Recap:

1. Sufficient statistic
2. Factorization theorem.

Theorem (Factorization theorem)
Let $\underline{X}$ have joint p.d.f. (p.m.f) $f_\theta(\underline{x})$, where θ is the unknown parameter. A statistic $T(\underline{X})$ is called sufficient statistic for θ if and only if

$$f_\theta(\underline{x}) = g(T(\underline{x}),\theta)\, h(\underline{x}).$$

Pf:- only if part

$$P(\underline{X}=\underline{x}) = \sum_t P(\underline{X}=\underline{x} \mid T(\underline{X})=t)\, P(T(\underline{X})=t)$$

only one of these summands is nonzero.

$$= P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x}))\, P(T(\underline{X})=T(\underline{x}))$$

By the definition of sufficient statistic, $\underline{X} \mid T(\underline{X})$ is free of θ $\Rightarrow$ $P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x}))$ is free of θ

$P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x})) = h(\underline{x})$

also I call $P(T(\underline{X})=T(\underline{x})) = g(T(\underline{x}),\theta)$

$P(\underline{X}=\underline{x}) = f_\theta(\underline{x}) = g(T(\underline{x}),\theta)\, h(\underline{x})$.

if part

We are given $f_\theta(\underline{x}) = g(T(\underline{x}),\theta)\, h(\underline{x})$

We need to prove that $T(\underline{X})$ is sufficient stat.

$$P(T(\underline{X}) = t) = P\left(\{x : T(\underline{x}) = t\}\right) = \sum_{x : T(x) = t} f_\theta(\underline{x})$$

$$= \sum_{x : T(\underline{x}) = t} g(T(\underline{x}),\theta)\, h(\underline{x}) = g(t,\theta) \sum_{\underline{x} : T(\underline{x}) = t} h(\underline{x})$$

$$P(\underline{X} = \underline{x} \mid T(\underline{X}) = t) = \frac{P(\underline{X} = \underline{x},\ T(\underline{X}) = t)}{P(T(\underline{X}) = t)}$$

$$= 0 \quad \text{if } t \neq T(\underline{x})$$

$$= \frac{P(\underline{X} = \underline{x})}{P(T(\underline{X}) = t)} \qquad t = T(\underline{x})$$

$$= \frac{g(T(\underline{x}),\theta)\, h(\underline{x})}{g(t,\theta) \sum_{x : T(\underline{x}) = t} h(\underline{x})} = \frac{\cancel{g(t,\theta)}\, h(\underline{x})}{\cancel{g(t,\theta)} \sum_{\underline{x} : T(\underline{x}) = t} h(\underline{x})}$$

thus $\underline{X} \mid T(\underline{X})$ is free of θ.

$\Rightarrow T(\underline{X})$ is sufficient stat.

Some Important Facts:

(1) $T(\underline{X}) = (X_1, \ldots, X_n)$ is trivially a sufficient stat.

(2) If $X_1, \ldots, X_n \overset{iid}{\sim} f_\theta(x)$ then $(X_{(1)}, \ldots, X_{(n)})$ is also a sufficient statistic.

(3) Any one to one function of a sufficient statistic is also a sufficient statistic.

- $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ when σ^2 is known.

we have seen that $\sum_{i=1}^{n} X_i$ is a sufficient stat.

$\Rightarrow \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is also a sufficient stat.

- $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ both μ and σ^2 are unknown.

we have seen that $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is sufficient.

define a function $K(z_1, z_2) = \left(z_1/n, \frac{z_2}{n} - \frac{z_1^2}{n^2}\right)$

$K(z_1, z_2) = (h_1, h_2)$, $h_1 = \frac{z_1}{n}$, $h_2 = \frac{z_2}{n} - \frac{z_1^2}{n^2}$

$\Rightarrow z_1 = n h_1$, $h_2 = \frac{z_2}{n} - \frac{n^2 h_1^2}{n^2} \Rightarrow z_2 = (h_2 + h_1^2) n$

~~$K(z_1, z_2)$~~ $K(\cdot, \cdot)$ is a one-one function.

$K\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right) = \left(\bar{X}, \frac{\sum_{i=1}^{n} X_i^2}{n} - \frac{(n\bar{X})^2}{n^2}\right)$

$= \left(\bar{X}, \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right) = \left(\bar{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)$

thus mean and variance together becomes sufficient for (μ, σ^2).

$X_1, X_2, X_3 \overset{iid}{\sim} Ber(p)$

$$f_p(\underset{\sim}{x}) = p^{\sum_{i=1}^{3} x_i} (1-p)^{3-\sum_{i=1}^{3} x_i}$$

we have seen $T(\underset{\sim}{X}) = \sum_{i=1}^{3} X_i$

now, $\left(\sum_{i=1}^{2} X_i, X_3\right)$ is also a sufficient statistic by factorization theorem.

and (X_1, X_2, X_3) is trivially sufficient.

How far we can go in terms of summarizing without losing any information on the parameter.

We describe a concept known as the minimal sufficiency which answers this.

<u>Definition (Minimal sufficiency)</u>

A statistic $T(\underset{\sim}{X})$ is minimal sufficient if (a) it is sufficient (b) it is a function of every other sufficient statistic.

$$H\left(\sum_{i=1}^{2} X_i, X_3\right) = \sum_{i=1}^{3} X_i \quad H(z_1, z_2) = z_1 + z_2 \text{ then}$$

<u>Question</u>: How to find the minimal sufficient stat.

<u>Theorem</u>: Let $f_\theta(\underline{x})$ be the p.d.f. (or p.m.f.) of $\underline{X}$. Suppose there exists a statistic $T(\underline{X})$ s.t. for any two realizations $\underline{x}, \underline{y}$ of the sample $\underline{X}$, $T(\underline{x}) = T(\underline{y})$ if and only if $\dfrac{f_\theta(\underline{x})}{f_\theta(\underline{y})} = K$, where K is a fixed constant independent of θ. Then $T(\underline{X})$ is minimal sufficient for θ.

<u>Example</u>: $X_1, X_2, X_3 \overset{iid}{\sim} Ber(p)$

$$f_p(\underline{x}) = p^{\sum_{i=1}^3 x_i}(1-p)^{3-\sum_{i=1}^3 x_i}, \quad f_p(\underline{y}) = p^{\sum_{i=1}^3 y_i}(1-p)^{3-\sum_{i=1}^3 y_i}$$

$$\frac{f_p(\underline{x})}{f_p(\underline{y})} = \frac{p^{\sum_{i=1}^3 x_i}(1-p)^{3-\sum_{i=1}^3 x_i}}{p^{\sum_{i=1}^3 y_i}(1-p)^{3-\sum_{i=1}^3 y_i}} = \left(\frac{p}{1-p}\right)^{\sum_{i=1}^3 x_i - \sum_{i=1}^3 y_i}$$

this ratio is constant as a function of parameter p, if and only if $\sum_{i=1}^3 x_i = \sum_{i=1}^3 y_i$. thus by applying the theorem, $T(\underline{X}) = X_1 + X_2 + X_3$

$$T_1(\underline{X}) = (X_1 + X_2, X_3)$$

$$\frac{f_p(\underline{x})}{f_p(\underline{y})} = \left(\frac{p}{1-p}\right)^{\sum_{i=1}^3 x_i - \sum_{i=1}^3 y_i}$$ ~~if $T_1(\underline{x}) = T_1(\underline{y})$~~

if $T_1(\underline{x}) = T_1(\underline{y})$ then the ratio is constant.

But the only if part is not as we can have (x_1, x_2, x_3) and (y_1, y_2, y_3) s.t.

$x_1 + x_2 \neq y_1 + y_2$, $x_3 \neq y_3$ but $x_1 + x_2 + x_3 = y_1 + y_2 + y_3$

then $\dfrac{f_p(\underline{x})}{f_p(\underline{y})}$ is constant but $T_1(\underline{x}) \neq T_1(\underline{y})$

Example: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, μ, σ^2 are unknown.

$$\frac{f_{\mu,\sigma^2}(\underline{x})}{f_{\mu,\sigma^2}(\underline{y})} = \frac{\cancel{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (x_i-\mu)^2\right]\right\}}{\cancel{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (y_i-\mu)^2\right]\right\}}$$

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

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

this ratio is a fixed constant if and only if $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ and $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$

thus $T(\underline{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ is minimal sufficient.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} U(\theta, \theta+1)$, $-\infty < \theta < \infty$.

$$\frac{f_\theta(\underline{x})}{f_\theta(\underline{y})} = \frac{I(x_{(1)} > \theta,\ x_{(n)} - 1 < \theta)}{I(y_{(1)} > \theta,\ y_{(n)} - 1 < \theta)}$$

in different ranges of θ this ratio takes the value $0, 1$ or ∞.

the ratio can be made fixed if and only if $x_{(1)} = y_{(1)}$ and $x_{(n)} = y_{(n)}$.

Thus $T(\underline{X}) = (X_{(1)}, X_{(n)})$ is minimal sufficient.

if $x_{(1)} = y_{(1)}$ but $x_{(n)} \neq y_{(n)}$

$x_{(n)} < y_{(n)}$ or $x_{(n)} > y_{(n)}$

when $x_{(n)} < y_{(n)}$

$$\frac{f_\theta(\underline{x})}{f_\theta(\underline{y})} = \frac{I(x_{(1)} > \theta,\ x_{(n)} - 1 < \theta)}{I(y_{(1)} > \theta,\ y_{(n)} - 1 < \theta)}$$

$\theta > x_{(n)} - 1$ but $\theta < y_{(n)} - 1$ then $\dfrac{f_\theta(\underline{x})}{f_\theta(\underline{y})} = \infty$

when $\theta > x_{(n)} - 1$ and $\theta > y_{(n)} - 1$ then $\dfrac{f_\theta(\underline{x})}{f_\theta(\underline{y})} = 1$

thus only using $X_{(1)}$ is not minimal sufficient.

Property: Any one-one function of a minimal sufficient statistic is also minimal sufficient.

Thus minimal sufficient statistic is not unique.

Recap: ① Minimal sufficient
② Examples of how to ~~compute min~~ find a minimal sufficient statistic.

Ancillary Statistic

A statistic whose distribution does not depend on the unknown parameter θ is known as an ancillary statistic.
It seems that ancillary statistic has a distribution free of θ. Then why are we interested? We will discuss it later.

Finding ancillary statistics

Let $X_1, \ldots, X_n \overset{iid}{\sim} f(x-\theta)$, $-\infty < \theta < \infty$.

$$X \sim N(\mu, 1) \Rightarrow f(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2}\right\} = g(x-\mu)$$

If $X_i \sim f(x-\theta)$, $\quad Z_i = X_i - \theta$

If $X_i \sim N(\mu+5, 1) \sim Z_i \sim N(5,1)$

$X_i \sim f(x-\theta) \Rightarrow Z_i = X_i - \theta \sim f(x)$

$$P(Z_i \le z) = P(X_i - \theta \le z) = P(X_i \le \theta + z) = \int_{-\infty}^{\theta+z} f(x-\theta)\,dx$$

Let $h = x - \theta$

$$= \int_{-\infty}^{z} f(h)\,dh$$

$$\Rightarrow f_{Z_i}(z) = f(z)$$

$R = X_{(n)} - X_{(1)} \longrightarrow$ Range statistic

$$P(R \le r) = P\left(\max_{1\le i\le n} X_i - \min_{1\le i\le n} X_i \le r\right)$$

$$= P\left(\max_{1\le i\le n}(z_i+\theta) - \min_{1\le i\le n}(z_i+\theta) \le r\right)$$

$$= P\left(\max_{1\le i\le n} z_i + \cancel{\theta} - \cancel{\theta} - \min_{1\le i\le n} z_i \le r\right)$$

$$= P\left(Z_{(n)} - Z_{(1)} \le r\right)$$

Since $z_1, \dots, z_n \overset{iid}{\sim} f(x)$ free of θ

Thus $Z_{(n)}$ and $Z_{(1)}$ have distributions free of θ

Thus $P\left(Z_{(n)} - Z_{(1)} \le r\right)$ is free of θ

$\Rightarrow$ The distribution of $X_{(n)} - X_{(1)}$ is free of θ.

$\Rightarrow X_{(n)} - X_{(1)}$ is an ancillary statistic.

Other ancillary statistics

$z_i = x_i - \theta$ thus $x_i - x_j = z_i + \cancel{\theta} - z_j - \cancel{\theta} = z_i - z_j$

$\Rightarrow x_i - x_j$ has a distribution free of θ

$\Rightarrow x_i - x_j$ for any $i \ne j$ is ancillary.

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

$\Rightarrow Z_i = X_i - \theta \sim U(0,1)$ hence $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Scale family of distributions

Let $X_1, \dots, X_n \overset{iid}{\sim} f\left(\frac{x}{\sigma}\right)\frac{1}{\sigma}$, $\sigma > 0$.

this is called the scale family of distributions

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

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

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

this is a scale family of distributions.

When $X_1, \dots, X_n \overset{iid}{\sim} \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$, $\sigma > 0$

$Z_i = \frac{X_i}{\sigma} \sim f(x)$ Thus $Z_1, \dots, Z_n \overset{iid}{\sim} f(x)$

Take any $\frac{X_i}{X_j} = \frac{X_i/\sigma}{X_j/\sigma} = \frac{Z_i}{Z_j}$

thus $\frac{Z_i}{Z_j}$ and hence $\frac{X_i}{X_j}$ has a density free of σ.

$\Rightarrow \frac{X_i}{X_j}$ is an ancillary statistic.

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

The minimal sufficient statistic is $(X_{(1)}, X_{(n)})$.

$\Rightarrow \left(X_{(n)} - X_{(1)}, \frac{X_{(n)} + X_{(1)}}{2}\right)$ being a one to one function of $(X_{(1)}, X_{(n)})$ is also a minimal sufficient statistic.

this is a location family and hence $X_{(n)} - X_{(1)}$ is an ancillary statistic.

Thus ancillary statistic together with some other statistic contains all the information about θ.

<u>Example:</u> ~~X_1, X_2~~, X_1, X_2 are i.i.d drawn from the following discrete distribution

$$P(X=\theta) = P(X=\theta+1) = P(X=\theta+2) = \frac{1}{3} \quad \forall \theta.$$

$Z = X - \theta$, $P(Z=0) = P(Z=1) = P(Z=2) = \frac{1}{3}$.

thus this is also a location family and

Hence $X_{(2)} - X_{(1)}$ is an ancillary statistic.

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

$\Rightarrow \left(X_{(2)} - X_{(1)}, \dfrac{X_{(2)}+X_{(1)}}{2}\right)$ is minimal sufficient.

If I observe from the data the value for $\dfrac{X_{(1)}+X_{(2)}}{2} = m$, then what are the possible values of θ?

<u>there are 9 possible cases</u>

$X_1 = \theta+2, \quad X_2 = \theta+2 \Rightarrow \dfrac{X_{(1)}+X_{(2)}}{2} = \theta+2 = m \Rightarrow \theta = m-2.$

if $X_1 = \theta, \quad X_2 = \theta+2 \Rightarrow \dfrac{X_{(1)}+X_{(2)}}{2} = \theta+1 = m \Rightarrow \theta = m-1$

if $X_1 = \theta, \quad X_2 = \theta \Rightarrow \dfrac{X_{(1)}+X_{(2)}}{2} = \theta = m \Rightarrow \theta = m$

If I am given an additional information $X_{(2)} - X_{(1)} = 2$ then one of them has to be θ and the other has to be $\theta+2$

$$\Rightarrow \frac{X_{(1)}+X_{(2)}}{2} = \theta+1 = m \Rightarrow \theta = m-1$$

Clearly minimal sufficient statistic is not independent of ancillary statistic although one of them contains all information about θ and the other has a distribution free of θ. It appears that we need to put extra restriction on minimal sufficient stat.

Def: (complete statistic)

Let $f_\theta(t)$ be a family of p.d.fs (or p.m.fs) for a statistic $T(\underline{X})$. The family of distributions is called complete if $E_\theta[g(T(\underline{X}))] = 0 \quad \forall \theta$ implies $P_\theta(g(T(\underline{X})) = 0) = 1 \quad \forall \theta$.

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

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

Let g be a function s.t. $E_p[g(T(\underline{X}))] = 0 \quad \forall p$

$$E_p[g(T(\underline{X}))] = \sum_{t=0}^{n} g(t) \binom{n}{t} p^t (1-p)^{n-t}$$

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

$$\Rightarrow \sum_{t=0}^{n} g(t) \binom{n}{t} p^t (1-p)^{n-t} = 0 \quad \forall p$$

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

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n = 0 \quad \forall x$$

$$\Rightarrow a_0 = a_1 = \cdots = a_n = 0$$

$\sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t$ is a polynomial w.r.t. $\frac{p}{1-p}$ and this polynomial is $=0$ for all values of $\frac{p}{1-p}$.

$\Rightarrow g(0)\binom{n}{0} = g(1)\binom{n}{1} = \cdots = g(n)\binom{n}{n} = 0$

$\Rightarrow g(0) = \cdots = g(n) = 0$

$P_p(g(T(\underline{x})) = 0) = 1 \quad \forall p$

<u>example:</u> $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$.

$T(\underline{X}) = \max_{1 \le i \le n} X_i \qquad f_\theta(t)$ is the density of $T(\underline{X})$

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

If there is a g s.t.

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

$\Rightarrow E_\theta[g(T(\underline{X}))] = \int_0^\theta g(t)\, n \frac{t^{n-1}}{\theta^n} = 0 \quad \forall \theta > 0$

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

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

$\Rightarrow g(\theta) = 0 \quad \forall \theta > 0$

$\Rightarrow P_\theta[g(T(\underline{X})) = 0] = 1$

$\Rightarrow \max_{1 \le i \le n} X_i$ is a complete statistic.

Recap: ① Complete statistic.

Def: Let $f_\theta(t)$ be a family of pdfs (or pmfs) for a statistic $T(\underline{x})$. The family of dists. is called complete if $E_\theta[g(T(\underline{x}))] = 0 \ \forall \theta$ implies $P_\theta(g(T(\underline{x})) = 0) = 1 \ \forall \theta$.
$T(\underline{x})$ is called a complete statistic.

$X_1, X_2, X_3 \overset{iid}{\sim} Ber(p) \qquad T(\underline{x}) = X_1 - X_2$.

$E[X_1 - X_2] = 0 \ \forall p \qquad g(t) = t$

$\Rightarrow E[g(T(\underline{x}))] = 0 \ \forall p$ but $g \neq 0$.

$X_1 - X_2$ is not a complete statistic.

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

Pf:- Let $T(\underline{x})$ be a complete sufficient statistic and $S(\underline{x})$ be a minimal sufficient statistic.
Then $S(\underline{x})$ is a function of $T(\underline{x})$.

Now $E[T(\underline{x}) \mid S(\underline{x})] = g(S(\underline{x}))$

$\Rightarrow E[(T(\underline{x}) - g(S(\underline{x}))) \mid S(\underline{x})] = 0$

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

If $S(\underline{x}) = g_1(T(\underline{x})) \quad \Rightarrow E[T(\underline{x}) - g(g_1(T(\underline{x})))] = 0$

By the definition of completeness of $T(\underline{x})$,

$T(\underline{x}) = g(g_1(T(\underline{x}))) = g(S(\underline{x}))$

$\Rightarrow T(\underline{X})$ is also minimal sufficient.

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

Pf:- (Only in the discrete case)

Let $S(\underline{X})$ be any ancillary statistic. Then $P_\theta(S(\underline{X})=s)$ does not depend on θ. Since $T(\underline{X})$ is a sufficient statistic

$P_\theta(S(\underline{X})=s \mid T(\underline{X})=t)$ is ~~independent~~ free of θ.

Now,

$$P_\theta(S(\underline{X})=s) = \sum_t P_\theta(S(\underline{X})=s \mid T(\underline{X})=t)\, P_\theta(T(\underline{X})=t) \quad \text{---} (*)$$

Furthermore,

$$P_\theta(S(\underline{X})=s) = P_\theta(S(\underline{X})=s) \sum_t P_\theta(T(\underline{X})=t) \quad \text{---} (**)$$

$\cancel{\sum_t P_\theta(T(\underline{X})=t)=1}$

From $(*)$ and $(**)$

$$\sum_t P_\theta(S(\underline{X})=s \mid T(\underline{X})=t)\, P_\theta(T(\underline{X})=t) = \sum_t P_\theta(S(\underline{X})=s)\, P_\theta(T(\underline{X})=t)$$

$$\Rightarrow \sum_t \left\{ P_\theta(S(\underline{X})=s \mid T(\underline{X})=t) - P_\theta(S(\underline{X})=s) \right\} P_\theta(T(\underline{X})=t) = 0 \quad \forall \theta$$

$$\cancel{g(T(\underline{x})) = P_\theta(S(\underline{x}) \mid T(\underline{x})) - P_\theta}$$

$$g(t) = P_\theta(S(\underline{x}) = s \mid T(\underline{x}) = t) - P_\theta(S(\underline{x}) = s)$$

$$\Rightarrow E[g(T(\underline{x}))] = 0 \quad \forall \theta$$

Since $T(\underline{x})$ is a complete statistic,

$$P_\theta(g(T(\underline{x})) = 0) = 1 \quad \forall \theta$$

$$\Rightarrow P_\theta(S(\underline{x}) = s \mid T(\underline{x}) = t) = P_\theta(S(\underline{x}) = s) \quad \forall \theta.$$

$\Rightarrow$ $S(\underline{x})$ and $T(\underline{x})$ are independent.

<u>Example</u>: Let $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{E}xp(\theta)$.

Find $E_\theta\left[\dfrac{X_n}{\sum_{i=1}^n X_i}\right]$?

$$f_\theta(x) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right)$$

this is a scale family of distributions.

$$\frac{X_n}{\sum_{i=1}^n X_i} = \frac{1}{\sum_{i=1}^n \frac{X_i}{X_n}}$$

. Now, we have seen that for scale families $\frac{X_i}{X_n}$ $\forall i$ are ancillary statistic.

Thus $\dfrac{X_n}{\sum_{i=1}^n X_i}$ is also an ancillary statistic.

$f_\theta(x_1, \ldots, x_n) = \frac{1}{\theta^n} \exp\left(-\frac{\sum_{i=1}^n x_i}{\theta}\right)$ is the joint density.

thus by factorization then $\sum_{i=1}^n x_i$ is a sufficient statistic.

$\sum_{i=1}^n x_i$ is also a complete statistic (as we will show with a general result)

thus $\sum_{i=1}^n x_i$ is a complete sufficient statistic.

thus $g(\underset{\sim}{x}) = \frac{x_n}{\sum_{i=1}^n x_i}$ and $T(\underset{\sim}{x}) = \sum_{i=1}^n x_i$ are independent by Basu's theorem

$$E[g(\underset{\sim}{x})\,T(\underset{\sim}{x})] = E[g(\underset{\sim}{x})]\,E[T(\underset{\sim}{x})]$$

$$\Rightarrow E[x_n] = E\left[\frac{x_n}{\sum_{i=1}^n x_i}\right] E\left[\sum_{i=1}^n x_i\right]$$

$$\Rightarrow \theta = E\left[\frac{x_n}{\sum_{i=1}^n x_i}\right] n\theta$$

$$\Rightarrow E\left[\frac{x_n}{\sum_{i=1}^n x_i}\right] = \frac{1}{n}$$

Exponential family of distributions

A one parameter exponential family density is given by $f_\theta(x) = h(x)\, c(\theta) \exp(\omega(\theta) t(x))$, where $h(\cdot)$, $c(\cdot)$, $\omega(\cdot)$, $t(\cdot)$ are some functions.

Example: $X \sim Ber(p)$

$$f_p(x) = p^x (1-p)^{1-x} = \left(\frac{p}{1-p}\right)^x (1-p)$$

$$= \exp\left\{ x \log \frac{p}{1-p} \right\} (1-p)$$

So Bernoulli density belongs to the exponential family with $t(x) = x$, $\omega(\theta) = \log \frac{p}{1-p}$, $c(\theta) = (1-p)$ and $h(x) = 1$

Example: $X \sim Pois(\lambda)$

$$f_\lambda(x) = \exp(-\lambda) \frac{\lambda^x}{x!} = \exp(-\lambda) \exp(x \log \lambda) \frac{1}{x!}$$

$h(x) = \frac{1}{x!}$, $c(\theta) = \exp(-\lambda)$, $\omega(\theta) = \log \lambda$ and $t(x) = x$.

You can express Exponential, Normal, Gamma, Inverse gamma all as exponential family of densities.

$$\int_x f_\theta(x)\,dx = 1$$

$$\Rightarrow \int_x h(x)\,c(\theta)\exp\left(w(\theta)\,t(x)\right)dx = 1$$

$$\Rightarrow \frac{d}{d\theta}\int_x h(x)\,c(\theta)\exp\left(w(\theta)\,t(x)\right)dx = 0$$

$$\Rightarrow \int_x \frac{d}{d\theta}\left[h(x)\,c(\theta)\exp\left(w(\theta)\,t(x)\right)\right]dx = 0$$

$$\Rightarrow \int_x \left[h(x)\,c'(\theta)\exp\left(w(\theta)\,t(x)\right) + h(x)\,c(\theta)\,w'(\theta)\,t(x)\exp\left(w(\theta)\,t(x)\right)\right]dx = 0$$

$$\Rightarrow \int h(x)\,c'(\theta)\exp\left(w(\theta)\,t(x)\right)dx = -\int h(x)\,c(\theta)\,w'(\theta)\,t(x)\exp\left(w(\theta)\,t(x)\right)dx$$

$$\Rightarrow \frac{c'(\theta)}{c(\theta)}\int h(x)\,c(\theta)\exp\left(w(\theta)\,t(x)\right)dx = -w'(\theta)\int h(x)\,c(\theta)\,t(x)\exp\left(w(\theta)\,t(x)\right)dx$$

$$\Rightarrow \frac{c'(\theta)}{c(\theta)}\int f_\theta(x)\,dx = -w'(\theta)\int t(x)\,f_\theta(x)\,dx$$

$$\Rightarrow \frac{c'(\theta)}{c(\theta)} = -w'(\theta)\,E\left[t(X)\right]$$

$$\Rightarrow E\left[t(X)\right] = \frac{-c'(\theta)}{c(\theta)\,w'(\theta)}$$

By taking second derivative you can find closed form expressions for $E[t(X)^2]$ and $Var(t(X))$.

Similar to the one parameter exponential family, a multi-parameter exponential family has density

$$f_{\underline{\theta}}(x) = h(x)\, c(\underline{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\underline{\theta})\, t_i(x)\right)$$

check: $N(\mu, \sigma^2)$ with μ, σ^2 both unknown parameters can be written in the above form.

If $X_1, \ldots, X_n \overset{iid}{\sim} f_{\underline{\theta}}(x)$ where $f_{\underline{\theta}}(x)$ is from a multiparameter exponential family

$$f_{\underline{\theta}}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{\underline{\theta}}(x_i)$$

$$= \left[\prod_{i=1}^{n} h(x_i)\right]\left[c(\underline{\theta})\right]^n \exp\left\{\sum_{j=1}^{k} w_j(\theta) \sum_{i=1}^{n} t_j(x_i)\right\}$$

By factorization theorem

$$\left(\sum_{i=1}^{n} t_1(X_i),\ \sum_{i=1}^{n} t_2(X_i),\ \ldots,\ \sum_{i=1}^{n} t_k(X_i)\right) \text{ is}$$

a sufficient statistic for $\underline{\theta}$.