Minimal sufficiency and completeness

Recap: minimal sufficient statistics

Definition: A statistic $T(X_1, \ldots, X_n)$ is a *minimal sufficient statistic* if for any other sufficient statistic $T^*(X_1, \ldots, X_n)$, $T(X_1, \ldots, X_n)$ is a function of $T^*(X_1, \ldots, X_n)$.

Theorem: Let $X_1, \ldots, X_n$ be a sample with probability function $f(x_1, \ldots, x_n | \theta)$, and let $T(x_1, \ldots, x_n)$ be a statistic such that

$$\frac{f(x_1, \ldots, x_n | \theta)}{f(y_1, \ldots, y_n | \theta)}$$

is constant as a function of θ if and only if $T(x_1, \ldots, x_n) = T(y_1, \ldots, y_n)$. Then T is a minimal sufficient statistic for θ

Example

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} Uniform[\theta, \theta+1]$.

$$f(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} \mathbb{1}\{X_i \in [\theta, \theta+1]\}$$

$$= \mathbb{1}\{\theta \leq X_i \leq \theta+1 \ \forall i\}$$

$$= \mathbb{1}\{\theta \leq X_{(1)} \text{ and } X_{(n)} \leq \theta+1\}$$

$$= \mathbb{1}\{X_{(n)} - 1 \leq \theta \leq X_{(1)}\}$$

$$\frac{f(x_1, \ldots, x_n | \theta)}{f(y_1, \ldots, y_n | \theta)} \text{ is constant} \iff$$

$$\mathbb{1}\{x_{(n)} - 1 \leq \theta \leq x_{(1)}\} = \mathbb{1}\{y_{(n)} - 1 \leq \theta \leq y_{(1)}\}$$

for all θ

$$\iff x_{(1)} = y_{(1)} \text{ and } x_{(n)} = y_{(n)}$$

$\Rightarrow (X_{(1)}, X_{(n)})$ is a minimal sufficient statistic.

Example

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$.

Find a minimal sufficient statistic for λ.

$$\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad f(x_1,\ldots,x_n \mid \lambda) = \frac{e^{-n\lambda}\,\lambda^{\sum_i x_i}}{\prod_i x_i!}$$

$$\frac{f(x_1,\ldots,x_n \mid \lambda)}{f(y_1,\ldots,y_n \mid \lambda)} = \frac{\cancel{e^{-n\lambda}}\,\lambda^{\sum_i x_i} / \prod_i x_i!}{\cancel{e^{-n\lambda}}\,\lambda^{\sum_i y_i} / \prod_i y_i!}$$

$$= \left(\lambda^{\sum_i x_i - \sum_i y_i}\right) \frac{\prod_i y_i!}{\prod_i x_i!}$$

constant as a function $\iff \sum_i x_i = \sum_i y_i \implies T = \sum_i X_i$ is MSS

Recap: Rao-Blackwell

Rao-Blackwell theorem: Let θ be a parameter of interest, and $\hat{\tau}$ an unbiased estimator of $\tau(\theta)$. If T is a sufficient statistic for θ, then $\tau^* = \mathbb{E}[\hat{\tau} | T]$ is an unbiased estimator of $\tau(\theta)$, and $Var(\tau^*) \leq Var(\hat{\tau})$.

> If we condition on the "right" sufficient statistic, does this process find the best unbiased estimator?

Completeness

Def: Let $X_1, \dots, X_n$ be a sample and $T \equiv T(X_1, \dots, X_n)$ be a statistic. Let $f_T(t|\theta)$ denote the pdf or pmf of T, and suppose T has the property that if there exists a function g for which $\mathbb{E}_\theta[g(T)] = 0$ for all θ, then $P_\theta(g(T) = 0) = 1$ for all θ. Then T is a <u>complete</u> statistic

<u>Example</u>: $X_1, \dots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$. $T = \sum_i X_i$

$$T \sim \text{Poisson}(n\lambda) \Rightarrow \mathbb{E}_\lambda[g(T)] = \sum_{t=0}^{\infty} g(t) \frac{e^{-n\lambda}(n\lambda)^t}{t!}$$

Suppose $\mathbb{E}_\lambda[g(T)] = 0 \quad \forall \lambda$

$$\Rightarrow \sum_{t=0}^{\infty} g(t) \frac{(n\lambda)^t}{t!} = 0 \quad \forall \lambda$$

$$\Rightarrow \frac{g(t)}{t!} = 0 \quad \forall \lambda, \forall t$$

$$\Rightarrow g(T) = 0 \quad \forall \lambda \Rightarrow P_\lambda(g(T) = 0) = 1$$

Then T is complete

$X \sim \text{Uniform}[\theta, \theta+1]$

X is minimal sufficient for θ

Now let $g(x) = \sin(2\pi X)$

$$\forall \theta, \quad \mathbb{E}_\theta[g(X)] = \int_\theta^{\theta+1} \sin(2\pi x)\,dx = 0$$

But $g \not\equiv 0$, so $P_\theta(g(X) = 0) \neq 1$

So X is <u>not</u> a complete statistic

Lehmann-Scheffé theorem

Applying completeness to unbiased estimators:

Let T be complete, and suppose $\varphi(T)$ and $\psi(T)$ are two unbiased estimators of θ. Then

$\mathbb{E}_\theta[\varphi(T)] = \mathbb{E}_\theta[\psi(T)] = \theta \quad \forall \theta$, so

$\mathbb{E}_\theta[\varphi(T) - \psi(T)] = 0 \quad \forall \theta$. But T is complete,

so $\varphi(T) - \psi(T) = 0 \quad \forall \theta \implies \varphi(T) \equiv \psi(T)$

And so $\varphi(T)$ is <u>unique</u> (there is only one unbiased estimator of θ which is a function only of T).

<u>Theorem</u>: Suppose $T \equiv T(X_1, \dots, X_n)$ is a complete, sufficient statistic, and $\varphi(T)$ is an unbiased estimator of θ. Then $\varphi(T)$ is the best unbiased estimator of θ.