Recap:

1. Ancillary statistic.

A statistic is ancillary if its distribution does not depend on the parameters.

Ex: location family: $X_1, \ldots, X_n \overset{iid}{\sim} f(x-\theta)$

$X_i - X_j$ is an ancillary statistic.

Ex: Scale family of distributions,

$X_1, \ldots, X_n \overset{iid}{\sim} f\left(\frac{x}{\sigma}\right)$, $\sigma > 0$.

$X_1, \ldots, X_n \overset{iid}{\sim} N(0, \sigma^2)$,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} = f\left(\frac{x}{\sigma}\right)$$

$Z_i = \frac{X_i}{\sigma} \sim f(x)$

Thus, $\frac{Z_i}{Z_j} = \frac{X_i}{X_j}$ ~~the~~ the distribution of this quantity should be free of σ.

thus any $\frac{X_i}{X_j}$ is an ancillary statistic for σ.

$X_1, \ldots, X_n \overset{iid}{\sim} Exp(\theta) \Rightarrow \frac{X_i}{\theta} \overset{iid}{\sim} Exp(1)$

$X_1, \ldots, X_n \overset{iid}{\sim} U(\theta, \theta+1)$

$U(\theta, \theta+1)$ is a location family of distribution.

Thus, $X_{(n)} - X_{(1)}$ is an ancillary statistic for the location families.

The minimal sufficient statistic in this example for θ is $(X_{(1)}, X_{(n)})$

Since any one-to-one function of a minimal sufficient statistic is also a minimal sufficient statistic $\Rightarrow \left(\underbrace{X_{(n)} - X_{(1)}}_{\text{ancillary}}, \frac{X_{(n)} + X_{(1)}}{2}\right)$ is minimal sufficient for θ.

In this case an ancillary statistic together with another statistic will give us a minimal sufficient statistic.

Thus ancillary statistic can also have some information about θ.

<u>Example</u>: X_1, X_2 are i.i.d. with a p.m.f.

$$P(X=\theta) = P(X=\theta+1) = P(X=\theta+2) = \frac{1}{3},$$

θ is an unknown integer.

$$f_\theta(x_1, x_2) = \frac{1}{3} 1(\theta < x_1 < \theta+2,\ \theta < x_2 < \theta+2,\ x_1, x_2 \text{ both integers})$$

$(X_{(1)}, X_{(2)})$ is the minimal sufficient statistic

$\left(X_{(2)} - X_{(1)},\ \frac{X_{(2)} + X_{(1)}}{2}\right)$ is the minimal sufficient statistic.

Lets say we have been given the values of (M, m) where m is an integer.

Given that we have been only told the value of m and the fact that m is an integer,

$\theta = m, m-1, m-2$.

$X_1 = \theta$, $X_2 = \theta$; $X_1 = \theta+1$, $X_2 = \theta+1$; $X_1 = \theta+2$, $X_2 = \theta+2$; $X_1 = \theta$, $X_2 = \theta+2$; $X_2 = \theta$, $X_1 = \theta+2$.

we observe, $\frac{X_{(1)} + X_{(2)}}{2} = m$.

$m = \theta$; $m = \theta+1$; $m = \theta+2$; $m = \theta+1$; $m = \theta+1$

thus given only m, we can infer that $\theta = m, m-1, m-2$.

Now if we are additionally provided the information that $M = 2$

$\Rightarrow m = \theta+1 \Rightarrow \theta = m-1$.

thus in extremely toy example information of $(X_{(2)} - X_{(1)})$ helped us making an exact inference on θ.

An ancillary statistic is not independent of a minimal sufficient statistic.

<u>Example:</u> $X_1, \dots, X_n \overset{iid}{\sim} U(\theta, \theta+1)$

minimal suff. $\left(X_{(n)} - X_{(1)}, \; \dfrac{X_{(n)} + X_{(1)}}{2}\right)$

ancillary.

An ancillary statistic has a distribution independent of θ and a minimal sufficient statistic should have all information about θ. Intuitively they should be independent. But they are <u>not independent</u>.

<u>Qn:</u> What extra condition one can impose on a minimal sufficient statistic that would make it independent of any ancillary statistic?

The answer is completeness.

<u>Definition:</u> Let $f_\theta(t)$ be a family of pdfs (or pmfs) for a statistic $T(\underset{\sim}{X})$. The family of distribution is called complete if

$E_\theta[g(T)] = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ.

Equivalently we call $T(\underset{\sim}{X})$ a complete statistic.

Result: If $T(\underline{X})$ is a complete and sufficient statistic, then $T(\underline{X})$ has to be a minimal sufficient statistic.

Pf:- Let $S(\underline{X})$ be a minimal sufficient statistic in this problem. Clearly, $\exists$ a function g_1 s.t. $S(\underline{X}) = g_1(T(\underline{X}))$

Now, let $g_2(S(\underline{X})) = E[T(\underline{X}) | S(\underline{X})]$

this function is free of the parameter θ, since $S(\underline{X})$ is a sufficient statistic so $\underline{X} | S(\underline{X})$ by definition is free of $\theta \Rightarrow$ $T(\underline{X}) | S(\underline{X})$ is free of $\theta \Rightarrow E[T(\underline{X}) | S(\underline{X})]$ is free of θ.

$$g_2(S(X)) = E[T(\underline{X}) | S(\underline{X})]$$

$$\Rightarrow E[(T(\underline{X}) - g_2(S(\underline{X}))) | S(\underline{X})] = 0$$

$$\Rightarrow E[T(\underline{X}) - g_2(S(\underline{X}))] = 0$$

We know, $S(\underline{X}) = g_1(T(\underline{X}))$

$$\Rightarrow E[T(\underline{X}) - g_2 \circ g(T(\underline{X}))] = 0 \quad \forall \theta$$

Since $T(\underline{X})$ is complete $\Rightarrow T(\underline{X}) = g_2 \circ g_1(T(\underline{X}))$

$T(\underline{X}) = g_2 \circ g_1(T(\underline{X}))$ w.p. 1

$= g_2(S(\underline{X}))$ w.p. 1.

$\Rightarrow$ $T(\underline{X})$ is a statistic which is a function of the minimal sufficient statistic $S(\underline{X})$.

$\Rightarrow$ $T(\underline{X})$ has to be minimal sufficient.

$X_1, \ldots, X_n \overset{iid}{\sim} U(\theta, \theta+1)$

we have seen $(X_{(1)}, X_{(n)})$ is minimal sufficient.

$$P(X_{(1)} \le t) = 1 - P(X_{(1)} > t)$$
$$= 1 - P(X_1 > t) \cdots P(X_n > t)$$
$$= 1 - (\theta+1-t)^n$$

$$f_{X_{(1)}}(t) = n(\theta+1-t)^{n-1}, \quad \theta < t < \theta+1$$

$$E[X_{(1)}] = \int_\theta^{\theta+1} n\,t\,(\theta+1-t)^{n-1}\,dt$$

let $z = \theta+1-t$

$$= \int_0^1 n(\theta+1-z)\, z^{n-1}\,dz$$

$$= n(\theta+1)\int_0^1 z^{n-1}\,dz - n\int_0^1 z^{n}\,dz$$

$$= \frac{\cancel{n}(\theta+1)}{\cancel{n}} - \frac{n}{n+1} = \theta+1-\frac{n}{n+1}$$

back ①

$$P(X_{(n)} \leq t) = P(X_1 \leq t, \ldots, X_n \leq t)$$

$$= P(X_1 \leq t) \cdots P(X_n \leq t)$$

$$= (t-\theta)^n$$

$$f_{X_{(n)}}(t) = n(t-\theta)^{n-1}, \quad \theta < t < \theta+1$$

$$E[X_{(n)}] = \int_{\theta}^{\theta+1} n\, t\, (t-\theta)^{n-1}\, dt$$

$t - \theta = z$

$$= \int_0^1 n(z+\theta) z^{n-1}\, dz = n\int_0^1 z^n\, dz + n\theta \int_0^1 z^{n-1}\, dz$$

$$= \frac{n}{n+1} + \theta$$

$$E[X_{(1)}] = \theta + 1 - \frac{n}{n+1}, \quad E[X_{(n)}] = \frac{n}{n+1} + \theta$$

$$\cancel{\Rightarrow E\left[\frac{(n+1)}{n} X_{(1)} - \left\{X_{(n)} - \frac{n}{n+1}\right\}\right] = 0}$$

$$\Rightarrow E\left[\left\{X_{(1)} - \frac{1}{n+1}\right\} - \left\{X_{(n)} - \frac{n}{n+1}\right\}\right] = 0 \quad \forall \theta$$

So, let $g(z_1, z_2) = \left(z_1 - \frac{1}{n+1}\right) - \left(z_2 - \frac{n}{n+1}\right)$

$g \neq 0$

But $E[g(X_{(1)}, X_{(n)})] = 0 \quad \forall \theta$

$\Rightarrow$ $(X_{(1)}, X_{(n)})$ is not a complete statistic.

though $(X_{(1)}, X_{(n)})$ is minimal sufficient in this case.

back ②

Example: Let $X_1, \ldots, X_n \overset{iid}{\sim} Ber(p)$.

$T(\underline{X}) = \sum_{i=1}^{n} X_i$. We have already seen that $T(\underline{X})$ is minimal sufficient. Is it complete?

$$E_p\left(g(T(\underline{X}))\right) = 0 \quad \forall p$$

~~What~~ $\sum_{i=1}^{n} X_i = T(\underline{X}) \sim Bin(n, p)$

$$\Leftrightarrow \sum_{k=0}^{n} g(k)\, p^k (1-p)^{n-k} \binom{n}{k} = 0 \quad \forall p.$$

$$\Rightarrow \sum_{k=0}^{n} \left(\frac{p}{1-p}\right)^k \binom{n}{k} g(k) = 0 \quad \forall p$$

If $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$

and $f(x) = 0 \quad \forall x \Rightarrow a_0 = a_1 = a_2 = \cdots = a_n = 0$

this result in our example will imply that

$$\binom{n}{k} g(k) = 0 \quad \forall k = 0, 1, \ldots, n.$$

$$\Rightarrow g(k) = 0 \quad \forall k = 0, 1, \ldots, n.$$

$$P\left(g(T(\underline{X})) = 0\right) = 1 \quad \forall p.$$

Thus by definition $T(\underline{X})$ is a complete statistic.

back ③

Example: $X_1, \dots, X_n \overset{iid}{\sim} \text{Pois}(\lambda)$

$T(\underline{X}) = \sum_{i=1}^{n} X_i$ is sufficient. Check that this is also minimal sufficient. Is it complete?

$$\sum_{i=1}^{n} X_i = T(\underline{X}) \sim \text{Pois}(n\lambda)$$

$$E_\lambda\left[g(T(\underline{X}))\right] = 0 \quad \forall \lambda$$

$$\Leftrightarrow \sum_{k=0}^{\infty} g(k)\, e^{-n\lambda} \frac{(n\lambda)^k}{k!} = 0 \quad \forall \lambda$$

$$\Leftrightarrow \sum_{k=0}^{\infty} \frac{g(k)\, n^k}{k!} \lambda^k = 0 \quad \forall \lambda$$

$$\Rightarrow g(k) = 0 \quad \forall\ k = 0, 1, \dots$$

$$\Rightarrow P\left(g(T(\underline{X})) = 0\right) = 1 \quad \forall \lambda.$$

Example: $X_1, \dots, X_n \overset{iid}{\sim} U(0, \theta)$.

Remember $X_{(n)} = T(\underline{X})$ is the minimal sufficient statistic. Is it complete?

$$P(T \leq t) = P(X_1 \leq t, \dots, X_n \leq t)$$

$$= P(X_1 \leq t) \cdots P(X_n \leq t) = \left(\frac{t}{\theta}\right)^n$$

$$f_{X_{(n)}}(t) = \frac{n t^{n-1}}{\theta^n}, \quad 0 < t < \theta.$$

$$E_\theta\left[g(T(\underline{X}))\right] = 0 \quad \forall \theta$$

back (4)

$$E_\theta[g(T(\underline{X}))] = E_\theta[g(X_{(n)})] = \int_0^\theta g(t)\,\frac{n t^{n-1}}{\theta^n}\,dt$$

$$E_\theta[g(T(\underline{X}))] = 0 \quad \forall\, \theta$$

$$\Rightarrow \int_0^\theta g(t)\,\frac{n t^{n-1}}{\theta^n}\,dt = 0 \quad \forall\, \theta$$

$$\Rightarrow \int_0^\theta g(t)\, t^{n-1}\,dt = 0 \quad \forall\, \theta$$

$$\Rightarrow \frac{d}{d\theta}\int_0^\theta g(t)\, t^{n-1}\,dt = 0 \quad \forall\, \theta \quad \text{by Newton-Leibnitz rule.}$$

$$g(\theta)\,\theta^{n-1} = 0 \quad \forall\, \theta \quad \Rightarrow \quad g(\theta) = 0 \quad \forall\, \theta .$$

hence $X_{(n)}$ is a complete statistic.

<u>Basu's theorem</u>: If $T(\underline{X})$ is a complete and sufficient, then $T(\underline{X})$ is independent of any ancillary statistic.

back ⑤