Biostatistics 602 - Statistical Inference
Lecture 02
Sufficient Statistics
Factorization Theorem

Hyun Min Kang

January 15th, 2013

Last Lecture

Definition 6.2.1

A statistic $T(\mathbf{X})$ is a *sufficient statistic* for θ if the conditional distribution of sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

Example

- Suppose $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p),\ 0 < p < 1.$
- $T(X_1, \cdots, X_n) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for p.

A Theorem for Sufficient Statistics

Theorem 6.2.2

- Let $f_{\mathbf{X}}(\mathbf{x}|\theta)$ is a joint pdf or pmf of X
- and $q(t|\theta)$ is the pdf or pmf of $T(\mathbf{X})$.
- Then $T(\mathbf{X})$ is a sufficient statistic for θ,
- if, for every $\mathbf{x} \in \mathcal{X}$,
- the ratio $f_{\mathbf{X}}(\mathbf{x}|\theta)/q(T(\mathbf{x})|\theta)$ is constant as a function of θ.

Proof of Theorem 6.2.2 - discrete case

$$\begin{aligned}
\Pr(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t) &= \frac{\Pr(\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = t)}{\Pr(T(\mathbf{X}) = t)} \\
&= \begin{cases} \dfrac{\Pr(\mathbf{X} = \mathbf{x})}{\Pr(T(\mathbf{X}) = t)} & \text{if } T(\mathbf{x}) = t \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} \dfrac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)} & \text{if } T(\mathbf{x}) = t \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

which does not depend on θ by assumption. Therefore, $T(\mathbf{X})$ is a sufficient statistic for θ.

Example 6.2.3 - Binomial Sufficient Statistic

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$, $0 < \theta < 1$.
- Show that $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

This is the same problem from the last lecture, but we would like to solve is using Theorem 6.2.2.

Example 6.2.3 - Binomial Sufficient Statistic

Proof

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}|p) &= p^{x_1}(1-p)^{1-x_1}\cdots p^{x_n}(1-p)^{1-x_n} \\
&= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \\
T(\mathbf{X}) &\sim \text{Binomial}(n,p) \\
q(t|p) &= \binom{n}{t} p^t (1-p)^{n-t} \\
\frac{f_{\mathbf{X}}(\mathbf{x}|p)}{q(T(\mathbf{x})|p)} &= \frac{p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}{\binom{n}{t} p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}} \\
&= \frac{1}{\binom{n}{\sum_{i=1}^n x_i}} = \frac{1}{\binom{n}{T(\mathbf{x})}}
\end{aligned}
$$

By theorem 6.2.2. $T(\mathbf{X})$ is a sufficient statistic for p.

Example 6.2.4 - Normal Sufficient Statistic

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$
- Assume that σ^2 is known.
- Show that the sample mean $T(\mathbf{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is a sufficient statistic for μ.

Example 6.2.4 - Proof

$f_{\mathbf{X}}(\mathbf{x}|\mu)$

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}|\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^n \frac{(x_i-\overline{x}+\overline{x}-\mu)^2}{2\sigma^2}\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^n (x_i-\overline{x})^2 + n(\overline{x}-\mu)^2}{2\sigma^2}\right)
\end{aligned}
$$

Example 6.2.4 - Proof (cont'd)

$q(T(\mathbf{x}|\mu))$

Remember from BIOSTAT601 that $T(\mathbf{X}) = \overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.

$$q(T(\mathbf{x}|\mu)) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left(-n(\overline{x}-\mu)^2/(2\sigma^2)\right)$$

Example 6.2.4 - Proof

Putting things together

$$\begin{aligned}\frac{f_{\mathbf{X}}(\mathbf{x}|\mu)}{q(T(\mathbf{x}|\mu))} &= \frac{(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{\sum_{i=1}^n(x_i-\overline{x})^2+n(\overline{x}-\mu)^2}{2\sigma^2}\right)}{(2\pi\sigma^2/n)^{-1/2}\exp\left(-\frac{n(\overline{x}-\mu)^2}{2\sigma^2}\right)} \\ &= n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2}\exp\left(-\frac{\sum_{i=1}^n(x_i-\overline{x})^2}{2\sigma^2}\right)\end{aligned}$$

which does not depend on μ. By Theorem 6.2.2, the sample mean is a sufficient statistic for μ.

Factorization Theorem

Theorem 6.2.6 - Factorization Theorem

- Let $f_{\mathbf{X}}(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$.
- A statistic $T(\mathbf{X})$ is a sufficient statistic for θ, if and only if
  - There exists function $g(t|\theta)$ and $h(\mathbf{x})$ such that,
  - for all sample points $\mathbf{x}$,
  - and for all parameter points θ,
  - $f_{\mathbf{X}}(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$.

Factorization Theorem : Proof

The proof below is only for discrete distributions.

only if part

- Suppose that $T(\mathbf{X})$ is a sufficient statistic
- Choose $g(t|\theta) = \Pr(T(\mathbf{X}) = t|\theta)$
- and $h(\mathbf{x}) = \Pr(\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = T(\mathbf{x}))$
- Because $T(\mathbf{X})$ is sufficient, $h(\mathbf{x})$ does not depend on θ.

$$\begin{aligned}f_{\mathbf{X}}(\mathbf{x}|\theta) &= \Pr(\mathbf{X}=\mathbf{x}|\theta) \\ &= \Pr(\mathbf{X}=\mathbf{x} \wedge T(\mathbf{X})=T(\mathbf{x})|\theta) \\ &= \Pr(T(\mathbf{X})=T(\mathbf{x})|\theta)\Pr(\mathbf{X}=\mathbf{x}|T(\mathbf{X})=T(\mathbf{x}),\theta) \\ &= \Pr(T(\mathbf{X})=T(\mathbf{x})|\theta)\Pr(\mathbf{X}=\mathbf{x}|T(\mathbf{X})=T(\mathbf{x})) \\ &= g(T(\mathbf{X})|\theta)h(\mathbf{x})\end{aligned}$$

Factorization Theorem : Proof

if part

- Assume that the factorization $f_{\mathbf{X}}(\mathbf{x}|\theta) = g(T(\mathbf{X})|\theta)h(\mathbf{x})$ exists.
- Let $q(t|\theta)$ be the pmf of $T(\mathbf{X})$
- Define $A_t = \{\mathbf{y} : T(\mathbf{y}) = t)\}$.

$$\begin{aligned} q(t|\theta) &= \Pr(T(\mathbf{X}) = t|\theta) \\ &= \sum_{\mathbf{y} \in A_t} f_{\mathbf{X}}(\mathbf{y}|\theta) \end{aligned}$$

Factorization Theorem : Proof

if part (cont'd)

$$\begin{aligned} \frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)} &= \frac{g(T(\mathbf{x})|\theta)h(\mathbf{x})}{q(T(\mathbf{x})|\theta)} = \frac{g(T(\mathbf{x})|\theta)h(\mathbf{x})}{\sum_{\mathbf{y} \in A_{T(\mathbf{x})}} f_{\mathbf{X}}(\mathbf{y}|\theta)} \\ &= \frac{g(T(\mathbf{x})|\theta)h(\mathbf{x})}{\sum_{\mathbf{y} \in A_{T(\mathbf{x})}} g(T(\mathbf{y})|\theta)h(\mathbf{y})} = \frac{g(T(\mathbf{x})|\theta)h(\mathbf{x})}{g(T(\mathbf{x})|\theta)\sum_{A_{\mathbf{y} \in T(\mathbf{x})}} h(\mathbf{y})} \\ &= \frac{h(\mathbf{x})}{\sum_{A_{T(\mathbf{x})}} h(\mathbf{y})} \end{aligned}$$

Thus, $T(\mathbf{X})$ is a sufficient statistic for θ, if and only if $f_{\mathbf{X}}(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$.

Example 6.2.7 - Factorization of Normal Distribution

From Example 6.2.4, we know that

$$f_{\mathbf{X}}(\mathbf{x}|\mu) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^n (x_i - \overline{x})^2 + n(\overline{x} - \mu)^2}{2\sigma^2}\right)$$

We can define $h(\mathbf{x})$, so that it does not depend on μ.

$$h(\mathbf{x}) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^n (x_i - \overline{x})^2}{2\sigma^2}\right)$$

Because $T(\mathbf{X}) = \overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, we have

$$g(t|\mu) = \Pr(T(\mathbf{X}) = t|\mu) = \exp\left(-\frac{n(t-\mu)^2}{2\sigma^2}\right)$$

Then $f_{\mathbf{X}}(\mathbf{x}|\mu) = h(\mathbf{x})g(T(\mathbf{x})|\mu)$ holds, and $T(\mathbf{X}) = \overline{X}$ is a sufficient statistic for μ by the factorization theorem.

Example 6.2.8 - Uniform Sufficient Statistic

Problem

- $X_1, \cdots, X_n$ are iid observations uniformly drawn from $\{1, \cdots, \theta\}$.

$$f_X(x|\theta) = \begin{cases} \frac{1}{\theta} & x = 1, 2, \cdots, \theta \\ 0 & \text{otherwise} \end{cases}$$

- Find a sufficient statistic for θ using factorization theorem.

Example 6.2.8 - Uniform Sufficient Statistic

Joint pmf

The joint pmf of $X_1, \cdots, X_n$ is

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \begin{cases} \theta^{-n} & \mathbf{x} \in \{1, 2, \cdots, \theta\}^n \\ 0 & \text{otherwise} \end{cases}$$

Define $h(\mathbf{x})$

$$h(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \{1, 2, \cdots\}^n \\ 0 & \text{otherwise} \end{cases}$$

Note that $h(\mathbf{x})$ is independent of θ.

Example 6.2.8 - Uniform Sufficient Statistic

Define $T(\mathbf{X})$ and $g(t|\theta)$

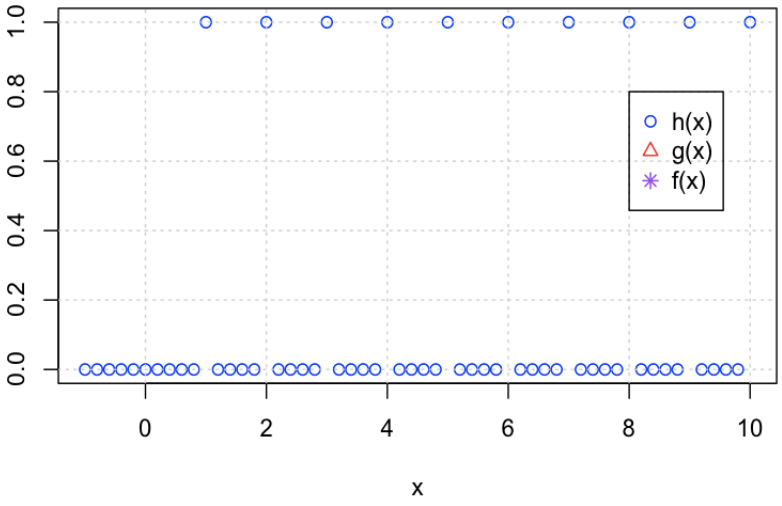
Define $T(\mathbf{X}) = \max_i x_i$, then

$$g(t|\theta) = \Pr(T(\mathbf{x}) = t|\theta) = \Pr(\max_i x_i = t|\theta) = \begin{cases} \theta^{-n} & t \le \theta \\ 0 & \text{otherwise} \end{cases}$$

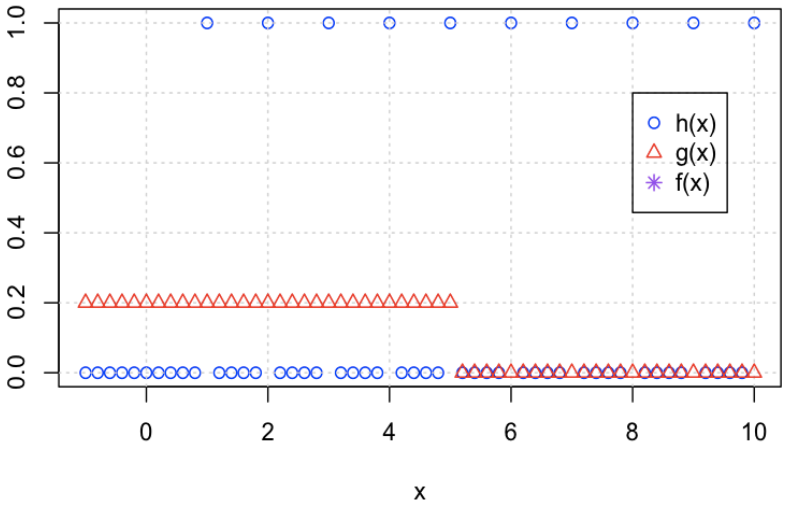
Putting things together

- $f_{\mathbf{X}}(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$ holds.
- Thus, by the factorization theorem, $T(\mathbf{X}) = \max_i X_i$ is a sufficient statistic for θ.

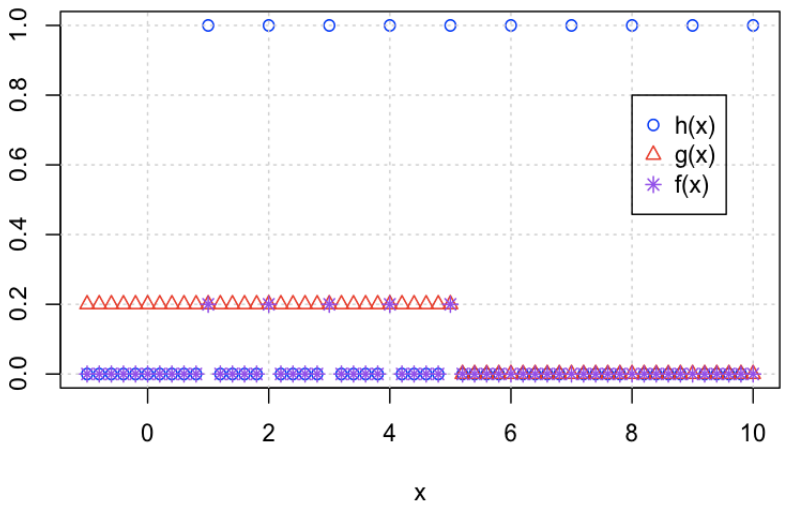
Example of $h(\mathbf{x})$ when $\theta = 5,\ n = 1$



Example of $g(\mathbf{x})$ when $\theta = 5,\ n = 1$



Example of $f(\mathbf{x})$ when $\theta = 5,\ n = 1$



Alternative Solution - Using Indicator Functions

- $I_A(x) = 1$ if $x \in A$, and $I_A(x) = 0$ otherwise.
- $\mathbb{N} = \{1, 2, \cdots\}$, and $\mathbb{N}_\theta = \{1, 2, \cdots, \theta\}$

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \frac{1}{\theta} I_{\mathbb{N}_\theta}(x_i) = \theta^{-n} \prod_{i=1}^{n} I_{\mathbb{N}_\theta}(x_i) \\
\prod_{i=1}^{n} I_{\mathbb{N}_\theta}(x_i) &= \left(\prod_{i=1}^{n} I_{\mathbb{N}}(x_i) \right) I_{\mathbb{N}_\theta}\left[\max_i x_i \right] = \left(\prod_{i=1}^{n} I_{\mathbb{N}}(x_i) \right) I_{\mathbb{N}_\theta}\left[T(\mathbf{x}) \right] \\
f_{\mathbf{X}}(\mathbf{x}|\theta) &= \theta^{-n} I_{\mathbb{N}_\theta}\left[T(\mathbf{x}) \right] \prod_{i=1}^{n} I_{\mathbb{N}}(x_i)
\end{aligned}
$$

$f_{\mathbf{X}}(\mathbf{x}|\theta)$ can be factorized into $g(t|\theta) = \theta^{-n} I_{\mathbb{N}_\theta}(t)$ and $h(\mathbf{x}) = \prod_{i=1}^{n} I_{\mathbb{N}}(x_i)$, and $T(\mathbf{x}) = \max_i x_i$ is a sufficient statistic.

Summary

Today

- Using Theorem 6.2.2 to show a statistic is sufficient
  - Binomial distribution : sum of observations
  - Normal distribution : sample mean
- Factorization Theorem
  - $f_{\mathbf{X}}(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$
  - Necessary and sufficient condition of a sufficient statistic
  - Uniform sufficient statistic : maximum of observations

Next Lecture

- More on Factorization Theorem
- Minimal Sufficient Statistics