Lecture 19: Sufficiency

Mathematical Statistics I, MATH 60061/70061

Tuesday November 30, 2021

Reference: Casella & Berger, 6.1-6.2.1

Factorization Theorem

So far, we've used the definition of sufficiency directly by showing that the conditional distribution of X given T is free of θ. What if we need to find a sufficient statistic?

Factorization Theorem: A statistic $T = T(\boldsymbol{X})$ is **sufficient** for θ if and only if there exists functions $g(t \mid \theta)$ and $h(\boldsymbol{x})$ such that

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = g(t \mid \theta)h(\boldsymbol{x}),$$

for all sample points $\boldsymbol{x} \in \mathcal{X}$ and all $\theta \in \Theta$.

Normal sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid from $\mathcal{N}(\mu, \sigma^2)$, with parameter $\theta = (\mu, \sigma^2)$. The joint PDF is

$$
\begin{aligned}
f_X(x \mid \boldsymbol{\theta}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right) \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)
\end{aligned}
$$

where s^2 is the realization of the sample variance $S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

$$f_X(x \mid \boldsymbol{\theta}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{(n-1)s^2}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)$$

By the Factorization Theorem, $(\bar{X}, S^2)$ is a sufficient statistic for $\theta = (\mu, \sigma^2)$.

- If σ^2 is known, then $\bar{X}$ is sufficient for μ.
- If μ is known, then S^2 is sufficient for σ^2.
- If both μ and σ^2 are unknown, we cannot say that $\bar{X}$ is sufficient for μ (or S^2 is sufficient for σ^2); the correct statement is that $\bar{X}$ and S^2 together is sufficient for μ and σ^2.
- We can also say that $(\bar{X}, S^2)$ is sufficient for μ (or σ^2).

Sufficient statistics in the Exponential family

Suppose $X_1, \ldots, X_n$ are iid from the **Exponential family**

$$f_X(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, $d \leq k$. Then

$$\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is sufficient for $\boldsymbol{\theta}$.

Binomial sufficient statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(\theta)$ with parameter $0 < \theta < 1$. Then $T(\boldsymbol{X}) = X_1 + \cdots + X_n$ is a sufficient statistic for θ.

The PMF of $\boldsymbol{X}$ is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}.$$

Note that $T(\boldsymbol{X})$ counts the number of X_i's that equal 1, so $T(\boldsymbol{X})$ has a $\mathrm{Bin}(n, \theta)$ distribution,

$$f_T(t \mid \theta) = \binom{n}{t}\theta^t(1-\theta)^{n-t}.$$

With $t = \sum_{i=1}^n x_i$, the conditional distribution

$$f_{\boldsymbol{X}|T}(\boldsymbol{x} \mid t) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_T(t \mid \theta)} = \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{\sum x_i}},$$

which is free of θ. Therefore, $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic.

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(\theta)$ with parameter $0 < \theta < 1$. For $x = 0, 1$, the PMF of X is

$$\begin{aligned} f_X(x \mid \theta) &= \theta^x (1-\theta)^{1-x} \\ &= (1-\theta)\left(\frac{\theta}{1-\theta}\right)^x \\ &= (1-\theta)\exp\left(\log\left(\frac{\theta}{1-\theta}\right)x\right) \\ &= h(x)c(\theta)\exp(w_1(\theta)t_1(x)), \end{aligned}$$

where $h(x) = 1$, $c(\theta) = 1 - \theta$, $w_1(\theta) = \log(\theta/(1-\theta))$, and $t_1(x) = x$.

Therefore,

$$T = T(\boldsymbol{X}) = \sum_{i=1}^n t_1(X_i) = \sum_{i=1}^n X_i$$

is sufficient.