

Biostatistics 602 - Statistical Inference

Lecture 02

Factorization Theorem

Hyun Min Kang

January 15th, 2013

Last Lecture - Key Questions

- ① What is the key difference between BIOSTAT601 and BIOSTAT602?
- ② What is the difference between random variable and data?
- ③ What is a statistic?
- ④ What is a sufficient statistic for θ ?
- ⑤ How do we show that a statistic is sufficient for θ ?

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

Recap - 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 .

Recap - Example 6.2.3 - Binomial Sufficient Statistic

Proof

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

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

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

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 (cont'd)

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 (cont'd)

if part (cont'd)

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

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 6.2.8 - Uniform Sufficient Statistic

Problem

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

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

- Find a sufficient statistic for θ using factorization theorem.

Example 6.2.7 - Factorization of Normal Distribution

From Example 6.2.4, we know that

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

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

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

Because $T(\mathbf{X}) = \bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, 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 $T(\mathbf{X}) = \bar{X}$ is a sufficient statistic for μ by the factorization theorem.

Example 6.2.8 - Uniform Sufficient Statistic

Joint pmf

The joint pmf of X_1, \dots, X_n is

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

Define $h(\mathbf{x})$

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

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

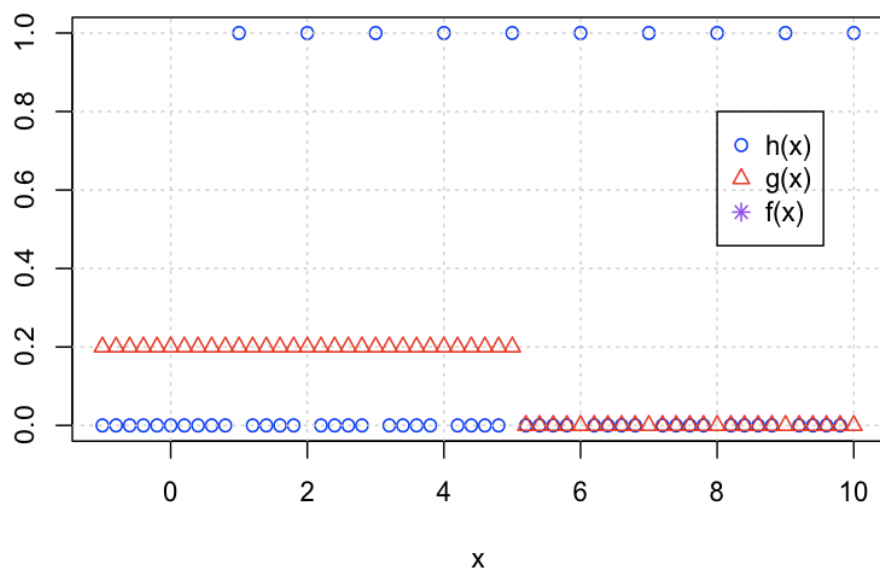
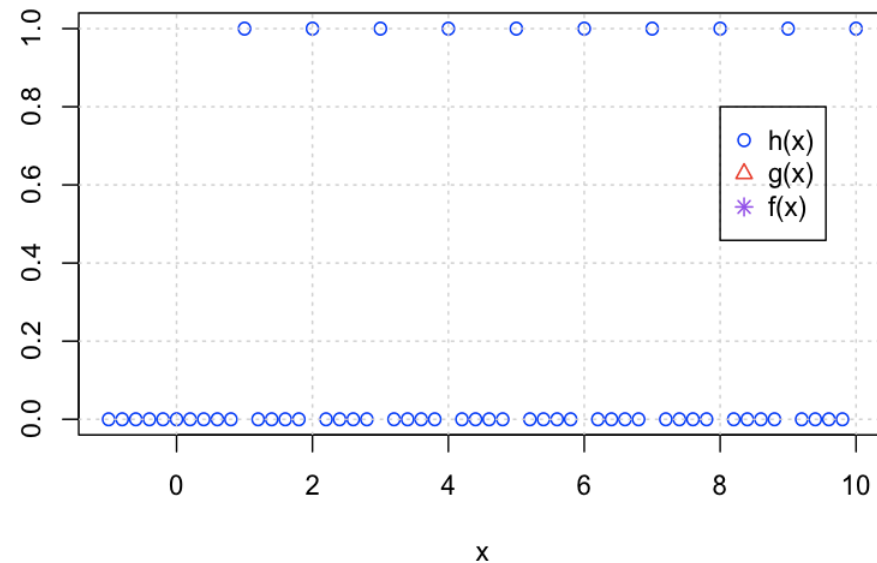
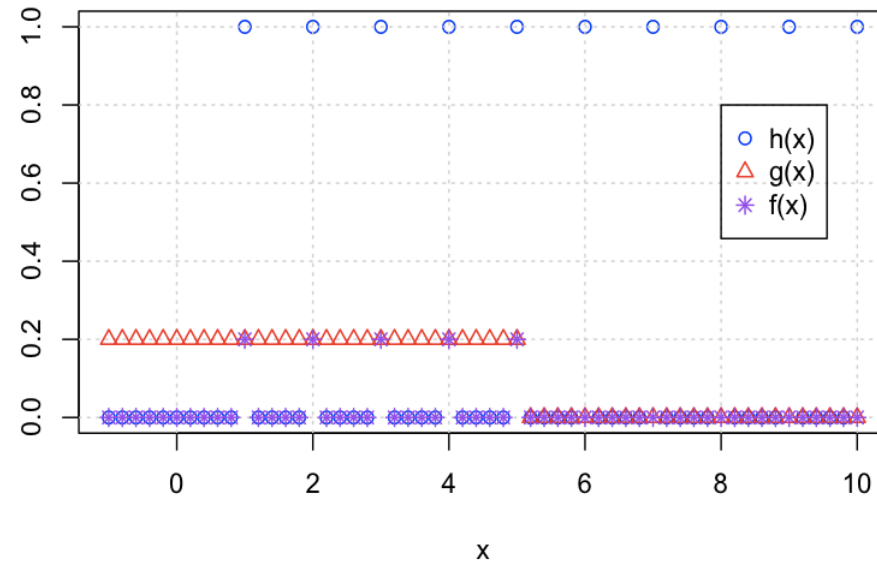
Example 6.2.8 - Uniform Sufficient Statistic

Define $T(\mathbf{X})$ and $g(t|\theta)$ Define $T(\mathbf{X}) = \max_i x_i$, then

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

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 $g(\mathbf{x})$ when $\theta = 5$, $n = 1$ Example of $h(\mathbf{x})$ 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.
- $\mathbb{N} = \{1, 2, \dots\}$, and $\mathbb{N}_\theta = \{1, 2, \dots, \theta\}$

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^n \frac{1}{\theta} I_{\mathbb{N}_\theta}(x_i) = \theta^{-n} \prod_{i=1}^n I_{\mathbb{N}_\theta}(x_i)$$

$$\prod_{i=1}^n I_{\mathbb{N}_\theta}(x_i) = \left(\prod_{i=1}^n I_{\mathbb{N}}(x_i) \right) I_{\mathbb{N}_\theta} \left[\max_i x_i \right] = \left(\prod_{i=1}^n I_{\mathbb{N}}(x_i) \right) I_{\mathbb{N}_\theta} [T(\mathbf{x})]$$

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \theta^{-n} I_{\mathbb{N}_\theta} [T(\mathbf{x})] \prod_{i=1}^n I_{\mathbb{N}}(x_i)$$

$f_{\mathbf{X}}(\mathbf{x}|\theta)$ 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.

Example 6.2.9 - Solution

Decomposing $f_{\mathbf{X}}(\mathbf{x}|\mu, \sigma^2)$ - Similarly to Example 6.2.4

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

Example 6.2.9 - Normal Sufficient Statistic

Problem

- $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$
- Both μ and σ^2 are unknown
- The parameter is a vector : $\boldsymbol{\theta} = (\mu, \sigma^2)$.
- The problem is to use the Factorization Theorem to find the sufficient statistics for $\boldsymbol{\theta}$.

How to solve it

- Propose $\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}))$ as sufficient statistic for μ and σ^2 .
- Use Factorization Theorem to decompose $f_{\mathbf{X}}(\mathbf{x}|\mu, \sigma^2)$.

Example 6.2.9 - Solution

Propose a sufficient statistic

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|\mu, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right) \\ \mathbf{T}(\mathbf{X}) &= (T_1(\mathbf{X}), T_2(\mathbf{X})) \\ T_1(\mathbf{x}) &= \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ T_2(\mathbf{x}) &= \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Example 6.2.9 - Solution

Factorize $f_{\mathbf{X}}(\mathbf{x}|\mu, \sigma^2)$

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

$$h(\mathbf{x}) = 1$$

$$g(t_1, t_2|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} t_2 + \frac{n}{2\sigma^2} (t_1 - \mu)^2\right)$$

$$f_{\mathbf{X}}(\mathbf{x}|\mu, \sigma^2) = g(T_1(\mathbf{x}), T_2(\mathbf{x})|\mu, \sigma^2)h(\mathbf{x})$$

Thus, $\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{x}), T_2(\mathbf{x})) = (\bar{x}, \sum_{i=1}^n (x_i - \bar{x})^2)$ is a sufficient statistic for $\theta = (\mu, \sigma^2)$.

One parameter, two-dimensional sufficient statistic

Factorization

$$h(\mathbf{x}) = 1$$

$$T_1(\mathbf{x}) = \min_i x_i$$

$$T_2(\mathbf{x}) = \max_i x_i$$

$$g(t_1, t_2|\theta) = I(t_1 > \theta \wedge t_2 < \theta + 1)$$

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = I\left(\min_i x_i > \theta \wedge \max_i x_i < \theta + 1\right)$$

$$= g(T_1(\mathbf{x}), T_2(\mathbf{x})|\theta)h(\mathbf{x})$$

Thus, $\mathbf{T}(\mathbf{x}) = (T_1(\mathbf{x}), T_2(\mathbf{x})) = (\min_i x_i, \max_i x_i)$ is a sufficient statistic for θ .

One parameter, two-dimensional sufficient statistic

Problem

Assume $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(\theta, \theta + 1)$, $-\infty < \theta < \infty$. Find a sufficient statistic for θ .

Rewriting $f_{\mathbf{X}}(\mathbf{x}|\theta)$

$$f_X(x|\theta) = \begin{cases} 1 & \text{if } \theta < x < \theta + 1 \\ 0 & \text{otherwise} \end{cases} = I(\theta < x < \theta + 1)$$

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|\theta) &= \prod_{i=1}^n I(\theta < x_i < \theta + 1) \\ &= I(\theta < x_1 < \theta + 1, \dots, \theta < x_n < \theta + 1) \\ &= I\left(\min_i x_i > \theta \wedge \max_i x_i < \theta + 1\right) \end{aligned}$$

Sufficient Order Statistics

Problem

- $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} f_X(x|\theta)$.
- $f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^n f_X(x_i|\theta)$
- Define order statistics $x_{(1)} \leq \dots \leq x_{(n)}$ as an ordered permutation of \mathbf{x}
- Is the order statistic a sufficient statistic for θ ?

$$\begin{aligned} \mathbf{T}(\mathbf{x}) &= (T_1(\mathbf{x}), \dots, T_n(\mathbf{x})) \\ &= (x_{(1)}, \dots, x_{(n)}) \end{aligned}$$

Factorization of Order Statistics

$$\begin{aligned}
 h(\mathbf{x}) &= 1 \\
 g(t_1, \dots, t_n | \theta) &= \prod_{i=1}^n f_X(t_i | \theta) \\
 f_{\mathbf{X}}(\mathbf{x} | \theta) &= g(T_1(\mathbf{x}), \dots, T_n(\mathbf{x}) | \theta) h(\mathbf{x})
 \end{aligned}$$

(Note that $(T_1(\mathbf{x}), \dots, T_n(\mathbf{x}))$ is a permutation of (x_1, \dots, x_n))
Therefore, $\mathbf{T}(\mathbf{x}) = (x_{(1)}, \dots, x_{(n)})$ is a sufficient statistics for θ .

Summary

Today : 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
- Normal distribution : multidimensional sufficient statistic
- One parameter, two dimensional sufficient statistics

Next Lecture

- Minimal Sufficient Statistics

Exercise 6.1

Problem

Let X be one observation from a $\mathcal{N}(0, \sigma^2)$. Is $|X|$ a sufficient statistic for σ^2 ?

Solution

$$f_X(x | \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Define

$$\begin{aligned}
 h(x) &= 1 \\
 T(x) &= |x| \\
 g(t | \theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right)
 \end{aligned}$$