Estimation ::

Definition :- let $n \geq 1$. let $X_1, X_2, \dots X_n$ be i.i.d X

let $g : \mathbb{R}^n \to \mathbb{R}$. Then

$g(X_1, X_2, \dots, X_n)$ is the point estimator

from the sample $X_1, X_2, \dots, X_n$.

A particular realisation is called an estimate.

Example:. $g(x) = \frac{1}{n} \sum_{i=1}^{n} x_i$

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

$\mu = E[X]$ $\sigma^2 = Var[X]$

$g(X_1, \dots, X_n) = \overline{X}$

(Sample mean)

(unbiasedness)

$E[\overline{X}] = \mu$

$Var[\overline{X}] = \frac{\sigma^2}{n}$

(Consistency)

Q:- $X \sim$ Bernoulli (p) Estimate p ?

Take sample $X_1, X_2, \dots, X_{100}$ i.i.d. X

$$\overline{X} = \frac{\sum_{i=1}^{100} x_i}{100} := \hat{p} \quad \text{estimate of } p.$$

How good is the estimate?

Confidence interval for p

Hypothesis testing from estimate

Two methods :- ① Method of moments

② maximum likelihood estimate.

① Method of moments :-

let $X_1, X_2, \ldots, X_n$ be i.i.d X.

$$m_k(X) = E[X^k] := k^{th} \text{ moment of } X.$$

$$\mu_k : \mathbb{R}^n \longrightarrow \mathbb{R}$$

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

Assumption : Suppose X is a r.v. such that its p.m.f or p.d.f is a function of d-parameters say $p_1, p_2, \ldots, p_d$.

Q: How to estimate $p_1 \ldots p_d$?

Example: ① $X \sim \text{Normal}(\mu, \sigma^2)$ $d=2$ $p_1 = \mu$, $p_2 = \sigma$

② $X \sim \text{Bernoulli}(p)$ $d=1$ $p_1 = p$

Step 1: Compute $\mu_k(X_1, \ldots, X_n) := \frac{1}{n}\sum_{i=1}^{n} X_i^k$ from the sample step2 :- Compute from p.m.f or p.d.f

$$m_k(x) = E[X^k] \equiv g(k; \theta_1 \dots \theta_d)$$

step3 :-

$$m_k = \mu_k(x_1 \dots x_n) \qquad k = 1, \dots, d$$

Solve for $\theta_1, \theta_2, \dots, \theta_d$

Benefits / limitations :-

- We may get an estimate
- no guarantee of a solution
- no guarantee that solution will make sense

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

$$\mu = E[X] := m_1(X)$$

$$\sigma^2 = Var[X] \equiv E[X^2] - E[X]^2$$
$$= m_2(X) - (m_1(X))^2$$

$$\Rightarrow \quad m_1 = \mu \quad \text{and} \quad m_2 = \sigma^2 + \mu^2$$

- $\mu_1(X_1 \ldots X_n) = \bar{X}$

$$\mu_2(X \ldots X_n) = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

- Set $\mu_1 = \mu$

  $\mu_2 = \sigma^2 + \mu^2$ $\Rightarrow$

$$\boxed{\hat{\mu} = \mu_1} \quad \ldots \quad \bar{X}$$

$$\boxed{\hat{\sigma} = \sqrt{\mu_2 - (\mu_1)^2}} \quad \ldots \quad \text{" } S \text{ "}$$

② maximum likelihood estimate.

$d \geq 1$ $\quad \mathcal{P} \subseteq \mathbb{R}^d$ – Parameter

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

X has p.m.f / p.d.f $f(\cdot \mid p)$

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from X.

<u>Definition:</u> The likelihood function

$L : \mathcal{P} \to \mathbb{R}$ from the sample is given by

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

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

Suppose $\hat{p} \equiv \hat{p}(X_1, \ldots, X_n)$ is a point in $\mathcal{P}$ such that

$L(p \mid X_1, \ldots, X_n)$ attains its maximum as a function of p at $\hat{p}$. Then $\hat{p}$ is called the MAXIMUM LIKELIHOOD ESTIMATE of p given the sample $X_1, \ldots, X_n$

Example :- $n \geq 1$.

$X_1, X_2, \ldots, X_n$ be iid. X

$X \sim \text{Normal}(\mu, 1)$

$$\boxed{\begin{matrix} \mathcal{P} = \mathbb{R} \\ \mu \in \mathbb{R} \end{matrix}, \ d=1}$$

- Likelihood function :

$$L(\mu \mid X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} f(x_i \mid \mu, 1)$$

$$= \prod_{i=1}^{n} \frac{\exp\left(-\frac{(X_i - \mu)^2}{2}\right)}{\sqrt{2\pi}}$$

$$L(\mu \mid X_1 \ldots X_n) = \frac{\exp\left(-\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{2}\right)}{(2\pi)^{n/2}}$$

Suppress notation :

$$L(\mu) = \frac{\exp\left(-\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{2}\right)}{(2\pi)^{n/2}}$$

$$L'(\mu) = L(\mu) \cdot \left[-\sum_{i=1}^{n} (x_i - \mu) \right]$$

Set $L'(\mu) = 0$

to get $\hat{\mu} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ Ex: Global maximum of $L(\cdot)$

$\therefore$ Maximum likelihood Estimate for μ when given sample $X_1 \ldots X_n$ from Normal $(\mu, 1)$

Is $\hat{\mu} = \overline{X}$

Example: $X \sim$ Bernoulli (p) ; $f(x) = p^x (1-p)^{1-x}$
$x \in \{0, 1\}$

$X_1, X_2, \ldots, X_n$ i.i.d X

$$L(p \mid X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} f(x_i \mid p)$$

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

Suppress $X_1 \ldots, X_n$ from notation

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

$$T(p) = \log(L(p))$$ [log - likelihood function]

$$= \left(\sum_{i=1}^{n} x_i\right)\log(p) + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-p)$$

$$T'(p) = \frac{\left(\sum_{i=1}^{n} x_i\right)}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

$$= \begin{cases} -\dfrac{n}{1-p} & \sum_{i=1}^{n} x_i = 0 \quad *_1 \\ \dfrac{n}{p} & \sum_{i=1}^{n} x_i = n \quad *_2 \\ \dfrac{a}{p} + \dfrac{a-n}{1-p} & \sum_{i=1}^{n} x_i = a \end{cases}$$

Assume $\sum_{i=1}^{n} x_i = a \qquad 0 < a < n$

$$T'(p) = 0 \Rightarrow \frac{a-n}{1-p} = -\frac{a}{p}$$

$$\Rightarrow \hat{p} = \frac{a}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}$$

$*_1$ $\sum_{i=1}^{n} x_i = 0$, $T'(p) = -\dfrac{n}{1-p} = 0$ ... No such p

$T(p) = n\log(1-p)$ $\equiv$ maximized $p=0$ $\equiv$ M.L.E

$\hat{p} = \overline{X}$

$*_2$ $\sum_{i=1}^{n} X_i = n$, $\Rightarrow$ $\frac{n}{p} = 0$ .... No such p

$T(p) = n\log(p)$ $\equiv$ maximized $p=1$ $\equiv$ MLE

$\hat{p} = \overline{X}$

$\therefore$ M.L.E. for p $= \overline{X}$.

$\square$