Recap

Factorization Theorem

Summary

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 $\theta$?
5. How do we show that a statistic is sufficient for $\theta$?

Definition 6.2.1

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

Example

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

Theorem 6.2.2

- Let $f_X(x|\theta)$ is a joint pdf or pmf of $X$
- and $q(t|\theta)$ is the pdf or pmf of $T(X)$.
- Then $T(X)$ is a sufficient statistic for $\theta$,
- if, for every $x \in \mathcal{X}$,
- the ratio $f_X(x|\theta)/q(T(x)|\theta)$ is constant as a function of $\theta$. 
Recap - Example 6.2.3 - Binomial Sufficient Statistic

**Proof**

\[
f_X(x|p) = \frac{p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n}}{\binom{n}{t} p^t(1-p)^{n-t}}
\]

\[T(X) \sim \text{Binomial}(n, p)\]

\[
g(t|p) = \binom{n}{t} p^t(1-p)^{n-t}
\]

\[
\frac{f_X(x|p)}{g(T(x)|p)} = \frac{p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n}}{(\sum_{i=1}^n x_i) p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}
\]

By theorem 6.2.2. \(T(X)\) is a sufficient statistic for \(p\).

The proof below is only for discrete distributions.

only if part

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

\[
f_X(x|\theta) = \Pr(X = x|\theta) = \Pr(X = x \land T(X) = T(x)|\theta)
= \Pr(T(X) = T(x)|\theta) \Pr(X = x|T(X) = T(x), \theta)
= \Pr(T(X) = T(x)|\theta) \Pr(X = x|T(X) = T(x))
= g(T(x)|\theta)h(x)
\]

Factorization Theorem : Proof (cont’d)

if part

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

\[
q(t|\theta) = \Pr(T(X) = t|\theta) = \sum_{y \in A_t} f_X(y|\theta)
\]
Recap

Factorization Theorem

Summary

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 $\theta$ using factorization theorem.

Example 6.2.8 - Uniform Sufficient Statistic

Joint pmf

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

\[
  f_X(x|\theta) = \begin{cases} 
    \theta^{-n} & x \in \{1, 2, \cdots, \theta\}^n \\
    0 & \text{otherwise}
  \end{cases}
\]

Define $h(x)$

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

Note that $h(x)$ is independent of $\theta$. 

Example 6.2.7 - Factorization of Normal Distribution

From Example 6.2.4, we know that

\[
  f_X(x|\mu) = (2\pi\sigma^2)^{-n/2} \exp \left( -\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right)
\]

We can define $h(x)$, so that it does not depend on $\mu$.

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

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

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

Then $f_X(x|\mu) = h(x)g(T(x)|\mu)$ holds, and $T(X) = \bar{X}$ is a sufficient statistic for $\mu$ by the factorization theorem.
Example 6.2.8 - Uniform Sufficient Statistic

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

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

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

Putting things together

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

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

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

Example of $f(x)$ when $\theta = 5$, $n = 1$
Alternative Solution - Using Indicator Functions

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

\[
f_X(x|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{\mathbb{N}\theta}(x_i) = \theta^{-n} \prod_{i=1}^{n} I_{\mathbb{N}\theta}(x_i)
\]

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

\[
f_X(x|\theta) = \theta^{-n} I_{\mathbb{N}\theta} [T(x)] \prod_{i=1}^{n} I_{\mathbb{N}}(x_i)
\]

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

Example 6.2.9 - Normal Sufficient Statistic

**Problem**

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

**How to solve it**

- Propose \( T(X) = (T_1(X), T_2(X)) \) as sufficient statistic for \( \mu \) and \( \sigma^2 \).
- Use Factorization Theorem to decompose \( f_X(x|\mu, \sigma^2) \).

Decomposing \( f_X(x|\mu, \sigma^2) \) - Similarly to Example 6.2.4

\[
f_X(x|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{(x_i - \mu)^2}{2\sigma^2} \right)
\]

\[
= (2\pi \sigma^2)^{-n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)
\]

\[
= (2\pi \sigma^2)^{-n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mu)^2 \right)
\]

\[
= (2\pi \sigma^2)^{-n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right)
\]

**Propose a sufficient statistic**

\[
f_X(x|\mu, \sigma^2) = (2\pi \sigma^2)^{-n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right)
\]

\[
T(X) = (T_1(X), T_2(X))
\]

\[
T_1(X) = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
T_2(X) = \sum_{i=1}^{n} (x_i - \bar{x})^2
\]
Example 6.2.9 - Solution

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

\[
\begin{align*}
    f_{X}(x | \mu, \sigma^{2}) &= (2\pi\sigma^{2})^{-n/2} \exp \left( -\frac{1}{2\sigma^{2}} \sum_{i=1}^{n} (x_{i} - \bar{x})^{2} - \frac{n}{2\sigma^{2}} (\bar{x} - \mu)^{2} \right) \\
    h(x) &= 1 \\
    g(t_{1}, t_{2} | \mu, \sigma^{2}) &= (2\pi\sigma^{2})^{-n/2} \exp \left( -\frac{1}{2\sigma^{2}} t_{2} - \frac{n}{2\sigma^{2}} (t_{1} - \mu)^{2} \right) \\
    f_{X}(x | \mu, \sigma^{2}) &= g(T_{1}(x), T_{2}(x) | \mu, \sigma^{2}) h(x)
\end{align*}
\]

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

One parameter, two-dimensional sufficient statistic

Factorization

\[
\begin{align*}
    h(x) &= 1 \\
    T_{1}(x) &= \min_{i} x_{i} \\
    T_{2}(x) &= \max_{i} x_{i} \\
    g(t_{1}, t_{2} | \theta) &= I(t_{1} > \theta \land t_{2} < \theta + 1) \\
    f_{X}(x | \theta) &= I \left( \min_{i} x_{i} > \theta \land \max_{i} x_{i} < \theta + 1 \right) \\
    f_{X}(x | \theta) &= g(T_{1}(x), T_{2}(x) | \theta) h(x)
\end{align*}
\]

Thus, $T(X) = (T_{1}(x), T_{2}(x)) = (\min_{i} x_{i}, \max_{i} x_{i})$ is a sufficient statistic for $\theta$.

One parameter, two-dimensional sufficient statistic

Problem

Assume $X_{1}, \cdots, X_{n}$ i.i.d. Uniform($\theta, \theta + 1$), $-\infty < \theta < \infty$. Find a sufficient statistic for $\theta$.

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

\[
\begin{align*}
    f_{X}(x | \theta) &= \begin{cases} 
        1 & \text{if } \theta < x < \theta + 1 \\
        0 & \text{otherwise}
    \end{cases} \\
    &= I(\theta < x_{1} < \theta + 1) \\
    &= I(\theta < x_{1} < \theta + 1, \cdots, \theta < x_{n} < \theta + 1) \\
    &= I \left( \min_{i} x_{i} > \theta \land \max_{i} x_{i} < \theta + 1 \right)
\end{align*}
\]

Sufficient Order Statistics

Problem

- $X_{1}, \cdots, X_{n}$ i.i.d. $f_{X}(x | \theta)$.
- $f_{X}(x | \theta) = \prod_{i=1}^{n} f_{X}(x_{i} | \theta)$
- Define order statistics $x_{(1)} \leq \cdots \leq x_{(n)}$ as an ordered permutation of $x$
- Is the order statistic a sufficient statistic for $\theta$?

Thus, $T(X) = (T_{1}(x), \cdots, T_{n}(x)) = (x_{(1)}, \cdots, x_{(n)})$
Factorization of Order Statistics

\[ h(x) = 1 \]
\[ g(t_1, \ldots, t_n|\theta) = \prod_{i=1}^{n} f_X(t_i|\theta) \]
\[ f_X(x|\theta) = g(T_1(x), \ldots, T_n(x)|\theta)h(x) \]

(Note that \((T_1(x), \ldots, T_n(x))\) is a permutation of \((x_1, \ldots, x_n))\)

Therefore, \(T(x) = (x(1), \ldots, x(n))\) is a sufficient statistics for \(\theta\).

Exercise 6.1

**Problem**

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

**Solution**

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

Define

\[ h(x) = 1 \]
\[ T(x) = |x| \]
\[ g(t|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \]

Then \(f_X(x|\theta) = g(T(x)|\theta)h(x)\) holds, and \(T(X) = |X|\) is a sufficient statistic by the Factorization Theorem.

Summary

**Today : Factorization Theorem**

- \(f_X(x|\theta) = g(T(x)|\theta)h(x)\)
- 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