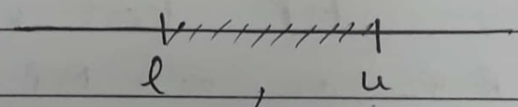
31-10-17

Recap:-

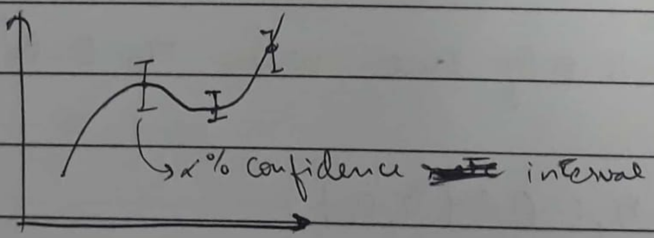
- Point estimators
  - Unbiased ~~e~~estimators
  - UMVUE
  - Maximum likelihood estimator
- Hypothesis Testing
  - Neyman - Pearson lemma.
  - ~~Interval estimation~~

⇒ Interval estimation

$X_1, \ldots X_n \sim f_\theta$, $\theta \in \Theta \subseteq \mathbb{R}$



θ can lie in this interval with probability $> \alpha$ (say 0.99)



~~find~~ Find $L: \mathcal{X} \to \mathbb{R}$

$U: \mathcal{X} \to \mathbb{R}$

$S(\vec{x}) = [L(\vec{x}), U(\vec{x})] \subseteq \Theta$

A random interval

<u>Problem</u>:- Find $[L(\bar{x}), U(\bar{x})] = S(\bar{x})$ s.t

$$P_\theta(\theta \in S(\bar{x})) \geq \underbrace{1-\alpha}_{\text{confidence level}} \quad \forall \theta \in \Theta$$

$L(\bar{x})$: lower confidence bound for θ at confidence level $1-\alpha$

$U(\bar{x})$: Upper confidence ~~bou~~ bound "

OR

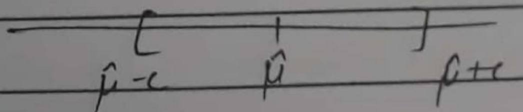
A family of random sets $\{S(\bar{x}) : \bar{x} \in \mathcal{X}\}$ of $\Theta \subseteq \mathbb{R}$ is said to constitute a family of confidence sets at conf. level $(1-\alpha)$ if

$$P_\theta(\theta \in S(\bar{x})) \geq 1-\alpha, \ \forall \theta \in \Theta$$

<u>Example</u>:- $X_1 \ldots X_n \sim G(\mu, 1)$ $\quad \mu \in \mathbb{R}$

α is specified.

UMVUE, $\quad T(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$



$$P_\mu(\hat{\mu}-c \leq \mu \leq \hat{\mu}+c) = P_\mu(\hat{\mu} \leq \mu+c, \hat{\mu} \geq \mu-c)$$

$$= P_\mu(\mu-c \leq \hat{\mu} \leq \mu+c)$$

$$= P_\mu\Big(\mu-c \leq \underbrace{\frac{1}{N}\sum X_i}_{\to G(0, 1/n)} \leq \mu+c\Big)$$

2-11-17

PAGE No.

DATE / /

Recap: Confidence interval

$X_1, X_2 \ldots X_n \sim f_\theta(\cdot)$ iid $\theta \in \mathbb{R}$.

Find $L(\bar{x})$ and $U(\bar{x})$, equivalently, $S(\bar{x}) = [L(\bar{x}), U(\bar{x})]$ s.t

$$P_\theta(\theta \in S(\bar{x})) \geq 1 - \alpha \quad \forall \theta$$

and given value of α

Quiz - 5

Q-4 $X_1 \ldots X_n \sim$

~~the~~ unbiased estimator for p

$$E_p[1 - \mathbb{1}_{\{X_1 = 1\}}] = 1 - P(X_1 = 1)$$

$$= 1 - (1-p)$$

$$= p$$

PAGE No.
DATE / /

Example:- $X_1 \dots X_n \sim C(\mu, 1)$

$$U(\bar{x}) = \frac{1}{n} \sum x_i + c$$

$$L(\bar{x}) = \frac{1}{n} \sum x_i - c \qquad c > 0$$

Result:- By choosing appropriate value of c, you can achieve any level of confidence α.
$[L(\bar{x}), U(\bar{x})] \to$ class of intervals, depends on the observation.

Uniformly Most Accurate (UMA)

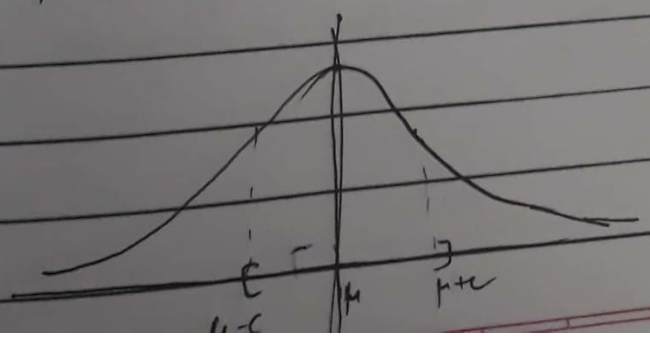
The collection of $S(\bar{x})$ is said to be UMA if

$$P_\theta(\theta' \in S(\bar{x})) \leq P_\theta(\theta' \in S'(\bar{x}))$$

$\forall \; \theta' \neq \theta$ & interval collection $S'(\bar{x})$

$$P\left(\mu' \in \left[\frac{1}{n}\sum x_i - c_1, \frac{1}{n}\sum x_i + c_2\right]\right) \geq P\left(\mu \in \left[\frac{1}{n}\sum x_i - c, \frac{1}{n}\sum x_i + c\right]\right)$$

where c_1, c_2, c are chose to meet α



PAGE No. DATE / /

Pivot Method :-

$$T : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$$

Random variable is called pivot if the distribution of T is independent of θ.

→ T takes θ as input but distribution of T ~~T can depen take θ as input~~. is independent of θ.

⇒ Sufficient statistic cannot take θ as input but its distribution can depend on θ.

Example :- $X_1 \ldots X_n \sim G(\mu, 1)$

$$T(\bar{x}, \mu) = \frac{1}{n} \sum x_i - \mu \sim G(0, 1/n)$$

Find c_1, c_2 s.t

$$P_\theta (c_1 \le T(\bar{x}, \theta) \le c_2) \geq 1 - \alpha$$

$$= P_\mu \left(c_1 \le \frac{1}{n} \sum x_i - \mu \le c_2 \right)$$

$$= P_\mu \left(\mu + c_1 \le \frac{1}{n} \sum x_i \le \mu + c_2 \right)$$

$\to [L(\bar{x}), U(\bar{x})]$ is called α-confidence interval

<u>Method of Pivots</u>

$T(\bar{x}, \theta)$ is called Pivot if the distribution of $T(\bar{x}, \theta)$ does not depend on θ.

<u>Ex</u>:- $X_1 \ldots X_n \sim G(\mu, 1)$

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \mu = T(\bar{x}, \mu)$$

<u>Note</u>:- $\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right) \sim G\left(0, \frac{1}{n}\right)$

$$P(-c \leq \bar{X} - \mu \leq c) = P(\bar{X} - c \leq \mu \leq \bar{x} + c)$$

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

— x —

variance of ΣX_i is dependent on θ so this pdf changes with θ.

$X + c \rightarrow$ distribution shifts, mean changes, var constant

$aX \rightarrow$ variance changes, mean remains same,

Try sufficient stat:-

$$Y = \max\{X_1 \ldots X_n\}$$

$$T(\bar{x}, \theta) = \max\{ x_1/\theta, \ldots x_n/\theta \}$$

$$T(\bar{x}, \theta) = \frac{Y}{\theta}$$

$$P(\max(X_1 \ldots X_n\} \leq x \} = \begin{cases} 0 & \text{if } x < 0 \\ x^n & \text{if } x \in (0, 1] \\ 1 & \text{o.w.} \end{cases}$$

$$f_Y(x) = \begin{cases} 0 & \text{o.w} \\ n x^{n-1} & \text{if } x \in (0, 1) \\ 0 & \text{o.w.} \end{cases}$$

$$\boxed{f_Y(x) = \begin{cases} n x^{n-1} & \text{if } x \in [0, 1) \\ 0 & \text{o.w} \end{cases}}$$

ϵ^2

$X_1 \ldots X_n \sim \text{Bernoulli}(p)$, $p \in [0,1]$

$P \left(\left| \underbrace{\frac{1}{n} \sum x_i}_{\text{Take this as } Y} - p \right| > t \right)$

$\therefore \quad 1 - P\left(\left| \frac{1}{n} \sum x_i - p \right| > c \right) \geq 1 - \alpha$

$P\left(\left| \frac{1}{n} \sum x_i - p \right| > c \right) \leq \alpha$

$\therefore$ It is enough to ensure $\frac{1}{4nc^2} \leq \alpha$

$\Rightarrow \boxed{c \geq \frac{1}{\sqrt{4n\alpha}}}$

Using central limit Theorem

$$\frac{\sum_{i=1}^{n} x_i - np}{\sqrt{np(1-p)}} \sim G(0,1)$$

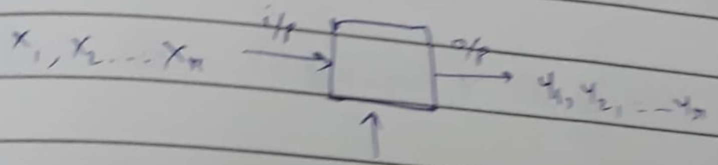
Scanned by CamScanner

7-11-17

Recap:-

- Representation of data
- Parameter estimation (Point estimate)
  - UMVUE
  - MLE
- Hypothesis testing
  - Neyman Pearson theory
- Confidence interval estimation

Regression:-



What's in here?

Linear regression

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \underset{\substack{\uparrow \\ \text{random error}}}{\epsilon} \sim \text{zero mean } 1, \sigma$$

Has "memory", depends on previous observations

→ multiple regression

Simple regression:- $\boxed{Y_i = \alpha + \beta X_i + \epsilon_i}$

~~→ Does not depend on~~ → Does not depend on previous inputs.

Estimate / find α and β such that

$$\mathcal{E}(A,B) = \sum_{i=1}^{n} \left(y_i - (A + B x_i) \right)^2$$

estimate for α — estimate for β

Pick A, B s.t. $\mathcal{E}(A,B)$ is minimised

$$\frac{\partial \mathcal{E}}{\partial A} = -\sum_{i=1}^{n} 2\left(y_i - (A + B x_i) \right) = 0$$

$$\frac{\partial \mathcal{E}}{\partial B} = -\sum_{i=1}^{n} 2 x_i \left(y_i - (A + B x_i) \right) = 0$$

$$A = \bar{y} - B\bar{x}, \quad \bar{y} = \frac{\sum y_i}{n}, \quad \bar{x} = \frac{\sum x_i}{n}$$

$$B = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

E_i's are assumed to be iid Gaussian$(0, \sigma^2)$

$y_i \sim G(\alpha + \beta x_i, \sigma^2)$ independent (NOT iid)