We have

$$P((X_1^\circ, X_2^\circ) = (0,0)) = P((X_1^\circ, X_2^\circ) = (0,0) \cap T(X_1, X_2) = 0)$$

$$= P((X_1^\circ, X_2^\circ) = (0,0) \mid T(X_1, X_2) = 0) \cdot P(T(X_1, X_2) = 0) = (1-\theta)^2 =$$

$$P((X_1, X_2) = (0,0)).$$

Similarly for $(X_1^\circ, X_2^\circ) = (1,0), (0,1)$ and $(1,1)$

Hence (X_1, X_2) and (X_1°, X_2°) have the same distribution, but are normally not identical, but inference about θ can be equally well made from (X_1°, X_2°)

Definition 6.2.1

$T(\underline{X})$ is a sufficient for θ if the conditional distribution of $\underline{X}$ given $T(\underline{X})$ does not depend on θ.

A sufficient statistics for a parameter θ is a statistic that in a certain sense capture all the information about θ in a sample.

Discrete case

In general experimenter 2 who knows $T(\underline{X}) = t$ can generate a distribution $P(\underline{X} = \underline{y} \mid T(\underline{X}) = t)$ on the set $A_{T(\underline{X})} = \{\underline{y} : T(\underline{y}) = t\}$

A random variable $\underline{Y}$ with this distribution satisfies

$$P(\underline{Y} = \underline{y} \mid T(\underline{X}) = t) = P(\underline{X} = \underline{y} \mid T(\underline{X}) = t)$$

and $P(\underline{X}=\underline{x}) = P(\underline{X}=\underline{x} \cap T(\underline{X})=t)$

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

$= P(\underline{Y}=\underline{x} \mid T(\underline{X})=t) \cdot P(T(\underline{X})=t)$

$= P(\underline{Y}=\underline{x} \cap T(\underline{X})=t) = P(\underline{Y}=\underline{x})$

We have $P(\underline{X}=\underline{x} \mid T(\underline{X})=T(\underline{x})=t) = \dfrac{P(\underline{X}=\underline{x} \cap T(\underline{X})=T(\underline{x}))}{P(T(\underline{X})=T(\underline{x}))}$

$$= \frac{P(\underline{X}=\underline{x})}{P(T(\underline{X})=T(\underline{x}))} = \frac{p(\underline{x}|\theta)}{q(T(\underline{x})|\theta)}$$

This gives Theorem 6.2.2

If $p(\underline{x}|\theta)$ is the joint pdf/pmf of $\underline{X}$ and $q(t|\theta)$ is the pdf/pmf of $T(\underline{X})$, then $T(\underline{X})$ is a sufficient statistic for θ if, for every $\underline{x}$ in the sample space $\dfrac{p(\underline{x}|\theta)}{q(T(\underline{x})|\theta)}$ is constant as a function of θ.

Example. $X_1, X_2, \ldots, X_n$ iid Bernoulli θ, $T = \sum_{i=1}^{n} X_i$

$$\frac{p(\underline{x}|\theta)}{q(T(\underline{x})|\theta)} = \frac{\theta^{\overset{t}{\sum x_i}}(1-\theta)^{n-\overset{t}{\sum x_i}}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}$$ ~~independent of θ~~ is constant as a function of θ.

Example 6.2.4

$X_1, \dots, X_m$ i.i.d. $N(\mu, \sigma^2)$, σ^2 known. $T(\underline{X}) = \bar{X}$

$$f(\underline{x} \mid \mu) = \prod_{i=1}^{m} (2\pi\sigma^2)^{-\frac{1}{2}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{m}{2}} \cdot e^{-\sum_{i=1}^{m} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{-\frac{m}{2}} \cdot e^{-\sum_{i=1}^{m} \frac{(x_i - \bar{x} + \bar{x} - \mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{m}{2}} e^{-\sum_{i=1}^{m} \frac{(x_i-\bar{x})^2}{2\sigma^2} - \frac{m(\bar{x}-\mu)^2}{2\sigma^2}}$$

$$g(T(\underline{x}) \mid \mu) = g(\bar{x} \mid \mu) = \left(2\pi \frac{\sigma^2}{m}\right)^{-\frac{1}{2}} \cdot e^{-\frac{(\bar{x}-\mu)^2 \cdot m}{2\sigma^2}}$$

$$\Rightarrow \frac{f(\underline{x} \mid \mu)}{g(\bar{x} \mid \mu)} = \frac{(2\pi\sigma^2)^{-\frac{m}{2}} \cdot e^{-\sum_{i=1}^{m} \frac{(x_i-\bar{x})^2}{2\sigma^2}}}{\left(\frac{2\pi\sigma^2}{m}\right)^{-\frac{1}{2}}}$$ constant as a function of μ.

Theorem 6.2.6. Factorization theorem

Let $f(\underline{x} \mid \theta)$ be the joint pmf/pdf of a sample $\underline{X}$

$T(\underline{X})$ is a sufficient statistic for $\theta \Leftrightarrow$ There exist a $g(t \mid \theta)$ such that $\forall \underline{x}$ and all values of θ

$$f(\underline{x} \mid \theta) = g(T(\underline{x}) \mid \theta)\, h(\underline{x})$$

Proof, $\underline{X}$ discrete.

$\Rightarrow$ $T(\underline{X})$ sufficient. Let $g(t \mid \theta) = P(T(\underline{X}) = t)$ and $h(\underline{x}) = P(\underline{X} = \underline{x} \mid T(\underline{X}) = T(\underline{x}) = t)$. Then $f(\underline{x} \mid \theta) = P(\underline{X} = \underline{x}) = P(\underline{X} = \underline{x} \cap T(\underline{X}) = T(\underline{x}))$

$= P(\underline{X} = x \mid T(\underline{X}) = T(\underline{x})) \cdot P(T(\underline{X}) = T(\underline{x})) = h(\underline{x})\, g(T(\underline{x}) \mid \theta)$, $\forall x$ and $\forall \theta$

$\Leftarrow$ $f(\underline{x} \mid \theta) = g(T(\underline{x}) \mid \theta)\, h(\underline{x})$ and define $A_{T(\underline{x})} = \{\underline{y} : T(\underline{y}) = T(\underline{x}) = t\} = A_t$

Then $\dfrac{f(\underline{x}|\theta)}{q(T(\underline{x})|\theta)} = \dfrac{g(T(\underline{x})|\theta)\cdot h(\underline{x})}{q(T(\underline{x})|\theta)} = \dfrac{g(T(\underline{x})|\theta)\cdot h(\underline{x})}{\sum_{A_{T(\underline{x})}} g(T(\underline{y})|\theta)\cdot h(\underline{y})}$

$= \dfrac{g(T(\underline{x})|\theta)\cdot h(\underline{x})}{g(T(\underline{x})|\theta)\sum_{A_{T(\underline{x})}} h(\underline{y})} = \dfrac{h(\underline{x})}{\sum_{A_{T(\underline{x})}} h(\underline{y})}$ is a constant as a function of θ.

Examples. $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$, σ^2 known. $T(\underline{X}) = \bar{X}$

$$f(\underline{x}/\mu) = \underbrace{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}}_{h(\underline{x})} \cdot \underbrace{e^{-\frac{n(\bar{x}-\mu)}{2\sigma^2}}}_{g(T(\underline{x})|\mu)}$$

Example.

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & x \in (0,\theta) \\ 0, & \text{elsewhere} \end{cases} = \frac{1}{\theta} I_{(0,\theta)}(x)$$

Sample $X_1, \dots, X_n$ with observations $x_1, \dots, x_n$

$$f(\underline{x}|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{(0,\theta)}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(0,\theta)}(x_i)$$

Let $T = \max_{i=1,2,\dots,n} x_i$

Then $$f(\underline{x}|\theta) = \underbrace{\frac{1}{\theta^n} I_{(0,\theta)}(t)}_{g(T(\underline{x})|\theta)} \underbrace{\prod_{i=1}^{n} I_{(0,\infty)}(x_i)}_{h(\underline{x})}$$

$\Rightarrow T = \max_i x_i$ is sufficient for θ

$T'(\underline{X}) = \underline{X}$ is sufficient

Let $T(\underline{X})$ be sufficient and define $T^{\circ}(\underline{X}) = r(T(\underline{X}))$, $\forall X$

Then $f(\underline{x}|\theta) = g(T(\underline{x})|\theta) = g(r^{-1}(T^{\circ}(\underline{x}))|\theta)\, h(\underline{x})$

$\Rightarrow T^{\circ}(\underline{x})$ is also sufficient.

<u>Theorem 6.2.10</u>

$X_1, \dots, X_n$ iid from a pmf/pdf $f(x|\underline{\theta})$

where $f(x|\underline{\theta}) = h(x)\, c(\underline{\theta})\, e^{\sum_{i=1}^{k} w_i(\underline{\theta}) t_i(x)}$

and $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta d)$, $d \le k$. Then

$T(\underline{x}) = \left(\sum_{j=1}^{n} t_1(X_j), \dots, \sum_{j=1}^{n} t_k(X_j)\right)$ is a sufficient statistic for $\underline{\theta}$

<u>Proof.</u>

$$f(\underline{x}|\underline{\theta}) = \prod_{j=1}^{n} h(x_j)\, c(\underline{\theta})\, e^{\sum_{i=1}^{k} w_i(\underline{\theta}) t_i(x_j)}$$

$$= \underbrace{\prod_{j=1}^{n} h(x_j)}_{h(x)} \underbrace{(c(\underline{\theta}))^{n}\, e^{\sum_{i=1}^{k} w_i(\theta) \sum_{j=1}^{n} t_i(x_j)}}_{g(T_1(\underline{x}), \dots, T_k(x)|\theta)}$$

<u>Example</u>

$$f(x|\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma}\, e^{-\frac{\mu^2}{2\sigma^2} - \frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}}$$

$t_1(x) = x$, $w_1(\mu,\sigma^2) = \frac{\mu}{\sigma^2}$, $t_2(x) = x^2$, $w_2(\mu,\sigma^2) = -\frac{1}{2\sigma^2}$

$\Rightarrow$ for $X_1, \dots, X_n$ that.

$\sum_{j=1}^{n} X_j$ and $\sum_{j=1}^{n} X_j^2$ are sufficient for (μ,σ^2).