No class next week

Next week :- Suppose to be midterm week.

In ISCD class we will have

20% weightage ⟷ { - 2 quizzes Thursday March 16th

9:00 am start | 11:30 am start

$\boxed{Q1}$ 11:00 am end | 1:30 pm start $\boxed{Q2}$

- Assignment format

+ online format

Rules:

1. Individual work
2. R-studio or notes or class website

Do not use internet

Estimation :-

Recall :- • $X_1, X_2, \ldots, X_n$ i.i.d. sample from an unknown distribution and discussed how to use this to estimate some aspects of the distribution.

• Behaviour of these estimates as $n \to \infty$. [asymptotic behaviour]

Here we will discuss specific methods to estimate parameters of the distribution from the sample.

Example :- Coin - assume probability p of Heads

To gather information about p :-

- Toss the coin 100 times.
- View Results : $X_1, X_2, \ldots X_{100}$ i.i.d. Bernoulli (p).

We observe (say) that $\sum_{i=1}^{100} X_i = 67$

- Can we use this information to estimate p?

We will discuss two methods first

① Method of moments

② Maximum likelihood estimate

- Both of these methods will produce a number as estimate for p.

Later : ③ Interval Estimation

- Provide a range which will the contain the true value of p with some accuracy.

④ Next topic (follows Estimation) - Hypothesis testing.

Assumptions :

① $X_1, X_2, \ldots, X_n$ are i.i.d. from X and

X <
- Discrete Random variable (p.m.f.)
- Continuous Random variable (p.d.f.)

— $f(\cdot)$

② $f(\cdot)$ - will depend on unknown parameters $(p_1, p_2, \ldots, p_d) \in \mathbb{R}^d$, $d \geq 1$

Notation: $f(\cdot \mid p_1, p_2, \ldots, p_d)$.

abbreviate $f(\cdot \mid p)$

$$p = (p_1, p_2, \ldots, p_d) \in \mathbb{R}^d ; \; d \geq 1.$$

Example :

$X \sim \text{Bernoulli}(p)$

$$f(x \mid p) = \begin{cases} p^x (1-p)^{1-x} & x \in \{0,1\} \\ 0 & \text{o.w.} \end{cases}$$

$X \sim \text{Normal}(\mu, \sigma^2)$

$$f(x \mid \mu, \sigma) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}, \; x \in \mathbb{R}$$

Notation :- $\cdot\ \mathcal{P} \subseteq \mathbb{R}^d$ $\quad p = (p_1, \ldots, p_d) \in \mathcal{P}$

Definition E.1.1 :- $X_1, X_2, \ldots, X_n$ be i.i.d. X.

$X \sim f(\cdot \mid p)$. Suppose we are interested in estimating $\theta(p)$

$$\theta : \mathcal{P} \longrightarrow \mathbb{R}$$

Let $g : \mathbb{R}^n \to \mathbb{R}$. Then $g(X_1, \ldots, X_n)$ is defined as a point estimator of $\theta(p)$.

Its value from a particular realisation of $X_1, \ldots, X_n$ is called an estimate

Example E.1.2 :- $\cdot$ X is random variable with $E[X] = \mu$ $\quad Var[X] = \sigma^2$

Sample mean - point estimator for μ - true mean	
	$\cdot\ X \sim f(\cdot \mid \mu, p_2, \ldots, p_d)$ $\quad d \geq 1$
	$\cdot\ \theta(\mu, p_2, \ldots, p_d) = \mu$
	$\cdot\ X_1, \ldots, X_n$ i.i.d X
	$\cdot\ g : \mathbb{R}^n \to \mathbb{R}$ $\quad g(x_1, \ldots, x_n) = \frac{\sum_{i=1}^{n} x_i}{n}$

Seen before:

$$E[g(X_1, \ldots, X_n)] = \mu$$

$$Var[g(X_1, \ldots, X_n)] = \frac{\sigma^2}{n}$$

E.2 Method of Moments

- $X_1, X_2, \ldots, X_n$ i.i.d sample from X
- $X \sim f(\cdot \mid p)$; $p = (p_1, p_2, \ldots p_d) \in \mathbb{R}^d$. $d \geq 1$

let $k \geq 1$, $m_k : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by

$$m_k(x) = \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

Note :-

$$m_k(X_1, X_2, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

(Ex.) → is the k^{th} moment of the Empirical distribution based on sample $X_1, \ldots, X_n$.

refer to it as the k^{th} moment of the sample

$\mu_k = E[X^k] = k^{th}$ moment of the distribution of X.

Note :- $X \sim f(\cdot \mid p_1, \ldots, p_d)$; we may view

$\mu_k \equiv$ function of $(p_1, \ldots, p_d) \equiv E[X^k]$

$\equiv \mu_k(p_1, \ldots, p_d)$

The Method of moments estimator for $(p_1, \ldots, p_d)$ is obtained by equating the first d moments of the sample to the corresponding moments of the distribution.

Specifically :-

$$(*) - \quad \mu_k(p_1, \ldots, p_d) = m_k(x_1, \ldots, x_n) \qquad k = 1, \ldots, d$$

Note :-

- $(*)$ - d equations in d unknowns given a sample $x_1, x_2, \ldots, x_n$
- no guarantee of a unique solution
- no guarantee of computing it.

Denote the solution from $(*)$ as $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_d$. and they will be written in terms of realised $m_k(x_1, \ldots x_n)$ $k = 1, \ldots, d$

Example E.2.1 :

Suppose we have $X_1, X_2, \ldots, X_{10}$ i.i.d. from X and $X \sim \text{Binomial}(N, p)$

- Neither N, p are known.

Assume that the realised values are

$$\underset{\substack{\parallel \\ x_1}}{8}, \underset{\substack{\parallel \\ x_2}}{7}, \ldots 6, 11, \underset{\substack{\parallel \\ x_5}}{8}, 5, 3, \underset{\substack{\parallel \\ x_8}}{7}, 6, \underset{\substack{\parallel \\ x_{10}}}{9}.$$

- It is immediate to see : (from Sample)

$$m_1(x_1, \ldots, x_{10}) = \underline{\sum_{i=1}^{10} x_i} = 7$$

$$m_2(x_1, \ldots, x_{10}) = \underline{\sum_{i=1}^{10} x_i^2} = 53.4.$$

- It is also well - known (from Binomial Distribution)

$$f(k \mid N, p) = \binom{N}{k} p^k (1-p)^{N-k} \qquad k = 0, 1, \ldots N$$

$$\mu_1 = E[X] = Np$$

$$\mu_2 = E[X^2] = Var[X] + E[X]^2 = Np(1-p) + (Np)^2$$

Equate:

$$\left.\begin{aligned} \mu_1 &= m_1(x_1, \ldots, x_n) \\ \mu_2 &= m_2(x_1, \ldots x_n) \end{aligned}\right\} \text{Method of moments}$$

to obtain

$$\left.\begin{aligned} Np &= 7 \\ Np(1-p) + (Np)^2 &= 53.4 \end{aligned}\right\} \quad (\ddagger)$$

Use elementary algebra :

$\therefore \hat{N} \simeq 19$ and $\hat{p} \simeq 0.375$

So from the sample :-

$$\underset{\substack{\parallel\\ x_1}}{8}, \ \underset{\substack{\parallel\\ x_2}}{7}, \ldots 6, \ 11, \ \underset{\substack{\parallel\\ x_5}}{8}, \ 5, \ 3, \ \underset{\substack{\parallel\\ x_8}}{7}, \ 6, \ \underset{\substack{\parallel\\ x_{10}}}{9}.$$

the Method of moments estimate

$\hat{N} \simeq 19$ and $\hat{p} \simeq 0.375$

Abstraction :- $X \sim \text{Binomial}(N, p)$

$$\mu_1 \equiv Np = m_1 \equiv m_1(X_1 \ldots X_n)$$

$$\mu_2 \equiv Np(1-p) + (Np)^2 = m_2 \equiv m_2(X_1 \ldots X_n)$$

$\Rightarrow$

$$\hat{N} = \frac{m_1^2}{m_1 - (m_2 - m_1^2)} \quad \text{and}$$

$$\hat{p} = \frac{m_1 - (m_2 - m_1^2)}{m_1}$$

Example E.2.2 :- $X \sim \text{Normal}(\mu, \sigma^2)$

$\mu_1 \equiv E[X] = \mu$

$\mu_2 \equiv E[X^2] = \mu^2 + \sigma^2$

$X_1, \ldots X_n$ from X i.i.d.

and sample moments

$m_1 \equiv m_1(X_1 \ldots X_n)$

$m_2 \equiv m_2(X_1 \ldots X_n)$

$\therefore$ we Equate :

$$\left.\begin{aligned} \mu &= m_1 \\ \mu^2 + \sigma^2 &= m_2 \end{aligned}\right\} \text{ - two equations obtained}$$

Solving :

$$\hat{\mu} = m_1 \equiv \overline{X}$$

$$\hat{\sigma} = \sqrt{m_2 - m_1^2} \equiv \sqrt{\frac{n-1}{n}}\, S$$

$\therefore$ Sample mean $\overline{X}$ and $\sqrt{\text{Sample variance } S^2}$ are the Method of moment estimate for

μ and σ

from the sample $X_1, X_2, \ldots, X_n$

Recall :- Estimation — point estimators

- $X_1, X_2, \ldots, X_n$ i.i.d from a population $X \sim f(\cdot \mid p)$
- Interest :- Estimating $\theta(p) \in \mathbb{R}$; $p \in \mathbb{R}^d$.
- $g : \mathbb{R}^n \to \mathbb{R}$; compute $g(X_1, \ldots, X_n)$ - point estimator for $\theta(p)$

Method of Moments

Start: $X_1, X_2, \ldots, X_n$ i.i.d. $f(\cdot \mid p_1, \ldots, p_d)$

$k = 1, \ldots, d$ compute $m_k(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i^k$

Sample moments

find in terms of p $\mu_k(p_1, \ldots, p_d) = E[X^k]$

True moments

Equate:

$$\mu_k = m_k(X_1, \ldots, X_n) \qquad k = 1, \ldots, d$$

Solve the d - equations in d-unknowns.

Remarks :-

- no solutions
- many solutions
- solution may not make sense.

Today :-

3 Maximum likelihood Estimate.

let $X_1, X_2, \dots, X_n$ be i.i.d. X and $X \sim f(\cdot \mid p_1, \dots, p_d)$

[either p.m.f. or p.d.f.]

$(p_1, p_2, \dots, p_d) \in \mathbb{P} \subseteq \mathbb{R}^d$.

Definition E.3.1 :- The likelihood function for the sample $X_1, X_2, \dots, X_n$ is the function

$$L : \mathbb{P} \times \mathbb{R}^n \longrightarrow \mathbb{R} \quad \text{given by}$$

$$L(p; X_1, X_2, \dots X_n) := \prod_{i=1}^{n} f(X_i \mid p)$$

For a given $(X_1, X_2, \dots X_n)$ suppose $\hat{p} \equiv \hat{p}(X_1, X_2, \dots X_n)$ is a point at which

$L(p; X_1, X_2, \dots X_n)$ attains its maximum as a function of p.

Then $\hat{p}$ is called a maximum likelihood estimator of p (or abbreviated as MLE of p) given the sample $X_1, X_2, \dots X_n$.

Remarks:-

- a unique $\hat{p}$ may not exist ; several variable calculus machinery may be required to solve the problem.

- $\hat{p}$ if unique :- can be thought of as the most likely value of the parameter p for the given realisation of $X_1, X_2, \dots X_n$.

Example E.3.2 :- $X \sim \text{Normal}(p, 1)$ $\quad p \in \mathbb{R}$

$X_1, X_2, \dots, X_n$ be i.i.d. X.

The likelihood function

$$L : P \times \mathbb{R}^n \to \mathbb{R}$$

$$L(p; x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} \frac{e^{-\frac{(x_i - p)^2}{2}}}{\sqrt{2\pi}}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i - p)^2}{2}}$$

To find MLE :-

- Treat $x_1, x_2, \dots, x_n$ as fixed.
- One needs to maximize $L(\ )$ as a function of p.

Solution:

Observe :-

- 

- $\left(\frac{1}{\sqrt{2\pi}}\right)^n$ is a constant

To maximize $L(p; x_1, x_2, \dots, x_n)$ as a function of p is equivalent to

minimising $g : \mathbb{R} \to \mathbb{R}$ where

$$g(p) = \sum_{i=1}^{n} \frac{(x_i - p)^2}{2}$$

Method 1 :-

$$g(p) = \sum_{i=1}^{n} (x_i - p)^2$$

$$= \sum_{i=1}^{n} \left[(x_i - \bar{x}) + (\bar{x} - p) \right]^2$$

Algebra Exercise

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - p)^2$$

Observe:

- $\sum_{i=1}^{n} (x_i - \bar{x})^2 \leq g(p) \quad \forall\ p \in \mathbb{R}$
- $g(p) = \sum_{i=1}^{n} (x_i - \bar{x})^2$ when $p = \bar{x}$

$\Rightarrow$ MLE of p given $x_1, x_2, \ldots, x_n$

is $\hat{p} = \bar{x}$.

Method 2 :-

$$g'(p) = -2 \sum_{i=1}^{n} (x_i - p)$$

$$g''(p) = +2$$

Set $g'(p) = 0$

$\Rightarrow \sum_{i=1}^{n} (x_i - p) = 0$

$\Rightarrow p = \bar{x}$

& $g''(\bar{x}) \geq 0 \Rightarrow$

$p = \bar{x}$ is a local minimum. $\longrightarrow$ global maximum.

(Exercise) $g(\cdot)$ – quadratic in p

$\hat{p} = \bar{x}$ is the MLE of p given $x_1, x_2, \ldots, x_n$

$\square$

Normal distribution :

$X \sim \text{Normal}(\mu, \sigma^2)$ then X has p.d.f.

given by

$$f(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} \; ; \; x \in \mathbb{R}$$

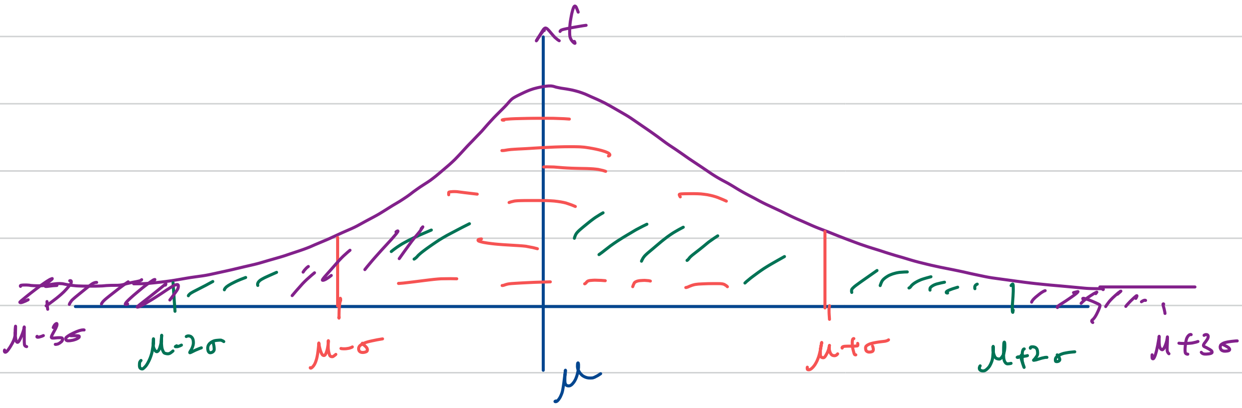
$$\mathbb{P}(X \le y) = \int_{-\infty}^{y} f(x)\,dx$$

$$= \int_{-\infty}^{y} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}\,dx$$

Can't evaluate integrals explicitly

Normal Tables

R - pnorm (.)



Observations :-

· 68 - 95 - 99 rule

$$P(|X-\mu| < \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)\,dx \approx 0.68$$

$$\mathbb{P}(|X-\mu| < 2\sigma) = \int_{\mu-2\sigma}^{\mu+2\sigma} f(x)\,dx \approx 0.95$$

$$\mathbb{P}(|X-\mu| < 3\sigma) = \int_{\mu-3\sigma}^{\mu+3\sigma} f(x)\,dx \approx 0.99$$

- Skewness & kurtosis :

$$\sum_{i=1}^{n} \left(\frac{x_i-\bar{x}}{\sigma_x}\right)^3$$

if skewness $= 0$
then the distribution
is symmetric

$$\sum_{i=1}^{n} \left(\frac{x_i-\bar{x}}{\sigma_x}\right)^4$$

Kurtosis measures
the peak of the
distribution

Normal : Kurtosis $= 3$

- if Data $(x_1, \dots, x_n)$ has skewness
far from 0 and kurtosis far from 3
then we conclude
Data is not Normal.

Q-Q plot

Consider data $= (x_1, x_2, \ldots, x_n)$

order them : $X_{(1)}, X_{(2)}, \ldots X_{(k)}, X_{(n)}$.

$\frac{k}{n+1}$ – Sample quantile

For $\alpha \in (0,1)$ find z_α such that

$$\mathbb{P}(X \leq z_\alpha) = \alpha$$

(α^{th} – quantile of Normal)

Q-Q plot :- $\left(z_{\frac{k}{n+1}}, X_{(k)}\right)$

If plot is Not a straight line then

Data is most likely not normal.