24.11:

$$X_{m+1} = A X_m + W_m \qquad W \sim N(0,Q)$$

$$Y_m = H X_m + V_m \quad X_0 \sim N(m, C)$$
$$V_m \sim N(0, R)$$

Know. A, H, Q, R, C, m,

Data Y_m.

$$X_1^F = A X_0 + W_1$$

Forecast $\begin{cases} m_1^F = Am \\ P_1^F = ACA^T + Q \end{cases}$

Kalman gain $K_1 = P_1^F H^T (H P^F H^T + R)^{-1}$

Analysis $\begin{cases} X_1^A = X_1^F - K(Y_1 - H X_1^F) \\ m_1^A = m_1^F - K(Y_1 - H m_1^F) \\ P_1^A = (I - KH) P_1^F \end{cases}$

$X_m \in \mathbb{R}^2$

$X_m = \begin{bmatrix} X_m^1 \\ X_m^2 \end{bmatrix}$

$Y_m \in \mathbb{R}^1$

$\Rightarrow H \in \mathbb{R}^{1\times 2}$

$H = [1 \ 0]$

$Y_m = H X_m = X_m^1$

Ex. $H = [1 \ 1]$

$Y_m = X_1^m + X_2^m$

Ensemble Kalman Filter:

$$X_{jk}^F = A X_{jk}^A + W_{jk}$$

Forecast. $\begin{cases} m_j^F = \frac{1}{n} \sum_{n=1}^{n} X_{jk}^F \\ P_j^F = \frac{1}{n-1} \sum_{n=1}^{n} (X_{jk}^F - m_j^F)(X_{jk}^F - m_j^F)^T \end{cases}$

$$K_j = P_j^F H^T (H P_j^F H^T + R)$$

$$X_{jk}^A = X_{jk}^F - K(Y_{jk} - H X_{jk}^F)$$

$$m_j^A = \frac{1}{n} \sum_{n=1}^{n} X_{jk}^A$$

$$P_j^A = \frac{1}{n-1} \sum_{n=1}^{n} (X_{jk}^A - m_j^A)(X_{jk}^A - m_j^A)^T$$

$$Y_{jk} = Y_j + \varepsilon_k$$

Sheet 5:

$$\lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$$; Θ= Parameter space e.g. $\mathbb{R}$

$H_0: \theta \in \Theta_0$

$H_1: \theta \in \Theta_0^c$

$\Theta = \Theta_0 \cup \Theta_0^c$

==x==

01.12:

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

$$f_X(x|\theta) = \exp(i\theta - x) \cdot \mathbb{1}_{\{x \geq i\theta\}}(x)$$

$$\exp(i\theta - x) \cdot \begin{cases} 1 & x \geq i\theta \\ 0 & x < i\theta \end{cases}$$

$$1_A(x) \cdot 1_B(x)$$

$$= \begin{cases} 1 \cdot 1 & x \in A, x \in B \\ 0 \cdot 1 & x \notin A, x \in B \\ 1 \cdot 0 & x \in A, x \notin B \\ 0 \cdot 0 & x \notin A, x \notin B \end{cases} = \begin{cases} 1, & x \in A \cap B \\ 0, & \text{else} \end{cases}$$

$$= 1_{A \cap B}(x)$$

$X, Y \in \mathbb{R}$

$1_{[a,b]}(x) \cdot 1_{[a,b]}(Y)$

$$= \begin{cases} 1 \cdot 1 & a \leq x, Y \leq b \\ 0 \cdot 1 & a \leq y \leq b,\ a \notin [a,b] \\ 1\ 0 & x \in [a,b],\ Y \notin [a,b] \\ 0\ 0 & X, Y \in [a,b] \end{cases}$$

$$= \begin{cases} 1, & X, Y \in [a,b] \\ 0, & \text{else} \end{cases}$$

$\min\{X, Y\} \in [a,b], \quad \max\{X, Y\} \in [a,b]$

$\Rightarrow X, Y \in [a,b]$

$1_{\{x \geq \theta\}}\left(\min \frac{X_i}{i}\right)$

$1_{\{x \geq i\theta\}}(x)$

$= 1_{\{\frac{x}{i} \geq \theta\}}(x)$

$= 1_{\{x \geq \theta\}}\left(\frac{x}{i}\right)$

$f_{x_i}(x|\theta) = \exp(i\theta - x) \cdot 1_{\{x \geq i\theta\}}(x)$

$$f(x\ x_m|\theta) = \prod_{i=1}^{n} \exp(i\theta - x_i) \cdot \prod_{i=1}^{n} 1_{\{x \geq \theta\}}\left(\frac{x_i}{i}\right)$$

$$\boxed{\begin{aligned} &\exp\left(\sum i\theta - \sum x_i\right) \\ &= \exp\left(\theta \cdot \frac{n(n-1)}{2} - \sum x_i\right) \\ &\prod_{i=1}^{n} 1_{\{x \geq \theta\}}\left(\frac{x_i}{i}\right) \\ &= 1_{\{x \geq \theta\}} \min \frac{x_i}{i} \end{aligned}}$$

$$= \exp\left(\theta \cdot \frac{n(n-1)}{2} - \sum x_i\right) \cdot 1_{\{x \geq \theta\}} \min \frac{x_i}{i}$$

$$= \exp\left(\theta \cdot \frac{n(n-1)}{2}\right) \cdot 1_{\{x \geq \theta\}}\left(T(x_{i-1}, x_n)\right) \cdot \exp\left(-\sum x_i\right)$$

$$= \underbrace{g\left(T(x_{i-1}\ x_n)|\theta\right)} \cdot \underbrace{h(x_{i-1}, x_n)}.$$

$X_1 \ldots X_n \underset{iid}{\sim} f(x|\theta)$

$X_1, \ldots X_n \sim f_{1..n}(x_1 \ldots x_n|\theta)$

$f_{1..n}(x_{1..n}, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

$= f(x_1|\theta) \cdot f(x_2|\theta) \ldots f(x_n|\theta)$

$= \prod_{i=1}^{n} f(x_{(i)}|\theta)$

$= g(T(x_1 \ldots x_n)|\theta) \cdot h((x_1 \ldots x_n)|\theta)$

$$\frac{f(x_1 \ldots x_n|\theta)}{g(T(x_1 \ldots x_n|\theta)} = \frac{\prod_{i=1}^{n} f(x_{(i)}|\theta)}{n \prod_{i=1}^{n} f(x_{(i)}|\theta)}$$

$T(x_1, x_2, x_3) = (x_{(1)}, x_{(2)}, x_{(3)})$

$$= \frac{1}{n!} = P_0(x_{1,\ldots,n}|T(\ldots))$$

08 December:

$$g(\mu) = P_\mu\left(|Z| > z_{\alpha/2}\right)$$

$$= 1 - P_\mu\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right)$$

$$= 1 - P\left(\left|\frac{\bar{x} - \mu + \mu - \theta_0}{\sigma}\sqrt{n}\right| \leq z_{\alpha/2}\right.$$

$$= 1 - P\left(-z_{\alpha/2} + \sqrt{n}\,\frac{\theta_0 - \mu}{\sigma} \leq Z \leq z_{\alpha/2} + \sqrt{n}\,\underbrace{\frac{\theta_0 - \mu}{\sigma}}_{= v}\right)$$

$$\int_{-z_{\alpha/2}+\sqrt{n}v}^{z_{\alpha/2}+\sqrt{n}v} p(x)\cdot dx = (*)$$

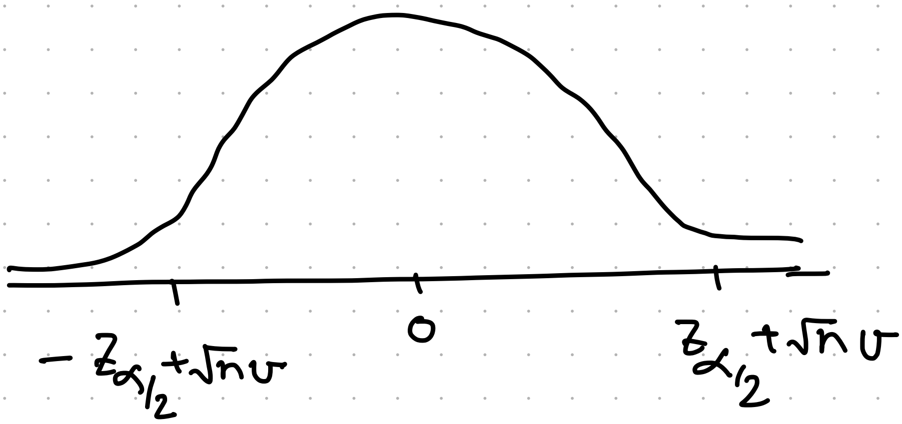
$$\frac{d}{dx}(*) = \frac{d}{dx}\left(F(z_{\alpha/2}+\sqrt{n}v) - F(-z_{\alpha/2}+\sqrt{n}v)\right)$$

$$= f(z_{\alpha/2}+\sqrt{n}v)\cdot\frac{v}{\sqrt{n}}$$

$$- f(-z_{\alpha/2}+\sqrt{n}v)\frac{v}{\sqrt{n}})$$

$$\longrightarrow 0,\ n \to \infty$$

$$= 1 - \underbrace{P(-z_{\alpha/2} + \sqrt{n}\,v \le Z \le z_{\alpha/2} + \sqrt{n}\,v)}_{F(z_{\alpha/2} + \sqrt{n}\,v) - F(z_{\alpha/2} + \sqrt{n}\,v)}$$



$$\frac{d}{dn} g(\mu) > 0$$

$$\text{Ideal}, \quad g(\mu) = \begin{cases} 0, & z \in \mathbb{R} \\ 1, & z \in R^c \end{cases}$$

$$\beta^*_{N+1}(\beta^*_N, X_N, Y_N)$$

proof of induction

$$f H: \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

$$f S;\ n = 1, \quad 1 = \frac{1(1+1)}{2} = 1.$$

Step: $n \rightarrow n+1$

$$\sum_{i=1}^{n+1} i = \underbrace{\sum_{i=1}^{n} i}_{} + n+1$$

$$= \frac{n(n+1)}{2} + (n+1)$$

$$= \frac{n(n+1)}{2} + \frac{2(n+1)}{2}$$

$$= \frac{(n+2)(n+1)}{2}$$

— x —

$$E = \sum O_K - t_K$$

$$\frac{\partial E}{\partial x^H} = \frac{\partial E}{\partial O_K} \cdot \frac{\partial O_K}{\partial a_K^H} \cdot \frac{\partial a_K^H}{\partial z^H} \cdot \frac{\partial z^H}{\partial x^H}$$

$$O_K = \sigma(\underbrace{w^H x^H + b^H}_{z^H})$$

$$a^H = \sigma(zH).$$

Extra:

$$Y = AX$$

$Y \in \mathbb{R}^m$ $\quad$ $A \in \mathbb{R}^{m \times n}$

$X \in \mathbb{R}^n$

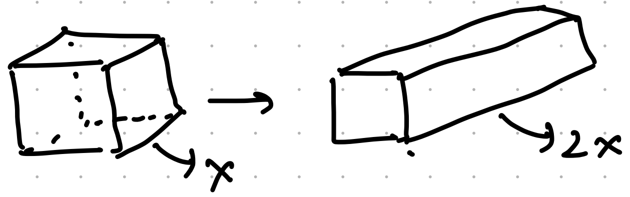
vol V of point cloud.

$A \in \mathbb{R}^{m \times n}$

vol W

$W = \det A \cdot V$

$V =$



$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \det A = 2$$

v is an EV with Eigen value λ

$\Leftrightarrow A v = \lambda v$

$v \in \mathbb{R}^m, \lambda \in \mathbb{R}$

$[v_{1, \dots, m}] = V$ orthonormal.

$$A = V^T \Sigma V \cdot \quad ,$$

$$\uparrow$$

$$\begin{bmatrix} \lambda_1 & \cdots & & 0 \\ \vdots & \lambda_2 & & \vdots \\ & & \ddots & \\ 0 & \cdots & & \lambda_n \end{bmatrix}$$

$$\det A = \prod_{i=1}^{n} \lambda_i$$

$$\det A = \det (V^T \Sigma V)$$

$$= \det V^T \cdot \det \Sigma , \det V$$

$$= (\det V)^{-1} \cdot \det \Sigma \cdot \det V$$

$$= \det \Sigma$$

$$= \prod_{i=1}^{n} \lambda_i$$

(1) Exercise 5.3:

$$LRT = \begin{cases} \exp(n(\theta - X_{(1)})) , & X_{(1)} > \theta \\ 1, & , X_{(1)} \leq \theta . \end{cases}$$

(2) $\beta = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

By calculation, $\hat{\beta} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$

$$\hat{\beta}_{Ridge} = \begin{bmatrix} 7 \\ 2 \end{bmatrix}$$

which one is good $\hat{\beta}$ or $\hat{\beta}_{Ridge}$??

$$\|\beta - \hat{\beta}\|_2 \,/\, \|\beta - \hat{\beta}_{Ridge}\|_2$$

15 December:

Problem 1: $B_N^* = (x_N^T x_N)^{-1} x_N^T Y_N$

$$x_N = \begin{bmatrix} 1 & x_{1,1} \cdots x_{1,n} \\ \vdots & \vdots \\ 1 & x_{m,1} \cdots x_{m,n} \end{bmatrix}$$

$$\beta^*_{N+1} = (x_{N+1}^T x_{N+1})^{-1} x_{N+1}^T Y_N$$

$$\Xi_{N+1} = \begin{bmatrix} x_N \\ x_{N+1} \end{bmatrix} \qquad Y_{N+1} = \begin{bmatrix} Y_N \\ Y_{N+1} \end{bmatrix}$$

$$(x_{N+1}^T x_{N+1})^{-1} = (x_N^T x_N + x_{N+1}^T x_{N+1})^{-1}$$

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} u v^T A^{-1}}{1 + u^T A^{-1} v}$$

[Sherman Morrwood Lobrry]

$$(\Xi_{N+1}^T \Xi_{N+1})^{-1} = (\Xi_N^T \Xi_N + x_{N+1}^T x_{N+1})^{-1}$$

$$= \underbrace{(\Xi_N^T \Xi_N)^{-1}}_{A} - \alpha (\Xi_N^T \Xi_N)^{-1} x_{N+1}^T x_{N+1} (\Xi_N^T \Xi_N)^{-1}$$

$$\alpha = \left(1 + x_{N+1} (X_N^T X_N)^{-1} x_{N+1}^T\right)$$

$$\beta_{N+1}^* = (X_{N+1}^T X_{N+1})^{-1} X_{N+1}^T Y_{N+1}$$

$$= (X_N^T X)^{-1} (X_N Y_N + x_{N+1}^T Y_{N+1})$$

$$- \alpha (X_N^T X_N)^{-1} x_{N+1}^T x_{N+1} (X_N^T X)^{-1}$$

$$(X_N^T Y_N + x_{N+1}^T Y_{N+1})$$

$$K := \alpha (x_N^T x_N)^{-1} x_{N+1}^T$$

$$\beta_{N+1}^* = \beta_N^* + K (Y_{N+1} - x_{N+1} \beta_N^*)$$

$$\begin{bmatrix} X_N \\ x_{N+1} \end{bmatrix}^T \begin{bmatrix} X_N \\ x_{N+1} \end{bmatrix}$$

$$= \begin{bmatrix} X_N^T & x_{N+1}^T \end{bmatrix} \begin{bmatrix} X_N \\ x_{N+1} \end{bmatrix}$$

$$= X_N^T X_N + x_{N+1}^T x_{N+1}.$$

Problem 2:

$$E = \frac{1}{2} \sum_{k \in N_0} (o_k - t_k)^2$$

$$= \frac{1}{2} \| o - k \|_2^2$$

sigmoid function

$$O = \sigma(w^H x^H + k^H)$$

$$z^H = w^H x^H + k^H$$

$$\Rightarrow O = \sigma(z^H)$$

$$\frac{\partial E}{\partial w_{ij}^H} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial z^H} \cdot \frac{\partial z^H}{\partial w_{ij}^H}$$

'$\circ$' instead of '$\cdot$'

↳ Hadamard Production

$$a \circ b = \begin{matrix} a_1 \cdot b_1 \\ a_2 \cdot b_2 \\ \vdots \\ a_n \cdot b_n \end{matrix}$$

$$\frac{\partial E}{\partial o_k} = \frac{1}{2} \cdot 2 \cdot 1 (o_k - t_k) = o_k - t_k$$

$$\nabla_o E = \begin{pmatrix} \frac{\partial E}{\partial o_1} \\ \frac{\partial E}{\partial o_{N_o}} \end{pmatrix} = o - k \qquad \boxed{\frac{\partial E}{\partial O} = o - t}$$

$$O = \sigma(z^H)$$

$$\frac{\partial o_k}{\partial z_k^H} = \sigma'(z_k^H)$$

$$= \sigma(z_k^H) \cdot (1 - \sigma(z_k^H))$$

$$\frac{\partial O}{\partial Z} = \sigma(z^H) \circ (1 - \sigma(z^H))$$

$$\frac{\partial o_k}{\partial z_j^H} = \begin{cases} 0 & j \neq k \\ \sigma'(z_k^H) & j = k \end{cases}$$

$$O \in \mathbb{R}_1^{N_o}, \quad z^H \in \mathbb{R}^{N_o}$$

$$\begin{matrix} \frac{\partial o_1}{\partial z_1^H} & \cdots & \frac{\partial o_1}{\partial z_{N_o}^H} \\ \frac{\partial o_2}{\partial z_1^H} & \cdots & \frac{\partial o_2}{\partial z_{N_o}^H} \\ \vdots & & \\ \frac{\partial o_{N_o}}{\partial z_1^H} & \cdots & \frac{\partial o_{N_o}}{\partial z_{N_o}^H} \end{matrix}$$

$$z^H = w^H x^H + b^H$$

$$\frac{\partial z_k^H}{\partial w_{ij}^H} = \frac{\partial}{\partial w_{ij}^H} \sum_{l=1}^{N_H} w_{kl}^H \cdot x_l^H + b_k^H$$

$$= \begin{cases} x_j & \text{if } k = 1 \\ 0 & k \neq 1 \end{cases}$$

$$\frac{\partial z}{\partial w_{ij}^{H}} = x_j^{H} \cdot e_i \; ; \; e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \rightarrow \text{ith row}$$

$$\frac{\partial E}{\partial w_{ij}^{H}} = (o - t) \circ \sigma(z^{H}) \circ (1 - \sigma(z^{H})) \circ e_i \circ x_j^{H}$$

* Deep Learning: An Introduction for Applied Mathematicians

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}$$

eigen value,
vector
space

$$\det(A - \lambda I) = (\lambda - 8)(\lambda + 1)(\lambda + 1)$$

$$\therefore \lambda_1 = 8, \quad \lambda_2 = \lambda_3 = -1$$

eigen vector, $X = Z$, $Y = \frac{1}{2} Z$ $\Rightarrow$ eigen space $\begin{bmatrix} z \\ \frac{1}{2} z \\ z \end{bmatrix}$

if $z = 1$, $\begin{bmatrix} 1 \\ \frac{1}{2} \\ 1 \end{bmatrix}$

eigen vector, $\lambda = \lambda_2 = \lambda_3 = -1$

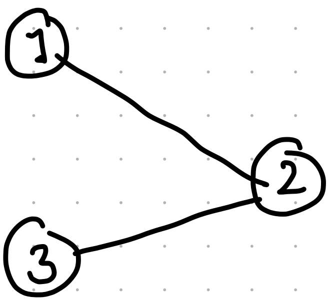
$$x + \frac{1}{2}Y = z$$

$$\begin{bmatrix} x \\ Y \\ x + \frac{1}{2}Y \end{bmatrix} = x \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + Y \begin{bmatrix} 0 \\ 1 \\ 1/2 \end{bmatrix}$$

05.01.2023:

$$\text{Ratio cut}(A, \bar{A}) = \frac{\text{cut}(A, \bar{A})}{|A|} + \frac{\text{cut}(\bar{A}, A)}{|\bar{A}|}$$

$$\text{Ncut}(A, \bar{A}) = \frac{\text{cut}(A, \bar{A})}{\text{vol}(A)} + \frac{\text{cut}(A, \bar{A})}{\text{vol}(\bar{A})}$$



$A = \{1\}$, $\bar{A} = \{2, 3\}$
$\text{cut}(A, \bar{A}) = 1$
$\text{vol}(A) = 1$
$\text{vol}(\bar{A}) = 3$

$A = \{2\}$, $\bar{A} = \{1, 3\}$
$\text{cut}(A, \bar{A}) = 2$
$\text{col}(A) = 2$
$\text{vol}(\bar{A}) = 2$

$$\begin{cases} \text{Ratio cut}(A, \bar{A}) = \dfrac{1}{1} + \dfrac{1}{2} = \dfrac{3}{2} \\ \text{Ncut}(A, \bar{A}) = \dfrac{1}{1} + \dfrac{1}{3} = \dfrac{4}{3} \end{cases}$$

$$\begin{cases} \text{Ratio cut}(A, \bar{A}) = \dfrac{2}{1} + \dfrac{2}{2} = 3 \\ \text{Ncut}(A, \bar{A}) = \dfrac{2}{2} + \dfrac{2}{2} = 2 \end{cases}$$

A	$\{1\}$	$\{1,2\}$	$\{1,3\}$	$\{2\}$	$\{2,3\}$	$\{3\}$
$\bar{A}$	$\{2,3\}$	$\{3\}$	$\{2\}$	$\{1,3\}$	$\{1\}$	$\{1,2\}$
RC	$3/2$	$3/2$	3	3	$3/2$	$3/2$
NC	$4/3$	$4/3$	2	2	$4/3$	$4/3$

$$L = D - W$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

Eig. Value	Eigen Vector
$\lambda_0 = 0$	$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$
$\lambda_1 = 1$	$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$
$\lambda_2 = 3$	$\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$

→ Algebraic connectivity of graphs.
Miroslav Fiedler.

12-01-2023:

$A \in \mathbb{R}^{n \times n}$ is diag

if $\exists\ S \in \mathbb{R}^{n \times n}$ $\det S = 0$

$$A = S^{-1} D S = \begin{bmatrix} * & & 0 \\ & * & \\ & & \ddots & \\ 0 & & & * \end{bmatrix}$$

$$A = U \Sigma V^T$$

$$U = S$$

$$S^{-1} = U^T$$

(λ_1, v_1) are eigen val, eigen vec. of A

if $\det(A - \lambda I)$ has k roots of λ_1: λ_1 has algebraic multiplicity k.

$$\det(A - \lambda I) = \lambda^2 (\lambda - 1)^9 (\lambda - 3)^1.$$

	A.M.
$\lambda_1 =$	2
$\lambda_2 =$	9
$\lambda_3 =$	3

$\dim(\mathrm{Eig}(A, \lambda_1)) = K$

then λ_1 has geometric multiplicity K

1) a) $\lambda_1 = 1 + i \quad \left\langle \begin{matrix} 1 \\ -i \end{matrix} \right\rangle$

$\lambda_2 = 1 - i \quad \left\langle \begin{matrix} 1 \\ i \end{matrix} \right\rangle$

b) $\lambda_1 = -6 \longrightarrow \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$

$\lambda_2 = 0 \longrightarrow \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$

$\lambda_3 = 4 \longrightarrow \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

c) $\lambda_1 = 1 \quad \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

$\lambda_2 = 2 \quad \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

$\lambda_3 = 3 \quad \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$

$$\mathrm{Ker}(A) = \{ x \in \mathbb{R}^n : Ax = 0 \}$$

$$A \in \mathbb{R}^{m \times n}$$

Pdf : $f(x, \lambda) = \lambda x^{\lambda - 1}$

$$L(x, \lambda) = \prod_{i=1}^{n} \lambda x_i^{\lambda - 1} = \lambda^n \sum x_i^{\lambda - 1}$$

log likelihood:

$$L(x, \lambda) = n \log \lambda + (\lambda - 1) \underbrace{\sum \log x_i}_{Q_1}$$

$$\frac{\partial L}{\partial \lambda} = \frac{n}{\lambda} + \sum \log x_i = 0$$

$$\Rightarrow \hat{\lambda} = - \frac{n}{\sum \log x_i}$$

Q_2 : log / ln $\rightarrow$ whichone.

Q_3 : $\sum \log x_i = \log x_1 + \log x_2 + \cdots + \log x_n$

$\log \sum x_i = \log (x_1 + x_2 + \sim + x_n)$

$$f_Y(y|\theta) = (\theta+1)y^{\theta}$$

$$L(y|\theta) = \prod_{i=1}^{n} (\theta+1)\, y_i^{\theta}$$

$$= (\theta+1)^n \sum y_i^{\theta}$$

log likelihood

$$L(y|\theta) = n\log(\theta+1) + \theta \ln\left(\sum y_i\right)$$

$$\frac{\partial L}{\partial \theta} = \frac{n}{\theta+1} + \ln\left(\sum y_i\right) \stackrel{!}{=} 0$$

$$\Rightarrow \quad n + \theta \ln\left(\sum y_i\right) + \ln\left(\sum y_i\right) = 0$$

$$\Rightarrow \quad \hat{\theta} = -\frac{n + \ln\left(\sum y_i\right)}{\ln\left(\sum y_i\right)}$$

$\blacksquare$

- [ ] Ensemble Kalman Filter.
- [ ]

$$\lambda^n \sum x_i^{\lambda - 1}$$

$\lambda^n \sum x_i \lambda - 1$

$n \log \lambda + (\lambda - 1) \log \sum x_i$

$\frac{n}{\lambda} + \log \sum x_i = 0$

$\Rightarrow n + \lambda \log \sum x_i = 0$

$\Rightarrow \hat{\lambda} = -\frac{n}{\log \sum x_i}$

$\log \sum (x_i)$

$= \log(x_1 + x_2 + \cdots + x_i)$

$\lambda^n \sum x_i \lambda - 1$

$n \log \lambda + (\lambda - 1) \sum \log x_i$

$\frac{n}{\lambda} + \sum \log x_i = 0$

$\Rightarrow n + \lambda \sum \log x_i = 0$

$\Rightarrow \hat{\lambda} = -\frac{n}{\sum \log x_i}$

$\sum \log x_i$

$= \log x_1 + \log x_2 + \cdots + \log x_i$

26-01-2023:

x is σ-sg

$$E[e^{\lambda x}] \leq e^{\lambda^2\sigma^2/2} \quad \forall \lambda \in \mathbb{R}$$

$$M_x(\lambda) = E[e^{\lambda x}] = E\left[\sum_{k=0}^{\infty} \frac{(\lambda x)^k}{k!}\right]$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^k E[x^k]}{k!}$$

$$\frac{\partial}{\partial \lambda}(M_x(\lambda)) = \frac{\partial}{\partial \lambda}\left[1 + \lambda E[x] + \frac{\lambda^2}{2} E[x^2] + \cdots\right]$$

$$= 0 + E[x] + \lambda E[x^2] + \cdots$$

$$\frac{\partial}{\partial \lambda} M_x(\lambda)\Big|_{\lambda=0} = E[x]$$

$$\frac{\partial^2}{\partial \lambda^2} M_x(\lambda)\Big|_{\lambda=0} = E[x^2]$$

$$\frac{\partial^k}{\partial x^k} M_x(\lambda)\Big|_{x=0} = E[x^k]$$

$$= 1 + \frac{\lambda}{1} E[x] + \frac{\lambda^2}{2} E[x^2] + \cdots$$

$$0 \leq \sum_{k=1}^{\infty} \left(\frac{\lambda^2\sigma^2}{2}\right)^k \frac{1}{k!} - \sum_{k=1}^{\infty} \frac{\lambda^k E[x^k]}{k!}$$

$$\lambda E[x] \leq \underbrace{\sum_{k=1}^{\infty} \left(\frac{\lambda^2\sigma^2}{2}\right)^k \cdot \frac{1}{k!}}_{O(\lambda^2)} - \underbrace{\sum_{k=2}^{\infty} \frac{\lambda^k E x^k}{k!}}_{O(\lambda^2)}$$

$$e^{\lambda^2\sigma^2/2} = \sum_{k=0}^{\infty} \left(\frac{\lambda^2\sigma^2}{2}\right)^k \frac{1}{k!}$$

$$= 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^4\sigma^4}{4} \cdot \frac{1}{2} + \cdots$$

$$\lambda E[x] \leq O(\lambda^2)$$

$\lambda \neq 0$ and divide by λ

$\lambda > 0, \quad E[x] \leq O(\lambda)$

$\lambda < 0, \quad E[x] \geq O(\lambda)$

$\lambda \to 0$, Sandwich theorem

$\Rightarrow \; 0 \leq E[x] \leq 0$

$\Rightarrow \; E[x] = 0$

$$0 \leq \sum_{k=1}^{\infty} \left(\frac{\lambda^2\sigma^2}{2}\right)^k \frac{1}{k!} - \sum_{k=2}^{\infty} \frac{\lambda^k E[x^k]}{k!}$$

$$\frac{\lambda^2}{2} E[x^2] \leq \sum_{k=1}^{\infty} \left(\frac{\lambda^2\sigma^2}{2}\right)^k \cdot \frac{1}{k} - \sum_{k=3}^{\infty} \frac{\lambda^k E[x^k]}{k!}$$

$$\frac{\lambda^2}{2} E[x^2] \leq \frac{\lambda^2\sigma^2}{2} \cdot \frac{1}{1} + \frac{\lambda^2\sigma^4}{4} \cdot \frac{1}{2} + \cdots - \left(\frac{\lambda^3 E[x^3]}{6} + \cdots\right)$$

divided by $\frac{\lambda^2}{2}$

$$E[x^2] \leq \sigma^2 + \frac{\lambda^2 \sigma^4}{4} + \cdots$$
$$- \left(\lambda \frac{E[x^3]}{3} + \cdots\right)$$

$\overset{\lambda \to 0}{\Longrightarrow} \quad E[x^2] \leq \sigma^2$

(2)

$$E[e^{\lambda c x}] = E[e^{\tilde{\lambda} x}]$$

X subgaussian

$$\leq e^{\tilde{\lambda}^2 \sigma^2 / 2} = e^{\frac{\lambda^2 c^2 \sigma^2}{2}} \quad c^2\sigma^2 \to \tilde{\sigma}^2$$

$$\tilde{\sigma} \to \sqrt{c^2 \sigma^2} = |c|\sigma$$

(3)

X_1 is σ_1 subgau. $\quad$ X_1, X_2 independent

X_2 is σ_2 subgau.

$$E[e^{\lambda(X_1 + X_2)}] = E[e^{\lambda X_1}] \cdot E[e^{\lambda X_2}]$$
$$\leq e^{\lambda^2 \sigma_1^2 / 2} \cdot e^{\lambda^2 \sigma_2^2 / 2}$$
$$= e^{\frac{\lambda^2 (\sigma_1^2 + \sigma_2^2)}{2}} \quad \sigma_1^2 + \sigma_2^2 \to \tilde{\sigma}^2$$
$$\Rightarrow \tilde{\sigma} = \sqrt{\sigma_1^2 + \sigma_2^2} \;\Big|\; \Rightarrow X_1 + X_2 \text{ is } \sqrt{\sigma_1^2 + \sigma_2^2} \text{ SG.}$$

$\{x_1, x_2, x_3, \ldots x_k\} \subset \mathbb{R}^n.$

$\{Y_1, \ldots, Y_k\} \subset \mathbb{R}^n$

$\langle Y_i, Y_j \rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$

$Y_i^T \cdot Y_j$

$\|Y_j\|_2^2 = \langle Y_j, Y_j \rangle$

Gram - Schmidt:

① $\tilde{Y}_1 = x_1$

$Y_1 = \dfrac{\tilde{Y}_1}{\|\tilde{y}_1\|}$

$x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 0 \\ 3 \\ 6 \end{pmatrix}$

① $\tilde{Y}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad Y_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot \dfrac{1}{1}$

② $\tilde{Y}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} - \left\langle \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} . \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

$= \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} \qquad Y_2 = \dfrac{1}{2} \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

3 $\tilde{Y}_3 = X_3 - \langle X_3 \cdot \cancel{X}_1 \rangle \cdot Y_1 - \langle X_3 \cdot Y_2 \rangle Y_2$

$$= \begin{pmatrix} 0 \\ 3 \\ 6 \end{pmatrix} - 0 \cdot Y_1 - 3 \cdot Y_2$$

$$= \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}, \quad Y_3 = \frac{1}{6}\begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$Y_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad Y_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad Y_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

02-01-2023:

$$E[\hat{\theta}] = \theta.$$

$$Y = X\beta + \epsilon. \quad y \in \mathbb{R}^n, \ \beta \in \mathbb{R}^{n+1}.$$

$$E[\epsilon] = 0, \quad var[\epsilon] = \sigma^2$$
$$cov(\epsilon_i, \epsilon_j) = 0, \ i \neq j$$

$$X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,n} \\ 1 & \vdots & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,n} \end{bmatrix} \qquad Y = \beta_0 + \beta_1 X$$

$$\hat{\beta} = \underbrace{(X^T X)^{-1} X^T}_{\text{Pseudo inverse.}} y$$

$$E[\hat{\beta}] = E[(X^T X)^{-1} X^T Y]$$
$$= E[(X^T X)^{-1} X^T (X\beta + \epsilon)]$$
$$= E[\beta + (X^T X)^{-1} X^T \epsilon]$$
$$= \beta \quad ; \ E[\epsilon] = 0$$

$X : \; -1 \quad 1$

$Y : \; -0.9 \quad 2.8$

$$Y = \begin{bmatrix} -0.9 \\ 2.8 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$X^T = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} ; \quad X^T X = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2 . I$$

$$(X^T X)^{-1} = \frac{1}{2} I$$

$$(X^T X)^{-1} X^T = \frac{1}{2} I . X^T = \frac{1}{2} X^T$$

$$\hat{\beta} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -0.9 \\ 2.8 \end{bmatrix}$$

$$= \begin{bmatrix} 0.95 \\ 1.85 \end{bmatrix}$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\frac{1}{4} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$X_1, \ldots, X_n$ Sample

$H_0 : \theta \in \Phi_0$

$H_1 : \theta \in \Phi_1$

Design an Label α test so chose from

$H_0 : \mu = \mu_0 \leftarrow$ Sample Mean

$H_1 : \mu \neq \mu_0$

given $X_1 \ldots X_n \sim N(\mu, \sigma^2)$

σ^2 known.

		Decision	
		H_0	H_1
Truth	H_0	✓	× ← Type I error
	H_1	× ↑ Type II error.	✓

$\alpha \geq P(\text{Type I error})$

$P(\text{Type I error}) = P(H_1 | H_0)$

$\Rightarrow \mathbb{P}(|Z| > c \mid H_0)$

$= 2\underbrace{\mathbb{P}(Z > c \mid H_0)}_{\Phi(c)}$

$= 1 - \Phi(c)$

$Z = \dfrac{\bar{X} - \mu}{\sigma} \sqrt{n}$

$\sim N(0, 1)$

$Z \in R = \{x \in \mathbb{R}, |x| > c\}$

$$\alpha \geq \boxed{2(1-\Phi(c))}$$

$$\Rightarrow \alpha/2 \geq 1 - \Phi(c)$$

$$\Rightarrow \Phi(c) \geq \alpha/2 - 1 = z_{\alpha/2}$$

Look up c.

$P(Z \leq c)$

$= \Phi(c)$

Type II error:

$$B(\mu) = P(H_0 | H_1)$$

$$P(H_0 | H_1) = P(|Z| \leq c \mid H_1)$$

$$= P\left(-c \leq \frac{\bar{x}-\mu_0}{\sigma}\sqrt{n} \leq c \mid H_1\right)$$

$$= P\left(\mu_0 - \frac{c\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + \frac{c\sigma}{\sqrt{n}} \mid H_1\right)$$

$$= P\left(\frac{\mu_0-\mu}{\sigma}\sqrt{n} - c \leq \frac{\bar{x}-\mu}{\sigma}\sqrt{n} \leq \frac{\mu_0-\mu}{\sigma}\sqrt{n} + c\right)$$

$N(0,1)$

Proof

- দেখা লাগবে

$$= \Phi\left(\frac{\mu_0-\mu}{\sigma}\sqrt{n} + c\right) - \Phi\left(\frac{\mu_0-\mu}{\sigma}\sqrt{n} - c\right)$$

Sample $x_1 \cdots x_n$

① $x_{(1)} --- x_{(n)}$

② $D_i = F_\theta(x_{(i)}) - F_\theta(x_{i-1})$

③ $\hat{\theta} = \underset{\theta}{\arg\max} \prod_{i=1}^{n+1} D_i^\theta$

$$D_1^\theta = F_\theta(x_{(1)}) - \underbrace{F(x_{(0)})}_{=0}$$

$$D_{n+1}^\theta = 1 - F(x_n)$$

Ex: $x_{(1)} = 2, \quad x_{(2)} = 4$

$\sim \exp(\lambda)$

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$F_\lambda(x) = 1 - e^{-\lambda x}$$

$$F(x_{(1)}) = 1 - e^{-2\lambda}$$

$$F(x_{(2)}) = 1 - e^{-4\lambda}$$

$$D_1 = 1 - e^{-2\lambda} - 0$$

$$D_2 = 1 - e^{-4\lambda} - 1 + e^{-2\lambda}$$

$$D_3 = 1 - 1 + e^{-4\lambda}$$

$$\hat{\lambda} = \underset{\lambda}{\arg\max} \left[(1 - e^{-2\lambda})(e^{-2\lambda} - e^{-4\lambda}) e^{-4\lambda} \right]$$

$$x := e^{-2\lambda}$$

$$\max (1-x)(x - x^2) x^2$$

$$= x^5 - 2x^4 + x^3$$

$$\boxed{x = 0.6}$$

$$e^{-2\lambda} = 0.6$$

$$\Rightarrow -2\lambda = \ln 0.6$$

$$\Rightarrow \lambda = -\frac{\ln 0.6}{2}.$$

Samples $X_1, \ldots, X_n$

$T(X_1, \ldots, X_n)$ is sufficient for θ if

$$P_\theta(X_1, \ldots, X_n \mid T(X_1, \ldots X_n))$$

$$= P(X_1, \ldots, X_n \mid T(X_1, \ldots X_n))$$

Factorization theorm:

T a statistic for θ is sufficient iff.

$$P(X_1, \ldots X_n \mid \theta)$$

$$= g(T(X_1 \ldots X_n) \mid \theta) \cdot h(X_1, \ldots, X_n)$$

Example:

$$X_1, \ldots X_n \overset{i.i.d}{\sim} Poi(\lambda)$$

Find a sufficient test Statistics for λ ?

$$f(x_1, \ldots, x_n \mid \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$= \frac{e^{-n\lambda} \lambda^{\sum x_i}}{x_1! \cdots x_n!}$$

$$= e^{-n\lambda} \lambda^{\sum x_i} \cdot \underbrace{\boxed{\frac{1}{x_1! \cdots x_n!}}}_{h(x)}$$

Define $\sum x_i = T(x_1, \ldots, x_n)$

$$= \underbrace{e^{-n\lambda} \lambda^{T(x_1, \ldots x_n)}}_{g} \cdot h(x_1 \ldots x_n)$$

$$= g(T(x_1, \ldots, x_n) \mid \lambda) \cdot h(x_1 \ldots x_n)$$

$\Rightarrow T(x_1 \ldots x_n) = \sum x_i$

is sufficiet for λ

One side:

$$P(\text{type I error} \mid \mu) = P(\text{Reject } H_0 \mid \mu)$$

$$= P(W > c \mid \mu)$$

$$= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > c \mid \mu\right)$$

$$= P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} + \frac{\mu - \mu_0}{\sigma/\sqrt{n}} > c \mid \mu\right)$$

$$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > c + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} \mid \mu\right)$$

$$\leq P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > c \mid \mu\right)$$

$$= 1 - \Phi(c)$$

given μ, $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$\alpha = 1 - \Phi(c)$$

$$c = z_\alpha$$

Accept H_0 if $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq z_\alpha$

$$\beta = P(\text{type II error})$$

$$= P(\text{accept } H_0 \mid H_1)$$

$$= P(W \leq c \mid H_1)$$

$$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c \mid H_1\right)$$

$= \Phi(c)$

Maximum spacing $- 2\pi$.

// VC Dimensions – 3 points

// Linear Regression – Proof.

// Ensemble k.p – unbiased.

// SVD – PCA

// Reinforcement learning.