Nonparametric estimation using wavelet methods. L1

Dominique Picard

Université Paris-Diderot
Laboratoire Probabilités et Modèles Aléatoires

Examples of nonparametrics

Estimate a density of probability

We observe $X_1, \ldots, X_n$ i.i.d. with probability P having a density f w.r. to Lebesgue measure. Our aim is to estimate f.

Regression framework

We observe $(X_1, Y_1) \ldots, (X_n, Y_n)$ i.i.d.
X_i Uniform on [0,1],

$$Y_i = f(X_i) + \epsilon_i, \quad \epsilon_i \sim N(0,1).$$

Our aim is to estimate f.

Examples of nonparametrics

White noise model

$$dY_t^{\epsilon} = f(t)dt + \epsilon dW_t,\ t \in [0,1], \epsilon = 1/\sqrt{n}.$$

We observe : $\forall\ \phi \in\ {}_2([0,1]),\ Y_\phi = \int_0^1 \phi(t)f(t)dt + \epsilon\xi_\phi,$
$(\xi_\phi, \xi_\eta) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \|\phi\|^2 & <\phi,\eta> \\ <\phi,\eta> & \|\eta\|^2 \end{pmatrix}\right)$ Our aim is to estimate f.

Examples of nonparametrics (more involved)

EDS

$$dX_t = b(t)dt + f(t)dW_t.$$

We observe $X_i = X_{i\Delta},\ i = 1, \dots, n$.
Our aim is to estimate f.

Examples of nonparametrics (more involved)

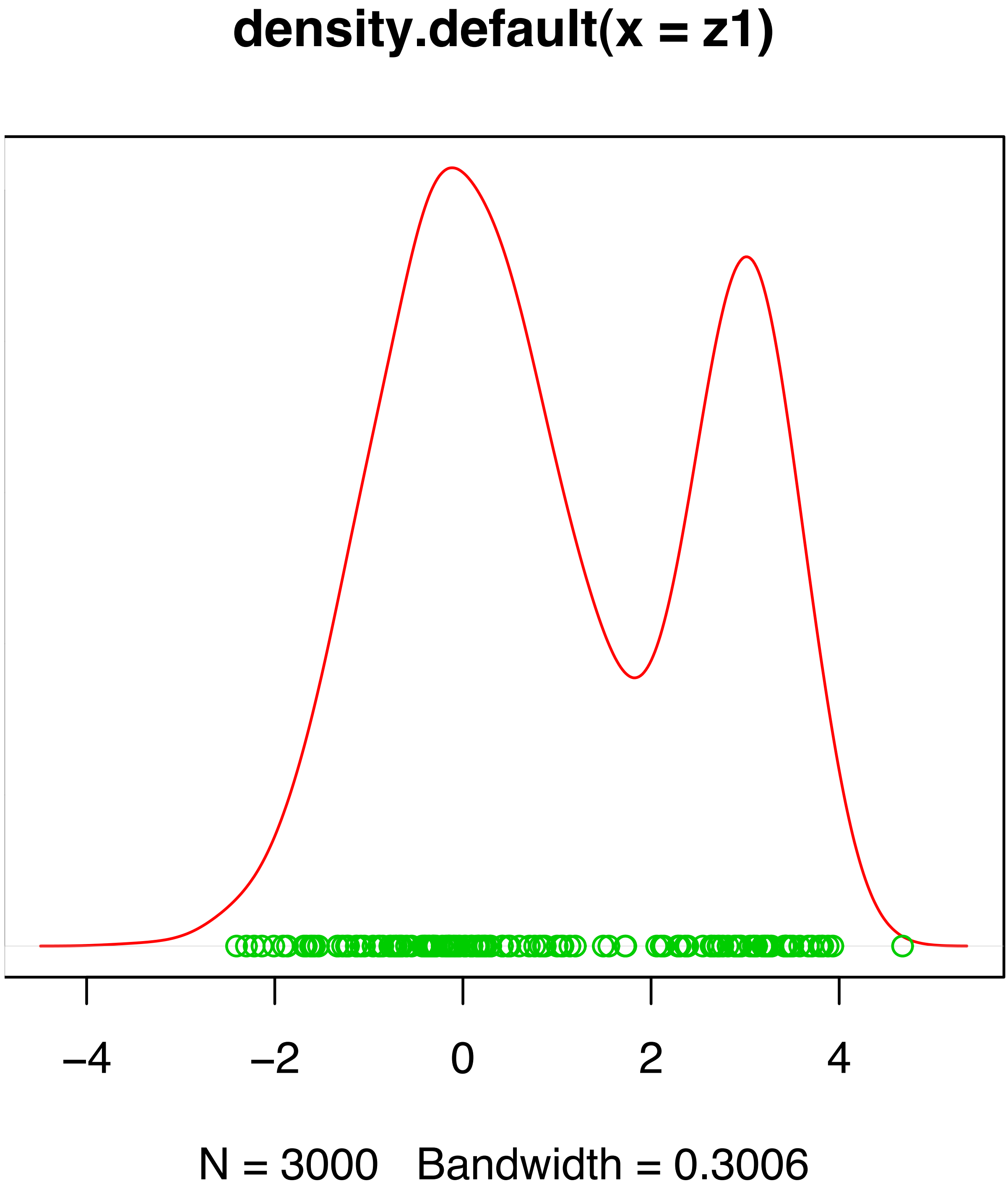
Inverse models

$$dY_t^{\epsilon} = Af(t)dt + \epsilon dW_t,\ t \in [0,1], \epsilon = 1/\sqrt{n}.$$

We observe : $\forall\ \phi \in \ {}_2([0,1]),\ Y_\phi = \int_0^1 \phi(t)Af(t)dt + \epsilon\xi_\phi,\ A$ known linear operator, for instance

$$Af(s) = \int g(s-t)f(t)dt,\ A \text{ Radon transform...}$$

Our aim is to estimate f.



Why is it difficult ?

Estimate a density of probability

We observe $X_1, \ldots, X_n$ i.i.d. with probability P having a density f w.r. to Lebesgue measure. Our aim is to estimate f.
Easy : Estimate $F(x) = P(X_i \leq x)$:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=0}^{n} 1_{]-\infty,x]}(X_i), \quad \hat{F}_n(x) - F(x) = \frac{\xi_n(x)}{\sqrt{n}}$$

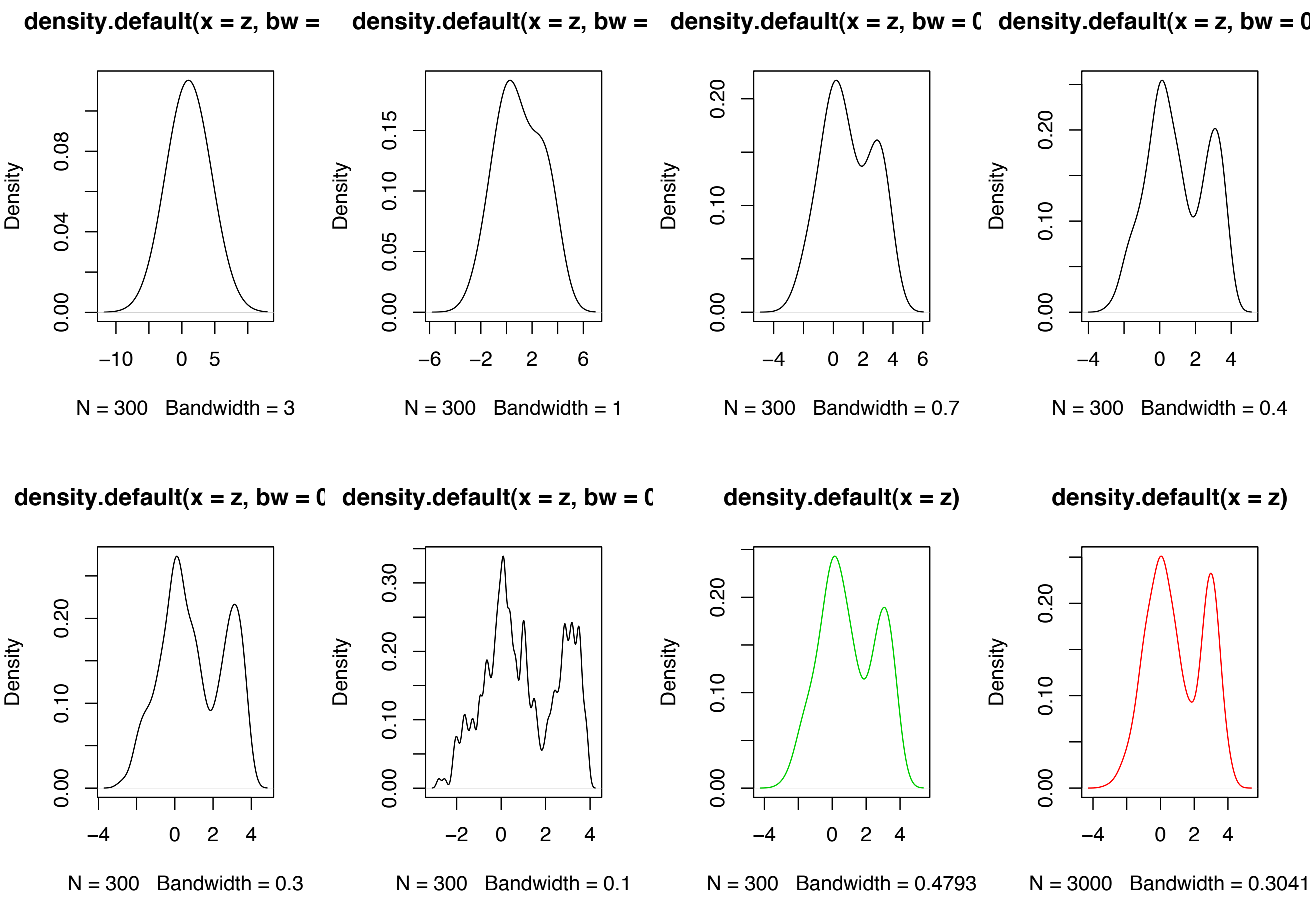
$$\{\xi_n(x),\ x \in \quad\} \overset{weakly}{\rightarrow} \{B_0(F(x)),\ x \in \quad\}$$

Kolmogorov Smirnov.

2 obstructions to differentiation :
- $\hat{F}_n(x)$ is not differentiable.
- $B_0(F(x))$ is not differentiable either.

Parzen kernel method

$$
\begin{aligned}
\hat{K}_{h_n}(x) &= \int \frac{1}{h_n} K\left(\frac{x-y}{h_n}\right) dF_n(y) \\
&= \frac{1}{nh_n} \sum_{i=0}^{n} K\left(\frac{x-X_i}{h_n}\right)
\end{aligned}
$$



Different bandwidths

Minimax framework

- Our aim is to estimate $f \in V$
- We have a loss function $l(\hat{f}, f)$
  (for instance $l(\hat{f}, f) = \|\hat{f} - f\|_p^p, \ \|\hat{f} - f\|_\infty$)

$f^\star$ is minimax (exactly) if

$$\sup_{V} \ {}_f\, l(f^\star, f) = \inf_{\hat{f}} \sup_{V} \ {}_f\, l(\hat{f}, f).$$

$f_n^\star$ is minimax (up to constants) if

$$c \inf_{\hat{f}_n} \sup_{V} \ {}^n_f\, l(\hat{f}_n, f) \leq \sup_{V} \ {}^n_f\, l(f_n^\star, f) \leq C \inf_{\hat{f}_n} \sup_{V} \ {}^n_f\, l(\hat{f}_n, f) \ \forall \ n.$$

Minimax framework

Two steps :
Find a rate $r(n)$

1. Lower bound $\sup_V \quad {}^n_f l(\hat{f}_n, f) \geq c\ r(n)$
2. Upper bound : construct an estimation method $f_n^\star$ with $\sup_V \quad {}^n_f l(f_n^\star, f) \leq C\ r(n)$.

Functional constraints and losses

Models : density-1, regression-2, white noise model-3

$$V = V_\alpha(L) = \{f : [0,1] \mapsto \quad , \sup_{|x-y|\leq\delta} |f(x) - f(y)| \leq L\delta^\alpha, \ \forall\ \delta,$$

$$|f(0)| \leq L\}$$

$$l(\hat{f}, f) = \|\hat{f} - f\|_p^p$$

Upper bound

$$\hat{K}_{h_n}(x) = \frac{1}{nh_n}\sum_{i=0}^{n} K(\frac{x - X_i}{h_n}) \quad \text{density } (\textit{Rosenblatt } 1956)$$

$$\hat{K}_{h_n}(x) = \frac{1}{nh_n}\sum_{i=0}^{n} K(\frac{x - X_i}{h_n})Y_i \quad \text{regression } (\textit{NadarayaWatson } 1964)$$

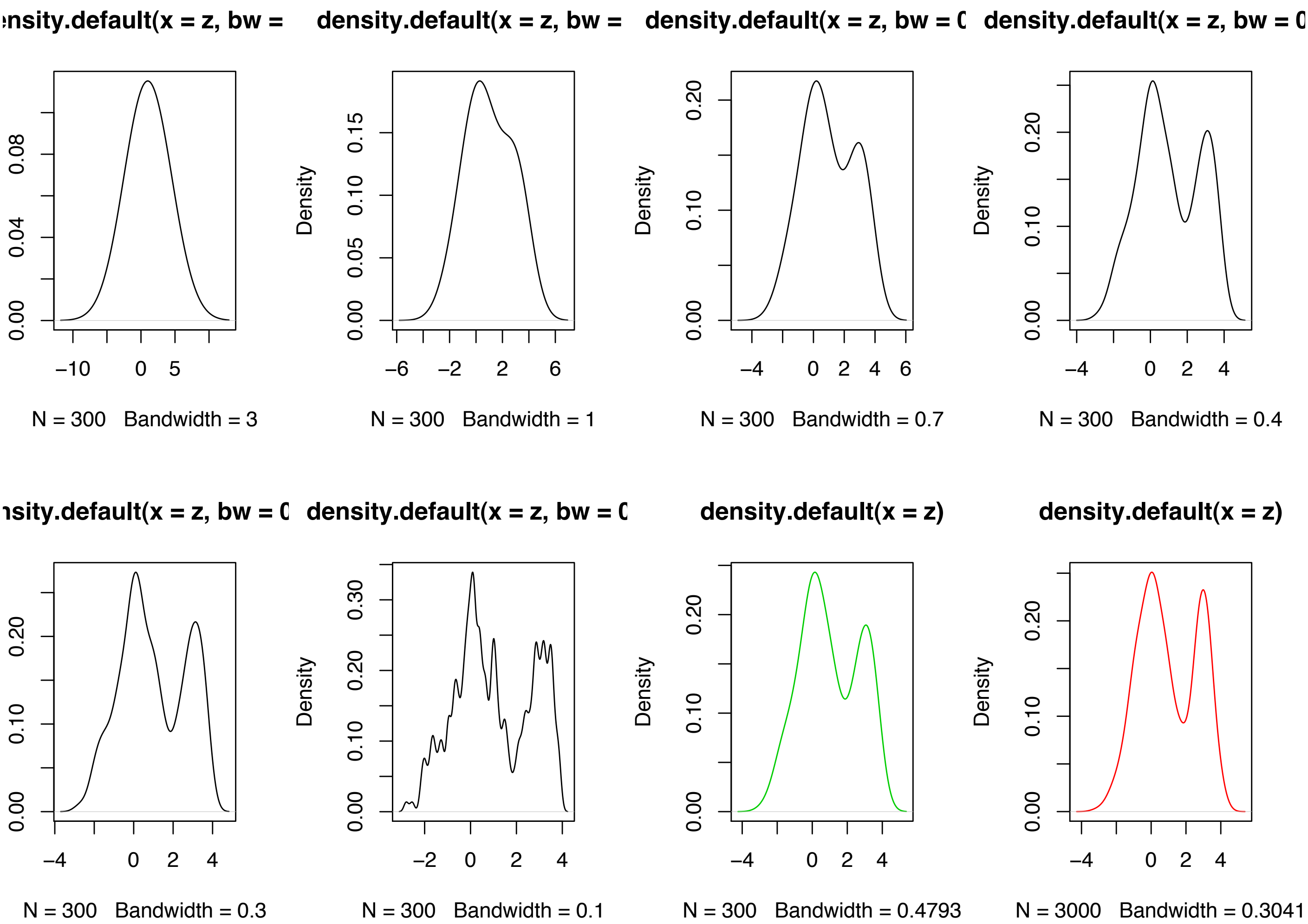
$$\hat{K}_{h_n}(x) = \int_0^1 \frac{1}{h_n}K(\frac{x - t}{h_n})dY_t^{\epsilon} \quad \text{white noise}$$

Theorem 2

For $0 \leq \alpha \leq 1$, in models 1,2 or 3, $\int K(x)dx = 1$, K compactly supported $[-M, M]$, $\|K\|_\infty < \infty$.

$$f_n^\star = \hat{K}_{h_n}, \quad h_n = n^{\frac{-1}{1+2\alpha}}$$

$$\sup_{V_\alpha(L)} {}^n_f\|f_n^\star - f\|_2^2 \leq Cn^{\frac{-2\alpha}{1+2\alpha}} = Cr(n)$$



ensity.default(x = z, bw =
N = 300 Bandwidth = 3
density.default(x = z, bw =
N = 300 Bandwidth = 1
density.default(x = z, bw = 0
N = 300 Bandwidth = 0.7
density.default(x = z, bw = 0
N = 300 Bandwidth = 0.4
nsity.default(x = z, bw = 0
N = 300 Bandwidth = 0.3
density.default(x = z, bw = 0
N = 300 Bandwidth = 0.1
density.default(x = z)
N = 300 Bandwidth = 0.4793
density.default(x = z)
N = 3000 Bandwidth = 0.3041
Density

Upper bound, density model

$$ {}^{n}_{f}(\hat{K}_{h_n}(x) - f(x))^2 = \mathrm{E}^{n}_{f}(\hat{K}_{h_n}(x) - K_{h_n}f(x))^2 + (K_{h_n}f(x) - f(x))^2 $$

Balance bias, variance

Variance

$$ \begin{aligned} {}^{n}_{f}(\hat{K}_{h_n}(x) - K_{h_n}f(x))^2 &\leq \frac{1}{n}\int K_{h_n}(x)^2 f(x)dx \\ &\leq \frac{L}{nh_n}\int K(x)^2 dx \\ &\leq \frac{L}{nh_n}2M\|K\|_{\infty}^2 \end{aligned} $$

Balance bias, variance

Bias

$$
\begin{aligned}
|K_{h_n} f(x) - f(x)| &= |\int K_{h_n}(u)[f(x-u) - f(x) du| \\
&\leq \int |K_{h_n}(u)| \sup_{|u/h_n| \leq M} |f(x-u) - f(x)| du \\
&\leq 2M \|K\|_\infty L(2Mh_n)^\alpha
\end{aligned}
$$

Balance bias, variance

$$
\begin{aligned}
{}^{n}_{f}(\hat{K}_{h_n}(x) - f(x))^2 &= \mathrm{E}^{n}_{f}(\hat{K}_{h_n}(x) - K_{h_n}f(x))^2 + (K_{h_n}f(x) - f(x))^2 \\
&\leq \frac{L}{nh_n} 2M\|K\|_\infty^2 + \quad 4M^2\|K\|_\infty^2 L^2 (2Mh_n)^{2\alpha}
\end{aligned}
$$

$$
{}^{n}_{f}\|\hat{K}_{h_n} - f\|_2^2 \leq c[\frac{L}{nh_n} 2M\|K\|_\infty^2 + \quad 4M^2\|K\|_\infty^2 L^2 (2Mh_n)^{2\alpha}]
$$

$$
c \leq 2M + 1
$$

Optimized for

$$
h_n^* \sim n^{\frac{-1}{2\alpha+1}}
$$

$$
{}^{n}_{f}\|\hat{K}_{h_n^*} - f\|_2^2 \leq Cn^{\frac{-2\alpha}{2\alpha+1}}
$$

Questions

- What about the lower bound ? (Is $n^{\frac{-2\alpha}{2\alpha+1}}$ the minimax rate ?)
- What is the behavior for a general L_p loss ?
- How to choose h_n in practice ?

Orthogonal series methods

$f \in \ _2([0,1])$, $\mathcal{E} = \{\psi_i, i \in \ \}$ orthonormal basis of $\mathrm{L}^2([0,1], dt)$,

$$f = \sum \theta_i \psi_i, \quad x_i = \int \psi_i dY, \ i \in \quad ,$$

General estimator

$$\hat{f} = \sum_{i \in A} \hat{\theta}_i \psi_i.$$

Two choices : A, $\hat{\theta}_i$.

Orthogonal series methods

$$A = \text{(generally) } \{0, \ldots, K\}$$

$$\hat{\theta}_i = \frac{1}{n}\sum_{i=0}^{n} \psi_i(X_i) \quad \text{density}$$

$$\hat{\theta}_i = \frac{1}{n}\sum_{i=0}^{n} \psi_i(X_i) Y_i \quad \text{regression}$$

$$\hat{\theta}_i = \int_0^1 \psi_i(t) dY_t^{\epsilon} \quad \text{white noise}$$

$$\rightarrow \hat{f}_K$$

Upper bounds

- If we assume f belongs to a polynomially tail compact domain :
  For $s > 0$, fixed,

$$V = \{f = \sum \theta_k \psi_k, \ \sum_{k>K} \theta_k^2 \leq LK^{-2s}, \ \forall K, \ \|f\|_\infty \leq L\}$$

Upper bounds

- $$f \in V(s,M) = \{f = \sum \theta_k \psi_k,\ \sum_{k>K} \theta_k^2 \le L^2 K^{-2s},\ \forall K,\ \|f\|_\infty \le L\}$$

- $$\begin{aligned} {}^{n}_{f}\|\hat{f}_K - f\|^2 &= \sum_{k\le K} {}^{n}_{f}(\hat{\theta}_k - \theta_k)^2 + \sum_{k>K} \theta_k^2 \\ &\le (K+1)\frac{L}{n} + \sum_{k>K} \theta_k^2 \\ &\le (K+1)\frac{L}{n} + L^2 K^{-2s} \end{aligned}$$

Optimized for $K_s^* = c[n]^{\frac{1}{1+2s}} \le K_0^* = cn$
(decreasing in s)

$$\sup_{f \in V(s,M)} E\|\hat{f}_{K*} - f\|^2 \le c' n^{\frac{-2s}{1+2s}}.$$

Kernels versus series

Easier calculation for series (for proof and computation)

Tuning parameters $K \sim h^{-1}$ gives an interpretation of the bandwidth parameter and the 'dimension' of the problem.

Space V depends on the basis, on the numbering in the basis, only allows a L_2 loss function.

Bases and functional spaces

Trigonometric basis and Sobolev spaces

- : $\mathrm{L}_2([0,1])$ of periodic functions

$$\psi_0 = 1$$
$$\psi_{2k}(x) = \sqrt{2}\cos 2k\pi x$$
$$\psi_{2k+1}(x) = \sqrt{2}\sin 2k\pi x$$

Let $\beta \in \ _*$, the following Sobolev space,

$$W(\beta, L) = \{f : [0,1] \mapsto \quad : f^{\beta-1} \ absolut. \ continuous \ \int (f^{\beta})^2(x)dx \leq L$$

$$W^{per}(\beta, L) = \{f \in W(\beta, L), \ periodic\}$$

Trigonometric basis and Sobolev spaces

Let

$$\Theta((a_j), Q) = \{\theta \in l^2 \ : \ \sum_j a_j^2 \theta_j^2 \leq Q^2\}$$

We have,

$$\begin{aligned} W^{per}(\beta, L) = \Theta((a_j), Q) &:= \Theta(\beta, Q), \\ a_j &= j^{\beta}, \ j \ even \\ a_j &= (j-1)^{\beta}, \ j \ odd \\ Q &= \frac{L}{\pi^{\beta}} \end{aligned}$$

Trigonometric basis and Sobolev spaces

$$\Theta(\beta, Q) = \{\theta \in l^2 \ : \ \sum_j j^{2\beta}\theta_j^2 \leq Q^2\}$$

$$\Theta(\beta, Q) \subset V(\beta, Q)$$

$$\sum_{j \geq K} \theta_j^2 \leq \sum_{j \geq K} [\frac{j}{K}]^{2\beta}\theta_j^2 \leq K^{-2\beta} \sum_j j^{2\beta}\theta_j^2$$