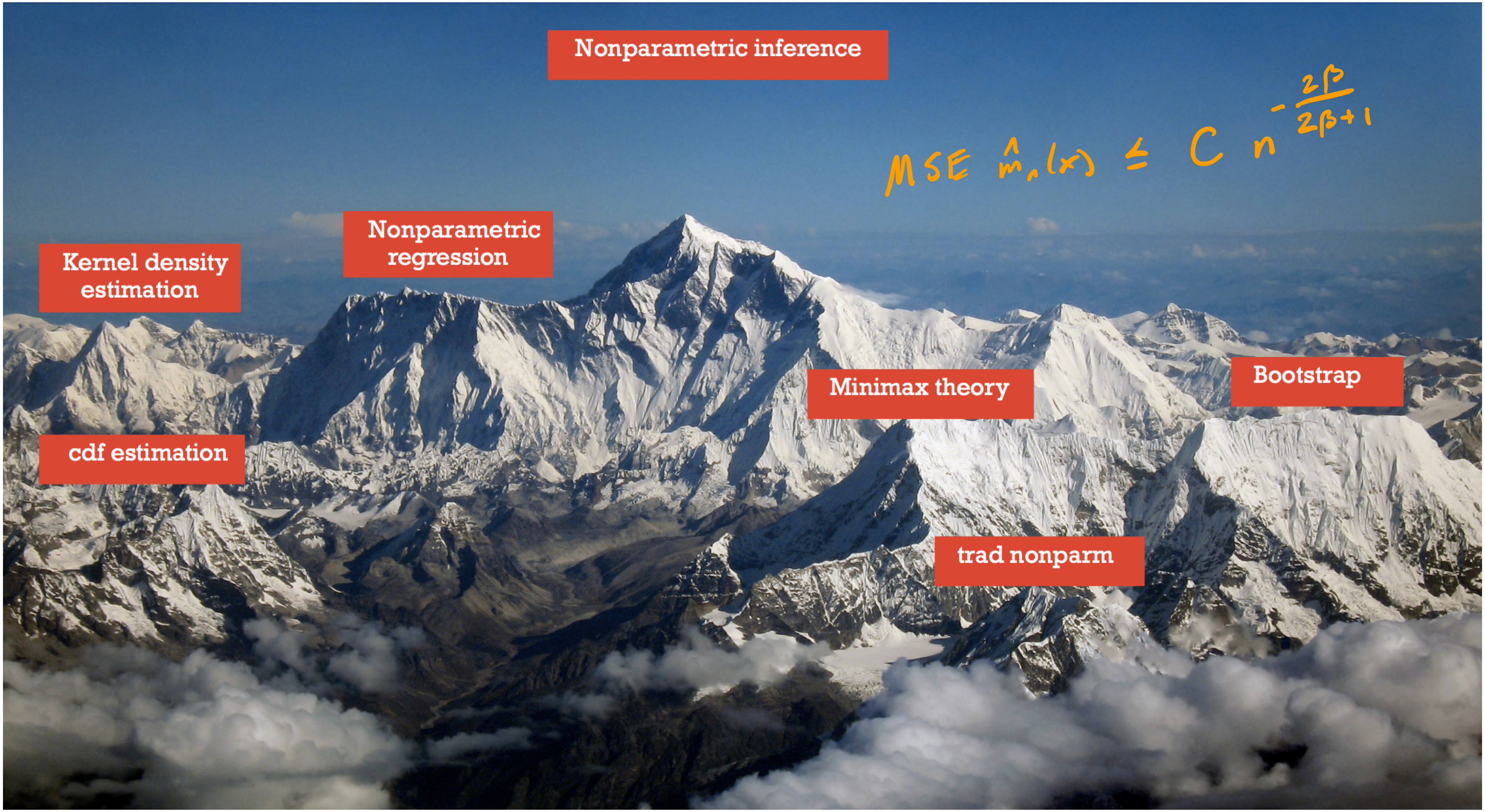
STAT 824 sp 2025 Lec 7 slides

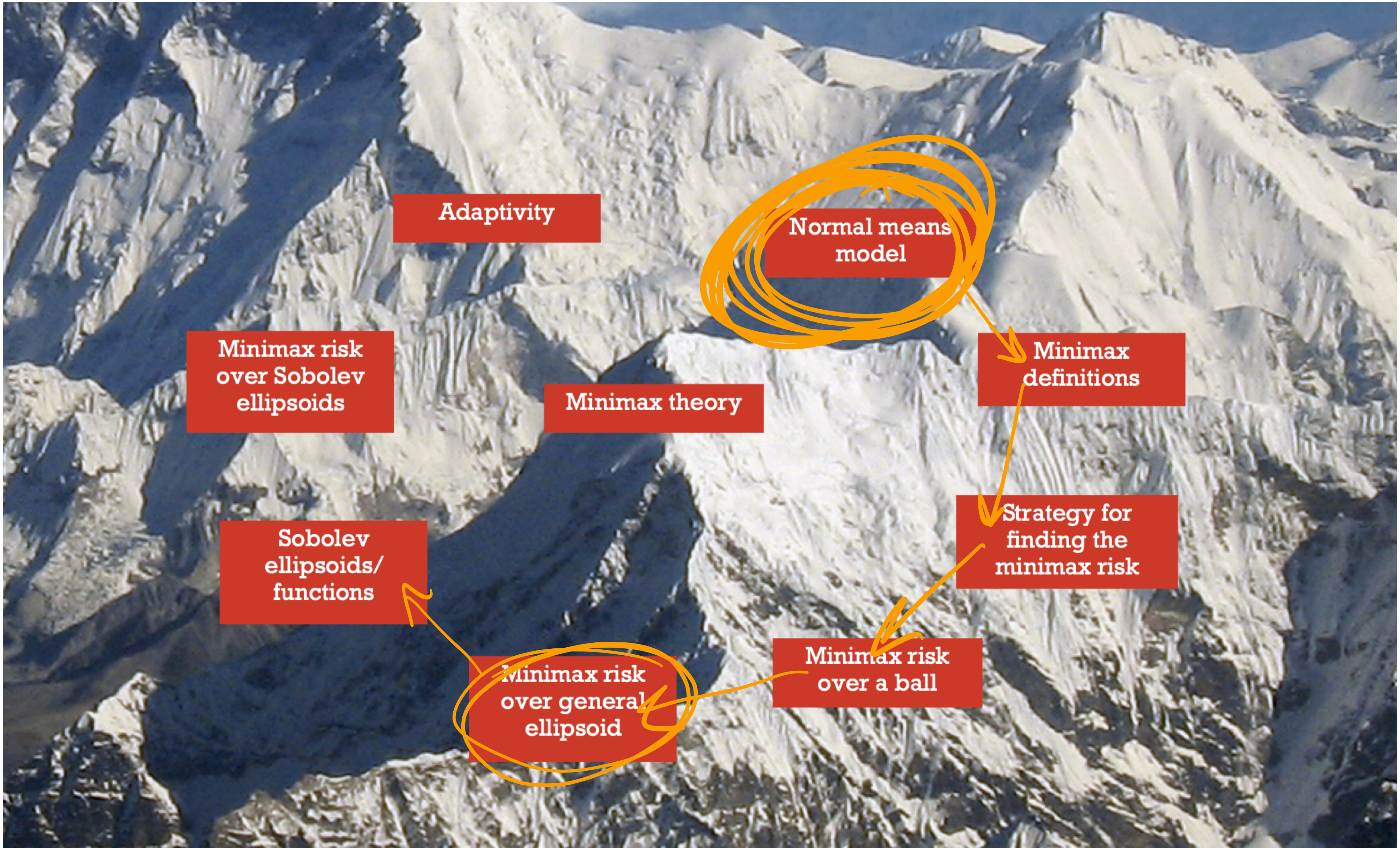
Minimax theory

Karl Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.


Nonparametric inference
$MSE\ \hat{m}_n(x) \leq C n^{-\frac{2\beta}{2\beta+1}}$
Nonparametric regression
Kernel density estimation
Minimax theory
Bootstrap
cdf estimation
trad nonparm


Adaptivity
Normal means model
Minimax risk over Sobolev ellipsoids
Minimax theory
Minimax definitions
Strategy for finding the minimax risk
Sobolev ellipsoids/ functions
Minimax risk over a ball
Minimax risk over general ellipsoid

Suppose we observe $Y_1, \ldots, Y_n$ arising as

$$Y_i = m(i/n) + \varepsilon_i, \quad i = 1, \ldots, n$$

where $m : [0,1] \to \mathbb{R}$ and $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2)$, for some $\sigma^2 > 0$.

Set design points $x_i = i/n$ for $i = 1, \ldots, n$ for the sake of later convenience.

We consider orthogonal series estimators. . .

These notes closely follow [1].

I want to write $m(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x)$

square integrable functions on $[0,1]$

- Let $L_2([0,1]) = \left\{g : [0,1] \to \mathbb{R} : \int_0^1 |g(x)|^2 dx < \infty\right\}$.
- Let $\varphi_1, \varphi_2, \dots : [0,1] \to \mathbb{R}$ be a collection of functions such that

$\int_0^1 \varphi_j(x)^2 dx = 1$

$\int_0^1 \varphi_i(x)\varphi_j(x)dx = 0$

$$\int_0^1 \varphi_i(x)\varphi_j(x)dx = \begin{cases} 1 & i = j \\ 0 & i \neq j, \end{cases}$$

orthonormal.

for all i and j and such that for any function $g \in L_2([0,1])$

$$\int_0^1 g(x)\varphi_j(x)dx = 0 \quad \forall j \iff g(x) = 0.$$

- Then $\varphi_1, \varphi_2, \dots$ comprise an orthonormal basis for $L_2([0,1])$.

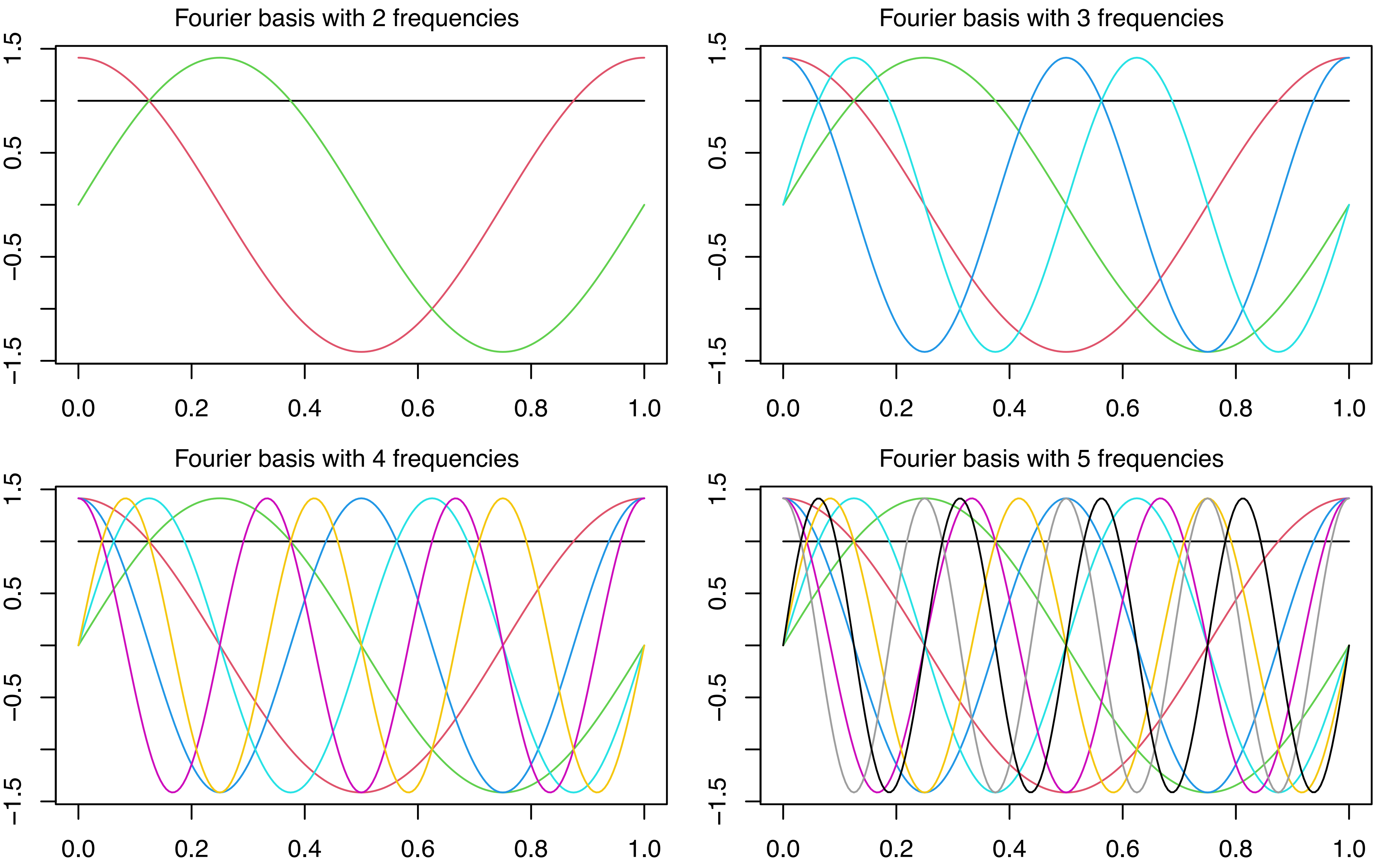
So $g \in L_2([0,1])$ can be written $g(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x)$.

Fourier basis

The Fourier basis on $[0, 1]$ is the collection of functions $\phi_1(x) = 1$ and

$$\begin{aligned} \phi_{2k}(x) &= \sqrt{2}\cos(2\pi kx) \\ \phi_{2k+1}(x) &= \sqrt{2}\sin(2\pi kx) \end{aligned}$$

for $k = 1, 2, \dots$


Fourier basis with 2 frequencies
Fourier basis with 3 frequencies
Fourier basis with 4 frequencies
Fourier basis with 5 frequencies
−1.5
−0.5
0.5
1.5
0.0
0.2
0.4
0.6
0.8
1.0

Think of vectors: $\underset{\sim}{g} \in \mathbb{R}^N$, $\underset{\sim}{\varphi}_1, \ldots, \underset{\sim}{\varphi}_N \in \mathbb{R}^N$, $\underset{\sim}{\varphi}_j^T \underset{\sim}{\varphi}_{j'} = \begin{cases} 1 & j = j' \\ 0 & j \neq j' \end{cases}$

and $\underset{\sim}{g}^T \underset{\sim}{\varphi}_j = 0 \ \forall \ j = 1, \ldots, N \iff \underset{\sim}{g} = \underset{\sim}{0} \in \mathbb{R}^N$, then $(\underset{\sim}{\varphi}_1, \ldots, \underset{\sim}{\varphi}_N)$ is a basis for $\mathbb{R}^N$.

I can write

$$\underset{\sim}{g} = \sum_{j=1}^N \underset{\sim}{\varphi}_j \underbrace{(\underset{\sim}{\varphi}_j^T \underset{\sim}{\varphi}_j)^{-1}}_{1} \underset{\sim}{\varphi}_j^T \underset{\sim}{g}$$

$$= \sum_{j=1}^N \underset{\sim}{\varphi}_j \underset{\sim}{\varphi}_j^T \underset{\sim}{g}$$

$$= \sum_{j=1}^N \theta_j \underset{\sim}{\varphi}_j,$$

where $\theta_j = \underset{\sim}{\varphi}_j^T \underset{\sim}{g}$.

- For $g \in L_2([0,1])$ we may write

any

$$g(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x),$$

where $\boxed{\theta_j = \int_0^1 g(x)\varphi_j(x)dx,}$ $j = 1, 2, \ldots$

- Moreover we have $\int_0^1 |g(x)|^2 dx = \sum_{j=1}^{\infty} \theta_j^2$, which is called Parseval's identity.

$$\int_0^1 (g(x))^2 dx = \int_0^1 \left(\sum_{j=1}^{\infty} \theta_j \varphi_j(x) \right)^2 dx = \int_0^1 \left(\sum_{j=1}^{\infty} \theta_j^2 \varphi_j^2(x) + \sum_{j \neq j'} \theta_j \theta_{j'} \varphi_j(x) \varphi_{j'}(x) \right) dx$$

$$= \sum_{j=1}^{\infty} \theta_j^2.$$

$Y_i = m(i/n) + \varepsilon_i, \quad i = 1, \dots, n.$

- Assume $m \in L_2([0,1])$ and suppose

$m(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x)$

$$m(x) \approx \sum_{j=1}^{n} \theta_j \varphi_j(x)$$

for some $\theta_1, \dots, \theta_n$ and orth. $\varphi_1, \dots, \varphi_n$.

- Consider estimating $\theta_1, \dots, \theta_n$ based on $Y_1, \dots, Y_n$ with

$\theta_j = \int_0^1 m(x) \varphi_j(x) dx$

$$\boxed{\hat{\theta}_j = \frac{1}{n} \sum_{i=1}^{n} Y_i \varphi_j(i/n),} \quad j = 1, \dots, n.$$

- Then set $\boxed{\hat{m}_n(x) = \sum_{j=1}^{n} \hat{\theta}_j \varphi_j(x).}$

Normal means model.

Exercise: Show that estimating m is approximately like estimating $\theta_1, \ldots, \theta_n$ in

$$Z_j = \theta_j + n^{-1/2}\sigma\xi_j, \quad j = 1, \ldots, n,$$

where $\xi_1, \ldots, \xi_n \overset{\text{ind}}{\sim} \mathcal{N}(0,1)$. Consider estimating each θ_j and the MISE of $\hat{m}_n$.

$$\hat{\theta}_j = \frac{1}{n}\sum_{i=1}^{n} Y_i \, \varphi_j(i/n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \left(m(i/n) + \varepsilon_i\right) \varphi_j(i/n)$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n} m(i/n)\,\varphi_j(i/n)}_{\int_0^1 m(x)\varphi_j(x)\,dx \;\approx\; \theta_j} + \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i\,\varphi_j(i/n)$$

$$\approx \theta_j + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i\,\varphi_j(i/n)}_{} \overset{\text{approx}}{\sim} N\left(0, \frac{\sigma^2}{n}\right)$$

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\,\varphi_j(i/n)\right) = \frac{1}{n^2}\sum_{i=1}^{n}\varphi_j^2(i/n)\,\sigma^2$$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_j^2(i/n) \approx \int_0^1 \varphi_j^2(x)\,dx = 1 \qquad \approx \frac{\sigma^2}{n}$$

$\Rightarrow$

$$\hat\theta_j \overset{D}{\approx} \theta_j + n^{-1/2}\sigma\,\xi_j, \qquad \xi_j \sim N(0,1).$$

$$\text{MISE } \hat m_n = \mathbb{E}\int_0^1 \left[\hat m_n(x) - m(x)\right]^2 dx$$

$$= \mathbb{E}\int_0^1 \left[\sum_{j=1}^{n}\hat\theta_j\varphi_j(x) - \sum_{j=1}^{\infty}\theta_j\varphi_j(x)\right]^2 dx$$

$$\approx \mathbb{E}\int_0^1 \left[\sum_{j=1}^{n}\hat\theta_j\varphi_j(x) - \sum_{j=1}^{n}\theta_j\varphi_j(x)\right]^2 dx$$

$$= \mathbb{E}\int_0^1 \sum_{j=1}^{n}(\hat\theta_j - \theta_j)^2\,\varphi_j^2(x)\,dx$$

$$= \sum_{j=1}^{n}\mathbb{E}(\hat\theta_j - \theta_j)^2$$

$Y_i = m(i/n) + \varepsilon_i$, $m(x) \approx \sum_{j=1}^{n} \theta_j \varphi_j(x)$, $\hat{\theta}_j = \frac{1}{n}\sum_{i=1}^{n} Y_i \varphi_j(i/n)$

$j=1,\dots,n$.

unknown

so $MISE\ \hat{m}_n(x) = \mathbb{E}\int [\hat{m}_n(x) - m(x)]^2 dx \approx \sum_{j=1}^{n} \mathbb{E}(\hat{\theta}_j - \theta_j)^2$,

The Normal means model

$\left(\hat{\theta}_j \overset{\text{approx}}{\sim} N\left(\theta_j, \frac{\sigma^2}{n}\right)\right)$

Observe $Z_1, \dots, Z_n$ from

$$Z_i = \theta_i + n^{-1/2}\sigma\xi_i, \quad i = 1, \dots, n,$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in \Theta \subset \mathbb{R}^n$, $\xi_1, \dots, \xi_n \overset{\text{ind}}{\sim} \mathcal{N}(0, 1)$, and $\sigma > 0$.

We ask how well we can estimate $\boldsymbol{\theta}$ depending on the space Θ.

Want to estimate $\underset{\sim}{\theta} \in \Theta \subset \mathbb{R}^n = (-\infty, \infty)^n$

- *Loss (squared error)*: Represents the cost of estimation error.

$$L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = \sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2.$$

- *Risk*: The risk is the expected value of the loss.

$$R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \mathbb{E} L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^n \mathbb{E}(\hat{\theta}_i - \theta_i)^2.$$

- *Minimax risk*: The best worst performance of any estimator of $\boldsymbol{\theta} \in \Theta$.

$$M(\Theta) = \inf_{\hat{\boldsymbol{\theta}}} \sup_{\boldsymbol{\theta} \in \Theta} R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}),$$

worst-case risk

where the infimum is taken over all estimators.

π is a distribution on $\mathbb{R}^n$

- *Integrated risk*: with respect to a prior π is defined as

$$I_\pi(\hat{\boldsymbol{\theta}}) = \int_{\mathbb{R}^n} R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}).$$

- *Bayes estimator*: under the prior π is defined as

$$\hat{\boldsymbol{\theta}}_\pi = \underset{\hat{\boldsymbol{\theta}}}{\operatorname{argmin}} \int_{\mathbb{R}} R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}),$$

Note: Under squared loss, the Bayes estimator is the posterior mean!

- *Integrated Bayes risk*: The integrated risk of the Bayes estimator

$$I_\pi = I_\pi(\hat{\boldsymbol{\theta}}_\pi) = \int_{\mathbb{R}^n} R(\hat{\boldsymbol{\theta}}_\pi, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}).$$

$M(\Theta) = \inf_{\hat\theta} \boxed{\sup_{\theta\in\Theta} R(\hat\theta,\theta)}$

"Average cannot exceed the maximum"

- For any estimator $\hat{\theta}$ we may write

average risk over Θ according π

$$
\begin{aligned}
\boxed{\sup_{\theta\in\Theta} R(\hat{\theta},\theta)} &\geq \int_{\Theta} R(\hat{\theta},\theta)d\pi(\theta) \\
&= \int_{\mathbb{R}^n} R(\hat{\theta},\theta)d\pi(\theta) - \int_{\Theta^c} R(\hat{\theta},\theta)d\pi(\theta) \\
&\geq I_\pi - \int_{\Theta^c} R(\hat{\theta},\theta)d\pi(\theta),
\end{aligned}
$$

Worst-case risk for estimator $\hat\theta$

Bayes estimator minimizes this term.

$$\int_{\mathbb{R}^n} R(\hat\theta,\theta)d\pi(\theta) = \int_{\Theta} R(\hat\theta,\theta)d\pi(\theta) + \int_{\Theta^c} R(\hat\theta,\theta)d\pi(\theta)$$

- Taking the infimum of both sides of the above over all estimators $\hat{\theta}$ gives

$$\inf_{\hat\theta} \sup_{\theta\in\Theta} R(\hat\theta,\theta) = M(\Theta) \geq \underbrace{I_\pi - \sup_{\hat{\theta}} \int_{\Theta^c} R(\hat{\theta},\theta)d\pi(\theta),}_{\text{Always a lower bound for } M(\Theta).}$$

- Second term vanishes under a seq. of priors that concentrates on Θ.

$$\inf_{\hat\theta} \sup_{\theta\in\Theta} R(\hat\theta,\theta) \leq \sup_{\theta\in\Theta} R(\hat\theta,\theta) \leq M^*$$

Strategy for finding the minimax risk

To find the minimax risk $M(\Theta)$, propose a candidate value M^* and then:

1. Find an estimator with worst-case risk equal to (or $\leq$) M^*. Shows $M(\Theta) \leq M^*$.
2. Find a prior (or a sequence of priors) such that the integrated Bayes risk over Θ is equal to (or converges to) M^*. Shows $M(\Theta) \geq M^*$.

Steps 1 and 2 give $M(\Theta) = M^*$.

Define the ball with radius c centered at the origin

$$\Theta_n(c) = \left\{ \boldsymbol{\theta} \in \mathbb{R}^n : \|\boldsymbol{\theta}\| \leq c \right\}.$$

Two results

In the Normal means model we have the minimax risks

1. $M(\mathbb{R}^n) = \sigma^2$ $\quad \Theta = \mathbb{R}^n \quad Z_j = \theta_j + \frac{\sigma}{\sqrt{n}} \xi_j, \quad j = 1, \dots, n.$
2. $\liminf_{n \to \infty} M(\Theta_n(c)) = \dfrac{\sigma^2 c^2}{\sigma^2 + c^2} = \sigma^2 \left(\dfrac{c^2}{\sigma^2 + c^2} \right)$

Exercise: Work through the proofs of the above.

$Z_j = \theta_j + \frac{\sigma}{\sqrt{n}} \xi_j \quad , j = 1, \dots, n, \qquad \xi_j \sim N(0,1)$

(1) **Step 1:** Propose $M^* = \sigma^2$.

Find an estimator such that $\sup_{\theta \in \Theta} R(\hat{\theta}, \theta) \overset{(\le)}{=} M^*$

Try $\hat{\underset{\sim}{\theta}} = \underset{\sim}{Z} = (Z_1, \dots, Z_n)^T$.

Then
$$R(\underset{\sim}{Z}, \underset{\sim}{\theta}) = \mathbb{E} \| \underset{\sim}{Z} - \underset{\sim}{\theta} \|^2$$
$$= \mathbb{E} \sum_{j=1}^n (Z_j - \theta_j)^2$$
$$= \sum_{j=1}^n \operatorname{Var} Z_j$$
$$= \sum_{j=1}^n \frac{\sigma^2}{n}$$
$$= \sigma^2 .$$

So
$$\sup_{\theta \in \underset{\underset{\Theta = \mathbb{R}^n}{\uparrow}}{\Theta}} R(\underset{\sim}{Z}, \underset{\sim}{\theta}) = \sigma^2 .$$

Therefore

$$\underset{\underset{\mathbb{R}^n}{\uparrow}}{M(\Theta)} = \inf_{\hat{\theta}} \sup_{\theta \in \underset{\underset{\mathbb{R}^n}{\uparrow}}{\Theta}} R(\underset{\sim}{Z}, \underset{\sim}{\theta}) \le \sup_{\theta \in \Theta} R(\underset{\sim}{Z}, \underset{\sim}{\theta}) = \sigma^2 .$$

So
$$M(\underset{\underset{\mathbb{R}^n}{\uparrow}}{\Theta}) \le \sigma^2 .$$

Step 2: $\pi_n = N(\underset{\sim}{0}, n^{-1}\tau^2 I_n)$, some $\tau^2 > 0$.

↑ Has support on $\mathbb{R}^n$.

Now we need I_{π_n}, the integrated Bayes Risk.

Since $\underset{\sim}{Z} \mid \underset{\sim}{\theta} \sim N(\underset{\sim}{\theta}, n^{-1}\sigma^2 I_n)$, ...

$\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, n^{-1}\tau^2 I_n)$

steps :

$$\hat{\theta}_{\pi_n} = \mathbb{E}[\underset{\sim}{\theta} \mid \underset{\sim}{Z}] = \frac{\tau^2}{\tau^2+\sigma^2}\underset{\sim}{Z} + \frac{\sigma^2}{\tau^2+\sigma^2}\underset{\sim}{0}$$

We have

$$\begin{aligned}
R(\hat{\underset{\sim}{\theta}}_{\pi_n}, \underset{\sim}{\theta}) &= \mathbb{E}\left[\|\hat{\underset{\sim}{\theta}}_{\pi_n} - \underset{\sim}{\theta}\|^2 \,\middle|\, \underset{\sim}{\theta}\right] \\
&= \mathbb{E}\left[\left\|\frac{\tau^2}{\tau^2+\sigma^2}\underset{\sim}{Z} - \underset{\sim}{\theta}\right\|^2 \,\middle|\, \underset{\sim}{\theta}\right] \\
&= \mathbb{E}\left[\left\|\frac{\tau^2}{\tau^2+\sigma^2}(\underset{\sim}{Z} - \underset{\sim}{\theta}) - \left(1 - \frac{\tau^2}{\tau^2+\sigma^2}\right)\underset{\sim}{\theta}\right\|^2 \,\middle|\, \underset{\sim}{\theta}\right] \\
&= \sum_{j=1}^n \mathbb{E}\left\{\left[\frac{\tau^2}{\tau^2+\sigma^2}(Z_j - \theta_j) - \left(\frac{\sigma^2}{\tau^2+\sigma^2}\right)\theta_j\right]^2 \,\middle|\, \underset{\sim}{\theta}\right\} \\
&= \sum_{j=1}^n \left\{\left(\frac{\tau^2}{\tau^2+\sigma^2}\right)^2 \mathbb{E}\left[(Z_j - \theta_j)^2 \mid \underset{\sim}{\theta}\right] + \left(\frac{\sigma^2}{\tau^2+\sigma^2}\right)^2 \theta_j^2\right\}
\end{aligned}$$

$$= \left(\frac{\tau^2}{\tau^2+\sigma^2}\right)^2 \sigma^2 + \left(\frac{\sigma^2}{\tau^2+\sigma^2}\right)^2 \|\underset{\sim}{\theta}\|^2$$

Now get integrated Bayes' Risk:

$$I_{\pi_n} = \int_{\mathbb{R}^n} R\left(\hat{\underset{\sim}{\theta}}_{\pi_n}, \underset{\sim}{\theta}\right) d\pi_n(\underset{\sim}{\theta})$$

$$= \int_{\mathbb{R}^n} \left[\left(\frac{\tau^2}{\tau^2+\sigma^2}\right)^2 \sigma^2 + \left(\frac{\sigma^2}{\tau^2+\sigma^2}\right)^2 \|\underset{\sim}{\theta}\|^2\right] d\pi(\underset{\sim}{\theta})$$

$$= \left(\frac{\tau^2}{\tau^2+\sigma^2}\right)^2 \sigma^2 + \left(\frac{\sigma^2}{\tau^2+\sigma^2}\right)^2 \underbrace{\int_{\mathbb{R}^n} \|\underset{\sim}{\theta}\|^2 d\pi(\theta)}_{\int_{\mathbb{R}^n} \sum_{j=1}^n \theta_j^2 \, d\pi(\theta) = \sum_{j=1}^n \mathbb{E}\,\theta_j^2 = \sum_{j=1}^n \frac{\tau^2}{n} = \tau^2}$$

$$= \left(\frac{\tau^2}{\tau^2+\sigma^2}\right)^2 \sigma^2 + \left(\frac{\sigma^2}{\tau^2+\sigma^2}\right)^2 \tau^2$$

$$\frac{\tau^4\sigma^2 + \sigma^4\tau^2}{(\tau^2+\sigma^2)^2}$$

$$= \frac{\tau^2\sigma^2(\tau^2+\sigma^2)}{(\tau^2+\sigma^2)^2}$$

$$= \frac{\tau^2\sigma^2}{\tau^2+\sigma^2}$$

"squeeze!"

So we have $\underbrace{\left(\frac{\tau^2}{\tau^2+\sigma^2}\right)}_{<1 \ \forall \tau^2>0.}\sigma^2 \leq M(\Theta) \leq \sigma^2 \quad \forall \tau^2 > 0.$

$\Rightarrow \quad M(\Theta) = \sigma^2.$

② Let $\quad \Theta_n(c) = \left\{ (\theta_1, \dots, \theta_n) \in \mathbb{R}^n : \sum_{j=1}^{n} \theta_j^2 \le c^2 \right\}$

$\left(\|\underset{\sim}{\theta}\| \le c \right)$

Step 1: Propose $\quad M^* = \dfrac{\sigma^2 c^2}{\sigma^2 + c^2}$.

Find an estimator $\hat{\underset{\sim}{\theta}}$ for which $\sup_{\theta \in \Theta} R(\hat{\underset{\sim}{\theta}}, \underset{\sim}{\theta}) \overset{\le}{=} \dfrac{\sigma^2 c^2}{\sigma^2 + c^2}$.

Consider estimators of the form $\hat{\underset{\sim}{\theta}}_\lambda = \lambda \underset{\sim}{Z} = (\lambda Z_1, \dots, \lambda Z_n)^T$ some $\lambda \in \mathbb{R}$.

Find λ such that $\sup_{\theta \in \Theta} R(\hat{\theta}_\lambda, \theta) = \dfrac{\sigma^2 c^2}{\sigma^2 + c^2}$?

First

$$
\begin{aligned}
R(\hat{\theta}_\lambda, \theta) &= \mathbb{E} \|\hat{\theta}_\lambda - \theta\|^2 \\
&= \sum_{j=1}^{n} \mathbb{E}(\lambda Z_j - \theta_j)^2 \\
&= \sum_{j=1}^{n} \mathbb{E}\left[\lambda (Z_j - \theta_j) - (1-\lambda)\theta_j\right]^2 \\
&= \sum_{j=1}^{n} \lambda^2 \overset{\sigma^2/n}{\mathrm{Var}\, Z_j} + (1-\lambda)^2 \sum_{j=1}^{n} \theta_j^2 \\
&= \lambda^2 \sigma^2 + (1-\lambda)^2 \|\underset{\sim}{\theta}\|^2 .
\end{aligned}
$$

Then
$$\sup_{\theta \in \Theta} R(\hat{\theta}_\lambda, \theta) = \lambda^2 \sigma^2 + (1-\lambda)^2 c^2 .$$

Consider
$$\inf_{\lambda \in \mathbb{R}} \sup_{\theta \in \Theta} R(\hat{\underset{\sim}{\theta}}_\lambda, \underset{\sim}{\theta}) = \inf_{\lambda \in \mathbb{R}} \left\{ \lambda^2 \sigma^2 + (1-\lambda)^2 c^2 \right\}.$$

We have
$$\frac{\partial}{\partial \lambda} \sup_{\theta \in \Theta} R(\hat{\underset{\sim}{\theta}}_\lambda, \underset{\sim}{\theta}) = 2\lambda\sigma^2 - 2(1-\lambda)c^2 \overset{\text{set}}{=} 0$$
$$\Rightarrow \lambda(\sigma^2 + c^2) - c^2 = 0$$
$$\Rightarrow \lambda = \frac{c^2}{\sigma^2 + c^2} .$$

plug in

$$\inf_{\lambda \in \mathbb{R}} \sup_{\theta \in \Theta} R(\hat{\underset{\sim}{\theta}}_\lambda, \underset{\sim}{\theta}) = \left(\frac{c^2}{\sigma^2 + c^2}\right)\sigma^2 + \left(\frac{\sigma^2}{\sigma^2 + c^2}\right)c^2$$
$$= \frac{\sigma^2 c^2}{\sigma^2 + c^2} .$$

So... the estimator $\hat{\underset{\sim}{\theta}}_\lambda = \left(\frac{c^2}{\sigma^2 + c^2}\right) \underset{\sim}{Z}$ has worst-case risk equal to $M^* = \dfrac{\sigma^2 c^2}{\sigma^2 + c^2}$.

$$\Rightarrow \quad M(\theta) \le \frac{\sigma^2 c^2}{\sigma^2 + c^2}.$$

<u>Step 2</u>: Find lower bound for $M(\theta)$.

$$\underset{\sim}{z} \mid \underset{\sim}{\theta} \sim N(\underset{\sim}{\theta}, n^{-1}\sigma^2 I_n)$$

$$\underset{\sim}{\theta} \sim N(\underset{\sim}{0}, n^{-1}\tau^2 c^2 I_n) \leftarrow \text{choose this prior.}$$

Then .... $$\mathbb{E}[\underset{\sim}{\theta} \mid \underset{\sim}{z}] = \frac{\tau^2 c^2}{\sigma^2 + \tau^2 c^2} \underset{\sim}{z} + \frac{\sigma^2}{\sigma^2 + \tau^2 c^2} \underset{\sim}{0}$$

$$\Rightarrow \quad \underset{\sim}{\hat{\theta}}_\pi = \frac{\tau^2 c^2}{\sigma^2 + \tau^2 c^2} \underset{\sim}{z}.$$

$$\Rightarrow \quad R(\underset{\sim}{\hat{\theta}}_\pi, \underset{\sim}{\theta}) = \mathbb{E}\left[\left\| \frac{\tau^2 c^2}{\sigma^2 + \tau^2 c^2} \underset{\sim}{z} - \underset{\sim}{\theta} \right\|^2 \middle| \underset{\sim}{\theta} \right]$$

$\vdots$ like previous work

$$= \left(\frac{\tau^2 c^2}{\tau^2 c^2 + \sigma^2} \right)^2 \sigma^2 + \left(\frac{\sigma^2}{\tau^2 c^2 + \sigma^2} \right)^2 \|\underset{\sim}{\theta}\|^2$$

Then

$$I_\pi = \int_{\mathbb{R}^n} R(\underset{\sim}{\hat{\theta}}_\pi, \underset{\sim}{\theta})\, d\pi(\theta) = \left(\frac{\tau^2 c^2}{\tau^2 c^2 + \sigma^2} \right)^2 \sigma^2 + \left(\frac{\sigma^2}{\tau^2 c^2 + \sigma^2} \right)^2 \tau^2 c^2$$

$$= \frac{\tau^2 c^2 \sigma^2}{\tau^2 c^2 + \sigma^2}$$

Now ... $M(\Theta) \ge I_\pi - \sup_{\hat{\boldsymbol{\theta}}} \int_{\Theta^c} R(\hat{\boldsymbol{\theta}}$[redacted]$, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}),$

So

$$M(\Theta) \ge \frac{\tau^2 c^2 \sigma^2}{\tau^2 c^2 + \sigma^2} - \sup_{\hat{\boldsymbol{\theta}}} \int_{\Theta^c} R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta})$$

$\Rightarrow$

$$\frac{\tau^2 c^2 \sigma^2}{\tau^2 c^2 + \sigma^2} - \sup_{\hat{\boldsymbol{\theta}}} \int_{\{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| > c\}} R(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) d\pi(\boldsymbol{\theta}) \le M(\Theta) \le \frac{c^2 \sigma^2}{c^2 + \sigma^2}$$

$\downarrow 0$ as $n \to \infty$

ball

Make $\tau^2 \uparrow 1$

Can we make this disappear?

$$\Rightarrow \liminf_{n \to \infty} M(\Theta_n(c)) = \frac{c^2 \sigma^2}{c^2 + \sigma^2}.$$

Ball

To bring out minimax risk results which look like the nonparametric rates of convergence we have seen, e.g. $n^{-2\beta/(2\beta+1)}$, we consider Sobolev functions...

Sobolev class and periodic Sobolev class

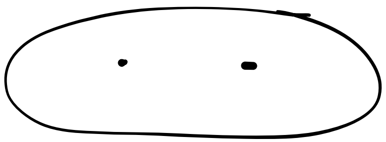
For β a positive integer and $L > 0$, define the *Sobolev class* of functions as

$$\mathcal{W}(\beta, L) = \left\{ m : [0,1] \to \mathbb{R} : \begin{array}{c} m^{(\beta-1)} \text{ is absolutely continuous} \\ \text{and } \int_0^1 (m^{(\beta)}(x))^2 dx \leq L^2 \end{array} \right\}.$$

Moreover, define the *periodic Sobolev class* of functions $W_{\text{per}}(\beta, L)$ as

$$\mathcal{W}_{\text{per}}(\beta, L) = \{m \in W(\beta, L) : m^{(\ell)}(0) = m^{(\ell)}(1) \text{ for } \ell = 0, \ldots, \beta - 1\}.$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \le c .$$



Sobolev ellipsoid

Define the *Sobolev ellipsoid* $\Theta(\beta, c)$ as

$$\Theta_{\text{Sob}}(\beta, L) = \left\{ (\theta_1, \theta_2, \dots) : \sum_{j=1}^{\infty} \theta_j^2 < \infty \text{ and } \sum_{j=1}^{\infty} a_j^2 \theta_j^2 \le L^2/\pi^{2\beta} \right\},$$

where $a_1 = 0$, $a_{2m} = a_{2m+1} = (2m)^\beta$, $m = 1, 2, \dots$

Definition of $a_1, a_2, \dots$ equivalent to

$$a_j = \begin{cases} j^\beta, & j \text{ even} \\ (j-1)^\beta, & j \text{ odd.} \end{cases}$$

Result: Fourier basis as a basis for periodic Sobolev functions

We have

$$\mathcal{W}_{\text{per}}(\beta, L) = \left\{ m : [0,1] \to \mathbb{R} : m(x) = \sum_{j=1}^{\infty} \theta_j \phi_j(x),\ (\theta_1, \theta_2, \dots) \in \Theta_{\text{Sob}}(\beta, L) \right\}.$$

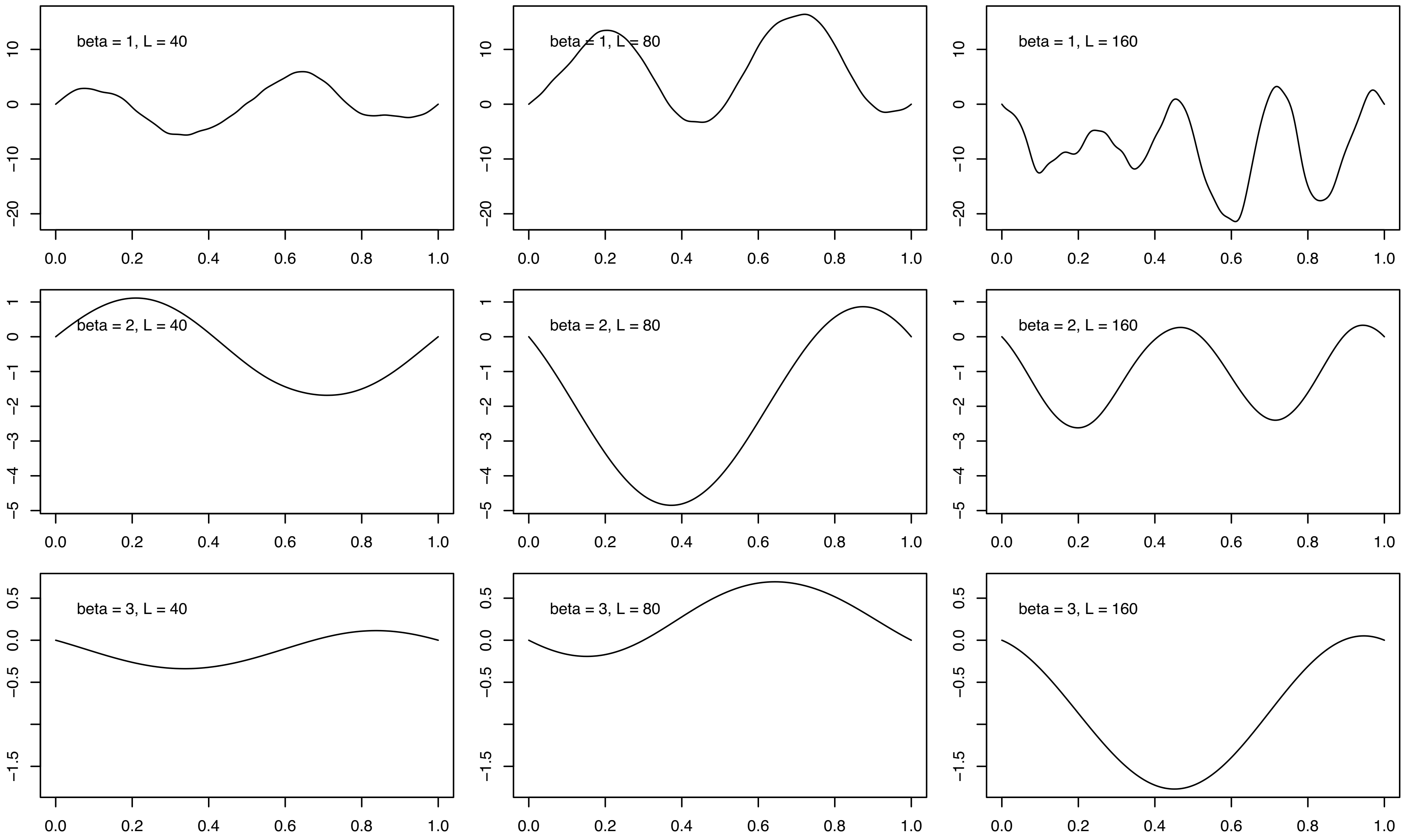
- So we can construct any periodic Sobolev function from the Fourier basis.
- Estimating $m \in \mathcal{W}_{\text{per}}(\beta, L)$ is essentially like estimating $\boldsymbol{\theta} \in \Theta_{\text{Sob}}(\beta, L)$ in the normal Means model.

What do periodic Sobolev functions look like?

Exercise: Can generate some functions belonging to $\mathcal{W}_{\text{per}}(\beta, L)$ in these steps:

1. Draw $\theta_1, \ldots \theta_N \overset{\text{ind}}{\sim} \text{Normal}(0, 1)$ with $N = 50$, say.
2. Minimize $\sum_{i=1}^{N} (\theta_j - w_j)^2$ subject to $\sum_{j=1}^{N} a_j^2 w_j^2 = L^2/\pi^{2\beta}$, where the a_j are those which define the Sobolev ellipsoid $\Theta_{\text{Sob}}(\beta, L)$.
3. Set $m(x) = \sum_{j=1}^{N} \hat{w}_j \phi_j(x)$, where $\hat{w}_1, \ldots, \hat{w}_N$ are from step 2 and $\{\phi_1, \phi_2, \ldots\}$ is the Fourier basis.

Next slide plots several after subtracting $m(0)$ (to aid comparison).


beta = 1, L = 40
beta = 1, L = 80
beta = 1, L = 160
beta = 2, L = 40
beta = 2, L = 80
beta = 2, L = 160
beta = 3, L = 40
beta = 3, L = 80
beta = 3, L = 160
0.0
0.2
0.4
0.6
0.8
1.0

Infinite-dimensional Normal means model

Let

$$Z_j = \theta_j + \sigma\xi_j, \quad j = 1, 2, \ldots,$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$ is unknown, $\xi_1, \xi_2, \ldots, \overset{\text{ind}}{\sim} \mathcal{N}(0, 1)$, and $\sigma > 0$.

Want to find minimax risk for estimating $\boldsymbol{\theta}$ in the above model when

1. $\boldsymbol{\theta}$ lies in a general ellipsoid

$$\Theta(c, a_1, a_2, \ldots) = \left\{(\theta_1, \theta_2, \ldots) \in \mathbb{R} : \sum_{j=1}^{\infty} \theta_j^2 < \infty \text{ and } \sum_{j=1}^{\infty} a_j^2 \theta_j^2 \leq c^2\right\}.$$

2. $\boldsymbol{\theta} \in \Theta_{\text{Sob}}(\beta, L)$.

- Consider only linear estimators, i.e. of the form

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}} = (\lambda_1 Z_1, \lambda_2 Z_2, \dots) \quad \text{for } \lambda = (\lambda_1, \lambda_2, \dots).$$

- Let

$$M_{\text{lin}}(\Theta(c, a_1, a_2, \dots)) = \inf_{\boldsymbol{\lambda}} \sup_{\boldsymbol{\theta} \in \Theta(c, a_1, a_2, \dots)} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}}, \boldsymbol{\theta})$$

denote the *linear minimax risk* over $\Theta(c, a_1, a_2, \dots)$, where

$$R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}}, \boldsymbol{\theta}) = \sum_{j=1}^{\infty} \mathbb{E}(\lambda_j Z_j - \theta_j)^2.$$

Linear minimax risk over a general ellipsoid

Let $a_1, a_2, \ldots$ be an increasing seq. such that $|\{j : a_j = 0\}| < \infty$ and $a_j \to +\infty$. Then a unique solution to

$$\eta^{-1}\sigma^2 \sum_{i=1}^{\infty} a_j(1 - \eta a_j)_+ = c^2 \tag{1}$$

over $\eta > 0$ exists such that, setting $\ell_j = (1 - \eta a_j)_+$ for $j = 1, 2, \ldots$ and $\boldsymbol{\ell} = (\ell_1, \ell_2, \ldots)$, we have

$$M_{\text{lin}}(\Theta(c, a_1, a_2, \ldots)) = \sup_{\boldsymbol{\theta} \in \Theta(c, a_1, a_2, \ldots)} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\ell}}, \boldsymbol{\theta}) = \sigma^2 \sum_{j=1}^{\infty} \ell_j,$$

provided the sum is finite.

The values $\ell_1, \ell_2, \ldots$ are called the *Pinsker weights*.

Exercise: To see why one should consider $\sigma^2 \sum_{j=1}^{\infty} \ell_j$ as a candidate for the linear minimax risk, show the following:

1. We have
$$\inf_{\boldsymbol{\lambda}} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}}, \boldsymbol{\theta}) = \sum_{j=1}^{\infty} \frac{\sigma^2 \theta_j^2}{\sigma^2 + \theta_j^2}.$$

2. An equation like that in (1) arises if we solve
$$\text{maximize} \sum_{j=1}^{\infty} \frac{\sigma^2 \theta_j^2}{\sigma^2 + \theta_j^2} \quad \text{subject to} \sum_{j=1}^{\infty} a_j^2 \theta_j^2 = c^2.$$

3. The maximum above is of the same form as $\sigma^2 \sum_{j=1}^{\infty} \ell_j$.

- Prove the linear minimax risk result by establishing

$$\sup_{\boldsymbol{\theta}\in\Theta(c,a_{\mathbf{1}},a_{\mathbf{2}},\dots)} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\ell}},\boldsymbol{\theta}) \leq \sigma^2 \sum_{j=1}^{\infty} \ell_j$$

and

$$\sup_{\boldsymbol{\theta}\in\Theta(c,a_{\mathbf{1}},a_{\mathbf{2}},\dots)} \inf_{\boldsymbol{\lambda}} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}},\boldsymbol{\theta}) \geq \sigma^2 \sum_{j=1}^{\infty} \ell_j.$$

- Explain why this is sufficient.

Adapted from Lemma 3.3 on pages 144–145 of Tsybakov [1]

Let $a_1 = 0$, $a_{2m} = a_{2m+1} = (2m)^\beta$, $m = 1, 2, \ldots$ and let η be the solution to

$$\eta^{-1}\sigma^2 \textstyle\sum_{i=1}^\infty a_j(1 - \eta a_j)_+ = L^2/\pi^{2\beta}$$

over $\eta > 0$ and set $\ell_j = (1 - \eta a_j)_+$, $j = 1, 2, \ldots$ Then

1. $\eta = (\beta^{-1}\pi^{-2\beta} L(\beta+1)(2\beta+1))^{-\frac{\beta}{2\beta+1}} \sigma^{\frac{2\beta}{2\beta+1}}(1 + o(1))$
2. $\sigma^2 \sum_{j=1} \ell_j = C\sigma^{\frac{4\beta}{2\beta+1}}(1 + o(1))$

as $\sigma \to 0$, where $C = L^{\frac{2}{2\beta+1}}(\beta^{-1}\pi(\beta+1))^{-\frac{2\beta}{2\beta+1}}(2\beta+1)^{\frac{1}{2\beta+1}}$.

Linear minimax risk over a Sobolev ellipsoid

Under the infinite-dimensional Normal means model we have

$$M_{\text{Lin}}(\Theta_{\text{Sob}}(\beta, L)) = \sup_{\boldsymbol{\theta} \in \Theta_{\textbf{Sob}}(\beta, L)} R(\hat{\boldsymbol{\theta}}_{\boldsymbol{\ell}}, \boldsymbol{\theta}) = C\sigma^{\frac{4\beta}{2\beta+1}}(1 + o(1))$$

as $\sigma \to 0$, where $\boldsymbol{\ell}$ and C are as on the previous slide.

Think of replacing σ^2 with σ^2/n. Then we obtain a minimax risk like

$$\tilde{C} n^{-\frac{2\beta}{2\beta+1}}(1 + o(1)) \quad \text{as } n \to \infty,$$

which resembles the nonparametric rates we have encountered before.

Alexandre B Tsybakov.
Introduction to nonparametric estimation.
Springer Science & Business Media, 2008.