Chapter 3. Statistical Inference – Point Estimation

Problem in statistics:
A random variables X with p.d.f. of the form $f(x, \theta)$ where function f is known but parameter θ is unknown. We want to gain knowledge about θ.
What we have for inference:
There is a random sample $X_1, \ldots, X_n$ from $f(x, \theta)$.

$$\text{Statistical inferences} \begin{cases} \text{Estimation} \begin{cases} \text{Point estimation: } \hat{\theta} = \hat{\theta}(X_1, \ldots, X_n) \\ \text{Interval estimation:} \\ \text{Find statistics } T_1 = t_1(X_1, \ldots, X_n), T_2 = t_2(X_1, \ldots, X_n) \\ \text{ such that } 1 - \alpha = P(T_1 \le \theta \le T_2) \end{cases} \\ \text{Hypothesis testing: } H_0 : \theta = \theta_0 \text{ or } H_0 : \theta \ge \theta_0. \\ \text{ Want to find a rule to decide if we accept or reject } H_0. \end{cases}$$

Def. *We call a statistic* $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ *an estimator of parameter* θ *if it is used to estimate* θ. *If* $X_1 = x_1, \ldots, X_n = x_n$ *are observed, then* $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ *is called an* ***estimate*** *of* θ.

Two problems are concerned in estimation of θ :

(a) How can we evaluate an estimator $\hat{\theta}$ for its use in estimation of θ ?
Need criterion for this estimation.

(b) Are there general rules in deriving estimators ? We will introduce two methods for deriving estimator of θ .

Def. *We call an estimator* θ ***unbiased*** *for* θ *if it satisfies*

$$E_\theta(\hat{\theta}(X_1, \ldots, X_n)) = \theta, \forall \theta.$$

$$\mathrm{E}_\theta(\hat{\theta}(X_1, \ldots, X_n)) = \begin{cases} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \hat{\theta}(x_1, \ldots, x_n) f(x_1, \ldots, x_n, \theta) dx_1 \cdots dx_n \\ \int_{-\infty}^{\infty} \theta^* f_{\hat{\theta}}(\theta^*) d\theta^* \text{ where } \hat{\theta} = \hat{\theta}(X_1, \ldots, X_n) \text{ is a r.v. with pdf } f_{\hat{\theta}}(\theta^*) \end{cases}$$

Def. *If* $E_\theta(\hat{\theta}(X_1, \ldots, X_n)) \neq \theta$ *for some* θ, *we said that* $\hat{\theta}$ *is a* ***biased*** *estimator.*

Example : $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, Suppose that our interest is μ, X_1,
$\mathrm{E}_\mu(X_1) = \mu$, is unbiased for μ,
$\frac{1}{2}(X_1 + X_2), \mathrm{E}(\frac{X_1+X_2}{2}) = \mu$, is unbiased for μ,
$\overline{X}, \mathrm{E}_\mu(\overline{X}) = \mu$, is unbiased for μ,

► $a_n \overset{n\to\infty}{\longrightarrow} a$, if , for $\epsilon > 0$, there exists $N > 0$ such that $|a_n - a| < \epsilon$ if $n \geq N$.

$\{X_n\}$ is a sequence of r.v.'s. How can we define $X_n \longrightarrow X$ as $n \longrightarrow \infty$?

Def. *We say that X_n* ***converges*** *to X, a r.v. or a constant, in probability if for $\epsilon > 0$,*

$$P(|X_n - X| > \epsilon) \longrightarrow 0, \text{ as } n \longrightarrow \infty.$$

In this case, we denote $X_n \overset{P}{\longrightarrow} X$.

Thm.

If $E(X_n) = a$ or $E(X_n) \longrightarrow a$ and $Var(X_n) \longrightarrow 0$, then $X_n \overset{P}{\longrightarrow} a$.

Proof.

$$\begin{aligned}\mathrm{E}[(X_n - a)^2] &= \mathrm{E}[(X_n - \mathrm{E}(X_n) + \mathrm{E}(X_n) - a)^2] \\ &= \mathrm{E}[(X_n - \mathrm{E}(X_n))^2] + \mathrm{E}[(\mathrm{E}(X_n) - a)^2] + 2\mathrm{E}[(X_n - \mathrm{E}(X_n))(\mathrm{E}(X_n) - a)] \\ &= \mathrm{Var}(X_n) + \mathrm{E}((X_n) - a)^2\end{aligned}$$

Chebyshev's Inequality :

$$P(|X_n - X| \geq \epsilon) \leq \frac{\mathrm{E}(X_n - X)^2}{\epsilon^2} \text{ or } P(|X_n - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

For $\epsilon > 0$,

$$\begin{aligned}0 \leq P(|X_n - a| > \epsilon) &= P((X_n - a)^2 > \epsilon^2) \\ &\leq \frac{\mathrm{E}(X_n - a)^2}{\epsilon^2} = \frac{\mathrm{Var}(X_n) + (\mathrm{E}(X_n) - a)^2}{\epsilon^2} \longrightarrow 0 \text{ as } n \longrightarrow \infty.\end{aligned}$$

$$\Rightarrow P(|X_n - a| > \epsilon) \longrightarrow 0, \text{ as } n \longrightarrow \infty. \Rightarrow X_n \overset{P}{\longrightarrow} a.$$

□

Thm. ***Weak Law of Large Numbers(WLLN)***
If $X_1, \ldots, X_n$ is a random sample with mean μ and finite variance σ^2, then $\overline{X} \overset{P}{\longrightarrow} \mu$.

Proof.

$$\mathrm{E}(\overline{X}) = \mu, \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \longrightarrow 0 \text{ as } n \longrightarrow \infty. \Rightarrow \overline{X} \stackrel{P}{\longrightarrow} \mu.$$

□

Def. *We sat that $\hat{\theta}$ is a* ***consistent*** *estimator of θ if $\hat{\theta} \stackrel{P}{\longrightarrow} \theta$.*

Example : $X_1, \ldots, X_n$ is a random sample with mean μ and finite variance σ^2.Is X_1 a consistent estimator of μ ?
E(X_1)=μ, X_1 is unbiased for μ.
Let $\epsilon > 0$,

$$\begin{aligned} P(|X_1 - \mu| > \epsilon) &= 1 - P(|X_1 - \mu| \leq \epsilon) = 1 - P(\mu - \epsilon \leq X_1 \leq \mu + \epsilon) \\ &= 1 - \int_{\mu-\epsilon}^{\mu+\epsilon} f_X(x)dx > 0, \nrightarrow 0 \text{ as } n \longrightarrow \infty. \end{aligned}$$

$\Rightarrow X$ is not a consistent estimator of μ

$$\begin{aligned} &\mathrm{E}(\overline{X}) = \mu, \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \longrightarrow 0 \text{ as } n \longrightarrow \infty. \\ &\Rightarrow \overline{X} \stackrel{P}{\longrightarrow} \mu. \\ &\Rightarrow \overline{X} \text{ is a consistent estimator of } \mu. \end{aligned}$$

► Unbiasedness and consistency are two basic conditions for good estimator.

Moments :
Let X be a random variable having a p.d.f. $f(x, \theta)$, the population k_{th} moment is defined by

$$\mathrm{E}_\theta(X^k) = \begin{cases} \sum\limits_{\text{all } x} x^k f(x, \theta) & \text{, discrete} \\ \int_{-\infty}^{\infty} x^k f(x, \theta)dx & \text{, continuous} \end{cases}$$

The sample k_{th} moment is defined by $\frac{1}{n}\sum\limits_{i=1}^{n} {X_i}^k$.
<u>Note</u> :

$$\mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} {X_i}^k) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}({X_i}^k) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_\theta(X^k) = \mathrm{E}_\theta(X^k)$$

$\Rightarrow$ Sample k_{th} moment is unbiased for population k_{th} moment.

$$\text{Var}(\frac{1}{n}\sum_{i=1}^{n} {X_i}^k) = \frac{1}{n^2}\text{Var}(\sum_{i=1}^{n} {X_i}^k) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}({X_i}^k) = \frac{1}{n}\text{Var}(X^k) \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

$\Rightarrow \frac{1}{n}\sum\limits_{i=1}^{n} {X_i}^k \xrightarrow{P} \text{E}_\theta(X^k)$.

$\Rightarrow \frac{1}{n}\sum\limits_{i=1}^{n} {X_i}^k$ is a consistent estimator of $\text{E}_\theta(X^k)$.

Let $X_1, \ldots, X_n$ be a random sample with mean μ and variance σ^2.The sample variance is defined by $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ Want to show that S^2 is unbiased for σ^2.

$$\text{Var}(X) = \text{E}[(X-\mu)^2] = \text{E}[X^2 - 2\mu X + \mu^2] = \text{E}(X^2) - \mu^2$$

$$\Rightarrow \text{E}(X^2) = \text{Var}(X) + \mu^2 = \text{Var}(X) + (\text{E}(X))^2$$

$$\text{E}(\overline{X}) = \mu, \text{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

$$\begin{aligned}
\text{E}(S^2) &= \text{E}(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2) = \frac{1}{n-1}\text{E}(\sum_{i=1}^{n} X_i^2 - 2\overline{X}\sum_{i=1}^{n} X_i + n\overline{X}^2) \\
&= \frac{1}{n-1}\text{E}(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2) = \frac{1}{n-1}[\sum_{i=1}^{n}\text{E}(X_i^2) - n\text{E}(\overline{X}^2)] \\
&= \frac{1}{n-1}[n\sigma^2 + n\mu^2 - n(\frac{\sigma^2}{n} + \mu^2)] = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2
\end{aligned}$$

$\Rightarrow S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ is unbiased for σ^2.

$$S^2 = \frac{1}{n-1}[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2] = \frac{n}{n-1}[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2] \xrightarrow{P} \text{E}(X^2) - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

$$\begin{pmatrix} X_1, \ldots, X_n \text{ are iid with mean } \mu \text{ and variance } \sigma^2 \\ {X_1}^2, \ldots, {X_n}^2 \text{ are iid r.v.'s with mean } \text{E}(X^2) = \mu^2 + \sigma^2 \\ \text{By WLLN }, \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \xrightarrow{P} \text{E}(X^2) = \mu^2 + \sigma^2 \end{pmatrix}$$

$\Rightarrow s^2 \xrightarrow{P} \sigma^2$

Def. *Let $X_1, \ldots, X_n$ be a random sample from a distribution with p.d.f. $f(x, \theta)$*

(a) If θ is univariate, the method of moment estimator $\hat{\theta}$ solve θ for $\overline{X} = E_\theta(X)$

(b) If $\theta = (\theta_1, \theta_2)$ is bivariate, the method of moment estimator $(\hat{\theta_1}, \hat{\theta_2})$ solves (θ_1, θ_2) for

$$\overline{X} = E_{\theta_1,\theta_2}(X), \frac{1}{n}\sum_{i=1}^{n} {X_i}^2 = E_{\theta_1,\theta_2}(X^2)$$

(c) If $\theta = (\theta_1, \ldots, \theta_k)$ is k-variate, the method of moment estimator $(\hat{\theta_1}, \ldots, \hat{\theta_k})$ solves $\theta_1, \ldots, \theta_k$ for

$$\frac{1}{n}\sum_{i=1}^{n} {X_i}^j = E_{\theta_1,\ldots,\theta_k}(X^j), j = 1, \ldots, k$$

Example :

(a) $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p)
Let $\overline{X} = \mathrm{E}_p(X) = p$
$\Rightarrow$ The method of moment estimator of p is $\hat{p} = \overline{X}$
By WLLN, $\hat{p} = \overline{X} \xrightarrow{P} \mathrm{E}_p(X) = p \Rightarrow \hat{p}$ is consistent for p.
$\mathrm{E}(\hat{p}) = \mathrm{E}(\overline{X}) = \mathrm{E}(X) = p \Rightarrow \hat{p}$ is unbiased for p.

(b) Let $X_1, \ldots, X_n$ be a random sample from Poisson(λ)
Let $\overline{X} = \mathrm{E}_\lambda(X) = \lambda$
$\Rightarrow$ The method of moment estimator of λ is $\hat{\lambda} = \overline{X}$
$\mathrm{E}(\hat{\lambda}) = \mathrm{E}(\overline{X}) = \lambda \Rightarrow \hat{\lambda}$ is unbiased for λ.
$\hat{\lambda} = \overline{X} \xrightarrow{P} \mathrm{E}(X) = \lambda \Rightarrow \hat{\lambda}$ is consistent for λ.

(c) Let $X_1, \ldots, X_n$ be a random sample with mean μ and variance σ^2.
$\theta = (\mu, \sigma^2)$
Let $\overline{X} = \mathrm{E}_{\mu,\sigma^2}(X) = \mu$
$\frac{1}{n}\sum_{i=1}^{n} {X_i}^2 = \mathrm{E}_{\mu,\sigma^2}(X^2) = \sigma^2 + \mu^2$
$\Rightarrow$ Method of moment estimator are $\hat{\mu} = \overline{X}$,

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} {X_i}^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
$\overline{X}$ is unbiased and consistent estimator for μ.
$\mathrm{E}(\hat{\sigma}^2) = \mathrm{E}(\frac{1}{n}\sum(X_i - \overline{X})^2) = \frac{n-1}{n}\mathrm{E}(\frac{1}{n-1}\sum(X_i - \overline{X})^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$
$\Rightarrow \hat{\sigma}^2$ is not unbiased for σ^2
$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} {X_i}^2 - \overline{X}^2 \xrightarrow{p} \mathrm{E}(X^2) - \mu^2 = \sigma^2$
$\Rightarrow \hat{\sigma}^2$ is consistent for σ^2.

Maximum Likelihood Estimator :
Let $X_1, \ldots, X_n$ be a random sample with p.d.f. $f(x, \theta)$.
The joint p.d.f. of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta), x_i \in R, i = 1, \ldots, n$$

Let Θ be the space of possible values of θ. We call Θ the **parameter space**.

Def. *The likelihood function of a random sample is defined as its joint p.d.f. as*

$$L(\theta) = L(\theta, x_1, \ldots, x_n) = f(x_1, \ldots, x_n, \theta), \theta \in \Theta.$$

which is considered as a function of θ.
For $(x_1, \ldots, x_n)$ fixed, the value $L(\theta, x_1, \ldots, x_n)$ is called the likelihood at θ.

Given observation $x_1, \ldots, x_n$, the likelihood $L(\theta, x_1, \ldots, x_n)$ is considered as the probability that $X_1 = x_1, \ldots, X_n = x_n$ occurs when θ is true.

Def. *Let $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ be any value of θ that maximizes $L(\theta, x_1, \ldots, x_n)$. Then we call $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ the* ***maximum likelihood estimator*** *(m.l.e) of θ. When $X_1 = x_1, \ldots, X_n = x_n$ is observed, we call $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ the maximum likelihood estimate of θ.*

Note :

(a) Why m.l.e ?
When $L(\theta_1, x_1, \ldots, x_n) \geq L(\theta_2, x_1, \ldots, x_n)$,
we are more confident to believe $\theta = \theta_1$ than to believe $\theta = \theta_2$

(b) How to derive m.l.e ?

$\frac{\partial \ln x}{\partial x} = \frac{1}{x} > 0 \Rightarrow \ln x$ is $\nearrow$ in x

$\Rightarrow$If $L(\theta_1) \geq L(\theta_2)$, then $\ln L(\theta_1) \geq \ln\ L(\theta_2)$

If $\hat{\theta}$ is the m.l.e., then $L(\hat{\theta}, x_1, \ldots, x_n) = \max_{\theta \in \Theta} L(\theta, x_1, \ldots, x_n)$ and $\ln\ L(\hat{\theta}, x_1, \ldots, x_n) = \max_{\theta \in \Theta} \ln L(\theta, x_1, \ldots, x_n)$

Two cases to solve m.l.e. :

(b.1) $\frac{\partial \ln L(\theta)}{\partial \theta} = 0$

(b.2) $L(\theta)$ is monotone. Solve $\max_{\theta \in \Theta} L(\theta, x_1, \ldots, x_n)$ from monotone property.

Order statistics:

Let $(X_1, \ldots, X_n)$ be a random sample with d.f. F and p.d.f. f.
Let $(Y_1, \ldots, Y_n)$ be a permutation $(X_1, \ldots, X_n)$ such that $Y_1 \leq Y_2 \leq \cdots Y_n$.
Then we call $(Y_1, \ldots, Y_n)$ the **order statistic** of $(X_1, \ldots, X_n)$ where Y_1 is the first (smallest) order statistic, Y_2 is the second order statistic,..., Y_n is the largest order statistic.

If $(X_1, \ldots, X_n)$ are independent, then

$$\begin{aligned} P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) &= \int_{A_n} \cdots \int_{A_1} f(x_1, \ldots, x_n) dx_1 \cdots dx_n \\ &= \int_{A_n} f_n(x_n) dx_n \cdots \int_{A_1} f_1(x_1) dx_1 \\ &= P(X_n \in A_n) \cdots P(X_1 \in A_1) \end{aligned}$$

Thm. *Let $(X_1, \ldots, X_n)$ be a random sample from a "continuous distribution" with p.d.f. $f(x)$ and d.f $F(x)$. Then the p.d.f. of $Y_n = \max\{X_1, \ldots, X_n\}$ is*

$$g_n(y) = n(F(y))^{n-1} f(y)$$

and the p.d.f. of $Y_1 = \min\{X_1, \ldots, X_n\}$ is

$$g_1(y) = n(1 - F(y))^{n-1} f(y)$$

Proof. This is a $R^n \to R$ transformation. Distribution function of Y_n is

$$\begin{aligned} G_n(y) = P(Y_n \leq y) &= P(\max\{X_1, \ldots, X_n\} \leq y) = P(X_1 \leq y, \ldots, X_n \leq y) \\ &= P(X_1 \leq y) P(X_2 \leq y) \cdots P(X_n \leq y) = (F(y))^n \end{aligned}$$

$\Rightarrow$ p.d.f. of Y_n is $g_n(y) = D_y(F(y))^n = n(F(y))^{n-1}f(y)$
Distribution function of Y_1 is

$$\begin{aligned} G_1(y) &= P(Y_1 \leq y) = P(\min\{X_1, \ldots, X_n\} \leq y) = 1 - P(\min\{X_1, \ldots, X_n\} > y) \\ &= 1 - P(X_1 > y, X_2 > y, \ldots, X_n > y) = 1 - P(X_1 > y)P(X_2 > y) \cdots P(X_n > y) \\ &= 1 - (1 - F(y))^n \end{aligned}$$

$\Rightarrow$ p.d.f. of Y_1 is $g_1(y) = D_y(1 - (1 - F(y))^n) = n(1 - F(y))^{n-1}f(y)$

□

Example : Let $(X_1, \ldots, X_n)$ be a random sample from $U(0, \theta)$.
Find m.l.e. of θ. Is it unbiased and consistent ?
sol: The p.d.f. of X is

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \leq x \leq \theta \\ 0 & \text{elsewhere.} \end{cases}$$

Consider the indicator function

$$I_{(a,b)}(x) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{elsewhere.} \end{cases}$$

Then $f(x, \theta) = \frac{1}{\theta} I_{[0,\theta]}(x)$.
The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{[0,\theta]}(x_i)$$

Let $Y_n = \max\{X_1, \ldots, X_n\}$
Then $\prod_{i=1}^{n} I_{[0,\theta]}(x_i) = 1 \Leftrightarrow 0 \leq x_i \leq \theta$, for all $i = 1, \ldots, n \Leftrightarrow 0 \leq y_n \leq \theta$
We then have

$$L(\theta) = \frac{1}{\theta^n} I_{[0,\theta]}(y_n) = \frac{1}{\theta^n} I_{[y_n,\infty]}(\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \geq y_n \\ 0 & \text{if } \theta < y_n \end{cases}$$

$L(\theta)$ is maximized when $\theta = y_n$. Then m.l.e. of θ is $\hat{\theta} = Y_n$
The d.f. of x is

$$F(x) = P(X \leq x) = \int_0^x \frac{1}{\theta} dt = \frac{x}{\theta}, 0 \leq x \leq \theta$$

The p.d.f. of Y is

$$g_n(y) = n(\frac{y}{\theta})^{n-1}\frac{1}{\theta} = n\frac{y^{n-1}}{\theta^n}, 0 \le y \le \theta$$

$E(Y_n) = \int_0^\theta yn\frac{y^{n-1}}{\theta^n}dy = \frac{n}{n+1}\theta \neq \theta \Rightarrow$ m.l.e. $\hat{\theta} = Y_n$ is not unbiased.
However, $E(Y_n) = \frac{n}{n+1}\theta \to \theta$ as $n \to \infty$, m.l.e. $\hat{\theta}$ is asymptotically unbiased.

$$E(Y_n^2) = \int_0^\theta y^2 n\frac{y^{n-1}}{\theta^n}dy = \frac{n}{n+2}\theta^2$$

$\mathrm{Var}(Y_n) = E(Y_n^2) - (EY_n)^2 = \dfrac{n}{n+2}\theta^2 - (\dfrac{n}{n+1})^2\theta^2 \longrightarrow \theta^2 - \theta^2 = 0$ as $n \longrightarrow \infty$.

$\Rightarrow Y_n \xrightarrow{P} \theta \Rightarrow$ m.l.e. $\hat{\theta} = Y_n$ is consistent for θ .
Is there unbiased estimator for θ ?

$$E(\frac{n+1}{n}Y_n) = \frac{n+1}{n}E(Y_n) = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta$$

$\Rightarrow \frac{n+1}{n}Y_n$ is unbiased for θ.
Example :

(a) $Y \sim b(n, p)$
The likelihood function is

$$L(p) = f_Y(y, p) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$\ln L(p) = \ln \binom{n}{y} + y \ln p + (n-y)\ln(1-p)$$

$\dfrac{\partial \ln L(p)}{\partial p} = \dfrac{y}{p} - \dfrac{n-y}{1-p} = 0 \Leftrightarrow \dfrac{y}{p} = \dfrac{n-y}{1-p} \Leftrightarrow y(1-p) = p(n-y) \Leftrightarrow y = np$
$\Rightarrow$ m.l.e. $\hat{p} = \frac{Y}{n}$
$E(\hat{p}) = \frac{1}{n}E(Y) = p \Rightarrow$ m.l.e. $\hat{p} = \frac{Y}{n}$ is unbiased.
$\mathrm{Var}(\hat{p}) = \frac{1}{n^2}\mathrm{Var}(Y) = \frac{1}{n}p(1-p) \longrightarrow 0$ as $n \longrightarrow \infty$
$\Rightarrow$ m.l.e. $\hat{p} = \frac{Y}{n}$ is consistent for p.

(b) $X_1, \ldots, X_n$ are a random sample from $N(\mu, \sigma^2)$. Want m.l.e.'s of μ and σ^2
The likelihood function is

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

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

$$\frac{\partial \ln\, L(\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \Rightarrow \hat{\mu} = \overline{X}$$

$$\frac{\partial \ln\, L(\hat{\mu},\sigma^2)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\overline{x})^2 = 0 \Rightarrow \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

$\mathrm{E}(\hat{\mu}) = \mathrm{E}(\overline{X}) = \mu$ (unbiased),$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \longrightarrow 0$ as $n \longrightarrow \infty$
$\Rightarrow$ m.l.e. $\hat{\mu}$ is consistent for μ.
$\mathrm{E}(\hat{\sigma}^2) = \mathrm{E}(\frac{1}{n}\sum(X_i-\overline{X})^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$ (biased).
$\mathrm{E}(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \longrightarrow \sigma^2$ as $n \longrightarrow \infty \Rightarrow \hat{\sigma}^2$ is asymptotically unbiased.

$$\begin{aligned}\mathrm{Var}(\hat{\sigma}^2) &= \mathrm{Var}(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2) = \frac{1}{n^2}\mathrm{Var}(\sigma^2\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sigma^2}) \\ &= \frac{\sigma^4}{n^2}\mathrm{Var}(\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sigma^2}) = \frac{2(n-1)}{n^2}\sigma^4 \longrightarrow 0 \text{ as } n \longrightarrow \infty\end{aligned}$$

$\Rightarrow$ m.l.e. $\hat{\sigma^2}$ is consistent for σ^2.

Suppose that we have m.l.e. $\hat{\theta} = \hat{\theta}(x_1,\ldots,x_n)$ for parameter θ and our interest is a new parameter $\tau(\theta)$, a function of θ.
What is the m.l.e. of $\tau(\theta)$?
The space of $\tau(\theta)$ is $T = \{\tau : \exists\theta \in \Theta \text{ s.t } \tau = \tau(\theta)\}$

Thm. *If $\hat{\theta} = \hat{\theta}(x_1,\ldots,x_n)$ is the m.l.e. of θ and $\tau(\theta)$ is a 1-1 function of θ, then m.l.e. of $\tau(\theta)$ is $\tau(\hat{\theta})$*

Proof. The likelihood function for θ is $L(\theta,x_1,\ldots,x_n)$. Then the likelihood function for $\tau(\theta)$ can be derived as follows :

$$\begin{aligned}L(\theta,x_1,\ldots,x_n) &= L(\tau^{-1}(\tau(\theta)),x_1,\ldots,x_n) \\ &= M(\tau(\theta),x_1,\ldots,x_n) \\ &= M(\tau,x_1,\ldots,x_n), \tau \in T\end{aligned}$$

$$
\begin{aligned}
M(\tau(\hat\theta), x_1, \dots, x_n) &= L(\tau^{-1}(\tau(\hat\theta), x_1, \dots, x_n) \\
&= L(\hat\theta, x_1, \dots, x_n) \\
&\ge L(\theta, x_1, \dots, x_n), \forall \theta \in \Theta \\
&= L(\tau^{-1}(\tau(\theta)), x_1, \dots, x_n) \\
&= M(\tau(\theta), x_1, \dots, x_n), \forall \theta \in \Theta \\
&= M(\tau, x_1, \dots, x_n), \tau \in T
\end{aligned}
$$

$\Rightarrow \tau(\hat\theta)$ is m.l.e. of $\tau(\theta)$.
This is the invariance property of m.l.e. □

Example :

(1)If $Y \sim b(n, p)$, m.l.e of p is $\hat p = \frac{Y}{n}$

$\tau(p)$	m.l.e of $\tau(p)$
p^2	$\widehat{p^2} = (\frac{Y}{n})^2$
$\sqrt{p}$	$\widehat{\sqrt{p}} = \sqrt{\frac{Y}{n}}$
e^p	$\widehat{e^p} = e^{\frac{Y}{n}}$
e^{-p}	$\widehat{e^{-p}} = e^{-\frac{Y}{n}}$

$p(1-p)$ is not a 1-1 function of p.

(2) $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, m.l.e.'s of (μ, σ^2) is $(\overline{X}, \frac{1}{n}\sum(X_i - \overline{X})^2)$.
m.l.e.'s of (μ, σ) is $(\overline{X}, \sqrt{\frac{1}{n}\sum(X_i - \overline{X})^2})$ $(\because \sigma \in (0, \infty) \therefore \sigma^2 \longrightarrow \sigma$ is 1-1)
You can also solve

$$
\begin{aligned}
\frac{\partial \ln\, L(\mu, \sigma^2, x_1, \dots, x_n)}{\partial \mu} &= 0 \\
\frac{\partial \ln\, L(\mu, \sigma^2, x_1, \dots, x_n)}{\partial \sigma} &= 0 \text{ for } \mu, \sigma
\end{aligned}
$$

(μ^2, σ) is not a 1-1 function of (μ, σ^2).
$(\because \mu \in (-\infty, \infty) \therefore \mu \longrightarrow \mu^2$ isn't 1-1)

Best estimator :

Def. *An unbiased estimator $\hat\theta = \hat\theta(X_1, \dots, X_n)$ is called a uniformly minimum variance unbiased estimator (UMVUE) or best estimator if for any unbiased estimator $\hat{\theta^*}$,we have*

$$Var_\theta\, \hat\theta \le Var_\theta\, \hat{\theta^*}, \text{ for } \theta \in \Theta$$

($\hat\theta$ is uniformly better than $\hat{\theta^*}$ in variance.)

There are several ways in deriving UMVUE of θ.
Cramer-Rao lower bound for variance of unbiased estimator :
Regularity conditions :

(a) Parameter space Θ is an open interval. $(a, \infty), (a, b), (b, \infty)$, a,b are constants not depending on θ.

(b) Set $\{x : f(x, \theta) = 0\}$ is independent of θ.

(c) $\int \frac{\partial f(x,\theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int f(x, \theta) dx = 0$

(d) If $T = t(x_1, \ldots, x_n)$ is an unbiased estimator, then

$$\int t \frac{\partial f(x, \theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int t f(x, \theta) dx$$

Thm. ***Cramer-Rao (C-R)***
Suppose that the regularity conditions hold.
If $\hat{\tau}(\theta) = t(X_1, \ldots, X_n)$ *is unbiased for* $\tau(\theta)$*, then*

$$Var_\theta\, \hat{\tau}(\theta) \geq \frac{(\tau'(\theta))^2}{nE_\theta\, [(\frac{\partial \ln\, f(x,\theta)}{\partial \theta})^2]} = \frac{(\tau'(\theta))^2}{-nE_\theta\, [(\frac{\partial^2 \ln\, f(x,\theta)}{\partial \theta^2})]} \quad for\ \theta \in \Theta$$

Proof. Consider only the continuous distribution.

$$\begin{aligned} \mathrm{E}[\frac{\partial \ln\, f(x, \theta)}{\partial \theta}] &= \int_{-\infty}^{\infty} \frac{\partial \ln\, f(x, \theta)}{\partial \theta} f(x, \theta) dx = \int_{-\infty}^{\infty} \frac{\partial f(x, \theta)}{\partial \theta} dx \\ &= \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(x, \theta) dx = 0 \end{aligned}$$

$$\tau(\theta) = \mathrm{E}_\theta \hat{\tau}(\theta) = \mathrm{E}_\theta (t(x_1, \ldots, x_n)) = \int \cdots \int t(x_1, \ldots, x_n) \prod_{i=1}^n f(x_i, \theta) \prod_{i=1}^n dx_i$$

Taking derivatives both sides.

$$\begin{aligned} \tau'(\theta) &= \frac{\partial}{\partial \theta} \int \cdots \int t(x_1, \ldots, x_n) \prod_{i=1}^n f(x_i, \theta) \prod_{i=1}^n dx_i - \tau(\theta) \frac{\partial}{\partial \theta} \int \cdots \int \prod_{i=1}^n f(x_i, \theta) \prod_{i=1}^n dx_i \\ &= \int \cdots \int t(x_1, \ldots, x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i, \theta) \prod_{i=1}^n dx_i - \int \cdots \int \tau(\theta) \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i, \theta) \prod_{i=1}^n dx_i \\ &= \int \cdots \int (t(x_1, \ldots, x_n) - \tau(\theta)) (\frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i, \theta)) \prod_{i=1}^n dx_i \end{aligned}$$

Now,

$$\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i,\theta) = \frac{\partial}{\partial\theta}[f(x_1,\theta)f(x_2,\theta)\cdots f(x_n,\theta)]$$

$$= (\frac{\partial}{\partial\theta}f(x_1,\theta))\prod_{i\neq 1} f(x_i,\theta) + \cdots + (\frac{\partial}{\partial\theta}f(x_n,\theta))\prod_{i\neq n} f(x_i,\theta)$$

$$= \sum_{j=1}^{n}\frac{\partial}{\partial\theta}f(x_j,\theta)\prod_{i\neq j} f(x_i,\theta)$$

$$= \sum_{j=1}^{n}\frac{\partial \ln\ f(x_j,\theta)}{\partial\theta}f(x_j,\theta)\prod_{i\neq j} f(x_i,\theta)$$

$$= \sum_{j=1}^{n}\frac{\partial \ln\ f(x_j,\theta)}{\partial\theta}\prod_{j=1}^{n} f(x_i,\theta)$$

Cauchy-Swartz Inequality

$$[\mathrm{E}(XY)]^2 \leq \mathrm{E}(X^2)\mathrm{E}(Y^2)$$

Then

$$\tau'(\theta) = \int\cdots\int(t(x_1,\ldots,x_n)-\tau(\theta))(\sum_{j=1}^{n}\frac{\partial \ln\ f(x_j,\theta)}{\partial\theta})\prod_{i=1}^{n} f(x_i,\theta)\prod_{i=1}^{n} dx_i$$

$$= \mathrm{E}[(t(x_1,\ldots,x_n)-\tau(\theta))\sum_{j=1}^{n}\frac{\partial \ln\ f(x_j,\theta)}{\partial\theta}]$$

$$(\tau'(\theta))^2 \leq \mathrm{E}[(t(x_1,\ldots,x_n)-\tau(\theta))^2]\,\mathrm{E}[(\sum_{j=1}^{n}\frac{\partial \ln\ f(x_j,\theta)}{\partial\theta})^2]$$

$$\Rightarrow \mathrm{Var}(\hat{\tau}(\theta)) \geq \frac{(\tau'(\theta))^2}{\mathrm{E}[(\sum_{j=1}^{n}\frac{\partial \ln f(x_j,\theta)}{\partial\theta})^2]}$$

Since

$$\mathrm{E}[(\sum_{j=1}^{n}\frac{\partial \ln f(x_j,\theta)}{\partial\theta})^2] = \sum_{j=1}^{n}\mathrm{E}(\frac{\partial \ln f(x_j,\theta)}{\partial\theta})^2 + \sum_{i\neq j}\mathrm{E}(\frac{\partial \ln f(x_j,\theta)}{\partial\theta}\frac{\partial \ln f(x_i,\theta)}{\partial\theta})$$

$$= \sum_{j=1}^{n}\mathrm{E}(\frac{\partial \ln f(x_j,\theta)}{\partial\theta})^2$$

$$= n\,\mathrm{E}(\frac{\partial \ln f(x_j,\theta)}{\partial\theta})^2$$

Then, we have

$$\operatorname{Var}_\theta \hat{\tau}(\theta) \geq \frac{(\tau'(\theta))^2}{n\mathrm{E}_\theta\left[\left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2\right]}$$

You may further check that

$$\mathrm{E}_\theta\left(\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2}\right) = -\mathrm{E}_\theta\left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2$$

□

Thm. *If there is an unbiased estimator $\hat{\tau}(\theta)$ with variance achieving the Cramer-Rao lower bound $\frac{(\tau'(\theta))^2}{-nE_\theta\left[\left(\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2}\right)\right]}$,then $\hat{\tau}(\theta)$ is a UMVUE of $\tau(\theta)$.*

<u>Note</u>:
If $\tau(\theta) = \theta$, then any unbiased estimator $\hat{\theta}$ satisfies

$$\operatorname{Var}_\theta(\hat{\theta}) \geq \frac{(\tau'(\theta))^2}{-n\mathrm{E}_\theta\left(\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2}\right)}$$

Example:

(a) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda), \mathrm{E}(X) = \lambda, \operatorname{Var}(X) = \lambda$.
MLE $\hat{\lambda} = \overline{X}, \mathrm{E}(\hat{\lambda}) = \lambda, \operatorname{Var}(\hat{\lambda}) = \frac{\lambda}{n}$.
p.d.f. $f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, \ldots$

$$\Rightarrow \ln f(x,\lambda) = x\ln\lambda - \lambda - \ln x!$$
$$\Rightarrow \frac{\partial}{\partial \lambda}\ln f(x,\lambda) = \frac{x}{\lambda} - 1$$
$$\Rightarrow \frac{\partial^2}{\partial \lambda^2}\ln f(x,\lambda) = -\frac{x}{\lambda^2}$$
$$\mathrm{E}\left(\frac{\partial^2}{\partial \lambda^2}\ln f(x,\lambda)\right) = \mathrm{E}\left(-\frac{x}{\lambda^2}\right) = -\frac{\mathrm{E}(X)}{\lambda^2} = -\frac{1}{\lambda}$$

Cramer-Rao lower bound is

$$\frac{1}{-n(-\frac{1}{\lambda})} = \frac{\lambda}{n} = \operatorname{Var}(\hat{\lambda})$$

$\Rightarrow$ MLE $\hat{\lambda} = \overline{X}$ is the UMVUE of λ.

(b) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p), \mathrm{E}(X) = p, \mathrm{Var}(X) = p(1-p)$.

Want UMVUE of p.

p.d.f $f(x,p) = p^x(1-p)^{1-x}$

$$\Rightarrow \ln f(x,p) = x \ln p + (1-x)\ln(1-p)$$

$$\frac{\partial}{\partial p} \ln f(x,p) = \frac{x}{p} - \frac{1-x}{1-p}$$

$$\frac{\partial^2}{\partial p^2} \ln f(x,p) = -\frac{x}{p^2} + \frac{1-x}{(1-p)^2}$$

$$\mathrm{E}(\frac{\partial^2}{\partial p^2} \ln f(X,p)) = \mathrm{E}(-\frac{X}{p^2} + \frac{1-X}{(1-p)^2}) = -\frac{1}{p} + \frac{1}{1-p} = -\frac{1}{p(1-p)}$$

C-R lower bound for p is

$$\frac{1}{-n(-\frac{1}{p(1-p)})} = \frac{p(1-p)}{n}$$

m.l.e. of p is $\hat{p} = \overline{X}$

$\mathrm{E}(\hat{p}) = \mathrm{E}(\overline{X}) = p, \mathrm{Var}(\hat{p}) = \mathrm{Var}(\overline{X}) = \frac{p(1-p)}{n} =$ C-R lower bound.

$\Rightarrow$ MLE $\hat{p}$ is the UMVUE of p.