Statistics 860 Lecture 3

The bottom line from lecture 2:

$$y_n = \underbrace{X_{n\times p}}_{K}\underbrace{\beta_p}_{f} + \epsilon \qquad \epsilon_n \sim \mathcal{N}(0, \sigma^2 I_n) \quad \beta \sim \mathcal{N}(0, b\Sigma)$$

$$\begin{aligned} \hat{\beta} = E(\beta \mid y) &= b\Sigma X'(bX\Sigma X' + \sigma^2 I)^{-1} y \\ &= \Sigma X'(X\Sigma X' + \tfrac{\sigma^2}{b} I)^{-1} y \end{aligned} \qquad (*)$$

$$\min \frac{1}{n} \underbrace{\|y - X\beta\|^2}_{\text{Euclidean norm}} + \lambda \beta' \Sigma^{-1} \beta \qquad (**)$$

$$\beta_\lambda = (X'X + n\lambda\Sigma^{-1})X'y = \Sigma X'(X\Sigma X' + n\lambda I)^{-1} y$$

by using the lemma

$$(X'X + n\lambda\Sigma^{-1})X' \equiv \Sigma X'(X\Sigma X' + n\lambda I)^{-1}.$$

Let $n\lambda = \sigma^2/b$. **Then the minimizer of** $(**)$ **is the same as the Bayes estimate** $(*)$**.**

Inner Product Notation: ($l_2, \mathcal{L}_2$ to be defined)

$$\begin{array}{ll} (,) & E_n, l_2, \mathcal{L}_2 \\ \langle , \rangle_{\mathcal{H}} & \mathcal{H} \quad any \quad other \quad Hilbert \quad space. \end{array}$$

Let $\langle , \rangle_{\mathcal{H}}$ be the inner product in $\mathcal{H}$

Let $x_1, ..., x_n \in \mathcal{H}$ with Gram matrix

$$\left(\begin{array}{lll} \langle x_1, x_1 \rangle_{\mathcal{H}} & ... & \langle x_1, x_n \rangle_{\mathcal{H}} \\ ... & ... & ... \\ \langle x_n, x_1 \rangle_{\mathcal{H}} & & \langle x_n, x_n \rangle_{\mathcal{H}} \end{array} \right) = G = \left\{ g_{ij} \right\}, \qquad say$$

G is always non-negative definite, if it is of full rank (strictly positive definite) then $x_1, ..., x_n$ are linearly independent.

$$G \succ 0 \Rightarrow \quad \sum_{i,j} a_i a_j g_{ij} = \sum_{i,j} a_i a_j \left\langle x_i, x_j \right\rangle_{\mathcal{H}}$$
$$= \left\langle \sum_i a_i x_i, \sum_j a_j x_j \right\rangle_{\mathcal{H}} = ||\sum_i a_i x_i||^2_{\mathcal{H}} > 0$$

$$\mathcal{H}_n = subspace \quad spanned \quad by \quad x_1, ..., x_n.$$

Let $x \in \mathcal{H}$

$$P_n x = projection \quad of \quad x \quad onto \quad \mathcal{H}_n : \; P_n x = \sum a_i x_i$$

$$\min \left\| x - \sum a_i x_i \right\|_{\mathcal{H}}^2$$

$$\left(\left\{\left\langle x_i, x_j \right\rangle_{\mathcal{H}}\right\}\right) \begin{pmatrix} a_1 \\ ... \\ a_n \end{pmatrix} = \begin{pmatrix} \langle x, x_1 \rangle \\ ... \\ \langle x, x_n \rangle \end{pmatrix}$$

(the ”normal equations”)

$\mathcal{H}$ Hilbert space

L_i a bounded linear functional on $\mathcal{H}$
Very Important Concept

$L_i(f_1 + f_2) = L_i f_1 + L_i f_2$ Linear
$L_i f =$ a real number depending on f with
$|L_i f| \leq M_i \, ||f||$ for some M_i (depending on L_i) Bounded

Special Case

E_n Euclidean n-space

$$x \in E_n, \quad L_i x = l_i' x = (l_i, x) \qquad \qquad where \quad l_i \in E_n$$

$$\left| l_i' x \right| \leq \underbrace{||l_i||}_{M_i} \times ||x||$$

$$f \in \mathcal{L}_2, \int f^2(s) ds < \infty \qquad L_i f = \int k_i(s) f(s) \, ds$$

$$k_i \in \mathcal{L}_2, \qquad \qquad L_i f = (k_i, f)_{\mathcal{L}_2}$$

$$y_i = L_i f + \varepsilon_i, \qquad\qquad i = 1, ..., n$$

$$\frac{1}{n}\sum (y_i - L_i f)^2 + \underbrace{\lambda \int_0^1 \left(f''(x)\right)^2 dx}_{quadratic\ penalty}$$

$$L_i f = < \eta_i, f >_{\mathcal{H}}$$

$$||y - X\beta||^2 + \lambda \beta' \Sigma^{-1} \beta$$

$$\sum (y_i - x_i' \beta)^2 + \lambda \beta' \Sigma^{-1} \beta$$

$L_i \beta = x_i' \beta$, notion of a bounded linear functional is a trivial idea in finite dimensional Euclidean space - maps β onto the real line. Not so trivial in a Hilbert space. (note: ”bounded linear functional” and ”linear functional” being used interchangeably. See Section 16. ”The Theorem of F. Riesz” for ”the Riesz Representation Theorem” - will return to this.

(Ordinary) Euclidean Space

$$\begin{aligned} E_n &: \\ x &= (x(1), ..., x(n))' \qquad\qquad colmn \\ y &= (y(1), ..., y(n))' \qquad\qquad vectors \end{aligned}$$

$$(x, y) = \sum_{i=1}^{n} x(i)\, y(i)\,, \qquad ||x||^2 = (x, x)$$

(Special) Euclidean space

$E_n(\Sigma)$ $\quad\quad$ $\Sigma > 0$ (positive definite)

$$\begin{aligned} x &= (x(1), ..., x(n)) \\ y &= (y(1), ..., y(n)) \end{aligned}$$

$$\langle x, y \rangle_{\Sigma} = x' \Sigma^{-1} y \qquad ||x||^2 = x' \Sigma^{-1} x$$

Hilbert spaces

$$l_2 \qquad\qquad\qquad ''little \quad l_2''$$

$$x = (x(1), x(2), ... \qquad \to \infty$$

$$y = (y(1), y(2), ... \qquad \to \infty$$

$$(x, y) = \sum_{i=1}^{\infty} x(i)\, y(i) \qquad\qquad ||x||^2 = \sum_{i=1}^{\infty} [x(i)]^2$$

Note $x \in l_2$ if and only if

$$\sum_{i=1}^{\infty} [x(i)]^2 < \infty$$

By the famous Cauchy-Schwartz inequality

$$\left| \sum_{i=1}^{\infty} x(i)\, y(i) \right| \leq \sqrt{\sum_{i=1}^{\infty} (x(i))^2} \sqrt{\sum_{i=1}^{\infty} (y(i))^2}$$

$$x, y \in l_2 \Rightarrow |(x, y)| \leq ||x||\, ||y||$$

$$\frac{(x, y)}{||x||\, ||y||} \qquad ''cosine\ of\ the\ angle\ between\ x\ and\ y''$$

$$l_2(\Lambda), \qquad \Lambda = (\lambda_1, \lambda_2, ... \qquad \lambda_v > 0$$

$$x \in l_2(\Lambda) \qquad iff \qquad \sum x^2(i)/\lambda_i < \infty$$

$$\langle x, y \rangle = \sum_{i=1}^{\infty} \frac{x(i)\, y(i)}{\lambda_i}$$

If $\lambda_v \to 0$, $l_2(\Lambda)$ will be "smaller" than l_2

If $\lambda_v \to \infty$, $l_2(\Lambda)$ will be "bigger" than l_2

If $0 < a \leq \lambda_v \leq b < \infty$, then $x \in l_2$ iff $x \in l_2(\Lambda)$

c.o.n.s.   ”complete orthonormal sequence”

$$\phi_1, \phi_2, ... \in \mathcal{H}$$

Orthonormal:

$$\begin{aligned} \left\langle \phi_i, \phi_j \right\rangle &= 1 \qquad\qquad i = j \\ &= 0 \qquad\qquad i \neq j \end{aligned}$$

Complete iff

$$x \in \mathcal{H} \qquad\qquad P_n x \equiv \sum_{i=1}^{n} \langle x, \phi_i \rangle \, \phi_i$$

$$\lim_{n \to \infty} ||x - P_n x||^2 \to 0 \quad norm\ convergence$$

$$\begin{aligned} l_2 \qquad\qquad x_1 &= (1, 0, 0, ... \\ x_2 &= (0, 1, 0, ... \\ x_3 &= (0, 0, 1, ... \\ &\quad ... \qquad\qquad\qquad c.o.n.s. \end{aligned}$$

$$\begin{aligned} l_2\,(\Lambda) \qquad\qquad x_1 &= (\sqrt{\lambda_1}, 0, 0, ... \\ x_2 &= (0, \sqrt{\lambda_2}, 0, ... \\ x_3 &= (0, 0, \sqrt{\lambda_3}, ... \\ &\quad ... \qquad\qquad\qquad c.o.n.s. \end{aligned}$$

$\mathcal{L}_2\,[0,1]$ "Big $\mathcal{L}_2$"

f Lebesgue measurable function on $[0,1]$,
$\int_0^1 f^2\,(t)\,dt < \infty$ well defined

$$(f,g) = \int_0^1 f\,(t)\,g\,(t)\,dt \qquad ||f||^2 = \int_0^1 f^2\,(t)\,dt$$

$$c.o.n.s \qquad \phi_1, \phi_2, ...$$

if

$$\int \phi_i\,(s)\,\phi_j\,(s)\,ds = \delta_{ij}$$

and

$$P_n f = \sum_{v=1}^{n} (f, \phi_v)\,\phi_v$$

satisfies

$$(f, \phi_v) = \int_0^1 f\,(s)\,\phi_v\,(s)\,ds$$

$$\lim_{n\to\infty} ||f - P_n f|| \to 0 \qquad for \quad all \quad f \in \mathcal{L}_2$$

$\mathcal{L}_2[0,1]$ $\qquad f \in \mathcal{L}_2[0,1] \iff \int_0^1 (f(t))^2\,dt < \infty$

Example: Kaplan advanced calculus (Chapter 7)

$$\begin{aligned}
\phi_v(t) &= 1 \qquad\qquad v = 0, \qquad t \in [0,1] \\
&= \sqrt{2}\cos 2\pi v t \\
&= \sqrt{2}\sin 2\pi v t \qquad v = 1, 2, \ldots
\end{aligned}$$

Fourier series expansion for $f \in \mathcal{L}_2[0,1]$

$$f(t) \sim a_0 + \sqrt{2}\sum a_v \sin 2\pi v t + \sqrt{2}\sum b_v \cos 2\pi v t$$

$$\begin{aligned}
a_v &= (f, \phi_v) = \sqrt{2}\int_0^1 f(t)\sin 2\pi v t dt \\
b_v &= (f, \phi_v) = \sqrt{2}\int_0^1 f(t)\cos 2\pi v t dt
\end{aligned}$$

Parseval's Theorem

$$||f||^2 = \lim_{n\to\infty} ||P_n f||^2 = \lim_{n\to\infty} \underbrace{\sum_{v=1}^{n} (f, \phi_v)^2}_{||P_n f||^2}$$

$$P_n f = \sum_{v=1}^{n} (f, \phi_v)\, \phi_v$$

$$\begin{aligned} ||f - P_n f||^2 &= ||f||^2 - 2(f, P_n f) + ||P_n f||^2 \\ &= ||f||^2 - ||P_n f||^2 \to 0 \end{aligned}$$

(since $(f, P_n f) = (P_n f, P_n f)$)

$$\begin{aligned} ||P_n f||^2 &= a_0^2 + \sum_{v=1}^{n} \left(a_v^2 + b_v^2\right) \\ & \quad (with \quad some \quad abuse \quad of \quad notation) \\ &\to \ ||f||^2 = \int_0^1 (f\,(t))^2\, dt \end{aligned}$$

Remark:

$$f(t) \sim a_0 + \sqrt{2}\sum a_v \sin 2\pi v t + \sqrt{2}\sum b_v \cos 2\pi v t$$

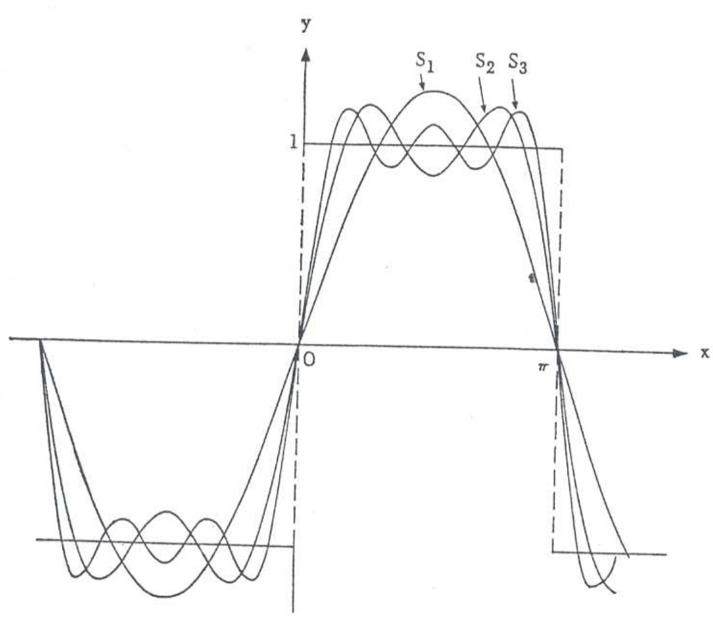
means

$$||f - P_n f||^2 \to 0$$

it does not necessarily mean

$$f(t) - (P_n f)(t) \to 0$$

Fourier series expansion of the step function. Sum of first 1, 2 and 3 (odd) components.Converges to halfway up the jump, not pointwise. (Gibbs effect)



$\mathcal{L}_2\left[\{\phi_v\}, \Lambda\right] \quad \Lambda = (\lambda_1, \lambda_2, ... \quad \lambda_v > 0$

$f \in \mathcal{L}_2\left[\{\phi_v\}, \Lambda\right]$

$$||f||^2 = \sum_{v=1}^{\infty} \frac{(f, \phi_v)^2}{\lambda_v} < \infty$$

$$\langle f, g \rangle = \sum_{v=1}^{\infty} \frac{(f, \phi_v)(g, \phi_v)}{\lambda_v}$$

c.o.n.s. is $\left\{\sqrt{\lambda_v}\phi_v\right\}$ Analogue of

$$\begin{aligned} E_n &: (x, y) = \sum x(i)\, y(i) \\ E_n(\Sigma) &: (x, y) = x'\Sigma^{-1} y \end{aligned}$$

If $\Sigma = \Gamma D \Gamma'$

$$D = \begin{pmatrix} \lambda_1 & 0 \\ ... & .. \\ 0 & \lambda_n \end{pmatrix}$$

$\phi_1, ..., \phi_n$ columns of Γ

$$\begin{aligned} x'\Sigma^{-1} y &= x'\Gamma D^{-1} \Gamma' y \\ &\equiv \sum_{v=1}^{n} \frac{1}{\lambda_v} (x, \phi_v)(y, \phi_v) \end{aligned}$$

Example $\mathcal{L}_2\left[\{\phi_v\}, \Lambda\right]$

$$\begin{aligned} \phi_0 &= 1 \\ \phi_v &= \sqrt{2}\sin 2\pi v t \\ &= \sqrt{2}\cos 2\pi v t \qquad v = 1, 2, ... \end{aligned}$$

$$\begin{aligned} \mathcal{H} \ : \ & \lambda_0 = 1 \\ & \lambda_v = (2\pi v)^{-2m} \qquad v = 1, 2, ... \end{aligned}$$

$$f(t) \sim a_0 + \sqrt{2}\sum a_v \sin 2\pi v t + \sqrt{2}\sum b_v \cos 2\pi v t$$

$$\begin{aligned} \sum \frac{(f, \phi_v)^2}{\lambda_v} &< \infty \\ &= \left[\int_0^1 f(t)\,dt\right]^2 + \underbrace{\sum_{v=1}^{\infty} (2\pi v)^{2m}\left(a_v^2 + b_v^2\right)}_{if\ finite} \\ &= \left[\int_0^1 f(t)dt\right]^2 + \int_0^1 \left(f^{(m)}(t)\right)^2 dt \end{aligned}$$

$$\frac{d}{dt}\sin 2\pi v t = 2\pi v \cos 2\pi v t, ... \qquad etc.$$

m even

$$f(t) = a_0 + \sqrt{2}\sum a_v \sin 2\pi vt + \sqrt{2}\sum b_v \cos 2\pi vt$$

$$f^{(m)}(t) = (-1)^{m/2}(2\pi v)^m \qquad \{\sum a_v \sin 2\pi vt + \sum b_v \cos 2\pi vt\}$$

$$\left\|f^{(m)}\right\|_{L_2}^2 = \sum (2\pi v)^{2m}\left(a_v^2 + b_v^2\right)$$

By Parseval's Theorem (if it is finite)

Note: $\mathcal{L}_2[\{\phi_r\}, \Lambda]$ with these eigenvalues and eigenfunctions does not include π_{m-1} (polynomials of degree $m-1$). Observe that

$$f^{(v)}(1) - f^{(v)}(0) = 0, \qquad v = 0, 1, ..., m-1.$$

A more general approach to see why $\pi_{m-1} \notin \mathcal{L}_2[\{\phi_v\}, \Lambda]$ is:

Let $f(t) = t^r$. We have that $\int_0^1 t^r cos 2\pi vt dt$ and $\int_0^1 t^r sin 2\pi vt dt$ behave like $\frac{1}{v^r}$. Recall $\lambda_v = v^{2m}$. Therefore it is necessary that

$$\sum_{v=1}^{\infty} \frac{v^{2m}}{v^{2r}} < \infty$$

for t^r to be in $\mathcal{L}_2[\{\phi_v\}, \Lambda$, that is, $r > m + 1/2$. Thus, $\pi_{m-1} \notin \mathcal{L}_2[\{\phi_v\}, \Lambda]$, and, actually, neither is is t^m.

Bernoulli Polynomials

$x \in [0, 1]$

$$\begin{aligned} B_0(x) &= 1 \\ B_1(x) &= x - \frac{1}{2} \\ B_2(x) &= x^2 - x + \frac{1}{6} \\ &\ldots \end{aligned}$$

$$B_n'(x) = nB_{n-1}(x) \qquad \int_0^1 B_n(x)\,dx = 0, \quad n > 0$$

$$\begin{aligned} \frac{B_{2k}(x)}{(2k)!} &= \frac{(-1)^{k-1}}{(2\pi)^{2k}} \sum_{v=1}^{\infty} 2\frac{\cos 2\pi v x}{v^{2k}} \quad 0 \leq x \leq 1 \\ \frac{B_{2k-1}(x)}{(2k-1)!} &= \frac{(-1)^{k}}{(2\pi)^{2k-1}} \sum_{v=1}^{\infty} 2\frac{\sin 2\pi v x}{v^{2k-1}} \quad 0 < x < 1 \end{aligned}$$

(Abromowitz and Stegun)