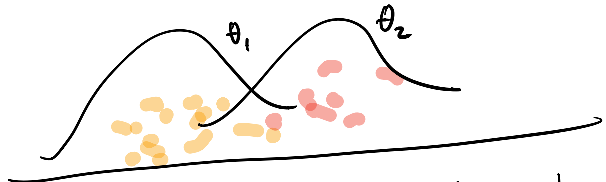
$\hat{\Theta}$ sufficient estimator for $\hat{\theta}$ if no additional info about θ can be obtained from X_i's not already obtainable from $\hat{\Theta}$.

$f(x_1, \ldots, x_n; \theta)$ $\qquad\qquad$ $f(x;\theta)$



If we can obtain info about θ from observations, $f(x;\theta)$ must depend on θ.

<u>Def</u> We say that $\hat{\Theta}$ is a sufficient estimator for θ if for any $\hat{\theta}$ a value of $\hat{\Theta}$ the conditional distribution

$$f(x_1, \ldots, x_n; \theta \mid \hat{\Theta} = \hat{\theta})$$ doesn't depend on θ.

$$\theta = \mu \qquad f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

$$f(x_1, \dots, x_n; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum (x_i-\mu)^2}$$

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$\sum (x_i-\mu)^2 = \sum_{i=1}^{n} \left[(x_i-\bar{x}) + (\bar{x}-\mu)\right]^2$$

$$= \sum (x_i-\bar{x})^2 + (\bar{x}-\mu)^2 + \underbrace{\cancel{2(x_i-\bar{x})(\bar{x}-\mu)}} \to 0$$

note: $$\sum_{i=1}^{n} (x_i-\bar{x})(\bar{x}-\mu) = 0 \; !$$

$$= \sum_{i=1}^{n} x_i\bar{x} - \bar{x}^2 - x_i\mu + \bar{x}\mu$$

$$= \bar{x}\sum_{i=1}^{n} x_i - n\bar{x}^2 - \mu\sum x_i + \sum \bar{x}\mu$$

$$= \bar{x}\cancel{n\bar{x}} - \cancel{n\bar{x}^2} - \mu\cancel{n\bar{x}} + \cancel{n\mu\bar{x}} = 0.$$

$$f(x_1, \dots, x_n; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum (x_i-\mu)^2}$$

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

$$f(x_1,\dots,x_n;\mu,\sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum (x_i-\bar{x})^2} e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2}$$

$\hat{\mu} = \bar{X}$ sufficient?

$f(x_1,\dots,x_n;\mu,\sigma^2 \mid \bar{X}=\bar{x})$

doesn't depend on μ.

$\bar{x} = \frac{1}{n}\sum x_i \qquad x_n = \bar{x} - \frac{1}{n}\sum_{i=1}^{n-1} x_i$

$x_1,\dots,x_n$

$\underline{x_1,\dots,x_{n-1},\bar{x}}$

$f(x,y)$

$f(x \mid Y=y_0)$

$\approx \dfrac{f(x,y_0)}{f_Y(y_0)}$

$$\underbrace{f\left(x_1,\dots,x_{n-1}, \bar{x}-\frac{1}{n}\sum_{i=1}^{n-1} x_i ;\mu,\sigma^2\right)}_{\text{constant}}$$

$$f(x_1,\dots,x_n;\mu,\sigma^2) = \underbrace{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum (x_i-\bar{x})^2}}_{\text{involve } x_1,\dots,x_{n-1}} \underbrace{e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2}}_{\text{constant in terms of } x_1,\dots,x_{n-1}}$$

$$f(x_1,\dots,x_n;\mu,\sigma^2 \mid \bar{X}=\bar{x}) = C e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n-1}(x_i-\bar{x})^2 + \left(\left(\bar{x}-\sum_{i=1}^{n-1}x_i\right)-\bar{x}\right)^2\right]}$$

doesn't depend on μ $\Rightarrow$ sufficient!

$f(x;\theta) \qquad f_\theta(x)$

$$\hat{\Theta} = \hat{\Theta}(X_1,\dots,X_n) = k(X_1,\dots,X_n)$$

<u>Theorem</u> "Factorization theorem"

$\hat{\Theta}$ is a sufficient estimator for θ if and only if we can write $f(x_1,\dots,x_n;\theta) = g(\hat{\theta},\theta)\,h(x_1,\dots,x_n)$

no θ

$$\hat{\theta} = k(x_1,\dots,x_n)$$

for some fns g & h

<u>Pf</u>: (a bit of) assume vars X_i are discrete.

$$p(\vec{x};\theta) = P_\theta(\vec{X}=\vec{x})$$

suppose $\hat{\Theta}$ is sufficient: $P_\theta(\vec{X}=\vec{x} \mid \hat{\Theta}=\hat{\theta})$ independent of θ.

$$\hat{\theta} = k(\vec{x})$$

$$P_\theta(\vec{X}=x \mid \hat{\Theta}=\hat{\theta}) = \frac{P_\theta(\vec{X}=\vec{x}, \hat{\Theta}=\hat{\theta})}{P_\theta(\hat{\Theta}=\hat{\theta})}$$

$$= \frac{P_\theta(\vec{X}=\vec{x})}{P_\theta(\hat{\Theta}=\hat{\theta})}$$

$$\Rightarrow P_\theta(\vec{X}=\vec{x}) = P_\theta(\vec{X}=\vec{x} \mid \hat{\Theta}=\hat{\theta})\, P_\theta(\hat{\Theta}=\hat{\theta})$$

$h(\vec{x})$

doesn't depend on θ

fcn of $\hat{\theta}$, θ

$g(\hat{\theta}, \theta)$

$\square$.