Mathmatical Statisticals

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:
$(1) P(A) \geq 0, \forall A \subset S.$
$(2) P(S) = 1.$
$(3) P(\bigcup_1^\infty A_i) = \sum_1^\infty P(A_i), if A_i \cap A_j = \varnothing, \forall i \neq j.$

Random variable X:
$X : S \to R$
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.

Let X be a r.v. with p.d.f $f(x)$. Let $g : R \to R$
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

$$G(y) = P(Y \le y) = P(g(X) \le y)$$

If G is differentiable then the p.d.f. of $Y = g(X)$ is $g(y) = G'(y)$.
(b) mgf method :(moment generating function)

$$E[e^{tx}] = \begin{cases} \sum e^{tx} f(x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & \text{(continuous)} \end{cases}$$

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

$$f(x_1, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n), \forall \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n$$

If $X_1, \ldots, X_n$ are continuous r.v.'s, there exists f such that

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n) dt_1 \ldots dt_n, \text{for} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in R^n$$

We call f the joint p.d.f. of $X_1, \ldots, X_n$.

If X is continuous, then

$$F(x) = \int_{-\infty}^{x} f(t)dt \text{ and } P(X = x) = \int_{x}^{x} f(t)dt = 0, \forall x \in R.$$

Marginal p.d.f's:
Discrete:

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

Continuous:

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

Events A and B are independent if $P(A \cap B) = P(A)P(B)$.
Q: If $A \cap B = \varnothing$, 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

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

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

$$\begin{aligned} g(y_1,\dots,y_m) &= P(Y_1=y_1,\dots,Y_m=y_m) \\ &= P(\{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : g_1(x_1,\dots,x_n)=y_1,\dots,g_m(x_1,\dots,x_n)=y_m\}) \\ &= \sum_{\{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : g_1(x_1,\dots,x_n)=y_1,\dots,g_m(x_1,\dots,x_n)=y_m\}} f(x_1,\dots,x_n) \end{aligned}$$

Example: joint p.d.f. of X_1, X_2, X_3 is

(x_1,x_2,x_3)	$(0,0,0)$	$(0,0,1)$	$(0,1,1)$	$(1,0,1)$	$(1,1,0)$	$(1,1,1)$
$f(x_1,x_2,x_3)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$$Y_1 = X_1+X_2+X_3, Y_2 = |X_3-X_2|$$

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

(y_1,y_2)	$(0,0)$	$(1,1)$	$(2,0)$	$(2,1)$	$(3,0)$
$g(y_1,y_2)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

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*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right| & y \in g(A) \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. The d.f. of Y is

$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$

(a) If g is $\nearrow$, g^{-1} is also $\nearrow$.$(\frac{dg^{-1}}{dy} > 0)$

$$F_Y(y) = P(X \le g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(x)dx$$

$\Rightarrow$ p.d.f. of Y is

$$\begin{aligned} f_Y(y) &= D_y \int_{-\infty}^{g^{-1}(y)} f_X(x)dx \\ &= f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} \\ &= f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}| \end{aligned}$$

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

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

$\Rightarrow$ p.d.f. of Y is

$$\begin{aligned} f_Y(y) &= D_y(1 - \int_{-\infty}^{g^{-1}(y)} f_X(x)dx) \\ &= -f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} \\ &= f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}| \end{aligned}$$

$\square$

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

$$f_Y(y) = f_X(g^{-1}(y))|\frac{dy}{dx}| = \frac{1}{2}e^{-\frac{y}{2}}, y > 0$$

$$(X \sim U(a,b) \text{ if } f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{elsewhere.} \end{cases})$$

$\Rightarrow Y \sim \chi^2(2)$

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

Continuous n-r.v.-to-m-r.v., $n > m$, case :

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longrightarrow \begin{cases} Y_1 = g_1(X_1, \ldots, X_n) \\ \vdots \\ Y_m = g_m(X_1, \ldots, X_n) \end{cases} \quad R^n \overset{\begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix}}{\longrightarrow} R^m$$

Q : What are the marginal p.d.f. of $Y_1, \cdots, Y_m$
A : We need to define $Y_{m+1} = g_{m+1}(X_1, \ldots, X_n), \cdots, Y_n = g_n(X_1, \ldots, X_n)$ such that $\begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix}$ is 1-1 from R^n to R^n.

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 $y_1 = g_1(x_1, \ldots, x_n), \cdots, y_n = g_n(x_1, \ldots, x_n)$ be a $1-1$ function with inverse $x_1 = w_1(y_1, \ldots, y_n), \cdots, x_n = w_n(y_1, \ldots, y_n)$ and Jacobian

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

Then

$$\begin{aligned} &\int \cdots \int f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \cdots dx_n \\ &= \int \cdots \int f_{X_1,\ldots,X_n}(w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n))|J| dy_1 \cdots dy_n \end{aligned}$$

Hence, joint p.d.f. of $Y_1, \cdots, Y_n$ is

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = f_{X_1,\ldots,X_n}(w_1, \ldots, w_n)|J|$$

Thm. *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*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} w_1(Y_1, Y_2) \\ w_2(Y_1, Y_2) \end{pmatrix} \text{ and Jacobian } J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

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

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

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 $\binom{Y_1}{Y_2} = \binom{g_1}{g_2}(A)$

Example : For $X_1, X_2 \overset{iid}{\sim} U(0, 1)$, 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

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

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

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

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

$$(1-1 \text{ transformation})$$

Jacobian is

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$$

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

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(w_1, w_2)|J|, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in B$$

Marginal p.d.f. of Y_1, Y_2 are

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

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

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

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

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 θ is called a* ***statistic****.*

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

For example, $N(\mu, \sigma^2)$: μ, σ^2 are parameters.
Poisson(λ) : λ is a parameter.

Example of statistics :
$X_1, \ldots, X_n$ are iid r.v.'s $\Rightarrow \overline{X}$ and S^2 are statistics.

<u>Note</u> : If $X_1, \ldots, X_n$ are r.v.'s, the m.g.f of $X_1, \ldots, X_n$ is

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

m.g.f

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

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

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,

$$\begin{aligned} M_{X_1,X_2}(t_1,t_2) &= \mathrm{E}(e^{t_1X_1+t_2X_2}) \\ &= \int\int e^{t_1x_1+t_2x_2} f(x_1,x_2)dx_1dx_2 \\ &= \int_{-\infty}^{\infty} e^{t_1x_1} f_{X_1}(x_1)dx_1 \int_{-\infty}^{\infty} e^{t_2x_2} f_{X_2}(x_2)dx_2 \\ &= \mathrm{E}(e^{t_1X_1})\mathrm{E}(e^{t_2X_2}) \\ &= M_{X_1}(t_1)M_{X_2}(t_2) \end{aligned}$$

$\Leftarrow$)

$$M_{X_1,X_2}(t_1,t_2) = \mathrm{E}(e^{t_1X_1+t_2X_2}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x_1+t_2x_2} f(x_1,x_2)dx_1dx_2$$

$$\begin{aligned} M_{X_1}(t_1)M_{X_2}(t_2) &= \mathrm{E}(e^{t_1X_1})\mathrm{E}(e^{t_2X_2}) \\ &= \int_{-\infty}^{\infty} e^{t_1x_1} f_{X_1}(x_1)dx_1 \int_{-\infty}^{\infty} e^{t_2x_2} f_{X_2}(x_2)dx_2 \\ &= \int_{-\infty}^{\infty} e^{t_1x_1+t_2x_2} f(x_1,x_2)dx_1dx_2 \end{aligned}$$

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. $\square$

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

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

Note :

(a) If $(X_1,\dots,X_n)$ and $(Y_1,\dots,Y_m)$ are independent, then $g(X_1,\dots,X_n)$ and $h(Y_1,\dots,Y_m)$ are also independent.

(b) If X , Y are independent, then

$$\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)].$$

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

$$\begin{aligned}
&(a)\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \\
&(b)\overline{X} \text{ and } S^2 \text{ are independent .} \\
&(c)\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)
\end{aligned}$$

Proof. (a) m.g.f. of $\overline{X}$ is

$$\begin{aligned}
M_{\overline{X}}(t) &= \mathrm{E}(e^{t\overline{X}}) = \mathrm{E}(e^{t\frac{1}{n}\sum_{i=1}^{n}X_i}) \\
&= \mathrm{E}(e^{\frac{t}{n}X_1}e^{\frac{t}{n}X_2}\cdots e^{\frac{t}{n}X_n}) \\
&= \mathrm{E}(e^{\frac{t}{n}X_1})\mathrm{E}(e^{\frac{t}{n}X_2})\mathrm{E}(e^{\frac{t}{n}X_n}) \\
&= M_{X_1}(\frac{t}{n})M_{X_2}(\frac{t}{n})\cdots M_{X_n}(\frac{t}{n}) \\
&= (e^{\mu\frac{t}{n}+\frac{\sigma^2}{2}(\frac{t}{n})^2})^n \\
&= e^{\mu t+\frac{\sigma^2/n}{2}t^2}
\end{aligned}$$

$\Rightarrow \overline{X} \sim (\mu, \frac{\sigma^2}{n})$
(b) First we want to show that $\overline{X}$ and $(X_1-\overline{X}, X_2-\overline{X},\dots,X_n-\overline{X})$ are

independent. Joint m.g.f. of $\overline{X}$ and $(X_1-\overline{X}, X_2-\overline{X}, \ldots, X_n-\overline{X})$ is

$$
\begin{aligned}
&M_{\overline{X},X_1-\overline{X},X_2-\overline{X},\ldots,X_n-\overline{X}}(t,t_1,\ldots,t_n)\\
&= \mathrm{E}[e^{t\overline{X}+t_1(X_1-\overline{X})+\cdots+t_n(X_n-\overline{X})}]\\
&= \mathrm{E}[e^{\frac{t}{n}\sum_{i=1}^n X_i+\sum_{i=1}^n t_iX_i-\sum_{i=1}^n t_i\frac{\sum X_i}{n}}]\\
&= \mathrm{E}[e^{\sum_{i=1}^n(\frac{t}{n}+t_i-\bar{t})X_i}], \bar{t}=\frac{1}{n}\sum_{i=1}^n t_i\\
&= \mathrm{E}[e^{\sum_{i=1}^n \frac{n(t_i-\bar{t})+t}{n}X_i}]\\
&= \mathrm{E}[\prod_{i=1}^n e^{\frac{n(t_i-\bar{t})+t}{n}X_i}]\\
&= \prod_{i=1}^n e^{\mu\frac{n(t_i-\bar{t})+t}{n}+\frac{\sigma^2}{2}\frac{(n(t_i-\bar{t})+t)^2}{n^2}}\\
&= e^{\frac{\mu}{n}\sum\limits_{i=1}^n(n(t_i-\bar{t})+t)+\frac{\sigma^2}{2n^2}\sum\limits_{i=1}^n(n(t_i-\bar{t})+t)^2}\\
&= e^{\mu t+\frac{\sigma^2/n}{2}t^2+\mu\sum(t_i-\bar{t})+\frac{\sigma^2}{2}\sum(t_i-\bar{t})^2+\frac{\sigma^2}{n^2}nt\sum(t_i-\bar{t})}\\
&= e^{\mu t+\frac{\sigma^2/n}{2}t^2}e^{\frac{\sigma^2}{2}\sum(t_i-\bar{t})^2}\\
&= M_{\overline{X}}(t)M_{(X_1-\overline{X},X_2-\overline{X},\ldots,X_n-\overline{X})}(t_1,\ldots,t_n)
\end{aligned}
$$

$\Rightarrow \overline{X}$ and $(X_1-\overline{X}, X_2-\overline{X}, \ldots, X_n-\overline{X})$ are independent.
$\Rightarrow \overline{X}$ and $S^2=\frac{1}{n-1}\sum\limits_{i=1}^n(X_i-\overline{X})^2$ are independent.

(c)

(1) $Z\sim N(0,1), \Rightarrow Z^2\sim\chi^2(1)$

(2)

$$X\sim\chi^2(r_1) \text{ and } Y\sim\chi^2(r_2) \text{ are independent. } \Rightarrow X+Y\sim\chi^2(r_1+r_2)$$

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

$$
\begin{aligned}
M_{X+Y}(t) &= \mathrm{E}(e^{t(X+Y)}) = \mathrm{E}(e^{tX+tY}) = \mathrm{E}(e^{tX})\mathrm{E}(e^{tY}) = M_X(t)M_Y(t)\\
&= (1-2t)^{-\frac{r_1}{2}}(1-2t)^{-\frac{r_2}{2}} = (1-2t)^{-\frac{r_1+r_2}{2}}
\end{aligned}
$$

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

(3)

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

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

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

$$\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2} = \frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\begin{aligned}
(1-2t)^{-\frac{n}{2}} &= M_{\frac{\sum(X_i-\mu)^2}{\sigma^2}}(t) = \mathrm{E}(e^{t\frac{\sum(X_i-\mu)^2}{\sigma^2}}) \\
&= \mathrm{E}(e^{t\frac{\sum(X_i-\overline{X}+\overline{X}-\mu)^2}{\sigma^2}}) = \mathrm{E}(e^{t\frac{\sum(X_i-\overline{X})^2+n(\overline{X}-\mu)^2}{\sigma^2}}) \\
&= \mathrm{E}(e^{t\frac{(n-1)s^2}{\sigma^2}} e^{t\frac{(\overline{X}-\mu)^2}{\sigma^2/n}}) \\
&= \mathrm{E}(e^{t\frac{(n-1)s^2}{\sigma^2}})\mathrm{E}(e^{t\frac{(\overline{X}-\mu)^2}{\sigma^2/n}}) \\
&= M_{\frac{(n-1)s^2}{\sigma^2}}(t) M_{\frac{(\overline{X}-\mu)^2}{\sigma^2/n}}(t) \\
&= M_{\frac{(n-1)s^2}{\sigma^2}}(t)(1-2t)^{-\frac{1}{2}}
\end{aligned}$$

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

□

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 θ is unknown. We want to gain knowledge about θ.
What we have for inference:
There is a random sample $X_1, \ldots, X_n$ from $f(x, \theta)$.

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

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

Two problems are concerned in estimation of θ :

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

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

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

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

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

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

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

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

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

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

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

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

Thm.

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

Proof.

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

Chebyshev's Inequality :

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

For $\epsilon > 0$,

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

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

□

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

Proof.

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

□

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example :

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

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

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

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

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

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

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

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

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

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

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

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

<u>Note</u> :

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

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

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

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

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

Two cases to solve m.l.e. :

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

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

Order statistics:

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

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

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

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

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

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

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

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

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

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

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

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

□

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

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

Consider the indicator function

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

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

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

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

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

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

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

The p.d.f. of Y is

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

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

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

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

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

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

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

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

The likelihood function is

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

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

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

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

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

The likelihood function is

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

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

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

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

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

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

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

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

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

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

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

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

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

Example :

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

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

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

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

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

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

Best estimator :

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

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

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

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

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

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

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

(d) If $T = t(x_1,\dots,x_n)$ 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 $\hat\tau(\theta) = t(X_1,\dots,X_n)$ *is unbiased for* $\tau(\theta)$*, then*
$$Var_\theta\,\hat\tau(\theta) \ge \frac{(\tau'(\theta))^2}{nE_\theta\,[(\frac{\partial\ln\,f(x,\theta)}{\partial\theta})^2]} = \frac{(\tau'(\theta))^2}{-nE_\theta\,[(\frac{\partial^2\ln\,f(x,\theta)}{\partial\theta^2})]}\ \textit{for}\ \theta\in\Theta$$

Proof. Consider only the continuous distribution.

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

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

Taking derivatives both sides.

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

Now,

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

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

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

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

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

Cauchy-Swartz Inequality

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

Then

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

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

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

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

Since

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

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

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

Then, we have

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

You may further check that

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

□

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

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

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

Example:

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

$$\begin{aligned} &\Rightarrow \ln f(x,\lambda) = x\ln\lambda - \lambda - \ln x! \\ &\Rightarrow \frac{\partial}{\partial\lambda}\ln f(x,\lambda) = \frac{x}{\lambda} - 1 \\ &\Rightarrow \frac{\partial^2}{\partial\lambda^2}\ln f(x,\lambda) = -\frac{x}{\lambda^2} \\ \mathrm{E}(&\frac{\partial^2}{\partial\lambda^2}\ln f(x,\lambda)) = \mathrm{E}(-\frac{x}{\lambda^2}) = -\frac{\mathrm{E}(X)}{\lambda^2} = -\frac{1}{\lambda} \end{aligned}$$

Cramer-Rao lower bound is

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

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

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

Want UMVUE of p.

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

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

C-R lower bound for p is

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

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

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

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

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 θ, we have a random sample $X_1, \ldots, X_n$ from p.d.f $f(x, \theta)$. The information we have about θ 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 $U = u(X)$

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

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

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

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

Q: Why sufficiency?
A: We want a statistic with dimension as small as possible and contains information about θ 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 θ.
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 $Y = \sum_{i=1}^{n} X_i$ 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 $(X_1,\dots,X_n)$ is $\{(x_1,\dots,x_n):\sum_{i=1}^n x_i=y\}$
The conditional p.d.f of $X_1,\dots,X_n$ given $Y=y$ is

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

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 $X_1,\dots,X_n$ be a random sample from uniform distribution $U(0,\theta)$.
Want to show that the largest order statistic $Y_n=\max\{X_1,\dots,X_n\}$ is a sufficient statistic.
The joint p.d.f of $X_1,\dots,X_n$ is

$$f(x_1,\dots,x_n,\theta)=\prod_{i=1}^n\frac{1}{\theta}I(0<x_i<\theta)=\frac{1}{\theta^n}\prod_{i=1}^n I(0<x_i<\theta)$$
$$=\begin{cases}\frac{1}{\theta^n} & \text{if } 0<x_i<\theta, i=1,\dots,n\\ 0 & \text{otherwise.}\end{cases}$$

The p.d.f of Y_n is

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

When $Y_n=y$ is given, $X_1,\dots,X_n$ be values with $0<x_i\le y, i=1,\dots,n$
The conditional p.d.f of $X_1,\dots,X_n$ given $Y_n=y$ is

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

$\Rightarrow$ independent of θ.
So, $Y_n=\max\{X_1,\dots,X_n\}$ is a sufficient statistic of θ.

Q:

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

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

$T = t(X_1, \dots, X_n)$ is sufficient for θ if conditional p.d.f $f(x_1, \dots, x_n, \theta|t)$ is indep. of θ.
Independence:
1.function $f(x_1, \dots, x_n, \theta|t)$ not depend on θ.
2.domain of $X_1, \dots, X_n$ not depend on θ.

Thm. ***Factorization Theorem.***
Let $X_1, \dots, X_n$ be a random sample from a distribution with p.d.f $f(x, \theta)$. A statistic $U = u(X_1, \dots, X_n)$ is sufficient for θ iff there exists functions $K_1, K_2 \geq 0$ such that the joint p.d.f of $X_1, \dots, X_n$ may be formulated as $f(x_1, \dots, x_n, \theta) = K_1(u(X_1, \dots, X_n), \theta)K_2(x_1, \dots, x_n)$ where K_2 is not a function of θ.

Proof. Consider only the continuous r.v's.
$\Rightarrow$) If U is sufficient for θ, then

$$\begin{aligned} f(x_1, \dots, x_n, \theta|u) &= \frac{f(x_1, \dots, x_n, \theta)}{f_U(u, \theta)} \text{ is not a function of } \theta \\ \Rightarrow f(x_1, \dots, x_n, \theta) &= f_U(u(X_1, \dots, X_n), \theta)f(x_1, \dots, x_n|u) \\ &= K_1(u(X_1, \dots, X_n), \theta)K_2(x_1, \dots, x_n) \end{aligned}$$

$\Leftarrow$) Suppose that $f(x_1, \dots, x_n, \theta) = K_1(u(X_1, \dots, X_n), \theta)K_2(x_1, \dots, x_n)$
Let $Y_1 = u_1(X_1, \dots, X_n), Y_2 = u_2(X_1, \dots, X_n), \dots, Y_n = u_n(X_1, \dots, X_n)$ be a 1-1 function with inverse functions $x_1 = w_1(y_1, \dots, y_n), x_2 = w_2(y_1, \dots, y_n), \dots, x_n = w_n(y_1, \dots, y_n)$ and Jacobian

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} \text{ (not depend on } \theta.)$$

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

$$\begin{aligned} f_{Y_1, \dots, Y_n}(y_1, \dots, y_n, \theta) &= f(w_1(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n), \theta)|J| \\ &= K_1(y_1, \theta)K_2(w_1(y_1, \dots, y_n), \dots, w_n(y_1, \dots, y_n), \theta)|J| \end{aligned}$$

The marginal p.d.f of $U = Y_1$ is

$$f_U(y_1,\theta) = K_1(y_1,\theta)\underbrace{\int\cdots\int K_2(w_1(y_1,\ldots,y_n),\ldots,w_n(y_1,\ldots,y_n))|J|dy_2\cdots dy_n}_{\text{not depend on } \theta.}$$

Then the conditional p.d.f of $X_1,\ldots,X_n$ given $U = u$ is

$$\begin{aligned} f(x_1,\ldots,x_n,\theta|u) &= \frac{f(x_1,\ldots,x_n,\theta)}{f_U(u,\theta)} \\ &= \frac{K_2(x_1,\ldots,x_n)}{\int\cdots\int K_2(w_1(y_1,\ldots,y_n),\ldots,w_n(y_1,\ldots,y_n),\theta)|J|dy_2\cdots dy_n} \end{aligned}$$

which is independent of θ.
This indicates that U is sufficient for θ. □

Example :

(a) $X_1,\ldots,X_n$ is a random sample from Poisson(λ).Want sufficient statistic for λ.
Joint p.d.f of $X_1,\ldots,X_n$ is

$$\begin{aligned} f(x_1,\ldots,x_n,\lambda) = \prod_{i=1}^{n}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!} &= \frac{\lambda^{\sum x_i}e^{-n\lambda}}{\prod\limits_{i=1}^{n}x_i!} = \lambda^{\sum x_i}e^{-n\lambda}\frac{1}{\prod\limits_{i=1}^{n}x_i!} \\ &= K_1(\sum_{i=1}^{n}x_i,\lambda)K_2(x_1,\ldots,x_n) \end{aligned}$$

$\Rightarrow \sum\limits_{i=1}^{n}X_i$ is sufficient for λ.
We also have

$$f(x_1,\ldots,x_n,\lambda) = \lambda^{n\overline{x}}e^{-n\lambda}\frac{1}{\prod\limits_{i=1}^{n}x_i!} = K_1(\overline{x},\lambda)K_2(x_1,\ldots,x_n)$$

$\Rightarrow \overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n}X_i$ is sufficient for λ.
We also have

$$f(x_1,\ldots,x_n,\lambda) = \lambda^{n(\overline{x}^2)^{\frac{1}{2}}}e^{-n\lambda}\frac{1}{\prod\limits_{i=1}^{n}x_i!} = K_1(\overline{x}^2,\lambda)K_2(x_1,\ldots,x_n)$$

$\Rightarrow \overline{X}^2$ is sufficient for λ.

(b)Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$.Want sufficient statistic for (μ, σ^2).

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

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

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-\mu)^2 = \sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\mu)^2 = (n-1)s^2+n(\overline{x}-\mu)^2$$

$$(s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2)$$

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

$\Rightarrow (\overline{X}, s^2)$ is sufficient for (μ, σ^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

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

The random conditional expectation $\mathrm{E}(Y|X)$ is function $\mathrm{E}(Y|x)$ with x replaced by X.
Conditional variance of Y given $X = x$ is

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

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

Thm. *Let Y and X be two r.v.'s.*
(a) $E[E(Y|x)] = E(Y)$
(b) $Var(Y) = E(Var(Y|x)) + Var(E(Y|x)$

Proof. (a)

$$\begin{aligned}
\mathrm{E}[\mathrm{E}(Y|x)] &= \int_{-\infty}^{\infty} \mathrm{E}(Y|x) f_X(x)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(y|x)dyf_X(x)dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy \\
&= \int_{-\infty}^{\infty} y(\int_{-\infty}^{\infty} f(x,y)dx)dy \\
&= \int_{-\infty}^{\infty} yf_Y(y)dy \\
&= \mathrm{E}(Y)
\end{aligned}$$

(b)

$$\begin{aligned}
\mathrm{Var}(Y|x) &= \mathrm{E}(Y^2|x) - (\mathrm{E}(Y|x))^2 \\
\Rightarrow \mathrm{E}(\mathrm{Var}(Y|x)) &= \mathrm{E}[\mathrm{E}(Y^2|x)] - \mathrm{E}[(\mathrm{E}(Y|x))^2] = \mathrm{E}(Y^2) - \mathrm{E}[(\mathrm{E}(Y|x))^2] \\
\text{Also}, \mathrm{Var}(\mathrm{E}(Y|x) &= \mathrm{E}[(\mathrm{E}(Y|x))^2] - \mathrm{E}[(\mathrm{E}(Y|x))]^2 \\
&= \mathrm{E}[(\mathrm{E}(Y|x))^2] - (\mathrm{E}(Y))^2
\end{aligned}$$

$$\Rightarrow \mathrm{E}(\mathrm{Var}(Y|x)) + \mathrm{Var}(\mathrm{E}(Y|x) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 = \mathrm{Var}(Y)$$

□

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

Proof. (a)

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

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

$$\begin{aligned}
\mathrm{Var}_\theta(\hat{\tau}(X_1, \ldots, X_n)) &= \mathrm{E}_\theta[\mathrm{Var}_\theta(\hat{\tau}(X_1, \ldots, X_n)|U)] + \mathrm{Var}_\theta[\mathrm{E}_\theta(\hat{\tau}(X_1, \ldots, X_n)|U) \\
&\geq \mathrm{Var}_\theta[\mathrm{E}_\theta(\hat{\tau}(X_1, \ldots, X_n)|U)], \forall\theta \in \Theta.
\end{aligned}$$

□

Conclusions:

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

(b) However, $\mathrm{E}_\theta[\hat{\tau}(X_1, \ldots, X_n)|U]$ 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 θ, then $\mathrm{E}_\theta[\hat{\tau}(X_1, \ldots, X_n)|u]$ is independent of θ. So, $\mathrm{E}_\theta[\hat{\tau}(X_1, \ldots, X_n)|U]$ 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 θ, then $\mathrm{E}_\theta[\hat{\tau}(X_1, \ldots, X_n)|u]$ is a function of u and θ. And $\mathrm{E}_\theta[\hat{\tau}(X_1, \ldots, X_n)|u]$ 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:

$$\begin{aligned}
\mathrm{Var}_\theta(\mathrm{E}[\hat{\tau}(\theta)|U_1]) &\leq \mathrm{Var}_\theta\hat{\tau}(\theta) \\
\mathrm{Var}_\theta\mathrm{E}(\mathrm{E}(\hat{\tau}(\theta)|U_1)|U_2) &\leq \mathrm{Var}_\theta\mathrm{E}(\hat{\tau}(\theta)|U_1) \\
\mathrm{Var}_\theta\mathrm{E}[\mathrm{E}(\mathrm{E}(\hat{\tau}(\theta)|U_1)|U_2)|U_3] &\leq \mathrm{Var}_\theta\mathrm{E}(\mathrm{E}(\hat{\tau}(\theta)|U_1)|U_2) \\
&\vdots
\end{aligned}$$

Will this process ends with Cramer-Rao lower bound ?
This can be solved with "complete statistic".

<u>Note</u>: Let U be a statistic and h is a function.

(a) If $h(U) = 0$ then $\mathrm{E}_\theta(h(U)) = \mathrm{E}_\theta(0) = 0, \forall\theta \in \Theta$.

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

$$f_{h(U)}(h)=\begin{cases}1 & ,\text{if } h=0\\ 0 & ,\text{otherwise.}\end{cases} \quad \text{Then } \mathrm{E}_\theta(h(U))=\sum_{\text{all } h} h f_{h(U)}(h)=0$$

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 $P_\theta(h(U)=0)=1, for\ \theta\in\Theta$.*

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 $0=\mathrm{E}_\theta(h(U)), \forall\theta\in\Theta$.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 $\mathrm{E}_\theta(h(U))=0, \forall\theta\in\Theta$ and $P_\theta(h(U)=0)<1$, for some $\theta\in\Theta$.

Examples:

(a)$X_1,\ldots,X_n \overset{iid}{\sim}$ 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 $0=\mathrm{E}_p h(Y), \forall 0<p<1$
Now,

$$\begin{aligned}0=\mathrm{E}_p h(Y) &= \sum_{y=0}^n h(y)\binom{n}{y}p^y(1-p)^{n-y}\\ &=(1-p)^n\sum_{y=0}^n h(y)\binom{n}{y}(\frac{p}{1-p})^y, \forall 0<p<1\end{aligned}$$

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

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

$$\Leftrightarrow 0=\sum_{y=0}^n h(y)\binom{n}{y}\theta^y, 0<\theta<\infty$$

An order $n+1$ polynomial equation cannot have infinite solutions except that coefficients are zero's.

$$\Rightarrow h(y)\binom{n}{y} = 0, y = 0, \dots, n \text{ for } 0 < \theta < \infty$$
$$\Rightarrow h(y) = 0, y = 0, \dots, n \text{ for } 0 < p < 1.$$
$$\Rightarrow 1 \geq P_p(h(Y) = 0) \geq P_p(Y = 0, \dots, n) = 1$$
$$\Rightarrow Y = \sum_{i=1}^{n} X_i \text{ is complete}$$

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

$$\mathrm{E}_p Z = \mathrm{E}_p(X_1 - X_2) = \mathrm{E}_p X_1 - \mathrm{E}_p X_2 = p - p = 0, \forall 0 < p < 1$$

$$\begin{aligned} P_p(Z = 0) = P_p(X_1 - X_2 = 0) &= P_p(X_1 = X_2 = 0 \text{ or } X_1 = X_2 = 1) \\ &= P_p(X_1 = X_2 = 0) + P_p(X_1 = X_2 = 1) \\ &= (1-p)^2 + p^2 < 1 \text{ for } 0 < p < 1. \end{aligned}$$

$\Rightarrow Z = X_1 - X_2$ is not complete.

(b)Let $(X_1, \dots, X_n)$ be a random sample from $U(0, \theta)$.
We have to show that $Y_n = \max\{X_1, \dots, X_n\}$ is a sufficient statistic.
Here we use Factorization theorem to prove it again.

$$\begin{aligned} f(x_1, \dots, x_n, \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} I(0 < x_i < \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 < x_i < \theta, i = 1, \dots, n) \\ &= \frac{1}{\theta^n} I(0 < y_n < \theta) \cdot 1 \end{aligned}$$

$\Rightarrow Y_n$ is sufficient for θ
Now, we prove it complete.
The p.d.f of Y_n is

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

Suppose that $h(Y_n)$ satisfies $0 = \mathrm{E}_\theta h(Y_n), \forall 0 < \theta < \infty$

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

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

Taking differentiation both sides with θ.

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

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

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

$\Rightarrow Y_n = \max\{X_1, \dots, X_n\}$ 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 θ, then we say that f belongs to an exponential family.

Thm. *Let $X_1, \dots, X_n$ 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 $X_1, \dots, X_n$ be a random sample from $f(x, \theta)$. Suppose that $U = u(X_1, \dots, X_n)$ is a complete and sufficient statistic. If $\hat{\tau} = t(U)$ 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

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

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

$\Rightarrow \hat{\tau}^* = \hat{\tau}$, unbiased function of U is unique.
If T is any unbiased estimator of $\tau(\theta)$ then Rao-Blackwell theorem gives:

(a) $\mathrm{E}(T|U)$ is unbiased estimator of $\tau(\theta)$.
By uniqueness, $\mathrm{E}(T|U) = \hat{\tau}$ with probability 1.
(b) $\mathrm{Var}_\theta(\hat{\tau}) = \mathrm{Var}_\theta(\mathrm{E}(T|U)) \leq \mathrm{Var}_\theta(T), \forall\theta \in \Theta$.
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 $\mathrm{E}(T|U)$ 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 $\mathrm{E}(U) = c\cdot\theta$, then $T = \frac{1}{c}U$ is the UMVUE of θ.

Example :

(a)Let $X_1, \ldots, X_n$ be a random sample from $U(0, \theta)$.
Want UMVUE of θ.
sol: $Y_n = \max\{X_1, \ldots, X_n\}$ is a complete and sufficient statistic .
The p.d.f of Y_n is

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

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

We then have $\mathrm{E}(\frac{n+1}{n}Y_n) = \frac{n+1}{n}\mathrm{E}(Y_n) = \theta$.
So, $\frac{n+1}{n}Y_n$ is the UMVUE of θ.

(b) Let $X_1, \ldots, X_n$ be a random sample from Bernoulli(p).
Want UMVUE of θ.
sol: The p.d.f is

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

$\Rightarrow \sum_{i=1}^n X_i$ is complete and sufficient.
$\mathrm{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{E}(X_i) = np$
$\Rightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}$ is UMVUE of p.

(c)$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$.

Want UMVUE of μ.

sol: The p.d.f of X is

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

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

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

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

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

(d)$X_1, \ldots, X_n \overset{iid}{\sim}$ Possion(λ).

Want UMVUE of $e^{-\lambda}$.

sol: The p.d.f of X is

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

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

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

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

$\Rightarrow \mathrm{E}(I(X_1 = 0) | \sum_{i=1}^{n} X_i)$ 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_α and $z_{\frac{\alpha}{2}}$ that satisfy

$$\alpha = P(Z \le -z_\alpha) = P(Z \ge z_\alpha) \text{ and } 1-\alpha = P(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}).$$

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

$1-\alpha$	$z_{\frac{\alpha}{2}}$
0.8	1.28 ($z_{0.1}$)
0.9	1.645 ($z_{0.05}$)
0.95	1.96 ($z_{0.025}$)
0.99	2.58 ($z_{0.005}$)
0.9973	3 ($z_{0.00135}$)

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*

$$1-\alpha = P(T_1 \le \theta \le T_2)$$

We call the random interval (T_1,T_2) a $100(1-\alpha)\%$ confidence interval of parameter θ. 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 θ

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 θ*

With a pivotal quantity $q(X_1,\ldots,X_n,\theta)$, there exists a,b such that

$$1-\alpha = P(a \le q(X_1,\ldots,X_n,\theta) \le b), \forall \theta \in \Theta.$$

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

$$a \le q(X_1,\ldots,X_n,\theta) \le b \text{ iff } T_1 \le \theta \le T_2$$

Then we have $1-\alpha = P(T_1 \le \theta \le T_2)$ and (T_1,T_2) is a $100(1-\alpha)\%$ C.I. for θ
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 μ.
(I) $\sigma = \sigma_0$ is known

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

$$\begin{aligned}
1 - \alpha &= P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}), Z \sim N(0,1) \\
&= P(-z_{\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma_0/\sqrt{n}} \leq z_{\frac{\alpha}{2}}) \\
&= P(-z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \leq \overline{X} - \mu \leq z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}) \\
&= P(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}})
\end{aligned}$$

$\Rightarrow (\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}})$ is a $100(1-\alpha)\%$ C.I. for μ.
ex: $n = 40, \sigma_0 = \sqrt{10}, \overline{x} = 7.164$ ($X_1, \ldots, X_{40} \overset{iid}{\sim} N(\mu, 10)$.)
Want a 80% C.I. for μ.
sol: A 80% C.I. for μ. is

$$\begin{aligned}
(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}) &= (7.164 - 1.28\frac{\sqrt{10}}{\sqrt{40}}, 7.164 + 1.28\frac{\sqrt{10}}{\sqrt{40}}) \\
&= (6.523, 7.805)
\end{aligned}$$

$$P(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}) = 1 - \alpha = 0.8$$
$$P(6.523 \leq \mu \leq 7.805) = 1 \text{ or } 0$$

(II)σ 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

$$f_T(t) = \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})} \frac{1}{\sqrt{r\pi}(1 + \frac{t^2}{r})^{\frac{r+1}{2}}}, -\infty < t < \infty$$

$\because f_T(-t) = f_T(t)$
$\therefore$ t-distribution is symmetric at 0.
Now $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. We have

$$\begin{cases} \overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \\ \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \end{cases} indep. \Rightarrow \begin{cases} \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \\ \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \end{cases} indep.$$

$$T = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2(n-1)}}} = \frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

Let $t_{\frac{\alpha}{2}}$ satisfies

$$\begin{aligned} 1-\alpha &= P(-t_{\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu}{s/\sqrt{n}} \leq t_{\frac{\alpha}{2}}) \\ &= P(-t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \leq \overline{X}-\mu \leq t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}) \\ &= P(\overline{X}-t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X}+t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}) \end{aligned}$$

$\Rightarrow (\overline{X}-t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \overline{X}+t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}})$ is a $100(1-\alpha)\%$ C.I. for μ.

ex: Suppose that we have $n = 10, \overline{x} = 3.22$ and $s = 1.17$. We also have $t_{0.025} = 2.262$. Want a 95% C.I. for μ.
sol: A 95% C.I. for μ is

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