Biostatistics 602 - Statistical Inference
Lecture 04
Ancillary Statistics

Hyun Min Kang

January 22th, 2013

Recap from last lecture

1. Is a sufficient statistic unique?

Recap from last lecture

1. Is a sufficient statistic unique?
2. What are examples obvious sufficient statistics for any distribution?

Recap from last lecture

1. Is a sufficient statistic unique?
2. What are examples obvious sufficient statistics for any distribution?
3. What is a minimal sufficient statistic?

Recap from last lecture

1. Is a sufficient statistic unique?
2. What are examples obvious sufficient statistics for any distribution?
3. What is a minimal sufficient statistic?
4. Is a minimal sufficient statistic unique?

Recap from last lecture

1. Is a sufficient statistic unique?
2. What are examples obvious sufficient statistics for any distribution?
3. What is a minimal sufficient statistic?
4. Is a minimal sufficient statistic unique?
5. How can we show that a statistic is minimal sufficient for θ?

Minimal Sufficient Statistic

Definition 6.2.11

A sufficient statistic $T(\mathbf{X})$ is called a *minimal sufficient statistic* if, for any other sufficient statistic $T'(\mathbf{X})$, $T(\mathbf{X})$ is a function of $T'(\mathbf{X})$.

Why is this called "minimal" sufficient statistic?

- The sample space $\mathcal{X}$ consists of every possible sample - finest partition
- Given $T(\mathbf{X})$, $\mathcal{X}$ can be partitioned into A_t where $t \in \mathcal{T} = \{t : t = T(\mathbf{X}) \text{ for some } \mathbf{x} \in \mathcal{X}\}$
- Maximum data reduction is achieved when $|\mathcal{T}|$ is minimal.
- If size of $\mathcal{T}' = t : t = T'(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathcal{X}$ is not less than $|\mathcal{T}|$, then $|\mathcal{T}|$ can be called as a minimal sufficient statistic.

Theorem for Minimal Sufficient Statistics

Theorem 6.2.13

- $f_{\mathbf{X}}(\mathbf{x})$ be pmf or pdf of a sample $\mathbf{X}$.
- Suppose that there exists a function $T(\mathbf{x})$ such that,
- For every two sample points $\mathbf{x}$ and $\mathbf{y}$,
- The ratio $f_{\mathbf{X}}(\mathbf{x}|\theta)/f_{\mathbf{X}}(\mathbf{y}|\theta)$ is constant as a function of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$.
- Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.

In other words..

- $f_{\mathbf{X}}(\mathbf{x}|\theta)/f_{\mathbf{X}}(\mathbf{y}|\theta)$ is constant as a function of $\theta \implies T(\mathbf{x}) = T(\mathbf{y})$.
- $T(\mathbf{x}) = T(\mathbf{y}) \implies f_{\mathbf{X}}(\mathbf{x}|\theta)/f_{\mathbf{X}}(\mathbf{y}|\theta)$ is constant as a function of θ

Exercise from the textbook

Problem

$X_1, \cdots, X_n$ are iid samples from

$$f_X(x|\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}, -\infty < x < \infty, -\infty < \theta < \infty$$

Find a minimal sufficient statistic for θ.

Exercise from the textbook

Problem

$X_1, \cdots, X_n$ are iid samples from

$$f_X(x|\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}, -\infty < x < \infty, -\infty < \theta < \infty$$

Find a minimal sufficient statistic for θ.

Solution

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^{n} \frac{\exp(-(x_i-\theta))}{(1+\exp(-(x_i-\theta)))^2}$$

Exercise from the textbook

Problem

$X_1, \cdots, X_n$ are iid samples from

$$f_X(x|\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}, -\infty < x < \infty, -\infty < \theta < \infty$$

Find a minimal sufficient statistic for θ.

Solution

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^{n} \frac{\exp(-(x_i-\theta))}{(1+\exp(-(x_i-\theta)))^2} = \frac{\exp\left(-\sum_{i=1}^n (x_i-\theta)\right)}{\prod_{i=1}^n (1+\exp(-(x_i-\theta)))^2}$$

Exercise from the textbook

Problem

$X_1, \cdots, X_n$ are iid samples from

$$f_X(x|\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2}, -\infty < x < \infty, -\infty < \theta < \infty$$

Find a minimal sufficient statistic for θ.

Solution

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

Solution (cont'd)

Applying Theorem 6.2.13

$$\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} = \frac{\exp\left(-\sum_{i=1}^n x_i\right)\exp(n\theta)\prod_{i=1}^n(1+\exp(-(y_i-\theta)))^2}{\exp\left(-\sum_{i=1}^n y_i\right)\exp(n\theta)\prod_{i=1}^n(1+\exp(-(x_i-\theta)))^2}$$

Solution (cont'd)

Applying Theorem 6.2.13

$$\begin{aligned}
\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} &= \frac{\exp\left(-\sum_{i=1}^n x_i\right)\exp(n\theta)\prod_{i=1}^n(1+\exp(-(y_i-\theta)))^2}{\exp\left(-\sum_{i=1}^n y_i\right)\exp(n\theta)\prod_{i=1}^n(1+\exp(-(x_i-\theta)))^2} \\
&= \frac{\exp\left(-\sum_{i=1}^n x_i\right)\prod_{i=1}^n(1+\exp(-(y_i-\theta)))^2}{\exp\left(-\sum_{i=1}^n y_i\right)\prod_{i=1}^n(1+\exp(-(x_i-\theta)))^2}
\end{aligned}$$

Solution (cont'd)

Applying Theorem 6.2.13

$$\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} = \frac{\exp\left(-\sum_{i=1}^n x_i\right)\exp(n\theta)\prod_{i=1}^n(1+\exp(-(y_i-\theta)))^2}{\exp\left(-\sum_{i=1}^n y_i\right)\exp(n\theta)\prod_{i=1}^n(1+\exp(-(x_i-\theta)))^2}$$

$$= \frac{\exp\left(-\sum_{i=1}^n x_i\right)\prod_{i=1}^n(1+\exp(-(y_i-\theta)))^2}{\exp\left(-\sum_{i=1}^n y_i\right)\prod_{i=1}^n(1+\exp(-(x_i-\theta)))^2}$$

The ratio above is constant to θ if and only if $x_1, \cdots, x_n$ are permutations of $y_1, \cdots, y_n$. So the order statistic $\mathbf{T}(\mathbf{X}) = (X_{(1)}, \cdots, X_{(n)})$ is a minimal sufficient statistic.

Ancillary Statistics

Definition 6.2.16

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Ancillary Statistics

Definition 6.2.16

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Examples of Ancillary Statistics

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known.

Ancillary Statistics

Definition 6.2.16

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Examples of Ancillary Statistics

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known.

- $s_{\mathbf{X}}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ is an ancillary statistic

Ancillary Statistics

Definition 6.2.16

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Examples of Ancillary Statistics

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known.

- $s^2_{\mathbf{X}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is an ancillary statistic
- $X_1 - X_2 \sim \mathcal{N}(0, 2\sigma^2)$ is ancillary.

Ancillary Statistics

Definition 6.2.16

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Examples of Ancillary Statistics

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known.

- $s^2_{\mathbf{X}} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ is an ancillary statistic
- $X_1 - X_2 \sim \mathcal{N}(0, 2\sigma^2)$ is ancillary.
- $(X_1 + X_2)/2 - X_3 \sim \mathcal{N}(0, 1.5\sigma^2)$ is ancillary.

Ancillary Statistics

Definition 6.2.16

A statistic $S(\mathbf{X})$ is an *ancillary statistic* if its distribution does not depend on θ.

Examples of Ancillary Statistics

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known.

- $s^2_{\mathbf{X}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is an ancillary statistic
- $X_1 - X_2 \sim \mathcal{N}(0, 2\sigma^2)$ is ancillary.
- $(X_1 + X_2)/2 - X_3 \sim \mathcal{N}(0, 1.5\sigma^2)$ is ancillary.
- $\frac{(n-1)s^2_{\mathbf{X}}}{\sigma^2} \sim \chi^2_{n-1}$ is ancillary.

More Examples of Ancillary Statistics

Examples with normal distribution at zero mean

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown

More Examples of Ancillary Statistics

Examples with normal distribution at zero mean

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown

- $\mathbf{Y} = \mathbf{X}/\sigma$ is an ancillary statistic because $Y_i \sim \mathcal{N}(0, 1)$.

More Examples of Ancillary Statistics

Examples with normal distribution at zero mean

$X_1, \cdots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown

- $\mathbf{Y} = \mathbf{X}/\sigma$ is an ancillary statistic because $Y_i \sim \mathcal{N}(0, 1)$.
- $\frac{X_1}{X_2} = \frac{\sigma Y_1}{\sigma Y_2} = \frac{Y_1}{Y_2}$ also follows a cauchy distribution and is an ancillary statistic.

More Examples of Ancillary Statistics

Examples with normal distribution at zero mean

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown

- $\mathbf{Y} = \mathbf{X}/\sigma$ is an ancillary statistic because $Y_i \sim \mathcal{N}(0,1)$.
- $\frac{X_1}{X_2} = \frac{\sigma Y_1}{\sigma Y_2} = \frac{Y_1}{Y_2}$ also follows a cauchy distribution and is an ancillary statistic.
- Any joint distribution of $Y_1, \cdots, Y_n$ does not depend on σ^2, and thus is an ancillary statistic.

More Examples of Ancillary Statistics

Examples with normal distribution at zero mean

$X_1, \cdots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown

- $\mathbf{Y} = \mathbf{X}/\sigma$ is an ancillary statistic because $Y_i \sim \mathcal{N}(0, 1)$.
- $\frac{X_1}{X_2} = \frac{\sigma Y_1}{\sigma Y_2} = \frac{Y_1}{Y_2}$ also follows a cauchy distribution and is an ancillary statistic.
- Any joint distribution of $Y_1, \cdots, Y_n$ does not depend on σ^2, and thus is an ancillary statistic.
- For example, the following statistic is also ancillary.

$$\frac{\text{median}(X_i)}{\overline{X}}$$

More Examples of Ancillary Statistics

Examples with normal distribution at zero mean

$X_1, \cdots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown

- $\mathbf{Y} = \mathbf{X}/\sigma$ is an ancillary statistic because $Y_i \sim \mathcal{N}(0,1)$.
- $\frac{X_1}{X_2} = \frac{\sigma Y_1}{\sigma Y_2} = \frac{Y_1}{Y_2}$ also follows a cauchy distribution and is an ancillary statistic.
- Any joint distribution of $Y_1, \cdots, Y_n$ does not depend on σ^2, and thus is an ancillary statistic.
- For example, the following statistic is also ancillary.

$$\frac{\text{median}(X_i)}{\overline{X}} = \frac{\sigma \text{median}(Y_i)}{\sigma \overline{Y}}$$

More Examples of Ancillary Statistics

Examples with normal distribution at zero mean

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ where σ^2 is unknown

- $\mathbf{Y} = \mathbf{X}/\sigma$ is an ancillary statistic because $Y_i \sim \mathcal{N}(0, 1)$.
- $\frac{X_1}{X_2} = \frac{\sigma Y_1}{\sigma Y_2} = \frac{Y_1}{Y_2}$ also follows a cauchy distribution and is an ancillary statistic.
- Any joint distribution of $Y_1, \cdots, Y_n$ does not depend on σ^2, and thus is an ancillary statistic.
- For example, the following statistic is also ancillary.

$$\frac{\text{median}(X_i)}{\overline{X}} = \frac{\sigma \text{median}(Y_i)}{\sigma \overline{Y}} = \frac{\text{median}(Y_i)}{\overline{Y}}$$

Range Statistics

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x - \theta)$.
- Show that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Range Statistics

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x - \theta)$.
- Show that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

- Let $Z_i = X_i - \theta$.

Range Statistics

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x - \theta)$.
- Show that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

- Let $Z_i = X_i - \theta$.
- $f_Z(z) = f_X(z + \theta - \theta) \left| \frac{dx}{dz} \right| = f_X(z)$

Range Statistics

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x-\theta)$.
- Show that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

- Let $Z_i = X_i - \theta$.
- $f_Z(z) = f_X(z+\theta-\theta)\left|\frac{dx}{dz}\right| = f_X(z)$
- $Z_1, \cdots, Z_n \overset{\text{i.i.d.}}{\sim} f_X(z)$ does not depend on θ.

Range Statistics

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x-\theta)$.
- Show that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

- Let $Z_i = X_i - \theta$.
- $f_Z(z) = f_X(z+\theta-\theta)\left|\frac{dx}{dz}\right| = f_X(z)$
- $Z_1, \cdots, Z_n \overset{\text{i.i.d.}}{\sim} f_X(z)$ does not depend on θ.
- $R = X_{(n)} - X_{(1)} = Z_{(n)} - Z_{(1)}$ does not depend on θ.

Uniform Ancillary Statistics

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(\theta, \theta + 1)$.
- Show that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Possible Strategies

Uniform Ancillary Statistics

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(\theta, \theta + 1)$.
- Show that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Possible Strategies

- Obtain the distribution of R and show that it is independent of θ.

Uniform Ancillary Statistics

Problem

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(\theta, \theta + 1)$.
- Show that $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Possible Strategies

- Obtain the distribution of R and show that it is independent of θ.
- Represent R as a function of ancillary statistics, which is independent of θ.

Proof : Method I - 1/4

R is a function of $(X_{(n)}, X_{(1)})$, so we need to derive the joint distribution of $(X_{(n)}, X_{(1)})$.

Proof : Method I - 1/4

R is a function of $(X_{(n)}, X_{(1)})$, so we need to derive the joint distribution of $(X_{(n)}, X_{(1)})$. Define

$$f_X(x|\theta) = I(\theta < x < \theta + 1)$$

Proof : Method I - 1/4

R is a function of $(X_{(n)}, X_{(1)})$, so we need to derive the joint distribution of $(X_{(n)}, X_{(1)})$. Define

$$f_X(x|\theta) = I(\theta < x < \theta + 1)$$

If $\theta < X_{(1)} \leq X_{(n)} < \theta + 1$,

$$f_{\mathbf{X}}(X_{(1)}, X_{(n)}|\theta) =$$

Proof : Method I - 1/4

R is a function of $(X_{(n)}, X_{(1)})$, so we need to derive the joint distribution of $(X_{(n)}, X_{(1)})$. Define

$$f_X(x|\theta) = I(\theta < x < \theta + 1)$$

If $\theta < X_{(1)} \leq X_{(n)} < \theta + 1$,

$$f_{\mathbf{X}}(X_{(1)}, X_{(n)}|\theta) = \frac{n!}{(n-2)!} \left(X_{(n)} - X_{(1)}\right)^{(n-2)}$$

Proof : Method I - 1/4

R is a function of $(X_{(n)}, X_{(1)})$, so we need to derive the joint distribution of $(X_{(n)}, X_{(1)})$. Define

$$f_X(x|\theta) = I(\theta < x < \theta + 1)$$

If $\theta < X_{(1)} \leq X_{(n)} < \theta + 1$,

$$f_{\mathbf{X}}(X_{(1)}, X_{(n)}|\theta) = \frac{n!}{(n-2)!}\left(X_{(n)} - X_{(1)}\right)^{(n-2)}$$

and $f_{\mathbf{X}}(X_{(1)}, X_{(n)}|\theta) = 0$ otherwise.

Proof : Method I - 2/4

Define R and M as follows

$$\left\{ \begin{array}{lcl} R & = & X_{(n)} - X_{(1)} \\ M & = & (X_{(n)} + X_{(1)})/2 \end{array} \right.$$

Proof : Method I - 2/4

Define R and M as follows

$$\begin{cases} R &= X_{(n)} - X_{(1)} \\ M &= (X_{(n)} + X_{(1)})/2 \end{cases}$$

Then

$$\begin{cases} X_{(1)} &= M - R/2 \\ X_{(n)} &= M + R/2 \end{cases}$$

Proof : Method I - 2/4

Define R and M as follows

$$\begin{cases} R &= X_{(n)} - X_{(1)} \\ M &= (X_{(n)} + X_{(1)})/2 \end{cases}$$

Then

$$\begin{cases} X_{(1)} &= M - R/2 \\ X_{(n)} &= M + R/2 \end{cases}$$

The Jacobian is

$$J = \begin{vmatrix} \frac{\partial X_{(1)}}{\partial M} & \frac{\partial X_{(1)}}{\partial R} \\ \frac{\partial X_{(n)}}{\partial M} & \frac{\partial X_{(n)}}{\partial R} \end{vmatrix} = \begin{vmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{vmatrix} = \frac{1}{2} - (-\frac{1}{2}) = 1$$

Proof : Method I - 3/4

The joint distribution of R and M is

$$f_{R,M}(r,m) = n(n-1)\left(\frac{2m+r}{2} - \frac{2m-r}{2}\right)^{(n-2)} = n(n-1)r^{(n-2)}$$

Proof : Method I - 3/4

The joint distribution of R and M is

$$f_{R,M}(r,m) = n(n-1)\left(\frac{2m+r}{2} - \frac{2m-r}{2}\right)^{(n-2)} = n(n-1)r^{(n-2)}$$

Because $\theta < X_{(1)} \leq X_{(n)} < \theta + 1$,

$$\theta < \frac{2m-r}{2} < \frac{2m+r}{2} < \theta + 1$$

Proof : Method I - 3/4

The joint distribution of R and M is

$$f_{R,M}(r,m) \quad = \quad n(n-1)\left(\frac{2m+r}{2} - \frac{2m-r}{2}\right)^{(n-2)} = n(n-1)r^{(n-2)}$$

Because $\theta < X_{(1)} \leq X_{(n)} < \theta + 1$,

$$\theta < \frac{2m-r}{2} < \frac{2m+r}{2} < \theta + 1$$

$$0 < r < 1$$

$$\theta + \frac{r}{2} < m < \theta + 1 - \frac{r}{2}$$

Proof : Method I - 4/4

The distribution of R is

$$f_R(r|\theta) = \int_{\theta+\frac{r}{2}}^{\theta+1-\frac{r}{2}} n(n-1)r^{(n-2)}dm$$

Proof : Method I - 4/4

The distribution of R is

$$
\begin{aligned}
f_R(r|\theta) &= \int_{\theta+\frac{r}{2}}^{\theta+1-\frac{r}{2}} n(n-1)r^{(n-2)}dm \\
&= n(n-1)r^{(n-2)}\left(\theta+1-\frac{r}{2}-\theta-\frac{r}{2}\right)
\end{aligned}
$$

Proof : Method I - 4/4

The distribution of R is

$$\begin{aligned}
f_R(r|\theta) &= \int_{\theta+\frac{r}{2}}^{\theta+1-\frac{r}{2}} n(n-1)r^{(n-2)}dm \\
&= n(n-1)r^{(n-2)}\left(\theta+1-\frac{r}{2}-\theta-\frac{r}{2}\right) \\
&= n(n-1)r^{(n-2)}(1-r) \qquad ,0<r<1
\end{aligned}$$

Proof : Method I - 4/4

The distribution of R is

$$\begin{aligned} f_R(r|\theta) &= \int_{\theta+\frac{r}{2}}^{\theta+1-\frac{r}{2}} n(n-1)r^{(n-2)}dm \\ &= n(n-1)r^{(n-2)}\left(\theta+1-\frac{r}{2}-\theta-\frac{r}{2}\right) \\ &= n(n-1)r^{(n-2)}(1-r) \qquad ,0<r<1 \end{aligned}$$

Therefore, $f_R(r|\theta)$ does not depend on θ, and R is an ancillary statistic.

Method II : Probably A Simpler Proof

$$f_X(x|\theta) \quad = \quad I(\theta < x < \theta + 1) = I(0 < x - \theta < 1)$$

Method II : Probably A Simpler Proof

$$f_X(x|\theta) = I(\theta < x < \theta + 1) = I(0 < x - \theta < 1)$$

Let $Y_i = X_i - \theta \sim \text{Uniform}(0,1)$. Then $X_i = Y_i + \theta$, $|\frac{dx}{dy}| = 1$ holds.

Method II : Probably A Simpler Proof

$$f_X(x|\theta) \quad = \quad I(\theta < x < \theta + 1) = I(0 < x - \theta < 1)$$

Let $Y_i = X_i - \theta \sim \text{Uniform}(0, 1)$. Then $X_i = Y_i + \theta$, $|\frac{dx}{dy}| = 1$ holds.

$$f_Y(y) \quad = \quad I(0 < y + \theta - \theta < 1)|\frac{dx}{dy}| = I(0 < y < 1)$$

Method II : Probably A Simpler Proof

$$f_X(x|\theta) = I(\theta < x < \theta + 1) = I(0 < x - \theta < 1)$$

Let $Y_i = X_i - \theta \sim \text{Uniform}(0, 1)$. Then $X_i = Y_i + \theta$, $|\frac{dx}{dy}| = 1$ holds.

$$f_Y(y) = I(0 < y + \theta - \theta < 1)|\frac{dx}{dy}| = I(0 < y < 1)$$

Then, the range statistic R can be rewritten as follows.

$$R = X_{(n)} - X_{(1)} = (Y_{(n)} + \theta) - (Y_{(1)} + \theta) = Y_{(n)} - Y_{(1)}$$

Method II : Probably A Simpler Proof

$$f_X(x|\theta) = I(\theta < x < \theta + 1) = I(0 < x - \theta < 1)$$

Let $Y_i = X_i - \theta \sim \text{Uniform}(0, 1)$. Then $X_i = Y_i + \theta$, $|\frac{dx}{dy}| = 1$ holds.

$$f_Y(y) = I(0 < y + \theta - \theta < 1)|\frac{dx}{dy}| = I(0 < y < 1)$$

Then, the range statistic R can be rewritten as follows.

$$R = X_{(n)} - X_{(1)} = (Y_{(n)} + \theta) - (Y_{(1)} + \theta) = Y_{(n)} - Y_{(1)}$$

As $Y_{(n)} - Y_{(1)}$ is a function of $Y_1, \cdots, Y_n$. Any joint distribution of $Y_1, \cdots, Y_n$ does not depend on θ. Therefore, R is an ancillary statistic.

A brief review on location and scale family

Theorem 3.5.1

Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constant, then,

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf.

A brief review on location and scale family

Theorem 3.5.1

Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constant, then,

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf.

Proof

Because $f(x)$ is a pdf, then $f(x) \geq 0$, and $g(x|\mu, \sigma) \geq 0$ for all x.

A brief review on location and scale family

Theorem 3.5.1

Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constant, then,

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

is a pdf.

Proof

Because $f(x)$ is a pdf, then $f(x) \geq 0$, and $g(x|\mu, \sigma) \geq 0$ for all x.
Let $y = (x - \mu)/\sigma$, then $x = \sigma y + \mu$, and $dx/dy = \sigma$.

A brief review on location and scale family

Theorem 3.5.1

Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constant, then,

$$g(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf.

Proof

Because $f(x)$ is a pdf, then $f(x) \geq 0$, and $g(x|\mu,\sigma) \geq 0$ for all x.
Let $y = (x-\mu)/\sigma$, then $x = \sigma y + \mu$, and $dx/dy = \sigma$.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma} f(y) \sigma \, dy = \int_{-\infty}^{\infty} f(y) \, dy = 1$$

A brief review on location and scale family

Theorem 3.5.1

Let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constant, then,

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf.

Proof

Because $f(x)$ is a pdf, then $f(x) \geq 0$, and $g(x|\mu,\sigma) \geq 0$ for all x.
Let $y = (x-\mu)/\sigma$, then $x = \sigma y + \mu$, and $dx/dy = \sigma$.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx \quad = \quad \int_{-\infty}^{\infty} \frac{1}{\sigma} f(y) \sigma \, dy = \int_{-\infty}^{\infty} f(y) \, dy = 1$$

Therefore, $g(x|\mu,\sigma)$ is also a pdf.

Location Family and Parameter

Definition 3.5.2

Let $f(x)$ be any pdf. Then the family of pdfs $f(x - \mu)$, indexed by the parameter $-\infty < \mu < \infty$, is called the *location family with standard pdf* $f(x)$, and μ is called the *location parameter* for the family.

Location Family and Parameter

Definition 3.5.2

Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\mu)$, indexed by the parameter $-\infty < \mu < \infty$, is called the *location family with standard pdf* $f(x)$, and μ is called the *location parameter* for the family.

Example

- $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \sim \mathcal{N}(0,1)$
- $f(x-\mu) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} \sim \mathcal{N}(\mu,1)$

Location Family and Parameter

Definition 3.5.2

Let $f(x)$ be any pdf. Then the family of pdfs $f(x - \mu)$, indexed by the parameter $-\infty < \mu < \infty$, is called the *location family with standard pdf* $f(x)$, and μ is called the *location parameter* for the family.

Example

- $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \sim \mathcal{N}(0, 1)$
- $f(x - \mu) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} \sim \mathcal{N}(\mu, 1)$
- $f(x) = I(0 < x < 1) \sim \mathrm{Uniform}(0, 1)$
- $f(x - \theta) = I(\theta < x < \theta + 1) \sim \mathrm{Uniform}(\theta, \theta + 1)$

Scale Family and Parameter

Definition 3.5.4

Let $f(x)$ be any pdf. Then for any $\sigma > 0$ the family of pdfs $f(x/\sigma)/\sigma$, indexed by the parameter σ is called the *scale family with standard pdf* $f(x)$, and σ is called the *scale parameter* for the family.

Scale Family and Parameter

Definition 3.5.4

Let $f(x)$ be any pdf. Then for any $\sigma > 0$ the family of pdfs $f(x/\sigma)/\sigma$, indexed by the parameter σ is called the *scale family with standard pdf* $f(x)$, and σ is called the *scale parameter* for the family.

Example

- $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \sim \mathcal{N}(0, 1)$
- $f(x/\sigma)/\sigma = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \sim \mathcal{N}(0, \sigma^2)$

Location-Scale Family and Parameters

Definition 3.5.5

Let $f(x)$ be any pdf. Then for any $\mu, -\infty < \mu < \infty$, and any $\sigma > 0$ the family of pdfs $f((x-\mu)/\sigma)/\sigma$, indexed by the parameter (μ, σ) is called the *location-scale family with standard pdf* $f(x)$, and μ is called the *location parameter* and σ is called the *scale parameter* for the family.

Location-Scale Family and Parameters

Definition 3.5.5

Let $f(x)$ be any pdf. Then for any $\mu, -\infty < \mu < \infty$, and any $\sigma > 0$ the family of pdfs $f((x-\mu)/\sigma)/\sigma$, indexed by the parameter (μ, σ) is called the *location-scale family with standard pdf* $f(x)$, and μ is called the *location parameter* and σ is called the *scale parameter* for the family.

Example

- $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \sim \mathcal{N}(0,1)$
- $f((x-\mu)/\sigma)/\sigma = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \sim \mathcal{N}(\mu, \sigma^2)$

Theorem for location and scale family

Theorem 3.5.6

- Let $f(\cdot)$ be any pdf.
- Let μ be any real number.
- Let σ be any positive real number.
- Then X is a random variable with pdf $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$
- if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.

Ancillary Statistics for Location Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x - \mu)$ where $-\infty < \mu < \infty$. Show that the range $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Ancillary Statistics for Location Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x-\mu)$ where $-\infty < \mu < \infty$. Show that the range $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

Assume that cdf is $F(x-\mu)$. Using Theorem 3.5.6, $Z_1 = X_1 - \mu, \cdots, Z_n = X_n - \mu$ are iid observations from pdf $f(x)$ and cdf $F(x)$.

Ancillary Statistics for Location Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x - \mu)$ where $-\infty < \mu < \infty$. Show that the range $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

Assume that cdf is $F(x - \mu)$. Using Theorem 3.5.6, $Z_1 = X_1 - \mu, \cdots, Z_n = X_n - \mu$ are iid observations from pdf $f(x)$ and cdf $F(x)$. Then the cdf of the range statistic R becomes

Ancillary Statistics for Location Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x-\mu)$ where $-\infty < \mu < \infty$. Show that the range $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

Assume that cdf is $F(x-\mu)$. Using Theorem 3.5.6, $Z_1 = X_1 - \mu, \cdots, Z_n = X_n - \mu$ are iid observations from pdf $f(x)$ and cdf $F(x)$. Then the cdf of the range statistic R becomes

$$F_R(r|\mu) \quad = \quad \Pr(R \leq r|\mu) = \Pr(X_{(n)} - X_{(1)} \leq r|\mu)$$

Ancillary Statistics for Location Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x-\mu)$ where $-\infty < \mu < \infty$. Show that the range $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

Assume that cdf is $F(x-\mu)$. Using Theorem 3.5.6, $Z_1 = X_1 - \mu, \cdots, Z_n = X_n - \mu$ are iid observations from pdf $f(x)$ and cdf $F(x)$. Then the cdf of the range statistic R becomes

$$\begin{aligned} F_R(r|\mu) &= \Pr(R \leq r|\mu) = \Pr(X_{(n)} - X_{(1)} \leq r|\mu) \\ &= \Pr(Z_{(n)} + \mu - Z_{(1)} - \mu \leq r|\mu) = \Pr(Z_{(n)} - Z_{(1)} \leq r|\mu) \end{aligned}$$

Ancillary Statistics for Location Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x - \mu)$ where $-\infty < \mu < \infty$. Show that the range $R = X_{(n)} - X_{(1)}$ is an ancillary statistic.

Solution

Assume that cdf is $F(x - \mu)$. Using Theorem 3.5.6, $Z_1 = X_1 - \mu, \cdots, Z_n = X_n - \mu$ are iid observations from pdf $f(x)$ and cdf $F(x)$. Then the cdf of the range statistic R becomes

$$\begin{aligned} F_R(r|\mu) &= \Pr(R \leq r|\mu) = \Pr(X_{(n)} - X_{(1)} \leq r|\mu) \\ &= \Pr(Z_{(n)} + \mu - Z_{(1)} - \mu \leq r|\mu) = \Pr(Z_{(n)} - Z_{(1)} \leq r|\mu) \end{aligned}$$

which does not depend on μ because $Z_1, \cdots, Z_n$ does not depend on μ. Therefore, R is an ancillary statistic.

Ancillary Statistics for Scale Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x/\sigma)/\sigma$ where $\sigma > 0$. Show that the following statistic $\mathbf{T}(\mathbf{X})$ is ancillary.

$$\mathbf{T}(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n)$$

Ancillary Statistics for Scale Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x/\sigma)/\sigma$ where $\sigma > 0$. Show that the following statistic $\mathbf{T}(\mathbf{X})$ is ancillary.

$$\mathbf{T}(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n)$$

Solution

Assume that cdf is $F(x/\sigma)$, and let $Z_1 = X_1/\sigma, \cdots, Z_n = X_n/\sigma$ be iid observations from pdf $f(x)$ and cdf $F(x)$.

Ancillary Statistics for Scale Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x/\sigma)/\sigma$ where $\sigma > 0$. Show that the following statistic $\mathbf{T}(\mathbf{X})$ is ancillary.

$$\mathbf{T}(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n)$$

Solution

Assume that cdf is $F(x/\sigma)$, and let $Z_1 = X_1/\sigma, \cdots, Z_n = X_n/\sigma$ be iid observations from pdf $f(x)$ and cdf $F(x)$. Then the joint cdf of the $\mathbf{T}(\mathbf{X})$ is

Ancillary Statistics for Scale Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x/\sigma)/\sigma$ where $\sigma > 0$. Show that the following statistic $\mathbf{T}(\mathbf{X})$ is ancillary.

$$\mathbf{T}(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n)$$

Solution

Assume that cdf is $F(x/\sigma)$, and let $Z_1 = X_1/\sigma, \cdots, Z_n = X_n/\sigma$ be iid observations from pdf $f(x)$ and cdf $F(x)$. Then the joint cdf of the $\mathbf{T}(\mathbf{X})$ is

$$F_{\mathbf{T}}(t_1, \cdots, t_{n-1}|\sigma) = \Pr(X_1/X_n \le t_1, \cdots, X_{n-1}/X_n \le t_{n-1}|\sigma)$$

Ancillary Statistics for Scale Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x/\sigma)/\sigma$ where $\sigma > 0$. Show that the following statistic $\mathbf{T}(\mathbf{X})$ is ancillary.

$$\mathbf{T}(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n)$$

Solution

Assume that cdf is $F(x/\sigma)$, and let $Z_1 = X_1/\sigma, \cdots, Z_n = X_n/\sigma$ be iid observations from pdf $f(x)$ and cdf $F(x)$. Then the joint cdf of the $\mathbf{T}(\mathbf{X})$ is

$$\begin{aligned} F_{\mathbf{T}}(t_1, \cdots, t_{n-1}|\sigma) &= \Pr(X_1/X_n \leq t_1, \cdots, X_{n-1}/X_n \leq t_{n-1}|\sigma) \\ &= \Pr(\sigma Z_1/\sigma Z_n \leq t_1, \cdots, \sigma Z_{n-1}/\sigma Z_n \leq t_{n-1}|\sigma) \end{aligned}$$

Ancillary Statistics for Scale Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x/\sigma)/\sigma$ where $\sigma > 0$. Show that the following statistic $\mathbf{T}(\mathbf{X})$ is ancillary.

$$\mathbf{T}(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n)$$

Solution

Assume that cdf is $F(x/\sigma)$, and let $Z_1 = X_1/\sigma, \cdots, Z_n = X_n/\sigma$ be iid observations from pdf $f(x)$ and cdf $F(x)$. Then the joint cdf of the $\mathbf{T}(\mathbf{X})$ is

$$\begin{aligned} F_{\mathbf{T}}(t_1, \cdots, t_{n-1}|\sigma) &= \Pr(X_1/X_n \leq t_1, \cdots, X_{n-1}/X_n \leq t_{n-1}|\sigma) \\ &= \Pr(\sigma Z_1/\sigma Z_n \leq t_1, \cdots, \sigma Z_{n-1}/\sigma Z_n \leq t_{n-1}|\sigma) \\ &= \Pr(Z_1/Z_n \leq t_1, \cdots, Z_{n-1}/Z_n \leq t_{n-1}|\sigma) \end{aligned}$$

Ancillary Statistics for Scale Family

Problem

Let $X_1, \cdots, X_n$ be iid from a location family with pdf $f(x/\sigma)/\sigma$ where $\sigma > 0$. Show that the following statistic $\mathbf{T}(\mathbf{X})$ is ancillary.

$$\mathbf{T}(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n)$$

Solution

Assume that cdf is $F(x/\sigma)$, and let $Z_1 = X_1/\sigma, \cdots, Z_n = X_n/\sigma$ be iid observations from pdf $f(x)$ and cdf $F(x)$. Then the joint cdf of the $\mathbf{T}(\mathbf{X})$ is

$$\begin{aligned}
F_{\mathbf{T}}(t_1, \cdots, t_{n-1}|\sigma) &= \Pr(X_1/X_n \leq t_1, \cdots, X_{n-1}/X_n \leq t_{n-1}|\sigma) \\
&= \Pr(\sigma Z_1/\sigma Z_n \leq t_1, \cdots, \sigma Z_{n-1}/\sigma Z_n \leq t_{n-1}|\sigma) \\
&= \Pr(Z_1/Z_n \leq t_1, \cdots, Z_{n-1}/Z_n \leq t_{n-1}|\sigma)
\end{aligned}$$

Because $Z_1, \cdots, Z_n$ does not depend on σ, $\mathbf{T}(\mathbf{X})$ is an ancillary statistic.

Summary

Today

- Minimal Sufficient Statistics
  - Recap from last lecture
  - Example from the textbook
- Ancillary Statistics
  - Definition
  - Examples
  - Location-scale family and parameters

Summary

Today

- Minimal Sufficient Statistics
  - Recap from last lecture
  - Example from the textbook
- Ancillary Statistics
  - Definition
  - Examples
  - Location-scale family and parameters

Next Lecture

- Complete Statistics