Module-II

Sufficiency:

$\underline{X} = (X_1, \ldots, X_n)$: a random sample from a population with d.f. $F \in \mathcal{P}$.

Here $\mathcal{P}$ may be a parametric family (i.e. functional form of $F \equiv F_\theta$ is known except for unknown parameter $\theta \in \Theta$) or a non-parametric family (F is completely unknown, except possibly that F is AC or discrete).

Goal in Statistical Inference Problems: To use the information contained in $\underline{X}$ (or its realization $\underline{x}$) to make inferences about F.

When n is large, it may be difficult to interpret $\underline{x} = (x_1, \ldots, x_n)$ and the statistician may wish to summarize the information contained in $\underline{x}$ by computing some statistic(s) (e.g., the sample mean, the sample variance etc.)

Any statistic $T: \mathcal{X} \to \mathbb{R}^k$ defines a form of data reduction or data summary.

An experimenter who uses the statistic $T(\underline{x})$ rather than the observed sample $\underline{x}$ will treat as equal two samples $\underline{x}$ and $\underline{y}$ that satisfy $T(\underline{x}) = T(\underline{y})$.

Data Reduction Through a Statistic T:

$$T: \mathcal{X} \to \mathbb{R}^k \quad : \quad \text{a given statistic}$$

Define

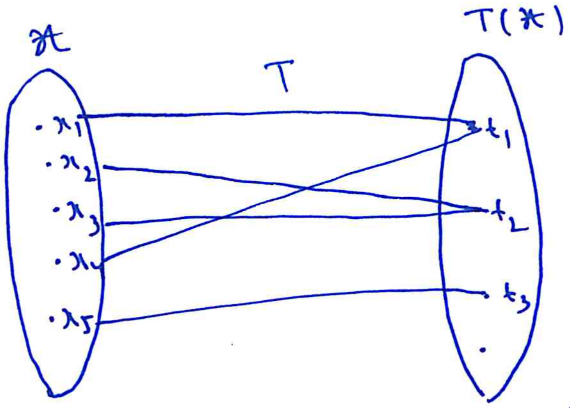
$$\mathcal{T} = T(\mathcal{X}) = \{t \in \mathbb{R}^k : T(\underline{x}) = t, \text{ for some } \underline{x} \in \mathcal{X}\}.$$

Then $T(\underline{x})$ partitions $\mathcal{X}$ into sets

$$A_t = \{\underline{x} \in \mathcal{X} : T(\underline{x}) = t\}, \quad t \in \mathcal{T} = T(\mathcal{X}).$$

The collection $\{A_t : t \in \mathcal{T}\}$ is called the partition induced by statistic T. Clearly every statistic T induces a partition of the sample space $\mathcal{X}$.

Example:



Here $A_{t_1} = \{x_1, x_4\}$, $A_{t_2} = \{x_2, x_3\}$, $A_{t_3} = \{x_5\}$

$A_{t_i} \cap A_{t_j} = \phi$, $\forall\, i \neq j$, $\bigcup_{i=1}^{3} A_{t_i} = \mathcal{X}$

The statistic $T(\underline{X})$ summarizes the data. Rather than reporting the entire sample x it only reports $T(x)$ or equivalently A_t.

Our goal is to summarize the data (i.e. data reduction) x through a statistic such that:

(a) no important information F that is contained in x is discarded;

(b) all irrelevant information contained in x about F is discarded.

Definition: Let $\underline{X} = (X_1, \ldots, X_n)$ have the joint d.f. $F \in \mathcal{P}$. A statistic $T \equiv T(\underline{X})$ (or the partition $\{A_t : t \in \mathcal{T}\}$ induced by T) is said to be sufficient for F or $\mathcal{P}$, if, for any $t \in \mathcal{T} = T(\mathcal{X})$, the conditional distribution of $\underline{X}$ given $T = t$ (or conditional distribution of $\underline{X}$ given $\underline{X} \in A_t$) does not depend on $F \in \mathcal{P}$ (i.e. it is the same for any $F \in \mathcal{P}$).

When $\mathcal{P} = \{F_\theta : \theta \in \Theta\}$ and the functional form of F_θ is known for every $\theta \in \Theta$, we simply say that T is sufficient for θ.

Suppose that $T(\underline{X})$ is sufficient for $\mathcal{P}$ and

$Pr((X_1,\dots,X_n) \in B \mid T(\underline{X})=t) = \nu_t(B)$, $B \in \mathcal{B}_n$, $t \in \mathcal{T}$,

where $\mathcal{B}_n$ is the Borel σ-field of $\mathbb{R}^n$.

Then $\nu_t(\cdot)$ does not depend on F and thus is completely known. Let $G(\cdot)$ be the d.f. of T.

Then, for any $B \in \mathcal{B}_n$,

$$Pr((X_1,\dots,X_n) \in B) = \int Pr((X_1,\dots,X_n) \in B \mid T=t)\, dG(t)$$

$$= \int \nu_t(B)\, dG(t)$$

$\rightarrow$ completely known

Thus one can get the distribution of $(X_1,\dots,X_n)$ from distribution of $T(\underline{X})$

Generating observation (or sample) from the population using just the observed value of a sufficient statistic: A discussion for discrete distributions

Let $T \equiv T(\underline{X})$ be a sufficient statistic. Then for any $t \in \mathcal{T}$, the conditional distribution of $\underline{X}$ given $T(\underline{X})=t$ does not depend on F (or it is completely known). So using computers one can generate a observation (or a sample) from the conditional distribution of $\underline{X}$ given $T(\underline{X})=t$, for any $t \in \mathcal{T}$.

Suppose that $T(\underline{X})=t$ is observed. Generate one observation, say $\underline{y}$, from the conditional distribution of $\underline{X}$ given $T(\underline{X})=t$. Let $\underline{Y}$ denote the r.v. corresponding to $\underline{Y}$. Then, for any $\underline{x} \in \mathcal{A}$

$$\Pr(Y = x) = \sum_{t \in \mathcal{T}} \Pr(Y = x,\ T(\underline{x}) = t)$$

$$= \sum_{t \in \mathcal{T}} \Pr(Y = x \mid T(\underline{x}) = t)\ \Pr(T(\underline{x}) = t)$$

$$= \sum_{t \in \mathcal{T}} \Pr(X = x \mid T(\underline{x}) = t)\ \Pr(T(X) = t)$$

$$= \Pr(X = x)$$

$$\Rightarrow \quad Y \overset{d}{=} X$$

Thus using any relization of sufficient statistic $T(\underline{X})$ one can generate an observation $\underline{Y}$ from the original distribution ($Y \overset{d}{=} X$).

__Example__ Let X_1 and X_2 be i.i.d. $B(1, \theta)$ rvs, where $\theta \in \Theta = (0, 1)$ is unknown. Let $\underline{X} = (X_1, X_2)$. Then, for $\underline{x} = (x_1, x_2)$

$$P_\theta(\underline{X} = \underline{x}) = \theta^k (1-\theta)^{2-k}, \quad k = x_1 + x_2$$
$$= \text{\# of 1's among } x_1 \text{ and } x_2$$

Define

$$T(\underline{X}) = X_1 + X_2 = \text{\# of successes among two trials.}$$

Sample	T	$P_\theta(\underline{X}=\underline{x} \mid T(\underline{x})=0)$	$P_\theta(\underline{X}=\underline{x} \mid T(\underline{x})=1)$	$P_\theta(\underline{X}=\underline{x} \mid T(\underline{x})=2)$	$P_\theta(\underline{X}=\underline{x})$
(0, 0)	0	1	0	0	$(1-\theta)^2$
(0, 1)	1	0	$\frac{1}{2}$	0	$\theta(1-\theta)$
(1, 0)	1	0	$\frac{1}{2}$	0	$\theta(1-\theta)$
(1, 1)	2	0	0	1	θ^2

Note that conditional distribution of $\underline{X}$ given any particular value of T does not depend on θ.

$\Rightarrow$ T is sufficient for θ.

From T, to get $Y_1, Y_2 \overset{iid}{\sim} B(1, \theta)$ we proceed as follows:

If $T=0$ is observed choose $(Y_1, Y_2) = (0, 0)$ w.p. 1

If $T=2$ is observed choose $(Y_1, Y_2) = (1, 1)$ w.p. 1

If $T=1$ is observed choose $(Y_1, Y_2) = \begin{cases} (0, 1) & \text{w.p. } \frac{1}{2} \\ (1, 0) & \text{w.p. } \frac{1}{2} \end{cases}$

Then

$$\Pr((Y_1, Y_2) = (0, 0)) = \Pr((Y_1, Y_2) = (0, 0), T=0) + \Pr((Y_1, Y_2) = (0, 0), T=1) + \Pr((Y_1, Y_2) = (0, 0), T=2)$$

$$= \Pr((Y_1, Y_2) = (0, 0), T=0)$$

$$= \Pr((Y_1, Y_2) = (0, 0) \mid T=0) \Pr(T=0)$$

$$= \Pr(T=0) = \Pr((X_1, X_2) = (0, 0))$$

$$\Pr((Y_1, Y_2) = (1, 1)) = \Pr((Y_1, Y_2) = (1, 1), T=0) + \Pr((Y_1, Y_2) = (1, 1), T=1) + \Pr((Y_1, Y_2) = (1, 1), T=2)$$

$$= \Pr((Y_1, Y_2) = (1, 1), T=2) = \Pr(T=2) =$$

$$= \Pr((X_1, X_2) = (1, 1))$$

$$\Pr((Y_1, Y_2) = (0, 1)) = \Pr((Y_1, Y_2) = (0, 1), T=1)$$

$$= \Pr((Y_1, Y_2) = (0, 1) \mid T=1) \Pr(T=1)$$

$$= \frac{1}{2} \Pr(T=1) = \frac{1}{2} [\theta(1-\theta) + (1-\theta)\theta]$$

$$= \theta(1-\theta) = \Pr((X_1, X_2) = (0, 1))$$

Similarly

$$\Pr((Y_1, Y_2) = (1, 0)) = \theta(1-\theta) = \Pr((X_1, X_2) = (1, 0))$$

Clearly

$$(X_1, X_2) \overset{d}{=} (Y_1, Y_2).$$

Example Let $X_1, \dots, X_n$ be a random sample from $N(\theta, 1)$ where $\theta \in \Theta = \mathbb{R}$ is unknown. Let P be an orthogonal matrix with first row as $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$. Define

$$\underline{Y} = P\underline{X} = \begin{pmatrix} \sqrt{n}\,\bar{X} \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

Clearly

$$\underline{Y} \sim N_n\left(\begin{pmatrix} \sqrt{n}\,\mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}, I_n \right)$$

Then $\underline{X}$ and $\underline{Y}$ have the same information about μ (since $\underline{Y} = P\underline{X}$ and P is non-singular). Moreover $Y_1, Y_2, \dots, Y_n$ are independent with $Y_1 \sim N(\sqrt{n}\mu, 1)$ and $Y_2, \dots, Y_n$ are iid $N(0, 1)$. This suggests that all the information about μ is contained in Y_1. Interestingly, as the following arguments show, $\bar{X}$ is sufficient for μ.

$$\underline{X} \mid \bar{X} \overset{d}{=} P^t \underline{Y} \,\Big|\, \frac{Y_1}{\sqrt{n}}.$$

$(Y_1, Y_2, \dots, Y_n) \mid Y_1$ distribution does not depend on μ

$(Y_1, Y_2, \dots, Y_n) \mid Y_1$ distribution does not depend on μ

$\Rightarrow$ $P^t \underline{Y} \mid Y_1$ distribution does not depend on μ

$\Rightarrow$ $P^t \underline{Y} \mid \frac{Y_1}{\sqrt{n}}$ distribution does not depend on μ

$\Rightarrow$ $\underline{X} \mid \bar{X}$ distribution does not depend on μ

$\Rightarrow$ $\bar{X}$ is sufficient for μ.

Example Let $\Theta = \{\theta_1, \theta_2, \theta_3\}$, $\mathcal{X} = \{x_1, x_2, x_3\}$ and the p.m.f. of X is given by the following matrix

$p_\theta(x)$

θ \ x	x_1	x_2	x_3
θ_1	.4	.2	.4
θ_2	.6	.3	.1
θ_3	.2	.1	.7

$\boxed{8/2}$

We know that a partition $\mathcal{P} = \{ P_t : t \in \Lambda \}$ will be sufficient if $P_\theta (X = x \mid X \in P_t)$ does not depend on θ for every $t \in \Lambda$, and $x \in \mathcal{X}$.

Let $\mathcal{P} = \{ P_1, P_2 \}$, where $P_1 = \{x_1, x_2\}$ and $P_2 = \{x_3\}$.

Then

$$f_{\theta_1}(x_1 \mid P_1) = P_{\theta_1} [X = x_1 \mid X \in \{x_1, x_2\}] = \frac{P_{\theta_1} [X = x_1]}{P_{\theta_1} [X = x_1] + P_{\theta_1} [X = x_2]}$$

$$= \frac{0.4}{0.4 + 0.2} = \frac{2}{3}$$

$$f_{\theta_1}(x_2 \mid P_1) = 1 - \frac{2}{3} = \frac{1}{3}, \quad f_{\theta_1}(x_3 \mid P_1) = 0$$

$$f_{\theta_2}(x_1 \mid P_1) = \frac{.6}{.6 + .3} = \frac{2}{3}, \quad f_{\theta_2}(x_2 \mid P_1) = 1 - \frac{2}{3} = \frac{1}{3},$$

$$f_{\theta_2}(x_3 \mid P_1) = 0$$

$$f_{\theta_3}(x_1 \mid P_1) = \frac{.2}{.2 + .1} = \frac{2}{3}, \quad f_{\theta_3}(x_2 \mid P_1) = 1 - \frac{2}{3} = \frac{1}{3}$$

$$f_{\theta_3}(x_3 \mid P_1) = 0$$

$\Rightarrow$ $f_\theta(x \mid P_1)$ does not depend on $\theta \in \Theta$

Also

$$f_{\theta_1}(x_1 \mid P_2) = 0, \quad f_{\theta_1}(x_2 \mid P_2) = 0, \quad f_{\theta_1}(x_3 \mid P_2) = 1$$

$$f_{\theta_2}(x_1 \mid P_2) = 0 \quad f_{\theta_2}(x_2 \mid P_2) = 0 \quad f_{\theta_2}(x_3 \mid P_2) = 1$$

$$f_{\theta_3}(x_1 \mid P_2) = 0 \quad f_{\theta_3}(x_2 \mid P_2) = 0 \quad f_{\theta_3}(x_3 \mid P_2) = 1$$

$\Rightarrow$ $f_\theta(x \mid P_2)$ does not depend on $\theta \in \Theta$

Thus for any $i \in \{1, 2\}$ and $x \in \mathcal{X}$

$f_\theta(x \mid P_i)$ does not depend on $\theta \in \Theta$

$\Rightarrow$ $\mathcal{P} = \{ \{x_1, x_2\}, \{x_3\} \}$ is a sufficient partition

$\Rightarrow$ Any statistic T such that $T(x_1) = T(x_2) \neq T(x_3)$ is sufficient.

$\boxed{7/2}$

Consider the statistic

$$T(x) = \begin{cases} 0, & \text{if } x = x_1 \\ 1, & \text{if } x \in \{x_2, x_3\}. \end{cases}$$

Then

	x_1	x_2	x_3	
$f_{\theta_j}(x \mid T(x)=0)$	1	0	0	, $j = 1, 2, 3$
$f_{\theta_1}(x \mid T(x)=1)$	0	$1/3$	$2/3$	
$f_{\theta_2}(x \mid T(x)=1)$	0	$3/4$	$1/4$	

Here

$$f_{\theta_1}(x_2 \mid T(x)=1) \neq f_{\theta_2}(x_2 \mid T(x)=1)$$

$\Rightarrow$ $T(X)$ is not sufficient for $\theta \in \Theta$.

- Let $f_F(\cdot)$ and $g_F(\cdot)$, respectively, denote the p.m.f./p.d.f. of $\underline{X}$ and $T(\underline{X})$ when $\underline{X} \sim F \in \mathcal{P}$. Let $t \in \mathcal{T}$ be such that $\Pr(T(\underline{X})=t) > 0$. Then

$$\Pr(\underline{X}=\underline{x} \mid T(\underline{X})=t) = \begin{cases} 0, & \text{if } T(\underline{x}) \neq t \\ \dfrac{\Pr(\underline{X}=\underline{x})}{\Pr(T(\underline{X})=t)}, & \text{if } T(\underline{x}) = t \end{cases}$$

and thus $T(\underline{X})$ is sufficient for $F \in \mathcal{P}$ if for every $\underline{x} \in \mathcal{X}$

$$\frac{\Pr(\underline{X}=\underline{x})}{\Pr(T(\underline{X})=T(\underline{x}))}$$

does not depend on F, or equivalently $T(\underline{X})$ is sufficient for $\mathcal{P}$ if for every $F \in \mathcal{P}$ and $\underline{x} \in \mathcal{X}$ such that $g_F(T(\underline{x})) > 0$, $f_F(\underline{x}) / g_F(T(\underline{x}))$ is constant in $F \in \mathcal{P}$.

Theorem

(I) $T(\underline{X})$ is sufficient for $\mathcal{P}$ iff for every $\underline{x} \in \mathcal{X}$

$\dfrac{f_F(\underline{x})}{g_F(T(\underline{x}))}$ is constant in $F \in \{F \in \mathcal{P} : g_F(T(\underline{x})) > 0\}$

Factorization Theorem:

(II) $T(\underline{X})$ is sufficient for $\mathcal{P}$ iff there exist functions $\psi_F(\cdot)$ and $h(\cdot)$ such that for all $\underline{x} \in \mathcal{X}$ and all $F \in \mathcal{P}$

$$f_F(\underline{x}) = \psi_F(T(\underline{x}))\, h(\underline{x}), \quad \ldots\ldots \quad (*)$$

here $h(\cdot)$ does not depend on $F \in \mathcal{P}$.

Proof. (Discrete Case)

(I) Follows trivially from earlier discussion

(II) Firstly suppose that $T(\underline{X})$ is sufficient for $F \in \mathcal{P}$. Then, for any $F \in \mathcal{P}$ and $\underline{x} \in \mathcal{X}$

$$\begin{aligned} f_F(\underline{x}) &= P_r(\underline{X} = \underline{x} \mid F) \\ &= P_r(\underline{X} = \underline{x},\ T(\underline{X}) = T(\underline{x}) \mid F) \\ &= P_r(T(\underline{X}) = T(\underline{x}) \mid F)\ \underbrace{P_{r_F}(\underline{X} = \underline{x} \mid T(\underline{X}) = T(\underline{x}))}_{\substack{= h(\underline{x}),\ \text{say} \\ \text{does not depend on } F \in \mathcal{P} \\ \text{as } T(\underline{X}) \text{ is sufficient for } F \in \mathcal{P}}} \\ &= g_F(T(\underline{x}))\, h(\underline{x}) \\ &= \psi_F(T(\underline{x}))\, h(\underline{x}). \end{aligned}$$

Conversely suppose that the factorization (*) holds. Then, for any $\underline{x} \in \mathcal{X}$ and $F \in \mathcal{P}$ such that $g_F(T(\underline{x})) > 0$ we have

$$\begin{aligned} g_F(T(\underline{x})) &= P_r(T(\underline{X}) = T(\underline{x}) \mid F) \\ &= \sum_{\substack{\underline{y} \in \mathcal{X} \\ T(\underline{y}) = T(\underline{x})}} P_r(\underline{X} = \underline{y} \mid F) \end{aligned}$$

$$= \sum_{\substack{\underline{y} \in \mathcal{X} \\ T(\underline{y}) = T(\underline{x})}} \psi_F(T(\underline{y}))\, h(\underline{y})$$

$$= \psi_F(T(\underline{x})) \sum_{\substack{\underline{y} \in \mathcal{X} \\ T(\underline{y}) = T(\underline{x})}} h(\underline{y})$$

Thus

$$\frac{f_F(\underline{x})}{g_F(T(\underline{x}))} = \frac{h(\underline{x})}{\sum_{\substack{\underline{y} \in \mathcal{X} \\ T(\underline{y}) = T(\underline{x})}} h(\underline{y})}$$

does not depend on $F \in \mathcal{P}$

$\Rightarrow$ $T(\underline{X})$ is sufficient for $F \in \mathcal{P}$.

Let $\mathcal{P}_0 \subseteq \mathcal{P}$.

Remark: (i) If $T(\underline{X})$ is sufficient for $F \in \mathcal{P}$ then $T(\underline{X})$ is also sufficient for $F \in \mathcal{P}_0$

(ii) $T(\underline{X}) = \underline{X}$ is a sufficient statistic, i.e. complete sample $\underline{X}$ is a sufficient statistic (called the trivial sufficient statistic). (Follow using factorization theorem)

(iii) If $T(\underline{X})$ is sufficient for $F \in \mathcal{P}$ and $T(\underline{X})$ is a function of statistic $U(\underline{X})$ then $U(\underline{X})$ is also sufficient for $F \in \mathcal{P}$ (follows using factorization theorem)

Example: Let $X_1, \ldots, X_n$ be a random sample from $Bin(1, \theta)$ distribution, where $\theta \in \Theta = (0, 1)$ is unknown. Show that $T(\underline{X}) = \sum_{i=1}^{n} X_i$ is a sufficient statistic. Are $T_1(\underline{X}) = \overline{X}$, $T_2(\underline{X}) = \overline{X}^2$ and $T_3(\underline{X}) = \overline{X} + 5$ sufficient for θ.

Solution For $\underline{x} \in \mathcal{X} = \{0, 1\}^n$ and $\theta \in \Theta = (0, 1)$

$$f_\theta(\underline{x}) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Thus, for $\underline{x} \in \mathbb{R}^n$,

$$f_\theta(\underline{x}) = \psi_\theta\left(\sum_{i=1}^{n} x_i\right) h(\underline{x})$$

$$= \psi_\theta(T(\underline{x}))\, h(\underline{x}),$$

Where $h(\underline{x}) \equiv \begin{cases} 1, & \forall\, \underline{x} \in \mathcal{X} \\ 0, & \text{otherwise} \end{cases}$ $\quad \psi_\theta(t) = \begin{cases} \theta^t(1-\theta)^{n-t}, & t = 0, 1, \ldots, n,\ 0<\theta<1 \\ 0, & \text{otherwise} \end{cases}$

Thus $T(\underline{X}) = \sum_{i=1}^{n} X_i$ is sufficient for $\theta \in \Theta$. Note that $T(\underline{X}) \geq 0$, and $T_1(\underline{X}) = \bar{X}$, $T_2(\underline{X}) = \bar{X}^2$ and $T_3(\underline{X}) = \bar{X} + 5$ are 1-1 functions of $T(\underline{X})$. It follows that $T_1(\underline{X})$, $T_2(\underline{X})$ and $T_3(\underline{X})$ are also sufficient for $\theta \in \Theta$.

Example Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ r.v.s where $\underline{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ is unknown. Show that $T(\underline{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is a sufficient statistic for $\underline{\theta} \in \Theta$. Is $T_1(\underline{X}) = (\bar{X}, S^2)$ sufficient for $\theta \in \Theta$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Solution For $\underline{x} \in \mathcal{X} = \mathbb{R}^n$ and $\underline{\theta} = (\mu, \sigma) \in \Theta$,

$$f_{\underline{\theta}}(\underline{x}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2\right]}$$

Thus, for $\underline{x} \in \mathbb{R}^n$

$$f_\theta(\underline{x}) = \psi_\theta\left(\left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)\right) h(\underline{x}),$$

where $h(\underline{x}) = 1 \ \forall\, \underline{x} \in \mathbb{R}^n$ and $\psi_\theta(t_1, t_2) = \begin{cases} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}(t_2 - 2\mu t_1 + n\mu^2)}, & t_1 \in \mathbb{R},\ t_2 \geq 0 \\ 0 & \text{o.w.} \end{cases}$

$\boxed{11/2}$

$\Rightarrow$ $T(\underline{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is sufficient for $F \in \mathcal{P}$

Since $T(\underline{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is a function of $T_1(\underline{X}) = (\bar{X}, S^2)$, it follows that $T_1(\underline{X})$ is also sufficient for $F \in \mathcal{P}$.

Example Let $X_1, \dots, X_n$ be a random sample from a discrete uniform distribution on $\{1, 2, \dots, \theta\}$, where $\theta \in \Theta = \mathbb{N} = \{1, 2, \dots\}$ is an unknown parameter. Show that $T(\underline{X}) = X_{(n)} = \max\{X_1, \dots, X_n\}$ is sufficient for $\theta \in \Theta$. Is $T_1(\underline{X}) = e^{X_{(n)}}$ sufficient for $\theta \in \Theta$?

Solution For $\underline{x} \in \mathcal{X} = \{1, 2, \dots\}^n$ and $\theta \in \Theta$

$$f_\theta(\underline{x}) = \begin{cases} \frac{1}{\theta^n}, & \text{if } x_i \in \{1, 2, \dots, \theta\}, \ i = 1(1)n \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n}, & \text{if } x_i \in \{1, 2, \dots\}, \ i = 1(1)n, \ x_{(n)} \le \theta \\ 0, & \text{otherwise} \end{cases}$$

Thus, for $\underline{x} \in \mathbb{R}^n$

$$f_\theta(\underline{x}) = \psi_\theta(x_{(n)})\, h(\underline{x}), \quad h(\underline{x}) = \begin{cases} 1, & \text{if } x_i \in \{1, 2, \dots\}, \ i = 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

where ~~$h(\underline{x}) = 1, \ \forall \underline{x} \in \mathbb{R}^n$, and~~

$$\psi_\theta(t) = \begin{cases} 1 & \text{if } t \le \theta, \ t \in \mathbb{R} \\ \cancel{0} & \cancel{\text{otherwise}} \end{cases}$$

$\Rightarrow$ $T(\underline{X}) = X_{(n)}$ is sufficient for $\theta \in \Theta$. Since $T_1(\underline{X})$ is a 1-1 function of $T(\underline{X})$, $T_1(\underline{X})$ is also sufficient for $\theta \in \Theta$.

Example (Sufficient statistic may not always provide substantial reduction in data)

Let $X_1, \ldots, X_n$ be a random sample from a distribution $F \in \mathcal{P}$, where $\mathcal{P}$ is the class of all absolutely continuous distributions. Suppose that $m_F(\cdot)$ is the common p.d.f. of X_i, $i = 1, \ldots, n$ under $F \in \mathcal{P}$. Let $X_{(i)} =$ i-th smallest of $X_1, \ldots, X_n$ so that $X_{(1)} \le \cdots \le X_{(n)}$. Then, for $\underline{x} \in \chi$

$$f_F(\underline{x}) = \prod_{i=1}^{n} m_F(x_i)$$

$$= \prod_{i=1}^{n} m_F(x_{(i)})$$

$$= \psi_F(x_{(1)}, \ldots, x_{(n)})\, h(\underline{x}) \qquad (h(\underline{x}) \equiv 1)$$

$\Rightarrow$ $T(\underline{X}) = (X_{(1)}, \ldots, X_{(n)})$ is sufficient for $F \in \mathcal{P}$.

Example Let $X_1, \ldots, X_n$ be a random sample from $N(\theta, \sigma^2)$, where $\sigma^2 > 0$ is known and $\theta \in \Theta = \mathbb{R}$ is unknown. Show that $T(\underline{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is sufficient for θ.

Example (i) Let X_1, X_2 be a random sample from $N(0, \sigma^2)$ where $\sigma \in \Theta = (0, \infty)$ is unknown. Using definition show that $T(\underline{X}) = X_1^2 + X_2^2$ is a sufficient statistic for σ^2.

(ii) Let $X_1, \ldots, X_n$ be a random sample from $N(0, \sigma^2)$ where $\sigma \in \Theta = (0, \infty)$ is unknown. Using definition show that $T(\underline{X}) = \sum_{i=1}^{n} X_i^2$ is sufficient for σ. (Hint: Use spherical coordinates transformation)

Example Let $X_1 \ldots X_n$ be a random sample from $E(\mu, \sigma)$ distribution

$$f_{\underline{\theta}}(x) = \begin{cases} \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}}, & \text{if } x \ge \mu, \\ 0, & \text{otherwise} \end{cases}$$

where $\underline{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ is unknown. Using definition show that $T = (X_{(1)}, \sum_{i=2}^{n} X_{(i)})$ is sufficient for $\underline{\theta}$. Is $(X_{(1)}, \sum_{i=1}^{n}(X_i - X_{(1)}))$ sufficient for $\underline{\theta}$?

Example Consider the above example.

(i) If $\sigma > 0$ is known, show that $T_1(\underline{X}) = X_{(1)}$ is sufficient for $\mu \in \Theta = \mathbb{R}$

(ii) If $\mu > 0$ is known, show that $T_2(\underline{X}) = \sum_{i=1}^{n} X_i$ is sufficient for $\sigma \in \Theta = (0, \infty)$

Example Let $X_1, \ldots, X_n$ be a random sample from $U(-\theta, \theta)$, where $\theta \in \Theta = (0, \infty)$ is unknown. Find a sufficient statistic other than the trivial sufficient statistic.

Example Let $X_1, \ldots, X_n$ be a random sample from gamma distribution with pdf

$$f_{\underline{\theta}}(x) = \frac{1}{\sigma^{\alpha}\Gamma\alpha} x^{\alpha-1} e^{-\frac{x}{\sigma}}, \quad x > 0,\ \alpha > 0,\ \sigma > 0.$$

(a) By assuming α and σ both unknown, find a non-trivial sufficient statistic for (α, σ);

(b) By assuming α to be known, find a non-trivial sufficient statistic for σ;

(c) By assuming σ to be known, find a non-trivial sufficient statistic for α.

Factorization theorem suggests that for any given problem there are numerous sufficient statistics. In fact if $T(\underline{X})$ is a sufficient statistic and if another statistic $U(\underline{X})$ is a function of $T(\underline{X})$ then $U(\underline{X})$ is also sufficient. In particular all 1-1 functions of a sufficient statistic are sufficient. Thus a natural question is:

Question: Among various available sufficient statistics, which one should be preferred over others?

Since the purpose of using sufficient statistic is summarization of data without loss of information about F, a sufficient statistic which provides maximum reduction of data should be preferred. Such a sufficient statistic is called the minimal sufficient statistic.

Definition: A sufficient statistic $T(\underline{X})$ is said to be minimal sufficient if it is a function of every other sufficient statistic, i.e., $T(\underline{X})$ is minimal sufficient if $T(\underline{X})$ is sufficient and for any other sufficient statistic $U(\underline{X})$, $T(\underline{X})$ is a function of $U(\underline{X})$.

Remark: To say that $U(\underline{X})$ is a function of $T(\underline{X})$ simply means that if $T(\underline{x})=T(\underline{y})$ then $U(\underline{x})=U(\underline{y})$. Define

$$\mathcal{T} = T(\mathcal{X}) = \{T(\underline{x}): \underline{x} \in \mathcal{X}\},$$

and

$$\mathcal{U} = U(\mathcal{X}) = \{U(\underline{x}): \underline{x} \in \mathcal{X}\},$$

$$B_u = \{\underline{x} \in \mathcal{X}: U(\underline{x}) = u\}, \quad u \in \mathcal{U}$$

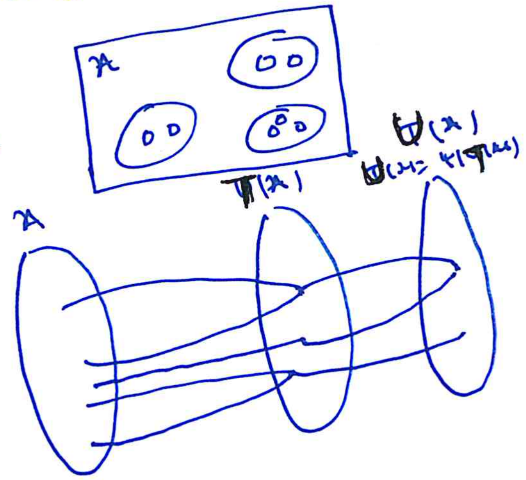
and

$$C_t = \{\underline{x} \in \mathcal{X}: T(\underline{x}) = t\}, \quad t \in \mathcal{T},$$

so that $\{B_u: u \in \mathcal{U}\}$ and $\{C_t: t \in \mathcal{T}\}$ are two partitions of $\mathcal{X}$. Then, clearly

$U(\underline{X})$ is a function of $T(\underline{X})$

$$\Longleftrightarrow \forall u \in \mathcal{U}, \quad B_u = \bigcup_{t \in \mathcal{T}_0} C_t \quad \text{for some } \mathcal{T}_0 \subseteq \mathcal{T}.$$



Thus, the partition associated with a minimal sufficient statistic is the coarsest possible partition for a sufficient statistic, and a minimal sufficient statistic achieves the greatest possible data reduction for a sufficient statistic.

Example Let $X_1, \ldots, X_n$ be a random sample from $N(\theta, \sigma_0^2)$, where $\theta \in \Theta = \mathbb{R}$ is unknown and $\sigma_0^2 > 0$ is known. Then

$$f_\theta(\underline{x}) = (2\pi\sigma_0)^{-\frac{n}{2}} \; e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\theta)^2}$$

$$= (2\pi\sigma_0)^{-\frac{n}{2}} \; e^{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2\right]}, \quad \underline{x} \in \mathcal{X} = \mathbb{R}^n$$

$\Rightarrow$ $T_1(\underline{X}) = \bar{X}$ is sufficient

$T_2(\underline{X}) = (\bar{X}, \sum_{i=1}^{n}(X_i-\bar{X})^2)$ is also sufficient (using factorization Theorem)

Thus T_1 and T_2 contain the same amount of information about θ. However $T_1(\underline{X})$ gives greater reduction of data than $T_2(\underline{X})$

Note that $T_1(\underline{X})$ is a function of $T_2(\underline{X})$ ($r(\underline{T_2}) = r(T_{21}, T_{22}) = T_{21} = T_1$)

Here the additional information about the value of S^2 does not add to our knowledge of θ, since σ_0^2 is known. Of course, when σ_0^2 is known $T_2(\underline{X})$ is not a sufficient statistic and $T_2(\underline{X}) = (\bar{X}, S^2)$ ($S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$, $n \geq 2$) contains more information about (θ, σ_0^2) than $T_1(\underline{X}) = \bar{X}$.

Theorem Let $\underset{\sim}{X} \sim F_{\underline{\theta}} \in \mathcal{P} = \{F_{\underline{\theta}} : \underline{\theta} \in \Theta\}$ and let f_θ be the joint p.d.f. / p.m.f. corresponding to $F_{\underline{\theta}}$, $\underline{\theta} \in \Theta$. Suppose that there exists a statistic $T(\underline{X})$ such that for any two sample points $\underline{x}$ and $\underline{y}$, the ratio $f_\theta(\underline{x}) / f_\theta(\underline{y})$ is independent of $\theta \in \Theta$ (constant as a function of θ) iff $T(\underline{x}) = T(\underline{y})$. Then $T(\underline{X})$ is a minimal sufficient statistic for θ.

Proof. For simplicity assume that $f_\theta(\underline{x})>0$, $\forall\, \underline{x}\in\mathcal{X}$ and $\theta\in\Theta$.

Sufficiency: Define

$$\mathcal{T}=T(\mathcal{X})=\{T(\underline{x}): \underline{x}\in\mathcal{X}\}$$

and $$A_t=\{\underline{x}\in\mathcal{X}: T(\underline{x})=t\} \qquad t\in\mathcal{T}$$

so that $\{A_t: t\in\mathcal{T}\}$ forms a partition of $\mathcal{X}$.

For each $t\in\mathcal{T}$, choose and fix one element $x_t\in A_t$. Then x_t depends only on t (for a given $T(\underline{x})$). Moreover, for each $\underline{x}\in\mathcal{X}$, $\underline{x}$ and $x_{T(\underline{x})}$ are in the same set A_t (where $t=T(\underline{x})$) and therefore $T(\underline{x})=x_{T(\underline{x})}$. Consequently

$$\frac{f_\theta(\underline{x})}{f_\theta(x_{T(\underline{x})})} \quad \text{is independent of } \theta\in\Theta$$

$$\Rightarrow \quad f_\theta(\underline{x})=f_\theta(x_{T(\underline{x})})\,h(\underline{x}), \qquad \text{for some } h(\cdot)$$

$\Rightarrow$ $T(\underline{x})$ is sufficient (by Factorization Theorem).

Minimal Sufficiency: Let $T_1(\underline{x})$ be any other sufficient statistic. Then, by the Factorization Theorem,

$$f_\theta(\underline{x})=\psi_\theta(T_1(\underline{x}))\,h_1(\underline{x}), \qquad \underline{x}\in\mathcal{X},$$

for some functions $\psi_\theta(\cdot)$ and $h_1(\cdot)$.

Let $\underline{x}$ and $\underline{y}$ be two sample points such that $T_1(\underline{x}) = T_1(\underline{y})$. Then, for any $\underline{x},\underline{y}\in\mathcal{X}$,

$$\frac{f_\theta(\underline{x})}{f_\theta(\underline{y})}=\frac{\psi_\theta(T_1(\underline{x}))\,h_1(\underline{x})}{\psi_\theta(T_1(\underline{y}))\,h_1(\underline{y})}=\frac{h_1(\underline{x})}{h_1(\underline{y})} \quad \text{does not depend on } \theta\in\Theta$$

$\Rightarrow$ $T(\underline{x})=T(\underline{y})$

$\Rightarrow$ T is minimal sufficient for $\theta\in\Theta$.

Corollary Consider the set-up of above theorem.

(i) If $T(\underline{x})$ is minimal sufficient for $\theta \in \Theta$ and $T_1(\underline{x})$ is 1-1 function of $T(\underline{x})$ then $T_1(\underline{x})$ is also sufficient for $\theta \in \Theta$

(ii) If $T(\underline{x})$ and $T_1(\underline{x})$ are minimal sufficient for $\theta \in \Theta$ then $T(\underline{x})$ and $T_1(\underline{x})$ are in 1-1 correspondence.

Proof. (i) $T_1(\underline{x})$ is clearly sufficient ($T(\underline{x})$ is sufficient and $T(\underline{x})$ is a function of $T_1(\underline{x})$). Let $T_2(\underline{x})$ be any other sufficient statistic. Then $T(\underline{x})$ is ~~sufficient for~~ ~~$T_2(\underline{x})$~~ a function of $T_2(\underline{x})$. Thus, for any $\underline{x}, \underline{y} \in \mathcal{X}$,

$$T_2(\underline{x}) = T_2(\underline{y}) \Rightarrow T(\underline{x}) = T(\underline{y}) \Rightarrow T_1(\underline{x}) = T_1(\underline{y})$$

$\Rightarrow$ T_2 is a function of T_1

$\Rightarrow$ Sufficient statistic T_1 is a function of any other sufficient statistic T_2

$\Rightarrow$ $T_1(\underline{x})$ is minimal sufficient.

(ii) Trivial.

Remark: (i) A minimal sufficient statistic may not be unique;

(ii) In general a minimal sufficient statistic may not exist (Bahadur (1954)). However a minimal sufficient statistic generally exists if $\mathcal{X} \subseteq \mathbb{R}^n$ and the distributions involved are discrete or absolutely continuous.

Example Let $X_1, \ldots, X_n$ $(n \geq 2)$ be a random sample from $N(\mu, \sigma^2)$, where $\underline{\theta} = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ is unknown. Show that $T(\underline{X}) = (\bar{X}, S^2)$ is minimal sufficient for $\underline{\theta} \in \Theta$, where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Also show that $T_1(\underline{X}) = \bar{X}$ and $T_2(\underline{X}) = S^2$ are not sufficient for $\underline{\theta} \in \Theta$.

Example Let $X_1, \ldots, X_n$ be a random sample from $U(\theta, \theta+1)$ distribution, where $\theta \in \Theta = \mathbb{R}$ is unknown. Show that $T(\underline{X}) = (X_{(1)}, X_{(n)})$ is minimal sufficient for $\theta \in \Theta$.

Example Let $X_1, \ldots, X_n$ be a random sample from a population having p.d.f./p.m.f. $g_\theta(\cdot)$, $\theta \in \Theta$. In each of the following cases find a minimal sufficient statistic $T(\underline{X})$:

(i) $g_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$, $x \in \mathcal{X} = \{0, 1, \ldots, n\}$, $\theta \in \Theta = (0, 1)$.

(ii) $g_\theta(x) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(x-\theta)^2}$, $x \in \mathcal{X} = \mathbb{R}$, $\theta \in \Theta = \mathbb{R}$, $\sigma_0 \in (0, \infty)$ is known.

Is $T_1(\underline{X}) = \overline{X}^2$ minimal sufficient? Is it sufficient

(iii) $g_\theta(x) = \frac{1}{\theta \sqrt{2\pi}} e^{-\frac{1}{2\theta^2}(x-\mu_0)^2}$, $x \in \mathcal{X} = \mathbb{R}$, $\theta \in \Theta = (0, \infty)$, $\mu_0 \in \mathbb{R}$ is known. Is $T_1(\underline{X}) = \sum_{i=1}^{n} X_i^2$ sufficient?

(iv) $g_\theta(x) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}$, $x \in \mathcal{X} = \mathbb{R}$, $\theta \in \Theta = \mathbb{R}$.

Ancillary Statistic

A complementary of a sufficient statistic (which contains all the information about $\mathcal{P}$) is the ancillary statistic (which does not contain any information about $\mathcal{P}$).

Definition. A statistic $T(\underline{X})$ is said to be a ancillary statistic if its distribution does not depend on the parameter $\theta \in \Theta$, here $\mathcal{P} = \{F_\theta : \theta \in \Theta\}$.

Example. Let $X_1, \ldots, X_n$ be a random sample from $U(\theta, \theta+1)$, $\theta \in \Theta = \mathbb{R}$. Show that $V_1(\underline{X}) = X_{(n)} - X_{(1)}$ and $V_2(\underline{X}) = X_2 - X_1$ are ancillary statistics.

Example (Location Family). Let $X_1, \ldots, X_n$ be a random sample with common d.f. $F(x-\theta)$, $-\infty<\theta<\infty$ and p.d.f. $f(x-\theta)$, $-\infty<\theta<\infty$. Let $V(\underline{X})$ be any statistic such that $V(X_1+c, \ldots, X_n+c) = V(X_1, \ldots, X_n)$, $\forall c \in \mathbb{R}$. Show that $V(\underline{X})$ is ancillary for $\theta \in \Theta = \mathbb{R}$. In particular show that $V_1(\underline{X}) = X_{(n)} - X_{(1)}$ and $V_2(\underline{X}) = \sum_{i=1}^{n}(X_i - \bar{X})^2$ are ancillary for $\theta \in \Theta$.

Example (Scale Family). Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with common d.f. $F(\frac{x}{\theta})$, $\theta>0$ and p.d.f. $\frac{1}{\theta} f(\frac{x}{\theta})$, $\theta>0$. Let $V(\underline{X})$ be any statistic such that $V(cX_1, \ldots, cX_n) = V(X_1, \ldots, X_n)$ $\forall c \in \mathbb{R}$ (or $\forall c > 0$). Show that $V(\underline{X})$ is ancillary for $\theta \in \Theta$. In particular show that $V_1(\underline{X}) = \frac{X_{(n)}}{X_{(1)}}$, $V_2(\underline{X}) = \frac{\bar{X}}{X_1}$ and $V_3(\underline{X}) = \frac{X_1}{X_2}$ are ancillary.

Example. Let $X_1, \ldots, X_n$ be a random sample from $U(\theta, \theta+1)$ where $\theta \in \Theta = \mathbb{R}$. Then $T_1(\underline{X}) = (X_{(1)}, X_{(n)})$ is minimal sufficient for $\theta \in \Theta$. Consequently $T_2(\underline{X}) = \left(X_{(n)} - X_{(1)}, \frac{X_{(1)} + X_{(n)}}{2}\right)$ is also minimal sufficient. However $T_{21}(\underline{X}) = X_{(n)} - X_{(1)}$ is ancillary for $\theta \in \Theta$. Thus, in this case, ancillary statistic is an important component of the minimal sufficient statistic. Certainly the ancillary statistic and the minimal sufficient statistic are not independent.

Remark Although an ancillary statistic by itself contains no information about the parameter θ but when it is used in conjuction with other statistics, sometimes does contain valuable information about $\theta \in \Theta$.

Example Let X_1 and X_2 be iid rvs with common p.m.f.

$$f_\theta(x) = \begin{cases} \frac{1}{3}, & \text{if } x \in \{\theta, \theta+1, \theta+2\} \\ 0, & \text{otherwise} \end{cases},$$

where $\theta \in \Theta = \{0, \pm 1, \pm 2, \ldots\}$ is unknown.

Here $T(\underline{X}) = (X_{(1)}, X_{(2)})$ and $T_1(\underline{X}) = \left(X_{(2)} - X_{(1)}, \frac{X_{(1)} + X_{(2)}}{2}\right)$ are minimal sufficient for $\theta \in \Theta$. Clearly $T_{11}(\underline{X}) = X_{(2)} - X_{(1)}$ is ancillary for $\theta \in \Theta$ (location family). Although $T_{11}(\underline{X})$ is ancillary, it gives information about θ. To see this consider a sample point (t_{11}, t_{12}), where t_{12} is an integer.

$T_{12} = t_{12} \Rightarrow \theta = t_{12}$ or $\theta = t_{12} - 1$ or $\theta = t_{12} - 2$

However with $t_{11} = 2$ and $t_{12} = 2$, $x_{(1)} = m-1$, and $x_{(2)} = m+1$

$\Rightarrow \theta = m-1$

Thus the knowledge about the ancillary $T_{11}(\underline{X}) = X_{(2)} - X_{(1)}$ has increased our knowledge about $\theta \in \Theta$

~~In many situations our~~ Our intuition suggests that minimal sufficient statistic and ancillary statistic should be independent. In many situations this intuction is proved. Describing such situations require the following definition.

Example. Let $X_1, \ldots, X_n$ be i.i.d. with common p.d.f.

$$f_\theta(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x-\theta)^2}, \quad -\infty < x < \infty, \quad \theta \in \Theta = \mathbb{R}.$$

(Cauchy distribution)

Show that: (i) $T(\underline{X}) = (X_{(1)}, \ldots, X_{(n)})$ is minimal sufficient. (ii) $V(\underline{X}) = (X_{(n)} - X_{(1)}, \ldots, X_{(n)} - X_{(n-1)})$ is ancillary.

$\boxed{21/2}$

~~Definition (Complete Statistic)~~

Completeness

Examples of Cauchy family and normal family show that the degree of reduction that can be achieved through sufficiency is extremely variable.

Question: What are the situations where sufficiency provides substantial reduction of data?

The amount of reduction of data by a minimal sufficient statistic seems to be related to the amount of ancillary information it contains.

An ancillary statistic by itself contains no information about F (or θ) but minimal sufficient statistic may still contain much ancillary information (e.g., in Cauchy family $(X_{(n)}-X_{(1)}, \ldots, X_{(n)}-X_{(n-1)})$ is ancillary statistic although it is a function of minimal sufficient statistic $(X_{(1)}, \ldots, X_{(n)})$).

A sufficient statistic T appears to be most successful in reducing the data if no non-constant function of T is ancillary or first order ancillary, i.e.

$$E_\theta(\psi(T)) = c, \quad \forall\, \theta \in \Theta \Rightarrow P_\theta(\psi(T)=c) = 1, \quad \forall\, \theta \in \Theta.$$

Definition The family of p.d.f.s / p.m.f.s $\{f_\theta(\cdot) : \theta \in \Theta\}$ is said to be complete (boundedly complete) if, for every function $\psi(\cdot)$ (bounded function), ~~the~~ ~~reducing~~ with

$$E_\theta(\psi(X)) = 0, \quad \forall\, \theta \in \Theta$$

we have

$$P_\theta(\psi(X)=0) = 1, \quad \forall\, \theta \in \Theta.$$

~~Exam~~ **Remark:** ~~A~~ Every complete family is boundedly complete. A boundedly complete family may not be complete.

Example: Let X be a r.v. with p.m.f. $f_\theta(x) = \begin{cases} \theta, & x=-1 \\ (1-\theta)^2\theta^x, & x=0,1,2,\ldots \\ 0, & \text{otherwise} \end{cases}$, where $\theta \in \Theta = (0,1)$. Then $\{f_\theta : \theta \in \Theta\}$ is boundedly complete but not complete.

$\boxed{22/2}$

Definition A statistic $T(\underline{X})$ is said to be complete (boundedly complete) if the family of distributions of T is complete (boundedly complete).

Example Let $X_1, \dots, X_n$ be i.i.d. Poisson(θ) r.v.s, where $\theta \in \Theta = (0, \infty)$. Find a minimal sufficient statistic for θ and check if it is complete.

Example For a fixed positive integer m, let $X_1, \dots, X_n$ be i.i.d. Bin(m, θ) where $\theta \in \Theta = (0, 1)$. Find a minimal sufficient statistic for θ and check if it is complete.

Example Let $X_1, \dots, X_n$ be i.i.d. $U(0, \theta)$, where $\theta \in \Theta = (0, \infty)$. Find a minimal sufficient statistic for θ and verify if it is complete.

Example Let $X_1, \dots, X_n$ be i.i.d. $N(0, \theta)$, where $\theta \in \Theta = (0, \infty)$. Show that the statistic $T_1(\underline{X}) = \sum_{i=1}^{n} X_i$ is not complete, whereas the statistic $T_2(\underline{X}) = \sum_{i=1}^{n} X_i^2$ is complete.

Example Show that the family $\mathcal{P} = \{N(\theta, 1) : -\infty < \theta < \infty\}$ is complete.

Example Let $X_1, \dots, X_n$ be i.i.d. $N(\theta, \theta^2)$ where $\theta \in \Theta = \mathbb{R}$. Show that the statistic $T(\underline{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is minimal sufficient for θ but is not complete.

Example Let $X_1, \dots, X_n$ be i.i.d. with p.m.f.

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}, & \text{if } x \in \{1, 2, \dots, \theta\} \\ 0, & \text{otherwise} \end{cases},$$

where $\theta \in \Theta = \{1, 2, \dots\} = \mathbb{N}$.

(i) Show that the family $\{f_\theta : \theta \in \Theta\}$ is complete.

(ii) Let $n_0 \in \mathbb{N}$. Show that the family $\{f_\theta : \theta \in \Theta - \{n_0\}\}$ is not complete.

(iii) Show that the statistic $T(\underline{X}) = X_{(n)}$ is complete.

Definition (Exponential Family of Distributions). A family $\{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$ is said to form a k-parameter exponential family if the d.f.s F_θ have pd.fs / pm.fs of the form

$$f_\theta(x) = \exp\left[\sum_{i=1}^{k} \eta_i(\theta)\, T_i(x) - B(\theta)\right] h(x),$$

where η_i and B are real-valued functions of the parameters and T_is are real-valued statistic

Theorem If a minimal sufficient statistic exists, then any complete-sufficient statistic is also a minimal sufficient statistic.

Proof. Let U be a minimal sufficient statistic and let T be a complete-sufficient statistic. Then $U = h(T)$ w.p.1, for some function $h(\cdot)$. (as minimal sufficient statistic is a function of any other sufficient statistic).

Define

$$\eta(U) = E_\theta(T \mid U),$$

which does not depend on $\theta \in \Theta$ as U is sufficient.

Then

$$\eta(h(T)) = E_\theta(T \mid U)$$

$$\Rightarrow \quad E_\theta(T - \eta(h(T))) = E_\theta^{U}\left[E_\theta^{T|U}\{(T - \eta(h(T)))\}\right] \quad (h(T) = U)$$

$$= E_\theta^{U}\left[E_\theta^{T|U}[T - \eta(U)]\right] \quad \forall\, \theta \in \Theta \quad (E_\theta^{T|U}(T) = \eta(U))$$

$$= 0,$$

$$\Rightarrow \quad T = \eta(h(T)) = \eta(U) \quad \text{w.p.1}$$

$\Rightarrow$ T is also minimal sufficient.

Theorem (Complete-Sufficient statistic for exponential family).

Let $X_1, \dots, X_n$ be a random sample from an exponential family with p.d.f. / p.m.f. given by

$$f_{\underline{\theta}}(\underline{x}) = \exp\left[\sum_{i=1}^{k} \eta_i(\underline{\theta})\, T_i(\underline{x}) - B(\underline{\theta})\right] h(\underline{x}), \quad \underline{x} \in \mathcal{X}, \quad \underline{\theta} \in \Theta \subseteq \mathbb{R}^k$$

where $k \le n$. Then

(a) $\underline{T} = (T_1, \dots, T_k)$ is a sufficient statistic for $\underline{\theta}$;

(b) if $\underline{\eta} = (\eta_1, \dots, \eta_k)$ and the range of $\underline{\eta}$ contains an open rectangle in $\mathbb{R}^k$, then $\underline{T} = (T_1, \dots, T_k)$ is a complete-sufficient (and hence minimal sufficient) statistic for $\underline{\theta}$.

Proof. (a) Follows using factorization theorem.

(b) For a general proof refer to Lehmann (TSH, Sections Theorem). It can be seen that The idea is as follows. The p.d.f. / p.m.f. of $\underline{T} = (T_1, \dots, T_k)$ is

$$g_{\underline{\theta}}(\underline{t}) = c(\underline{\theta}) \exp\left[\sum_{i=1}^{k} \eta_i t_i\right] H(\underline{t}), \quad \underline{t} \in \mathbb{R}^k,$$

for some functions $c(\cdot)$ and $H(\cdot)$.

Now $E_{\theta}(\psi(\underline{T})) = 0, \quad \forall \underline{\theta} \in \Theta$

$$\Rightarrow \int_{\mathbb{R}^k} \psi(\underline{t})\, c(\underline{\theta}) \exp\left[\sum_{i=1}^{k} \eta_i t_i\right] H(\underline{t})\, d\underline{t} = 0, \quad \forall \underline{\eta}$$

$$\Rightarrow \int_{\mathbb{R}^k} \psi(\underline{t})\, H(\underline{t}) \exp\left[\sum_{i=1}^{k} \eta_i t_i\right] d\underline{t} = 0, \quad \forall \underline{\eta}$$

$\rightarrow$ Laplace transform of vanishing in an open set

$$\Rightarrow \psi(\underline{t})\, H(\underline{t}) = 0, \quad \text{a.e.} \quad \forall \underline{\eta}$$

$$\Rightarrow \psi(\underline{t}) = 0 \quad \text{a.e.} \quad \forall \underline{\eta} \quad \Rightarrow P_{\underline{\theta}}(\psi(\underline{T}) = 0) = 1 \quad \forall \underline{\theta}$$

$\frac{25}{2}$

Note: (i) The above part (b) of theorem may not hold if the range of $\eta = (\eta_1, \dots, \eta_k)$ does not contain an open rectangle in $\mathbb{R}^k$;

(ii) A minimal sufficient statistic may not be complete.

Example Let $X_1, \dots, X_n$ be a random sample from $N(\theta, \theta^2)$, where $\theta \in \Theta = \mathbb{R} - \{0\}$ is unknown. Show that $T(\underline{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$ is minimal sufficient but not complete.

Solution For $\underline{x} \in A = \mathbb{R}^n$,

$$f_\theta(\underline{x}) = \frac{1}{(|\theta| \sqrt{2\pi})^n} e^{-\frac{1}{2\theta^2}\left[\sum_{i=1}^n x_i^2 - 2\theta \sum_{i=1}^n x_i + n\theta^2 \right]}.$$

Here

$$(\eta_1, \eta_2) = \left(-\frac{1}{2\theta^2}, \frac{1}{\theta} \right)$$

and the range of (η_1, η_2) does not contain an open rectangle in $\mathbb{R}^2$ $\left(\eta_1 = -\frac{\eta_2^2}{2} \right)$. Thus we can not use last theorem to conclude that $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$ is complete-sufficient. ~~clearly~~ clearly $T(\underline{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right) = (T_1, T_2)$ is sufficient. ~~clearly~~ We have

$$E_\theta\left(\frac{1}{n-1}\left(T_2 - n\left(\frac{T_1}{n}\right)^2 \right) \right) = E_\theta(S_X^2) = \theta^2, \quad \forall\, \theta \in \Theta$$

$$E_\theta\left(\left(\frac{T_1}{n}\right)^2 \right) = E_\theta(\bar{X}^2) = \frac{\theta^2}{n} + \theta^2 = \frac{n+1}{n}\theta^2$$

$$\Rightarrow E_\theta\left(\frac{1}{n-1}\left(T_2 - n\left(\frac{T_1}{n}\right)^2 \right) - \frac{1}{n(n+1)} T_1^2 \right) = 0, \quad \forall\, \theta \in \Theta$$

$$\Rightarrow E_\theta\left[\frac{1}{n-1} T_2 - \frac{2}{n^2-1} T_1^2 \right] = 0, \quad \forall\, \theta \in \Theta$$

But

$$P_\theta\left(\frac{T_2}{n-1} - \frac{2}{n^2-1} T_1^2 \right) = 0, \quad \forall\, \theta \in \Theta$$

$\Rightarrow$ (T_1, T_2) is not complete

For $\underline{x}, \underline{y} \in A$

$$\frac{f_\theta(\underline{x})}{f_\theta(\underline{y})} = e^{\frac{1}{2\theta^2}\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2\right) - \frac{1}{\theta}\left(\sum_{i=1}^n y_i - \sum_{i=1}^n x_i\right)}$$

is independent of $\theta \in \Theta$

$$\Leftrightarrow \eta^2\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2\right) - \sqrt{2}\eta\left(\sum_{i=1}^n y_i - \sum_{i=1}^n x_i\right)$$

does not depend on $\eta > 0$

$$\Leftrightarrow \sum_{i=1}^n y_i = \sum_{i=1}^n x_i \quad \text{and} \quad \sum_{i=1}^n y_i^2 = \sum_{i=1}^n x_i^2.$$

$\Rightarrow$ $T(\underline{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ is minimal sufficient for $\theta \in \Theta$

Example Let $X_1, \ldots, X_n$ be i.i.d. $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ rvs, where $\theta \in \Theta = \mathbb{R}$. Show that $\underline{T} = (X_{(1)}, X_{(n)})$ is minimal sufficient but it is not complete

Theorem (Basu's Theorem). If $T(\underline{X})$ is a boundedly complete and sufficient statistic for the family $\mathcal{P} = \{F_\theta : \theta \in \Theta\}$ then $T(\underline{X})$ and any ancillary statistic $V(\underline{X})$ are independent.

Proof. For any measurable set A and $t \in \mathcal{T}$, define

$$\eta_A(t) = P_\theta(V(\underline{X}) \in A \mid T(\underline{X}) = t),$$

which is independent of $\underline{\theta} \in \Theta$ by virtue of sufficiency of $T(\underline{X})$. Then

$$E_\theta(\eta_A(T)) = E(P_\theta(V(\underline{X}) \in A \mid T)) = P_\theta(V(\underline{X}) \in A)$$
$$= p_A \quad \text{(does not depend on } \theta \text{ as } V \text{ is ancillary)}$$

$\Rightarrow$ $E_\theta(\eta_A(T) - p_A) = 0, \quad \forall \theta \in \Theta$

$\Rightarrow$ $\eta_A(t) = p_A$ (a.e. $\mathcal{P}$) $\Rightarrow$ $P_\theta(V(\underline{X}) \in A \mid T(\underline{X}) = t)$
$= P_\theta(V(\underline{X}) \in A) \quad \forall \theta \in \Theta$

$\Rightarrow$ $V(\underline{X})$ and $T(\underline{X})$ are independent

Example Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, \sigma^2)$, where $\theta \in \mathbb{R}$ and $\sigma > 0$.

(i) Let $V(\underline{X})$ be any statistic such that

$$V(X_1+c, \ldots, X_n+c) = V(X_1, \ldots, X_n), \quad \forall\, c \in \mathbb{R}.$$

Show that $\bar{X}$ and $V(\underline{X})$ are statistically independent. In particular show that $\bar{X}$ and $\sum_{i=1}^{n}(X_i-\bar{X})^2$ are statistically independent.

(ii) Suppose θ is fixed at $\theta_0 \in \mathbb{R}$ and $V(\underline{X})$ is any statistic satisfying

$$V(cX_1, \ldots, cX_n) = V(X_1, \ldots, X_n) \quad \forall\, c > 0.$$

Show that $T(\underline{X}) = \sum_{i=1}^{n}(X_i-\theta_0)^2$ and $V(X_1-\theta_0, \ldots, X_n-\theta_0)$ are statistically independent. In particular show that $\dfrac{\bar{X}-\theta_0}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}$ and $\sum_{i=1}^{n}(X_i-\theta_0)^2$ are statistically independent. Also $\dfrac{\bar{X}-\theta_0}{\sqrt{\sum_{i=1}^{n}(X_i-\theta_0)^2}}$ and $\sum_{i=1}^{n}(X_i-\theta_0)^2$ are independent.

Example Let $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ be independent random samples from $N(\theta_1, \sigma_1^2)$ and $N(\theta_2, \sigma_2^2)$, respectively. Show that $T = (\bar{X}, \sum_{i=1}^{n} X_i^2, \bar{Y}, \sum_{i=1}^{n} Y_i^2)$ and

$$V(\underline{X}) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2 \sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$$

are statistically independent.

Example Let $X_1, \ldots, X_n$ be i.i.d. $U(0, \theta)$, $\theta \in \Theta = (0, \infty)$. Show that $\dfrac{X_{(1)}}{X_{(n)}}$ and $X_{(n)}$ are independent. Hence show that $E_\theta(X_{(1)} X_{(n)}) = E_\theta\left(\dfrac{X_{(1)}}{X_{(n)}}\right) E_\theta(X_{(n)}^2)$, $\forall\, \theta \in \Theta$.