Week 1:

- WHAT IS FDA?
  - What does 'functional data' look like?
  - What do we do with functional Data
- Smoothing and Interpolation.

1. What is FD?

- Every data point REPRESENTS a function. (curve)

$$f : [0,1] \to \mathbb{R}.$$

- Underlying 'data generating' distribution/process is a stochastic process:

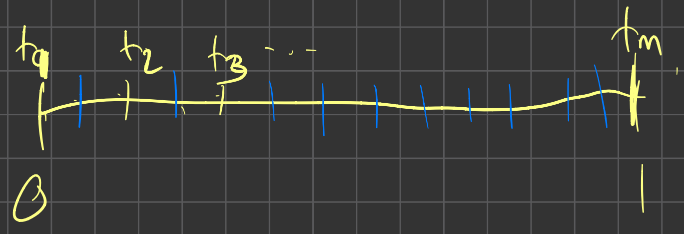
$$\left\{ \{ X(t,\omega) : t \in I \subseteq \mathbb{R}^d, \omega \in \Omega \} \right.$$

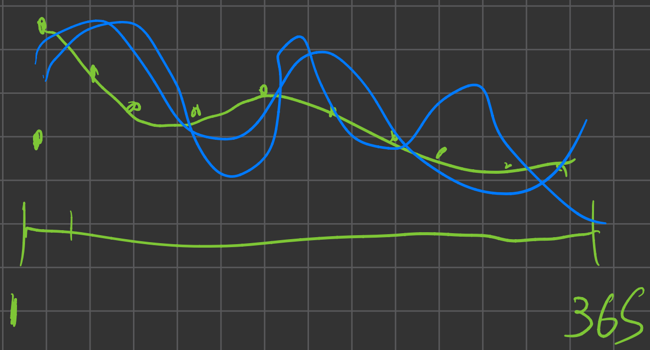
For each t: $X(t,\cdot)$ is a $\mathcal{F}$-meas. Random variable when $(\Omega, \mathcal{F}, \mathbb{P})$.

Assume $I = [0,1]$ "index set"

Typical data $\{ \underline{x}_1, \dots, \underline{x}_n \}$,

$$\underline{x}_i = (X_i(t_j))_{j=1}^{m}, \in \mathbb{R}^m$$





If m is the dimension of my data

N is the number of samples

In FDA I must allow that $M \gg N$

Ex Suppose we observe $\underline{X_1}, \dots, \underline{X_N}$ as above.

$X_1, \dots X_N$ IID with mean μ unknown.

Is $\mu = 0$?

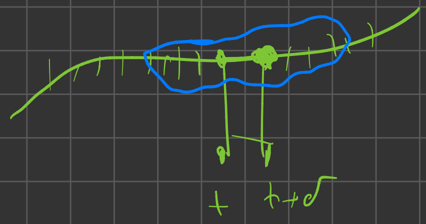
Natural test is the T-squared test:

$$T^2 = \underline{\bar{X}} \, \hat{\Sigma}^{-1} \, \underline{\bar{X}}$$

sample covariance of my data.

If $m > N$ then $\hat{\Sigma}$ is singular

Property of FD → Smoothness/Locality



GRIP: (Grid Refinement Invariance Principle)

- Develop methods/algorithm which work on functions

- Develop a method for finite dimensional data, then see if it has a well-defined limit.

Smoothing or Interpolation: Reconstruct a curve from discrete observations

- Basis Expansions (Parametric Approach)
- Kernel Based Smoothing (Nonparametric approaches)

① Model: Observations $(t_1, y_1), \dots, (t_J, y_J)$

$$y_j = X(t_j) + \varepsilon_j, \quad j = 1, \dots, J$$

$$x(t) = \sum_{m=1}^{M} c_m \Phi_m(t)$$

$\{\Phi_1, \Phi_2, \dots\}$ typically some basis system of functions.

Standard error model:

Find $\underline{c} \in \mathbb{R}^m$ which minimises $L(y|\underline{c})$

$$L(y|\underline{c}) = \sum_{j=1}^{J} \left\| y_j - \sum_{m=1}^{m} c_m \Phi_m(t_j) \right\|^2$$

$$\underline{\Phi}_m := (\phi_m(t_1), \dots, \phi_m(t_J))^T \in \mathbb{R}^J$$

$$\underline{\Phi} = [\underline{\Phi}_1 | \dots | \underline{\Phi}_m] \in \mathbb{R}^{J \times M}$$

$$L(y|\underline{c}) = \|y - \underline{\Phi}\underline{c}\|^2$$

$$\hat{\underline{c}} = (\underline{\Phi}^T \underline{\Phi})^{-1} \underline{\Phi}^T y$$

Suppose $\underline{t}^* = (t_1^*, \dots, t_K^*)$

Define $\underline{\Phi}_m^* = (\phi_m(t_1^*), \dots, \phi_m(t_K^*))^T$

$\underline{\underline{\Phi}}^*$ accordingly

Predicted value at $\underline{t}^*$:

$$\hat{y}^* = \underbrace{\boxed{\underline{\underline{\Phi}}^* (\underline{\underline{\Phi}}^T \underline{\underline{\Phi}})^{-1} \underline{\underline{\Phi}}^T}}_{S(\underline{t}^*)} \underline{y}$$

Linear smoother

$$\hat{y}^* = \boxed{\underline{\underline{S}}(\underline{t}^*)}\, \underline{y}$$

- Map from $\mathbb{R}^J \to \mathbb{R}^K$ linear

$$\underline{\underline{S}}(\cdot)\underline{y}$$

- Map from $\mathbb{R}^J \to$ function space

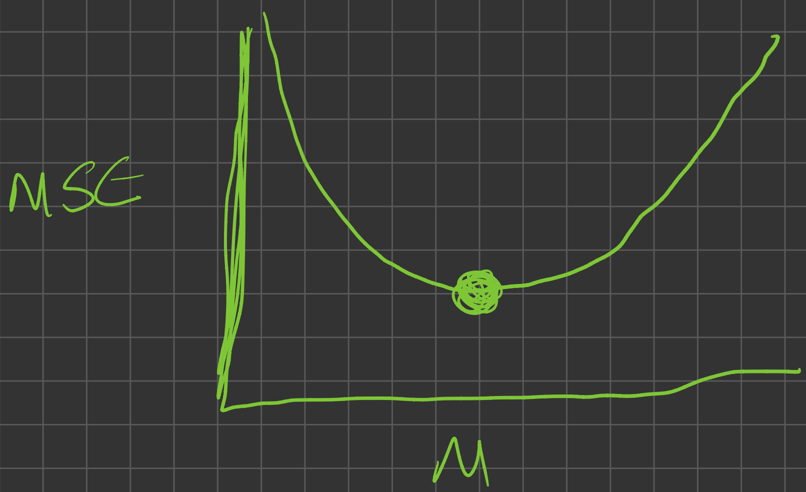
$$S(t)\, S(t) = S(t)$$

What value should M be?

As $M \to \infty$ bias $\to 0$

$$MSE[\hat{x}(t)] = \mathbb{E}|\hat{x}(t) - x(t)|^2$$

$$= \underbrace{Bias[\hat{x}(t)]}_{x(t) - \mathbb{E}[\hat{x}(t)]}^2 + Var(\hat{x}(t))$$



Kernel based smoothing:

$$\hat{x}(t) = \sum_{j=1}^{J} S_j(t)\, y_j$$

NADARYA-WATSON:

$$S_j(t) = \frac{k\left(\frac{t-t_j}{h}\right)}{\sum_r k\left(\frac{t-t_r}{h}\right)}$$

k  ex. $k(y) = \frac{e^{-y^2/2h^2}}{\sqrt{2\pi h^2}}$

h = Bandwith.

$$P_2(x) = \int \|D^2 x(s)\|^2 ds.$$

<u>Penalised Least squares:</u>

$$L_\lambda(y \mid x) = \|y - x(\underline{t})\|^2 + \lambda P_2(x)$$

Penalises misfit

Penalises roughness

If $x(t) = \sum c_m \phi_m(t) = \underline{\Phi}(t)\underline{c}$

Unpenalised estimator:

- $\hat{C} = (\underline{\Phi}^T \underline{\Phi})^{-1} \underline{\Phi}^T y$

- $$P_2(\underline{\Phi}\underline{c}) = \underline{c}^T \overbrace{\left[\int [D^2 \underline{\Phi}(s)][D^2 \underline{\Phi}(s)]^T ds \right]}^{R} \underline{c}$$

$$\underbrace{\hphantom{c^T R c}}_{\text{quadratic form}}$$

$$L_\lambda(y \mid \underline{c}) = \underline{\|y - \underline{\Phi}\underline{c}\|^2} + \lambda c^T \underline{R} c$$

$$\hat{c} = (\underline{\Phi}^T \underline{\Phi} + \lambda \underline{R})^{-1} \underline{\Phi}^T y$$

$$\hat{y}(t) = \Phi(t)\left(\Phi^T\Phi + \lambda R\right)^{-1}\Phi^T y$$

$$= S_\lambda(t)\, y$$

Generalized Cross-Validation (GCV)

Craven & Wahba 1978.