Parcametric Point Estimation

Let $X_1, \ldots, X_n$ be a random sample from a population described by family $\mathcal{F} = \{ f_{\theta}(\cdot) : \underline{\theta} \in \Theta \}$ of pdfs/pmfs where for each $\underline{\theta} \in \Theta$ form of $f_{\underline{\theta}}(\cdot)$ is known but $\underline{\theta} \in \Theta$ is unknown.

Here knowledge of unknown $\underline{\theta} \in \Theta$ yields knowledge of unknown $\underline{\theta} \in \Theta$ yields knowledge of the entire population. More over $\underline{\theta}$ itself may represent an importent characteristic of the population (such as population mean, variance etc) and there may be direct interest in obtaining a point estimate of $\underline{\theta}$. Some times there may be interset in obtaining point estimate of $g(\theta)$, ~~for~~ a given fun of $\underline{\theta}$.

<u>Goal</u>: Based on a random sample $X_1, \ldots, X_n$ from the population, find a good point estimate of $g(\underline{\theta})$

<u>Def^n</u>: A point estimator of $g(\theta)$ is a fun ~~of~~ $W(\underline{X})$ of the random sample

$$\underline{X} = (X_1, \ldots X_n)$$

Note : (1) An estimator $W(\underline{X})$ is a random variable whereas an estimate is an observed value of the estimator based on an observed sample.

Different Method of finding Estimator

(1) Method of Moment Estimator (MME)

Let $X_1, \dots, X_n$ be a random sample from a population with dist^n $F_{\underline{\theta}}$, $\underline{\theta} = (\theta_1, \dots, \theta_k) \in \Theta$

consider k - non-central moments

$$\left.\begin{aligned} \mu_1' &= E(X_1) = g_1(\underline{\theta}) \\ \mu_2' &= E(X_1^2) = g_2(\underline{\theta}) \\ \mu_3' &= E(X_1^3) = g_3(\underline{\theta}) \\ &\vdots \\ \mu_k' &= E(X_1^k) = g_k(\underline{\theta}) \end{aligned}\right\} \text{——— (1)}$$

Assume the system of equ^n (1) have sol^n and solving for $\theta_1, \dots, \theta_k$ we get

$$\left.\begin{aligned} \theta_1 &= h_1(\mu_1', \dots, \mu_k') \\ &\vdots \\ \theta_k &= h_k(\mu_1', \dots, \mu_k') \end{aligned}\right\} \text{——— (2)}$$

Define the 1^{st} k non central sample moments

$$\alpha_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \alpha_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \quad \ldots, \quad \alpha_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

In method of moments we estimate k^{th} population moment by k^{th} sample moment

i.e. $\hat{\mu}_j = \alpha_j, \quad j = 1, 2, \ldots k.$

Thus the method of moment estimators (MME) of $\theta_1, \ldots, \theta_k$ are

$$\hat{\theta}_1 = h_1(\alpha_1, \ldots, \alpha_k), \quad \ldots, \quad \hat{\theta}_k = h(\alpha_1, \ldots, \alpha_k).$$

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

Find MME of $\underline{\theta} = (\mu, \sigma^2)$, and $(\mu + \sigma) = g\,\Psi(\underline{\theta})$.

Solⁿ:
$$\left.\begin{aligned} \mu_1' &= E(X_1) = \mu. \\ \mu_2' &= E(X^2) = \sigma^2 + \mu^2 \end{aligned}\right\} \Rightarrow \begin{aligned} \mu &= \mu_1' \\ \sigma^2 &= \mu_2' - \mu_1'^2 \end{aligned}$$

So $\hat{\mu}_{MME} = \bar{X}$, $\quad \hat{\sigma}^2_{MME} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$$\boxed{\hat{\mu}_{MME} = \bar{X}, \quad \hat{\sigma}^2_{MME} = \frac{1}{n}\sum (X_i - \bar{X})^2}$$

MME of $\psi(\underline{\theta}) = \mu + \sigma$ is $\bar{x} + \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$

Ex: $X_1, \ldots, X_n \underset{\sim}{iid} \text{Bin}(m, p)$, $\underline{\theta} = (m, p) \in \Theta$

$\Theta = \{1, 2, \ldots\} \times (0,1)$.

$\mu_1' = E(X_1) = mp$ ——— (I)

$\mu_2' = E(X_1^2) = mp(1-p) + m^2p^2$ ——— (II)

Now we have $mp(1-p) = \mu_2' - \mu_1'^2$

$$\Rightarrow (1-p) = \frac{\mu_2' - \mu_1'^2}{\mu_1'}$$

$$\Rightarrow p = 1 - \frac{\mu_2' - \mu_1'^2}{\mu_1'} \text{ and } m = \frac{\mu_1'}{p}$$

$$\hat{p}_{MME} = 1 - \frac{\frac{1}{n}\sum X_i^2 - \bar{X}^2}{\bar{X}} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})^2}{\bar{X}}$$

$$\hat{m}_{MME} = \frac{\bar{X}}{\hat{p}_{MME}} = \frac{\bar{X}^2}{1 - \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Remark: (1) Sometimes for K-dimensional parameters we may have to consider more than K-equations.

Ex: $X_1, \ldots, X_n \overset{iid}{\sim} U[-\theta, \theta]$, $\theta > 0$.

$\mu_1' = E(X) = 0$, $\mu_2' = E(X_1^2) = \frac{\theta^2}{3} \Rightarrow \theta = \sqrt{3\mu_2'}$

$$\hat{\theta}_{MME} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2}$$

Here the 1^{st} moment did not give any solu^n so we are using 2^{nd} moment.

(ii) MME may not exist

(iii) MME is not unique.

Method of Maximum Likelihood Estimation :⇒

For a given observed sample point $\underline{x} = (x_1, \ldots, x_n)$ define

$$L_{\underline{x}}(\underline{\theta}) = \prod_{i=1}^{n} f(x_i \mid \underline{\theta}) \quad , \quad \underline{\theta} \in \Theta$$

as a function of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$

$L_{\underline{x}}(\underline{\theta})$: The probability that the observed sample point $\underline{x}$ came from population represented by pdf/pmf $f(\underline{x} \mid \underline{\theta})$, $\underline{\theta} \in \Theta$, i.e. the likelihood of observing r.s. $(x_1, \ldots, x_n)$ from $f(\cdot \mid \underline{\theta})$, $\theta \in \Theta$.

Def^n (1) For a given sample point $\underline{x}$, the fu^n $L_{\underline{x}}(\underline{\theta})$ as a fu^n of $\theta \in \Theta$ is called the likelihood function.

It makes sense to find $\hat{\underline{\theta}}$ that maximizes $L_{\underline{x}}(\underline{\theta})$ for a given sample point $\underline{x}$, as the corresponding population (represented by pdf/pmf) is most likely to have yielded the observed sample $\underline{x}$.

Def^n: For each $\underline{x} \in \mathcal{X}$, let $\hat{\theta} = \hat{\theta}(\underline{x})$ be such that

$$L_{\underline{x}}(\hat{\theta}) = \sup_{\underline{\theta} \in \Theta} L_{x}(\underline{\theta})$$

Then a maximum likelihood estimator (MLE) of the parameter $\underline{\theta} = (\theta_1, \dots \theta_k)$, based on random sample $\underline{X}$ is $\hat{\theta}(\underline{X})$.

Finding MLE:

$\hat{\theta}$: MLE if

$$L_{\underline{x}}(\hat{\theta}) = \sup_{\underline{\theta} \in \Theta} L_{\underline{x}}(\underline{\theta}).$$

Define $l_{\underline{x}}(\underline{\theta}) = \log L_{\underline{x}}(\underline{\theta})$, $\underline{\theta} \in \Theta$. This is called log-likelihood function.

MLE maximizes $L_{\underline{x}}(\underline{\theta})$ or equivalently $l_{\underline{x}}(\underline{\theta})$

If $l_{\underline{x}}(\underline{\theta})$ is differentiable and maximum at the interior of Θ then MLE $\hat{\theta}(\underline{x})$ satisfies

$$\left.\frac{\partial}{\partial \theta_i} l_{\underline{x}}(\underline{\theta})\right]_{\underline{\theta} = \hat{\theta}(\underline{x})} = 0, \quad i = 1, 2, \dots K. \longrightarrow (*)$$

Examples (1) Let $X \sim Bin(n, p)$, $0 \le p \le 1$, $\theta = p \in [0,1]$

Here n is known.

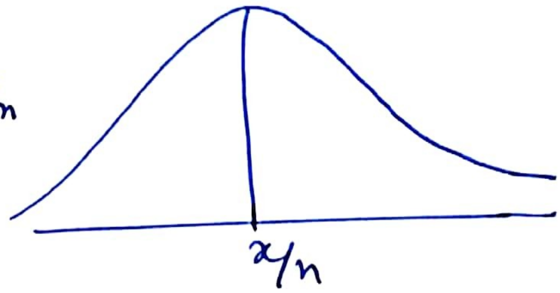
$$L_{\underset{\sim}{x}}(p) = f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots n$$

$$\log L_x(p) = l_{\underset{\sim}{x}}(p) = \log\binom{n}{x} + x \log p + (n-x)\log(1-p)$$

$$\frac{\partial l_x}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)} \begin{cases} <0 \text{ if } p > x/n \\ >0 \text{ if } p < x/n \end{cases}$$

$l_x(p) \uparrow$ if $p < x/n$

$\downarrow$ if $p > x/n$



So $\hat{p}_{ML} = \frac{X}{n}$.

(2) Let $X_1, \dots, X_n \overset{iid}{\sim} P(\lambda)$, $\lambda > 0$ and $\lambda \le \lambda_0$

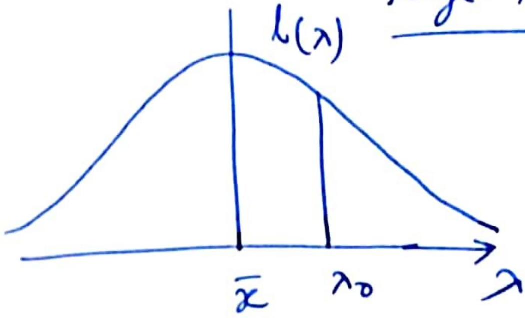
$$L_{\underline{x}}(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\Sigma x_i}}{\prod_{i=1}^{n}(x_i!)}$$

$$l_{\underline{x}}(\lambda) = \log L_x(\lambda) = -n\lambda + \Sigma x_i \log\lambda - \log\left\{\prod_{i=1}^{n}(x_i!)\right\}$$

$$\frac{\partial l_x(\lambda)}{\partial\lambda} = -n + \frac{\Sigma x_i}{\lambda} = \frac{\Sigma x_i - n\lambda}{\lambda} \begin{cases} >0 \text{ if } \lambda < \bar{x} \\ <0 \text{ if } \lambda > \bar{x} \end{cases}$$

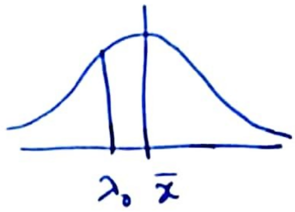
$l_x(\lambda)$ $\uparrow$ $\lambda < \bar{x}$

$\downarrow$ $\lambda > \bar{x}$



Given $\lambda \le \lambda_0$ i.e. $\Theta = (0, \lambda_0)$

$$\hat{\lambda}_{RML} = \begin{cases} \bar{x} & \text{if } \bar{x} \le \lambda_0 \\ \lambda_0 & \text{if } \bar{x} > \lambda_0 \end{cases}$$



This is restricted MLE.

Remark (1) MLE is not unique

Ex: $X_1, \ldots, X_n \overset{iid}{\sim} U[\theta - a, \theta + a]$, $a \in \mathbb{R}$, $a > 0$

where a is a known constant.

The likelihood function

$$L_{\underline{x}}(\theta) = \begin{cases} \left(\frac{1}{2a}\right)^n & \text{if } \theta - a \le x_{(1)} < x_{(2)} < \cdots < x_{(n)} \le \theta + a \\ 0, & \text{o/w.} \end{cases}$$

So $L_{\underline{x}}(\theta)$ is maximum when

$\theta - a \le x_{(1)}$ and $x_{(n)} < \theta + a$.

i.e. $x_{(n)} - a \le \theta \le x_{(1)} + a$.

So any value of θ between $x_{(n)}-a$ to $x_{(n)}+a$ is MLE of θ.

We may choose the midpoint i.e. $\frac{x_{(1)}-x_{(n)}}{2}$ as the MLE.

Ex

Let $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma > 0$

$$L_{\underline{x}}(\theta) = \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \right], \quad \mu \in \mathbb{R}, \sigma > 0, x_i \in \mathbb{R}$$

$i = 1(1)n$

$$= \frac{1}{\sigma^n (2\pi)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Now $l_{\underline{x}}(\theta) = \log L_{\underline{x}}(\underline{\theta}) = -n\log\sigma - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$

$$\frac{\partial l_{\underline{x}}(\theta)}{\partial\mu} = \frac{1}{\sigma^2}\sum(x_i-\mu) \quad \therefore \frac{\partial l_{\underline{x}}(\theta)}{\partial\mu} = 0 \Rightarrow \mu = \bar{x}$$

$$\frac{\partial l}{\partial\sigma^2} = 0 \Rightarrow \sigma^2 = \frac{1}{n}\sum(x_i-\mu)^2$$

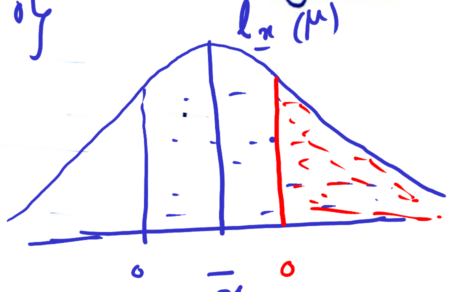
So $\hat{\mu}_{ML} = \bar{X}$ and $\sigma^2_{ML} = \frac{1}{n}\sum(X_i-\bar{X})^2$

Consider the MLE of μ & σ^2 when $\mu > 0$. In this case $\mu \in (0,\infty)$

We have $\frac{\partial l_x}{\partial\mu} = \frac{n(\bar{x}-\mu)}{\sigma}$ > 0 if $\mu < \bar{x}$

< 0 if $\mu > \bar{x}$

So $\hat{\mu}_{RML} = \begin{cases} \bar{X} & \text{if } \bar{X} > 0 \\ 0 & \text{if } \bar{X} \le 0 \end{cases} = \max\{\bar{X}, 0\}$



$$\hat{\sigma}^2_{RML} = \frac{1}{n}\sum(X_i - \hat{\mu}_{RML})^2 = \begin{cases} \frac{1}{n}\sum(X_i-\bar{X})^2, & \bar{X} > 0 \\ \frac{1}{n}\sum X_i^2 & \text{if } \bar{X} \le 0 \end{cases}$$

Ex: (i) Let $X_1, X_2, \dots X_n \underset{\sim}{iid} \text{Exp}(\mu, \sigma)$. Find MLE of μ & σ

(ii) Let $X_1, X_2, \dots X_n \underset{\sim}{iid} \text{Gamma}(r, \lambda)$. Find MLE of r, λ.

Ex: Let $X_1, X_2, \dots X_n \underset{\sim}{iid} U(0, \theta)$.

Then we have

$$L_{\underline{x}}(\theta) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_i < \theta, \quad i = 1(1)n \\ 0, & o/w \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n}, & 0 < x_{(1)} < x_{(2)} < \dots x_{(n)} < \theta \\ 0, & o/w \end{cases}$$

$L_{\underline{x}}(\theta)$ is the decreasing fun^n of θ. So $L_{\underline{x}}(\theta)$ attains its maximum when θ is minimum $\Rightarrow \hat{\theta}_{ML} = X_{(n)}$.

Ex. Let $X_1, \dots X_n \underset{\sim}{iid} U(\theta, \theta+1)$. Find MLE of θ.

Remark (3) Finding MLE requires maximization of likelihood function which is some times diffcult and may require numerical optimization techniques.

Remark ④: MLE is sometimes sensitive to data, a slightly different data may produce a vastly different MLE.

Invariance Property of MLE: Let $\hat{\theta} = \theta(\underset{\sim}{X})$ be a MLE of $\hat{\theta}$ then the of MLE of $\psi(\theta)$ is $\psi(\hat{\theta})$.

Ex (1) Let $X_1, \ldots X_n \overset{iid}{\sim} Bin(n, \theta)$. Then find a MLE of $\theta(1-\theta)$

(2) Let $X_1, \ldots X_m \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$. Then find the distn of The two sample aree independent.

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(m-1)S_1^2}{\sigma_1^2}+\frac{(n-1)S_2^2}{\sigma_2^2}}}\sqrt{\frac{m+n-2}{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$$

③ Let $X_1, \ldots X_n \overset{iid}{\sim} U(\theta_1, \theta_2)$. Find MLE of θ_1 & θ_2

Efficiency of Estimators

Let $g(\theta)$ be a parametric function and $\delta(\underline{x})$ be an estimator

Mean absolute Error : $E|\delta(\underline{x}) - g(\theta)|$

Mean squared Error : $E(\delta(\underline{x}) - g(\theta))^2$

Defn: We say that estimator δ_1 is better (more efficient) than δ_2 if

$$MSE(\delta_1) \le MSE(\delta_2) \quad \forall\ \theta \in \Theta$$

If $E(\delta(\underline{x})) = g(\theta)$ then

$$MSE(\delta) = E(\delta(\underline{x}) - g(\theta))^2 = Var(\delta(\underline{x}))$$

Unbiased estimator: A estimator $\delta(\underline{x})$ is said to be an unbiased estimator of $g(\theta)$ if

$$E(\delta(\underline{x})) = g(\theta) \quad \forall\ \theta \in \Theta$$

Ex: $X_1 \dots X_n \sim N(\mu, \sigma^2)$, $\delta_1(\underline{x}) = \overline{X}$.

$E(\delta_1(\underline{x})) = \mu \Rightarrow \delta_1(\underline{x})$ is unbiased for μ.