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 $(X_1, \ldots, X_n)$ 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

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

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 $X_1, \ldots, X_n$ 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

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

When $Y_n = y$ is given, $X_1, \ldots, X_n$ 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

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

$\Rightarrow$ independent of θ.
So, $Y_n = \max\{X_1, \ldots, X_n\}$ 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 ?

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

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 θ 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 θ.

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

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

$\Leftarrow$) 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

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

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

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

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)

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

(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".

Note: 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,

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

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

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

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

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

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

(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 $(X_1, \ldots, X_n)$ 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.

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

$\Rightarrow Y_n$ 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.

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

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

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

$\Rightarrow Y_n = \max\{X_1, \ldots, X_n\}$ 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 $X_1, \ldots, X_n$ 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 $\hat{\tau} = t(U)$ 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

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

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

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

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

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

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

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

sol: The p.d.f of X is

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

$\Rightarrow \sum\limits_{i=1}^{n} X_i$ 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 $e^{-\lambda}$

$\Rightarrow \mathrm{E}(I(X_1 = 0) | \sum\limits_{i=1}^{n} X_i)$ is UMVUE of $e^{-\lambda}$.