Week 3

Let $X \sim N(0, \Sigma)$ on $\mathbb{R}^n$.

- Find a matrix $R \in \mathbb{R}^{n \times n}$ s.t. $R^* R = \Sigma$.

Simulate RZ, $Z \sim N(0, I_{n \times n})$

i.e. X and RZ have the same distribution

- Suppose Σ is positive definite.

Σ has eigenvalues $\lambda_1, \dots, \lambda_n$ with associated eigenvectors $e_1, \dots, e_n$

then
$$\Sigma = \lambda_1 \underbrace{e_1 e_1^T}_{} + \dots + \lambda_n e_n e_n^T$$

Write
$$X = \sum_{i=1}^{n} \langle X, e_n \rangle e_n$$

Consider the random variables $\langle X, e_n \rangle = \xi_n$

$$\mathbb{E}[\xi_n \xi_m] = \mathbb{E}[\langle X, e_n \rangle \langle X, e_m \rangle]$$

$$= \mathbb{E}\left[e_n^T X^T X e_m \right]$$

$$= e_n^T \Sigma e_m$$

$$= e_n^T (\lambda_m e_m) = \begin{cases} \lambda_m & \text{if } m = n \\ 0 & \text{otherwise} \end{cases}$$

We know that $\xi_1, \dots, \xi_m$ are independent $N(0, \lambda_i)$ random variables.

Then writing

$$\lambda_1^{\frac{1}{2}} \chi_1 e_1 + \dots + \lambda_n^{\frac{1}{2}} \chi_n e_n$$

where $\chi_1, \dots, \chi_n \sim N(0, I)$

then this has the same dist as X.

KARHUNEN-LOÈVE EXPANSION /

FUNCTIONAL PRINC COMPONENT ANALYSIS

Let $f \in L^2([0,1])$, and let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis in $L^2([0,1])$.

Then,

$$f = \sum_{n=1}^{\infty} f_n e_n.$$

$$f_n = \langle f, e_n \rangle = \int_0^1 f(x) e_n(x)\, dx$$

i.e.

$$\lim_{N \to \infty} \left\| f - \sum_{n=1}^{N} f_n e_n \right\|_{L^2} = 0$$

Consider $f = X_t$ on $[0,1]$.

Assume: $\mathbb{E}[X_t] = m(t) = 0$

$Cov(X_t, X_s) = \mathbb{E}[X_t X_s] = \gamma(t,s)$

$$X_t = \sum_{n=1}^{\infty} \xi_n e_n(t)$$

$$\langle X_t, e_n \rangle$$

$$\boxed{\text{Suppose } \mathbb{E}[\xi_n \xi_m] = \lambda_n \delta_{n,m}}$$

numbers to be determined.

Consider the covariance

$$\gamma(t,s) = \mathbb{E}[X_t X_s] =$$

$$\mathbb{E}\left[\sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \xi_k e_k(t) \xi_\ell e_\ell(s)\right]$$

$$= \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \underbrace{\mathbb{E}[\xi_k \xi_\ell]}_{\lambda_k \delta_{k,\ell}} e_k(t) e_\ell(s)$$

$$= \boxed{\sum_{k=1}^{\infty} \lambda_k e_k(t) e_k(s)}$$

$$\underline{\gamma(t,s)} = \sum_{k=1}^{\infty} \lambda_k e_k(t)\, e_k(s)$$

Consider the covariance operator K for X

$$\underline{Ke_n} = \int_0^1 \gamma(t,s)\, e_n(s)\, ds$$

$$= \int_0^1 \sum_{k=1}^{\infty} \lambda_k e_k(t)\, \underline{e_k(s) e_n(s)}\, ds$$

$$= \sum_{k=1}^{\infty} \lambda_k e_k(t)\, \delta_{n,k}$$

$$= \underline{\lambda_n e_n(t)}$$

$\therefore$ 1 The $\{e_1, \ldots\}$ must nec. be the eigenfunctions associated with K.

2 The $\{\lambda_1, \ldots\}$ must be the assoc eigenvalues of K.

Thm (Mercer's theorem).

- γ is cts, symmetric, non-negative definite.

Then (1) there is an orthonormal basis $\{e_1, e_2, \ldots\}$ of $L^2([0,1])$ which are eigenfunctions of K, assoc with the non-zero eigenvals of K.

(2) These eigenfunctions are cts on $[0,1]$ &

$$\gamma(s,t) = \sum_{n=1}^{\infty} \lambda_n e_n(s) e_n(t)$$

Thm ① Let $\{X_t, t \in [0,1]\}$ be an $L^2([0,1])$ process with mean 0, cts covariance function γ.

② Let $\{\lambda_n, e_n\}_{n=1}^{\infty}$ be the eigenbasis associated with the covariance operator K.

i.e. $(Kf)(t) = \int_0^1 \gamma(t,s)\, f(s)\, ds$.

THEN

$$X_t = \sum_{n=1}^{\infty} \xi_n \underbrace{e_n(t)}_{\text{principal components}}, \quad t \in [0,1]$$

$$\xi_n := \int_0^1 X_t\, e_n(t)\, dt,$$

$$\boxed{\mathbb{E}\,\xi_n = 0, \quad \mathbb{E}[\xi_n \xi_m] = \lambda_n \delta_{n,m}}$$

(ξ_n) are uncorrelated.

Eg. If X_t is a Gaussian Process

$\Rightarrow (\xi_n)_{n \geq 1}$ are independent, s.t.

- $\xi_n \sim N(0, \lambda_n)$

then we can simulate X by

$$\boxed{X = \sum_{n=1}^{\infty} \sqrt{\lambda_n}\, X_n\, e_n,}$$

$X_1, X_2, \dots$ are iid $N(0,1)$

Ex The KL expansion for Brownian Motion.

W_t : BM is a Gaussian process with covariance

$$\gamma(t,s) = \min(t,s).$$

The eigenvalue process:

$$K e_n = \lambda_n e_n, \qquad Kf = \int_0^1 \underbrace{\gamma(t,s)}_{\min(t,s)} f(s)\,ds$$

$\vdots$

$$e_n''(t) = -\frac{1}{\lambda_n} e_n(t). \qquad e_n(0) = e_n'(1) = 0$$

$$\Rightarrow e_n(t) = \sqrt{2}\sin\left(\tfrac{1}{2}(2n-1)\pi t\right)$$

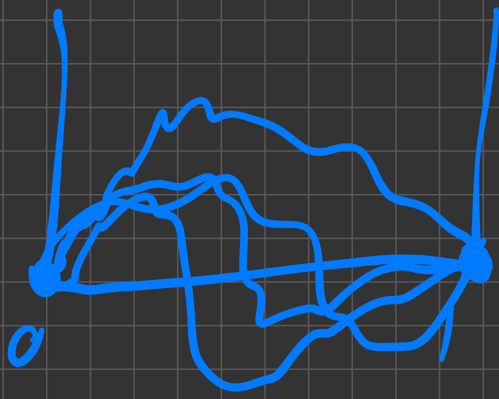
$$\lambda_n = \left(\frac{2}{(2n-1)\pi}\right)^2$$

$$\boxed{W_t = \sqrt{2}\sum_{n=1}^{\infty} \frac{2}{(2n-1)\pi} X_n \sin\left(\tfrac{1}{2}(2n-1)\pi t\right)}$$

where $X_1, X_2, \ldots$ are iid $N(0,1)$

Example: Brownian Bridge.

i.e. $\gamma(t,s) = \min(t,s) - ts$



Consider $\underline{\mathbb{E}\|X_t\|^2}$

$$= \mathbb{E}\left\langle \sum_n \underset{\sim}{\xi_n} e_n , \sum_m \underline{\xi_m} e_m \right\rangle$$

$$= \mathbb{E} \sum_{n,m} \xi_n \xi_m \delta_{nm}$$

$$= \sum_n \mathbb{E}[\xi_n^2] = \underbrace{\sum_n \lambda_n}_{\mathrm{Var}(\xi_n)} = \mathrm{tr}(K) < \infty$$

THE KL EXPANSION THROUGH OPTIMISATION

Define

$$S(u_1, \dots, u_m) = \mathbb{E} \left\| \underline{X} - \sum_{i=1}^{M} \langle X, u_i \rangle u_i \right\|^2$$

for $\{u_1, \dots, u_m\}$ being an orthonormal system.

$$S(u_1, \dots, u_m) =$$

$$\mathbb{E}\left[\left\langle X - \sum_{i=1}^{m} \langle X, u_i \rangle u_i ,\ X - \sum_{j=1}^{m} \langle X, u_j \rangle u_j \right\rangle \right]$$

$$= \mathbb{E}\|X\|^2 - \underbrace{\sum_{i=1}^{m} \mathbb{E}\langle X, u_i \rangle^2}_{\sum_{i=1}^{m} \langle K u_i, u_i \rangle}$$

$$\therefore S(\underline{u}) = \mathbb{E}\|X\|^2 - \sum_{i=1}^{M} \langle K u_i, u_i \rangle$$

Objective: maximise $\sum_{i=1}^{m} \langle K u_i, u_i \rangle$ over $u_1, \dots, u_m$ ONS

$$\boxed{\max_{u_1 : \|u_1\| = 1} \langle u_1, K u_1 \rangle = \lambda_1 = \text{largest e.v.}}$$

The maximising u_1 is e_1

If eigenvalues satisfy $\lambda_1 > \lambda_2 > \lambda_3 \cdots$

then

$u_1 = e_1$

$u_2 = e_2$

$\vdots$

Thm 1: If X is square integrable with mean 0 and K has eigenvalues $\lambda_1 > \lambda_2 > \lambda_3, \cdots$, then

$S(u_1, \ldots, u_m)$ is minimised for

$u_1, u_2, \ldots = e_1, e_2, \cdots$ which

are eigenf^{n}s assoc with $\lambda_1, \lambda_2, \cdots$

Suppose we have access to $X_1, \ldots, X_n$ iid realisations of X.
- Can compute $\hat{\gamma}$ of the covariance function.
- Can compute $\hat{K}$ the covariance operator.
- Can compute eigenvalues and eigenfunctions

$\{\hat{\lambda}_1, \dots, \hat{\lambda}_n, \dots\}$

$\{\hat{e}_1, \dots, \hat{e}_n, \dots\}$

empirical functional principal components.

We require that $\|\hat{e}_i\| = 1$, $\langle \hat{e}_i, \hat{e}_j \rangle = \delta_{ij}$

Only hope is that $\hat{c}_j \hat{e}_j \approx e_j$

where $\hat{c}_j = \operatorname{sign}(\langle \hat{e}_j, e_j \rangle)$.

The EMPIRICAL PRINC COMP PROBLEM:

$$\int \hat{\gamma}(t,s)\, \hat{e}_j(s)\, ds = \hat{\lambda}_j \hat{e}_j, \quad j = 1, \dots, m.$$

Thm: If $\mathbb{E}\|X\|^4 < \infty$ and the eigenvalues are ordered and distinct then

- $\limsup_{N \to \infty} N \mathbb{E}\|\hat{e}_j - e_j\|^2 < \infty$
- $\limsup_{N \to \infty} N \mathbb{E}|\hat{\lambda}_j - \lambda_j|^2 < \infty$

If $\mathbb{E}\|X\|^4 < \infty$

- $\sqrt{N}\,(\hat{\lambda}_j - \lambda_j) \xrightarrow[n\to\infty]{D} N(0, 2\lambda_j^2)$
- $\sqrt{N}\,(\hat{e}_j - e_j) \xrightarrow[n\to\infty]{D} N(0, C_j)$

Once we have FPCs $\hat{e}_1, \dots, \hat{e}_m$, then we represent X_i as

$$\mathbf{x}_i = [\langle X, e_1\rangle, \langle X, e_2\rangle, \dots \langle X_i, e_m\rangle]^T$$

Recall

$$\mathbb{E}\|X\|^2 = \sum_{j=1}^{\infty} \lambda_j = \mathrm{tr}(K) < \infty$$

$$100 \times \frac{\lambda_j}{\sum_{\ell=1}^{\infty} \lambda_\ell} = \% \text{ var explained by } \lambda_j$$

Cumulative Percentage of Total Variance

$$CPV(m) = \frac{\sum_{k=1}^{m} \hat{\lambda}_k}{\sum_{k=1}^{N} \hat{\lambda}_k}$$

Choose M s.t. $CPV(M) \geq 0.85$.