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 $\theta$ is unknown. We want to gain knowledge about $\theta$.

What we have for inference:
There is a random sample $X_1, \ldots, X_n$ from $f(x, \theta)$.

Statistical inferences

- **Point estimation**: $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$
- **Interval estimation**: Find statistics $T_1 = t_1(X_1, \ldots, X_n), T_2 = t_2(X_1, \ldots, X_n)$ such that $1 - \alpha = P(T_1 \leq \theta \leq T_2)$
- **Hypothesis testing**: $H_0 : \theta = \theta_0$ or $H_0 : \theta \geq \theta_0$.

Want to find a rule to decide if we accept or reject $H_0$.

**Def.** We call a statistic $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ an estimator of parameter $\theta$ if it is used to estimate $\theta$. 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 $\theta$.

Two problems are concerned in estimation of $\theta$:

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

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

**Def.** We call an estimator $\hat{\theta}$ unbiased for $\theta$ if it satisfies

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

$$E_\theta(\hat{\theta}(X_1, \ldots, X_n)) = \left\{ \begin{array}{ll}
\int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \hat{\theta}(x_1, \ldots, x_n) f(x_1, \ldots, x_n, \theta) dx_1 \cdots dx_n \\
\int_{-\infty}^\infty \theta^* f_{\hat{\theta}}(\theta^*) d\theta^* \end{array} \right.$$

where $\hat{\theta}(X_1, \ldots, X_n)$ is a r.v. with pdf $f_{\hat{\theta}}(\theta^*)$.

**Def.** If $E_\theta(\hat{\theta}(X_1, \ldots, X_n)) \neq \theta$ for some $\theta$, we said that $\hat{\theta}$ is a biased estimator.
Example: \( X_1, \ldots, X_n \sim_{\text{iid}} N(\mu, \sigma^2) \), Suppose that our interest is \( \mu, X_1 \),

\[ E_{\mu}(X_1) = \mu, \text{ is unbiased for } \mu, \]
\[ \frac{1}{2}(X_1 + X_2), E\left(\frac{X_1 + X_2}{2}\right) = \mu, \text{ is unbiased for } \mu, \]
\[ \bar{X}, E_{\mu}(\bar{X}) = \mu, \text{ is unbiased for } \mu, \]

\[ \Rightarrow a_n \xrightarrow{n \to \infty} a, \text{ if } \epsilon > 0, \text{ there exists } \exists N > 0 \text{ such that } |a_n - a| < \epsilon \text{ if } n \geq N. \]

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

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

\[ P(|X_n - X| > \epsilon) \to 0, \text{ as } n \to \infty. \]

In this case, we denote \( X_n \xrightarrow{P} X \).

**Thm.**

If \( E(X_n) = a \) or \( E(X_n) \to a \) and \( \text{Var}(X_n) \to 0 \), then \( X_n \xrightarrow{P} a \).

**Proof.**

\[
E([X_n - a]^2) = E[(X_n - E(X_n) + E(X_n) - a)^2] \\
= E[(X_n - E(X_n))^2] + E[(E(X_n) - a)^2] + 2E[(X_n - E(X_n))(E(X_n) - a)] \\
= \text{Var}(X_n) + E((X_n) - a)^2
\]

**Chebyshev’s Inequality :**

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

For \( \epsilon > 0 \),

\[
0 \leq P(|X_n - a| > \epsilon) = P((X_n - a)^2 > \epsilon^2) \\
\leq \frac{E(X_n - a)^2}{\epsilon^2} = \frac{\text{Var}(X_n) + (E(X_n) - a)^2}{\epsilon^2} \to 0 \text{ as } n \to \infty.
\]

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

**Thm. Weak Law of Large Numbers (WLLN)**

If \( X_1, \ldots, X_n \) is a random sample with mean \( \mu \) and finite variance \( \sigma^2 \), then \( \bar{X} \xrightarrow{P} \mu \).
Proof.

\[ E(\bar{X}) = \mu, \ Var(\bar{X}) = \frac{\sigma^2}{n} \rightarrow 0 \text{ as } n \rightarrow \infty. \Rightarrow \bar{X} \xrightarrow{P} \mu. \]

\[ \square \]

**Def.** We sat that \( \hat{\theta} \) is a **consistent** estimator of \( \theta \) if \( \hat{\theta} \xrightarrow{P} \theta \).

**Example:** \( X_1, \ldots, X_n \) is a random sample with mean \( \mu \) and finite variance \( \sigma^2 \). Is \( X_1 \) a consistent estimator of \( \mu \)?

\[ E(X_1) = \mu, \] \( X_1 \) is unbiased for \( \mu \).

Let \( \epsilon > 0 \),

\[ P(|X_1 - \mu| > \epsilon) = 1 - P(|X_1 - \mu| \leq \epsilon) = 1 - P(\mu - \epsilon \leq X_1 \leq \mu + \epsilon) \]
\[ = 1 - \int_{\mu-\epsilon}^{\mu+\epsilon} f_X(x) \, dx > 0, \rightarrow 0 \text{ as } n \rightarrow \infty. \]

\( \Rightarrow X \) is not a consistent estimator of \( \mu \)

\[ E(\bar{X}) = \mu, \ Var(\bar{X}) = \frac{\sigma^2}{n} \rightarrow 0 \text{ as } n \rightarrow \infty. \]
\[ \Rightarrow \bar{X} \xrightarrow{P} \mu. \]
\[ \Rightarrow \bar{X} \text{ is a consistent estimator of } \mu. \]

\( \blacktriangleright \) 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

\[ E_{\theta}(X^k) = \begin{cases} \sum_{\text{all } x} x^k f(x, \theta), & \text{discrete} \\ \int_{-\infty}^{\infty} x^k f(x, \theta) \, dx, & \text{continuous} \end{cases} \]

The sample \( k_{th} \) moment is defined by \( \frac{1}{n} \sum_{i=1}^{n} X_i^k \).

**Note:**

\[ E\left( \frac{1}{n} \sum_{i=1}^{n} X_i^k \right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i^k) = \frac{1}{n} \sum_{i=1}^{n} E_{\theta}(X^k) = E_{\theta}(X^k) \]
Sample $k_{th}$ moment is unbiased for population $k_{th}$ moment.

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

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

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

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

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

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

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

$$\mathbb{E}(S^2) = \mathbb{E}(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2) = \frac{1}{n-1} \mathbb{E}(\sum_{i=1}^{n} X_i^2 - 2\bar{X} \sum_{i=1}^{n} X_i + n\bar{X}^2)$$

$$= \frac{1}{n-1} \mathbb{E}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2) = \frac{1}{n-1} \left[ \sum_{i=1}^{n} \mathbb{E}(X_i^2) - n\mathbb{E}(\bar{X}^2) \right]$$

$$= \frac{1}{n-1} \left[ n\sigma^2 + n\mu^2 - n\left( \frac{\sigma^2}{n} + \mu^2 \right) \right] = \frac{1}{n-1} (n-1)\sigma^2 = \sigma^2$$

$$\Rightarrow S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \text{ is unbiased for } \sigma^2.$$

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

$$X_1, \ldots, X_n \text{ are iid with mean } \mu \text{ and variance } \sigma^2$$

$$X_1^2, \ldots, X_n^2 \text{ are iid r.v.'s with mean } \mathbb{E}(X^2) = \mu^2 + \sigma^2$$

By WLLN, $\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mathbb{E}(X^2) = \mu^2 + \sigma^2$

$$\Rightarrow s^2 \xrightarrow{P} \sigma^2$$
Def. Let $X_1, \ldots, X_n$ be a random sample from a distribution with p.d.f. $f(x, \theta)$

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

(b) If $\theta = (\theta_1, \theta_2)$ is bivariate, the method of moment estimator $(\hat{\theta}_1, \hat{\theta}_2)$ solves $(\theta_1, \theta_2)$ for

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

(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) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$

Let $\bar{X} = E_p(X) = p$  
⇒ The method of moment estimator of $p$ is $\hat{p} = \bar{X}$

By WLLN, $\hat{p} = \bar{X} \overset{P}{\rightarrow} E_p(X) = p$ ⇒ $\hat{p}$ is consistent for $p$.

$E(\hat{p}) = E(\bar{X}) = E(X) = p$ ⇒ $\hat{p}$ is unbiased for $p$.

(b) Let $X_1, \ldots, X_n$ be a random sample from Poisson($\lambda$)

Let $\bar{X} = E_\lambda(X) = \lambda$  
⇒ The method of moment estimator of $\lambda$ is $\hat{\lambda} = \bar{X}$

$E(\hat{\lambda}) = E(\bar{X}) = \lambda$ ⇒ $\hat{\lambda}$ is unbiased for $\lambda$.

$\hat{\lambda} = \bar{X} \overset{P}{\rightarrow} E(X) = \lambda$ ⇒ $\hat{\lambda}$ is consistent for $\lambda$.

(c) Let $X_1, \ldots, X_n$ be a random sample with mean $\mu$ and variance $\sigma^2$.

$\theta = (\mu, \sigma^2)$  
Let $\bar{X} = E_{\mu, \sigma^2}(X) = \mu$

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 = E_{\mu, \sigma^2}(X^2) = \sigma^2 + \mu^2$$

⇒ Method of moment estimator are $\hat{\mu} = \bar{X}$.
\[ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2. \]

\( \overline{X} \) is unbiased and consistent estimator for \( \mu \).

\[
E(\hat{\sigma}^2) = E(\frac{1}{n} \sum (X_i - \overline{X})^2) = \frac{n-1}{n} E(\frac{1}{n-1} \sum (X_i - \overline{X})^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2
\]

\[ \Rightarrow \hat{\sigma}^2 \text{ is not unbiased for } \sigma^2. \]

\[
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2 \rightarrow E(X^2) - \mu^2 = \sigma^2
\]

\[ \Rightarrow \hat{\sigma}^2 \text{ is consistent for } \sigma^2. \]

**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 \( \Theta \) be the space of possible values of \( \theta \). We call \( \Theta \) 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 \( \theta \).
For \( (x_1, \ldots, x_n) \) fixed, the value \( L(\theta, x_1, \ldots, x_n) \) is called the likelihood at \( \theta \).

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 \( \theta \) is true.

**Def.** Let \( \hat{\theta} = \hat{\theta}(x_1, \ldots, x_n) \) be any value of \( \theta \) 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 \( \theta \). 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 \( \theta \).

**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.?
\[ \frac{\partial \ln x}{\partial x} = \frac{1}{x} > 0 \Rightarrow \ln x \text{ is } \nearrow \text{ in } x \]
⇒ 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 \( Y_1 \leq Y_2 \leq \cdots \leq Y_n \).
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, \ldots, \( Y_n \) is the largest order statistic.

If \((X_1, \ldots, X_n)\) are independent, then
\[
P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \int_{A_n} \cdots \int_{A_1} f(x_1, \ldots, x_n) dx_1 \cdots dx_n
\]
\[
= \int_{A_n} f_n(x_n) dx_n \cdots \int_{A_1} f_1(x_1) dx_1
\]
\[
= P(X_n \in A_n) \cdots P(X_1 \in A_1)
\]

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
\[
G_n(y) = P(Y_n \leq y) = P(\max\{X_1, \ldots, X_n\} \leq y) = P(X_1 \leq y, \ldots, X_n \leq y)
\]
\[
= P(X_1 \leq y) P(X_2 \leq y) \cdots P(X_n \leq y) = (F(y))^n
\]
⇒ 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
\[
G_1(y) = P(Y_1 \leq y) = P(\min\{X_1, \ldots, X_n\} \leq y) = 1 - P(\min\{X_1, \ldots, X_n\} > y)
\]
\[
= 1 - P(X_1 > y, X_2 > y, \ldots, X_n > y) = 1 - P(X_1 > y)P(X_2 > y) \cdots P(X_n > y)
\]
\[
= 1 - (1 - F(y))^n
\]
⇒ 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 \( \theta \). 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 \( \prod_{i=1}^{n} I_{[0,\theta]}(x_i) = 1 \iff 0 \leq x_i \leq \theta, \) for all \( i = 1, \ldots, n \iff 0 \leq y_n \leq \theta \)
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 \( \theta \) 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

\[ g_n(y) = n\left(\frac{y}{\theta}\right)^{n-1} \frac{1}{\theta} = n\frac{y^{n-1}}{\theta^n}, \quad 0 \leq y \leq \theta \]

\[ E(Y_n) = \int_0^\theta y n\frac{y^{n-1}}{\theta^n} dy = \frac{n}{n+1} \theta \neq \theta \Rightarrow \text{m.l.e. } \hat{\theta} = Y_n \text{ is not unbiased.} \]

However, \[ E(Y_n) = \frac{n}{n+1} \theta \to \theta \text{ as } n \to \infty, \text{ m.l.e. } \hat{\theta} \text{ is asymptotically unbiased.} \]

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

\[ \text{Var}(Y_n) = E(Y_n^2) - (E(Y_n))^2 = \frac{n}{n+2} \theta^2 - \left(\frac{n}{n+1}\right)^2 \theta^2 \to \theta^2 - \theta^2 = 0 \text{ as } n \to \infty. \]

\[ \Rightarrow Y_n \overset{p}{\to} \theta \Rightarrow \text{m.l.e. } \hat{\theta} = Y_n \text{ is consistent for } \theta. \]

Is there unbiased estimator for $\theta$?

\[ E\left(\frac{n+1}{n} Y_n\right) = \frac{n+1}{n} E(Y_n) = \frac{n+1}{n} \frac{n}{n+1} \theta = \theta \]

\[ \Rightarrow \frac{n+1}{n} Y_n \text{ is unbiased for } \theta. \]

Example:

(a) \[ Y \sim b(n, p) \]

The likelihood function is

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

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

\[ \frac{\partial \ln L(p)}{\partial p} = \frac{y}{p} - \frac{n-y}{1-p} = 0 \iff \frac{y}{p} = \frac{n-y}{1-p} \iff y(1-p) = p(n-y) \iff y = np \]

\[ \Rightarrow \text{m.l.e. } \hat{p} = \frac{Y}{n} \]

\[ E(\hat{p}) = \frac{1}{n} E(Y) = p \Rightarrow \text{m.l.e. } \hat{p} = \frac{Y}{n} \text{ is unbiased.} \]

\[ \text{Var}(\hat{p}) = \frac{1}{n^2} \text{Var}(Y) = \frac{1}{n} p(1-p) \to 0 \text{ as } n \to \infty \]

\[ \Rightarrow \text{m.l.e. } \hat{p} = \frac{Y}{n} \text{ is consistent for } p. \]

(b) \[ X_1, \ldots, X_n \text{ are a random sample from } N(\mu, \sigma^2). \text{ Want m.l.e.'s of } \mu \text{ and } \sigma^2 \]

The likelihood function is

\[ L(\mu, \sigma^2) = \prod_{i=1}^{\infty} \frac{1}{\sqrt{2\pi}(\sigma^2)^{\frac{3}{2}}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(2\pi\right)^{-\frac{3}{4}}(\sigma^2)^{-\frac{3}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \]

21
\[
\ln L(\mu, \sigma^2) = \left(\frac{n}{2}\right) \ln (2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2
\]

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

\[
\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \bar{X})^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2
\]

E(\hat{\mu}) = E(\bar{X}) = \mu \text{ (unbiased), } Var(\hat{\mu}) = Var(\bar{X}) = \frac{\sigma^2}{n} \to 0 \text{ as } n \to \infty \Rightarrow \text{ m.l.e. } \hat{\mu} \text{ is consistent for } \mu.

E(\hat{\sigma}^2) = E\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2\right) = \frac{n-1}{n} \sigma^2 \neq \sigma^2 \text{ (biased).}

E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2 \to \sigma^2 \text{ as } n \to \infty \Rightarrow \hat{\sigma}^2 \text{ is asymptotically unbiased.}

Var(\hat{\sigma}^2) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2\right) = \frac{1}{n^2} \text{Var}\left(\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{\sigma^2}\right)

= \frac{\sigma^4}{n^2} \text{Var}\left(\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{\sigma^2}\right) = \frac{2(n-1)}{n^2} \sigma^4 \to 0 \text{ as } n \to \infty

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

Suppose that we have m.l.e. \( \hat{\theta} = \hat{\theta}(x_1, \ldots, x_n) \) for parameter \( \theta \) and our interest is a new parameter \( \tau(\theta) \), a function of \( \theta \).

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 \( \theta \) and \( \tau(\theta) \) is a 1-1 function of \( \theta \), then m.l.e. of \( \tau(\theta) \) is \( \tau(\theta) \)

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

\[
L(\theta, x_1, \ldots, x_n) = L(\tau^{-1}(\tau(\theta)), x_1, \ldots, x_n)
\]

\[
= M(\tau(\theta), x_1, \ldots, x_n)
\]

\[
= M(\tau, x_1, \ldots, x_n), \tau \in T
\]
\[ M(\tau(\hat{\theta}), x_1, \ldots, x_n) = L^{-1}(\tau(\hat{\theta}), x_1, \ldots, x_n) \]
\[ = L(\hat{\theta}, x_1, \ldots, x_n) \]
\[ \geq L(\theta, x_1, \ldots, x_n), \forall \theta \in \Theta \]
\[ = L^{-1}(\tau(\theta)), x_1, \ldots, x_n) \]
\[ = M(\tau(\theta), x_1, \ldots, x_n), \forall \theta \in \Theta \]
\[ = M(\tau, x_1, \ldots, x_n), \tau \in T \]

\[ \Rightarrow \tau(\hat{\theta}) \text{ is m.l.e. of } \tau(\theta). \]

This is the invariance property of m.l.e.

\[ \square \]

Example:

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

<table>
<thead>
<tr>
<th>( \tau(p) )</th>
<th>m.l.e. of ( \tau(p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p^2 )</td>
<td>( \hat{p}^2 = \left( \frac{Y}{n} \right)^2 )</td>
</tr>
<tr>
<td>( \sqrt{p} )</td>
<td>( \hat{\sqrt{p}} = \sqrt{\frac{Y}{n}} )</td>
</tr>
<tr>
<td>( e^p )</td>
<td>( \hat{e^p} = e^{\frac{Y}{n}} )</td>
</tr>
<tr>
<td>( e^{-p} )</td>
<td>( \hat{e^{-p}} = e^{-\frac{Y}{n}} )</td>
</tr>
</tbody>
</table>

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

(2) \( X_1, \ldots, X_n \) \( \text{iid} \sim N(\mu, \sigma^2) \), m.l.e.'s of \( (\mu, \sigma^2) \) is \( (\bar{X}, \sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2}) \).

m.l.e.'s of \( (\mu, \sigma) \) is \( (\bar{X}, \sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2}) \) (\( : \sigma \in (0, \infty) \). \( \because \sigma^2 \rightarrow \sigma \) is 1-1)

You can also solve

\[ \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2, x_1, \ldots, x_n) = 0 \]
\[ \frac{\partial}{\partial \sigma} \ln L(\mu, \sigma^2, x_1, \ldots, x_n) = 0 \text{ for } \mu, \sigma \]

(\( \mu^2, \sigma \)) is not a 1-1 function of \( (\mu, \sigma^2) \).

\( \because \mu \in (-\infty, \infty) \). \( \because \mu \rightarrow \mu^2 \) isn’t 1-1

Best estimator:

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

\[ \text{Var}_\theta \hat{\theta} \leq \text{Var}_\theta \hat{\theta^*}, \text{ for } \theta \in \Theta \]

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

23
There are several ways in deriving UMVUE of $\theta$.

Cramer-Rao lower bound for variance of unbiased estimator:

Regularity conditions:

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

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

(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

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

**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

$$\text{Var}_\theta \hat{\tau}(\theta) \geq \frac{(\tau'(\theta))^2}{n E_\theta \left[ (\frac{\partial \ln f(x, \theta)}{\partial \theta})^2 \right]} = \frac{(\tau'(\theta))^2}{-n E_\theta \left[ (\frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2}) \right]}$$

for $\theta \in \Theta$

**Proof.** Consider only the continuous distribution.

$$E_\theta \left[ \frac{\partial \ln f(x, \theta)}{\partial \theta} \right] = \int_{-\infty}^{\infty} \frac{\partial \ln f(x, \theta)}{\partial \theta} f(x, \theta) dx = \int_{-\infty}^{\infty} \frac{\partial f(x, \theta)}{\partial \theta} dx = 0$$

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

Taking derivatives both sides.

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

$$= \int \cdots \int t(x_1, \ldots, x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i, \theta) \prod_{i=1}^{n} dx_i - \tau(\theta) \frac{\partial}{\partial \theta} \int \cdots \int \prod_{i=1}^{n} f(x_i, \theta) \prod_{i=1}^{n} dx_i$$

$$= \int \cdots \int (t(x_1, \ldots, x_n) - \tau(\theta)) \frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i, \theta) \prod_{i=1}^{n} dx_i$$
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
\[
[E(XY)]^2 \leq E(X^2)E(Y^2)
\]

Then
\[
\tau'(\theta) = \int \cdots \int (t(x_1, \ldots, x_n) - \tau(\theta)) \left( \sum_{j=1}^{n} \frac{\partial \ln f(x_j, \theta)}{\partial \theta} \prod_{i=1}^{n} f(x_i, \theta) \prod_{i=1}^{n} dx_i \right)
\]
\[
= E[(t(x_1, \ldots, x_n) - \tau(\theta)) \sum_{j=1}^{n} \frac{\partial \ln f(x_j, \theta)}{\partial \theta}]
\]
\[
(\tau'(\theta))^2 \leq E[(t(x_1, \ldots, x_n) - \tau(\theta))^2] E[(\sum_{j=1}^{n} \frac{\partial \ln f(x_j, \theta)}{\partial \theta})^2]
\]
\[
\Rightarrow \text{Var}(\hat{\tau}(\theta)) \geq \frac{(\tau'(\theta))^2}{E[(\sum_{j=1}^{n} \frac{\partial \ln f(x_j, \theta)}{\partial \theta})^2]}
\]

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

Cauchy-Swartz Inequality
\[
[E(XY)]^2 \leq E(X^2)E(Y^2)
\]
Then, we have
\[
\text{Var}_\theta \hat{\tau}(\theta) \geq \frac{(\tau'(\theta))^2}{nE_\theta \left[\left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2\right]}
\]
You may further check that
\[
E_\theta \left(\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2}\right) = -E_\theta \left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2
\]

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

\textbf{Note:}
If \(\tau(\theta) = \theta\), then any unbiased estimator \(\hat{\theta}\) satisfies
\[
\text{Var}_\theta \hat{\theta}(\theta) \geq \frac{(\tau'(\theta))^2}{-nE_\theta \left[\left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2\right]}
\]

\textbf{Example:}
(a) \(X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda), E(X) = \lambda, \text{Var}(X) = \lambda\).

\(\text{MLE} \hat{\lambda} = \bar{X}, E(\hat{\lambda}) = \lambda, \text{Var}(\hat{\lambda}) = \frac{\lambda}{n}.\)

\(\text{p.d.f. } f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, \ldots\)

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

Cramer-Rao lower bound is
\[
\frac{1}{-n\left(-\frac{1}{\lambda}\right)} = \frac{\lambda}{n} = \text{Var}(\hat{\lambda})
\]

\[\Rightarrow \text{MLE} \hat{\lambda} = \bar{X} \text{ is the UMVUE of } \lambda.\]
(b) $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$, $E(X) = p$, $\text{Var}(X) = p(1-p)$.

Want UMVUE of $p$.

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

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

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

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

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

C-R lower bound for $p$ is

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

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

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

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