(4) $X_1, \ldots X_n$ random sample from $N(\mu, \sigma^2)$

$\mu \in \mathbb{R}, \ \sigma > 0$

$$\underset{\sim}{\theta} = (\mu, \sigma) \in \Theta = \{(\mu,\sigma) : \mu \in \mathbb{R}, \sigma > 0\}$$

$$f_{\underset{\sim}{\theta}}(\underline{x}) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2\right)$$

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

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

$$= h(\underset{\sim}{x}) \ g_{\underset{\sim}{\theta}}\left(\sum x_i, \sum x_i^2\right)$$

$T(\underset{\sim}{X}) = \left(\sum_1^n X_i, \sum_1^n X_i^2\right)$ is jointly sufficients for $\underset{\sim}{\theta}$

(5) $X_1, \ldots X_n$ is a random sample from $N(\theta, \theta^2)$ $\theta > 0$

jt p.d.f.

$$f_\theta(\underline{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(-\frac{1}{2\theta^2}[x_i - \theta)^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\theta^n} \exp\left(-\frac{1}{2\theta^2}\left(\Sigma x_i^2 + n\theta^2 - 2\theta \Sigma x_i\right)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\theta^n} e^{-\frac{\Sigma x_i^2}{2\theta^2}} e^{-\frac{n}{2}} e^{\frac{\Sigma x_i}{\theta}}$$

$$= \underbrace{\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-n/2}\right)}_{h(\underline{x})} \underbrace{\left(\frac{1}{\theta^n} e^{-\frac{\Sigma x_i^2}{2\theta^2}} e^{\frac{\Sigma x_i}{\theta}}\right)}_{g_\theta(\Sigma x_i,\, \Sigma x_i^2)}$$

By NFFT, $T(\underline{X}) = \left(\sum_{i=1}^{n} X_i, \Sigma X_i^2\right)$ is jointly sufficient for θ

(6) $X_1, \ldots, X_n$ is a random sample from $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ $\theta \in \mathbb{R}$

jt p.d.f.

$$f_\theta(\underline{x}) = \begin{cases} 1, & \theta - \frac{1}{2} < x_1, \ldots, x_n < \theta + \frac{1}{2} \\ 0, & o/w \end{cases}$$

$$= \begin{cases} 1, & \theta - \frac{1}{2} < x_{(1)} - \cdots < x_{(n)} < \theta + \frac{1}{2} \\ 0, & o/w \end{cases}$$

i.e. $f_\theta(\underset{\sim}{x}) = I(\theta - \frac{1}{2}, x_{(1)})\, I(x_{(n)}, \theta + \frac{1}{2})$

By NFFT, $T(\underline{X}) = (X_{(1)}, X_{(n)})$ is jointly sufficient for θ

$h(\underline{x}) = 1$

$g_\theta(x_{(1)}, x_{(n)}) = I(\theta - \frac{1}{2}, x_{(1)})\, I(x_{(n)}, \theta + \frac{1}{2})$

Remark: Any 1-1 f^n of sufficient statistic is sufficient

Remark: Any f^n of sufficient statistic is NOT necessarily a sufficient statistic

Remark: The factorization (as in NFFT) is not unique.

Remark: sufficient statistic is not unique. We look for the sufficient statistic that provides the maximum possible reduction of the data, this leads us to the concept of "minimal sufficient statistic

Ex: $X_1, \dots X_n$ r.s. from $N(\theta, 1)$ $\theta \in \mathbb{R}$

By NFFT, all the following statistics are sufficient for θ.

$T_1(\underline{X}) = (X_1, \dots, X_n)$

$T_2(\underline{X}) = (X_1 + X_2, X_3, X_4, \dots, X_n)$

$T_3(\underline{X}) = (X_1 + X_2, X_3 + X_4, X_5, \dots, X_n)$

$T_4(\underline{X}) = (X_1 + X_2 + X_3 + X_4, X_5, \dots, X_n)$

$$T_5(\underline{X}) = \left(X_1 + X_2, \sum_{i=3}^{n} X_i\right)$$

$$\vdots$$

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

Intuitively, among all the above sufficient statistics, $T(\underline{X}) = \sum_{i=1}^{n} X_i$ appears to be "best", giving maximum possible reduction.

Def^n: **Minimal sufficient statistic**

Let $X_1, \ldots, X_n$ be a random sample from a dist^n P_θ, $\theta \in \Theta$, having p.d.f. or p.m.f. $f_\theta(x)$. A statistic $T(\underline{X})$ is said to be minimal sufficient for θ if

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

& (ii) $T(\underline{X})$ is a function of every other sufficient statistics.

Note: Note that in the above example, $T(\underline{X}) = \sum_1^n X_i$ is a f^n of all other listed sufficient statistics.

An important result to find minimal sufficient statistic

Let $f_\theta(\underline{x})$ be the joint p.d.f (or p.m.f) of $X_1, \ldots, X_n$ from P_θ $(\theta \in \Theta)$, $\underset{\sim}{x} \in \mathcal{X}$. Suppose $\exists$ a function $T(.)$ such that for every two sample points $\underset{\sim}{x}$ and $\underset{\sim}{y}$ $(\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X})$, the ratio $f_\theta(\underset{\sim}{x}) / f_\theta(\underset{\sim}{y})$ is independent of θ iff $T(\underline{x}) = T(\underline{y})$. Then $T(\underline{X})$ is a minimal sufficient statistic for θ.

<u>Remark</u>: The above result can be used to find minimal sufficient statistic.

<u>Example</u>: $X_1, \ldots, X_n$ random sample from $N(\theta, 1)$; $\theta \in \mathbb{R}$

$$f_\theta(\underline{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_1^n (x_i - \theta)^2} \qquad \underset{\sim}{x} \in \mathbb{R}^n$$

$$f_\theta(\underline{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum x_i^2 - \frac{n}{2}\theta^2 + \theta \sum x_i}$$

$\forall\, \underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$, we have

$$\frac{f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{y})} = \frac{\cancel{\left(\frac{1}{\sqrt{2\pi}}\right)^n} e^{-\sum x_i^2/2} \; \cancel{e^{-n/2\,\theta^2}} \; e^{\theta \sum x_i}}{\cancel{\left(\frac{1}{\sqrt{2\pi}}\right)^n} e^{-\sum y_i^2/2} \; \cancel{e^{-n/2\,\theta^2}} \; e^{\theta \sum y_i}}$$

i.e. $$\frac{f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{y})} = \underbrace{e^{-\frac{1}{2}\left(\sum x_i^2 - \sum y_i^2\right)}}_{\text{indep of } \theta} \; e^{\theta\left(\sum_1^n x_i - \sum y_i\right)}$$

$$\Rightarrow \frac{f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{y})} \text{ is indep of } \theta \ \underline{\text{iff}} \ \sum_1^n x_i = \sum_1^n y_i$$

$\Rightarrow$ (by the previous result) $T(\underset{\sim}{X}) = \sum_{i=1}^n X_i$ is

minimal sufficient statistic for θ.