L03: Sufficient, ancillary and complete statistics

1. Order relation of statistics

   (1) Order of statistics
   Let $X_1, ..., X_n$ be a random sample. Then a statistic is a function of the sample (could be a vector-valued function). For two statistics $T_1 = T_1(X_1, .., X_n)$ and $T_2 = T_2(X_1, ..., X_n)$ we write $T_1 \succeq T_2$, equivalently $T_2 \preceq T_1$ if T_2 is a function of T_1. For example
   $$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \succeq \begin{pmatrix} X_1 + \cdots + X_{n-k} \\ X_{n-k+1} + \cdots + X_n \end{pmatrix} \succeq X_1 + \cdots + X_n \succeq \overline{X}$$

   (2) $\preceq$ is an order relation
   (i) $\preceq$ is reflective: $T \preceq T$ (ii) $\preceq$ is transitive: $T_1 \preceq T_2$ and $T_2 \preceq T_3 \Longrightarrow T_1 \preceq T_3$.

   (3) During the descending process, the structure of the statistic becomes simpler and simpler.

2. Sufficient statistic

   (1) Definition and iff condition
   $$S(X_1, .., X_n) \text{ is a sufficient statistic for } \theta \stackrel{def}{\iff} f_{X_1,..,X_n|S}(\cdot) \text{ is free of } \theta$$
   $$\iff f_{X_1,..,X_n}(\cdot) = g(s;\, \theta)h(x_1, .., x_n).$$

   Comments: By the definition all information on θ in sample are absorbed by S.
   The second $\Leftrightarrow$ is Fisher-Neyman Factorization Theorem. By this theorem inference on θ based on sample can be made based on S.

   Proof: $\Rightarrow$**:** $f_{X_1,..,X_n}(\cdot) = f_S(s;\, \theta) f_{X_1,..,X_n|S}(x_1, .., x_n) = g(s;\, \theta)h(x_1, .., x_n)$.
   $\Leftarrow$**:** $f_{X_1,..,X_n|S}(\cdot) = \frac{f_{X_1,..,X_n}(x_1,..,x_n;\theta)}{f_S(s;\theta)} = \frac{g(s;\theta)h(x_1,..,x_n)}{\iint g(s;\theta)h(x_1,..,x_n)dx_1\cdots dx_n} = \frac{h(x_1,..,x_n)}{\iint h(x_1,..,x_n)dx_1 \cdot dx_n}$

   Ex1: By Factorization theorem, sample vector $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ is always sufficient for θ since
   $f_{X_1,..,X_n}(\cdot) = f(x_1, .., x_n;\, \theta) \cdot 1$.

   (2) A property
   S is sufficient for θ and $T \succeq S \Longrightarrow T$ is sufficient.

   Proof: $f_{X_1,..,X_n}(\cdot) = g(s;\, \theta)h(x_1, .., x_n) = g(s(t);\, \theta)h(x_1, .., x_2) = g_1(t;\, \theta)h(x_1, .., x_n)$.
   Comment: This property explained why sample vector is sufficient.

   (3) Minimal sufficient statistic
   S is a minimal sufficient statistic if S is a sufficient statistic and $S \preceq T$ for all sufficient statistics T.
   Minimal sufficient statistic is the simplest statistic in structure among those containing all information on θ.

   Ex2: $X_1, ..., X_n$ is a random sample from Poisson distribution Poisson(λ) with pmf $f(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$. Thus $f(x_1, .., x_n;\, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}e^{-\lambda} = \left(\lambda^{\sum x_i} e^{n\lambda}\right) \frac{1}{\prod_i x_i!}$. By Factorization theorem $X_1 + \cdots + X_n \sim$ Poisson($n\lambda$) is sufficient for λ.
   Note that $X_1 + \cdots + X_n = n\overline{X}$, i.e., $X_1 + \cdots + X_n \preceq \overline{X}$. Hence $\overline{X}$ is also sufficient for λ. For $X_1 + \cdots + X_n$ the distribution is Poisson, for $\overline{X}$, $E(\overline{X}) = \lambda$.

3. Ancillary statistic and complete statistic

(1) Ancillary statistic: Definition
Statistic T is ancillary for θ if the distribution of T is free of θ.
Ancillary statistic does not contain any information on θ.

(2) Ancillary statistic: Property
If T_1 is ancillary and $T_1 \succeq T_2$, then T_2 is also ancillary since if the distribution of T_1 is free of θ, then the distribution of a function of T_1 is also free of θ.

Ex3: $X_1, .., X_n$ is a random sample from $N(\mu, 1^2)$. $X_1 - X_2$ is ancillary for μ since $X_1 - X_2 \sim N(0, 2)$.

(3) Complete statistic: Definition
Statistic T is complete for θ if $E(g(T)) = 0 \Longrightarrow P(g(T) = 0) = 1$.

Comments: If a statistic T contains redundant information on θ, then one may use two different functions $g_1(T)$ and $g_2(T)$ to get the same information on θ, i.e., $E(g_1(T)) = E(g_2(T))$.
On this sense a complete statistic contains no redundant information on θ since $E(g_1(T)) = E(g_2(T)) \Longrightarrow g_1(T) = g_2(T)$.

(4) Complete statistic: Property
If T_1 is complete and $T_1 \succeq T_2$, then T_2 is complete because if distribution of T_1 is free of θ, so is the distribution of a function of T_1.

L04: A sufficient condition for MVUE via complete and sufficient statistics

1. The law of total expectation

   (1) A Theorem: $E[E(Y|X)] = E(Y)$.

   Proof: Suppose $X \in R^{p\times q}$ and $Y \in R^{m\times n}$. It suffices to show $E[E(y_{st}|X)] = E(y_{st})$. We consider continuous distribution only.
   $E(y_{st}|X) = \int_R y_{st} f_{y_{st}|X}(y_{st})dy_{st}$.
   $$\begin{aligned} E[E(y_{st}|X)] &= \iint_{R^{p\times q}} E(y_{st}|x)f_X(x)\,dx = \iint_{R^{p\times q}}[\int_R y_{st}f_{y_{st}|x}(y_{st})dy_{st}]f_X(x)dx \\ &= \iint_{R,\,R^{p\times q}} y_{st}f_{y_{st},X}(y_{st},x)dy_{st}dx = E(y_{st}).\square \end{aligned}$$

   (2) Application I: $E[P(Y \in A|X)] = P(Y \in A)$.

   Proof: $P(Y \in A) = \iint_A f_Y(y)dy = \iint_{R^{p\times q}} I_A(y)f_y(y)dy = E[I_A(Y)]$.
   $P(Y \in A|X) = \iint_A f_{Y|X}(y)dy = \iint I_A(y)f_{Y|X}(y)dy = E[I_A(Y)|X]$.
   By the law of total expectation, $E[I_A(Y)] = E[E(I_A(Y)|X]$, i.e.,
   $P(Y \in A) = E[P(Y \in A|X)]$. $\square$

   (3) Application II: For $X \in R^p$ and $Y \in R^m$, $\text{Cov}(E(Y|X)) = \text{Cov}(Y) - E[\text{Cov}(Y|X)]$.

   Proof: By the formula for $\text{Cov}(\cdot)$ and the law of total expectation,
   $$\begin{aligned} &\text{Cov}(E(Y|X)) = E\{[E(Y|X)][E(Y|X)]'\} - E[E(Y|X)]\{E[E(Y|X)]\}' \\ =\ & E\{[E(Y|X)][E(Y|X)]'\} - [E(Y)][E(Y)]' \\ =\ & E\{[E(Y|X)][E(Y|X)]'\} - [E(Y)][E(Y)]' + E(YY') - E[E(YY'|X)] \\ =\ & \text{Cov}(Y) - E\{E(YY'|X) - E(Y|X)[E(Y|X)]'\} = \text{Cov}(Y) - E[\text{Cov}(Y|X)].\square \end{aligned}$$

   Comment: $E(Y|X)]$ has the same mean of Y, but "small" covariance matrix of Y.

2. Basu Theorem

   (1) A lemma
   $S \in R^p$ and $T \in R^q$ are two random vectors. If $P(T \in A) = P(T \in A|S = s)$ for all $s \in R^p$ and $A \subset R^q$, then S and T are independent.

   Proof: For discrete distributions with $A = t \in R^q$ let $f_t(t)$ and $f_{T|s}(t)$ be a marginal and a conditional pmfs. Then $f_T(t) = f_{T|s}(t)$ for all t and s. Hence T and S are independent.
   For continuous distributions with $A = \{-\infty < T_1 \le t_1, ..., -\infty < T_q \le t_q\}$ let $f_T(t)$ and $f_{T|s}(t)$ be a marginal and a conditional pdfs. Then
   $$\int_{-\infty}^{t_1}\cdots\int_{-\infty}^{t_q} f_T(t)dt_1,..dt_q = \int_{-\infty}^{t_1}\cdots\int_{-\infty}^{t_q} f_{T|s}(t)dt_1,..dt_q.$$
   By taking derivatives with respect to $t_1, ..., t_q$, $f_T(t) = f_{T|s}(t)$ for all t and s.
   Hence T and S are independent. $\square$

   (2) Basu Theorem
   If S is a complete and sufficient statistic for θ and T is an ancillary statistic for θ, then T and S are independent.

   Proof: T is ancillary $\Longrightarrow P_\theta(T \in A) = P(T \in A)$ is free of θ.
   S is sufficient $\Longrightarrow P_\theta(T \in A|S = s) = P(T \in A|S = s)$ is free of θ.
   By (2) of 1, $E_\theta[P(T \in A|S = s) - P(T \in A)] = P(T \in A) - P(T \in A) = 0$.
   But S is complete. Thus $P(T \in A|S = s) = P(T \in A)$.
   Hence T and S are independent. $\square$

Comment: By using Basu Theorem one may show the independence without going through tedious details of joint and marginal distributions. But one will need tool to identify sufficient and complete statistics.

3. A sufficient condition for MVUE

(1) Complete and sufficient statistics in an exponential family

Theorem: If the joint pdf or pmf of sample $X: X_1, ..., X_n$ is

$$f(x;\, \theta) = \exp\left[a_1(\theta)b_1(x) + \cdots + a_t(\theta)b_t(x) + c(\theta) + d(X)\right] \text{ where } x:\, x_1, ..., x_n,$$

then $\begin{pmatrix} b_1(X) \\ \vdots \\ b_t(X) \end{pmatrix}$ is a complete sufficient statistic for θ.

Proof: Skipped

Comment: If a population has pdf or pmf

$$f(x;\, \theta) = \exp\left[a_1(\theta)b_1(x) + \cdots + a_t(\theta)b_t(x) + c(\theta) + d(x)\right],$$

then $\begin{pmatrix} \sum_i b_1(X_i) \\ \vdots \\ \sum_i b_t(X_i) \end{pmatrix}$ is complete and sufficient for θ.

Ex1: $X \sim$ Bernoulli(p) has pmf

$f(x;\, p) = p^x(1-p)^{1-x} = \left(\frac{p}{1-p}\right)^x (1-p) = \exp\left[x \ln \frac{p}{1-p} + \ln(1-p)\right]$ where $x = 0,\, 1$.

So $\sum_{i=1}^n X_i$ is complete and sufficient. But $\sum_i X_i = n\overline{X}$ and $\overline{X} = \frac{\sum_i X_i}{n}$. Thus $\overline{X} = \widehat{p}$ is also complete and sufficient.

(2) A Theorem

If S is sufficient and complete, and $E_\theta[g(S)] = \eta = \phi(\theta)$, then $g(S)$ is a MVUE for η

Proof: $E_\theta[g(S)] = \eta = \phi(\theta) \Longrightarrow g(S)$ is an U-estimator for η.

Suppose $E_\theta(U) = \eta = \pi(\theta)$. We show that $\mathrm{Cov}_\theta[g(S)] \leq \mathrm{Cov}_\theta(U)$.

S is sufficient $\Longrightarrow T = E_\theta(U|S)$ is a function of S free of θ.

By the law of total expectation,

$$E_\theta(T) = E\theta[E_\theta(U|S)] = E_\theta(U) = \eta = E_\theta[g(S)], \text{ i.e., } E_\theta[T - g(S)] = 0.$$

S is complete $\Longrightarrow T = g(S)$. By (3) of 1,

$$\mathrm{Cov}_\theta[g(S)] = \mathrm{Cov}_\theta(T) = \mathrm{Cov}_\theta[E_\theta(U|S)] \leq \mathrm{Cov}_\theta(U). \ \square$$

Ex2: In Ex1 $\widehat{p} = \overline{X}$ is a complete and sufficient statistics wrt p in Bernoulli(p). But $E(\widehat{p}) = p$. So $\widehat{p} = \overline{X}$ is MVUE for p.