CHAPTER 2

SUFFICIENCY

§2.1 INTRODUCTION

Let $X_1,\ldots,X_n$ be a random sample from a density $f(\cdot\,;\,\theta)$, and let Ω, called the parameter space, denote the set of possible valuse that the parameter can assume. Recall that a statistic is a function of the sample. Any statistic, $T = \not{t}(X_1,\ldots,X_n)$, defines a from of data reduction or data summary; it condenses the n random variables $X_1,\ldots,X_n$ into a single random variable. Data reduction in terms of a statistic can also be viewed another way. Let $\mathfrak{X}$ be the range of values that $(X_1,\ldots,X_n)$ can take on. Then, a particular statistic induces or defines a partition of $\mathfrak{X}$.

EX 2.1 : A random sample of size 3 is selected from a Bernoulli distribution, then

$$\mathfrak{X} = \left\{ \left(X_1, X_2, X_3 \right) : X_i = 0 \text{ or } 1,\ i = 1,2,3 \right\}$$

$$= \left\{ (0,0,0), (0,0,1), \ldots, (1,1,1) \right\}$$

and Card($\mathfrak{X}$)=8. Let $\mathscr{T}\left(x_1, x_2, x_3\right) = x_1 + x_2 + x_3$.

Then the partition of $\mathfrak{X}$ induced by $\mathscr{T}(\cdot,\cdot,\cdot)$ is

$$\left\{ \left(x_1, x_2, x_3 \right) : \mathscr{T}\left(x_1, x_2, x_3 \right) = 0,1,2,3 \right\}$$

$$= \left\{ (0,0,0) \right\} \cup \left\{ (0,0,1), (0,1,0), (1,0,0) \right\} \cup$$

$$\left\{ (0,1,1), (1,0,1), (1,1,0) \right\} \cup \left\{ (1,1,1) \right\}$$

§2.2 SUFFICIENT STATISTIC

A sufficient statistic is a particular statistic that condenses $\mathfrak{X}$ so that no information about θ is lost. It is good for making inferences about θ.

<u>Def 2.1</u> : A statistic $T = t\left(X_1,\cdots,X_n\right)$ is defined to be a sufficient statistic if the conditional distribution of $X_1,\ldots,X_n$ given $T = t$ does not dependent on θ for any value t of T.

- To use the above definition to verify a statistic T is a sufficient statistic for θ, we need to verify that the conditional probability $P\left(X_1 = x_1, \ldots, X_n = x_n \middle| T = t\right)$ does not depend on θ for all fixed values of $\left(x_1, \ldots, x_n\right)$ and t.

Since $\left\{X_1 = x_1, \ldots, X_n = x_n\right\}$ is a subset of $\left\{T = t\right\}$,

$$P\left(X_1 = x_1, \ldots, X_n = x_n \middle| T = t\right)$$

$$= \frac{P\left(X_1 = x_1, \ldots, X_n = x_n \text{ and } T = t\right)}{P\left(T = t\right)}$$

$$= \frac{P\left(X_1 = x_1, \ldots, X_n = x_n\right)}{P\left(T = t\right)}$$

$$= \frac{f_{X_1,\cdots,X_n}\left(x_1,\ldots,x_n;\theta\right)}{g_T\left(t;\theta\right)},$$

where $f_{X_1,\ldots,X_n}\left(x_1,\ldots,x_n;\theta\right)$ is the joint pmf of the sample and $g_T\left(t;\theta\right)$ is the pmf of T.

- Note that Equation (2.1) is still appropriate to use if $\left(X_1,\ldots,X_n\right)$ and T have continuous distributions.

EX 2.2 : $X_1,\ldots,X_n \sim iid$ Bernoulli(θ)

$f(x;\theta)=\theta^x(1-\theta)^{1-x}$, $x=0,1,\ 0\le\theta\le 1$. Let $T=\sum_{i=1}^{n} X_i$. Then

$T\sim B(n,\theta)$ and its pmf is $g(t;\theta)=\binom{n}{t}\theta^t(1-\theta)^{n-t}$, $t=0,\ldots,n$

$$\therefore\ P\left(X_1=x_1,\ldots,X_n=x_n \middle| T=t\right)=\frac{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{g_T(t)}$$

$$=\frac{\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}}=\frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}}=\frac{1}{\binom{n}{t}},$$

and it does not depent on θ. T is a sufficient statistic.

Def 2.2 : Let $X_1,\ldots,X_n$ be a random sample from a density $f(\cdot\,;\theta)$, where θ may be a vector. The statistics $T_1,\ldots,T_r$ are defined to be jointly sufficient if the conditional distribution of $X_1,\ldots,X_n$ given $T_1=t_1,\ldots,T_r=t_r$ does not depend on θ.

Thm 2.1 : If $T_1=\ell_1(X_1,\ldots,X_n),\ldots,T_r=\ell_r(X_1,\ldots,X_n)$ is a set of jointly sufficient statistics, then any set of one-to-one functions of $T_1,\ldots,T_r$ is also sufficient.

EX 2.3 : $X_1, \ldots, X_n \sim \ iid \ N \ (\mu, \sigma^2)$

$\sum X_i$ and $\sum X_i^2$ are jointly sufficient.

$\Rightarrow \bar{X}$ and $\sum \left(X_i - \bar{X} \right)^2$ are also jointly sufficient.

Thm 2.2 :(Factorization theorem)

Let $X_1, \ldots, X_n$ be a random sample from a density $f\left(\cdot \, ; \theta \right)$, where θ may be a vector. A set of statistics $T_1 = \ell_1(X_1, \ldots, X_n), \ldots, T_r = \ell_r(X_1, \ldots, X_n)$ is jointly sufficient iff there exists functions $g\left(t_1, \ldots, t_r ; \theta \right)$ and $h\left(x_1, \ldots, x_n \right)$ such that for all $x_1, \ldots, x_n$ and θ,

$$f_{X_1, \ldots, X_n}\left(x_1, \ldots, x_n ; \theta \right) = g\left(t_1, \ldots, t_r ; \theta \right) h\left(x_1, \ldots, x_n \right).$$

EX 2.4 : $X_1,\ldots,X_n$ ~ Bernoulli(θ)

$f(x;\theta) = \theta^x (1-\theta)^{1-x}$, $x = 0,1$, $0 \le \theta \le 1$. Then

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \prod_{i=1}^{n} f(x;\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

Set $t = T(x_1,\ldots,x_n) = \sum x_i$ and take

$g(t;\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$ and $h(x_1,\ldots,x_n) = 1$.

So $T = \sum X_i$ is a sufficient statistic.

EX 2.5 : $X_1, \ldots, X_n \sim \ iid \ N\left(\mu, \sigma^2\right), \ \theta = \left(\mu, \sigma^2\right)$

$$\prod_{i=1}^{n} f\left(x_i;\theta\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}\left(x_i-\mu\right)^2\right]$$

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

$$= \frac{1}{\left(2\pi\right)^{n/2}\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\left(\sum x_i^{\,2} - 2\mu\sum x_i + n\mu^2\right)\right]$$

Set $t_1 = \sum x_i$ and $t_2 = \sum x_i^{\,2}$. Then $\sum X_i$ and $\sum X_i^{\,2}$ are jointly sufficient statistics. By Thm2.1, $\bar{X} = \frac{1}{n}\sum X_i$ and $S^2 = \frac{1}{n-1}\sum\left(X_i - \bar{X}\right)^2$ are also jointly sufficient.

EX 2.6 : $X_1,\ldots,X_n$: random sample from a pdf

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1} & , x\in(0,1)\\ 0 & , \text{o.w.}\end{cases}$$

Then

$$\prod_{i=1}^{n} f(x_i;\theta)=\prod_{i=1}^{n}\theta x_i^{\theta-1}=\theta^n\left(x_1\cdots x_n\right)^{\theta-1}$$

$$=\theta^n\left(x_1\cdots x_n\right)^{\theta}\cdot\left(\frac{1}{x_1\cdots x_n}\right)$$

Set $t=x_1\cdots x_n$ and take $g(t;\theta)=\theta^n\left(x_1\cdots x_n\right)^{\theta}$

and $h(x_1\cdots x_n)=\dfrac{1}{x_1\cdots x_n}$. Then $T=\prod_{i=1}^{n}X_i$ is a sufficient statistic.

EX 2.7 : $X_1,\ldots,X_n \sim \; iid \; U(\theta_1,\theta_2),\; \theta \in [\theta_1,\theta_2]$.

Then

$$f(x;\theta) = \frac{1}{\theta_2 - \theta_1} \cdot I_{[\theta_1,\theta_2]}(x),$$

and

$$\prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} \cdot I_{[\theta_1,\theta_2]}(x_i)$$

$$= \frac{1}{(\theta_2 - \theta_1)^n} \cdot I_{[\theta_1,y_n]}(y_1) \cdot I_{[y_1,\theta_2]}(y_n).$$

Set $t_1 = y_1,\; t_2 = y_n$, and $h(x_1,\ldots,x_n) = 1$. Then Y_1 and Y_n are jointly sufficient.

Further, if $\theta_1 = 0$ and $\theta_2 = \theta$, then

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

$$= \frac{1}{\theta^n} I_{[0,y_n]}\left(y_1\right) \cdot I_{[0,\theta]}\left(y_n\right)$$

$$= \frac{1}{\theta^n} I_{[0,\theta]}\left(y_n\right) \cdot I_{[0,y_n]}\left(y_1\right).$$

Set $t = y_n$, then Y_n is a sufficient statistic.

Def 2.3 : A family of densities $f\left(\cdot\ ;\theta_1,\ldots,\theta_k\right)$ is called an exponential family if it can be expressed as

$$f\left(x\ ;\theta_1,\ldots,\theta_k\right)=a\left(\theta_1,\ldots,\theta_k\right)b\left(x\right)\exp\left[\sum_{j=1}^{k}c_j\left(\theta_1,\ldots,\theta_k\right)d_j\left(x\right)\right],$$

for $-\infty < x < \infty$ and for all $\theta \in \Omega$, and for a suitable choice of functions $a\left(\cdot,\ldots,\cdot\right)$, $b\left(\cdot\right)$, $c_j\left(\cdot,\ldots,\cdot\right)$ and $d_j\left(\cdot\right)$.

EX 2.8 : $f\left(x;\theta\right)=\theta e^{-\theta x}\cdot I_{\left(0,\infty\right)}(x)$

$\Rightarrow f\left(x;\theta\right)$ belongs to the exponential family.

EX 2.9 : $X \sim$ $iid\ N\ (\mu,\sigma^2),\ \theta=(\theta_1,\theta_2)=(\mu,\sigma^2)$

$$f(x;\mu,\sigma)=\phi_{\mu,\sigma^2}(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

$$=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2\sigma^2}\left(x^2-2\mu x+\mu^2\right)\right]$$

$$=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)\exp\left[-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x\right]$$

$\Rightarrow \left\{N\ (\mu,\sigma^2)\right\}$ belongs to the exponential family.

Remark 2.1 : Any family of densities for which the range of valuse where the density is nonnegative depends on θ does not belong to the exponential family.

EX 2.10 :

(1) A family of uniform densities does not belong to the exponential family.

(2) A family of Binomial densities with n, p unknown does not belong to the exponential family. If n is known and $\theta = p$, then

$$f(x,p) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\ldots,n\}}(x)$$

$$= \binom{n}{x} I_{\{0,1,\cdots,n\}}(x) \exp\left[x\ln p + (c-x)\ln(1-p)\right].$$

So $\{B(n,p)\}$ with n known belongs to the exponential family.

<u>Thm 2.3</u> : Let $X_1,\ldots,X_n$ be a random sample from a pdf or pmf $f\left(x\,;\theta_1,\ldots,\theta_k\right)$ that belongs to an exponential family given by

$$f\left(x\,;\theta_1,\ldots,\theta_k\right)=a\left(\theta_1,\ldots,\theta_k\right)b\left(x\right)\exp\left[\sum_{j=1}^{k}c_j\left(\theta_1,\ldots,\theta_k\right)d_j\left(x\right)\right].$$

Then $\sum_{i=1}^{n}d_1\left(X_i\right),\ldots,\sum_{i=1}^{n}d_k\left(X_i\right)$ is a set of jointly sufficient statistics.

EX2.11 : $X \sim \ iid \ Beta(\theta_1, \theta_2)$. Then

$$f\left(x;\theta_1,\theta_2\right) = \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x^{\theta_1-1} x^{\theta_2-1} I_{[0,1]}(x)$$

$$= \frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} I_{[0,1]}(x) \exp\left[\left(\theta_1-1\right)\ln x + \left(\theta_2-1\right)\ln\left(1-x\right)\right]$$

So $\left\{Beta(\theta_1,\theta_2)\right\}$ belongs to the exponential family.

By Thm2.3, $\sum_{i=1}^{n} \ln\left(X_i\right)$ and $\sum_{i=1}^{n} \ln\left(1-X_i\right)$ are jointly sufficient statistics.

§2.3 MINIMAL SUFFICIENT STATISTICS

Recall that the purpose of a sufficient statistics is to achieve data reduction without losing any important information about the unknown parameter θ. Thus, a statistic that achieves the most data condensation while still retaining all the information about θ might be preferable. A formal definition of such a statistic is given below.

Def 2.4 : A sufficient statistic $T = t\left(X_1,\dots,X_n\right)$ is called a minimal sufficient statistic if it is a function of any other sufficient statistic $T^* = t^*\left(X_1,\dots,X_n\right)$.

- To say that $T(x_1,\ldots,x_n)$ is a function of $T^*(x_1,\ldots,x_n)$, if $T^*(x_1,\ldots,x_n) = T^*(x_1',\ldots,x_n')$, then $T(x_1,\ldots,x_n) = T^*(x_1',\ldots,x_n')$.

- In terms of partition sets, if $\{B_{t^*}\}$ are the partition sets for T^* and $\{A_t\}$ are the partition sets for T, then every B_{t^*} is a subset of some A_t.

Thm 2.4 : Let $X_1,\ldots,X_n$ be a random sample from a pdf or pmf $f(x;\theta)$. For every two sample values $x_1,\ldots,x_n$ and $x_1',\ldots,x_n'$, if there exists a function $T=\ell(X_1,\ldots,X_n)$ such that the ratio

$$\frac{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta)}{f_{X_1,\ldots,X_n}(x_1',\ldots,x_n';\theta)}$$

does not depend on θ iff $\ell(x_1,\ldots,x_n)=\ell(x_1',\ldots,x_n')$, then T is a minimal sufficient statistic for θ.

EX 2.12 : $X_1,\ldots,X_n \sim iid\ N(\mu,\sigma^2)$

(i) If μ and σ^2 are unknown, then

$$\frac{\prod_{i=1}^{n} f(x_i;\mu,\sigma^2)}{\prod_{i=1}^{n} f(x_i';\mu,\sigma^2)} = \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}(x_i-\mu)^2\right]}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}(x_i'-\mu)^2\right]}$$

$$= \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^{\,2} + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i\right]}{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i'^{\,2} + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i'\right]}$$

$$= \exp\left\{\frac{\mu}{\sigma^2}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i'\right) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^{\,2} - \sum_{i=1}^{n} x_i'^{\,2}\right)\right\} \cdots\cdots\cdots (*)$$

$(*)$ does not depend on μ and σ^2 if and only if

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i' \text{ and } \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i'^2.$$

So $\sum_{i=1}^{n} X_i$ and $\sum_{i=1}^{n} X_i^2$ are jointly minimal sufficient statistics for μ and σ^2.

(ii) If σ^2 is known and μ is unknown, then $(*)$ does not depend on μ if and only if $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i'$. So $\sum_{i=1}^{n} X_i$ is a minimal sufficient statistic.

(iii) If μ is known and σ^2 is unknown, then

$$\frac{\prod_{i=1}^{n} f\left(x_i;\sigma^2\right)}{\prod_{i=1}^{n} f\left(x_i';\sigma^2\right)} = \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i'-\mu\right)^2\right]}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}\left(x_i-\mu\right)^2-\sum_{i=1}^{n}\left(x_i'-\mu\right)^2\right]\right\}\cdots\cdots(**)$$

$(**)$ does not depend on σ^2 if and only if

$$\sum_{i=1}^{n}\left(x_i-\mu\right)^2=\sum_{i=1}^{n}\left(x_i'-\mu\right)^2$$

So $\sum_{i=1}^{n}\left(X_i-\mu\right)^2$ is a minimal sufficient statistic.

§2.4 ANCILLARY STATISTICS

Def 2.5 : A statistic $T = t(X_1, \dots, X_n)$ is called an ancillary statistic if its distribution does not depend on the parameter θ.

EX 2.13 : $X_1, \dots, X_n \sim \ iid \ U(\theta, \theta+1), \ -\infty < x < \infty$.
$Y_1, \dots, Y_n$: order statistics corresponding to the sample.
As in EX 1.5 , Sec1.4 , we can show that the pdf of the sample range R is $f_R(r) = n(n-1)(1-r)r^{n-2}, \ 0 < r < 1$.
i.e. $R \sim Beta(n-1, 2)$. Thus, the distribution of R does not depend on θ and R is an ancillary statistic.

Def 2.6 : Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\theta)$, indexed by θ, $\theta \in (-\infty, \infty)$, is called the location family with standard pdf $f(x)$ and θ is called the location parameter for the family.

- The above definition states that, if W is a random variable with pdf $f(w)$, then the pdf of the random variable $X = W + \theta$ is $f(x-\theta)$.

EX 2.14 : Suppose $X \sim N\,(0,1)$. To form a location family, we replace x with $x-\mu$ to obtain

$$f(x;\mu)=\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x-\mu)^2\right].$$

So $\{N\,(\mu,1)\}$ is a location family.

EX 2.15 : Suppose $X_1,\ldots,X_n$ is a random sample such that $X_i = W_i + \theta,\ i=1,\ldots,n,$ where $-\infty < \theta < \infty$ and $W_1,\ldots,W_n$ are iid r.v.'s with p.d.f $f(w)$ which does not depend on θ. Then the common pdf of X_i is $f(x-\theta)$. Let $Y_1,\ldots,Y_n$ be the order statistics in the sample. The cdf of the sample range $R = Y_n - Y_1$ is

$$
\begin{aligned}
F_R(r;\theta) &= P\left(Y_n - Y_1 < r\right) \\
&= P\left(\max_i X_i - \min_i X_i \le r\right) \\
&= P\left(\max_i (W_i + \theta) - \min_i (W_i + \theta) \le r\right) \\
&= P\left(\max_i W_i + \theta - \min_i W_i - \theta \le r\right) \\
&= P\left(\max_i W_i - \min_i W_i \le r\right).
\end{aligned}
$$

So the c.d.f of R does not depend on θ and hence R is an ancillary statistic.

Def 2.7 : A statistic $T = \ell\left(X_1, \ldots, X_n\right)$ is defined to be location-invariant iff $\ell\left(x_1 + c, \ldots, x_n + c\right) = \ell\left(x_1, \ldots, x_n\right)$ for all values $x_1, \ldots, x_n$ and all real c.

<u>EX 2.16</u> : Assume that $X_1,\ldots,X_n$ is a random sample such that $X_i = W_i + \theta,\ i = 1,\ldots,n,$ where $-\infty < \theta < \infty$ is a parameter and $W_1,\ldots,W_n$ are iid r.v.'s with pdf $f(w)$ which does not depend on θ. Then the common pdf of X_i is $f(x-\theta)$. Let T be a location-invariant statistic. Then

$$
\begin{aligned}
T &= t(x_1,\ldots,x_n) \\
&= t(w_1+\theta,\ldots,w_n+\theta) \\
&= t(w_1,\ldots,w_n)
\end{aligned}
$$

is a function of $W_1,\ldots,W_n$ alone. Hence T must have a distribution that does not dependent on θ ; i.e T is an ancillary statistic.

Def 2.8 : Let $f(x)$ be any pdf. Then for any $\theta > 0$, the family of pdfs $\frac{1}{\theta} f\left(\frac{x}{\theta}\right)$, indexed by θ, is called the scale family with standard pdf $f(x)$ and θ is called the scale parameter for the family.

- The above definition states that, if W is a random variable with pdf $f(w)$, then the pdf of the random variable $X = \theta \cdot W$ is $\frac{1}{\theta} f\left(\frac{1}{\theta}\right)$.

Def 2.9 : Let $f(x)$ be any pdf. Then for any $-\infty < \theta_1 < \infty$ and $\theta_2 > 0$, the family of pdfs $\frac{1}{\theta_2} f\left(\frac{x-\theta_1}{\theta_2}\right)$, indexed by the parameters θ_1 and θ_2, is called the location-scale family with standard pdf $f(x)$.

Def 2.10 : A statistic $T = \ell(X_1, \ldots, X_n)$ is defined to be scale-invariant if and only if

$$\ell(cx_1, \ldots, cx_n) = \ell(x_1, \ldots, x_n)$$

for all values $x_1, \ldots, x_n$ and all real $c > 0$.

EX 2.17 : $X_1,\ldots,X_n \sim \; iid \; F\left(\frac{x}{\theta}\right)$, $\theta > 0$. Then any statistic that depend on the sample only through the $n-1$ values $\frac{x_1}{x_n},\ldots,\frac{x_{n-1}}{x_n}$ is an ancillary statistic.

For example,

$$\frac{x_1+\cdots+x_n}{x_n} = \frac{x_1}{x_n}+\cdots+\frac{x_{n-1}}{x_n}+1$$

is an ancillary statistic. To see this fact, let $W_1,\ldots,W_n \sim \; iid \; F(w)$ with $x_i = \theta w_i$.

The joint c.d.f of $\frac{x_1}{x_n},\ldots,\frac{x_{n-1}}{x_n}$ is

$$F\left(z_1,\ldots,z_n;\theta\right)=P_\theta\left(\frac{x_1}{x_n}\le z_1,\ldots,\frac{x_{n-1}}{x_n}\le z_{n-1}\right)$$

$$=P_\theta\left(\frac{\theta w_1}{\theta w_n}\le z_1,\ldots,\frac{\theta w_{n-1}}{\theta w_n}\le z_{n-1}\right)$$

$$=P_\theta\left(\frac{w_1}{w_n}\le z_1,\ldots,\frac{w_{n-1}}{w_n}\le z_{n-1}\right),$$

which does not depend on θ.

<u>EX 2.18</u> : Suppose $X_1,\ldots,X_n$ is a random sample such that $X_i = \theta W_i,\ i = 1,\ldots,n,$ where $\theta > 0$ and $W_1,\ldots,W_n$ are iid r.v.'s with p.d.f $f(w)$ which does not depend on θ. Then the common pdf of X_i is $\frac{1}{\theta} f(x-\theta)$. Let $T = t(X_1,\ldots,X_n)$ be a scale-invariant statistic. Then

$$T = t(x_1,\ldots,x_n) = t(\theta w_1,\ldots,\theta w_n) = t(w_1,\cdots,w_n).$$

Since neither the joint dist. of w_i nor T contain θ, the distribution of T does not depend on θ. T is an ancillary statistic.

§2.5 Sufficiency, completeness, and independence

EX 2.19 : Let X_1 and X_2 be iid observations from the discrete distribution

$$P_\theta\left(x=\theta\right)=P\left(x=\theta+1\right)=P\left(x=\theta+2\right)=\frac{1}{3},$$

where θ is an integer. Let $Y_1 \le Y_2$ be the order statistics in the sample. We can show that $R = Y_2 - Y_1$ and $M = \dfrac{Y_1 + Y_2}{2}$ are jointly minimal sufficient statistics.

Consider a sample point, where m is an integer.

If we only comsider m, then $\theta = m$ or $\theta = m-1$ or $\theta = m-2$. Suppose we know $R = 2$, then $Y_1 = m-1$ and $Y_2 = m+1$, with this additional information, the only possible value for θ is $m-1$.

<u>Def 2.11</u> : Let $X_1,\ldots,X_n$ be a random sample from a pdf or pmf $f\left(x\,;\theta\right)$, and let $T = \not{\ell}\left(X_1,\ldots,X_n\right)$ be a statistic. The family of probability distributions of T is defined to be complete iff $E_\theta\left[z\left(T\right)\right] = 0$ for all θ implies $P\left[z\left(T\right) = 0\right] = 1$ for all θ. Equivalently, the statistic $T = \not{\ell}\left(X_1,\ldots,X_n\right)$ is called a complete statistic.

EX 2.20 : $X \sim N$ (0,1), $z(x) = x$. Then

$$E[z(x)] = E(x) = 0.$$

But $P[g(x) = 0] = P(x = 0) = 0$. This is a particular distribution, not a family of distributions. If $X \sim N$ $(\theta,1)$, We can show that if $E(x) = 0$, then $P(x = 0) = 1$.

EX 2.21 : $X_1,\ldots,X_n \sim$ iid Bernoulli(θ), $0 < \theta < 1$.

Let $T = \sum_{i=1}^{n} X_i$ and let z be a function of T such that $E_\theta[z(T)] = 0$. Then

$$0 = E_\theta\left[z\left(T\right)\right] = \sum_{t=0}^{n} z\left(t\right)\binom{n}{t}\theta^t\left(1-\theta\right)^{n-t}$$

$$= \left(1-\theta\right)^n \sum_{t=0}^{n} z\left(t\right)\binom{n}{t}\left(\frac{\theta}{1-\theta}\right)^t.$$

Since $\left(1-\theta\right)^n \neq 0,\ \forall\ 0<\theta<1,$

$$0 = \sum_{i=0}^{n} z\left(t\right)\binom{n}{t}\left(\frac{\theta}{1-\theta}\right)^t = \sum_{i=0}^{n} z\left(t\right)\binom{n}{t}\alpha^t,\ \alpha = \frac{\theta}{1-\theta}.$$

So $\binom{n}{t}z\left(t\right) = 0,\ \forall t = 0,\ldots,n$. But $\binom{n}{t} \neq 0$ and so $z\left(t\right) = 0,\ \forall t = 0,\ldots,n$. i.e. $P\left[z\left(T\right) = 0\right] = 1$. Thus, T is a complete statistic.

EX 2.22 : $X_1, \ldots, X_n \sim$ *iid* $U(0,\theta)$, $\theta > 0$. Then Y_n is a sufficient statistic. Suppose $z(Y_n)$ is a function of Y_n such that $E_\theta\left[z(Y_n) \right] = 0$. Then

$$0 = E_\theta\left[Z(Y_n) \right] = \int_0^\theta z(y_n) f_{Y_n}(y_n)\, dy_n$$

$$= \int_0^\theta z(y_n) \cdot n \cdot \left(\frac{y_n}{\theta} \right)^{n-1} \frac{1}{\theta} dy_n = \frac{n}{\theta^n} \int_0^\theta z(y_n) y_n^{\,n-1} dy_n$$

$$\Rightarrow \int_0^\theta z(y_n) y_n^{\,n-1} dy_n = 0 \Rightarrow \frac{d}{d\theta} \int_0^\theta z(y_n) y_n^{\,n-1} dy_n = 0$$

$$\Rightarrow z(\theta) \cdot \theta^{n-1} = 0 \Rightarrow z(\theta) = 0, \ \forall \theta > 0.$$

Thus, Y_n is complete.

Thm 2.5 : If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic.

Thm 2.6 : Let $X_1,\ldots,X_n$ be a random sample from a pdf or pmf $f\left(x;\theta_1,\ldots,\theta_k\right)$ that belongs to an exponential family given by

$$f\left(x;\theta_1,\ldots,\theta_k\right)=a\left(\theta_1,\ldots,\theta_k\right)b\left(x\right)\exp\left[\sum_{j=1}^{k}c_j\left(\theta_1,\ldots,\theta_k\right)d_j\left(x\right)\right].$$

Then $\sum_{i=1}^{n}d_1\left(x_i\right),\ldots,\sum_{i=1}^{n}d_k\left(x_i\right)$ is a set of jointly complete sufficient statistics if $\left\{\left(c_1\left(\theta_1,\ldots,\theta_k\right),\ldots,c_k\left(\theta_1,\ldots,\theta_k\right)\right):\theta\in\Omega\right\}$ contains an open set in $\mathbb{R}^k$.

Thm 2.7 : (Basu's theorem) If $T = \ell(X_1, \ldots, X_n)$ is a complete (and minimal) sufficient statistic, then T is independent of every ancillary statistic.

EX 2.23 : $X_1, \ldots, X_n \sim \; iid \; N(\mu, \sigma^2)$. We know that the sample mean $\bar{X}$ is, for every known σ^2, a complete sufficient statistic for μ, $-\infty < \mu < \infty$.

Note that the sample variance $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a location-invariant statistic. To see this, for any real c , let $T = \ell(X_1, \ldots, X_n) = S^2$.

Then

$$
\begin{aligned}
&\ell\left(X_1 + c, \ldots, X_n + c\right) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i + c) - \frac{1}{n}\sum_{i=1}^{n}(X_i + c)\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i + c) - (\bar{X} + c)\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2 \\
&= \ell\left(X_1, \ldots, X_n\right).
\end{aligned}
$$

Thus, S^2 is an ancillary statistic.

EX 2.24 : $X_1,\ldots,X_n \sim$ *iid* Exp(θ), $f\left(x;\theta\right)=\dfrac{1}{\theta}e^{-x/\theta}$.

The pdf is of the form $\dfrac{1}{\theta}f\left(\dfrac{x}{\theta}\right)$, where $f\left(w\right)=e^{-w}$,

$0<w<\theta$; i.e. the exponential distribution form a scale family. Consider the statistic

$$T_1=\frac{X_n}{X_1+\cdots+X_n}=t_1\left(X_1,\ldots,X_n\right).$$

Then

$$t_1\left(cX_1,\ldots,cX_n\right)=\frac{cX_n}{cX_1+\cdots+cX_n}=\frac{X_n}{X_1+\cdots+X_n}$$

$$=t_1\left(X_1,\ldots,X_n\right).$$

So T_1 is a scale invariant statistic and hence an ancillary statistic. Since $T=\sum_{i=1}^{n} X_i$ is a complete statistic. We conclude T and T_1 are independent Thus, we have

$$\theta = E_\theta\left(X_n\right) = E_\theta\left[\sum_{i=1}^{n} X_i \cdot {}^{X_n}\Big/_{\sum_{i=1}^{n} X_i}\right] = E_\theta\left(T\right)\cdot E_\theta\left(T_1\right)$$

$$\Rightarrow E_\theta\left(T_1\right) = \frac{1}{n}.$$