$$\frac{1}{2b}\left[\int_{\mu} e^{-\left(\frac{x-\mu}{b}\right)} dx + \int_{-\infty} e \, dx\right]$$

$$= \frac{1}{2b}\left[b \times 1 + b \times 1\right] = 1$$

$\{x_1, x_2, \dots, x_n\}$

Assumption :- — univariate
— independent

AIM :- What is the best fitted distribution corresponding to the data?

i) Histogram   2) Q-Q plot

$$\hat{f}(x) = \frac{1}{2nh}\sum\left(F(x+h) - F(x-h)\right)$$

$$= \frac{1}{2nh}\sum_{i=1}^{n} K(\cdot) \geq 0$$

Double exponential — $f(x) = \frac{1}{2b} e^{-\left|\frac{x-\mu}{b}\right|}$ $-\infty < x < \infty$

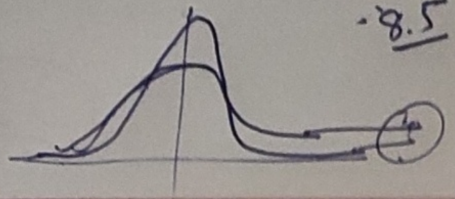
Laplace f^n —

$$F(x;\mu,b) = \begin{cases} \frac{1}{2} e^{\left(\frac{x-\mu}{b}\right)} & x < \mu \\ \frac{1}{2} + \frac{1}{2} e^{-\left(\frac{x-\mu}{b}\right)} & x > \mu \end{cases}$$

Skewness $S(x) = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right)^3$

Kurtosis $K(x) = \mathbb{E}\left(\frac{X-\mu}{\sigma}\right)^4$ (Peakness)

For Normal distribution, $K(x) = 3$



$8.5 \xrightarrow[0.57]{} 7.93 \xrightarrow[0.07]{} 7.86 \xrightarrow[0.05]{} 7.81$

$\frac{d^2 r}{dt^2} > 0$

$$\frac{f(t)}{1-F(t)}$$

For measuring heavy-tailedness, mean & var shud be same.

13-15 types of bivariate paretos are available.

19/1/24

Multinomial

Multi$(10, 0.2, 0.2, 0.2)$ is 4-D.

$$\mathbb{P}\left(\sum_{i=1}^{k-1} p_i < U_i \le \sum_{i=1}^{k} p_i\right) = \mathbb{P}\left[U_i \le \sum_{i=1}^{k} p_i\right] - \mathbb{P}\left[U_i \le \sum_{i=1}^{k-1} p_i\right]$$

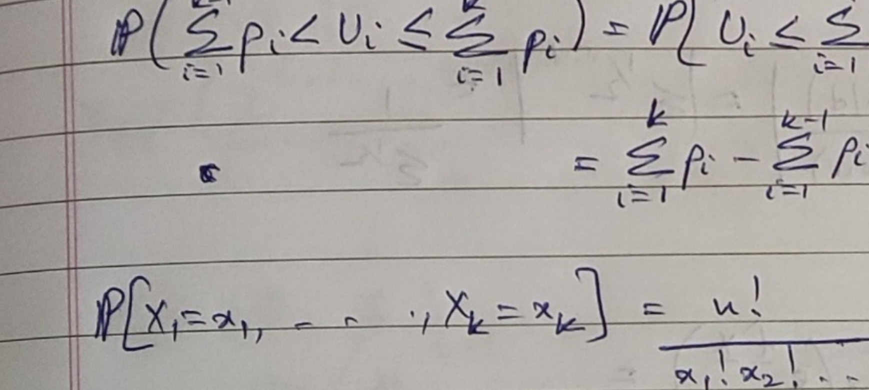
$$= \sum_{i=1}^{k} p_i - \sum_{i=1}^{k-1} p_i = p_k$$

$$\mathbb{P}[X_1 = x_1, \ldots, X_k = x_k] = \frac{n!}{x_1!\, x_2! \ldots x_k!}\, p_1^{x_1} \ldots (1-p_1-\ldots p_k)^{n-\sum x}$$

$$\mathbb{P}[X_k = x_k] = \sum \sum_{x_{k-1}} \mathbb{P}[X_1 = x_1, \ldots, X_k = x_k] \qquad \sum x_i = n$$

$\sim \text{Bin}(n, p_k)$

Box plot -



$$IQR = Q_3 - Q_1$$

Example of cov = 0 but X, Y are independent.

See multivar normal distributⁿ

$$X = \mu + \Sigma^{1/2} Z$$

$$\Sigma = P \Lambda P'$$

$$= P \Lambda^{1/2} \Lambda^{1/2} P'$$

$$= \underbrace{P \Lambda^{1/2} P'}_{\Sigma^{1/2}} \underbrace{P \Lambda^{1/2} P}_{\Sigma^{1/2}}$$

Σ – positive definite

$$|J| = \left| \frac{\partial(\text{old})}{\partial(\text{new})} \right| = \left| \Sigma^{1/2^{-1}} \right| = \frac{1}{\Sigma^{1/2}}$$

28/1/24

$$M_{\underset{\sim}{X}}(\underline{t}) = E_{\underset{\sim}{X}}\left(e^{\underline{t}' \underline{X}}\right) =$$

classmate

Date

Page

If $\underset{\sim}{X} \sim N(\underline{\mu}, \underline{\Sigma})$,

$Y^{d\times 1} = A^{d\times d} X^{d\times 1} \sim N(A\underline{\mu}, A\Sigma A')$

Pf:

$$\mathbb{E}(e^{\underline{t}'Y}) = \mathbb{E}(e^{t'A\underline{X}}) \qquad t^{*'} = t'A$$

$$= \mathbb{E}(e^{t^{*'}\underline{x}})$$

$$= \exp\left(\mu t^{*} + \frac{1}{2} t^{*'} \Sigma t^{*}\right)$$

$$= \exp\left(\mu t'A + \frac{1}{2} A't \Sigma t'A\right)$$

$\rightarrow X = \begin{pmatrix} X_1^{d_1\times 1} \\ X_2^{(d-d_1)\times 1} \end{pmatrix}$

$$X_1 = \underbrace{\left(I^{d_1\times d_1} \quad 0^{d_1\times(d-d_1)}\right)}_{A} \begin{pmatrix} X_1^{d_1\times 1} \\ X_2^{(d-d_1)\times 1} \end{pmatrix}$$

$X_2 = (0 \quad I)$

$$\mu = \begin{pmatrix} \mu_1^{d_1\times 1} \\ \mu_2^{(d-d_1)\times 1} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11}^{d_1\times d_1} & \Sigma_{12}^{d} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$\rightarrow$ Prove that $\Sigma_{12} = 0 \Rightarrow \underset{\sim}{X_1}$ & $\underset{\sim}{X_2}$ are ind.

Pf:-

$$\mathbb{E}[e^{\underline{t}'\underline{X}}] = \mathbb{E}\left[e^{(t_1'X_1 + t_2'X_2)}\right]$$

$$= \left[e^{(t_1' \; t_2')\begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix} + (t_1' \; t_2')\begin{pmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{pmatrix}t_1 \\ t_2\end{pmatrix}}\right]$$

$$= e^{(t_1'\mu_1 + t_2'\mu_2 + t_1'\Sigma_{11}t_1 + t_1\Sigma_{12}t_2 \; - \; - \; -)}$$

$$= \cancel{M_{X_1}(t_1)}\, e^{(t_2'\mu_2 + t_2\Sigma_2 t_2)} \cdot e^{(t_1'\mu_1 + \; \cdot \; \cdot \; \cdot)}$$

$$= M_{X_1}(t_1) \cdot M_{X_2}(t_2)$$

classmate
Date
Page

<u>05/1/04</u> $(X_1 | X_2 = x_2) \sim \text{Multinormal}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)\right),$

$$E\begin{pmatrix} W \\ Y \end{pmatrix} = \begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix} \begin{pmatrix} W \\ \underline{X} \end{pmatrix}$$

$$= A\underline{X} = \begin{pmatrix} I^{d_1 \times d_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I^{d-d_1 \times d-d_1} \end{pmatrix}$$

चूतिया
साला

Statistical Inference → Parameter based ⟶ distributional
Statistical Inference → Non-parametric approach

assumpt^n

Frequentist based
Bayesian Approach ⟹ MAP, Bayes Estimate

θ = fixed
θ ~ random

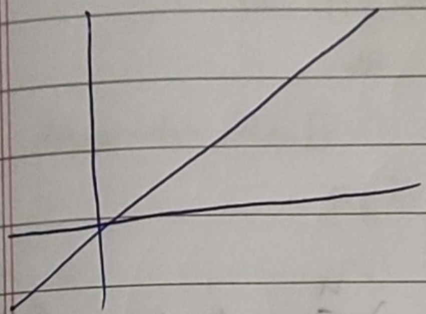
Uniformly min var unbiased Estimate
Max likelihood estimate
Method of Moments
Min max algo
Parametric

Quiz-1

heavy tailed

Heavy tailed :- (than exponential by def^n).

$\lim_{x\to\infty} e^{tx}\left(1-F(x)\right) = \infty$ where F is CDF



$$\begin{bmatrix}\bar{W}\\ \bar{X}_2\end{bmatrix} = \begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1}\\ 0 & I_p\end{bmatrix}\begin{bmatrix}X_1\\ X_2\end{bmatrix}$$

4/9/24

$$f(x;\beta,\theta) = \beta\theta^{\beta} x^{\beta-1} e^{-(\theta x)^{\beta}} \; ; x \geq 0, \theta > 0, \beta > 0$$

$$(\hat{\beta},\hat{\theta}) = \underset{\beta,\theta}{\arg\max} \ln L(\beta,\theta)$$

$$\ln L = \ln \prod_{i=1}^{n} f(x;\beta,\theta)$$

$$= \sum_{i=1}^{n} \ln(f(x_i,\beta,\theta))$$

$$= n\ln\beta + n\beta\ln\theta + (\beta-1)\sum_{i=1}^{n}\ln(x_i) - \theta^{\beta}\sum_{i=1}^{n} x_i^{\beta}$$

$$\frac{n\beta}{\theta} - \beta\theta^{\beta-1}\sum_{i=1}^{n} x_i^{\beta} = 0 \Rightarrow \theta = \left[\frac{n}{\sum_{i=1}^{n} x_i^{\beta}}\right]^{1/\beta} = f_1(\underline{x},\beta)$$

$\ln L(f; \;)$ = profile log-likelihood f^n

$$\frac{n}{\beta} + n\ln\theta + \sum\ln x_i - \sum(\theta x_i)^{\beta}\ln(\theta x_i) = 0$$

$$\rightarrow f(x;\theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x_i \leq \theta \\ 0 & \text{o.w} \end{cases}$$

Assume $x_{(1)} < \dots < x_n$

$$L(\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \begin{cases} \underline{\underline{\frac{1}{\theta^n} I(0,x_{(1)}) I(x_n,\theta)}} & \\ 0 & \text{o.w} \end{cases}$$

$$\therefore \hat{\theta} = X_{(n)}$$

$\rightarrow f(x;\theta) = \begin{cases} 1 & \theta \le x \le \theta+1 \\ 0 & \text{o.w} \end{cases}$

$$\prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} 1 & \theta \le x_{(1)} < \dots x_{(n)} \le \theta+1 \\ 0 & \text{o.w} \end{cases}$$

$$= I(\theta; x_{(1)})\, I(x_{(n)}; \theta+1)$$

~~Max~~ MLE $\qquad \theta < x_1 < \dots x_n < \theta+1$

$x_1 \ge \theta$ & $x_n \le (\theta+1)$

$\therefore \theta \in [x_n - 1, x_1]$

MLE is not unique

In a data, some part follows f_1-distribution while other follows f_2-distributⁿ.

$\{X_1, X_2, \dots, X_n\} \sim p f_1(x; \mu_1, \sigma_1^2) + (1-p) f_2(x; \mu_2, \sigma_2^2)$.

$$\ln L(p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = \sum_{i=1}^{n} \ln\left(f(x_i; p, \mu, \sigma_1^2, \mu_2, \sigma_2^2)\right)$$

EM $\geq$ Gradient descent $\rightarrow$ faster convergence & also for f^n non cont f^ns.

HMM uses EM algo.

EM algo:-

1) "Missing obs within the sys".

$\underset{z_i|X,\theta}{E} f(X_i, z_i) =$ ~~complete~~ pseudo likelihood f^n. $= f(x; \theta, \theta^t)$

$z_i \equiv$ unknown / hidden

Complete likelihood f^n -?.

$p(z_i|x,\theta^t) \equiv$ posterior of $\underline{z}$ given $\underline{x}, \theta^t$ - known

Expectat^n step:-

$$Q(\theta,\theta^t) = \mathbb{E}_{z|x,\theta^t}\left(\ln f(x,z;\theta)\right)$$

Maximizat^n step:-

$$\hat{\theta} = \arg\max_{\theta} Q(\theta,\theta^t)$$

6/2/24

$L(\theta) = \ln f(x,\theta) =$ likelihood f^n based on incomplete data

$$\ln L(\theta) - \ln L(\theta^t) = \ln \frac{f(x|\theta)}{f(x|\theta^t)}$$

z = unknown r.v

$$q(\cdot|x,\theta^t) = \frac{f(z,x|\theta^t)}{f(x|\theta^t)}$$

$$= \ln\left(\frac{\int f(x;z|\theta)\,dz}{f(x;\theta^t)}\right)$$

$$= \ln\left(\int \frac{f(x,z|\theta)}{f(x|\theta^t)}\,dz\right)$$

$$= \ln\left(\int \frac{f(x,z|\theta)\,q(z|x;\theta^t)}{\underbrace{f(x|\theta^t)\,q(z|x;\theta^t)}_{f(x,z|\theta^t)}}\,dz\right)$$

(Jensen inequality) $\geq \underbrace{\int \left[\ln \frac{f(x,z|\theta)}{f(x,z|\theta^t)}\right] q(z|x,\theta^t)\,dz}_{=\Delta L(\theta,\theta^t)}$

$q(E(X)) \geq \mathbb{E}(q(X))$

$$\ln L(\theta) - \ln(\theta^t) \geq \Delta L(\theta,\theta^t)$$

$$\cancel{\Delta L(\theta,\theta^t)} = \int \left[\ln \frac{f(x,z|\theta)}{f(x,z|\theta^t)} \right] \cdot f(z|x,\theta^t) \cdot dz$$

$$= \int \ln$$

$$\underset{\theta}{\arg\max}\ \Delta L(\theta,\theta^t) = \underset{\theta}{\arg\max} \int \ln f(x,z|\theta)\, q(z|x,\theta^t) \,.\, dz$$

$$= \underset{\theta}{\arg\max}\ \underset{q(z|x,\theta^t)}{E} \big(f(x,z|\theta) \big)$$

$$L(\theta) - L(\theta^t) = \log \left[\frac{\int p(x,z|\theta) \,.\, dz \ \cancel{p(z|x,\theta)}}{p(x|\theta^t)} \right]$$

$$= \log \left[\int \frac{p(z|x,\theta)\, p(x,z|\theta)}{p(z|\theta^t,x)\, p(x|\theta^t)} dz \right]$$

$$\geq \int \log \left(\frac{p(x,z|\theta)}{p(z|\theta^t,x)\, p(x|\theta^t)} \right) \cdot p(z|x,\theta) \,.\, dz$$

EM algo in Estimating the parameter of two normal distributⁿ mixture.

$$f(x; p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2) = p f_1(x; \mu_1, \sigma_1^2) + (1-p) f_2(x; \mu_2, \sigma_2^2)$$

$$z_i = \begin{cases} 1 & X_i \sim f_1 \\ 0 & X_i \sim f_2 \end{cases}$$

(X_i, z_i) ; $i = 1(1)n$ → complete obs

$$f(x_i, z_i) = (f_1(x_i))^{z_i} (f_2(x_i))^{1-z_i}$$

$$P[x_i \le X_i \le x_i + \Delta x_i, Z = z_i] = P[x_i \le X \le x_i + \Delta x_i | Z_i = z_i] P[Z_i = z_i]$$

$$= (p f_1(x_i))^{z_i} ((1-p) f_2(x_i))^{1-z_i}$$

$$\text{Complete likelihood} = \prod_{i=1}^{n} f(x_i, z_i)$$

$$= \prod_{i=1}^{n} (p f_1(x_i))^{z_i} ((1-p) f_2(x_i))^{1-z_i}$$

$$\ln L = \sum_{i=1}^{n} z_i [\ln p + \ln f_1(x_i, \mu_1, \sigma_1^2)] + (1-z_i) [\ln(1-p) + \ln f_2(x_i, \mu_2, \sigma_2^2)]$$

pseudolikelihood $= E_{(\underline{Z}|\underline{X},\theta^t)}(\ln L(\theta))$

$$= \sum_{i=1}^{n} \left[(\ln p) E_{Z|X,\theta^t}(Z_i) + \sum_{i=1}^{n} E_{Z|X,\theta^t}\left(Z_i\left[-\ln \sigma_1 + \frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]\right.\right.$$

$$+ \sum_{i=1}^{n}(1-p) E_{Z|X,\theta^t}(1-Z_i) + \sum_{i=1}^{n} E_{Z|X,\theta^t}\left((1-Z_i)\ln f_2(x_i;\mu_2,\sigma_2^2)\right)$$

$$m_i^t = E_{Z_i|X_i,\theta^t}(Z_i) = P\left[Z_i=1 \mid X_i=x,\theta^t\right] = \frac{p^t f_1(x;\mu_1^t,\sigma_1^{2t})}{p^t f_1(x;\mu_1^t,\sigma_1^2) + (1-p^t) f_2(x;\mu_2^t,\sigma_2^2)}$$

Take derivative w.r.t $p, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2$

$$\hat{p}^{t+1} = \frac{1}{n}\sum_{i=1}^{n} m_i^t \qquad \hat{\mu}_1^{t+1} = \frac{1}{\sum_{i=1}^{n} m_i^t}\sum_{i=1}^{n} m_i^t x_i \qquad \hat{\sigma}_1^{2\,t+1} = \frac{1}{\sum_{i=1}^{n} m_i}\sum_{i=1}^{n} m_i^t (x_i-\hat{\mu}_1)^2$$

$$\hat{\mu}_2^{t+1} = \frac{1}{n}\sum_{i=1}^{n}(1-m_i^t)x_i \qquad \hat{\sigma}_2^{2\,t+1} = \frac{1}{\sum_{i=1}^{n}(1-m_i^t)}\sum_{i=1}^{n}(1-m_i^t)(x_i-\hat{\mu}_2^t)^2$$

$\frac{1}{(1-2t)^{k/2}}$

$f(x) - x^{k/2-1} e^{-x/2}$

classmate
Date
Page

13/2/24 **Vector Calculus**

$f: \mathbb{R}^n \to \mathbb{R}$

Def^n of derivative w.r.t $x^{n\times 1}$

$$\frac{\partial f}{\partial x^{n\times 1}} = \left[\frac{\partial f}{\partial x_1} \; \frac{\partial f}{\partial x_2} \; \cdots \; \frac{\partial f}{\partial x_n}\right]_{1\times n}$$

$$f(\underline{x}) = A^{m\times n} x^{n\times 1}$$

$$D_f = \lim_{\|h\|\to 0} \frac{f(x+he_i) - f(x) - L_x}{\|h\|}$$

$$\frac{\partial f}{\partial x} = A$$

$$f(\underline{x}) = x^{T\,1\times n} A^{n\times n} x^{n\times 1}$$

$$\frac{\partial f(\underline{x})}{\partial x^{n\times 1}} = \left[\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right] \qquad \frac{\partial f}{\partial x_i} = (x^T A + x^T A^T)$$

$$= x^T A + x^T A^T \qquad A \text{ is symmetric}$$

$$= 2x^T A$$

Matrix Calculus

$f: \mathbb{R}^{m\times n} \to \mathbb{R}$

Derivative of f will be a $n\times m$ matrix.

$$\left(\left(\frac{\partial f}{\partial a_{ij}}\right)\right)_{\substack{j=1(1)n \\ i=1(1)m}}$$

$a_{11} b_{11} + a_{12} b_{21}$ $a_{21} b_{12} + a_{12} b_{22}$

Date
Page

$$\frac{d\left(tr(AB)\right)}{dA_n} = ?\ (b_{ij})_{\substack{i=1(1)m \\ j=1(1)n}} = B$$

$$\frac{\partial}{\partial a_{ji}} \sum_{k=1} \sum_{i=1}^{m} a_{ki} b_{ik} = b_{ij}$$

always cool

from the scratch

If A is symmetric,

$$\frac{d \ln|A|}{dA} = A^{-1}$$

$$|A| = |A^T| = \sum (-1)^{i+j} a_{ji} M_{ji}$$

scratch

$$\frac{d|A^T|}{da_{ji}} = (-1)^{i+j} M_{ji} = \breve{A}\ (adj(A))$$

$$\frac{d(\ln|A|)}{dA} = \frac{1}{|A|} \cdot \frac{d}{dA}|A^T| = \frac{1}{|A|} \breve{A} = A^{-1}$$

can the derivative will make sense?

It makes sense.

I am so f**king bored

Kalyan is going to hit me.

15/2/24

$$\ln L(\mu, \Sigma) = \ln\left[\frac{1}{(2\pi)^{Nd/2}|\Sigma|^{N/2}} e^{-\frac{1}{2}\sum_{i=1}^{N}(\underline{x_i}-\underline{\mu})^T \Sigma^{-1}(\underline{x_i}-\underline{\mu})}\right]$$

$$= \frac{Nd}{2}\ln(2\pi) - \underbrace{\frac{N}{2}\ln|\Sigma|}_{= \frac{N}{2}\ln|\Sigma^{-1}|} - \frac{1}{2}\sum_{i=1}^{N}(\underline{x_i}-\underline{\mu})^T \Sigma^{-1}(\underline{x_i}-\underline{\mu})$$

$$\frac{\partial}{\partial \mu}\ln L(\mu, \Sigma) = 0 \rightarrow$$

$$\frac{-1}{2} \times 2 \sum_{i=1}^{N}(\underline{x_i}-\underline{\mu})^T \Sigma^{-1} = 0$$

$$\boxed{\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \underline{x_i}}$$

Instead of $\frac{\partial}{\partial \Sigma}$, we take $\frac{\partial}{\partial \Sigma^{-1}}$ as $f(x) = x^{-1}$ is one-one f^n

$$\frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}(\underline{x_i}-\hat{\underline{\mu}})(\underline{x_i}-\hat{\underline{\mu}})^T = 0$$

Gaussian Mixture Model

$$\underset{\sim}{X} \sim \sum_{k=1}^{K} \pi_k f(\underline{x}, \mu_k, \Sigma_k)$$

$$\pi_k = \frac{1}{N}\sum_{i=1}^{N} m_{ik}^{t}$$

$$\underset{\sim}{\mu_k} = \frac{1}{\sum_{i=1}^{N} m_{ik}^{t}} \sum_{i=1}^{N} m_{ik}^{t} x_i \longrightarrow (1)$$

$$\hat{\Sigma}_k = \frac{1}{\sum_{i=1}^{N} m_{ik}} \sum_{i=1}^{N} m_{ik}^{t} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T$$

$$z_{ik} = \begin{cases} 1 & \underset{\sim}{x_i} \sim f_k \\ 0 & x_i \not\sim f_k \end{cases}$$

K-Means Algorithm

$$r_{ik} = \begin{cases} 1 & x_i \text{ obs falls into } k^{th} \text{ cluster} \\ 0 & \text{O.W} \end{cases}$$

$$J = \frac{1}{N} \sum_i \sum_k r_{ik} \|x_i - \mu_k\|^2$$

Step 1:- fix r_{ik}, compute best μ_k by minimizing J.

Step 2:- fix μ_k, compute r_{ik} " "

$$\frac{\partial J}{\partial \mu_k} = 2 \sum_{i=1}^{n} r_{ik} (\underline{x_i} - \underline{\mu_k})^T = 0 \Rightarrow \boxed{\hat{\mu}_k = \frac{\sum_{i=1}^{n} r_{ik} \underline{x_i}}{\sum_{i=1}^{n} r_{ik}}}$$

Similiar to (1)

∴ K-means & EM algo are connected

$$\sum_{k=1}^{K} m_{ik} = 1 \Rightarrow \sum_{k=1}^{K} \underbrace{P\left[z_k = k / x_i, \theta^t\right]}_{m_{ik}} \quad k = 1(1)K$$

K Means	GMM
1) Minimizing MSE	1) Minimizing log-likelihood
2) ~~Soft~~ Hard thresholding	2) ~~Hard~~ Soft thresholding
3) Useful for Spherical clustering	3) Non spherical clustering

7/2/24

MLE

$\hat{\theta}$ is an MLE of θ

$g(\ )$ is onto

$g(\hat{\theta})$ is an MLE of $g(\theta)$

~~Statistical~~

$X_1, X_2, \ldots X_n$ are iid bernoulli(p) ; $0 < p < 1$

What is the MLE of $p(1-p) = g(p)$

$$L(p/x) = p^{\sum x_i} (1-p)^{n - \sum x_i}$$

$$\ln(L) = \sum x_i \ln p + (n - \sum x_i) \ln(1-p)$$

$$\frac{\partial L}{\partial p} = 0 \Rightarrow \frac{\sum x_i}{p} = \frac{n - \sum x_i}{1-p} = \frac{n}{1}$$

$$p = \cancel{\frac{n}{\sum x_i}} \quad \frac{\sum x_i}{n}$$

Informat^n f^n

$I(\theta) = E_x\left[\frac{-\partial^2}{\partial\theta^2}\ln f(x;\theta)\right] = E\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2$ but how is not mentioned

Score f^n

$\underbrace{\frac{\partial}{\partial\theta}\ln f(x;\theta)}$

= Expected value of Hessian f^n of $\ln f(x,\theta)$

$= Var\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)$

$E\left(\frac{\partial}{\partial\theta}\ln f(x;\theta)\right)$

$= \int_x \frac{1}{\cancel{f(x;\theta)}}\left[\frac{\partial}{\partial\theta}f(x;\theta)\right]\cancel{f(x;\theta)}.dx$ (Under some regularity condi^{n}s)

$= \int_x \frac{\partial}{\partial\theta}\int_x f(x;\theta).dx$ $=1$

$= 0$

① $f(x;\theta)$ is an bdd f^n in x & domain of x does not depend on θ

② $f(x;\theta)$ has infinite supp, cont diff & integral uniformly converges for all θ

~~For~~ ~~wait~~ ① & ② $\Leftrightarrow$ $|f(x;\theta)| \leq g(x)$

Uniform distribut^n doesn't satisfy ①

$$\int -\left[\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta)\right] f(x,\theta).dx$$

$$= \int \frac{\partial}{\partial\theta}\left[\frac{1}{f(x;\theta)}\frac{\partial f(x;\theta)}{\partial\theta}\right].dx$$

$$= \int_X \left(\frac{1}{f(x;\theta)} \frac{\partial^2 f(x;\theta)}{\partial \theta^2} - \frac{1}{f^2} \left(\frac{\partial f(x;\theta)}{\partial \theta} \right)^2 \right) . dx$$

$$= \mathbb{E} \left(\frac{\partial}{\partial \theta} \ln f(x;\theta) \right)^2$$

$X_1, X_2, \cancel{X_3}, \dots X_n \overset{iid}{\sim} f(x;\theta)$

$I_n(\theta) = n I(\theta)$

$$I_n(\theta) = E\left[\frac{-\partial^2}{\partial \theta^2} \sum_{i=1}^{n} \ln(f(x;\theta)) \right]$$

$$= \sum_{i=1}^{n} \mathbb{E}\left(\frac{-\partial^2}{\partial \theta^2} \ln(f(x,\theta)) \right)$$

$$= n I(\theta)$$

$Var(\hat{\theta}_{MLE}) \simeq \frac{1}{n I(\theta)}$ for large n

$T(x)$ is unbiased Estimate of $g(\theta)$

$$\boxed{V(T(x)) \geq \frac{(g'(\theta))^2}{n I(\theta)}}$$ — Crammer Rao Lower Bound

19/2/24 Crammer - Rao Lower bound

$T(x)$ is u.e of $g(\theta)$

$$V(T(x)) \geq \frac{\{g'(\theta)\}^2}{nI(\theta)}$$

supp(X) is finite & domain X does not depend on θ.

or

supp(X) is infinite & $f(\ ;\theta)$ is cont diff & the integral is uniformly convgt for all θ

$E(T(x)) = g(\theta) \quad \forall \ \theta \in \Theta$

$$\int T(x) f(x,\theta) . dx = g(\theta)$$

$$\int T(x) \left[\frac{\frac{\partial}{\partial\theta} f(x,\theta)}{f(x,\theta)}\right] . f(x,\theta) . dx$$

$$= g'(\theta)$$

$$\int (T(x) - g(\theta)) \left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right) f(x,\theta) . dx = g'(\theta)$$

$$\left(\because \int g(\theta) \left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right) . f(x,\theta) . dx = 0 \right.$$ proved in last class)

$$\{g'(\theta)\}^2 = E^2\left[(T(x) - g(\theta)) \left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)\right]$$

$$\leq E\left(T(x) - g(\theta)\right)^2 E\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2$$

$$= V(T(x)) \ I_X(\theta)$$

$$\boxed{V(T(x)) \geq \frac{\{g'(\theta)\}^2}{I_X(\theta)}}$$

$$\sqrt{n}\left(\hat{\theta}_{MLE} - \theta\right) \xrightarrow{d} N\left(0, \frac{1}{I_X(\theta_0)}\right) \quad n \to \infty$$

$$\Rightarrow \cancel{\sqrt{n}}\,\hat{\theta}_{MLE} \to N\left(\theta_0, \frac{1}{\cancel{n} I_X(\theta_0)}\right)$$

$$\hat{\theta}_{MLE} \xrightarrow{d} N\left(\theta_0, \frac{1}{n I_X(\theta_0)}\right) \quad \text{as } n \to \infty$$

Estimation through Bootstrap

Aim :- To reduce variance

1st bootstrap $\{x_{11}, \ldots, x_{1n}\} \to \hat{\theta}_1^{\,b}$ — Bagging

$\to \hat{\theta}_2^{\,b}$

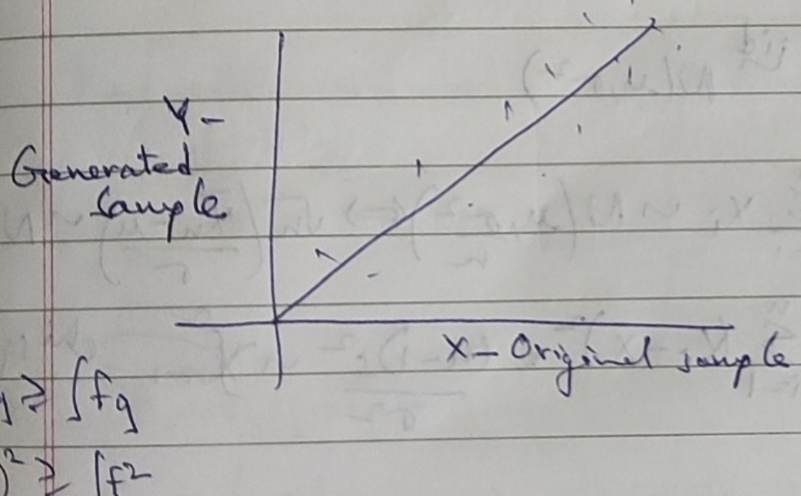
$$\theta_B = \frac{1}{B}\sum_{i=1}^{B} \theta_i^{\,b}$$

avg is best to reduce var

B^{th} bootstrap $\{x_{B1}, \ldots, x_{Bn}\} \to \hat{\theta}_3^{\,b}$

K-fold Cross-validation — Imp for exam

Delete-1 Jack Knife



$\int f \int g \geq \int fg$

$(\int f)^2 \geq \int f^2$

Date
Page

Interval Estimate

19/ 1) Width of the interval
2) Confidence coeff = chance of including the true parameter θ should be large
$= 1-\alpha$

$$\mathbb{P}(\mu \in I) = 100(1-\alpha)\%$$

"Construxn Method"

$\mathcal{D}: \{X_1, X_2, \ldots, X_n\} \sim f(x;\theta)$

$\{X_{11}, X_{12}, \ldots, X_{1n}\} \longrightarrow [\hat{a}_1, \hat{b}_1]$

$\vdots$

$\{X_{21}, X_{22}, \ldots, X_{2n}\} \longrightarrow [\hat{a}_2, \hat{b}_2]$

$$\frac{\#\, i \ni \mu \in [\hat{a}_i, \hat{b}_i]}{\#\text{total } i} = \text{Confidence Interval/ Coverage prob}$$

20/2/24

$X_1, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \sim N(0,1)$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \text{ chi square}$$

classmate

Date

Page

$$E\left(\frac{(n-1)s^2}{\sigma^2}\right) = (n-1)$$

$\Rightarrow E(s^2) = \sigma^2$

$\Rightarrow s^2$ is an unbiased estimate (u.e) of σ^2

$f(x) =$

$$Y_1 = \frac{1}{\sqrt{n}}(X_1 - \mu) + \dots + \frac{1}{\sqrt{n}}(X_n - \mu)$$

$\boxed{Y = PX}$

$$Y_2 = \frac{1}{\sqrt{2}}(X_1 - \mu) - \frac{1}{\sqrt{2}}(X_2 - \mu) + \dots$$

$PP' = I$

$$Y_i = \frac{1}{\sqrt{i(i-1)}}(X_1 - \mu) + \dots + \frac{(X_{i-1} - \mu)}{\sqrt{i(i-1)}} - \frac{(i-1)(X_i - \mu)}{\sqrt{i(i-1)}}$$

$\vdots$

$$Y_n = \frac{1}{\sqrt{n(n-1)}}(X_1 - \mu) + \dots + \frac{X_{n-1} - \mu}{\sqrt{n(n-1)}} - \frac{(n-1)(X_n - \mu)}{\sqrt{n(n-1)}}$$

$$P = \begin{bmatrix} \frac{1}{\sqrt{n}} & \dots \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \dots & \\ & \dots \end{bmatrix}$$

$$-\frac{(X_i - \mu)(X_j - \mu)}{i(i-1)} - \frac{(X_i - \mu)(X_j - \mu)}{\sqrt{i(i-1)}}$$

$$\cancel{-\frac{(X_i - \mu)(X_j - \mu)}{\sqrt{ij(i-1)(j-1)}}}$$

$$\sum_{i=k}^{n} \frac{1}{k(k-1)}(X_i - \mu)(X_j - \mu) - \frac{1}{i(i-1)}(X_i - \mu)(X_j - \mu) + \frac{1}{n}(X_i - \mu)(X_j - \mu)$$

$$\cancel{+\frac{1}{i}} + \frac{1}{i} + \dots$$

$$\frac{1}{i} \cdot \frac{1}{j} + \frac{1}{i} - \frac{1}{i-1} - \frac{1}{j(j-1)}(X_i - \mu)(X_j - \mu) + \sum_{i=2}^{n} \frac{1}{i(i-1)}$$

$$+ \frac{1}{j} - \frac{1}{j-1} \quad (X_i - \mu)(X_j - \mu)$$

$1 - \frac{1}{n}$

Date
Page

$$f(\underline{x}) = \prod_{i=1}^{n} f(x_i, \mu, \sigma_x^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{\frac{-1}{2\sigma^2}(\underline{X}-\mu)^T(\underline{X}-\mu)}$$

19.

$$f(\underline{y}) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2}\sum_{i=1}^{n} \frac{y_i^2}{\sigma^2}}$$

$J = \mathbb{J} = \frac{\partial_{old}}{\partial_{new}}$

$Y = PX$

$P'Y = X$

$(X-\mu)^T P' P (X-\mu)$

$= (P(X-\mu))'(P(X-\mu))$

$\frac{Y_i}{\sigma} \overset{iid}{\sim} N(0,1)$, $i = 1(1)n$

$$\frac{Y_1}{\sigma} = \sqrt{n}\left[\frac{\Sigma(X_i-\mu)}{n\sigma}\right]$$

$$= \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \sim N(0,1)$$

$Y = X_i^2$

$X_i = \sqrt{y}$

$-\sqrt{y}$

$$\mathbb{E}(e^{tY}) = \frac{1}{(1-2t)^{1/2}}$$

$T \sim \chi_n^2$

$$\mathbb{E}(e^{tT}) = \frac{1}{(1-2t)^{n/2}}$$

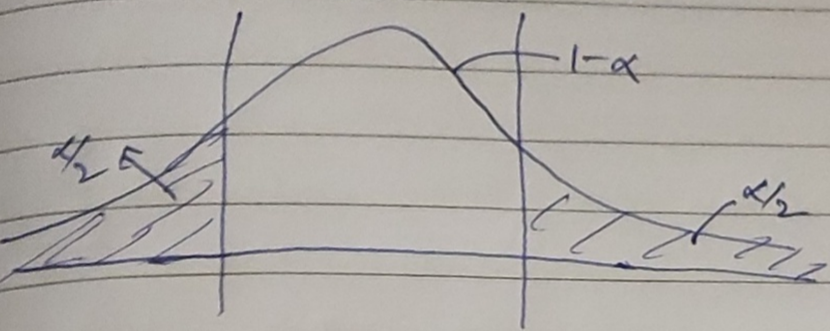
$$\frac{\sum_{i=2}^{n} y_i^2}{\sigma^2} = \frac{\sum_{i=1}^{n} y_i^2}{\sigma^2} - \frac{y_1^2}{\sigma^2}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

Construct^n of Confidence Interval

$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$

$T = \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right)$



$$P\left[-z_{\alpha/2} \le \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \le z_{\alpha/2}\right] = 1-\alpha$$