Chapter 3 : Functional Principal components . (FPC)

- FPC's allow us to reduce the dimension of ∞ - dimensional functional data to a small (p) finite dimension in an optimal way.

① maximizing variability.

② optimal orthonormal basis.

§ 3.1 A maximization problem

$\boxed{\text{Thm 3.1}}$ (Principal axis theorem).

Suppose A is a sym. $p \times p$ matrix. Then there is an orthonormal matrix $U = [u_1 \cdots u_p]$ whose columns are the eigenvectors of A.

so, $U^T U = I$ and $A \cdot u_j = \lambda_j \cdot u_j$

$\Rightarrow$ $U^T A U = \Lambda = \text{diag}(\lambda_1 \cdots \lambda_p)$. ☒

Thm 3.1 implies that $A = U \cdot \Lambda \cdot U^T$

Question :

Suppose A be symm. and pd. with $\lambda_1 > \lambda_2 \cdots > \lambda_p$.

Find X s.t $X^T A X$ is maximum and $\|X\| = 1$.

$X^T A X = X^T U \Lambda U^T X = Y^T \cdot \Lambda \cdot Y$ where $Y = U^T \cdot X$

$\Rightarrow X = U \cdot Y$

Since $\|Y\| = \|X\| = 1$, find unit length vector y first and $X = U \cdot Y$.

$Y^T \Lambda \cdot Y = \sum_{j=1}^{p} \lambda_j \cdot Y_j^2 \Rightarrow Y = [1, 0 \cdots 0]^T \Rightarrow X = u_1$

Extend this idea to a separable Hilbert space.

Suppose ψ is a sym. pd Hilbert-Schmidt operator in $\mathcal{L}^2$.

RECALL:

$$\langle \psi(x), y \rangle = \langle x, \psi(x) \rangle \quad , x, y \in \mathcal{L}^2$$

$$\langle \psi(x), x \rangle \geq 0 \quad , x \in \mathcal{L}^2$$

$$\psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j \quad , x \in \mathcal{L}^2 \qquad (2.4).$$

where orthonormal v_j and $\psi(v_j) = \lambda_j \cdot v_j$.

$$\Rightarrow \psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j \quad , \quad x \in \mathcal{L}^2 \qquad \text{by (2.4)}$$

Then,

maximize $\langle \psi(x), x \rangle$ subj. to $\|x\| = 1$

a. $$\langle \psi(x), x \rangle = \langle \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j , x \rangle$$
$$= \sum_{j=1}^{\infty} \lambda_j \langle \langle x, v_j \rangle \cdot v_j , x \rangle$$
$$= \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle^2$$

b. $\|x\|^2 = \sum_{j=1}^{\infty} | \langle x, v_j \rangle |^2 = 1$ by Parseval's equality

$\Leftrightarrow$ maximize $\sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle^2$ subj. to $\sum_{j=1}^{\infty} |\langle x, v_j \rangle|^2 = 1$.

Suppose $\lambda_1 > \lambda_2 > \cdots > \cdots$

So, take $\langle x, v_1 \rangle = 1$ and $\langle x, v_j \rangle = 0 \ \ni \ \forall j > 1$

$\Rightarrow \max \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle^2 = \lambda_1$ and $x = v_1$ (or $-v_1$)

Next, maximize $\langle \Psi(x), x \rangle$ subj. to $\|x\| = 1$, $\langle x, v_1 \rangle = 0$

then, $\max \sum_{j=2}^{\infty} \lambda_j \langle x, v_j \rangle^2$ & $\sum_{j=2}^{\infty} |\langle x, v_j \rangle|^2 = 1$.

$\Rightarrow x = v_2$ & maximum value $= \lambda_2$

$\boxed{\text{Thm 3.2}}$

Suppose Ψ is a symm. pd Hilbert - Schmidt operator with eigenfunctions v_j and eigenvalues λ_j satisfying $\lambda_1 > \lambda_2 > \cdots > \lambda_{p+1}$,

Then,

$\sup\{ \langle \Psi(x), x \rangle : \|x\| = 1, \langle x, v_j \rangle = 0, 1 \le j \le i-1, i < p \} = \lambda_i$

and supremum is reached if $x = v_i$

§ 3.2 Optimal Empirical Orthogonal basis

Approach in § 3.1 can be applied to the problem to find an orthonormal basis $u_1 \cdots u_p$ s.t

$$\hat{S}^2 = \sum_{i=1}^{N} \| x_i - \sum_{k=1}^{p} \langle x_i, u_k \rangle u_k \|^2 \text{ is minimized.}$$

- $x_1 \cdots x_N$: observed realizations of random functions in $\mathcal{L}^2$.
- $p < N$ (fixed)

$\Rightarrow X_i \simeq \sum_{k=1}^{p} \langle x_i, u_k \rangle u_k$

u_j : optimal empirical orthonormal basis

or natural orthonormal components

The functions $u_1 \cdots u_p$ minimizing $\hat{S}^2$ are equal to the normalized eigenfunctions of the sample covariance operator $\hat{C}$.

proof)

Suppose $p=1$. Find u with $\|u\|=1$.

$$\hat{S}^2 = \sum_{i=1}^{N} \| x_i - \langle x_i, u \rangle u \|^2$$

$$= \sum^{N} \| x_i \|^2 - 2 \sum \langle x_i, u \rangle^2 + \sum \langle x_i, u \rangle^2 \| u \|^2$$

$$= \sum \| x_i \|^2 - \sum \langle x_i, u \rangle^2$$

RECALL:

$$\hat{C}(u) = \frac{1}{N} \sum_{i=1}^{N} \langle x_i, u \rangle u.$$

$$\Rightarrow \langle \hat{C}u, u \rangle = \langle \frac{1}{N} \sum_{i=1}^{N} \langle x_i, u \rangle x_i, u \rangle$$

$$= \frac{1}{N} \sum \langle x_i, u \rangle^2$$

$$\Rightarrow \min \hat{S}^2 \Leftrightarrow \max \sum \langle x_i, u \rangle^2 = \max \langle \hat{C}u, u \rangle$$

$\max \langle \hat{C}u, u \rangle = \lambda_1$ and $u = v_1$ by Thm 3.2 $\square$

Generally,

$$\min \hat{S}^2 = \sum_{i=1}^{N} \| x_i \|^2 - \sum_{i=1}^{N} \sum_{k=1}^{p} \langle x_i, u_k \rangle^2$$

$$\Leftrightarrow \max \sum_{i=1}^{N} \sum_{k=1}^{p} \langle x_i, u_k \rangle^2$$

$$= \max \sum_{k=1}^{p} \langle \hat{C}u_k, u_k \rangle$$

$$= \max \sum_{k=1}^{p} \langle \sum_{j=1}^{\infty} \hat{\lambda}_j \langle u_k, \hat{v}_j \rangle \hat{v}_j, u_k \rangle$$

$$= \max \sum_{k=1}^{p} \sum_{j=1}^{\infty} \hat{\lambda}_j \langle u_k, \hat{v}_j \rangle^2 \leq \sum_{k=1}^{p} \hat{\lambda}_k \quad \text{by thm 3.2}$$

and maximum is attained if $u_1 = \hat{v}_1 \cdots u_p = \hat{v}_p$

Section 3.3 Functional Principal Components

Section 3.3, we identif[y] the funct. princ. scor[es] with the eigenfunction[s] of the covariance operat[or] and show how its eigenv[alues] decompose the variance [of] the functional data.

Suppose $X_1, X_2, \dots, X_N$ are functional observations.

$\boxed{\text{EFPC}}$ → Empirical Functional Principal Components
(The eigenfunctions of the sample covariance operator.)

1

* If these observations have the ~~same~~ have the same distribution as a square integrable L^2 valued random function X, we define the FPCs as the eigenfunctions of the covariance operator.

Recall: If X is square integrable, $E\|X\|^2 = E\int X^2(t)dt < \infty$

and $E(X)=0$, then $C(y) = E[\langle X,y\rangle, X]$, $y \in L^2$.

$\mu(t) = E(X(t))$ (mean function)

$c(t,s) = E[(X(t)-\mu(t))\cdot(X(s)-\mu(s))]$ (covariance function)

$C = E[\langle X-\mu, \cdot\rangle (X-\mu)]$ (covariance operator)

$\hat{\mu}(t) = \frac{1}{N}\cdot\sum_{i=1}^{N} X_i(t)$ (sample mean function)

$\hat{c}(t,s) = \frac{1}{N}\cdot\sum_{i=1}^{N}(X_i(t)-\hat{\mu}(t))(X_i(s)-\hat{\mu}(s))$ (sample cov. function)

$\hat{C}(x) = \frac{1}{N}\cdot\sum_{i=1}^{N}\langle X_i-\hat{\mu}, x\rangle (X_i-\hat{\mu})$, $x \in L^2$ (sample cov. operator)

* Parseval's Eq. ⇒ $\sum_{j=1}^{\infty} |\langle X, v\rangle|^2 = \|X\|^2$

---

* Let v_j, λ_j, $j \geq 1$ be the eigenfunctions and eigenvalues of the covariance operator C. The relation $Cv_j = \lambda_j v_j$ implies that

$$\boxed{\lambda_j = \langle Cv_j, v_j\rangle}$$

$$\lambda_j = \langle C v_j, v_j \rangle = \langle E[\langle X, v_j \rangle X], v_j]$$

$c(v_j), v_j$

Cv_j, v_j

$$= \langle E\left[\int X(t)X(s)v_j(t)dt\right], v_j \rangle$$

$$= \int E\left[\int X(t)X(s)v_j(t)dt\right] v_j(s)ds$$

$$= \iint E\left[X(t)X(s)\right] v_j(t)v_j(s)dtds$$

$$= E\left[\iint X(t)X(s)v_j(t)v_j(s)dtds\right]$$

$\star$ Section 2.3 pg. 23

If $\psi \in \mathcal{L}$ and X is integrable, then $E[\psi(X)] = \psi[E(X$

$$= E\left[\int X(t)v_j(t)dt\right]^2 = E\left[\langle X, v_j \rangle^2\right]$$

$\star\star$ Section 3.2 explains that the EFPCs can be interpreted as an optimal orthonormal basis.

(2) $\langle X_i, \hat{v}_j \rangle = \int X_i(t) \cdot \hat{v}_j(t)dt \Rightarrow$ j-th score of X_i

$\star$ Weight of the contribution of the FPC $\hat{v}_j$ to the curve

Recall: $\hat{C}(x) = \frac{1}{N} \cdot \sum_{i=1}^{N} \langle X_i - \hat{\mu}, x \rangle (X_i - \hat{\mu})$, $x \in L^2$

$\langle \hat{C}x, x \rangle$ or $\langle \hat{C}(x), x \rangle$

$$\langle \hat{C}x, x \rangle = \left\langle \frac{1}{N} \cdot \sum_{i=1}^{N} \langle X_i - \hat{\mu}, x \rangle (X_i - \hat{\mu}), x \right\rangle$$

Assume $E(X_i) = 0$

$$\langle \hat{C}x, x \rangle = \left\langle \frac{1}{N} \cdot \sum_{i=1}^{N} \langle X_i, x \rangle X_i, x \right\rangle$$

$$= \frac{1}{N} \left\langle \sum_{i=1}^{N} \int X_i(t) x(t) dt \cdot X_i(s), x \right\rangle$$

$$= \ldots$$

$$= \frac{1}{N} \cdot \sum_{i=1}^{N} \iint X_i(t) X_i(s) x(t) x(s) dt ds$$

$$= \frac{1}{N} \cdot \sum_{i=1}^{N} \left[\int X_i(t) x(t) dt \right] \cdot \left[\int X_i(s) x(s) ds \right]$$

$$= \frac{1}{N} \cdot \sum_{i=1}^{N} \left[\int X_i(t) x(t) dt \right]^2$$

$$= \frac{1}{N} \cdot \sum_{i=1}^{N} \langle X_i, u \rangle^2$$

Hence, $\underbrace{\frac{1}{N} \cdot \sum_{i=1}^{N} \langle X_i, x \rangle^2 = \langle \hat{C}x, x \rangle}$

Sample variance of the data in the direction x.
of the function

We must find x by maximizing $\langle \hat{C}x, x \rangle$ subject to

$\|x\| = 1 \Rightarrow$ By Th. 3.2, $\underbrace{x = \hat{v}_1}_{\text{The first EFPC}}$.

Observe that since $\hat{v}_j$, $j = 1, \dots, N$ form a basis in $\mathbb{R}^N$,

$$\frac{1}{N} \cdot \sum_{i=1}^{N} \|x_i\|^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{j=1}^{N} \langle x_i, \hat{v}_j \rangle^2$$

Parseval's eq. $\sum_{j=1}^{\infty} |\langle X, v \rangle|^2 = \|X\|$

$$= \sum_{j=1}^{N} \frac{1}{N} \cdot \sum_{i=1}^{N} \langle x_i, \hat{v}_j \rangle^2$$

$$= \frac{1}{N} \sum_{j=1}^{N} (C\hat{v}_j, \hat{v}_j)$$

$$= \frac{1}{N} \cdot \sum_{j=1}^{N} \hat{\lambda}_j$$

RESULT: Variance in the direction $\hat{v}_j$ is $\hat{\lambda}_j$.

$\hat{v}_j$ explains the fraction of the total sample variance equal to $\dfrac{\hat{\lambda}_j}{\sum_{k=1}^{N} \hat{\lambda}_k}$

$$X_n(t) = \sum_{j=1}^{p} a_j \cdot Z_{jn} \cdot e_j(t)$$

a_j : real numbers for every n

$Z_{jn} \overset{iid}{\sim} (0,1)$

$e_j(t)$ orth. functions with $\|e_j\| = 1$

★ Denote by X a random function with the same distribution as $X_n(t)$.

$$X(t) = \sum_{j=1}^{p} a_j \cdot Z_j \cdot e_j(t)$$

By definiton;

$$C(x)t = E\left[\left(\int X(s) \cdot x(s)\,ds\right) X(t)\right]$$

$$= E\left[\int X(t) X(s) x(s)\,ds\right] = \int \underbrace{E[X(t)X(s)]}_{c(t,s) \text{ cov. function}} x(s)\,ds$$

$$c(t,s) = E[X(t)X(s)] = E\left[\sum_{j=1}^{p} a_j \cdot Z_j \cdot e_j(t) \sum_{i=1}^{p} a_i \cdot Z_i \cdot e_j(s)\right]$$

Spectral Decomposition

$$= E\left[\sum_{j=1}^{p} a_j^2 \cdot Z_j^2 \cdot e_j(t) e_j(s)\right]$$

$$= \sum_{j=1}^{p} a_j^2 \cdot \underbrace{E(Z_j^2)}_{1} \cdot e_j(t) e_j(s) = \sum_{j=1}^{p} a_j^2 e_j(t) e_j(s)$$

$$\boxed{\lambda_j = a_j^2 \quad , \quad v_j = e_j}$$