound ‡ holds by the law of large numbers and since $\theta^*_d \in \Theta(Q)$. Finally, note that for each d large enough, $\eta := \langle \theta^*_d, \xi\rangle / \|\theta^*_d\| \sim N(0,1)$ and hence

$$\sum_{d=1}^{\infty} \mathsf{E}\left|\frac{1}{d}\langle \theta^*_d, \xi\rangle\right|^4 \le \sum_{d=1}^{\infty} \frac{Q^2}{d^2}\mathsf{E}|\eta|^4 < \infty.$$

The Borel-Cantelli lemma now implies $\overline{\lim}_{d\to\infty} \frac{2}{d}\langle \theta^*_d, \xi\rangle = 0$ P-a.s. and the upper bound follows:

$$\overline{\lim_{d\to\infty}}\, R^*_d\big(\Theta(Q)\big) \le 1 - \frac{1}{Q+1} = \frac{Q}{Q+1}.$$

The upper bound II.

Consider the linear estimator $\hat{\theta}(\lambda)$ with the weights λ_j, $j = 1, ..., d$ and the corresponding risk

$$R(\lambda, \theta) = \frac{1}{d}\mathsf{E}_\theta \sum_{j=1}^{d} (\lambda_j y_j - \theta_j)^2 = \frac{1}{d}\sum_{j=1}^{d}(\lambda_j - 1)^2\theta_j^2 + \frac{1}{d}\sum_{j=1}^{d}\lambda_j^2.$$

To find the maximal risk over $\Theta(Q)$, note that for $\theta \in \Theta(Q)$,

$$\frac{1}{d}\sum_{j=1}^{d}(\lambda_j - 1)^2\theta_j^2 \le \max_{j\le d}(\lambda_j - 1)^2 \frac{1}{d}\sum_{j=1}^{d}\theta_j^2 \le \max_{j\le d}(\lambda_j - 1)^2 Q,$$

and this bound is attained on $\theta \in \Theta(Q)$ with

$$\theta_j = \begin{cases} \sqrt{dQ} & j = j^* \\ 0 & j \ne j^* \end{cases}$$

where $j^* = \operatorname{argmax}_{j\le d}(\lambda_j - 1)^2$. Hence

$$\sup_{\theta\in\Theta(Q)} R(\lambda, \theta) = \max_{j\le d}(\lambda_j - 1)^2 Q + \frac{1}{d}\sum_{j=1}^{d}\lambda_j^2.$$

The entries of the minimizer of the latter expression over $\lambda \in \mathbb{R}^d$ are confined to the interval $[0, 1]$, since otherwise the risk can be reduced. Further, for any $\lambda \in [0, 1]^d$ the risk can be reduced by decreasing all the entries to be equal to the minimal one. Hence the minimum is attained by $\lambda \in [0, 1]^d$ with constant entries i.e.:

$$\inf_{\lambda\in\mathbb{R}^d}\sup_{\theta\in\Theta(Q)} R(\lambda, \theta) = \inf_{t\in[0,1]}\Big((t-1)^2 Q + t^2\Big) = \frac{Q}{Q+1}.$$

To recap, the linear minimax (Pinsker) estimator has constant weights

$$\ell_j = \frac{Q}{Q+1}, \quad j = 1, ..., d$$

and the corresponding risk is

$$\inf_{\lambda\in\mathbb{R}^d}\sup_{\theta\in\Theta(Q)} R(\lambda, \theta) = R(\ell, \theta) = \frac{Q}{Q+1}.$$

The lower bound.

To derive the lower bound, define the prior density

$$\mu(\theta) = \prod_{i=1}^d \mu_{\delta Q}(\theta_i), \quad \theta \in \mathbb{R}^d,$$

where $\mu_\sigma(x)$ is the density of $N(0,\sigma^2)$ distribution. Then

$$R^*_d\big(\Theta(Q)\big) = \inf_{\hat\theta\in\mathbb{R}^d} \sup_{\theta\in\Theta(Q)} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2 \stackrel{\dagger}{=}$$
$$\inf_{\hat\theta\in\Theta(Q)} \sup_{\theta\in\Theta(Q)} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2 \ge \inf_{\hat\theta\in\Theta(Q)} \int_{\Theta(Q)} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2\mu(\theta)d\theta =$$
$$\inf_{\hat\theta\in\Theta(Q)} \left(\int_{\mathbb{R}^d} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2\mu(\theta)d\theta - \int_{\Theta^c(Q)} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2\mu(\theta)d\theta\right) \ge$$
$$\inf_{\hat\theta\in\mathbb{R}^d} \int_{\mathbb{R}^d} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2\mu(\theta)d\theta - \sup_{\hat\theta\in\Theta(Q)} \int_{\Theta^c(Q)} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2\mu(\theta)d\theta := I - R,$$

where $\dagger$ holds since $\Theta(Q)$ is closed and compact and hence projecting $\hat\theta$ onto $\Theta(Q)$ only reduces[4]the risk. The term I contributes the main asymptotic:

$$I = \inf_{\hat\theta\in\mathbb{R}^d} \int_{\mathbb{R}^d} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2\mu(\theta)d\theta \ge \frac{1}{d}\sum_{i=1}^d \inf_{\hat\theta} \int_{\mathbb{R}^d} \mathsf{E}_\theta(\theta_i-\hat\theta_i)^2\mu(\theta)d\theta =$$
$$\frac{1}{d}\sum_{i=1}^d \inf_{\hat\theta_i} \int_{\mathbb{R}^d} \mathsf{E}_\theta(\theta_i-\hat\theta_i)^2\mu(\theta)d\theta \stackrel{\dagger}{\ge} \frac{1}{d}\sum_{i=1}^d \frac{1}{\delta Q+1} = \frac{\delta Q}{\delta Q+1},$$

where in $\dagger$ we used the explicit formula for the Bayes risk in the problem of estimating $\theta_i \sim N(0,\delta Q)$ given $y_j = \theta_j + \xi_j$, $j = 1,...,d$ with independent θ_j's and ξ_j's and $\xi_j \sim N(0,1)$.

Next we shall bound the residual term:

$$R = \sup_{\hat\theta\in\Theta(Q)} \int_{\Theta^c(Q)} \mathsf{E}_\theta \frac{1}{d}\|\theta-\hat\theta\|^2\mu(\theta)d\theta \le$$
$$\int_{\Theta^c(Q)} \frac{2}{d}\|\theta\|^2\mu(\theta)d\theta + \int_{\Theta^c(Q)} \sup_{\hat\theta\in\Theta(Q)} \frac{2}{d}\mathsf{E}_\theta\|\hat\theta\|^2\mu(\theta)d\theta \le$$
$$\frac{2}{d}\mathbb{E}_\mu\|\theta\|^2\mathbf{1}_{\{\theta\in\Theta^c\}} + 2Q\mathbb{P}_\mu(\Theta^c) \le \frac{2}{d}\sqrt{\mathbb{E}_\mu\|\theta\|^4}\sqrt{\mathbb{P}_\mu(\theta\in\Theta^c)} + 2Q\mathbb{P}_\mu(\Theta^c) \le$$
$$6\delta Q\sqrt{\mathbb{P}_\mu(\theta\in\Theta^c)} + 2Q\mathbb{P}_\mu(\Theta^c),$$

[4]see the exact explanation following eq. (3.36) page 149 [**1**]

where in the last inequality we used the bound

$$\mathbb{E}_\mu \|\theta\|^4 = \mathbb{E}_\mu \left(\sum_{i=1}^d \theta_i^2 \right)^2 = \sum_{i \neq j} \mathbb{E}_\mu \theta_i^2 \mathbb{E}_\mu \theta_j^2 + \sum_{i=1}^d \mathbb{E}_\mu \theta_i^4 =$$

$$\sum_{i \neq j} (\delta Q)^2 + \sum_{i=1}^d \mathbb{E}_\mu 3(\delta Q)^2 \leq 3d^2 (\delta Q)^2.$$

For any $0 < \delta < 1$,

$$\mathbb{P}_\mu(\Theta^c) = \mathbb{P}_\mu \left(\frac{1}{d} \sum_{i=1}^d \theta_i^2 > Q \right) = \mathbb{P}_\mu \left(\frac{1}{d} \sum_{i=1}^d (\theta_i^2 - \delta Q) > (1-\delta) Q \right) \leq$$

$$\mathbb{P}_\mu \left(\left| \frac{1}{d} \sum_{i=1}^d (\theta_i^2 - \delta Q) \right| > (1-\delta) Q \right) \leq \frac{1}{(1-\delta)^2 Q^2} \frac{1}{d^2} \mathbb{E}_\mu \left(\sum_{i=1}^d (\theta_i^2 - \delta Q) \right)^2 =$$

$$\frac{1}{(1-\delta)^2 Q^2} \frac{1}{d^2} \mathbb{E}_\mu \sum_{i=1}^d (\theta_i^2 - \delta Q)^2 = \frac{1}{d} \frac{2(\delta Q)^2}{(1-\delta)^2 Q^2} \xrightarrow{d \to \infty} 0,$$

and hence

$$\varliminf_{d \to \infty} R_d^*\big(\Theta(Q)\big) \geq \frac{\delta Q}{\delta Q + 1}.$$

The claimed asymptotic follows by taking $\delta \to 1$.

Exerecise 3.3. Consider the setting of Exercise 3.2

(1) Prove that the Stein estimator

$$\hat{\theta}_S = \left(1 - \frac{d}{\|y\|^2} \right) y,$$

as well as the positive part Stein estimator

$$\hat{\theta}_{S+} = \left(1 - \frac{d}{\|y\|^2} \right)_+ y,$$

are adaptive in the exact minimax sense over the family of classes $\{\Theta(Q), Q > 0\}$, that is, for all $Q > 0$

$$\varlimsup_{d \to \infty} \sup_{\theta \in \Theta(Q)} \mathsf{E}_\theta \left(\frac{1}{d} \|\hat{\theta} - \theta\|^2 \right) \leq \frac{Q}{Q+1},$$

with $\hat{\theta} = \hat{\theta}_S$ or $\hat{\theta} = \hat{\theta}_{S+}$. (Here, we deal with adaptation at an unknown radius Q of the ball $\Theta(Q)$).

Hint: apply Lemma 3.10

(2) Prove that the linear minimax estimator on $\Theta(Q)$ (the Pinsker estimator) is inadmissible on the class $\Theta(Q')$ such that $0 < Q' < Q$ for all $d > d_1$, where d_1 depends only on Q and Q'.

<u>Solution</u>

1. By Lemma 3.10, for all $d \geq 4$ and $\theta \in \Theta(Q)$,

$$\frac{1}{d}\mathsf{E}_\theta\|\hat{\theta}_S - \theta\|^2 \leq \frac{\frac{1}{d}\|\theta\|^2}{\frac{1}{d}\|\theta\|^2 + 1} + \frac{4}{d} \leq \frac{Q}{Q+1} + \frac{4}{d},$$

which proves the claim (the same bound holds for $\hat{\theta}_{S+}$).

2. In Exercise 3.2(2) we saw that the linear minimax (Pinsker) estimator has constant weights

$$\ell_j = \frac{Q}{1+Q}, \quad j = 1, ..., d.$$

The corresponding risk function is

$$R(\ell,\theta) = \frac{1}{d}\mathsf{E}_\theta\|\hat{\theta}(\ell) - \theta\|^2 = \frac{1}{d}\sum_{j=1}^d \Big((\ell_j - 1)^2\theta_j^2 + \ell_j^2\Big) =$$

$$\frac{1}{(Q+1)^2}\frac{1}{d}\sum_{j=1}^d \theta_j^2 + \left(\frac{Q}{1+Q}\right)^2 =$$

$$\frac{1}{(Q+1)^2}Q + \left(\frac{Q}{1+Q}\right)^2 + \frac{1}{(Q+1)^2}\left(\frac{1}{d}\sum_{j=1}^d \theta_j^2 - Q\right) =: \frac{Q}{Q+1} + r,$$

where

$$r = \frac{1}{(Q+1)^2}\left(\frac{1}{d}\sum_{j=1}^d \theta_j^2 - Q\right) \leq \frac{1}{(Q+1)^2}\big(Q' - Q\big) < 0, \quad \forall\theta \in \Theta(Q').$$

Hence for any $\theta \in \Theta(Q')$ and all sufficiently large d's

$$R(\ell,\theta) \geq R(\hat{\theta}_S,\theta) - \frac{4}{d} + r > R(\hat{\theta}_S,\theta).$$

Exerecise 3.4. Consider the Model 1 of Section 3.4. Let $\tilde{\tau} > 0$.

(1) Show that the hard thresholding estimator $\hat{\theta}_{HT}$ with the components

$$\hat{\theta}_{j,HT} = \mathbf{1}_{\{|y_j| > \tilde{\tau}\}} y_j, \quad j = 1, ..., d,$$

is the solution of the minimization problem

$$\min_{\theta \in \mathbb{R}^d}\left\{\sum_{j=1}^d (y_j - \theta_j)^2 + \tilde{\tau}^2 \sum_{j=1}^d \mathbf{1}_{\{\theta_j \neq 0\}}\right\}.$$

(2) Show that the soft thresholding estimator $\hat{\theta}_{ST}$ with the components

$$\hat{\theta}_{j,ST} = \left(1 - \frac{\tilde{\tau}}{|y_j|}\right)_+ y_j, \quad j = 1, ..., d$$

is the solution of the minimization problem

$$\min_{\theta\in\mathbb{R}^d}\left\{\sum_{j=1}^d (y_j-\theta_j)^2+2\tilde\tau\sum_{j=1}^d|\theta_j|\right\}.$$

Solution

1. Since

$$(y_j-\theta_j)^2+\tilde\tau^2\mathbf{1}_{\{\theta_j\ne 0\}}=\Big((y_j-\theta_j)^2+\tilde\tau^2\Big)\mathbf{1}_{\{\theta_j\ne 0\}}+y_j^2\mathbf{1}_{\{\theta_j=0\}}\ge \\ \tilde\tau^2\mathbf{1}_{\{\theta_j\ne 0\}}+y_j^2\mathbf{1}_{\{\theta_j=0\}}\ge \tilde\tau^2\wedge y_j^2,$$

it follows that for any $\theta\in\mathbb{R}^d$

$$\sum_{j=1}^d (y_j-\theta_j)^2+\tilde\tau^2\sum_{j=1}^d\mathbf{1}_{\{\theta_j\ne 0\}}=\sum_{j=1}^d\Big((y_j-\theta_j)^2+\tilde\tau^2\mathbf{1}_{\{\theta_j\ne 0\}}\Big)\ge\sum_{j=1}^d\tilde\tau^2\wedge y_j^2$$

This lower bound is attained at the suggested estimator, since

$$(y_j-\hat\theta_{j,HT})^2+\tilde\tau^2\mathbf{1}_{\{\hat\theta_{j,HT}\ne 0\}}=\mathbf{1}_{\{|y_j|>\tilde\tau\}}\tilde\tau^2+\mathbf{1}_{\{|y_j|\le\tilde\tau\}}y_j^2=\tilde\tau^2\wedge y_j^2.$$

2. As before, the minimization can be carried out componentwise. The scalar function $t\mapsto\psi(t):=(y_j-t)^2+2\tilde\tau|t|$ is smooth, except for $t=0$. Hence it's local minima over $\mathbb{R}\setminus\{0\}$ must satisfy

$$\frac{d}{dt}\psi(t)=-2(y_j-t)+2\tilde\tau\text{sign}(t)=0.$$

The latter has two solutions: $t_+:=y_j-\tilde\tau$, if $y_j>\tilde\tau$, and $t_-:=y_j+\tilde\tau$, if $y_j<-\tilde\tau$. For $y_j>\tilde\tau$

$$\psi(t_+)=\tilde\tau^2+2\tilde\tau(y_j-\tilde\tau)=y_j^2-(y_j-\tilde\tau)^2<y_j^2=\psi(0)$$

and hence t_+ is the global minimum in the case $y_j>\tilde\tau$. Similarly, t_- is the minimum if $y_j<-\tilde\tau$. When $|y_j|\le\tilde\tau$, the function $\psi(t)$ doesn't have any extrema on $\mathbb{R}\setminus\{0\}$ and hence the global minimum is at the origin. To recap, the minimum is given at

$$\theta_j^*=(y_j-\text{sign}(y_j)\tilde\tau)\mathbf{1}_{\{|y_j|>\tilde\tau\}}=\left(1-\frac{\tilde\tau}{|y_j|}\right)y_j\mathbf{1}_{\{|y_j|>\tilde\tau\}}=\left(1-\frac{\tilde\tau}{|y_j|}\right)_+y_j,$$

as claimed.

Exerecise 3.5. Consider Model 1 of Section 3.4. Using Stein's lemma, show that the statistic

$$J_1(\tilde\tau)=\sum_{j=1}^d\big(2\varepsilon^2+\tilde\tau^2-y_j^2\big)\mathbf{1}_{\{|y_j|\ge\tilde\tau\}}$$

is an unbiased estimator of the risk of the soft thresholding estimator $\hat\theta_{ST}$, up to the additive term $\|\theta\|^2$ that does not depend on $\tilde\tau$:

$$\mathsf{E}_\theta J_1(\tilde\tau)=\mathsf{E}_\theta\|\hat\theta_{ST}-\theta\|^2-\|\theta\|^2.$$

Based on this, suggest a data-driven choice of the threshold $\tilde\tau$.

<u>Solution</u>

On one hand, we have

$$\mathsf{E}_\theta\|\hat\theta_{ST}-\theta\|^2 = \mathsf{E}_\theta\sum_{j=1}^d\left(\left(1-\frac{\tilde\tau}{|y_j|}\right)_+ y_j-\theta_j\right)^2 =$$

$$\sum_{j=1}^d \theta_j^2 \mathsf{P}_\theta(|y_j|<\tilde\tau) + \mathsf{E}_\theta\sum_{j=1}^d\Big(y_j-\tilde\tau\,\mathrm{sign}(y_j)-\theta_j\Big)^2\mathbf{1}_{\{|y_j|\ge\tilde\tau\}} =$$

$$\sum_{j=1}^d \theta_j^2 + \mathsf{E}_\theta\sum_{j=1}^d\Big(y_j^2-2(y_j-\theta_j)\tilde\tau\,\mathrm{sign}(y_j)+\tilde\tau^2-2\theta_j y_j\Big)\mathbf{1}_{\{|y_j|\ge\tilde\tau\}} =$$

$$\sum_{j=1}^d \theta_j^2 + \mathsf{E}_\theta\sum_{j=1}^d\big(\tilde\tau^2-y_j^2\big)\mathbf{1}_{\{|y_j|\ge\tilde\tau\}} - 2\sum_{j=1}^d\mathsf{E}_\theta(\theta_j-y_j)f(y_j),$$

where

$$f(u)=\big(u-\tilde\tau\,\mathrm{sign}(u)\big)\mathbf{1}_{\{|u|\ge\tilde\tau\}},\quad u\in\mathbb{R}.$$

Note that $f(u)=\int_0^u \mathbf{1}_{\{|s|\ge\tilde\tau\}}ds$ and hence $f(u)$ is absolutely continuous and by the Stein lemma

$$\mathsf{E}_\theta(\theta_j-y_j)f(y_j) = -\varepsilon^2\mathsf{E}_\theta\mathbf{1}_{\{|y_j|\ge\tilde\tau\}},$$

which verifies the claim.

Ideally we would choose $\tilde\tau$ so that the risk or, equivalently, $\mathsf{E}_\theta J_1(\tilde\tau)$ is minimized. This is impractical, since such choice depends on the unknown θ and hence it is not unreasonable to choose $\tilde\tau$ to minimize $J_1(\tilde\tau)$. A close look reveals that J_1 doesn't have a minimum, but its version with the strict inequality in the indicator

$$J_1(\tilde\tau) := \sum_{j=1}^d\big(2\varepsilon^2+\tilde\tau^2-y_j^2\big)\mathbf{1}_{\{|y_j|>\tilde\tau\}}$$

does and the optimal value $\tilde\tau^*$ belongs to the data set $\{y_1,...,y_d\}$, since otherwise the value of J_1 can be decreased by setting it to the greatest y_j less than $\tilde\tau^*$. Moreover, $\tilde\tau^*$ does not exceed the greatest y_j smaller than $\sqrt{2}\varepsilon$, which is checked by contradiction. These properties reduce finding $\tilde\tau^*$ to a simple computationally efficient search.

Exerecise 3.6. Consider Model 1 of Section 3.4. Let $\tau>0$.

(1) Show that the global hard thresholding estimator

$$\hat\theta_{GHT}=\mathbf{1}_{\{\|y\|>\tau\}}y$$

is a solution of the minimization problem

$$\min_{\theta\in\mathbb{R}^d}\left\{\sum_{j=1}^d(y_j-\theta_j)^2+\tau^2\mathbf{1}_{\{\|\theta\|\ne0\}}\right\}.$$

(2) Show that the global soft thresholding estimator

$$\hat{\theta}_{GST} = \left(1 - \frac{\tau}{\|y\|}\right)_+ y$$

is a solution of the minimization problem

$$\min_{\theta \in \mathbb{R}^d} \left\{ \sum_{j=1}^d (y_j - \theta_j)^2 + 2\tau\|\theta\| \right\}.$$

Solution

1. We have

$$\sum_{j=1}^d (y_j - \theta_j)^2 + \tau^2 \mathbf{1}_{\{\|\theta\| \neq 0\}} = \sum_{j=1}^d y_j^2 \mathbf{1}_{\{\|\theta\|=0\}} + \Big(\tau^2 + \sum_{j=1}^d (y_j - \theta_j)^2\Big)\mathbf{1}_{\{\|\theta\| \neq 0\}} \geq \|y\|^2 \wedge \tau^2.$$

The inequality is saturated by the choice $\theta := \mathbf{1}_{\{\|y\| \geq \tau\}} y$, and the claim follows.

2. The function

$$g(\theta) := \sum_{j=1}^d (y_j - \theta_j)^2 + 2\tau\|\theta\|$$

is differentiable on $\mathbb{R}^d \setminus \{0\}$ and hence all of its extrema on this set satisfy $\nabla g(\theta) = 0$, i.e.

$$\nabla g(\theta) = -2(y - \theta) + 2\tau \frac{\theta}{\|\theta\|} = 0,$$

which yields

$$\theta \left(1 + \frac{\tau}{\|\theta\|}\right) = y.$$

This means that the extremum θ^* has the same direction as y, i.e. it has the form $\hat{\theta}^* = yt$, where $t > 0$ solves the scalar equation

$$t\left(1 + \frac{\tau}{t\|y\|}\right) = 1.$$

This equation doesn't have a solution when $\|y\| < \tau$ and has the unique solution $t = 1 - \frac{\tau}{\|y\|}$ otherwise. Hence when $\|y\| < \tau$ the only possible minima is outside $\mathbb{R}^d \setminus \{0\}$, i.e. at the origin. When $\|y\| \geq \tau$ the only extremum of $g(\theta)$ over $\mathbb{R}^d \setminus \{0\}$ is

$$\theta^* = \left(1 - \frac{\tau}{\|y\|}\right) y.$$

This point is clearly a local minimum, since $g(\theta) \to \infty$ as $\theta \to 0$. To decide when it is a global minimum, we shall compare the values of $g(0) = \|y\|^2$ and

$$\begin{aligned}
&g(\theta^*) = \|y\|^2 - 2\langle \theta^*, y\rangle + \|\theta^*\|^2 + 2\tau\|\theta^*\| = \\
&g(0) - 2\left(1 - \frac{\tau}{\|y\|}\right)\|y\|^2 + \left(1 - \frac{\tau}{\|y\|}\right)^2 \|y\|^2 + 2\tau\left(1 - \frac{\tau}{\|y\|}\right)\|y\| = \\
&g(0) - (\|y\| - \tau)^2.
\end{aligned}$$

Hence for $\|y\| \geq \tau$, the θ^* is the global minimum, which verifies the suggested solution.

EXERECISE 3.7. Consider the Model 1 of Section 3.4 Define a global hard thresholding estimator of the vector $\theta = (\theta_1, ..., \theta_d)$ as follows

$$\hat{\theta} = \mathbf{1}_{\{\|y\|>\tau\}} y,$$

where $\tau = 2\varepsilon\sqrt{d}$.

(1) Prove that for $\|\theta\|^2 \leq \varepsilon^2 d/4$ we have

$$\mathsf{P}_\theta(\hat{\theta} = y) \leq \exp(-c_0 d),$$

where $c_0 > 0$ is an absolute constant.

Hint: Use the following inequality

$$\mathsf{P}\left(\sum_{j=1}^{d}(\xi_j^2 - 1) \geq td\right) \leq \exp\left(-\frac{t^2 d}{8}\right), \quad t \in (0, 1].$$

(2) Based on (1) prove that

$$\mathsf{E}_\theta\|\hat{\theta} - \theta\|^2 \leq \|\theta\|^2 + c_1\varepsilon^2 d \exp(-c_0 d/2),$$

for $\|\theta\|^2 \leq \varepsilon^2 d/4$ with an absolute constat $c_1 > 0$.

(3) Show that, for all $\theta \in \mathbb{R}^d$,

$$\mathsf{E}_\theta\|\hat{\theta} - \theta\|^2 \leq 9\varepsilon^2 d.$$

(4) Combine (2) and (3) to prove the oracle inequality

$$\mathsf{E}_\theta\|\hat{\theta} - \theta\|^2 \leq c_2 \frac{d\varepsilon^2\|\theta\|^2}{d\varepsilon^2 + \|\theta\|^2} + c_1\varepsilon^2 d \exp(-c_0 d/2), \quad \theta \in \mathbb{R}^d,$$

where $c_2 > 0$ is an absolute constant.

Hint: $\min(a, b) \leq \frac{2ab}{a+b}$ for all $a \geq 0$, $b > 0$.

(5) We switch now to the Gaussian sequence model (3.10):

$$y_j = \theta_j + \varepsilon\xi_j, \quad j \geq 1.$$

Introduce the blocks B_j of size $\mathrm{card}(B_j) = j$ and define the estimators

$$\tilde{\theta}_k = \mathbf{1}_{\{\|y\|_{(j)}>\tau_j\}} y_k \quad \text{for } k \in B_j,\ j = 1, 2, ..., J,$$

where $\tau_j = 2\varepsilon\sqrt{j}$, $J \geq 1/\varepsilon^2$, and $\tilde{\theta}_k = 0$ for $k > \sum_{j=1}^{J} \mathrm{card}(B_j)$. Set $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, ...)$. Prove the oracle inequality

$$\mathsf{E}_\theta\|\tilde{\theta} - \theta\|^2 \leq c_3 \min_{\lambda \in \Lambda_{mon}} R(\lambda, \theta) + c_4\varepsilon^2, \quad \theta \in \ell_2(\mathbb{N}),$$

where $c_3 > 0$ and $c_4 > 0$ are absolute constants.

(6) Show that the estimator $\tilde\theta$ define in (5) is adaptive in the minimax sense on the family of classes $\{\Theta(\beta,Q),\beta>0,Q>0\}$, i.e. for all sufficiently small $\varepsilon>0$,

$$\sup_{\theta\in\Theta(\beta,Q)} \mathsf{E}_\theta\|\tilde\theta-\theta\|^2 \le C(\beta,Q)\varepsilon^{\frac{4\beta}{2\beta+1}},\quad \forall\beta>0,Q>0,$$

where $C(\beta,Q)$ is a constant depending only on β and Q.

Solution

1. For any $\theta\in\mathbb{R}^d$,

$$\mathsf{P}_\theta(\hat\theta=y)=\mathsf{P}_\theta(\|y\|>\tau)=\mathsf{P}_\theta(\|\theta+\varepsilon\xi\|>\tau)\le\mathsf{P}_\theta(\|\theta\|+\varepsilon\|\xi\|>\tau)=\mathsf{P}_\theta\left(\|\xi\|>\frac{\tau-\|\theta\|}{\varepsilon}\right)=$$
$$\mathsf{P}_\theta\left(\|\xi\|>\frac{2\varepsilon\sqrt{d}-\frac12\varepsilon\sqrt{d}}{\varepsilon}\right)=\mathsf{P}_\theta\left(\|\xi\|^2>\frac94 d\right)=\mathsf{P}_\theta\left(\sum_{j=1}^d(\xi_j^2-1)>\frac54 d\right)\le\exp\left(-\frac{25}{128}d\right),$$

i.e. the claim holds with $c_0=25/128$.

2. By the Cauchy-Schwarz inequality and the bound from (1)

$$\mathsf{E}_\theta\|\hat\theta-\theta\|^2=\|\theta\|^2\mathsf{P}_\theta(\|y\|\le\tau)+\varepsilon^2\mathsf{E}_\theta\|\xi\|^2\mathbf{1}_{\{\|y\|>\tau\}}\le\|\theta\|^2+\varepsilon^2\sqrt{\mathsf{E}_\theta\|\xi\|^4}\sqrt{\mathsf{P}(\|y\|>\tau)}$$

and the claim follows with $c_1:=\sqrt{3}$, since by the Jensen inequality

$$\sqrt{\mathsf{E}_\theta\|\xi\|^4}=\sqrt{\mathsf{E}_\theta\Big(\sum_{j=1}^d\xi_j^2\Big)^2}\le\sqrt{d\mathsf{E}_\theta\sum_{j=1}^d\xi_j^4}=\sqrt{3d^2}\le\sqrt{3}d.$$

3. For $\theta\in\mathbb{R}^d$,

$$\mathsf{E}_\theta\|\hat\theta-\theta\|^2=\|\theta\|^2\mathsf{P}_\theta(\|y\|<\tau)+\mathsf{E}_\theta\|y-\theta\|^2\mathbf{1}_{\{\|y\|\ge\tau\}}\le\|\theta\|^2\mathsf{P}_\theta(\|y\|<\tau)+\varepsilon^2 d.$$

For $\|\theta\|>\tau$, by the symmetry of the Gaussian distribution [5]

$$\mathsf{P}_\theta(\|y\|<\tau)=\mathsf{P}_\theta(\|\varepsilon\xi+\theta\|^2<\tau^2)=\mathsf{P}_\theta(\|\varepsilon\xi+v\|^2<\tau^2)=$$
$$\mathsf{P}_\theta\left((\varepsilon\xi_1+\|\theta\|)^2+\varepsilon^2\sum_{j=2}^d\xi_j^2<\tau^2\right)\le\mathsf{P}_\theta\left(\Big|\xi_1+\|\theta\|/\varepsilon\Big|<\tau/\varepsilon\right)=$$
$$\int_{-\tau/\varepsilon-\|\theta\|/\varepsilon}^{\tau/\varepsilon-\|\theta\|/\varepsilon}\varphi(x)dx\le\int_{-\infty}^{\tau/\varepsilon-\|\theta\|/\varepsilon}\varphi(x)dx=\Phi\big(\tau/\varepsilon-\|\theta\|/\varepsilon\big).$$

Hence

$$\|\theta\|^2\mathsf{P}_\theta(\|y\|<\tau)=\|\theta\|^2\mathsf{P}_\theta(\|y\|<\tau)\mathbf{1}_{\{\|\theta\|\le\tau\}}+\|\theta\|^2\mathsf{P}_\theta(\|y\|<\tau)\mathbf{1}_{\{\|\theta\|>\tau\}}\le$$
$$\tau^2+\sup_{\|\theta\|\ge\tau}\|\theta\|^2\Phi\big(\tau/\varepsilon-\|\theta\|/\varepsilon\big)=\tau^2+\varepsilon^2\sup_{x\ge\tau/\varepsilon}x^2\Phi\big(\tau/\varepsilon-x\big).$$

[5] v is the vector with all but one zero entries which equals $\|\theta\|$ and φ and Φ are the $N(0,1)$ density and the c.d.f. respectively

A calculation shows that for $d \geq 2$, the supremum is attained at $x := \tau/\varepsilon$ and hence

$$\|\theta\|^2 \mathsf{P}_\theta(\|y\| < \tau) \leq \tau^2 + \frac{1}{2}\tau^2.$$

For $d = 1$, the supremum is less than $\frac{1}{3}\tau^2$ and hence

$$\mathsf{E}_\theta\|\hat{\theta} - \theta\|^2 \leq \frac{3}{2}\tau^2 + \varepsilon^2 d \leq 7\varepsilon^2 d, \quad \theta \in \mathbb{R}^d,\ d \geq 1,$$

as claimed.

Remark. Only a slightly worse bound is obtained in a simpler way:

$$\mathsf{E}_\theta\|\hat{\theta} - \theta\|^2 \leq 2\mathsf{E}_\theta\|\hat{\theta} - y\|^2 + 2\mathsf{E}_\theta\|y - \theta\|^2 = 2\mathsf{E}_\theta\|y\|^2\mathbf{1}_{\{\|y\| \leq \tau\}} + 2\varepsilon^2\mathsf{E}_\theta\|\xi\|^2 = 2\tau^2 + 2\varepsilon^2\mathsf{E}_\theta\|\xi\|^2 = 10\varepsilon^2 d$$

4. Following the hint, for $\|\theta\|^2 \leq \varepsilon^2 d/4$

$$\mathsf{E}_\theta\|\hat{\theta} - \theta\|^2 \leq \|\theta\|^2 + c_1\varepsilon^2 d\exp(-c_0 d/2) \leq \min\big(\|\theta\|^2, \varepsilon^2 d\big) + c_1\varepsilon^2 d\exp(-c_0 d/2) \leq \frac{2\|\theta\|^2\varepsilon^2 d}{\|\theta\|^2 + \varepsilon^2 d} + c_1\varepsilon^2 d\exp(-c_0 d/2).$$

On the other hand, for any $\theta \in \mathbb{R}^d$,

$$9\frac{\|\theta\|^2\varepsilon^2 d}{\|\theta\|^2 + \varepsilon^2 d} + c_1\varepsilon^2 d\exp(-c_0 d/2) \geq 9\varepsilon^2 d \geq \mathsf{E}_\theta\|\hat{\theta} - \theta\|^2$$

and hence the claimed bound in fact holds with $c_2 := 9$ for any $\theta \in \mathbb{R}^d$.

5. Let $\theta_{(j)}$ and $\tilde{\theta}_{(j)}$ denote the restrictions of the sequences θ and $\tilde{\theta}$ to the block B_j and $N_{\max} := \sum_{j=1}^J \operatorname{card}(B_j) = \frac{1}{2}J(J+1)$. Recall that

$$\min_{t_j \in \mathbb{R}} \|t_j y_{(j)} - \theta_{(j)}\|^2 = \frac{\operatorname{card}(B_j)\varepsilon^2\|\theta_{(j)}\|^2}{\operatorname{card}(B_j)\varepsilon^2 + \|\theta_{(j)}\|^2}$$

and hence by the oracle inequality from (4),

$$\begin{aligned}
&\mathsf{E}_\theta\|\tilde{\theta} - \theta\|^2 = \sum_{j=1}^J \mathsf{E}_\theta\|\tilde{\theta}_{(j)} - \theta_{(j)}\|^2 + \sum_{k > N_{\max}} \theta_k^2 \leq \\
&c_2\sum_{j=1}^J \min_{t_j \in \mathbb{R}} \|t_j y_{(j)} - \theta_{(j)}\|^2 + \sum_{j=1}^J c_1\varepsilon^2 \operatorname{card}(B_j) e^{-\frac{c_0}{2}\operatorname{card}(B_j)} + \sum_{k > N_{\max}} \theta_k^2 = \\
&c_2 \inf_{\lambda \in \Lambda^*} R(\lambda, \theta) + c_1\varepsilon^2 \sum_{j=1}^J j e^{-\frac{c_0}{2}j} \stackrel{\dagger}{\leq} \\
&c_2\Big(1 + \eta\Big) \inf_{\lambda \in \Lambda_{mon}} R(\lambda, \theta) + c_2\varepsilon^2 T_1 + c_1\varepsilon^2 \sum_{j=1}^\infty j e^{-\frac{c_0}{2}j} = c_3 \inf_{\lambda \in \Lambda_{mon}} R(\lambda, \theta) + c_4\varepsilon^2
\end{aligned}$$

where $c_3 := 2c_2$ and

$$c_4 := c_2 + c_1\sum_{j=1}^\infty j e^{-\frac{c_0}{2}j} < \infty.$$

and the bound † holds[6] by Lemma 3.11 on page 175 [**1**], with $\eta := 1$ and $T_1 := 1$.

6.Pinsker's weights belong to Λ_{mon}. Hence for sufficiently small $\varepsilon > 0$

$$\min_{\lambda \in \Lambda_{mon}} R(\lambda, \theta) \leq R(\ell, \theta) = C^* \varepsilon^{\frac{4\beta}{2\beta+1}} (1 + o(1))$$

and the claim follows from the bound, obtained in (5).

[6]Note that Lemma 3.11 is valid for any $N_{\max}$ ($N_{\max} = [1/\varepsilon^2]$ is not assumed). In our case, $N_{\max} \propto J^2 = 1/\varepsilon^4$.

Bibliography

[1] A.Tsybakov, Introduction to Nonparametric Estimation, Springer, 2009