Optimal Estimation

Sufficient Statistics

Sufficient means Enough

Idea: If you know $T = T(X_1, \ldots, X_n)$ that's all you need. It contains all the information about Θ.

Def Let $X_1, \ldots, X_n$ be a random sample from a distribution with parameter Θ. The statistic $T = T(X_1, \ldots, X_n)$ is said to be sufficient for Θ if the conditional distribution of $X_1, \ldots, X_n$ Given T does not depend on Θ.

Comments

- T & Θ could be vectors
- T could be $\hat{\Theta}$, but not necessarily
- Easiest to see for discrete distributions

Ex $X_1, \dots, X_n \overset{iid}{\sim}$ Poisson (λ)

Know $T = \sum_{i=1}^n X_i \sim P(n\lambda)$ by MGF

Want conditional distribution of $X_1 \dots X_n \mid T = t$

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid \sum_{i=1}^n X_i = t)$$

$$= \frac{P(X_1 = x_1, \dots, X_n = x_n, \sum_{i=1}^n X_i = t)}{P(\sum_{i=1}^n X_i = t)}$$ = 0 unless $t = \sum_{i=1}^n x_i$

$$= \frac{P(\underbrace{\sum_{i=1}^n X_i = \sum_{i=1}^n x_i}_{\text{Redundant}}, X_1 = x_1, \dots, X_n = x_n)}{P(\sum_{i=1}^n X_i = \sum_{i=1}^n x_i)}$$

$$= \frac{\overbrace{P(\sum_{i=1}^n X_i = \sum_{i=1}^n x_i \mid X_1 = x_1, \dots X_n = x_n)}^{=1} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)}{P(\Sigma X_i = \Sigma x_i)}$$

$$= \frac{1 \cdot P(X_1 = x_1, \dots, X_n = x_n)}{P(\Sigma X_i = \Sigma x_i)} \overset{ind}{=} \frac{\prod_{i=1}^n P(X_i = x_i)}{P(\Sigma X_i = \Sigma x_i)}$$

$$= \frac{\prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}}{\frac{e^{-n\lambda} (n\lambda)^{\Sigma x_i}}{(\Sigma x_i)!}} = \frac{\cancel{e^{-n\lambda}} \cancel{\lambda^{\Sigma x_i}} / \prod_{i=1}^n x_i!}{\cancel{e^{-n\lambda}} n^{\Sigma x_i} \cancel{\lambda^{\Sigma x_i}} / (\sum_{i=1}^n x_i)!}$$

$$= \frac{(\Sigma x_i)!}{\prod_{i=1}^{n} x_i! \; n^{\Sigma x_i}} \prod_{i=1}^{n} I(x_i = 0, 1, \cdots)$$

This was invisible

Free of λ, so $\sum_{i=1}^{n} X_i$ is sufficient.

So is $\bar{X}_n$

So is <u>any</u> 1-1 function of ΣX_i

So is $\left(\Sigma X_i , \prod_{i=1}^{n} X_i ! \right)$

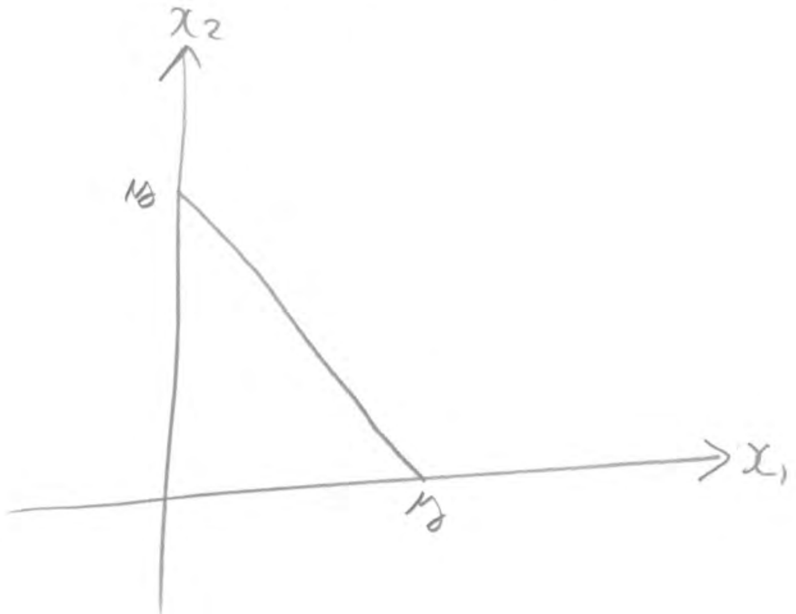
So is $T = (X_1, \cdots X_n)$

(But this answer will get you a zero if you are asked to find a sufficient statistic)

For continuous random variables the definition still makes sense, but it gets technical.

Ex $X_1, X_2 \overset{iid}{\sim} Exp(\lambda)$. Show $Y_1 = X_1 + X_2$ is sufficient for λ.

Need to look at the conditional distribution of X_1 & X_2 GIVEN $X_1 + X_2 = y \Leftrightarrow X_2 = y - X_1$



In 2-D, all the conditional probability is concentrated on the line $x_2 = y - x_1$

There is no (conditional) joint density of X_1 & X_2. In 2-D, the distribution is degenerate.

On the other hand, there IS a density of X_1 (or X_2) given $X_1 + X_2 = y$. One RV is redundant. obtain by subtraction.

So look at the conditional density of X_2 given $X_1 + X_2$.

$$Y_1 = X_1 + X_2 \Longrightarrow X_1 = Y_1 - Y_2 \qquad \begin{vmatrix} \frac{\partial x_1}{\partial y_1} = 1 & \frac{\partial x_1}{\partial y_2} = -1 \\ \frac{\partial x_2}{\partial y_1} = 0 & \frac{\partial x_2}{\partial y_2} = 1 \end{vmatrix}$$

$$Y_2 = X_2 \qquad\qquad X_2 = Y_2$$

$$= 1$$

And

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(x_1(y_1, y_2), x_2(y_1, y_2))\ |J|$$

$$= \lambda e^{-\lambda(y_1 - y_2)} I(y_1 - y_2 > 0)\ \lambda e^{-\lambda y_2} I(y_2 > 0) * 1$$

$$= \lambda^2 e^{-\lambda y_1} \cancel{e^{\lambda y_2}} I(y_1 > y_2)\ \cancel{e^{-\lambda y_2}} I(y_2 > 0)$$

$$= \lambda^2 e^{-\lambda y_1} I(0 < y_2 < y_1), \text{ and}$$

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_1 Y_2}(y_1, y_2)}{f_{Y_1}(y_1)} \qquad \left(\begin{array}{l} M_{Y_1}(t) = \left(1 - \frac{t}{\lambda}\right)^{-2} \\ Gamma(\alpha = 2, \lambda) \end{array} \right)$$

$$= \frac{\cancel{\lambda^2} \cancel{e^{-\lambda y_1}} I(0 < y_2 < y_1)}{\frac{\cancel{\lambda^2}}{\Gamma(2)} \cancel{e^{-\lambda y_1}} y_1^{2-1}} = \frac{1}{y_1} I(0 < y_2 < y_1)$$

where $y_1 > 0$

UNIFORM $(0, y_1)$. FREE of λ

$Y_1 = X_1 + X_2$ is sufficient

For $X_1, \ldots X_n$, same approach

$$Y_1 = X_1 \qquad X_1 = Y_1$$
$$Y_2 = X_2 \qquad X_2 = Y_2$$
$$\vdots$$
$$Y_{n-1} = X_{n-1}$$
$$Y_n = \sum_{i=1}^{n} X_i \qquad X_n = Y_n - \sum_{i=1}^{n-1} Y_i$$

$n \times n$
Jacobian

$$f(y_1, \ldots, y_{n-1} \mid y_n)$$

Inconvenient

Theorem (The Factorization Theorem)

The statistic $T = T(X_1, \ldots, X_n)$ is sufficient iff the joint density or pmf of $X_1, \ldots X_n$ can be factored so that

$$\prod_{i=1}^{n} f(x_i \mid \theta) = g(t, \theta)\, h(x_1, \ldots x_n)$$

where $g(t, \theta)$ is a function of $x_1, \ldots x_n$ only through $t(x_1, \ldots, x_n)$, and $h(\underset{\sim}{x})$ is not a function of θ.

Exponential Example

$$f(\underset{\sim}{x} \mid \theta) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} I(x_i \geq 0)$$

$$= \underbrace{\lambda^n e^{-\lambda \Sigma x_i}}_{g(t = \Sigma x_i,\, \lambda)} \underbrace{\prod_{i=1}^{n} I(x_i \geq 0)}_{h(\underset{\sim}{x})}$$

Normal Example

$$f(\underset{\sim}{x} \mid \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2}$$

$$= \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= \frac{1}{(\sigma^2)^{n/2}(2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\left[\Sigma(x_i - \bar{x})^2 + n(\bar{x}-\mu)^2\right]}$$

Letting $h(\underset{\sim}{x}) = (2\pi)^{-n/2}$, see

$\left(\bar{x}, \sum_{i=1}^{n}(x_i - \bar{x})^2\right)$ is sufficient for (μ, σ^2)

or $(\bar{x}, \hat{\sigma}^2)$

Proof for the discrete case

① Suppose $P_\theta(\underset{\sim}{X} = \underset{\sim}{x}) = g(t(\underset{\sim}{x}), \theta)\, h(\underset{\sim}{x})$.
Show $T = T(\underset{\sim}{X})$ is sufficient.

$$P(\underset{\sim}{X} = \underset{\sim}{x} \mid T = t) = \frac{P_\theta(\underset{\sim}{X} = \underset{\sim}{x},\, T = t)}{P_\theta(T = t)}$$

Note that t is a fixed constant.

$$= \frac{P_\theta(T(\underset{\sim}{X}) = t \mid \underset{\sim}{X} = \underset{\sim}{x})\, P_\theta(\underset{\sim}{X} = \underset{\sim}{x})}{P_\theta(T = t)}$$

$$= \frac{1 \cdot P_\theta(\underset{\sim}{X} = \underset{\sim}{x})}{P_\theta(T = t)} = \frac{g(t, \theta)\, h(\underset{\sim}{x})}{P_\theta(T = t)}$$

$$= \frac{g(t, \theta)\, h(\underset{\sim}{x})}{\sum\limits_{\underset{\sim}{y}: T(\underset{\sim}{y}) = t} g(t, \theta)\, h(\underset{\sim}{y})} = \frac{\cancel{g(t, \theta)}\, h(\underset{\sim}{x})}{\cancel{g(t, \theta)} \sum\limits_{\underset{\sim}{y}: T(\underset{\sim}{y}) = t} h(\underset{\sim}{y})}$$

$$= \frac{h(\underset{\sim}{x})}{\sum\limits_{\underset{\sim}{y}: T(\underset{\sim}{y}) = t} h(\underset{\sim}{y})}$$

Free of θ. Sufficient

(2) Suppose T is sufficient.

Show $P_\theta(\underset{\sim}{X}=\underset{\sim}{x}) = g(t,\theta)\, h(\underset{\sim}{x})$

$P(\underset{\sim}{X}=\underset{\sim}{x} \mid T=t)$ is free of θ

$$\|$$

$$\frac{P_\theta(T=t,\ \underset{\sim}{X}=\underset{\sim}{x})}{P_\theta(T=t)} = \frac{P(T=t \mid \underset{\sim}{X}=x)\, P_\theta(\underset{\sim}{X}=\underset{\sim}{x})}{P_\theta(T=t)}$$

$$= \frac{1 \cdot P_\theta(\underset{\sim}{X}=\underset{\sim}{x})}{P_\theta(T=t)} \overset{\text{start}}{=} P(\underset{\sim}{X}=\underset{\sim}{x} \mid T=t)$$

$$\Longrightarrow P_\theta(\underset{\sim}{X}=\underset{\sim}{x}) = \underbrace{P_\theta(T=t)}_{g(t,\theta)}\ \underbrace{P(\underset{\sim}{X}=\underset{\sim}{x} \mid T=t)}_{h(x)}$$

$\blacksquare$

$$f(\underset{\sim}{x} \mid \theta) \text{ or } p(\underset{\sim}{x} \mid \theta) = g(t, \theta)\, h(\underset{\sim}{x})$$

Maximum likelihood

$$\ell(\theta, x) = \ln f(\underset{\sim}{x} \mid \theta) = \ln\left(g(t,\theta)\, h(\underset{\sim}{x})\right)$$

$$= \ln g(t(\underset{\sim}{x}), \theta) + \ln h(\underset{\sim}{x})$$

The MLE depends on the data only through the value of a sufficient statistic

Bayes Methods

$$\pi(\theta \mid \underset{\sim}{x}) \propto f(x \mid \theta)\, \pi(\theta)$$

$$= g(t(\underset{\sim}{x}), \theta)\, h(x)\, \pi(\theta)$$

$$\propto g(t(\underset{\sim}{x}), \theta)\, \pi(\theta)$$

The posterior distribution depends on the data only through the value of a sufficient statistic.