Factorization Theorem

Theorem 1.1 *The joint pmf/pdf of* $X_1,\ X_2, \ldots,\ X_n$ *is said to satisfy the* factorization criterion *in terms of the statistic* $T = t(\mathbf{X})$ *if and only if it can be expressed as*

$$f_{\mathbf{X}}(\mathbf{x}, \theta) = g\left(t(\mathbf{x}), \theta\right) h(\mathbf{x})$$

Note that $g\left(t(\mathbf{x}), \theta\right)$ depends on $\mathbf{x}$ only through $t(\mathbf{x})$ and that $h(\mathbf{x})$ is independent of θ.

Proof 1.1

($\Leftarrow$)

Suppose that

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}, \theta) &= P\left(\mathbf{X} = \mathbf{x}\right) \\ &= p_{\mathbf{X}}(\mathbf{x}) \end{aligned}$$

and assume that $p_{\mathbf{X}}(\mathbf{x}) = g\left(t(\mathbf{x}), \theta\right) h(\mathbf{x})$*. Then*

$$\begin{aligned} \mathrm{P}(T = t_0) &= \sum_{t(\mathbf{x})=t_0} \mathrm{P}(\mathbf{X} = \mathbf{x}) && \textit{summing over those } \mathbf{x} \textit{ for which } t(\mathbf{x}) = t_0 \\ &= \sum_{t(\mathbf{x})=t_0} g\left(t(\mathbf{x}), \theta\right) h(\mathbf{x}) && \textit{by factorization (assumed)} \\ &= g(t_0, \theta) \sum_{t(\mathbf{x})=t_0} h(\mathbf{x}) && \textit{since } t(\mathbf{x}) = t_0 \end{aligned}$$

The conditional distribution of $\mathbf{X}$ *given* $T = t_0$ *is*

$$\begin{aligned} P\left(\mathbf{X} = \mathbf{x} | T = t_0\right) &= \frac{P\left(\mathbf{X} = \mathbf{x} \cap T = t_0\right)}{P\left(T = t_0\right)} \\ &= \begin{cases} \frac{P(\mathbf{X}=\mathbf{x})}{P(T=t_0)} & \textit{if } t(\mathbf{x}) = t_0 \\ 0 & \textit{else} \end{cases} \end{aligned}$$

If $t(\mathbf{x}) = t_0$ *then*

$$\begin{aligned} P\left(\mathbf{X} = \mathbf{x} | T = t_0\right) &= \frac{g(t_0, \theta) h(\mathbf{x})}{g(t_0, \theta) \sum_{t(\mathbf{x})=t_0} h(\mathbf{x})} \\ &= \frac{h(\mathbf{x})}{\sum_{t(\mathbf{x})=t_0} h(\mathbf{x})} \end{aligned}$$

which is independent of θ*. So,* T *is sufficient for* θ*.*

($\Rightarrow$)

Now suppose that T *is sufficient for* θ*. Then, by definition,* $\mathrm{P}(\mathbf{X} = \mathbf{x} | T = t_0)$ *is independent of* θ*. If we let* $g(t_0, \theta) = \mathrm{P}(T(\mathbf{X}) = t_0)$ *and* $h(\mathbf{x}) = \mathrm{P}(\mathbf{X} = \mathbf{x} | T(\mathbf{x} = t_0)$ *then*

$$\begin{aligned} \mathrm{P}(\mathbf{X} = \mathbf{x} | \theta) &= \mathrm{P}(\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = t_0) \\ &= \mathrm{P}(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t_0) \mathrm{P}(T(\mathbf{X}) = t_0 | \theta) \\ &= h(\mathbf{x}) g(t_0, \theta) \end{aligned}$$

Thus, if we can factor $f(\mathbf{x}, \theta)$*, then* $T(\mathbf{X}) = t(\mathbf{X})$ *is a sufficient statistic.*