Proof of Theorem (when X is discrete)

Notation: $T = T(X)$, $t = T(x)$.

Assume T is a suff. stat. for θ.

$$f(x|\theta) = P_\theta(X = x) \quad (\text{assume} > 0)$$

$\uparrow$ pmf

$$= \underbrace{P_\theta(T=t)}_{\substack{\text{This is a}\\ \text{function of } t\\ \text{and } \theta.\\ \text{Call it } g(t|\theta).}} \cdot \underbrace{P_\theta(X = x | T = t)}_{\substack{\text{This depends on } x,\\ \text{but not } \theta \text{ (by defn.}\\ \text{of suff. stat.).}\\ \text{Call it } h(x).}}$$

$$= g(t|\theta)\, h(x) \quad \text{and FC is true.}$$

Assume FC is true.

$$P_\theta(X = x | T = t) = \frac{P_\theta(X = x)}{P_\theta(T = t)} \quad \left(\begin{array}{c}\text{since}\\ \{X = x\} \subset \{T = t\}\end{array}\right)$$

$$= \frac{f(x|\theta)}{\sum_{z:T(z)=t} f(z|\theta)} = \frac{g(t|\theta)\,h(x)}{\sum_{z:T(z)=t} g(t|\theta)\,h(z)}$$

$$= \frac{h(x)}{\sum_{z:T(z)=t} h(z)}$$

which does not involve θ. Thus T is suff. stat. for θ.

Clarification: In this proof

$$f(x|\theta) = g(T(x)|\theta)\,h(x) = g(t|\theta)\,h(x)$$

and $f(z|\theta) = g(T(z)|\theta)\,h(z) = g(t|\theta)\,h(z)$

since we are summing over z such that

$T(z) = t \; (= T(x))$.

Example: Simple Linear Regression

$$X_i = \beta_0 + \beta_1 \underset{\substack{\uparrow \\ \text{known} \\ \text{constants}}}{z_i} + \underset{\substack{\uparrow \\ \text{iid } N(0, \sigma_0^2)}}{\varepsilon_i}, \quad i = 1, \dots, n$$

σ_0^2: assume known for now

Alternative statement of model:

$X_1, X_2, \dots, X_n$ independent

$X_i \sim N(\beta_0 + \beta_1 z_i, \sigma_0^2)$

Data is $X = (X_1, X_2, \dots, X_n)$.

($z_1, \dots, z_n$ are known constants)

Unknown parameter is $\theta = (\beta_0, \beta_1) \in \mathbb{R}^2$.

What are sufficient statistics for this model?

Use factorization criterion.

$$f(x|\theta) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x_i - \beta_0 - \beta_1 z_i)^2 / 2\sigma_0^2}}_{N(\beta_0 + \beta_1 z_i, \sigma_0^2) \text{ density}}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2} \underbrace{\sum_{i=1}^{n}(x_i - \beta_0 - \beta_1 z_i)^2}_{\text{call this } S}\right\}$$

$$S = \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i(\beta_0 + \beta_1 z_i) + \sum_{i=1}^{n}(\beta_0 + \beta_1 z_i)^2$$

$$= \sum x_i^2 - 2\beta_0 \sum x_i - 2\beta_1 \sum x_i z_i + \sum(\beta_0 + \beta_1 z_i)^2$$

Plug this back into the exponential and rearrange to get

$f(x|\theta) =$

$$\left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\left(-2\beta_0 \sum x_i - 2\beta_1 \sum x_i z_i + \sum(\beta_0 + \beta_1 z_i)^2\right)\right\}$$

times $\exp\left\{-\frac{1}{2\sigma_0^2}\sum x_i^2\right\}$

$$= g\left(\sum x_i, \sum x_i z_i, \beta_0, \beta_1\right) h(x)$$

$$= g(T(x), \theta)\, h(x)$$

where $T(x) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i z_i\right)$ and

$g(t,\theta) =$

$$\left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\left(-2\beta_0 t_1 - 2\beta_1 t_2 + \sum_{i=1}^n (\beta_0+\beta_1 z_i)^2\right)\right\}$$

with $t=(t_1,t_2)$ and

$$h(x) = \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2\right\}.$$

Continuation of Simple Linear Regression Example:

What if the variance σ^2 is unknown?

Now $\theta = (\beta_0, \beta_1, \sigma^2)$ and $\Theta = \mathbb{R}^2 \times (0, \infty)$.

(Change σ_0^2 to σ^2 in earlier formulas to indicate this.)

Now $\exp\left\{-\frac{1}{2\sigma^2}\sum_i x_i^2\right\}$ is not a function of x, but depends also on θ.

So we now factor the joint density as

$$
\begin{aligned}
f(x \mid \theta) & \\
= \quad & \left[\left(2\pi\sigma^2\right)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \left(\sum_i x_i^2 - 2\beta_0 \sum_i x_i - 2\beta_1 \sum_i z_i x_i \right.\right.\right. \\
& \left.\left.\left. + \sum_i (\beta_0 + \beta_1 z_i)^2 \right) \right\} \right] \cdot 1 \\
= \quad & g\left(\sum_i x_i^2, \sum_i x_i, \sum_i z_i x_i, \beta_0, \beta_1, \sigma^2\right) h(x) \\
= \quad & g\left(T(x), \theta\right) h(x)
\end{aligned}
$$

where

$$
\begin{aligned}
T(x) &= \left(\sum_i x_i^2, \sum_i x_i, \sum_i z_i x_i\right) = (t_1, t_2, t_3) \\
g(t, \theta) &= \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \left(t_1 - 2\beta_0 t_2 - 2\beta_1 t_3 \right.\right. \\
& \qquad \left.\left. + \sum_i (\beta_0 + \beta_1 z_i)^2 \right) \right\} .
\end{aligned}
$$

According to the FC, $T(X) = \left(\sum_i X_i^2, \sum_i X_i, \sum_i z_i X_i\right)$ is a sufficient statistic for $\theta = (\beta_0, \beta_1, \sigma^2)$.

Discussion:

We have described two models.

The model with σ^2 known (i.e., $\sigma^2 = \sigma_0^2$) can be regarded as a subset of the model where σ^2 is unknown.

$\Theta_1 = \left\{(\beta_0, \beta_1, \sigma^2) \,:\, \sigma^2 = \sigma_0^2\right\} = \mathbb{R}^2 \times \{\sigma_0^2\}$.

$\Theta_2 = \left\{(\beta_0, \beta_1, \sigma^2) \,:\, \sigma^2 > 0\right\} = \mathbb{R}^2 \times (0, \infty)$.

$\Theta_1 \subset \Theta_2$.

The sufficient statistics we found for these two models were different:

$T_1 \equiv \left(\sum_i X_i, \sum_i z_i X_i\right)$ is SS for Θ_1.

$T_2 \equiv \left(\sum_i X_i^2, \sum_i X_i, \sum_i z_i X_i\right)$ is SS for Θ_2.

Note: T_2 is also a SS for Θ_1, but it is not "minimal". This is discussed shortly.