Reminder : $T = T(\underline{X})$ sufficient statistic for $\theta \in \Theta$ if $P(X_1 = x_1, \dots, X_\nu = x_\nu)$ independent of θ

Example

Let $X_1, X_2, \dots, X_\nu$ r.s. of Poisson(λ). Prove that $T = \sum_{i=1}^{\nu} X_i$ is a sufficient statistic.

Solution

$$P(X_1 = x_1, \dots, X_\nu = x_\nu \mid T = t) = \begin{cases} \dfrac{P(X_1 = x_1, \dots, X_\nu = x_\nu, T = t)}{P(T = t)}, & \sum x_1 = t \\ 0, & \sum x_1 \neq t \end{cases}$$

If $\sum X_1 = t$: $P(X_1 = x_1, \dots, X_\nu = x_\nu, T = t) = P(X_1 = x_1) \cdot \dots \cdot P(X_\nu = x_\nu)$

Since $X_1, X_2, \dots X_\nu$ r.s. of Poisson(λ) : $T = T(\underline{X}) \sim$ Poisson($\nu\lambda$)

Hence : $$\frac{P(X_1 = x_1) \cdot \dots \cdot P(X_\nu = x_\nu)}{P(T = t)} = \frac{(e^{-\lambda}\lambda^{x_1}/x_1!) \cdot \dots \cdot (e^{-\lambda}\lambda^{x_\nu}/x_\nu!)}{e^{-\nu\lambda} \cdot (\nu\lambda)^{\sum x_i}/(\sum x_i)!} =$$

$$= \frac{e^{-\nu\lambda}\lambda^{\sum x_i}/\prod(x_i!)}{e^{-\nu\lambda}\nu^{\sum x_i} \cdot \lambda^{\sum x_i}/(\sum x_i)!} = \frac{(\sum x_i)!}{\nu^{\sum x_i} \cdot \prod(x_i!)}$$, which does not include λ.

Note

The statistics $T_1 = (X_1, \sum_{i=1}^{\nu} X_i) = T_1(\underline{X})$

$T_2 = (X_1, X_2, \sum_{i=1}^{\nu} X_i) = T_2(\underline{X})$

are also sufficient for λ. Indeed :

$$P(X_1 = x_1, \dots, X_\nu = x_\nu \mid T_1(\underline{X}) = \underline{t_1}) = \begin{cases} \dfrac{P(X_1 = x_1, X_2 = x_2, \dots, X_\nu = x_\nu, \underline{T_1} = \underline{t_1})}{P(\underline{T_1} = \underline{t_1})}, & \text{if } T_1 = t_1 \\ 0, & \text{else} \end{cases}$$

and $P(X_1 = x_1, X_2 = x_2, \dots, X_\nu = x_\nu \mid \underline{T_1} = \underline{t_1}) =$

$$= \frac{P(X_1 = x_1) \cdot \dots \cdot P(X_\nu = x_\nu)}{P(X_1 = x_1) \cdot P(\sum x_i = t_0)} = \dots = (\sum x_i)! / (\nu - 1)^{\sum x_i} \cdot \prod(x_i!)$$

which does not include λ.

It's clear that there are multiple sufficient statistics. We will be looking for a minimal one

Definition

A statistic $T = T(\underline{X})$ is called minimal sufficient statistic iff
- T is sufficient
- if $T'(\underline{X})$ is sufficient, $\exists f$ such that $T' = f(T)$

(From now on when we say "sufficient" we refer to the minimal)

Examples

- Let $X_1, \dots, X_n$ r.s. Bernoulli(θ). Prove that $T = \sum X_i$ sufficient for θ.

Solution

$f(x_i; \theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, $x_i = 0, 1, \dots$, $\theta \in (0,1)$

$T = \sum X_i$ so $T \sim Bin(n, \theta)$

Hence $P(X_1 = x_1, \dots, X_n = x_n \mid T = t) = \dfrac{P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n)}{P(T = t)} =$

$= \dfrac{\theta^{x_1}(1-\theta)^{x_1} \cdot \dots \cdot \theta^{x_n}(1-\theta)^{x_n}}{\binom{1}{t}\theta^t(1-\theta)^{n-t}} = \binom{n}{t}^{-1}$, which does not include θ

- Let X_1, X_2, X_3 r.s Bernoulli(θ), $T = X_1 + 2X_2 + X_3$. Then:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \mid T = t) = \frac{P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot P(X_3 = x_3)}{P(T = t)}$$

We suppose, as a counterexample, $x_1 = 1$, $x_2 = 0$, $x_3 = 1$ and we have $T = 2$ and $\dfrac{P(X_1 = 1) P(X_2 = 0) P(X_3 = 1)}{P(T = 2)} = \dfrac{\theta(1-\theta)\theta}{\theta(1-\theta)\theta + (1-\theta)\theta(1-\theta)} =$

$= \dfrac{\theta^2(1-\theta)}{\theta^2(1-\theta) + \theta(1-\theta)^2} = \dfrac{\theta}{\theta + 1 - \theta} = \theta$ includes θ

For the statistic to be sufficient is has to be independent of θ for every $\underline{x}$ and for $\underline{x} = (1, 0, 1)$ this is not true and hence T is not sufficient.

Theorem (Fischer-Neyman factorization criterion)

Let $\underline{X} = (X_1, \ldots, X_n)$ r.s of distribution with PDF $f(\underline{x}, \underline{\theta})$ $\underline{\theta} = (\theta_1, \ldots, \theta_s)$. Then the statistic $\underline{T}(\underline{x}) = (T_1(\underline{x}), T_2(\underline{x}), \ldots, T_k(\underline{x}))$ is sufficient for $\underline{\theta}$ iff $\exists g, h$ such that:

$$f(\underline{x} ; \underline{\theta}) = g(T(\underline{x}) ; \underline{\theta}) \cdot h(\underline{x})$$

(g depends on x only via $T(x)$ and h is independent of $\underline{\theta}$)