Example 1. Let $X \sim b(n, p)$. One observation on X is available, and it is known that n is either 2 or 3 and $p = \frac{1}{2}$ or $\frac{1}{3}$. Our object is to find an estimate of the pair (n, p). The following table gives the probability that $X = x$ for each possible pair (n, p):

x	$(2, \frac{1}{2})$	$(2, \frac{1}{3})$	$(3, \frac{1}{2})$	$(3, \frac{1}{3})$	Maximum Probability
0	$\frac{1}{4}$	$\frac{4}{9}$	$\frac{1}{8}$	$\frac{8}{27}$	$\frac{4}{9}$
1	$\frac{1}{2}$	$\frac{4}{9}$	$\frac{3}{8}$	$\frac{12}{27}$	$\frac{1}{2}$
2	$\frac{1}{4}$	$\frac{1}{9}$	$\frac{3}{8}$	$\frac{6}{27}$	$\frac{3}{8}$
3	0	0	$\frac{1}{8}$	$\frac{1}{27}$	$\frac{1}{8}$

The last column gives the maximum probability in each row, that is, for each value that X assumes. If the value $x = 1$, say, is observed, it is more probable that it came from the distribution $b(2, \frac{1}{2})$ than from any of the other distributions, and so on. The following estimate is, therefore, reasonable in that it maximizes the probability of the observed value:

$$(\hat{n}, \hat{p})(x) = \begin{cases} (2, \frac{1}{3}) & \text{if} \quad x = 0, \\ (2, \frac{1}{2}) & \text{if} \quad x = 1, \\ (3, \frac{1}{2}) & \text{if} \quad x = 2, \\ (3, \frac{1}{2}) & \text{if} \quad x = 3. \end{cases}$$

The *principle of maximum likelihood* essentially assumes that the sample is representative of the population and chooses as the estimate that value of the parameter that maximizes the pdf (pmf) $f_\theta(x)$.

Let $X_1, X_2, \ldots, X_n$ be 'n' random variables with their joint p.m.f. or p.d.f. $f_{\underset{\sim}{\theta}}(\underset{\sim}{x})$, where, $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)'$: realization vector with x_i being the realization of X_i, $\forall\ 1 \leq i \leq n$;

$\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_s)'$: unknown parametric vector $\in \Theta$.

If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} f_{\underset{\sim}{\theta}}(x)$, then, $\prod_{i=1}^{n} f_{\underset{\sim}{\theta}}(x_i) = f_{\underset{\sim}{\theta}}(\underset{\sim}{x})$ where, $\underset{\sim}{\theta}$ is known, but $\underset{\sim}{x}$ ie ~~(x_1, x_n)~~ $(x_1, x_2, \ldots, x_n)$ are unknown &

$\prod_{i=1}^{n} f_{\underset{\sim}{\theta}}(x_i) = L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$, if $\underset{\sim}{x}$ is fixed or given, but $\underset{\sim}{\theta}$ ie $\theta_1, \theta_2, \ldots, \theta_s$ is unknown. Then, $f_{\underset{\sim}{\theta}}(\underset{\sim}{x})$ is called the joint p.m.f. or p.d.f. of $x_1, \ldots, x_n$ provided $\underset{\sim}{\theta}$ is kept fixed & $L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$ is called the likelihood function of $\underset{\sim}{\theta}$ provided $\underset{\sim}{x}$ is kept fixed.

Similarly $\ln L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$ is called the log likelihood function of $\underset{\sim}{\theta}$ for given $\underset{\sim}{x}$. Now, the value or estimate of $\underset{\sim}{\theta}$ which maximizes $L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$ or $\ln L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$ is called the <u>maximum likelihood estimate</u> & the corresponding estimator as the <u>maximum likelihood estimator</u> (MLE) of $\underset{\sim}{\theta}$. Obviously, $L_{\underset{\sim}{x}}(\hat{\underset{\sim}{\theta}}_{MLE}) \geq L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$ [or, $\ln L_{\underset{\sim}{x}}(\hat{\underset{\sim}{\theta}}_{MLE}) \geq \ln L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$]

where, $\hat{\underset{\sim}{\theta}}_{MLE}$ denotes the MLE of $\underset{\sim}{\theta}$. So, we can say that $\hat{\underset{\sim}{\theta}}_{MLE}$ is the most likely value of $\underset{\sim}{\theta}$, for a given $\underset{\sim}{x}$.

Note: If the likelihood function of $\underset{\sim}{\theta}$ is twice differentiable w.r.t. $\underset{\sim}{\theta}$ then, using maxima-minima principle, we can maximize $L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$, i.e. $\ln L_{\underset{\sim}{x}}(\underset{\sim}{\theta})$. Then, $\frac{\partial}{\partial \theta} L_{\underset{\sim}{x}}(\underset{\sim}{\theta}) = 0$ is ~~the most likely value of $\underset{\sim}{\theta}$ for a given $\underset{\sim}{x}$~~ called the likelihood equation of $\underset{\sim}{\theta}$ & obviously

$$\frac{\partial^2}{\partial \theta^2} L_{\underset{\sim}{x}}(\underset{\sim}{\theta}) \Big|_{\underset{\sim}{\theta} = \hat{\theta}_{MLE}} < 0.$$

Example → 1: Let, $X_1, X_2, \dots, X_n \overset{iid}{\sim} Bin(m, p)$. Then find $\hat{p}_{MLE}$.

Solution: Since, $X_i \overset{iid}{\sim} Bin(m, p)$ $\forall\, i = 1(1)m$, so their common p.d.f. is

$$f_p(x) = \begin{cases} \binom{m}{x} p^x (1-p)^{m-x}, & x = 0, 1, 2, \dots, m \\ 0, & \text{o.w.} \end{cases} \quad 0 \le p \le 1.$$

∴ The likelihood function of p for given $\underset{\sim}{x}$ is

$$L_{\underset{\sim}{x}}(p) = \prod_{i=1}^{n} f_p(x_i) = \prod_{i=1}^{m} \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i}$$

$$L_{\underset{\sim}{x}}(p) = \begin{cases} \left(\prod_{i=1}^{n} \binom{m}{x_i} \right) p^{\sum_i x_i} (1-p)^{mn - \sum_i x_i}, & 0 \le p \le 1 \\ 0, & \text{o.w.} \end{cases}$$

$\therefore \ln L_{\underline{x}}(p) = \sum_{i=1}^{n} \ln \binom{m}{x_i} + \sum_{i=1}^{n} x_i \ln p + \left(mn - \sum_{i=1}^{n} x_i\right) \ln(1-p)$

$\therefore \dfrac{\partial \ln L_{\underline{x}}(p)}{\partial p} = \dfrac{\sum_{i=1}^{n} x_i}{p} - \dfrac{mn - \sum_{i=1}^{n} x_i}{1-p}$.

$\therefore$ for optimum 'p', we have,

$\dfrac{\partial \ln L_{\underline{x}}(p)}{\partial p} = 0 \Rightarrow \dfrac{\sum_{i=1}^{n} x_i}{p} = \dfrac{mn}{1-p} - \dfrac{\sum_{i=1}^{n} x_i}{1-p}$ or, $\sum_{i=1}^{n} x_i \left(\dfrac{1}{p} + \dfrac{1}{1-p}\right) = \dfrac{mn}{1-p}$.

or, $\sum_{i=1}^{n} x_i \cdot \dfrac{1}{p(1-p)} = \dfrac{mn}{1-p}$ or, $\dfrac{1}{p} \sum_{i=1}^{n} x_i = mn$ $\therefore \hat{p} = \dfrac{\sum_{i=1}^{n} x_i}{mn} = \dfrac{\bar{x}}{m}$.

Again, $\dfrac{\partial^2 \ln L_{\underline{x}}(p)}{\partial p^2} = -\dfrac{\sum_{i=1}^{n} x_i}{p^2} - \dfrac{mn - \sum_{i=1}^{n} x_i}{(1-p)^2} = -\left(\dfrac{m^2 \sum_{i=1}^{n} x_i}{\bar{x}^2} + \dfrac{mn - \sum_{i=1}^{n} x_i}{\left(1-\frac{\bar{x}}{m}\right)^2}\right)$

$= -\left(\dfrac{m^2 n \bar{x}}{\bar{x}^2} + \dfrac{m^2 n - mn\bar{x}}{(m-\bar{x})^2}\right) = -m^2 n \left(\dfrac{1}{\bar{x}} + \dfrac{1}{m-\bar{x}}\right)$.

Now, $0 \le x_i \le m$, $\forall i = 1(1)n$ i.e. $0 \le \sum_{i=1}^{n} x_i \le mn$.

i.e. $0 \le \bar{x} \le m$ $\Rightarrow \bar{x} \ge 0$ & $\bar{x} \le m$ i.e. $(m - \bar{x}) \ge 0$.

$\therefore \dfrac{1}{\bar{x}} + \dfrac{1}{m-\bar{x}} \ge 0$.

Hence, $p = \dfrac{\bar{x}}{m}$ maximizes $L_{\underline{x}}(p)$. $\therefore$ $\boxed{\hat{p}_{MLE} = \dfrac{\bar{X}}{m} = \dfrac{\sum_{i=1}^{n} X_i}{mn}}$

Following situations may arise for MLE.

① MLE may not be unbiased:

Example: Let, $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where, μ & σ^2 are unknown.

Then, $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

$\therefore E(\hat{\sigma}^2_{MLE}) = \frac{\sigma^2}{n} E\left[\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2}\right] = \frac{\sigma^2}{n} E(\chi^2_{n-1}) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$.

$\therefore \hat{\sigma}^2_{MLE}$ is not an unbiased estimator of σ^2, $\forall \sigma > 0$

Note: As $n \to \infty$, $\frac{n-1}{n} \to 1$ & then, $E(\hat{\sigma}^2_{MLE}) \to \sigma^2$.

② MLE may not exist at all. [C.U. → 2003].

Example: Let, $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Ber(p)$, $0 < p < 1$.

Then, the likelihood function of 'p' is,

$$L_x(p) = \begin{cases} p^{\sum_i x_i}(1-p)^{n-\sum_i x_i} & , 0 < p < 1 \\ 0, & o.w. \end{cases}$$

$\therefore \ln L_x(p) = \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$.

$$\therefore \frac{\partial \ln L}{\partial \ln p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\frac{\partial \ln L}{\partial p} = 0 \Rightarrow \frac{\sum_i x_i}{p} = \frac{n - \sum_i x_i}{1-p} \quad \text{or, } \sum_i x_i - p\sum_i x_i = np - p\sum_i x_i$$

$$\therefore p = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}. \quad \& \quad \frac{\partial^2}{\partial p^2} \ln L \Big|_{p=\bar{x}} < 0.$$

$\therefore$ $p = \bar{x}$ maximizes $L_{\underline{x}}(p)$. Hence, $\hat{p} = \hat{p}_{MLE} = \bar{X}$, if $0 < \hat{p} < 1$.

Now, for $\hat{p} = 0$ we have, $\bar{X} = 0$, i.e. $\sum_{i=1}^{n} X_i = 0 \Leftrightarrow X_i = 0 \ \forall i = 1(1)n$.

$\therefore$ $L_{\underline{x}}(p) = p^0 (1-p)^{n-0} = (1-p)^n \downarrow p$.

$\therefore$ For $\underline{x} = (0, 0, \ldots, 0)$, $L_{\underline{x}}(p)$ will be maximized at $p = 0$. But $0 < p$.

$\therefore$ For $\underline{x} = (0, 0, \ldots, 0)$, $\hat{p}_{MLE}$ does not exist.

Again, for $\hat{p} = 1$, we have, $\bar{x} = 1$, or $\sum_{i=1}^{n} X_i = n \Leftrightarrow X_i = 1, \ \forall i = 1(1)n$.

$\therefore$ $L_{\underline{x}}(p) = p^n \uparrow p$.

$\therefore$ For ~~~~ $\underline{x} = (1, 1, \ldots, 1)$, $L_{\underline{x}}(p)$ will be maximized at $p = 1$. But $p < 1$.

Hence, for $\underline{x} = \underline{1}$, $\hat{p}_{MLE}$ does not exist.

Hence, if $0 < p < 1$, then, $\nexists$ any MLE for p, but if $0 \le p \le 1$, then, $\hat{p}_{MLE} = \bar{x}$, always.

③ MLE may be worthless.

Example: Let, $X \sim Ber(p)$, $\forall \frac{1}{4} \le p \le \frac{3}{4}$. Then, p.m.f. of X is,

$$f_p(x) = p^x (1-p)^{1-x}, \quad x = 0, 1.$$

$\therefore$ The likelihood function of 'p' is, $L_x(p) = \begin{cases} p^x (1-p)^{1-x}, & \frac{1}{4} \le p \le \frac{3}{4} \\ 0, & \text{o.w} \end{cases}$

Now, for given $x=0$, $L_x(p) = p^0 (1-p)^{1-0} = 1-p$.

$\therefore$ For $\frac{1}{4} \le p \le \frac{3}{4}$, $L_x(p) = 1-p$ will be maximized at $p = \frac{1}{4}$.

$\therefore$ For $x=0$, $\hat{p}_{MLE} = \frac{1}{4}$.

Again, for given $x=1$, $L_x(p) = p^1 (1-p)^{1-1} = p$ which will be maximized at $p = \frac{3}{4}$. $\therefore$ For $x=1$, $\hat{p}_{MLE} = \frac{3}{4}$.

Now, for two points on $(x, \hat{p}_{MLE})$ viz. $(0, \frac{1}{4})$ & $(1, \frac{3}{4})$, we can form a linear relation betn $\hat{p}_{MLE}$ & 'x'.

Let it be, $\hat{p}_{MLE} = mx + c \cdots$ ①.

putting $x=0$, $\hat{p}_{MLE} = \frac{1}{4}$ in ①, we have, $c = \frac{1}{4}$.

putting $x=1$, $\hat{p}_{MLE} = \frac{3}{4}$ in ①, we have, $\frac{3}{4} = m + \frac{1}{4}$ $\therefore m = \frac{2}{4}$. $\therefore m = \frac{1}{2}$.

$\therefore \hat{p}_{MLE} = \frac{1}{2}x + \frac{1}{4} = \frac{2x+1}{4}$ ie $\hat{p}_{MLE} = \frac{2x+1}{4} = T$, say.

Now, for $T = \hat{p}_{MLE}$,

$$MSE(T) = E(T-p)^2 = E\left(\frac{2x+1}{4} - p\right)^2 = \sum_{x=0}^{1} \left(\frac{2x+1}{4} - p\right)^2 . f_p(x)$$

$$= \sum_{x=0}^{1} \left(\frac{2x+1}{4} - p\right)^2 . p^x . (1-p)^{1-x} = \left(\frac{1}{4} - p\right)^2 . (1-p) + \left(\frac{3}{4} - p\right)^2 . p$$

$$= \left(\frac{1}{16} + p^2 - \frac{p}{2}\right)(1-p) + \left(\frac{9}{16} + p^2 - \frac{3}{2}p\right)p.$$

$$= \frac{1}{16} + p^2 - \frac{p}{2} - \frac{p}{16} - p^3 + \frac{p^2}{2} + \frac{9p}{16} + p^3 - \frac{3}{2}p^2 = \frac{1}{16}, \forall p \in \left[\frac{1}{4}, \frac{3}{4}\right]$$

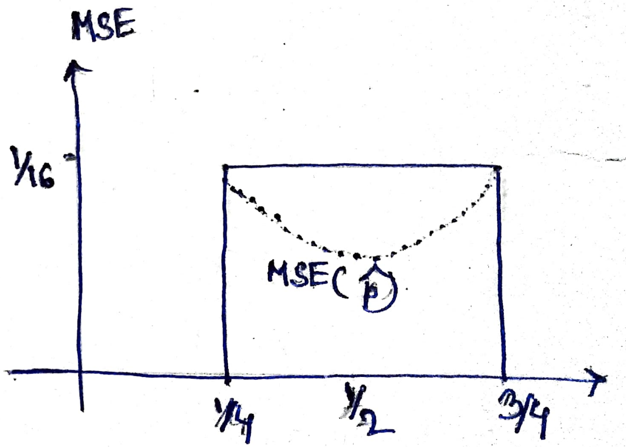
Let us now choose a trivial estimator of p as $\hat{p} = \frac{1}{2}$.

$$\therefore MSE(\hat{p}) = E\left(\frac{1}{2} - p\right)^2 = \left(\frac{1}{2} - p\right)^2$$

Now, $\frac{1}{4} \le p \le \frac{3}{4} \Rightarrow \frac{1}{4} - \frac{1}{2} \le p - \frac{1}{2} \le \frac{3}{4} - \frac{1}{2}$ $\therefore -\frac{1}{4} \le p - \frac{1}{2} \le \frac{1}{4} \Rightarrow (p-\frac{1}{2})^2 \le \frac{1}{16}$

$\Rightarrow (p-\frac{1}{2})^2 \le \frac{1}{16} = MSE(T) \ \forall\, p \in [\frac{1}{4}, \frac{3}{4}]$,

with strict inequality iff $p \neq \frac{1}{4}, \frac{3}{4}$.



Hence, it can be concluded that $MSE(\hat{p})$ is uniformly smaller than $MSE(T)$.

As such, performance of $\hat{p} = \frac{1}{2}$ is uniformly better than $T = \hat{p}_{MLE}$.

But, $\hat{p}$ is a trivial estimator. Hence, $\hat{p}_{MLE}$ is of no use here..

Properties of MLE:

① Suppose, T be a sufficient statistic for θ, where, $T = T(x_1, x_2, \ldots, x_n)$. Then, MLE of θ must be a function of T.

Proof: Since, T is a sufficient statistic for θ, so, by Neyman-Fisher factorizability criterion, we can say that,

$$L_{\underset{\sim}{x}}(\theta) = g_\theta(t) \cdot h(\underset{\sim}{x}),$$

where, 't' is a realization of T. & $h(\underset{\sim}{x})$ is independent of θ.

$$\therefore \ln L_{\underset{\sim}{x}}(\theta) = \ln g_\theta(t) + \ln h(\underset{\sim}{x}).$$

$$\therefore \frac{\partial \ln L_x(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \ln g_\theta(t) \cancel{+0}.$$

Hence, the likelihood eq^n will be $\frac{\partial}{\partial \theta} \ln L_{\underset{\sim}{x}}(\theta) = 0$, ie $\frac{\partial}{\partial \theta} \ln g_\theta(t) = 0$

~~Here~~ $\therefore$ The sol^n of the above eq^n will be a function of 't' only.

Hence, $\hat{\theta}_{MLE}$ is a function of T. [Proved].

② <u>Assymptotic property</u>: Suppose, $\hat{\theta}_{MLE}$ is the consistent maximum likelihood estimator of θ. Then, under certain ~~regularity~~ regularity conditions, it can be shown that,

$$\sqrt{n}\,(\hat{\theta}_{MLE} - \theta) \overset{a}{\sim} N\left[0, \frac{1}{E\left[\frac{\partial}{\partial\theta}\ln f_\theta(X)\right]^2}\right],$$

where, f is the common p.m.f. or p.d.f. of sample observations.

[C.U. 2000]

③ If for a parent population represented by the p.m.f. or p.d.f. f_θ an MVB estimator T of θ exists, then $\hat{\theta}_{MLE}$ will correspond to the estimate 't'.

④ If T is an MLE of θ, & $\psi(\theta)$ is a one-to-one function of θ, then $\psi(T)$ is the MLE of $\psi(\theta)$.

⑤ If $L(\theta)$ is differentiable in an interval including the true value of θ, then with probability unity, as $n\to\infty$, ~~MLE~~ $\hat{\theta}_{MLE}$ exists which is consistent for θ.

⑥. Under regularity conditions, any consistent estimator of θ provides ~~a~~ the maximum likelihood with probability tending to unity as the sample size tends to infinity. (*)

(*) (7) Invariance property of MLE: Let, $\hat{\theta}_{MLE}$ be the maximum likelihood estimator of θ in the density $f(x;\theta)$, where, θ is assumed to be ~~unidimens~~ unidimensional. If $T(\cdot)$ is a function with a single value inverse, ~~of θ~~, then, the MLE of $T(\theta)$ is, $T(\hat{\theta})$..

12 **Example:** We know that in the normal density with known μ_0, the MLE of σ^2 is, $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2$.

Hence, By the invariance property of maximum likelihood estimators we have,

i) the MLE of σ is, $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2}$;

& ii) the MLE of $\log_e \sigma^2$ is, $\ln\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2\right]$.

<u>Note:</u> Following Zehna, the invariance property of MLE can be extended in two dimensions, viz,

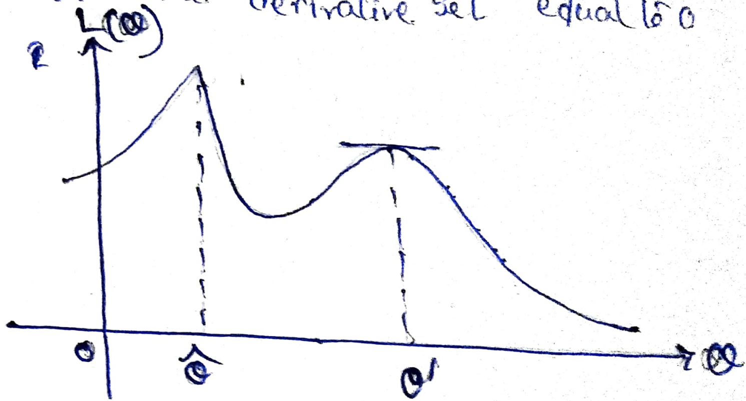
i) θ will be taken as k-dimensional rather than unidimentitional &

ii) the assumption that $T(\cdot)$ has a single valued inverse will be removed.

2) Unreliability of differentiation process to locate maximum.

It can be noted that differentiation process is not always ~~rely on~~ reliable to locate maximum.

The function $L(\alpha)$ may, for example be represented by the following curve where, the actual maxima is at $\hat{\alpha}$ but the derivative set equal to 0 would locate α' as the maxima.



It is also to be remembered that the equation $\frac{\partial L}{\partial \alpha} = 0$ locates minima as well as maxima, & hence, one must avoid using a root of the equation which actually locates a minima.

Q.1) Get the MLE of the parameter θ for random samples from rectangular distributions of the following types:

(a) $f_\theta(x) = \begin{cases} \frac{1}{\theta}, & \text{if } -\theta \le x \le 0 \\ 0, & \text{o.w.} \end{cases}$

Ans/ Since, $X_i \overset{iid}{\sim} R(-\theta, 0)$, $\forall\, i = 1(1)n$, so, their common p.d.f. is

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}, & -\theta \le x \le 0. \\ 0, & \text{o.w.} \end{cases}$$

Hence, the likelihood function of θ is

$$L_x(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \begin{cases} \frac{1}{\theta^n}, & \text{if } -\theta \le x_1, x_2, \dots, x_n \le 0. \\ 0, & \text{o.w.} \end{cases}$$

This can be written as, $L_x(\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{if } -\theta \le x_{(1)} \le 0 \\ 0, & \text{o.w} \end{cases}$ (*)

Now, $\forall\,\theta$, we have, $-\theta \le x_{(1)} \Leftrightarrow -\frac{1}{\theta} \ge \frac{1}{x_{(1)}} \Leftrightarrow \frac{1}{\theta} \le \frac{-1}{x_{(1)}}$.

$\Leftrightarrow \frac{1}{\theta^n} \le -\frac{1}{\{x_{(1)}\}^n} \therefore \Leftrightarrow L_x(\theta) \le - L_x(x_{(1)})$.

Hence by definition of MLE, we have, $\hat{\theta}_{MLE} = -x_{(1)} = -\min\{x_1, x_2, \dots, x_n\}$.

(*) r.a.y This can be also written as, $L_{\underline{x}}(\theta) = \frac{1}{\theta^n}$, $|x_{(1)}| \le |x_{(2)}| \le \dots \le |x_{(n)}| \le \theta$.

Now, we have, $\forall \theta$, ~~as~~ $|x_{(n)}| \le \theta \Rightarrow \frac{1}{\theta} \le \frac{1}{|x_{(n)}|}$

$\Rightarrow \frac{1}{\theta^n} \le \frac{1}{\{|x_{(n)}|\}^n}$. $\Leftrightarrow L_{\underline{x}}(\theta) \le L_{\underline{x}}(|x_{(n)}|)$.

Hence, by defn of MLE, we have, $\hat{\theta}_{MLE} = |X_{(n)}| = |\max\{X_1, \dots, X_n\}|$.

(b) $f_\theta(x) = \begin{cases} \frac{1}{2\theta}, & \text{if } -\theta \le x \le \theta \\ 0, & \text{o.w.} \end{cases}$

Ans/ Since, ~~are~~ $X_i \overset{iid}{\sim} R(-\theta, \theta)$ uniform, so, their common p.d.f is

$$f_\theta(x) = \begin{cases} \frac{1}{2\theta}, & -\theta \le x \le \theta \\ 0, & \text{o.w} \end{cases}$$

Hence, the likelihood function of θ is,

$$L_{\underline{x}}(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \begin{cases} \frac{1}{2^n \theta^n}, & \text{if } -\theta \le x_1, x_2, \ldots, x_n \le \theta \\ 0, & \text{o.w} \end{cases}$$

This can be written as, $L_{\underline{x}}(\theta) = \begin{cases} \frac{1}{2^n \theta^n}, & \text{if } -\theta \le x_{(1)} \le x_{(n)} \le \theta \\ 0, & \text{o.w} \end{cases}$

Thus, $L_{\underline{x}}(\theta)$ attains it's maximum provided that,

$-\theta \le x_{(1)}$ & $\theta \geq x_{(n)}$. ~~ie~~ when,

Now, $\forall \theta$, we have $\theta \geq x_{(n)} \Leftrightarrow \frac{1}{2\theta} \le \frac{1}{2x_{(n)}} \Leftrightarrow L_{\underline{x}}(\theta) \le L_{\underline{x}}(x_{(n)})$.

Hence, by definition of MLE, we have $\hat{\theta}_{MLE} = X_{(n)} = \max\{X_1, \ldots, X_n\}$.

(3.4) Let, $X_1, X_2, \dots, X_n$ be a random sample from ~~the~~ a distⁿ with p.m.f

$$f_N(x) = \begin{cases} \frac{1}{N}, & x = 1, 2, \dots, N \\ 0, & \text{o.w.} \end{cases}$$

Obtain the MLE of N.

A/ By question, ~~X_i's~~ iid the common ~~pdf~~ p.m.f. of $X_1, X_2, \dots, X_n$ is,

$$f_N(x) = \begin{cases} \frac{1}{N}, & x = 1, 2, \dots, N \\ 0, & \text{o.w.} \end{cases}$$

Hence, the likelihood function of N is,

$$L_{\underset{\sim}{x}}(N) = \prod_{i=1}^{n} f_N(x_i) = \frac{1}{N^n}, \text{ if } x_{(n)} \leq N \leq \infty$$

Now, $\forall N$, we have, $N \geq x_{(n)} \Leftrightarrow \frac{1}{N} \leq \frac{1}{x_{(n)}} \Leftrightarrow \frac{1}{N^n} \leq \frac{1}{\{x_{(n)}\}^n} \Leftrightarrow L_{\underset{\sim}{x}}(N) \leq L_{\underset{\sim}{x}}(x_{(n)})$, $\forall N$.

Hence, by definition of MLE, we have, $\hat{N}_{MLE} = X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$.

① Let, $X_i \overset{iid}{\sim} P(\lambda)$. Find $\hat{\lambda}_{MLE}$.

A/ Since, $X_i \overset{iid}{\sim} P(\lambda)$, $\forall i = 1(1)n$, so their common p.d.f. is given by

$$f_\lambda(x) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^x}{x!}, & x = 0,1,2,\ldots \\ & \lambda > 0. \\ 0, & o.w. \end{cases}$$

Then, the likelihood function of λ for given $\underline{x}$ is,

$$L_{\underline{x}}(\lambda) = \prod_{i=1}^{n} f_\lambda(x_i) = \prod_{i=1}^{n} \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!} = \begin{cases} \frac{e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i!)}, & x_i = 0,1,2,\ldots \ \forall i = 1(1)n. \ \lambda > 0 \\ 0, & o.w. \end{cases}$$

$$\therefore \ln L_{\underline{x}}(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln \lambda - \sum_{i=1}^{n} \ln x_i!$$

$$\therefore \frac{\partial \ln L_{\underline{x}}(\lambda)}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}.$$

Now, for optimum λ, we have, $\frac{\partial \ln L_{\underline{x}}(\lambda)}{\partial \lambda} = 0 \Rightarrow \frac{\sum_{i=1}^{n} x_i}{\lambda} = n \ \therefore \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$.

$$\therefore \frac{\partial^2 \ln L_{\underline{x}}(\lambda)}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} = -\frac{n\bar{x}}{\bar{x}^2} = -n < 0,$$ $\because$ 'n' being the no. of r.v.'s is always +ve.

Hence, $\lambda = \bar{x}$ maximizes $L_{\underline{x}}(\lambda)$. $\therefore \hat{\lambda}_{MLE} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(3) Let, $X_i \overset{iid}{\sim} f_\theta(x) = \theta(1-\theta)^x$, $x_i = 0,1,2,\ldots$; $\forall i=1(1)m$. Find $\hat{\theta}_{MLE}$.

A/ Since, $X_i \overset{iid}{\sim} f_\theta(x) = \begin{cases} \theta(1-\theta)^x, & x_i = 0,1,2,\ldots, \forall i=1(1)m \\ 0, & o.w \end{cases}$

Then, the likelihood function for given $\underline{x}$ is, $L_{\underline{x}}(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \theta^n (1-\theta)^{\sum_{i=1}^{n} x_i}$

& $0 < \theta < 1$.

$\therefore \ln L_{\underline{x}}(\theta) = n\ln\theta + \sum_{i=1}^{n} x_i \ln(1-\theta)$. $\therefore \dfrac{\partial \ln L_{\underline{x}}(\theta)}{\partial\theta} = \dfrac{n}{\theta} - \dfrac{\sum_{i=1}^{n} x_i}{1-\theta}$.

$\therefore$ For optimum θ, we have, $\dfrac{\partial \ln L_{\underline{x}}(\theta)}{\partial\theta} = 0$.

$\Rightarrow \dfrac{n}{\theta} = \dfrac{\sum_{i=1}^{n} x_i}{1-\theta}$ or, $n - n\theta - \theta\sum_{i=1}^{n} x_i = 0$ or, $n - n\theta\left(1 + \frac{1}{n}\sum_{i=1}^{n} x_i\right) = 0$.

$\Rightarrow n = n\theta(1+\bar{x})$ $\therefore$ $\theta = \dfrac{1}{1+\bar{x}}$

Now $\dfrac{\partial^2 \ln L_{\underline{x}}(\theta)}{\partial\theta^2} = -\dfrac{n}{\theta^2} - \dfrac{\sum_{i=1}^{n} x_i}{(1-\theta)^2} = -n\left[\dfrac{1}{\left(\frac{1}{1+\bar{x}}\right)^2} + \dfrac{\sum_{i=1}^{n} x_i}{\left(1-\frac{1}{1+\bar{x}}\right)^2}\right]$.

$= -n\left[(1+\bar{x})^2 + \dfrac{(1+\bar{x})^2 \cdot n\bar{x}}{\bar{x}^2}\right] = -n\left[(1+\bar{x})^2 + \dfrac{n(1+\bar{x})^2}{\bar{x}}\right] < 0$,

$\because x_i = 0,1,2,\ldots$, i.e. $x_i > 0$ $\therefore \bar{x} > 0$.

Hence, $\hat{\theta} = \dfrac{1}{1+\bar{x}}$ maximizes $L_{\underline{x}}(\theta)$. $\therefore \hat{\theta}_{MLE} = \dfrac{1}{1+\bar{X}}$

(7) Suppose, $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, μ, σ^2 unknown. Find the MLE of both μ & σ^2.

Ans: $\because$ $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ $\forall i = 1(1)n$, so their common p.d.f. is

$$f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \begin{matrix} -\infty < x < \infty \\ -\infty < \mu < \infty \quad \sigma > 0 \end{matrix} \quad ; \theta = (\mu, \sigma^2).$$

Then, the likelihood function of θ is

$$L_x(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma^*}\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^*}}, \text{ where, } \sigma^* = \sigma^2.$$

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

$$\therefore \ln L_x(\theta) = -\frac{n}{2}\ln\sigma^* - \ln(2\pi)^{\frac{n}{2}} - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^*}.$$

$$\therefore \frac{\partial \ln L}{\partial \mu} = -\frac{2}{2}(-1)\sum_{i=1}^{n}\frac{(x_i-\mu)}{\sigma^*} = \frac{\sum_i x_i - n\mu}{\sigma^*} = \frac{n\bar{x} - n\mu}{\sigma^*}.$$

$$\& \frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{n}{\sigma^*}$$

Again, $\frac{\partial \ln L}{\partial \sigma^*} = -\frac{n}{2\sigma^*} + \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^{*2}}$ & $\frac{\partial^2 \ln L}{\partial \sigma^{*2}} = \frac{n}{2\sigma^{*2}} - 2\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^{*3}}.$

& Also, $\frac{\partial^2 \ln L}{\partial \mu \partial \sigma^*} = -\frac{n(\bar{x}-\mu)}{\sigma^{*2}}.$

Now, doing $\frac{\partial}{\partial \mu}\ln L = 0$, we have, $\mu = \bar{x}$ & doing $\frac{\partial}{\partial \sigma^*}\ln L = 0$, we have, $\sigma^* = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = s^2$. Now ~~here~~ extrema exists if

$$\left(\frac{\partial^2}{\partial \mu^2}\ln L\right)\left(\frac{\partial^2 \ln L}{\partial \sigma^{*2}}\right) - \left(\frac{\partial^2 \ln L}{\partial \mu \partial \sigma^*}\right)^2 \Bigg|_{\substack{\mu = \bar{x} \\ \sigma^* = s^2}} > 0.$$

Now, in the present case, $-\frac{n}{s^2} \times \left(\frac{n}{2s^4} - \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{s^6}\right) - 0$ ~~$\times\left(\frac{n}{s^4}\right)$~~

$$= -\frac{n}{s^2}\left(\frac{n}{2s^4} - \frac{ns^2}{s^6}\right) = -\frac{n^2}{s^6}\left(\frac{1}{2}-1\right) = \frac{n^2}{2s^6} > 0.$$

$\therefore$ Here, extrema exists & as $\left.\frac{\partial^2}{\partial\mu^2}\ln L\right|_{\substack{\mu=\bar{x}\\ \sigma^{*}=s^2}} = -\frac{n}{s^2} < 0$

& $\left.\frac{\partial^2}{\partial\sigma^{*2}}\ln L\right|_{\substack{\mu=\bar{x}\\ \sigma^{*}=s^2}} = -\frac{n}{2s^4} < 0$, hence $\mu = \bar{x}$ & $\sigma^2 = \sigma^{*} = s^2$

maximize the likelihood function $L_x(\theta)$.

$\therefore \hat{\mu}_{MLE} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ & $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$.

...... MLE from Binomial Distribution

Let $x_1, x_2, \ldots, x_n$ be the observations of a random sample of size n. Let x_i follow a binomial distribution with parameters m and p, where the sample observations $x_1, x_2, \ldots, x_n$ are independent. Now, we have to find out the MLE of the parameter p where m is supposed to be known from the condition of random experiment.

Since the binomial distribution is a discrete distribution and its *p.m.f* of the sample observations x_i (for $i = 1, 2, \ldots, n$) is given by, $f_i(x_i) = {}^m C_{x_i}\, p^{x_i}\,(1-p)^{m-x_i}$ for $x = 0, 1, 2, \ldots, m$. The joint *p.m.f* is given by, $f(x_1, x_2, \ldots, x_n/p)$ and since the sample observations are independent, the joint *p.m.f* becomes.

$$f(x_1, x_2, \ldots, x_n/p) = \prod_{i=1}^{n} f_i(x_i).$$

For a given set of sample observations the joint *p.m.f* is looked upon as a function of the parameter p and is called the likelihood function of p and is denoted by $L(p)$.

$$\text{Here } L(p) = \prod_{i=1}^{n} f_i(x_i) = \prod_{i=1}^{n} {}^m C_{x_i}\, p^{x_i}\,(1-p)^{m-x_i}, \quad 0 < p < 1.$$

$$\text{Now, } \log L(p) = \sum_{i=1}^{n} x_i \log p + \sum_{i=1}^{n} (m - x_i)\log(1-p) + \text{constant term}.$$

$$= \sum_{i=1}^{n} x_i \log p + nm\log(1-p) - \sum_{i=1}^{n} x_i \log(1-p) + \text{constant term}\ldots\text{(A)}$$

Now, we have to find out the MLE of p.
So, $\log L$ is to be maximized *w.r.t.* p. Let us now maximize $\log L$ *w.r.t.p* and we get,

$$\frac{d \log L}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n} (m - x_i)}{1-p} = 0$$ (follows from first order condition of maximization).

$$\text{or, } \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n} (m - x_i)}{1-p}. \quad \text{or, } \frac{\sum_{i=1}^{n} x_i}{p} = \frac{nm - \sum_{i=1}^{n} x_i}{1-p}$$

$$\text{or, } \sum_{i=1}^{n} x_i - p \sum_{i=1}^{n} x_i = pnm - p \sum_{i=1}^{n} x_i \quad \text{or, } \sum_{i=1}^{n} x_i = pnm$$

$$\therefore p.n.m = \sum_{i=1}^{n} x_i$$

$$\text{or, } p = \hat{p} = \sum_{i=1}^{n} x_i / n.m = \frac{\bar{x}}{m} \quad \therefore \hat{p} = \frac{\bar{x}}{m}.$$

$$\text{Again, } \frac{d^2 \log L}{dp^2} = -\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{nm - \sum_{i=1}^{n} x_i}{(1-p)^2} < 0$$

It can be seen that if $\hat{p} = \frac{\bar{x}}{m}$ is such that $0 < \frac{\bar{x}}{m} < 1$, then $\frac{d^2 \log L}{dp^2} < 0$. Thus, $\hat{p} = \frac{\bar{x}}{m}$ is the MLE of p if $0 < \hat{p} < 1$.