Fantuan's Academia

Fantuan's Statistics Notes series Vol.3

Selected Topics in Statistical Decision Theory

Author: Jingxuan Xu
Lecturer: Prof. Harrison Zhou

February 1, 2025

Contents

All the Sections with * are hard sections and can be skipped without losing coherence.

This note is a scribe of S&DS611 course: Selected Topics in Statistical Decision Theory taught by Prof. Harrison Zhou at Yale University.

Chapter 1

Nonparametric Estimation

I also referenced to **Introduction to Nonparametric Estimation**[1] by Alexandre B. Tsybakov and **Gaussian estimation: Sequence and wavelet models** by Iain M. Johnstone when scribing the lecture.

1.1 Gaussian Sequence Model

1.1.1 Discrete Fourier Transform

Consider the nonparametric regression model

$$Y_i = f(X_i) + Z_i, \quad i = 1, 2, \cdots, n,\ Z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$$

where X_i are deterministic and $f \in L_2[0,1] : [0,1] \to \mathbb{R}$ is a periodic function. We want to estimate f. Consider the case $X_i = i/n$. Let θ_i be the Fourier coefficients of f w.r.t. orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$ of $L_2[0,1]$:

$$\theta_i = \int_0^1 f(x)\varphi_i(x)\,\mathrm{d}x, \quad f(x) = \sum_{i=1}^{\infty} \theta_i \varphi_i(x)$$

A natural estimator for θ_i is

$$\hat{\theta}_i = \frac{1}{n}\sum_{j=1}^{n} Y_j \varphi_i(X_j)$$

Here $X_i = i/n$ are nicely spread in $[0,1]$. Then, for large n,

$$\frac{1}{n}\sum_{i=1}^{n} \phi_j(X_i)\varphi_k(X_i) \approx \int_0^1 \varphi_j(x)\varphi_k(x)\,\mathrm{d}x = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \tag{1.1}$$

Hence $\{\varphi_i(X_j)\}_{j=1}^{\infty}$ approximately behaves like an orthonormal system in the discrete sense. The approximate expectation and variance is

$$\mathbb{E}\left[\hat{\theta}_i\right] = \frac{1}{n}\sum_{j=1}^{n} \varphi_i(X_j)\mathbb{E}[Y_j] = \frac{1}{n}\sum_{j=1}^{n} \varphi_i(X_j) f(X_j) \approx \int_0^1 \varphi_i(x) f(x) = \theta_i$$

$$\mathrm{Var}\left[\hat{\theta}_i\right] = \frac{1}{n^2}\sum_{j=1}^{n}\varphi_i(X_j)^2\mathrm{Var}[Y_j] = \frac{1}{n^2}\sum_{i=1}^{n}\phi_i(X_j)^2 \approx \frac{1}{n}$$

where the last step uses Equation 1.1. Put it together, by central limit theorem, we have

$$\hat{\theta}_i = \theta_i + \frac{1}{\sqrt{n}}Z_i, \quad i = 1, 2, \cdots, \; Z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$$

which is the form of **Gaussian sequence model**.

1.1.2 Sobolev Ellipsoid

We assume the regression function f is sufficiently smooth. We will assume that it belongs to the *periodic Sobolev class*

$$W^{\text{per}}(\alpha, L) = \left\{ f : [0,1] \to \mathbb{R} \text{ periodic } : f^{(\alpha-1)} \text{ is absolutely continuous and } \int_0^1 (f^{(\alpha)}(x))^2 \,\mathrm{d}x \leqslant L^2 \right\}$$

It can be proved that, this function space, after mapping to the sequence space using discrete Fourier transform with trigonometric basis $\phi_1(x) = 1, \phi_{2k}(x) = \sqrt{2}\cos(2\pi kx), \phi_{2k+1}(x) = \sqrt{2}\sin(2\pi kx),\ k = 1, 2, \cdots,$ is isomorphic to the ellipsoid

$$\Theta' = \left\{ \theta : \sum_{i=1}^{\infty} a_i^2\theta_i^2 \leqslant M \right\}$$

where $a_1 = 0, a_{2i} = a_{2i+1} = (2i)^\alpha$. This is called a **Sobolev ellipsoid**. For proof of this result, see Tsybakov[1] Lemma A.3. Often time, we set $a_i = i^\alpha$ so that

$$\Theta = \left\{ \theta : \sum_{i=1}^{\infty} i^{2\alpha}\theta_i^2 \leqslant M \right\}$$

One can show that these two sets' minimax rates are very close, and $\Theta \subseteq \Theta'$. See Johnstone Section 3.1.

1.1.3 Minimax Risk Upper and Lower Bound

Consider the Gaussian sequence model

$$Y_i = \theta_i + \frac{1}{\sqrt{n}}Z_i, \quad i = 1, 2, \cdots, \; Z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$$

with parameter space being the Sobolev ellipsoid

$$\Theta = \left\{ \theta : \sum_{i=1}^{\infty} i^{2\alpha}\theta_i^2 \leqslant M \right\}$$

Our goal is to find the lower bound and upper bound of minimax risk.

1. Upper Bound

Consider the estimator

$$\hat{\theta}_i = \begin{cases} Y_i, & i \leqslant k \\ 0, & i > k \end{cases}$$

The risk function w.r.t. the squared loss is then

$$\begin{aligned}
\mathbb{E}\left[\|\hat{\theta}-\theta\|_2^2\right] &= \mathbb{E}\left[\sum_{i=1}^{k}(Y_i-\theta_i)^2\right] + \mathbb{E}\left[\sum_{i=k+1}^{\infty}\theta_i^2\right] \\
&= \sum_{i=1}^{k}\mathbb{E}\left[\frac{1}{n}Z_i^2\right] + \sum_{i=k+1}^{\infty}\theta_i^2 && (Y_i-\theta_i=\tfrac{1}{\sqrt{n}}Z_i) \\
&\leqslant \frac{k}{n} + \frac{1}{(k+1)^{2\alpha}}\sum_{i=k+1}^{\infty} i^{2\alpha}\theta_i^2 \\
&\leqslant \frac{k}{n} + \frac{M}{k^{2\alpha}} && (\theta\in\Theta)
\end{aligned}$$

Now we want to get the tightest upper bound by choosing appropriate k. Set the variance and bias^2 equal, we have

$$\frac{k}{n} = \frac{M}{k^{2\alpha}} \quad\Longrightarrow\quad k=(Mn)^{\frac{1}{2\alpha+1}}$$

We then get the upper bound:

$$\mathbb{E}\left[\|\hat{\theta}-\theta\|_2^2\right] \leqslant \frac{(Mn)^{\frac{1}{2\alpha+1}}}{n} + \frac{M}{(Mn)^{\frac{2\alpha}{2\alpha+1}}} = 2M^{\frac{1}{2\alpha+1}}n^{-\frac{2\alpha}{2\alpha+1}} = Cn^{-\frac{2\alpha}{2\alpha+1}}$$

where C is a constant not depending on n. This is lower than typical parametric models, which have convergence rate n^{-1}. Since it holds for all $\theta\in\Theta$, we have the minimax upper bound

$$\inf_{\hat{\theta}}\sup_{\theta\in\Theta}\mathbb{E}_{Y|\theta}\left[\|\hat{\theta}-\theta\|_2^2\right] \leqslant Cn^{-\frac{2\alpha}{2\alpha+1}}$$

2. Lower Bound

To find the lower bound, we seek a sub-parameter space Θ_0. We choose

$$\Theta_0 = \left\{\theta:\theta_i=\frac{1}{\sqrt{n}},\ i\leqslant k,\quad \theta_i=0,\ i\geqslant k+1\right\}$$

where $k=(Mn)^{\frac{1}{1+2\alpha}}$. To show that it is indeed a subspace of Θ, note that

$$\sum_{i=1}^{\infty} i^{2\alpha}\theta^2 = \sum_{i=1}^{k}\frac{1}{n}i^{2\alpha} \leqslant \frac{k}{n}k^{2\alpha} = M$$

In this smaller parameter space, we can always truncate the tail from $k+1$ term, since we can always precisely estimate these 0's. For any estimator $\hat{\theta}$, we have

$$\sup_{\theta\in\Theta}\mathbb{E}_{Y|\theta}\left[\|\hat{\theta}-\theta\|_2^2\right] \geqslant \sup_{\theta\in\Theta_0}\mathbb{E}_{Y|\theta}\left[\|\hat{\theta}-\theta\|_2^2\right]$$

since $\Theta_0 \subseteq \Theta$. Take infimum on both sides, we have that

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{Y|\theta} \left[\|\hat{\theta} - \theta\|_2^2 \right] \geqslant \inf_{\hat{\theta}} \sup_{\theta \in \Theta_0} \mathbb{E}_{Y|\theta} \left[\|\hat{\theta} - \theta\|_2^2 \right] \tag{1.2}$$

Therefore, we only need to find the lower bound in this sub-paremeter space. Take prior distribution π of θ_i: $\mathbb{P}(\theta_i = 0) = \mathbb{P}(\theta_i = \frac{1}{\sqrt{n}}) = 1/2$. Let $\hat{\theta}_{i,Bayes}$ denotes the Bayes estimator of θ_i under this prior. Let $\varphi_{\alpha,\beta}$ be the normal density with mean α and variance β. Since minimax risk is lower bounded by the average risk, we have

$$\begin{aligned}
\sup_{\theta \in \Theta_0} \mathbb{E}_{Y|\theta} \left[\|\hat{\theta} - \theta\|_2^2 \right] &\geqslant \mathbb{E}_\theta \mathbb{E}_{Y|\theta} \left[\|\hat{\theta} - \theta\|_2^2 \right] && \text{(worst-case risk greater than average risk)} \\
&\geqslant \mathbb{E}_\theta \mathbb{E}_{Y|\theta} \left[\sum_{i=1}^{k} (\hat{\theta}_{i,Bayes} - \theta_i)^2 \right] && \text{(Truncate to } k\text{)} \\
&= \sum_{i=1}^{k} \left(\frac{1}{2} \mathbb{E}_{Y_i|\theta_i = 0} \left[\left(\hat{\theta}_{i,Bayes} - 0 \right)^2 \right] + \frac{1}{2} \mathbb{E}_{Y_i|\theta_i = \frac{1}{\sqrt{n}}} \left[\left(\hat{\theta}_{i,Bayes} - \frac{1}{\sqrt{n}} \right)^2 \right] \right) \\
&= \sum_{i=1}^{k} \left(\frac{1}{2} \int \left(\hat{\theta}_{i,Bayes} \right)^2 \varphi_{0,\frac{1}{n}}(x_i) \, \mathrm{d}x_i + \frac{1}{2} \left(\hat{\theta}_{i,Bayes} - \frac{1}{\sqrt{n}} \right)^2 \varphi_{\frac{1}{\sqrt{n}},\frac{1}{n}}(x_i) \, \mathrm{d}x_i \right) \\
&\geqslant \sum_{i=1}^{k} \frac{1}{2} \int \left(\hat{\theta}_{i,Bayes}^2 + \left(\hat{\theta}_{i,Bayes} - \frac{1}{\sqrt{n}} \right)^2 \right) \min \left\{ \varphi_{0,\frac{1}{n}}(x_i), \varphi_{\frac{1}{\sqrt{n}},\frac{1}{n}}(x_i) \right\} \mathrm{d}x_i \\
&\geqslant \sum_{i=1}^{k} \frac{1}{2} \int \left(\frac{1}{2} \left(\frac{1}{\sqrt{n}} \right)^2 \right) \min \left\{ \varphi_{0,\frac{1}{n}}(x_i), \varphi_{\frac{1}{\sqrt{n}},\frac{1}{n}}(x_i) \right\} \mathrm{d}x_i && (a^2 + b^2 \geqslant \tfrac{1}{2}(a-b)^2) \\
&= \frac{k}{4n} \int \min \left\{ \varphi_{0,\frac{1}{n}}(x), \varphi_{\frac{1}{\sqrt{n}},\frac{1}{n}}(x) \right\} \mathrm{d}x
\end{aligned}$$

(HW 1 Solution:)

Note that with change of variable $z = \sqrt{n}x$, we have $\mathrm{d}z = \sqrt{n}\,\mathrm{d}x$, and the integration becomes

$$\begin{aligned}
\int \min \left\{ \varphi_{0,\frac{1}{n}}(x), \varphi_{\frac{1}{\sqrt{n}},\frac{1}{n}}(x) \right\} \mathrm{d}x &= \int \min \left\{ \frac{1}{\sqrt{2\pi \frac{1}{n}}} \exp\left(-\frac{(x - \frac{1}{\sqrt{n}})^2}{2\frac{1}{n}} \right), \frac{1}{\sqrt{2\pi \frac{1}{n}}} \exp\left(-\frac{x^2}{2\frac{1}{n}} \right) \right\} \mathrm{d}x \\
&= \int \min \left\{ \frac{1}{\sqrt{2\pi \frac{1}{n}}} \exp\left(-\frac{(\frac{z}{\sqrt{n}} - \frac{1}{\sqrt{n}})^2}{2\frac{1}{n}} \right), \frac{1}{\sqrt{2\pi \frac{1}{n}}} \exp\left(-\frac{(\frac{z}{\sqrt{n}})^2}{2\frac{1}{n}} \right) \right\} \frac{1}{\sqrt{n}} \mathrm{d}z \\
&= \int \min \left\{ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z-1)^2}{2} \right), \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2} \right) \right\} \mathrm{d}z \\
&= \int \min \left\{ \varphi_{0,1}(x), \varphi_{1,1}(x) \right\} \mathrm{d}x = c'
\end{aligned}$$

where $c' > 0$ is a positive constant that is not dependent on n. Therefore, we can get the lower bound

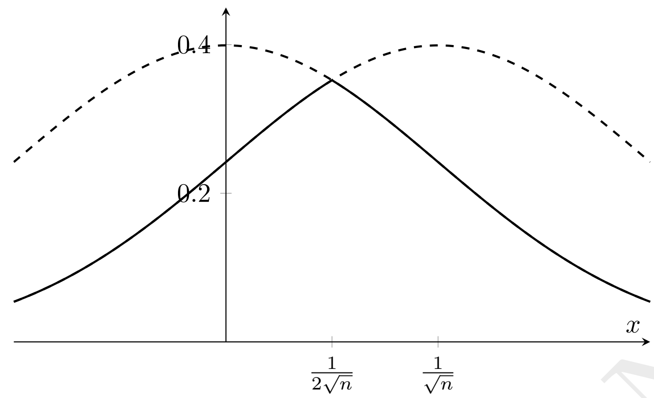
$$\sup_{\theta \in \Theta_0} \mathbb{E}_{Y|\theta} \left[\|\hat{\theta} - \theta\|_2^2 \right] \geqslant \frac{c'k}{4n} = \frac{c'}{4n} (Mn)^{\frac{1}{1+2\alpha}} = \frac{c'}{4} M^{\frac{1}{1+2\alpha}} n^{-\frac{2\alpha}{2\alpha+1}} = cn^{-\frac{2\alpha}{2\alpha+1}}$$

Since this is true for all $\hat{\theta}$, we take infimum on both sides, and get

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_0} \mathbb{E}_{Y|\theta} \left[\|\hat{\theta} - \theta\|_2^2 \right] \geqslant cn^{-\frac{2\alpha}{2\alpha+1}}$$

Then, combine this with Equation 1.2, we have the final lower bound

$$\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}_{Y|\theta}\left[\|\hat{\theta}-\theta\|_2^2\right] \geqslant \inf_{\hat{\theta}} \sup_{\theta\in\Theta_0} \mathbb{E}_{Y|\theta}\left[\|\hat{\theta}-\theta\|_2^2\right] \geqslant cn^{-\frac{2\alpha}{2\alpha+1}}$$



With the upper bound and the lower bound, we have

$$\boxed{\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}_{Y|\theta}\left[\|\hat{\theta}-\theta\|_2^2\right] \asymp n^{-\frac{2\alpha}{2\alpha+1}}}$$

where $f(n) \asymp g(n)$ denotes the case when there exists two positive constants c, C independent of n such that $cg(n) \leqslant f(n) \leqslant Cg(n)$.

1.2 Best Linear Procedure

To get a better bound, instead of just cutting the estimator discretely at a specific k (as what we did in upper bound calculation, where we set $\hat{\theta}_i = Y_i$ when $i \leqslant k$ and $\hat{\theta}_i = 0$ otherwise), we can make it gradually decreases to 0, where we introduce a linear procedure c_iY_i here. We consider a more general Gaussian sequence model

$$Y_i = \theta_i + \sigma Z_i, \quad i = 1, 2, \cdots, \; Z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$$

with variance σ^2 instead of $1/n$. Consider $c = (c_1, c_2, \cdots)$, where each $c_i \in [0,1]$ and the estimator $\hat{\theta}_i = c_iY_i$. Then, the minimax risk

$$\begin{aligned}
\inf_{c\in[0,1]^\infty} \sup_{\theta\in\Theta} \mathbb{E}\left[\sum_{i=1}^{\infty}(c_iY_i-\theta_i)^2\right] &= \inf_{c\in[0,1]^\infty} \sup_{\theta\in\Theta} \sum_{i=1}^{\infty} \mathbb{E}\left[(c_iY_i-\theta_i)^2\right] \\
&= \inf_{c\in[0,1]^\infty} \sup_{\theta\in\Theta} \sum_{i=1}^{\infty} \left(c_i^2\mathbb{E}[Y_i^2] - 2c_i\theta_i\mathbb{E}[Y_i] + \theta_i^2\right) \\
&= \inf_{c\in[0,1]^\infty} \sup_{\theta\in\Theta} \sum_{i=1}^{\infty} \left(c_i^2(\sigma^2+\theta_i^2) - 2c_i\theta_i^2 + \theta_i^2\right) \\
&= \inf_{c\in[0,1]^\infty} \sup_{\theta\in\Theta} \sum_{i=1}^{\infty} \left[\underbrace{(c_i-1)^2\theta_i^2}_{\text{bias}^2} + \underbrace{c_i^2\sigma^2}_{\text{variance}}\right] \qquad (1.3)
\end{aligned}$$

The goal here is to find the optimal c (find the *best linear procedure*) that attains this minimax risk. To make the calculation simpler, we use the concave-convex function version of minimax theorem.

Theorem 1.2.1: Minimax Theorem

Let X be a convex set and Y be a convex, compact set. If $f : X \times Y \to \mathbb{R}$ such that

- $f(\cdot, y)$ is continuous and convex on X, for any fixed $y \in Y$, and
- $f(x, \cdot)$ is continuous and concave on Y, for any fixed $x \in X$

Then,

$$\inf_{x\in X}\sup_{y\in Y} f(x,y) = \sup_{y\in Y}\inf_{x\in X} f(x,y)$$

In our case, if we set $\theta_i^2 = \omega_i$, then

$$X = [0,1]^\infty, \quad Y = \left\{\omega : \sum_{i=1}^{\infty} a_i^2\omega_i \leqslant M\right\}$$

Our $f(c,\omega) = \sum_{i=1}^{\infty}\left[(c_i-1)^2\omega_i + c_i^2\sigma^2\right]$. Then, it is convex on c (quadratic), and concave on ω (linear). Therefore, we can continue 1.3 and get

$$\inf_{c\in[0,1]^\infty}\sup_{\theta\in\Theta}\mathbb{E}\left[\sum_{i=1}^{\infty}(c_iY_i-\theta_i)^2\right] = \inf_{c\in[0,1]^\infty}\sup_{\theta\in\Theta}\sum_{i=1}^{\infty}\left[(c_i-1)^2\theta_i^2 + c_i^2\sigma^2\right] = \sup_{\theta\in\Theta}\inf_{c\in[0,1]^\infty}\sum_{i=1}^{\infty}\left[(c_i-1)^2\theta_i^2 + c_i^2\sigma^2\right]$$

To find the optimal c, we set derivative w.r.t. each c_i to zero:

$$2(c_i-1)\theta_i^2 + 2c_i\sigma^2 = 0 \quad\Longrightarrow\quad c_i = \frac{\theta_i^2}{\theta_i^2+\sigma^2}$$

Substitute this c_i into the equation, we have

$$\begin{aligned}
\inf_{c\in[0,1]^\infty}\sup_{\theta\in\Theta}\mathbb{E}\left[\sum_{i=1}^{\infty}(c_iY_i-\theta_i)^2\right] &= \sup_{\theta\in\Theta}\sum_{i=1}^{\infty}\left[\left(\frac{\theta_i^2}{\theta_i^2+\sigma^2}-1\right)^2\theta_i^2 + \left(\frac{\theta_i^2}{\theta_i^2+\sigma^2}\right)^2\sigma^2\right]\\
&= \sup_{\theta\in\Theta}\sum_{i=1}^{\infty}\left[\left(\frac{\sigma^2}{\theta_i^2+\sigma^2}\right)^2\theta_i^2 + \left(\frac{\theta_i^2}{\theta_i^2+\sigma^2}\right)^2\sigma^2\right]\\
&= \sup_{\theta\in\Theta}\sum_{i=1}^{\infty}\frac{\sigma^2\theta_i^2(\sigma^2+\theta_i^2)}{(\theta_i^2+\sigma^2)^2} = \sup_{\theta\in\Theta}\sum_{i=1}^{\infty}\frac{\sigma^2\theta_i^2}{\theta_i^2+\sigma^2}
\end{aligned}$$

Now, this becomes a constrained (on the set Θ) optimization problem. We introduce *Lagrangian multiplier*

$$L(\theta) = \sum_{i=1}^{\infty}\frac{\sigma^2\theta_i^2}{\theta_i^2+\sigma^2} - \frac{1}{\lambda^2}\left(\sum_{i=1}^{\infty}a_i^2\theta_i^2 - M\right)$$

Here we use $1/\lambda^2$ instead of λ just for some simplicity of representation later. Reparametrize by $\omega_i = \theta_i^2$ and set the derivative to zero:

$$\frac{\partial L(\omega)}{\partial\omega_i} = \frac{\sigma^2(\sigma^2+\omega_i)-\sigma^2\omega_i}{(\sigma^2+\omega_i)^2} - \frac{1}{\lambda^2}a_i^2 = \frac{\sigma^4}{(\sigma^2+\omega_i)^2} - \frac{1}{\lambda^2}a_i^2 = 0$$

$$\iff \frac{\sigma^2}{\sigma^2+\omega_i} = \frac{a_i}{\lambda} \quad \iff \quad \lambda\sigma^2 = a_i\sigma^2 + a_i\omega_i \quad \iff \quad \omega_i = \frac{\lambda\sigma^2 - a_i\sigma^2}{a_i} = \left(\frac{\lambda}{a_i} - 1\right)\sigma^2$$

Since $\omega_i \geqslant 0$, typically we choose

$$\theta_i^{*2} = \left(\frac{\lambda}{a_i} - 1\right)_+ \sigma^2 \tag{1}$$

Note that we directly take derivative of $L(\theta)$ since $\frac{\sigma^2\theta_i^2}{\theta_i^2+\sigma^2}$ is an increasing function w.r.t. each ω_i, thus the constraint (sobolev ball) is in active. Therefore, we choose λ at the boundary:

$$\lambda^* : \sum_{i=1}^{\infty} a_i^2 \left(\frac{\lambda^*}{a_i} - 1\right)_+ \sigma^2 = \sum_{i=1}^{\infty} a_i \left(\lambda^* - a_i\right)_+ \sigma^2 = M \tag{2}$$

The corresponding procedure is then:

$$c_i^* = \frac{\theta_i^{*2}}{\theta_i^{*2}+\sigma^2} = \frac{\left(\frac{\lambda^*}{a_i} - 1\right)_+ \sigma^2}{\left(\frac{\lambda^*}{a_i} - 1\right)_+ \sigma^2 + \sigma^2} = \begin{cases} 1 - \dfrac{a_i}{\lambda^*}, & a_i \leqslant \lambda^* \\ 0, & a_i > \lambda^* \end{cases} \tag{3}$$

The cooresponding best linear risk is then:

$$R_L^*(\Theta) = \sum_{i=1}^{\infty} \frac{\theta_i^{*2}\sigma^2}{\sigma^2+\theta_i^{*2}} = \sum_{i=1}^{\infty} c_i^*\sigma^2 \tag{4}$$

where the subscription 'L' means 'linear'.

1.2.1 Pinsker Upper Bound

(HW2 Solution:)

Solving λ^*

Since in sobolev ellipsoid, we make $a_i \to \infty$ for compactness, there must be some point such that $a_i > \lambda^*$, then $c_i^* = 0$ by formula, and the best linear procedure terminates. Therefore, we don't actually need to calculate the infinite sum when solving λ^*.

For example, if we recover $a_i = i^\alpha$ and $\sigma^2 = 1/n$, then

$$\begin{aligned} &\sum_{i=1}^{\infty} i^\alpha \left(\lambda^* - i^\alpha\right)_+ \frac{1}{n} = \sum_{i=1}^{I^*} i^\alpha (\lambda^* - i^\alpha)_+ \frac{1}{n} = M, \quad (I^*)^\alpha \leqslant \lambda^* < (I^*+1)^\alpha \\ \Longrightarrow &\lambda^* \left(\sum_{i=1}^{I^*} i^\alpha\right) - \sum_{i=1}^{I^*} i^{2\alpha} = Mn \end{aligned}$$

Use the integration to approximate the sum. Note that when I^* is sufficiently large,

$$\sum_{i=1}^{I^*} i^\alpha = \frac{1}{\alpha+1}(I^*)^{\alpha+1}(1+o(1))$$

by Euler-Maclaurin. Choose $\lambda^* = (I^*)^\alpha(1+o(1))$ (since they have at most difference $(I^*+1)^\alpha - (I^*)^\alpha$, and this difference cancelled out the highest order term), we have

$$\begin{aligned}
&\left[\lambda^* \frac{1}{\alpha+1}(I^*)^{\alpha+1} - \frac{1}{2\alpha+1}(I^*)^{2\alpha+1}\right](1+o(1)) = Mn \\
\implies \quad & (I^*)^{2\alpha+1}\left(\frac{1}{\alpha+1} - \frac{1}{2\alpha+1}\right) = Mn(1+o(1)) \\
\implies \quad & I^* = \left(\frac{Mn}{\frac{1}{\alpha+1} - \frac{1}{2\alpha+1}}\right)^{\frac{1}{2\alpha+1}}(1+o(1)) = \left(\frac{Mn(\alpha+1)(2\alpha+1)}{\alpha}\right)^{\frac{1}{2\alpha+1}}(1+o(1))
\end{aligned}$$

Therefore, the best linear procedure risk is

$$\begin{aligned}
R_L^*(\Theta) &= \frac{1}{n}\sum_{i=1}^{\infty} c^* = \frac{1}{n}\sum_{i=1}^{I^*}\left(1 - \frac{i^\alpha}{\lambda^*}\right) \\
&= \frac{1}{n}\left(I^* - \frac{1}{\lambda^*}\sum_{i=1}^{I^*} i^\alpha\right) \\
&= \frac{1}{n}\left(I^* - \frac{1}{(I^*)^\alpha}\frac{1}{\alpha+1}(I^*)^{\alpha+1}(1+o(1))\right) \\
&= \frac{1}{n}\frac{\alpha}{\alpha+1} I^*(1+o(1)) \\
&= \frac{1}{n}\frac{\alpha}{\alpha+1}\left(\frac{Mn(\alpha+1)(2\alpha+1)}{\alpha}\right)^{\frac{1}{2\alpha+1}}(1+o(1)) \\
&= n^{-\frac{2\alpha}{2\alpha+1}}\alpha^{\frac{2\alpha}{2\alpha+1}}(\alpha+1)^{-\frac{2\alpha}{2\alpha+1}}(2\alpha+1)^{\frac{1}{2\alpha+1}}M^{\frac{1}{2\alpha+1}}(1+o(1)) \\
&= \left(\frac{\alpha}{\alpha+1}\right)^{\frac{2\alpha}{2\alpha+1}}[(2\alpha+1)M]^{\frac{1}{2\alpha+1}}(1+o(1))n^{-\frac{2\alpha}{2\alpha+1}} \\
&= c_p(1+o(1))n^{-\frac{2\alpha}{2\alpha+1}}
\end{aligned}$$

where

$$c_p = \left(\frac{\alpha}{\alpha+1}\right)^{\frac{2\alpha}{2\alpha+1}}[(2\alpha+1)M]^{\frac{1}{2\alpha+1}}$$

is the *Pinsker constant*. Therefore, we get the minimax upper bound: For any $\hat{\theta}$, linear or nonlinear, we have

$$R_N(\Theta) = \inf_{\hat{\theta}}\sup_{\theta\in\Theta}\mathbb{E}_{Y|\theta}\left[\|\hat{\theta}-\theta\|_2^2\right] \leqslant R_L^*(\Theta) = c_p(1+o(1))n^{-\frac{2\alpha}{2\alpha+1}}$$

1.2.2 Pinsker Lower Bound

A lower bound is much more difficult. One typically constructs a prior, and use the average risk to bound. What prior should we choose? Naively, we can just choose the Gaussian prior as our first trial.

Naive attempt: Gaussian prior π

Let $\theta_i \sim N(0, \tau_i^2)$, where θ_i's are independent.

Bibliography

[1] Tsybakov, A. B. (2009). Nonparametric estimators. *Introduction to Nonparametric Estimation*, pages 1–76.