Mathematical Statistics

Chen, L.-A.

Chapter 4. Distribution of Function of Random variables

Sample space $S$: set of possible outcome in an experiment.

Probability set function $P$:
(1) $P(A) \geq 0, \forall A \subset S$.
(2) $P(S) = 1$.
(3) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), i f A_i \cap A_j = \emptyset, \forall i \neq j$.

Random variable $X$:
$X : S \rightarrow R$
Given $B \subset R, P(X \in B) = P(\{s \in S : X(s) \in B\}) = P(X^{-1}(B))$ where $X^{-1}(B) \subset S$.

$X$ is a discrete random variable if its range
$$X(s) = \{x \in R : \exists s \in S, X(s) = x\}$$
is countable. The probability density/mass function (p.d.f) of $X$ is defined as
$$f(x) = P(X = x), x \in R.$$

Distribution function $F$:
$$F(x) = P(X \leq x), x \in R.$$  

A r.v. is called a continuous r.v. if there exists $f(x) \geq 0$ such that
$$F(x) = \int_{-\infty}^{x} f(t)dt, x \in R.$$  

where $f$ is the p.d.f of continuous r.v. $X$. 

1
Let X be a r.v. with p.d.f. f(x). Let g : R → R
Q: What is the p.d.f. of g(x)? and is g(x) a r.v.? (Yes)
Answer:
(a) distribution method:
Suppose that X is a continuous r.v. Let Y = g(X)
The d.f (distribution function) of Y is
\[ G(y) = P(Y \leq y) = P(g(X) \leq y) \]
If G is differentiable then the p.d.f. of Y = g(X) is g′(y).
(b) mgf method: (moment generating function)
\[ E[e^{tX}] = \left\{ \begin{array}{ll} \sum e^{tx} f(x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{(continuous)} \end{array} \right. \]
Thm. m.g.f. \( M_x(t) \) and its distribution (p.d.f. or d.f.) forms a 1-1 functions.
ex:
\[ M_Y(t) = e^{\frac{1}{2}t} = M_{N(0,1)}(t) \Rightarrow Y \sim N(0,1) \]
Let \( X_1, \ldots, X_n \) be random variables.
If they are discrete, the joint p.d.f. of \( X_1, \ldots, X_n \) is
\[ f(x_1, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n), \forall \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \]
If \( X_1, \ldots, X_n \) are continuous r.v.’s, there exists \( f \) such that
\[ F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \ldots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n) dt_1 \ldots dt_n, \text{for } \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \]
We call \( f \) the joint p.d.f. of \( X_1, \ldots, X_n \).
If \( X \) is continuous, then
\[ F(x) = \int_{-\infty}^{x} f(t) dt \text{ and } P(X = x) = \int_{x}^{x} f(t) dt = 0, \forall x \in \mathbb{R}. \]
Marginal p.d.f.’s:
Discrete:

\[ f_{X_i}(x) = P(X_i = x) = \sum_{x_n} \cdots \sum_{x_{i+1}} \sum_{x_{i-1}} \sum_{x_1} f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) \]

Continuous:

\[ f_{X_i}(x) = \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n \]

Events A and B are independent if \( P(A \cap B) = P(A)P(B) \).
Q: If \( A \cap B = \emptyset \), are A and B independent?
A: In general, they are not.

Let X and Y be r.v.’s with joint p.d.f. \( f(x, y) \) and marginal p.d.f. \( f_X(x) \) and \( f_Y(y) \). We say that X and Y are independent if

\[ f(x, y) = f_X(x)f_Y(y), \forall \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \]

Random variables X and Y are identically distributed (i.d.) if marginal p.d.f.’s f and g satisfy \( f = g \) or d.f.’s F and G satisfy \( F = G \).

We say that X and Y are iid random variables if they are independent and identically distributed.

Transformation of r.v.’s (discrete case)
Univariate: \( Y = g(X) \), p.d.f. of Y is

\[ g(y) = P(Y = y) = P(g(x) = y) = P(\{x \in \text{Range of } X : g(x) = y\}) = \sum_{\{x:g(x)=y\}} f(x) \]

For random variables \( X_1, \ldots, X_n \) with joint p.d.f. \( f(x_1, \ldots, x_n) \), define transformations

\[ Y_1 = g_1(X_1, \ldots, X_n), \ldots, Y_m = g_m(X_1, \ldots, X_n). \]

The joint p.d.f. of \( Y_1, \ldots, Y_m \) is
\[ g(y_1, \ldots, y_m) = P(Y_1 = y_1, \ldots, Y_m = y_m) \\
= P(\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : g_1(x_1, \ldots, x_n) = y_1, \ldots, g_m(x_1, \ldots, x_n) = y_m \}) \\
= \sum \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \{ : g_1(x_1, \ldots, x_n) = y_1, \ldots, g_m(x_1, \ldots, x_n) = y_m \} \sum f(x_1, \ldots, x_n) \\
\]

Example: joint p.d.f. of \( X_1, X_2, X_3 \) is
\[
\begin{array}{c|cccccc}
(x_1, x_2, x_3) & (0, 0, 0) & (0, 0, 1) & (0, 1, 1) & (1, 0, 1) & (1, 1, 0) & (1, 1, 1) \\
f(x_1, x_2, x_3) & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\
\end{array}
\]
\[ Y_1 = X_1 + X_2 + X_3, Y_2 = |X_3 - X_2| \]
Space of \((Y_1, Y_2)\) is \{(0, 0), (1, 1), (2, 0), (2, 1), (3, 0)\}.
Joint p.d.f. of \( Y_1 \) and \( Y_2 \) is
\[
\begin{array}{c|cccc}
(y_1, y_2) & (0, 0) & (1, 1) & (2, 0) & (2, 1) & (3, 0) \\
g(y_1, y_2) & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & \frac{2}{8} & \frac{1}{8} \\
\end{array}
\]
Continuous one-to-one transformations:
Let \( X \) be a continuous r.v. with joint p.d.f. \( f(x) \) and range \( A = X(s) \).
Consider \( Y = g(x) \), a differentiable function. We want p.d.f. of \( Y \).

**Thm.** If \( g \) is 1-1 transformation, then the p.d.f. of \( Y \) is
\[ f_Y(y) = \begin{cases} 
  f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| & y \in g(A) \\
  0 & \text{otherwise.} 
\end{cases} \]

**Proof.** The d.f. of \( Y \) is
\[ F_Y(y) = P(Y \leq y) = P(g(X) \leq y) \]
(a) If \( g \) is \( \nearrow \), \( g^{-1} \) is also \( \nearrow \), \( \left( \frac{dg^{-1}}{dy} > 0 \right) \)
\[ F_Y(y) = P(X \leq g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(x) dx \]
\[ f_Y(y) = D_y \int_{-\infty}^{g^{-1}(y)} f_X(x) \, dx \]
\[ = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} \]
\[ = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \]

(b) If \( g \) is \( \searrow \), \( g^{-1} \) is also \( \searrow \). \( \left( \frac{dg^{-1}}{dy} < 0 \right) \)

\[ F_Y(y) = P(X \geq g^{-1}(y)) = \int_{g^{-1}(y)}^{\infty} f_X(x) \, dx = 1 - \int_{-\infty}^{g^{-1}(y)} f_X(x) \, dx \]

\[ \Rightarrow \text{p.d.f. of } Y \text{ is} \]
\[ f_Y(y) = D_y(1 - \int_{-\infty}^{g^{-1}(y)} f_X(x) \, dx) \]
\[ = -f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} \]
\[ = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \]

Example: \( X \sim U(0, 1), Y = -2 \ln(x) = g(x) \)

sol: p.d.f. of \( X \) is
\[ f_X(x) = \begin{cases} 
1 & \text{if } 0 < x < 1 \\
0 & \text{elsewhere.}
\end{cases} \]

\( A = (0, 1), g(A) = (0, \infty) \),

\[ x = e^{-\frac{y}{2}} = g^{-1}(y), \frac{dx}{dy} = -\frac{1}{2}e^{-\frac{y}{2}} \]

p.d.f. of \( Y \) is
\[ f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dy}{dx} \right| = \frac{1}{2}e^{-\frac{y}{2}}, y > 0 \]

\( (X \sim U(a, b) \text{ if } f_X(x) = \begin{cases} 
\frac{1}{b-a} & \text{if } a < x < b \\
0 & \text{elsewhere.}
\end{cases} \)
\[ Y \sim \chi^2(2) \]

\[ X \sim \chi^2(r) \text{ if } f_X(x) = \frac{1}{\Gamma\left(\frac{r}{2}\right)2^{\frac{r}{2}}}x^{\frac{r}{2}-1}e^{-\frac{x}{2}}, x > 0 \]

Continuous n-r.v.-to-m-r.v., \( n > m \), case:

\[
\begin{pmatrix}
  x_1 \\
  \vdots \\
  x_n
\end{pmatrix}
\rightarrow
\begin{cases}
  Y_1 = g_1(X_1, \ldots, X_n) \\
  \vdots \\
  Y_m = g_m(X_1, \ldots, X_n)
\end{cases}
\]

Q: What are the marginal p.d.f. of \( Y_1, \ldots, Y_m \)

A: We need to define \( Y_{m+1} = g_{m+1}(X_1, \ldots, X_n), \ldots, Y_n = g_n(X_1, \ldots, X_n) \)

such that \( \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix} \) is 1-1 from \( \mathbb{R}^n \) to \( \mathbb{R}^m \).

Theory for change variables:

\[
P(\begin{pmatrix}
  x_1 \\
  \vdots \\
  x_n
\end{pmatrix} \in A) = \int \cdots \int f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)dx_1 \cdots dx_n
\]

Let \( y_1 = g_1(x_1, \ldots, x_n), \ldots, y_n = g_n(x_1, \ldots, x_n) \) be a 1-1 function with

inverse \( x_1 = w_1(y_1, \ldots, y_n), \ldots, x_n = w_n(y_1, \ldots, y_n) \)

and Jacobian

\[
J = \begin{vmatrix}
  \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\
  \vdots & \ddots & \vdots \\
  \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n}
\end{vmatrix}
\]

Then

\[
\int \cdots \int f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)dx_1 \cdots dx_n
\]

\[= \int \cdots \int f_{X_1, \ldots, X_n}(w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n))|J|dy_1 \cdots dy_n\]

Hence, joint p.d.f. of \( Y_1, \ldots, Y_n \) is

\[ f_{Y_1, \ldots, Y_n}(y_1, \ldots, y_n) = f_{X_1, \ldots, X_n}(w_1, \ldots, w_n)|J| \]
Theorem. Suppose that $X_1$ and $X_2$ are two r.v.'s with continuous joint p.d.f. $f_{X_1, X_2}$ and sample space $A$.
If $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$ forms a $1 - 1$ transformation inverse function

$$
\begin{pmatrix}
X_1 \\
X_2
\end{pmatrix} = \begin{pmatrix}
W_1(Y_1, Y_2) \\
W_2(Y_1, Y_2)
\end{pmatrix}
$$

and Jacobian $J = \left| \begin{array}{cc}
\frac{\partial X_1}{\partial Y_1} & \frac{\partial X_1}{\partial Y_2} \\
\frac{\partial X_2}{\partial Y_1} & \frac{\partial X_2}{\partial Y_2}
\end{array} \right|$

the joint p.d.f. of $Y_1, Y_2$ is

$$
f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(w_1(y_1, y_2), w_2(y_1, y_2))|J|, \quad \left( \begin{array}{c}
y_1 \\
y_2
\end{array} \right) \in \left( \begin{array}{c}
g_1 \\
g_2
\end{array} \right)(A).
$$

Steps:
(a) joint p.d.f. of $X_1, X_2$, space $A$.
(b) check if it is $1 - 1$ transformation.
Inverse function $X_1 = w_1(Y_1, Y_2), X_2 = w_2(Y_1, Y_2)$
(c) Range of $(Y_1, Y_2) = (g_1, g_2)(A)$

Example: For $X_1, X_2 \sim U(0, 1)$, let $Y_1 = X_1 + X_2, Y_2 = X_1 - X_2$.
Want marginal p.d.f. of $Y_1, Y_2$
Sol: joint p.d.f. of $X_1, X_2$ is

$$
f_{X_1, X_2}(x_1, x_2) = \begin{cases} 
1 & \text{if } 0 < x_1 < 1, 0 < x_2 < 1 \\
0 & \text{elsewhere.}
\end{cases}
$$

$$
A = \left\{ \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} : 0 < x_1 < 1, 0 < x_2 < 1 \right\}
$$

Given $y_1, y_2$, solve $y_1 = x_1 + x_2, y_2 = x_1 - x_2$.

$$
\Rightarrow x_1 = \frac{y_1 + y_2}{2} = w_1(y_1, y_2), x_2 = \frac{y_1 - y_2}{2} = w_2(y_1, y_2)
$$

$(1 - 1$ transformation)

Jacobian is

$$
J = \left| \begin{array}{cc}
\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\
\frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2}
\end{array} \right| = \left| \begin{array}{cc}
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & -\frac{1}{2}
\end{array} \right| = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}
$$

The joint p.d.f. of $Y_1, Y_2$ is

$$
f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(w_1, w_2)|J|, \quad \left( \begin{array}{c}
y_1 \\
y_2
\end{array} \right) \in B
$$
Marginal p.d.f. of $Y_1, Y_2$ are

$$f_{Y_1}(y_1) = \begin{cases} \int_{y_1}^{y_1 + \frac{1}{2}} dy_2 = y_1, & 0 < y_1 < 1 \\ \int_{y_1 - \frac{1}{2}}^{1 - y_1} dy_2 = 1 - y_1, & 1 < y_1 < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

$$f_{Y_2}(y_2) = \begin{cases} \int_{-y_2 + \frac{1}{2}}^{y_2 + \frac{1}{2}} dy_1 = y_2 + 1, & -1 < y_2 < 0 \\ \int_{y_2 - \frac{1}{2}}^{1 - y_2} dy_1 = 1 - y_2, & 0 < y_2 < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

**Def.** If a sequence of r.v.’s $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.), then they are called a random sample.

If $X_1, \ldots, X_n$ is a random sample from a distribution with p.d.f. $f_0$, then the joint p.d.f. of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_0(x_i), \quad \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

**Def.** Any function $g(X_1, \ldots, X_n)$ of a random sample $X_1, \ldots, X_n$ which is not dependent on a parameter $\theta$ is called a statistic.

**Note:** If $X$ is a random sample with p.d.f. $f(x, \theta)$, where $\theta$ is an unknown constant, then $\theta$ is called a parameter.

For example, $N(\mu, \sigma^2): \mu, \sigma^2$ are parameters.

$\text{Poisson}(\lambda): \lambda$ is a parameter.

Example of statistics:

$X_1, \ldots, X_n$ are iid r.v.’s $\Rightarrow \bar{X}$ and $S^2$ are statistics.

**Note:** If $X_1, \ldots, X_n$ are r.v.’s, the m.g.f of $X_1, \ldots, X_n$ is

$$M_{X_1,\ldots,X_n}(t_1,\ldots,t_n) = \text{E}(e^{t_1X_1+\cdots+t_nX_n})$$

m.g.f

$$M_x(t) = \text{E}(e^{tx}) = \int e^{tx} f(x) dx$$

$$\Rightarrow D_t M_x(t) = D_t \text{E}(e^{tx}) = D_t \int e^{tx} f(x) dx = \int D_t e^{tx} f(x) dx$$
Lemma. $X_1$ and $X_2$ are independent if and only if

$$M_{X_1, X_2}(t_1, t_2) = M_{X_1}(t_1) M_{X_2}(t_2), \forall t_1, t_2.$$ 

Proof. $\Rightarrow$ If $X_1, X_2$ are independent,

$$M_{X_1, X_2}(t_1, t_2) = E(e^{t_1 X_1 + t_2 X_2})$$

$$= \int \int e^{t_1 x_1 + t_2 x_2} f(x_1, x_2) dx_1 dx_2$$

$$= \int e^{t_1 x_1} f_{X_1}(x_1) dx_1 \int e^{t_2 x_2} f_{X_2}(x_2) dx_2$$

$$= E(e^{t_1 X_1}) E(e^{t_2 X_2})$$

$$= M_{X_1}(t_1) M_{X_2}(t_2)$$

$\Leftarrow$ \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm}

$$M_{X_1, X_2}(t_1, t_2) = E(e^{t_1 X_1 + t_2 X_2}) = \int \int e^{t_1 x_1 + t_2 x_2} f(x_1, x_2) dx_1 dx_2$$

$$M_{X_1}(t_1) M_{X_2}(t_2) = E(e^{t_1 X_1}) E(e^{t_2 X_2})$$

$$= \int e^{t_1 x_1} f_{X_1}(x_1) dx_1 \int e^{t_2 x_2} f_{X_2}(x_2) dx_2$$

$$= \int e^{t_1 x_1 + t_2 x_2} f(x_1, x_2) dx_1 dx_2$$

With $1 - 1$ correspondence between m.g.f and p.d.f,
then $f(x_1, x_2) = f_1(x_1) f_2(x_2), \forall x_1, x_2$

$\Rightarrow X_1, X_2$ are independent. \hfill \Box$

$X$ and $Y$ are independent, denote by $X \perp Y$.

$$\begin{cases} 
X \sim N(\mu, \sigma^2) & , M_x(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}, \forall t \in R \\
X \sim \text{Gamma}(\alpha, \beta) & , M_x(t) = (1 - \beta t)^{-\alpha}, t < \frac{1}{\beta} \\
X \sim b(n, p) & , M_x(t) = (1 - p + p e^t)^n, \forall t \in R \\
X \sim \text{Poisson}(\lambda) & , M_x(t) = e^{\lambda (e^t - 1)}, \forall t \in R 
\end{cases}$$

Note :
(a) If \((X_1, \ldots, X_n)\) and \((Y_1, \ldots, Y_m)\) are independent, then \(g(X_1, \ldots, X_n)\) and \(h(Y_1, \ldots, Y_m)\) are also independent.

(b) If \(X, Y\) are independent, then
\[
E[g(X)h(Y)] = E[g(X)]E[h(Y)].
\]

**Thm.** If \((X_1, \ldots, X_n)\) is a random sample from \(N(\mu, \sigma^2)\), then

(a) \(\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\)

(b) \(\overline{X}\) and \(S^2\) are independent.

(c) \(\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)\)

**Proof.** (a) m.g.f. of \(\overline{X}\) is
\[
M_{\overline{X}}(t) = E(e^{t\overline{X}}) = E(e^{t \frac{1}{n} \sum_{i=1}^{n} X_i})
= E(e^{\frac{t}{n} X_1} e^{\frac{t}{n} X_2} \cdots e^{\frac{t}{n} X_n})
= E(e^{\frac{t}{n} X_1})E(e^{\frac{t}{n} X_2})E(e^{\frac{t}{n} X_n})
= M_{X_1}(\frac{t}{n})M_{X_2}(\frac{t}{n}) \cdots M_{X_n}(\frac{t}{n})
= (e^{\mu \frac{t}{n} + \frac{\sigma^2}{2n} \frac{t^2}{2}})^n
= e^{\mu t + \frac{\sigma^2}{2n} t^2}
\]

\(\Rightarrow \overline{X} \sim (\mu, \frac{\sigma^2}{n})\)

(b) First we want to show that \(\overline{X}\) and \((X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})\) are
independent. Joint m.g.f. of $X$ and $(X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X})$ is

$$M_{X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}} (t, t_1, \ldots, t_n) = E[e^{tX_1 + t_1(X_1 - \bar{X}) + \ldots + t_n(X_n - \bar{X})}]$$

$$= E[e^{\frac{t}{n} \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} t_i (X_i - \bar{X})}]$$

$$= E[e^{\sum_{i=1}^{n} \left( \frac{t}{n} + t_i \right) X_i}], \bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

$$= E\left[ e^{\sum_{i=1}^{n} \left( \frac{t}{n} + t_i \right) X_i} \right]$$

$$= E\left[ \prod_{i=1}^{n} e^{\frac{t}{n} + t_i X_i} \right]$$

$$= \prod_{i=1}^{n} e^{e^{\frac{t}{n} + t_i} + \frac{\sigma^2}{2} \frac{t^2}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$= e^{e^{\frac{t}{n} + \frac{\sigma^2}{2} \sum_{i=1}^{n} (X_i - \bar{X})^2}}$$

$$= M_{X} (t) M_{X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}} (t_1, \ldots, t_n)$$

$\Rightarrow \bar{X}$ and $(X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X})$ are independent.

$\Rightarrow \bar{X}$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ are independent.

(c)

1. $Z \sim N(0, 1) \Rightarrow Z^2 \sim \chi^2(1)$
2. $X \sim \chi^2(r_1)$ and $Y \sim \chi^2(r_2)$ are independent. $\Rightarrow X + Y \sim \chi^2(r_1 + r_2)$

Proof. m.g.f. of $X + Y$ is

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$$

$$= (1 - 2t)^{-\frac{r_1}{2}} (1 - 2t)^{-\frac{r_2}{2}} = (1 - 2t)^{-\frac{r_1 + r_2}{2}}$$

$\Rightarrow X + Y \sim \chi^2(r_1 + r_2)$

3. $(X_1, \ldots, X_n) \overset{iid}{\sim} N(\mu, \sigma)$

$$\frac{X_1 - \mu}{\sigma}, \frac{X_2 - \mu}{\sigma}, \ldots, \frac{X_n - \mu}{\sigma} \overset{iid}{\sim} N(0, 1)$$
\[
\frac{(X_1 - \mu)^2}{\sigma^2}, \frac{(X_2 - \mu)^2}{\sigma^2}, \ldots, \frac{(X_n - \mu)^2}{\sigma^2} \overset{iid}{\sim} \chi^2(1)
\]

\[
\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi^2(n)
\]

\[
\frac{(n - 1)s^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - X)^2}{\sigma^2} \sim \chi^2(n - 1)
\]

\[
(1 - 2t)^{-\frac{a}{2}} = M_{\frac{\sum(X_i - \mu)^2}{\sigma^2}}(t) = \mathbb{E}(e^{t\frac{\sum(X_i - \mu)^2}{\sigma^2}})
\]

\[
= \mathbb{E}(e^{t\frac{\sum(X_i - X + X - \mu)^2}{\sigma^2}}) = \mathbb{E}(e^{t\frac{\sum(X_i - X)^2 + n(X - \mu)^2}{\sigma^2}})
\]

\[
= \mathbb{E}(e^{t\frac{(n-1)s^2}{\sigma^2}} e^{t\frac{(X - \mu)^2}{\sigma^2/n}})
\]

\[
= \mathbb{E}(e^{t\frac{(n-1)s^2}{\sigma^2}}) \mathbb{E}(e^{t\frac{(X - \mu)^2}{\sigma^2/n}})
\]

\[
= M_{\frac{(n-1)s^2}{\sigma^2}}(t) M_{\frac{(X - \mu)^2}{\sigma^2/n}}(t)
\]

\[
= M_{\frac{(n-1)s^2}{\sigma^2}}(t)(1 - 2t)^{-\frac{1}{2}}
\]

\[
\Rightarrow M_{\frac{(n-1)s^2}{\sigma^2}}(t) = (1 - 2t)^{-\frac{n-1}{2}} \Rightarrow \frac{(n - 1)s^2}{\sigma^2} \sim \chi^2(n - 1)
\]
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

- **Estimation**
  - 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 $\theta$ **unbiased** for $\theta$ if it satisfies

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

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

**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: \(X_1, \ldots, X_n \sim iid N(\mu, \sigma^2)\), Suppose that our interest is \(\mu, X_1, \ldots, X_n\), \(E_{\mu}(X_1) = \mu\), is unbiased for \(\mu\),
\[
\frac{1}{2}(X_1 + X_2), E(\frac{X_1 + X_2}{2}) = \mu, \text{ is unbiased for } \mu,
\]
\[
\overline{X}, E_{\mu}(\overline{X}) = \mu, \text{ is unbiased for } \mu,
\]
\[\triangleright a_n \xrightarrow{n \to \infty} a, \text{ if } \forall \epsilon > 0, \text{ there exists } N > 0 \text{ such that } |a_n - a| < \epsilon \text{ if } n \geq N.\]

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

Def. We say that \(X_n \text{ converges to } X\), 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 \(\overline{X} \xrightarrow{P} \mu\).
Proof.

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

\[ \square \]

**Def.** We say 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 \( \bar{X} \) a consistent estimator of \( \mu \)?

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

\[ \text{Let } \epsilon > 0, \]

\[ P(|X_1 - \mu| > \epsilon) = 1 - P(|X_1 - \mu| \leq \epsilon) = 1 - P(\mu - \epsilon \leq X_1 \leq \mu + \epsilon) \]

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

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

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

\[ \Rightarrow \bar{X} \xrightarrow{P} \mu. \]

\[ \Rightarrow \bar{X} \text{ is a consistent estimator of } \mu. \]

\[ \triangleq \text{Unbiasedness and consistency are two basic conditions for good estimator.} \]

**Moments :**

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

\[ E_\theta(X^k) = \left\{ \begin{array}{ll}
\sum_{x} x^k f(x, \theta) & , \text{discrete} \\
\int_{-\infty}^{\infty} x^k f(x, \theta) dx & , \text{continuous}
\end{array} \right. \]

The sample \( k \text{th} \) 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.

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

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

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

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

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

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

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

\[
\text{E}(S^2) = \text{E}\left(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2\right) = \frac{1}{n-1} \text{E}\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X} \sum_{i=1}^{n} X_i + n\overline{X}^2\right)
\]

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

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

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

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

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

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

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

⇒ $S^2 \xrightarrow{P} \sigma^2$
**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 \( \bar{X} = E_{\theta}(X) \)

(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 \( \bar{X} = E_p(X) = p \)

\( \Rightarrow \) The method of moment estimator of \( p \) is \( \hat{p} = \bar{X} \)

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

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

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

Let \( \bar{X} = E_\lambda(X) = \lambda \)

\( \Rightarrow \) The method of moment estimator of \( \lambda \) is \( \hat{\lambda} = \bar{X} \)

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

\( \hat{\lambda} = \bar{X} \overset{p}{\rightarrow} E(X) = \lambda \Rightarrow \hat{\lambda} \) 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 \( \bar{X} = E_{\mu, \sigma^2}(X) = \mu \)

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

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

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

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

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

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

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

**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.?
\[
\frac{\partial \ln x}{\partial x} = \frac{1}{x} > 0 \quad \Rightarrow \quad \ln x \text{ is } \nearrow \text{ in } x \nabla
\]
⇒ 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
\[
\ln L(\hat{\theta}, x_1, \ldots, x_n) = \max_{\theta \in \Theta} \ln L(\theta, x_1, \ldots, x_n)
\]

Two cases to solve m.l.e.:
\[
(b.1) \quad \frac{\partial \ln L(\theta)}{\partial \theta} = 0
\]
\[
(b.2) \quad L(\theta) \text{ is monotone. Solve } \max_{\theta \in \Theta} L(\theta, x_1, \ldots, x_n) \text{ from monotone property.}
\]

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
\[
G_n(y) = P(Y_n \leq y) = P(\max\{X_1, \ldots, X_n\} \leq y) = P(X_1 \leq y, \ldots, X_n \leq y)
\]
\[
= P(X_1 \leq y)P(X_2 \leq y) \cdots P(X_n \leq y) = (F(y))^n
\]

19
\[ p.d.f. \text{ of } Y_n \text{ is } g_n(y) = D_y(F(y))^n = n(F(y))^{n-1}f(y) \]

Distribution function of \( Y_1 \) is

\[ G_1(y) = P(Y_1 \leq y) = P(\min\{X_1, \ldots, X_n\} \leq y) = 1 - P(\min\{X_1, \ldots, X_n\} > y) \]
\[ = 1 - P(X_1 > y, X_2 > y, \ldots, X_n > y) = 1 - P(X_1 > y)P(X_2 > y)\cdots P(X_n > y) \]
\[ = 1 - (1 - F(y))^n \]

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

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

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

Let \( Y_n = \max\{X_1, \ldots, X_n\} \)

Then \( \prod_{i=1}^{n} I_{[0,\theta]}(x_i) = 1 \iff 0 \leq x_i \leq \theta, \text{ for all } i = 1, \ldots, n \iff 0 \leq y_n \leq \theta \)

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

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

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

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

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

Is there unbiased estimator for \( \theta \)?

\[
E\left(\frac{n+1}{n} Y_n\right) = \frac{n+1}{n} E(Y_n) = \frac{n+1}{n} \frac{n}{n+1} \theta = \theta
\]
\( \Rightarrow \frac{n+1}{n} Y_n \) is unbiased for \( \theta \).

Example :

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

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

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

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

\( \Rightarrow \text{m.l.e. } \hat{p} = \frac{Y}{n} \)

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

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

\( \Rightarrow \text{m.l.e. } \hat{p} = \frac{Y}{n} \) 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 \( \mu \) and \( \sigma^2 \)

The likelihood function is
\[
L(\mu, \sigma^2) = \prod_{i=1}^{\infty} \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}
\]
\[
\ln L(\mu, \sigma^2) = \left(-\frac{n}{2}\right) \ln (2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \\
\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \Rightarrow \hat{\mu} = \bar{X} \\
\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \bar{X})^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2 \\
E(\hat{\mu}) = E(\bar{X}) = \mu \text{ (unbiased)}, \text{Var}(\hat{\mu}) = \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \rightarrow 0 \text{ as } n \rightarrow \infty \\
\Rightarrow \text{m.l.e. } \hat{\mu} \text{ is consistent for } \mu. \\
E(\hat{\sigma}^2) = E\left(\frac{1}{n} \sum (X_i - \bar{X})^2\right) = \frac{n-1}{n} \sigma^2 \neq \sigma^2 \text{ (biased).} \\
E(\hat{\sigma}^2) = \frac{1}{n} \sigma^2 \rightarrow \sigma^2 \text{ as } n \rightarrow \infty \Rightarrow \hat{\sigma}^2 \text{ is asymptotically unbiased.}
\]

\[
\text{Var}(\hat{\sigma}^2) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2\right) = \frac{1}{n^2} \text{Var}\left(\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{\sigma^2}\right) = \frac{\sigma^4}{n^2} \text{Var}\left(\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{\sigma^2}\right) = \frac{2(n-1)}{n^2} \sigma^4 \rightarrow 0 \text{ as } n \rightarrow \infty \\
\Rightarrow \text{m.l.e. } \hat{\sigma}^2 \text{ is consistent for } \sigma^2.
\]

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:

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

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

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

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

\[ \frac{\tau(p)}{\text{m.l.e. of } \tau(p)} \]
\[ p^2 \quad \hat{p}^2 = \left(\frac{Y}{n}\right)^2 \]
\[ \sqrt{p} \quad \sqrt{\hat{p}} = \sqrt{\frac{Y}{n}} \quad p(1-p) \text{ is not a 1-1 function of } p. \]
\[ e^p \quad e^\hat{p} = e^{\frac{Y}{n}} \]
\[ e^{-p} \quad e^{-\hat{p}} = e^{-\frac{Y}{n}} \]

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

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

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

\((\mu^2, \sigma)\) is not a 1-1 function of \((\mu, \sigma^2)\).
\( (\therefore \mu \in (-\infty, \infty) \), \( \therefore \mu \to \mu^2 \) isn’t 1-1)

Best estimator:

**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.)
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
\[
\int t \frac{\partial f(x, \theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int t f(x, \theta) dx
\]

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

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

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

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

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

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

25
Then, we have

\[ \text{Var}_{\theta} \hat{\tau}(\theta) \geq \frac{(\tau'(\theta))^2}{nE_{\theta}[(\frac{\partial \ln f(x,\theta)}{\partial \theta})^2]} \]

You may further check that

\[ E_{\theta}(\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2}) = -E_{\theta}(\frac{\partial \ln f(x,\theta)}{\partial \theta})^2 \]

\[ \square \]

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

**Note:**

If \( \tau(\theta) = \theta \), then any unbiased estimator \( \hat{\theta} \) satisfies

\[ \text{Var}_{\theta} \hat{\theta} \geq \frac{(\tau'(\theta))^2}{-nE_{\theta}[(\frac{\partial \ln f(x,\theta)}{\partial \theta})^2]} \]

**Example:**

(a) \( X_1, \ldots, X_n \sim \text{Poisson}(\lambda) \), \( E(X) = \lambda, \text{Var}(X) = \lambda \).

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

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

\[ \Rightarrow \ln f(x,\lambda) = x \ln \lambda - \lambda - \ln x! \]

\[ \Rightarrow \frac{\partial}{\partial \lambda} \ln f(x,\lambda) = x \frac{\lambda}{\lambda} - 1 \]

\[ \Rightarrow \frac{\partial^2}{\partial \lambda^2} \ln f(x,\lambda) = -\frac{x}{\lambda^2} \]

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

Cramer-Rao lower bound is

\[ \frac{1}{-n(-\frac{1}{\lambda})} = \frac{\lambda}{n} = \text{Var}(\hat{\lambda}) \]

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

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

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

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

Chapter 4. Continue to Point Estimation-UMVUE

Sufficient Statistic:
A, B are two events. The conditional probability of A given B is

\[ P(A|B) = \frac{P(A \cap B)}{P(B)}, A \subset S. \]

\( P(\cdot|B) \) is a probability set function with domain of subsets of sample space S.

Let \( X, Y \) be two r.v’s with joint p.d.f \( f(x, y) \) and marginal p.d.f’s \( f_X(x) \) and \( f_Y(y) \). The conditional p.d.f of \( Y \) given \( X = x \) is

\[ f(y|x) = \frac{f(x, y)}{f_X(x)}, y \in R \]

Function \( f(y|x) \) is a p.d.f satisfying \( \int_{-\infty}^{\infty} f(y|x)dy = 1 \)

In estimation of parameter \( \theta \), we have a random sample \( X_1, \ldots, X_n \) from p.d.f \( f(x, \theta) \). The information we have about \( \theta \) is contained in \( X_1, \ldots, X_n \).

Let \( U = u(X_1, \ldots, X_n) \) be a statistic having p.d.f \( f_U(u, \theta) \)

The conditional p.d.f \( X_1, \ldots, X_n \) given \( U = u \) is

\[ f(x_1, \ldots, x_n|u) = \frac{f(x_1, \ldots, x_n, \theta)}{f_U(u, \theta)}, \{(x_1, \ldots, x_n) : u(x_1, \ldots, x_n) = u\} \]

Function \( f(x_1, \ldots, x_n|u) \) is a joint p.d.f with

\[ \int_{u(x_1, \ldots, x_n)=u} \cdots \int f(x_1, \ldots, x_n|u)dx_1 \cdots dx_n = 1 \]

Let \( X \) be r.v. and \( U = u(X) \)

\[ f(x|U = u) = \frac{f(x, u)}{f_U(u)} = \begin{cases} \frac{f_X(x)}{f_U(u)} & \text{if } u(X) = u \\ 0 & \text{if } u(X) \neq u \end{cases} \]

If, for any \( u \), conditional p.d.f \( f(x_1, \ldots, x_n, \theta|u) \) is unrelated to parameter \( \theta \), then the random sample \( X_1, \ldots, X_n \) contains no information about \( \theta \) when \( U = u \) is observed. This says that \( U \) contains exactly the same amount of information about \( \theta \) as \( X_1, \ldots, X_n \).

**Def.** Let \( X_1, \ldots, X_n \) be a random sample from a distribution with p.d.f \( f(x, \theta), \theta \in \Theta \). We call a statistic \( U = u(X_1, \ldots, X_n) \) a **sufficient statistic** if, for any value \( U = u \), the conditional p.d.f \( f(x_1, \ldots, x_n|u) \) and its domain all not
depend on parameter $\theta$.
Let $U = (X_1, \ldots, X_n)$. Then

$$f(x_1, \ldots, x_n, \theta | u = (x_1^*, \ldots, x_n^*)) = \begin{cases} \frac{f(x_1, \ldots, x_n, \theta)}{f(x_1^*, \ldots, x_n^*, \theta)} & \text{if } x_1 = x_1^*, x_2 = x_2^*, \ldots, x_n = x_n^* \\ 0 & \text{if } x_i \neq x_i^* \text{ for some } i \text{'s.} \end{cases}$$

Then $(X_1, \ldots, X_n)$ itself is a sufficient statistic of $\theta$.

Q: Why sufficiency?
A: We want a statistic with dimension as small as possible and contains information about $\theta$ the same amount as $X_1, \ldots, X_n$ does.

Def. If $U = u(X_1, \ldots, X_n)$ is a sufficient statistic with smallest dimension, it is called the minimal sufficient statistic.

Example:

(a) Let $(X_1, \ldots, X_n)$ be a random sample from a continuous distribution with p.d.f $f(x, \theta)$. Consider the order statistic $Y_1 = \min \{X_1, \ldots, X_n\}, \ldots, Y_n = \max \{X_1, \ldots, X_n\}$. If $Y_1 = y_1, \ldots, Y_n = y_n$ are observed, sample $X_1, \ldots, X_n$ have equal chance to have values in

$$\{(x_1, \ldots, x_n) : (x_1, \ldots, x_n) \text{ is a permutation of } (y_1, \ldots, y_n)\}.$$ 

Then the conditional joint p.d.f of $X_1, \ldots, X_n$ given $Y_1 = y_1, \ldots, Y_n = y_n$ is

$$f(x_1, \ldots, x_n, \theta | y_1, \ldots, y_n) = \begin{cases} \frac{1}{n!} & \text{if } x_1, \ldots, x_n \text{ is a permutation of } y_1, \ldots, y_n. \\ 0 & \text{otherwise.} \end{cases}$$

Then order statistic $(Y_1, \ldots, Y_n)$ is also a sufficient statistic of $\theta$. Order statistic is not a good sufficient statistic since it has dimension $n$.

(b) Let $X_1, \ldots, X_n$ be a random sample from Bernoulli distribution. The joint p.d.f of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n, p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}, x_i = 0, 1, i = 1, \ldots, n.$$ 

Consider the statistic $Y = \sum X_i$ which has binomial distribution $b(n, p)$ with p.d.f

$$f_Y(y, p) = \binom{n}{y} p^y (1-p)^{n-y}, y = 0, 1, \ldots, n$$
If $Y = y$, the space of $(X_1, \ldots, X_n)$ is \{$(x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = y$\}

The conditional p.d.f of $X_1, \ldots, X_n$ given $Y = y$ is

$$f(x_1, \ldots, x_n, p | y) = \begin{cases} \frac{n \sum_{i=1}^{n} x_i (1-p)^{n-\sum_{i=1}^{n} x_i}}{n^p (1-p)^{n-y}} & \text{if } \sum_{i=1}^{n} x_i = y \\ 0 & \text{if } \sum_{i=1}^{n} x_i \neq y \end{cases}$$

which is independent of $p$.

Hence, $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic of $p$ and is a minimal sufficient statistic.

(c) Let $X_1, \ldots, X_n$ be a random sample from uniform distribution $U(0, \theta)$.

Want to show that the largest order statistic $Y_n = \max\{X_1, \ldots, X_n\}$ is a sufficient statistic.

The joint p.d.f of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(0 < x_i < \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 < x_i < \theta)$$

$$= \begin{cases} \frac{1}{\theta^n} & \text{if } 0 < x_i < \theta, i = 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

The p.d.f of $Y_n$ is

$$f_{Y_n}(y, \theta) = n \left(\frac{y}{\theta}\right)^{n-1} \frac{1}{\theta} = n \frac{y^{n-1}}{\theta^n}, 0 < y < \theta$$

When $Y_n = y$ is given, $X_1, \ldots, X_n$ be values with $0 < x_i \leq y, i = 1, \ldots, n$.

The conditional p.d.f of $X_1, \ldots, X_n$ given $Y_n = y$ is

$$f(x_1, \ldots, x_n | y) = \frac{f(x_1, \ldots, x_n, \theta)}{f_{Y_n}(y, \theta)} = \begin{cases} \frac{1}{n y^{n-1}} & 0 < x_i \leq y, i = 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

$\Rightarrow$ independent of $\theta$.

So, $Y_n = \max\{X_1, \ldots, X_n\}$ is a sufficient statistic of $\theta$.

Q:
(a) If $U$ is a sufficient statistic, are $U+5$, $U^2$, $\cos(U)$ all sufficient for $\theta$?

(b) Is there easier way in finding sufficient statistic?

$T = t(X_1, \ldots, X_n)$ is sufficient for $\theta$ if conditional p.d.f $f(x_1, \ldots, x_n, \theta|t)$ is

indep. of $\theta$.

Independence:

1. function $f(x_1, \ldots, x_n, \theta|t)$ not depend on $\theta$.
2. domain of $X_1, \ldots, X_n$ not depend on $\theta$.

**Thm. Factorization Theorem.**

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

A statistic $U = u(X_1, \ldots, X_n)$ is sufficient for $\theta$ iff there exists functions

$K_1, K_2 \geq 0$ such that the joint p.d.f of $X_1, \ldots, X_n$ may be formulated as

$f(x_1, \ldots, x_n, \theta) = K_1(u(X_1, \ldots, X_n), \theta)K_2(x_1, \ldots, x_n)$ where $K_2$ is not a function of $\theta$.

**Proof.** Consider only the continuous r.v.'s.

$\Rightarrow$) If $U$ is sufficient for $\theta$, then

$$f(x_1, \ldots, x_n, \theta|u) = \frac{f(x_1, \ldots, x_n, \theta)}{f_U(u, \theta)}$$

is not a function of $\theta$

$$\Rightarrow f(x_1, \ldots, x_n, \theta) = f_U(u(X_1, \ldots, X_n), \theta)f(x_1, \ldots, x_n|u)$$

$$= K_1(u(X_1, \ldots, X_n), \theta)K_2(x_1, \ldots, x_n)$$

$\Leftarrow$) Suppose that $f(x_1, \ldots, x_n, \theta) = K_1(u(X_1, \ldots, X_n), \theta)K_2(x_1, \ldots, x_n)$

Let $Y_1 = u_1(X_1, \ldots, X_n), Y_2 = u_2(X_1, \ldots, X_n), \ldots, Y_n = u_n(X_1, \ldots, X_n)$ be a

1-1 function with inverse functions $x_1 = w_1(y_1, \ldots, y_n), x_2 = w_2(y_1, \ldots, y_n), \ldots, x_n = w_n(y_1, \ldots, y_n)$ and Jacobian

$$J = \begin{vmatrix}
\frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n}
\end{vmatrix}$$

(not depend on $\theta$)

The joint p.d.f of $Y_1, \ldots, Y_n$ is

$$f_{Y_1, \ldots, Y_n}(y_1, \ldots, y_n, \theta) = f(w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n), \theta)|J|$$

$$= K_1(y_1, \theta)K_2(w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n), \theta)|J|$$
The marginal p.d.f of $U = Y_1$ is
\[
f_U(y_1, \theta) = K_1(y_1, \theta) \int \cdots \int K_2(w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n)) |J| dy_2 \cdots dy_n
\]
not depend on $\theta$.

Then the conditional p.d.f of $X_1, \ldots, X_n$ given $U = u$ is
\[
f(x_1, \ldots, x_n, \theta | u) = \frac{f(x_1, \ldots, x_n, \theta)}{f_U(u, \theta)} = \frac{K_2(x_1, \ldots, x_n)}{\int \cdots \int K_2(w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n), \theta) |J| dy_2 \cdots dy_n}
\]
which is independent of $\theta$.

This indicates that $U$ is sufficient for $\theta$.

Example:

(a) $X_1, \ldots, X_n$ is a random sample from Poisson($\lambda$). Want sufficient statistic for $\lambda$.

Joint p.d.f of $X_1, \ldots, X_n$ is
\[
f(x_1, \ldots, x_n, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} = \lambda^{\sum x_i} e^{-n\lambda} \frac{1}{\prod_{i=1}^{n} x_i!}
\]
\[
= K_1(\sum_{i=1}^{n} x_i, \lambda) K_2(x_1, \ldots, x_n)
\]
\[
\Rightarrow \sum_{i=1}^{n} X_i \text{ is sufficient for } \lambda.
\]

We also have
\[
f(x_1, \ldots, x_n, \lambda) = \lambda^{\bar{x}} e^{-n\lambda} \frac{1}{\prod_{i=1}^{n} x_i!} = K_1(\bar{x}, \lambda) K_2(x_1, \ldots, x_n)
\]
\[
\Rightarrow \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ is sufficient for } \lambda.
\]

We also have
\[
f(x_1, \ldots, x_n, \lambda) = \lambda^{\frac{1}{2}} e^{-n\lambda} \frac{1}{\prod_{i=1}^{n} x_i!} = K_1(\bar{x}^2, \lambda) K_2(x_1, \ldots, x_n)
\]
\[
\Rightarrow \bar{X}^2 \text{ is sufficient for } \lambda.
\]
(b) Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Want sufficient statistic for $(\mu, \sigma^2)$.

Joint p.d.f of $X_1, \ldots, X_n$ is

\[ f(x_1, \ldots, x_n, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^\frac{n}{2}(\sigma^2)^\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \]

\[ \sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2 = (n-1)s^2 + n(\bar{x}-\mu)^2 \]

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n}(x_i - \bar{x})^2 \]

\[ f(x_1, \ldots, x_n, \mu, \sigma^2) = \frac{1}{(2\pi)^\frac{n}{2}(\sigma^2)^\frac{n}{2}} e^{-\frac{(n-1)s^2 + n(\bar{x}-\mu)^2}{2\sigma^2}} \cdot 1 = K_1(\bar{x}, s^2, \mu, \sigma^2)K_2(x_1, \ldots, x_n) \]

$\Rightarrow (\bar{x}, s^2)$ is sufficient for $(\mu, \sigma^2)$.

What is useful with a sufficient statistic for point estimation?

Review: $X, Y$ r.v.’s with join p.d.f $f(x, y)$.

Conditional p.d.f

\[ f(y|x) = \frac{f(x, y)}{f_X(x)} \Rightarrow f(x, y) = f(y|x)f_X(x) \]

\[ f(x|y) = \frac{f(x, y)}{f_Y(y)} \Rightarrow f(x, y) = f(x|y)f_Y(y) \]

Conditional expectation of $Y$ given $X = x$ is

\[ E(Y|x) = \int_{-\infty}^{\infty} y f(y|x) dy \]

The random conditional expectation $E(Y|X)$ is function $E(Y|x)$ with $x$ replaced by $X$.

Conditional variance of $Y$ given $X = x$ is

\[ \text{Var}(Y|x) = E[(Y - E(Y|x))^2|x] = E(Y^2|x) - (E(Y|x))^2 \]

The conditional variance $\text{Var}(Y|X)$ is $\text{Var}(Y|x)$ replacing $x$ by $X$.

**Thm.** Let $Y$ and $X$ be two r.v.’s.

(a) $E[E(Y|x)] = E(Y)$

(b) $\text{Var}(Y) = E(\text{Var}(Y|x)) + \text{Var}(E(Y|x))$
Proof. (a)

\[
E[E(Y|x)] = \int_{-\infty}^{\infty} E(Y|x) f_X(x) \, dx \\
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(y|x) dy f_X(x) \, dx \\
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) \, dxdy \\
= \int_{-\infty}^{\infty} y \left( \int_{-\infty}^{\infty} f(x,y) \, dx \right) dy \\
= \int_{-\infty}^{\infty} y f_Y(y) dy \\
= E(Y)
\]

(b)

\[
\text{Var}(Y|x) = E(Y^2|x) - (E(Y|x))^2 \\
\Rightarrow E(\text{Var}(Y|x)) = E[E(Y^2|x)] - E[(E(Y|x))^2] = E(Y^2) - E[(E(Y|x))^2]
\]

Also,
\[
\text{Var}(E(Y|x)) = E[(E(Y|x))^2] - E[(E(Y|x))]^2 \\
= E[(E(Y|x))^2] - (E(Y))^2
\]

\[
\Rightarrow E(\text{Var}(Y|x)) + \text{Var}(E(Y|x)) = E(Y^2) - (E(Y))^2 = \text{Var}(Y)
\]

Now, we comeback to the estimation of parameter function \( \tau(\theta) \). We have a random sample \( X_1, \ldots, X_n \) from \( f(x, \theta) \).

Lemma. Let \( \hat{\tau}(X_1, \ldots, X_n) \) be an unbiased estimator of \( \tau(\theta) \) and \( U = u(X_1, \ldots, X_n) \) is a statistic. Then

(a) \( E_\theta[\hat{\tau}(X_1, \ldots, X_n)|U] \) is unbiased for \( \tau(\theta) \)

(b) \( \text{Var}_\theta(E[\hat{\tau}(X_1, \ldots, X_n)|U]) \leq \text{Var}_\theta(\hat{\tau}(X_1, \ldots, X_n)) \)

Proof. (a)

\[
E_\theta[E(\hat{\tau}(X_1, \ldots, X_n)|U)] = E_\theta(\hat{\tau}(X_1, \ldots, X_n)) = \tau(\theta), \forall \theta \in \Theta.
\]

Then \( E_\theta[\hat{\tau}(X_1, \ldots, X_n)|U] \) is unbiased for \( \tau(\theta) \).

(b)

\[
\text{Var}_\theta(\hat{\tau}(X_1, \ldots, X_n)) = E_\theta[\text{Var}_\theta(\hat{\tau}(X_1, \ldots, X_n)|U)] + \text{Var}_\theta[E_\theta(\hat{\tau}(X_1, \ldots, X_n)|U)] \\
\geq \text{Var}_\theta[E_\theta(\hat{\tau}(X_1, \ldots, X_n)|U)], \forall \theta \in \Theta.
\]
Conclusions:

(a) For any estimator $\hat{\tau}(X_1, \ldots, X_n)$ which is unbiased for $\tau(\theta)$, and any statistic $U$, $E_\theta[\hat{\tau}(X_1, \ldots, X_n)|U]$ is unbiased for $\tau(\theta)$ and with variance smaller than or equal to $\hat{\tau}(X_1, \ldots, X_n)$.

(b) However, $E_\theta[\hat{\tau}(X_1, \ldots, X_n)|U]$ may not be a statistic. If it is not, it cannot be an estimator of $\tau(\theta)$.

(c) If $U$ is a sufficient statistic, $f(x_1, \ldots, x_n, \theta|u)$ is independent of $\theta$, then $E_\theta[\hat{\tau}(X_1, \ldots, X_n)|u]$ is independent of $\theta$. So, $E_\theta[\hat{\tau}(X_1, \ldots, X_n)|U]$ is an unbiased estimator.

If $U$ is not a sufficient statistic, $f(x_1, \ldots, x_n, \theta|u)$ is not only a function of $u$ but also a function of $\theta$, then $E_\theta[\hat{\tau}(X_1, \ldots, X_n)|u]$ is a function of $u$ and $\theta$. And $E_\theta[\hat{\tau}(X_1, \ldots, X_n)|u]$ is not a statistic.

**Thm. Rao-Blackwell**

If $\hat{\tau}(X_1, \ldots, X_n)$ is unbiased for $\tau(\theta)$ and $U$ is a sufficient statistic, then

(a) $E_\theta[\hat{\tau}(X_1, \ldots, X_n)|U]$ is a statistic.

(b) $E_\theta[\hat{\tau}(X_1, \ldots, X_n)|U]$ is unbiased for $\tau(\theta)$.

(c) $Var_\theta(E[\hat{\tau}(X_1, \ldots, X_n)|U]) \leq Var_\theta(\hat{\tau}(X_1, \ldots, X_n)), \forall \theta \in \Theta$.

If $\hat{\tau}(\theta)$ is an unbiased estimator for $\tau(\theta)$ and $U_1, U_2, \ldots$ are sufficient statistics, then we can improve $\hat{\tau}(\theta)$ with the following fact:

$$Var_\theta(E[\hat{\tau}(\theta)|U_1]) \leq Var_\theta(\hat{\tau}(\theta))$$
$$Var_\theta E(E(\hat{\tau}(\theta)|U_1|U_2) \leq Var_\theta E(\hat{\tau}(\theta)|U_1)$$
$$Var_\theta E[E(\hat{\tau}(\theta)|U_1)|U_2]|U_3 \leq Var_\theta E(E(\hat{\tau}(\theta)|U_1)|U_2)$$

$$\vdots$$

Will this process ends with Cramer-Rao lower bound?

This can be solved with “complete statistic”.

**Note:** Let $U$ be a statistic and $h$ is a function.

(a) If $h(U) = 0$ then $E_\theta(h(U)) = E_\theta(0) = 0, \forall \theta \in \Theta$. 
(b) If $P_\theta(h(U) = 0) = 1, \forall \theta \in \Theta$, $h(U)$ has a p.d.f

$$f_{h(U)}(h) = \begin{cases} 1 & \text{if } h = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then $E_\theta(h(U)) = \sum_{all \ h} h f_{h(U)}(h) = 0$.

**Def.** $X_1, \ldots, X_n$ is random sample from $f(x, \theta)$. A statistic $U = u(X_1, \ldots, X_n)$ is a complete statistic if for any function $h(U)$ such that $E_\theta(h(U)) = 0, \forall \theta \in \Theta$, then $P_\theta(h(U) = 0) = 1, \forall \theta \in \Theta$.

Q: For any statistic $U$, how can we verify if it is complete or not complete?

A:

1. To prove completeness, you need to show that for any function $h(U)$ with $0 = E_\theta(h(U)), \forall \theta \in \Theta$, the following $1 = P_\theta(h(U) = 0), \forall \theta \in \Theta$ hold.

2. To prove in-completeness, you need only to find one function $h(U)$ that satisfies $E_\theta(h(U)) = 0, \forall \theta \in \Theta$ and $P_\theta(h(U) = 0) < 1, \text{ for some } \theta \in \Theta$.

Examples:

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

Find a complete statistic and in-complete statistic?

sol: (a.1) We show that $Y = \sum_{i=1}^n X_i$ is a complete statistic. $Y \sim b(n, p)$.

Suppose that function $h(Y)$ satisfies $0 = E_p h(Y), \forall 0 < p < 1$.

Now,

$$0 = E_p h(Y) = \sum_{y=0}^n h(y) \binom{n}{y} p^y (1-p)^{n-y}$$

$$= (1-p)^n \sum_{y=0}^n h(y) \binom{n}{y} \left( \frac{p}{1-p} \right)^y, \forall 0 < p < 1$$

$$\Leftrightarrow 0 = \sum_{y=0}^n h(y) \binom{n}{y} \left( \frac{p}{1-p} \right)^y, \forall 0 < p < 1$$

(Let $\theta = \frac{p}{1-p}, 0 < p < 1 \Leftrightarrow 0 < \theta < \infty$)

$$\Leftrightarrow 0 = \sum_{y=0}^n h(y) \binom{n}{y} \theta^y, 0 < \theta < \infty$$
An order \( n+1 \) polynomial equation cannot have infinite solutions except that coefficients are zero’s.

\[ \Rightarrow h(y) \binom{n}{y} = 0, y = 0, \ldots, n \text{ for } 0 < \theta < \infty \]

\[ \Rightarrow h(y) = 0, y = 0, \ldots, n \text{ for } 0 < p < 1. \]

\[ \Rightarrow 1 = P_p(h(Y) = 0) \geq P_p(Y = 0, \ldots, n) = 1 \]

\[ \Rightarrow Y = \sum_{i=1}^{n} X_i \text{ is complete} \]

(a.2) We show that \( Z = X_1 - X_2 \) is not complete.

\[ E_p Z = E_p(X_1 - X_2) = E_p X_1 - E_p X_2 = p - p = 0, \forall 0 < p < 1 \]

\[ P_p(Z = 0) = P_p(X_1 - X_2 = 0) = P_p(X_1 = X_2 = 0 \text{ or } X_1 = X_2 = 1) \]

\[ = P_p(X_1 = X_2 = 0) + P_p(X_1 = X_2 = 1) \]

\[ = (1 - p)^2 + p^2 < 1 \text{ for } 0 < p < 1. \]

\[ \Rightarrow Z = X_1 - X_2 \text{ is not complete.} \]

(b) Let \( (X_1, \ldots, X_n) \) be a random sample from \( U(0, \theta) \).

We have to show that \( Y_n = \max\{X_1, \ldots, X_n\} \) is a sufficient statistic.

Here we use Factorization theorem to prove it again.

\[ f(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(0 < x_i < \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 < x_i < \theta, i = 1, \ldots, n) \]

\[ = \frac{1}{\theta^n} I(0 < y_n < \theta) \cdot 1 \]

\[ \Rightarrow Y_n \text{ is sufficient for } \theta \]

Now, we prove it complete.

The p.d.f of \( Y_n \) is

\[ f_{Y_n}(y) = n \left(\frac{y}{\theta}\right)^{n-1} \frac{1}{\theta} = \frac{n}{\theta^n} y^{n-1}, 0 < y < \theta \]
Suppose that $h(Y_n)$ satisfies $0 = E_\theta h(Y_n), \forall 0 < \theta < \infty$

$$0 = E_\theta h(Y_n) = \int_0^\theta h(y) \frac{n}{\theta^n} y^{n-1} dy = \frac{n}{\theta^n} \int_0^\theta h(y) y^{n-1} dy$$

$$\Leftrightarrow 0 = \int_0^\theta h(y) y^{n-1} dy, \forall \theta > 0$$

Taking differentiation both sides with $\theta$.

$$\Leftrightarrow 0 = h(\theta) \theta^{n-1}, \forall \theta > 0$$

$$\Leftrightarrow 0 = h(y), 0 < y < \theta, \forall \theta > 0$$

$$\Leftrightarrow P_\theta(h(Y_n) = 0) = P_\theta(0 < Y_n < \theta) = 1, \forall \theta > 0$$

$$\Rightarrow Y_n = \max\{X_1, \ldots, X_n\} \text{ is complete.}$$

**Def.** If the p.d.f of r.v. $X$ can be formulated as

$$f(x, \theta) = e^{a(x)b(\theta)+c(\theta)+d(x)}, l < x < q$$

where $l$ and $q$ do not depend on $\theta$, then we say that $f$ belongs to an exponential family.

**Thm.** Let $X_1, \ldots, X_n$ be a random sample from $f(x, \theta)$ which belongs to an exponential family as

$$f(x, \theta) = e^{a(x)b(\theta)+c(\theta)+d(x)}, l < x < q$$

Then $\sum_{i=1}^n a(X_i)$ is a complete and sufficient statistic.

Note: We say that $X = Y$ if $P(X = Y) = 1$.

**Thm. Lehmann-Scheffe**

Let $X_1, \ldots, X_n$ be a random sample from $f(x, \theta)$. Suppose that $U = u(X_1, \ldots, X_n)$ is a complete and sufficient statistic. If $\tilde{\tau} = t(U)$ is unbiased for $\tau(\theta)$, then $\hat{\tau}$ is the unique function of $U$ unbiased for $\tau(\theta)$ and is a UMVUE of $\tau(\theta)$.

(Unbiased function of complete and sufficient statistic is UMVUE.)

**Proof.** If $\hat{\tau}^* = t^*(U)$ is also unbiased for $\tau(\theta)$, then

$$E_\theta(\hat{\tau} - \hat{\tau}^*) = E_\theta(\hat{\tau}) - E_\theta(\hat{\tau}^*) = \tau(\theta) - \tau(\theta) = 0, \forall \theta \in \Theta.$$

$$\Rightarrow 1 = P_\theta(\hat{\tau} - \hat{\tau}^* = 0) = P(\hat{\tau} = \hat{\tau}^*), \forall \theta \in \Theta.$$

$$\Rightarrow \hat{\tau}^* = \hat{\tau}, \text{ unbiased function of } U \text{ is unique.}$$

If $T$ is any unbiased estimator of $\tau(\theta)$ then Rao-Blackwell theorem gives:
(a) $E(T|U)$ is unbiased estimator of $\tau(\theta)$.
By uniqueness, $E(T|U) = \hat{\tau}$ with probability 1.
(b) $\text{Var}_\theta(\hat{\tau}) = \text{Var}_\theta(E(T|U)) \leq \text{Var}_\theta(T), \forall \theta \in \Theta$.
This holds for every unbiased estimator $T$.
Then $\hat{\tau}$ is UMVUE of $\tau(\theta)$.

Two ways in constructing UMVUE based on a complete and sufficient statistic $U$:

(a) If $T$ is unbiased for $\tau(\theta)$, then $E(T|U)$ is the UMVUE of $\tau(\theta)$.
This is easy to define but difficult to transform it in a simple form.

(b) If there is a constant such that $E(U) = c \cdot \theta$, then $T = \frac{1}{c}U$ is the UMVUE of $\theta$.

Example:

(a) Let $X_1, \ldots, X_n$ be a random sample from $U(0, \theta)$.
Want UMVUE of $\theta$.

sol: $Y_n = \max\{X_1, \ldots, X_n\}$ is a complete and sufficient statistic.
The p.d.f of $Y_n$ is
\[
f_{Y_n}(y, \theta) = n\left(\frac{y}{\theta}\right)^{n-1} \frac{1}{\theta} = n \frac{y^{n-1}}{\theta^n}, 0 < y < \theta
\]

$E(Y_n) = \int_0^\theta y^n \frac{y^{n-1}}{\theta^n} dy = \frac{n}{n+1} \theta$.
We then have $E(\frac{n+1}{n}Y_n) = \frac{n+1}{n} E(Y_n) = \theta$.
So, $\frac{n+1}{n} Y_n$ is the UMVUE of $\theta$.

(b) Let $X_1, \ldots, X_n$ be a random sample from Bernoulli($p$).
Want UMVUE of $\theta$.

sol: The p.d.f is
\[
f(x, p) = p^x (1-p)^{1-x} = (1-p)\left(\frac{p}{1-p}\right)^x = e^{x \ln(\frac{p}{1-p}) + \ln(1-p)}
\]

$\Rightarrow \sum_{i=1}^n X_i$ is complete and sufficient.
$E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = np$
$\Rightarrow \hat{p} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ is UMVUE of $p$. 39
(c) $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$.
Want UMVUE of $\mu$.

sol: The p.d.f of $X$ is

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x^2-2\mu x+\mu^2)}{2}} = e^{\mu x - \frac{x^2}{2} - \ln \sqrt{2\pi}}$$

$$\Rightarrow \sum_{i=1}^{n} X_i \text{ is complete and sufficient.}$$

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = n\mu$$

$$\Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X} \text{ is UMVUE of } \mu.$$

Since $X_1$ is unbiased, we see that $E(X_1 | \sum_{i=1}^{n} X_i) = \bar{X}$

(d) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Possion}(\lambda)$.
Want UMVUE of $e^{-\lambda}$.

sol: The p.d.f of $X$ is

$$f(x, \lambda) = \frac{1}{x!} \lambda^x e^{-\lambda} = e^{x \ln \lambda - \lambda - \ln x!}$$

$$\Rightarrow \sum_{i=1}^{n} X_i \text{ is complete and sufficient.}$$

$$E(I(X_1 = 0)) = P(X_1 = 0) = f(0, \lambda) = e^{-\lambda} \text{ where } I(X_1 = 0) \text{ is an indicator function.}$$

$$\Rightarrow I(X_1 = 0) \text{ is unbiased for } e^{-\lambda}$$

$$\Rightarrow E(I(X_1 = 0) | \sum_{i=1}^{n} X_i) \text{ is UMVUE of } e^{-\lambda}.$$
Chapter 5. Confidence Interval

Let $Z$ be the r.v. with standard normal distribution $N(0, 1)$
We can find $z_\alpha$ and $z_{\alpha/2}$ that satisfy
\[
\alpha = P(Z \leq -z_\alpha) = P(Z \geq z_\alpha) \quad \text{and} \quad 1 - \alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}).
\]

A table of $z_{\alpha/2}$ is the following:

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>$z_{\alpha/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>1.28 ($z_{0.1}$)</td>
</tr>
<tr>
<td>0.9</td>
<td>1.645 ($z_{0.05}$)</td>
</tr>
<tr>
<td>0.95</td>
<td>1.96 ($z_{0.025}$)</td>
</tr>
<tr>
<td>0.99</td>
<td>2.58 ($z_{0.005}$)</td>
</tr>
<tr>
<td>0.9973</td>
<td>3 ($z_{0.00135}$)</td>
</tr>
</tbody>
</table>

**Def.** Suppose that we have a random sample from $f(x, \theta)$. For $0 < \alpha < 1$, if there exists two statistics $T_1 = t_1(X_1, \ldots, X_n)$ and $T_2 = t_2(X_1, \ldots, X_n)$ satisfying
\[
1 - \alpha = P(T_1 \leq \theta \leq T_2)
\]
We call the random interval $(T_1, T_2)$ a $100(1 - \alpha)$% confidence interval of parameter $\theta$. If $X_1 = x_1, \ldots, X_n = x_n$ is observed, we also call $(t_1(X_1, \ldots, X_n), t_2(X_1, \ldots, X_n))$ a $100(1 - \alpha)$% confidence interval (C.I.) for $\theta$

Constructing C.I. by pivotal quantity:

**Def.** A function of random sample and parameter, $Q = q(X_1, \ldots, X_n, \theta)$, is called a pivotal quantity if its distribution is independent of $\theta$

With a pivotal quantity $q(X_1, \ldots, X_n, \theta)$, there exists $a, b$ such that
\[
1 - \alpha = P(a \leq q(X_1, \ldots, X_n, \theta) \leq b), \forall \theta \in \Theta.
\]

The interest of pivotal quantity is that there exists statistics $T_1 = t_1(X_1, \ldots, X_n)$ and $T_2 = t_2(X_1, \ldots, X_n)$ with the following 1-1 transformation
\[
a \leq q(X_1, \ldots, X_n, \theta) \leq b \iff T_1 \leq \theta \leq T_2
\]
Then we have $1 - \alpha = P(T_1 \leq \theta \leq T_2)$ and $(T_1, T_2)$ is a $100(1 - \alpha)$% C.I. for $\theta$

Confidence Interval for Normal mean:
Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. We consider the C.I. of
parameter \( \mu \).

(I) \( \sigma = \sigma_0 \) is known

\[
X \sim N(\mu, \frac{\sigma_0^2}{n}) \Rightarrow \frac{X - \mu}{\sigma_0/\sqrt{n}} \sim N(0, 1)
\]

\[
1 - \alpha = P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}), Z \sim N(0, 1)
\]

\[
= P(-z_{\frac{\alpha}{2}} \leq \frac{X - \mu}{\sigma_0/\sqrt{n}} \leq z_{\frac{\alpha}{2}})
\]

\[
= P(-z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \leq X - \mu \leq z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}})
\]

\[
= P(\frac{X - \mu}{\frac{\sigma_0}{\sqrt{n}}} \leq \frac{\sigma_0}{\sqrt{n}} \leq X + \frac{\sigma_0}{\sqrt{n}})
\]

\[
\Rightarrow (X - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, X + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}) \text{ is a } 100(1 - \alpha)\% \text{ C.I. for } \mu.
\]

ex: \( n = 40, \sigma_0 = \sqrt{10}, \bar{x} = 7.164 \) (\( X_1, \ldots, X_{40} \iid N(\mu, 10) \).)

Want a 80% C.I. for \( \mu \).

sol: A 80% C.I. for \( \mu \) is

\[
(X - z_{\frac{0.8}{2}} \frac{\sigma_0}{\sqrt{n}}, X + z_{\frac{0.8}{2}} \frac{\sigma_0}{\sqrt{n}}) = (7.164 - 1.28 \frac{\sqrt{10}}{\sqrt{40}}, 7.164 + 1.28 \frac{\sqrt{10}}{\sqrt{40}})
\]

\[
= (6.523, 7.805)
\]

\[
P(X - z_{\frac{0.8}{2}} \frac{\sigma_0}{\sqrt{n}} \leq \mu \leq X + z_{\frac{0.8}{2}} \frac{\sigma_0}{\sqrt{n}}) = 1 - \alpha = 0.8
\]

\[
P(6.523 \leq \mu \leq 7.805) = 1 \text{ or } 0
\]

(II) \( \sigma \) is unknown.

**Def.** If \( Z \sim N(0, 1) \) and \( \chi^2(r) \) are independent, we call the distribution of the r.v.

\[
T = \frac{Z}{\sqrt{\frac{\chi^2(r)}{r}}}
\]

a \( t \)-distribution with \( r \) degrees of freedom.

The p.d.f of \( t \)-distribution is

\[
f_T(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} \frac{1}{\sqrt{\pi r}(1 + \frac{t^2}{r})^{\frac{r+1}{2}}}, -\infty < t < \infty
\]
∵ \( f_T(-t) = f_T(t) \)
∴ t-distribution is symmetric at 0.

Now \( X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \). We have

\[
\begin{align*}
\{ & \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \\
& \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \} \overset{indep.}{\Rightarrow} \left\{ \begin{array}{l}
\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \\
\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)
\end{array} \right. \overset{indep.}{\Rightarrow} \\
T = \frac{\bar{X} - \mu}{\frac{(n-1)s^2}{\sigma^2(n-1)}} = \frac{\bar{X} - \mu}{s / \sqrt{n}} \sim t(n-1)
\end{align*}
\]

Let \( t_{\frac{\alpha}{2}} \) satisfies

\[
1 - \alpha = P(-t_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{s / \sqrt{n}} \leq t_{\frac{\alpha}{2}})
= P(-t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \bar{X} - \mu \leq t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}})
= P(\bar{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}})
\Rightarrow (\bar{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}) \text{ is a } 100(1 - \alpha)\% \text{ C.I. for } \mu.
\]

ex: Suppose that we have \( n = 10, \bar{x} = 3.22 \) and \( s = 1.17 \). We also have \( t_{0.025} = 2.262 \). Want a 95\% C.I. for \( \mu \).

sol: A 95\% C.I. for \( \mu \) is

\[
(3.22 - 2.262 \frac{1.17}{\sqrt{10}}, 3.22 + 2.262 \frac{1.17}{\sqrt{10}}) = (2.34, 4.10)
\]