STAT 583 Lecture 4

Linjun (Leon) Zhang

Department of Statistics
Rutgers University

Feb. 11th

Misc

- Midterm is scheduled at 6:50-8:50pm on March 3rd.
- TA: Youmeng Jiang
  - Office hours: 2:00-3:30pm every Monday.

Last time

- Parameter estimation for population mean μ and binomial parameter p
- CI for μ
  - Large n: z-interval
  - Small n: we need to assume the sample is normally distributed
    - when σ is unknown, use Student's t-distribution
    - when σ is known, use normal distribution
- CI for p
  - Large n: z-interval
- CI for $p_1 - p_2$, $\mu_X - \mu_Y$

Paired Sample

- Sometimes it appears we have data from two samples with the further feature that there is a natural “pairing” of the data between the two samples.

Paired Sample

- Sometimes it appears we have data from two samples with the further feature that there is a natural "pairing" of the data between the two samples.
- For example, suppose that the data consists on n brother-sister pairs, with blood pressures $X_1, ..., X_n$ for the n sisters and blood pressures $Y_1, ..., Y_n$ for their respective brothers.

CI for the difference between two means

Let's first consider the case where n is small.

- If $X_1, X_2, ..., X_n$ i.i.d. $\sim N(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n$ i.i.d. $\sim N(\mu_Y, \sigma_Y^2)$. In the paired samples' case, it's not realistic to assume that $\{X_i\}$ and $\{Y_i\}$ are independent.

CI for the difference between two means

Let's first consider the case where n is small.

- If $X_1, X_2, ..., X_n$ i.i.d. $\sim N(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n$ i.i.d. $\sim N(\mu_Y, \sigma_Y^2)$. In the paired samples' case, it's not realistic to assume that $\{X_i\}$ and $\{Y_i\}$ are independent.
- However, it's natural to assume $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ are n independent pairs.

CI for the difference between two means

Let's first consider the case where n is small.

- If $X_1, X_2, ..., X_n$ i.i.d. $\sim N(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n$ i.i.d. $\sim N(\mu_Y, \sigma_Y^2)$. In the paired samples' case, it's not realistic to assume that $\{X_i\}$ and $\{Y_i\}$ are independent.
- However, it's natural to assume $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ are n independent pairs.
- We denote $D_i = X_i - Y_i$, then $D_i\ i.i.d. \sim N(\mu_X - \mu_Y, \sigma_D^2)$

CI for the difference between two means

Let's first consider the case where n is small.

- If $X_1, X_2, ..., X_n$ i.i.d. $\sim N(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n$ i.i.d. $\sim N(\mu_Y, \sigma_Y^2)$. In the paired samples' case, it's not realistic to assume that $\{X_i\}$ and $\{Y_i\}$ are independent.
- However, it's natural to assume $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ are n independent pairs.
- We denote $D_i = X_i - Y_i$, then $D_i \ i.i.d. \sim N(\mu_X - \mu_Y, \sigma_D^2)$
- We estimate $\mu_X - \mu_Y$ by $\bar{D} = \bar{X} - \bar{Y}$

CI for the difference between two means

Let's first consider the case where n is small.

- If $X_1, X_2, ..., X_n$ i.i.d. $\sim N(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n$ i.i.d. $\sim N(\mu_Y, \sigma_Y^2)$. In the paired samples' case, it's not realistic to assume that $\{X_i\}$ and $\{Y_i\}$ are independent.
- However, it's natural to assume $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ are n independent pairs.
- We denote $D_i = X_i - Y_i$, then $D_i \ i.i.d. \sim N(\mu_X - \mu_Y, \sigma_D^2)$
- We estimate $\mu_X - \mu_Y$ by $\bar{D} = \bar{X} - \bar{Y}$
- A $(1-\alpha)$ confidence interval for $\mu_X - \mu_Y$ is

$$[\bar{D} - t_{\alpha/2, n-1}\frac{s_D}{\sqrt{n}}, \quad \bar{D} + t_{\alpha/2, n-1}\frac{s_D}{\sqrt{n}}],$$

where $s_D = \sqrt{\frac{(D_1 - \bar{D})^2 + \ldots + (D_n - \bar{D})^2}{n}}$.

CI for the difference between two means

Let's first consider the case where n is small.

- If $X_1, X_2, ..., X_n$ i.i.d. $\sim N(\mu_X, \sigma_X^2)$, $Y_1, Y_2, ..., Y_n$ i.i.d. $\sim N(\mu_Y, \sigma_Y^2)$. In the paired samples' case, it's not realistic to assume that $\{X_i\}$ and $\{Y_i\}$ are independent.
- However, it's natural to assume $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ are n independent pairs.
- We denote $D_i = X_i - Y_i$, then $D_i \; i.i.d. \sim N(\mu_X - \mu_Y, \sigma_D^2)$
- We estimate $\mu_X - \mu_Y$ by $\bar{D} = \bar{X} - \bar{Y}$
- A $(1-\alpha)$ confidence interval for $\mu_X - \mu_Y$ is

$$[\bar{D} - t_{\alpha/2,n-1}\frac{s_D}{\sqrt{n}}, \quad \bar{D} + t_{\alpha/2,n-1}\frac{s_D}{\sqrt{n}}],$$

  where $s_D = \sqrt{\frac{(D_1-\bar{D})^2+\ldots+(D_n-\bar{D})^2}{n}}$.
- If n is large, we can simply replace $t_{\alpha/2,n-1}$ with $z_{\alpha/2}$ and remove the normality assumptions.

Example

Suppose that we take the blood pressures of $n = 12$ women and their brothers, and get the following blood pressure reading:

Family	1	2	3	4	5	6	7	8	9	10	11	12
Sister	107	134	111	141	121	118	145	110	164	126	148	132
Brother	110	136	115	140	124	119	148	113	168	129	148	137

Construct the 95% CI for the difference of the mean blood pressure of men and women. ($s^2_{sister} = 307$, $s^2_{brother} = 299$, $s^2_{diff} = 3$, $t_{0.025,11} = 2.306$, $t_{0.05,11} = 1.860$, $t_{0.025,12} = 2.262$, $t_{0.05,12} = 1.833$)

Example

Suppose that we take the blood pressures of $n = 12$ women and their brothers, and get the following blood pressure reading:

Family	1	2	3	4	5	6	7	8	9	10	11	12
Sister	107	134	111	141	121	118	145	110	164	126	148	132
Brother	110	136	115	140	124	119	148	113	168	129	148	137

Construct the 95% CI for the difference of the mean blood pressure of men and women. ($s^2_{sister} = 307$, $s^2_{brother} = 299$, $s^2_{diff} = 3$, $t_{0.025,11} = 2.306$, $t_{0.05,11} = 1.860$, $t_{0.025,12} = 2.262$, $t_{0.05,12} = 1.833$)

Answer

Family	1	2	3	4	5	6	7	8	9	10	11	12
Difference	3	2	4	-1	3	1	3	3	4	3	0	5

- $\bar{D} = 2.5,\ s^2_D = 3,\ t_{0.025,11} = 2.306$

Example

Suppose that we take the blood pressures of $n = 12$ women and their brothers, and get the following blood pressure reading:

Family	1	2	3	4	5	6	7	8	9	10	11	12
Sister	107	134	111	141	121	118	145	110	164	126	148	132
Brother	110	136	115	140	124	119	148	113	168	129	148	137

Construct the 95% CI for the difference of the mean blood pressure of men and women. ($s^2_{sister} = 307$, $s^2_{brother} = 299$, $s^2_{diff} = 3$, $t_{0.025,11} = 2.306$, $t_{0.05,11} = 1.860$, $t_{0.025,12} = 2.262$, $t_{0.05,12} = 1.833$)

Answer

Family	1	2	3	4	5	6	7	8	9	10	11	12
Difference	3	2	4	-1	3	1	3	3	4	3	0	5

- $\bar{D} = 2.5,\ s^2_D = 3,\ t_{0.025,11} = 2.306$
- $2.5 \pm 2.306 \cdot \frac{\sqrt{3}}{\sqrt{12}} = [1.35, 3.65]$.

Overview

- General confidence intervals construction based on pivotal quantities
- General confidence intervals construction based on maximum likelihood estimators

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:
  - We randomly sample an observation from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:
  - We randomly sample an observation from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.
  - If $Y \sim Exp(\theta)$, then $f_Y(y) = \frac{1}{\theta} e^{-y/\theta}$ for $y \geq 0$.

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:
  - We randomly sample an observation from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.
  - If $Y \sim Exp(\theta)$, then $f_Y(y) = \frac{1}{\theta} e^{-y/\theta}$ for $y \geq 0$.
  - Let $U = \frac{Y}{\theta}$, we have $f_U(u) = f_Y(u \cdot \theta) \cdot \frac{dy}{du} = \frac{1}{\theta} e^{-u} \cdot \theta = e^{-u}$ for $u > 0$, that is, $U \sim Exp(1)$

Other confidence intervals

- We have now talked about CI for mean parameters: $\mu, p, \mu_1 - \mu_2, p_1 - p_2$
- For general CIs, a useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:
  - We randomly sample an observation from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.
  - If $Y \sim Exp(\theta)$, then $f_Y(y) = \frac{1}{\theta} e^{-y/\theta}$ for $y \geq 0$.
  - Let $U = \frac{Y}{\theta}$, we have $f_U(u) = f_Y(u \cdot \theta) \cdot \frac{dy}{du} = \frac{1}{\theta} e^{-u} \cdot \theta = e^{-u}$ for $u > 0$, that is, $U \sim Exp(1)$
  - Thus, we can use $U = \frac{Y}{\theta}$ as a pivotal quantity

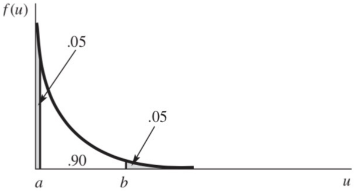
Example

- We need to find two numbers a and b such that

$$\mathbb{P}(a \leq U \leq b) = 90\%$$

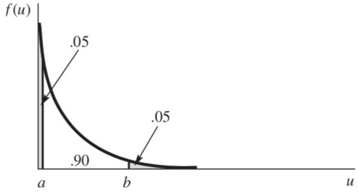
Example

- We need to find two numbers a and b such that
$$\mathbb{P}(a \leq U \leq b) = 90\%$$
- One way to do this is to choose a and b to satisfy $\mathbb{P}(U < a) = \mathbb{P}(U > b) = 5\%$



Example

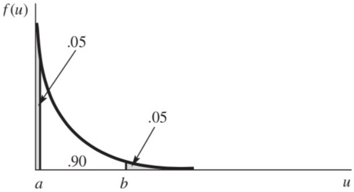
- We need to find two numbers a and b such that
$$\mathbb{P}(a \le U \le b) = 90\%$$
- One way to do this is to choose a and b to satisfy $\mathbb{P}(U < a) = \mathbb{P}(U > b) = 5\%$



- $1 - e^{-a} = .05$ and $e^{-b} = .05$, equivalently, $a = 0.051$, $b = 2.996$

Example

- We need to find two numbers a and b such that
$$\mathbb{P}(a \leq U \leq b) = 90\%$$
- One way to do this is to choose a and b to satisfy $\mathbb{P}(U < a) = \mathbb{P}(U > b) = 5\%$



- $1 - e^{-a} = .05$ and $e^{-b} = .05$, equivalently, $a = 0.051$, $b = 2.996$
- $0.9 = \mathbb{P}(0.051 \leq \frac{Y}{\theta} \leq 2.996) = \mathbb{P}(\frac{Y}{2.996} \leq \theta \leq \frac{Y}{0.051})$

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, ..., X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+...+\frac{1}{\theta_n}})$.

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, ..., X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+...+\frac{1}{\theta_n}})$.
- $\frac{X_1}{\theta}, ... \frac{X_{10}}{\theta} \overset{i.i.d.}{\sim} Exp(1)$

Example

Example

We randomly sample 10 observations $X_1, \ldots, X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, \ldots, X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+\ldots+\frac{1}{\theta_n}})$.
- $\frac{X_1}{\theta}, \ldots \frac{X_{10}}{\theta} \overset{i.i.d.}{\sim} Exp(1)$
- $U = \min\{\frac{X_1}{\theta}, \ldots \frac{X_{10}}{\theta}\} \sim Exp(\frac{1}{10})$

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, ..., X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+...+\frac{1}{\theta_n}})$.
- $\frac{X_1}{\theta}, ... \frac{X_{10}}{\theta} \overset{i.i.d.}{\sim} Exp(1)$
- $U = \min\{\frac{X_1}{\theta}, ... \frac{X_{10}}{\theta}\} \sim Exp(\frac{1}{10})$
- Find a, b such that $\mathbb{P}(U < a) = 1 - e^{-10a} = .05$ and $\mathbb{P}(U > b) = e^{-10b} = .05$, equivalently, $a = 0.005,\ b = 0.300$

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, ..., X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+...+\frac{1}{\theta_n}})$.
- $\frac{X_1}{\theta}, ... \frac{X_{10}}{\theta} \overset{i.i.d.}{\sim} Exp(1)$
- $U = \min\{\frac{X_1}{\theta}, ... \frac{X_{10}}{\theta}\} \sim Exp(\frac{1}{10})$
- Find a, b such that $\mathbb{P}(U < a) = 1 - e^{-10a} = .05$ and $\mathbb{P}(U > b) = e^{-10b} = .05$, equivalently, $a = 0.005,\ b = 0.300$
- $0.9 = \mathbb{P}(0.005 < U < 0.300) = \mathbb{P}(0.005 < \frac{\min\{X_1,...,X_{10}\}}{\theta} < 0.300) = \mathbb{P}(\frac{\min\{X_1,...,X_{10}\}}{0.300} < \theta < \frac{\min\{X_1,...,X_{10}\}}{0.005})$

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Answer

- Consider $U = \frac{Y}{\theta}$

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Answer

- Consider $U = \frac{Y}{\theta}$
- $U \sim U(0, 1)$

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Answer

- Consider $U = \frac{Y}{\theta}$
- $U \sim U(0, 1)$
- $\mathbb{P}(U < 0.95) = 0.95$

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Answer

- Consider $U = \frac{Y}{\theta}$
- $U \sim U(0, 1)$
- $\mathbb{P}(U < 0.95) = 0.95$
- $0.95 = \mathbb{P}(U < 0.95) = \mathbb{P}(\frac{Y}{\theta} < 0.95) = \mathbb{P}(\frac{Y}{0.95} \leq \theta)$

Example

Example

Let $Y_1, ..., Y_n$ denote a random sample of size n from $U(0, \theta)$. Let $Y_{(n)} = \max\{Y_1, Y_2, ..., Y_n\}$. Find a 95% lower confidence bound for θ.

Example

Example

Let $Y_1, ..., Y_n$ denote a random sample of size n from $U(0, \theta)$. Let $Y_{(n)} = \max\{Y_1, Y_2, ..., Y_n\}$. Find a 95% lower confidence bound for θ.

Answer

- Consider $U = \frac{Y_{(n)}}{\theta}$

Example

Example

Let $Y_1, ..., Y_n$ denote a random sample of size n from $U(0, \theta)$. Let $Y_{(n)} = \max\{Y_1, Y_2, ..., Y_n\}$. Find a 95% lower confidence bound for θ.

Answer

- Consider $U = \frac{Y_{(n)}}{\theta}$
- U has the cdf
  $F(u) = \mathbb{P}(\frac{\max\{Y_1, Y_2, \ldots, Y_n\}}{\theta} \leq u) = \mathbb{P}(\frac{Y_1}{\theta} \leq u, ..., \frac{Y_n}{\theta} \leq u) = \mathbb{P}(\frac{Y_1}{\theta} \leq u)^n$

Example

Example

Let $Y_1, \ldots, Y_n$ denote a random sample of size n from $U(0, \theta)$. Let $Y_{(n)} = \max\{Y_1, Y_2, \ldots, Y_n\}$. Find a 95% lower confidence bound for θ.

Answer

- Consider $U = \frac{Y_{(n)}}{\theta}$
- U has the cdf
  $F(u) = \mathbb{P}(\frac{\max\{Y_1, Y_2, \ldots, Y_n\}}{\theta} \leq u) = \mathbb{P}(\frac{Y_1}{\theta} \leq u, \ldots, \frac{Y_n}{\theta} \leq u) = \mathbb{P}(\frac{Y_1}{\theta} \leq u)^n$
- $F(u) = u^n$ for $u \in [0, 1]$.

Example

Example

Let $Y_1, \ldots, Y_n$ denote a random sample of size n from $U(0, \theta)$. Let $Y_{(n)} = \max\{Y_1, Y_2, \ldots, Y_n\}$. Find a 95% lower confidence bound for θ.

Answer

- Consider $U = \frac{Y_{(n)}}{\theta}$
- U has the cdf
  $F(u) = \mathbb{P}(\frac{\max\{Y_1, Y_2, \ldots, Y_n\}}{\theta} \leq u) = \mathbb{P}(\frac{Y_1}{\theta} \leq u, \ldots, \frac{Y_n}{\theta} \leq u) = \mathbb{P}(\frac{Y_1}{\theta} \leq u)^n$
- $F(u) = u^n$ for $u \in [0, 1]$.
- $0.95 = \mathbb{P}(U < (0.95)^{1/n}) = \mathbb{P}(\frac{Y_{(n)}}{\theta} < (0.95)^{1/n}) = \mathbb{P}(\frac{Y_{(n)}}{(0.95)^{1/n}} \leq \theta)$

Example

Example

If we have $n = 10$ samples i.i.d. from $U(0, \theta)$, with the max $Y_{\text{max}} = 5.7$. We are 95% confident that θ is at least:.

Example

Example

If we have $n = 10$ samples i.i.d. from $U(0, \theta)$, with the max $Y_{\max} = 5.7$. We are 95% confident that θ is at least:.

Answer

- $0.95 = \mathbb{P}(U < (0.95)^{1/n}) = \mathbb{P}(\frac{Y_{(n)}}{\theta} < (0.95)^{1/n}) = \mathbb{P}(\frac{Y_{(n)}}{(0.95)^{1/n}} \leq \theta)$

Example

Example

If we have $n = 10$ samples i.i.d. from $U(0, \theta)$, with the max $Y_{\max} = 5.7$. We are 95% confident that θ is at least:.

Answer

- $0.95 = \mathbb{P}(U < (0.95)^{1/n}) = \mathbb{P}(\frac{Y_{(n)}}{\theta} < (0.95)^{1/n}) = \mathbb{P}(\frac{Y_{(n)}}{(0.95)^{1/n}} \leq \theta)$
- $\frac{Y_{(n)}}{(0.95)^{1/n}} = \frac{5.7}{(0.95)^{1/10}} = 5.73$

Confidence interval for the variance

- Suppose $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown. We seek a $100(1-\alpha)\%$ CI for σ^2

Confidence interval for the variance

- Suppose $Y_1, \dots, Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown. We seek a $100(1-\alpha)\%$ CI for σ^2
- Recall that, for $S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$, we have the pivotal quantity
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Confidence interval for the variance

- Suppose $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown. We seek a $100(1-\alpha)\%$ CI for σ^2
- Recall that, for $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, we have the pivotal quantity
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$
- As a result,
$$\mathbb{P}(\chi^2_{1-(\alpha/2),n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-1}) = 1-\alpha$$

Confidence interval for the variance

- Suppose $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown. We seek a $100(1-\alpha)\%$ CI for σ^2
- Recall that, for $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, we have the pivotal quantity
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$
- As a result,
$$\mathbb{P}(\chi^2_{1-(\alpha/2),n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-1}) = 1-\alpha$$
- A $100(1-\alpha)\%$ Confidence Interval for σ^2:
$$\mathbb{P}(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2),n-1}}) = 1-\alpha.$$

Example

Example

Suppose the maturation times for a flower species are $N(\mu, \sigma^2)$. If a random sample of $n = 13$ seeds yielded $s^2 = 10.7$, then what is a 90% CI for σ^2?
$(\chi^2_{0.05,12} = 21.03, \chi^2_{0.95,12} = 5.23)$

Example

Example

Suppose the maturation times for a flower species are $N(\mu, \sigma^2)$. If a random sample of $n = 13$ seeds yielded $s^2 = 10.7$, then what is a 90% CI for σ^2? $(\chi^2_{0.05,12} = 21.03, \chi^2_{0.95,12} = 5.23)$

Answer

- $$\mathbb{P}\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2),n-1}}\right) = 1 - \alpha.$$

Example

Example

Suppose the maturation times for a flower species are $N(\mu, \sigma^2)$. If a random sample of $n = 13$ seeds yielded $s^2 = 10.7$, then what is a 90% CI for σ^2? $(\chi^2_{0.05,12} = 21.03, \chi^2_{0.95,12} = 5.23)$

Answer

- $$\mathbb{P}\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2),n-1}}\right) = 1 - \alpha.$$
- $\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} = \frac{12 \times 10.7}{21.03} = 6.11$

Example

Example

Suppose the maturation times for a flower species are $N(\mu, \sigma^2)$. If a random sample of $n = 13$ seeds yielded $s^2 = 10.7$, then what is a 90% CI for σ^2? ($\chi^2_{0.05,12} = 21.03, \chi^2_{0.95,12} = 5.23$)

Answer

- $$\mathbb{P}(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2),n-1}}) = 1 - \alpha.$$
- $\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} = \frac{12\times10.7}{21.03} = 6.11$
- $\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} = \frac{12\times10.7}{5.23} = 24.55$

Example

Example

Suppose the maturation times for a flower species are $N(\mu, \sigma^2)$. If a random sample of $n = 13$ seeds yielded $s^2 = 10.7$, then what is a 90% CI for σ^2? $(\chi^2_{0.05,12} = 21.03, \chi^2_{0.95,12} = 5.23)$

Answer

- $$\mathbb{P}\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2),n-1}}\right) = 1 - \alpha.$$
- $\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} = \frac{12 \times 10.7}{21.03} = 6.11$
- $\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} = \frac{12 \times 10.7}{5.23} = 24.55$
- The 90% CI is $[6.11, 24.55]$.

Maximum likelihood estimator

- Another general and important method to construct point estimation and confidence intervals are maximum likelihood estimator (MLE)

Likelihood function

- Suppose $Y_1, \ldots, Y_n$ are an i.i.d. random sample from pdf $f_Y(y;\theta)$

Likelihood function

- Suppose $Y_1, ..., Y_n$ are an i.i.d. random sample from pdf $f_Y(y; \theta)$
- Then we call

$$f(y_1, y_2, ..., y_n | \theta)$$

the joint density function. It is a function of $y_1, ..., y_n$ for a fixed (unknown) value of θ.

Likelihood function

- Suppose $Y_1, ..., Y_n$ are an i.i.d. random sample from pdf $f_Y(y; \theta)$
- Then we call

$$f(y_1, y_2, ..., y_n | \theta)$$

  the joint density function. It is a function of $y_1, ..., y_n$ for a fixed (unknown) value of θ.
- This function gives the probability or likelihood of observing the event $\{(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n)\}$ when the value of the parameter is θ.

Likelihood function

- Suppose $Y_1, ..., Y_n$ are an i.i.d. random sample from pdf $f_Y(y; \theta)$
- Then we call

$$f(y_1, y_2, ..., y_n|\theta)$$

  the joint density function. It is a function of $y_1, ..., y_n$ for a fixed (unknown) value of θ.
- This function gives the probability or likelihood of observing the event $\{(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n)\}$ when the value of the parameter is θ.
- By the i.i.d. assumption, we have

$$f(y_1, y_2, ..., y_n|\theta) = \Pi_{i=1}^{n} f(y_i \mid \theta)$$

Likelihood function

- We could also view $f(y_1, y_2, ..., y_n|\theta)$ as a function of θ for a given data set $y_1, ..., y_n$. In that case we write

$$L(\theta \mid y_1, ..., y_n) := f(y_1, y_2, ..., y_n|\theta).$$

Likelihood function

- We could also view $f(y_1, y_2, ..., y_n|\theta)$ as a function of θ for a given data set $y_1, ..., y_n$. In that case we write

$$L(\theta \mid y_1, ..., y_n) := f(y_1, y_2, ..., y_n|\theta).$$

- We call $L(\theta \mid y_1, ..., y_n)$ the likelihood function.

Likelihood function

- We could also view $f(y_1, y_2, ..., y_n|\theta)$ as a function of θ for a given data set $y_1, ..., y_n$. In that case we write

$$L(\theta \mid y_1, ..., y_n) := f(y_1, y_2, ..., y_n|\theta).$$

- We call $L(\theta \mid y_1, ..., y_n)$ the likelihood function.
- Note: the joint density and likelihood function are the same function, but the first is treated as a function of the data, and the second as a function of the parameter θ.

Maximum likelihood estimator

- Loosely speaking, the value of θ that maximizes $L(\theta \mid y_1, ..., y_n)$ is the parameter value that is "most likely" to have produced the data we observe.

Maximum likelihood estimator

- Loosely speaking, the value of θ that maximizes $L(\theta \mid y_1, ..., y_n)$ is the parameter value that is "most likely" to have produced the data we observe.
- The maximum likelihood estimator (MLE) of the parameter θ is

$$\hat{\theta} = \arg\max L(\theta \mid y_1, ..., y_n)$$

Maximum likelihood estimator

- Loosely speaking, the value of θ that maximizes $L(\theta \mid y_1, ..., y_n)$ is the parameter value that is "most likely" to have produced the data we observe.
- The maximum likelihood estimator (MLE) of the parameter θ is

$$\hat{\theta} = \arg\max L(\theta \mid y_1, ..., y_n)$$

- Note: θ can be taken as a vector $\theta = (\theta_1, ..., \theta_k)$

Maximum likelihood estimator

- Loosely speaking, the value of θ that maximizes $L(\theta \mid y_1, ..., y_n)$ is the parameter value that is "most likely" to have produced the data we observe.
- The maximum likelihood estimator (MLE) of the parameter θ is

$$\hat{\theta} = \arg\max L(\theta \mid y_1, ..., y_n)$$

- Note: θ can be taken as a vector $\theta = (\theta_1, ..., \theta_k)$
- Note 2: In many cases, it is easier to maximize $\log L(\theta)$ than to maximize $L(\theta)$ itself, since log is a strictly increasing function. We call $\log L(\theta)$ the log likelihood.

Maximum likelihood estimator

Often, the MLE is found by

1. Writing out the (log) likelihood asa function of the parameter (say, θ)
2. Taking the derivative with respect to θ
3. Setting the derivative equal to 0 and solving for $\hat{\theta}$
4. Checking that the second derivative is negative at $\hat{\theta}$ to ensure the solution is a maximum.

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p
- Solution

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p
- Solution
  - $f(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}$.

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p
- Solution
  - $f(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}$.
  - $L(p) = L(y_1, y_2, ..., y_n|p) = p^y(1-p)^{n-y}$, where $y = \sum_{i=1}^n y_i$.

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p
- Solution
  - $f(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}$.
  - $L(p) = L(y_1, y_2, ..., y_n|p) = p^y(1-p)^{n-y}$, where $y = \sum_{i=1}^n y_i$.
  - $\log L(p) = y \log p + (n-y)\log(1-p)$

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p
- Solution
  - $f(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}$.
  - $L(p) = L(y_1, y_2, ..., y_n|p) = p^y(1-p)^{n-y}$, where $y = \sum_{i=1}^n y_i$.
  - $\log L(p) = y \log p + (n-y)\log(1-p)$
  - $\frac{d(\log L(p))}{dp} = y(\frac{1}{p}) + (n-y)\frac{-1}{n-p}$

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p
- Solution
  - $f(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}$.
  - $L(p) = L(y_1, y_2, ..., y_n|p) = p^y(1-p)^{n-y}$, where $y = \sum_{i=1}^n y_i$.
  - $\log L(p) = y \log p + (n-y)\log(1-p)$
  - $\frac{d(\log L(p))}{dp} = y(\frac{1}{p}) + (n-y)\frac{-1}{n-p}$
  - We need to solve for $\hat{p}$ such that
$$\frac{y}{\hat{p}} = \frac{n-y}{1-\hat{p}}$$

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p
- Solution
  - $f(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}$.
  - $L(p) = L(y_1, y_2, ..., y_n|p) = p^y(1-p)^{n-y}$, where $y = \sum_{i=1}^n y_i$.
  - $\log L(p) = y \log p + (n-y)\log(1-p)$
  - $\frac{d(\log L(p))}{dp} = y(\frac{1}{p}) + (n-y)\frac{-1}{n-p}$
  - We need to solve for $\hat{p}$ such that
$$\frac{y}{\hat{p}} = \frac{n-y}{1-\hat{p}}$$
  - We then have $p = \frac{y}{n}$

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of p
- Solution
  - $f(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}$.
  - $L(p) = L(y_1, y_2, ..., y_n | p) = p^y(1-p)^{n-y}$, where $y = \sum_{i=1}^n y_i$.
  - $\log L(p) = y \log p + (n-y)\log(1-p)$
  - $\frac{d(\log L(p))}{dp} = y(\frac{1}{p}) + (n-y)\frac{-1}{n-p}$
  - We need to solve for $\hat{p}$ such that
  $$\frac{y}{\hat{p}} = \frac{n-y}{1-\hat{p}}$$
  - We then have $p = \frac{y}{n}$
  - Hence, the MLE of p is actually the intuitive estimator for p that we used before

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} U(0, \theta)$. Find the MLE of θ

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} U(0, \theta)$. Find the MLE of θ
- Solution

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} U(0, \theta)$. Find the MLE of θ
- Solution
  - $f(y_i \mid \theta) = I_{[0,\theta]}(y_i)$, , where $I_{[a,b]}(\cdot)$ is an indicator function such that $I_{a,b}(x) = 1$ if $x \in [a, b]$, and $= 0$ otherwise

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} U(0, \theta)$. Find the MLE of θ
- Solution
    - $f(y_i \mid \theta) = I_{[0,\theta]}(y_i)$, , where $I_{[a,b]}(\cdot)$ is an indicator function such that $I_{a,b}(x) = 1$ if $x \in [a, b]$, and $= 0$ otherwise
    - $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \prod_{i=1}^{n} I_{[0,\theta]}(y_i) = 1$ if $y_1, ..., y_n \in [0, \theta]$, and 0 otherwise

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} U(0, \theta)$. Find the MLE of θ
- Solution
  - $f(y_i \mid \theta) = I_{[0,\theta]}(y_i)$, , where $I_{[a,b]}(\cdot)$ is an indicator function such that $I_{a,b}(x) = 1$ if $x \in [a, b]$, and $= 0$ otherwise
  - $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \prod_{i=1}^n I_{[0,\theta]}(y_i) = 1$ if $y_1, ..., y_n \in [0, \theta]$, and 0 otherwise
  - $L(\theta)$ is maximized if $\hat{\theta} \geq \max_i y_i$

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} U(0, \theta)$. Find the MLE of θ
- Solution
    - $f(y_i \mid \theta) = I_{[0,\theta]}(y_i)$, , where $I_{[a,b]}(\cdot)$ is an indicator function such that $I_{a,b}(x) = 1$ if $x \in [a, b]$, and $= 0$ otherwise
    - $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \prod_{i=1}^{n} I_{[0,\theta]}(y_i) = 1$ if $y_1, ..., y_n \in [0, \theta]$, and 0 otherwise
    - $L(\theta)$ is maximized if $\hat{\theta} \geq \max_i y_i$
    - The MLE can be any value $\geq \max_i y_i$

Example of maximum likelihood estimator

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} U(0, \theta)$. Find the MLE of θ
- Solution
  - $f(y_i \mid \theta) = I_{[0,\theta]}(y_i)$, , where $I_{[a,b]}(\cdot)$ is an indicator function such that $I_{a,b}(x) = 1$ if $x \in [a, b]$, and $= 0$ otherwise
  - $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \prod_{i=1}^{n} I_{[0,\theta]}(y_i) = 1$ if $y_1, ..., y_n \in [0, \theta]$, and 0 otherwise
  - $L(\theta)$ is maximized if $\hat{\theta} \geq \max_i y_i$
  - The MLE can be any value $\geq \max_i y_i$
  - Hence, the MLE of θ is not unique

Example of maximum likelihood estimator

Problem

Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Exp(\theta)$. Find the MLE of θ.

Solution

- $f(y_i \mid \theta) = \frac{1}{\theta} e^{-y_i/\theta}$

Example of maximum likelihood estimator

Problem

Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Exp(\theta)$. Find the MLE of θ.

Solution

- $f(y_i \mid \theta) = \frac{1}{\theta} e^{-y_i/\theta}$
- $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \theta^{-n} e^{-(y_1+...+y_n)/\theta}$

Example of maximum likelihood estimator

Problem

Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Exp(\theta)$. Find the MLE of θ.

Solution

- $f(y_i \mid \theta) = \frac{1}{\theta} e^{-y_i/\theta}$
- $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \theta^{-n} e^{-(y_1+...+y_n)/\theta}$
- $\log L(\theta) = -n \log \theta - (y_1 + ... + y_n)/\theta$

Example of maximum likelihood estimator

Problem

Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Exp(\theta)$. Find the MLE of θ.

Solution

- $f(y_i \mid \theta) = \frac{1}{\theta} e^{-y_i/\theta}$
- $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \theta^{-n} e^{-(y_1+...+y_n)/\theta}$
- $\log L(\theta) = -n \log \theta - (y_1 + ... + y_n)/\theta$
- $\frac{d(\log L(\theta))}{d\theta} = -\frac{n}{\theta} + \frac{y_1+...+y_n}{\theta^2}$

Example of maximum likelihood estimator

Problem

Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Exp(\theta)$. Find the MLE of θ.

Solution

- $f(y_i \mid \theta) = \frac{1}{\theta} e^{-y_i/\theta}$
- $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \theta^{-n} e^{-(y_1+...+y_n)/\theta}$
- $\log L(\theta) = -n \log \theta - (y_1 + ... + y_n)/\theta$
- $\frac{d(\log L(\theta))}{d\theta} = -\frac{n}{\theta} + \frac{y_1+...+y_n}{\theta^2}$
- We need to solve for $\hat{\theta}$ such that $\frac{n}{\hat{\theta}} = \frac{y_1+...+y_n}{\hat{\theta}^2}$

Example of maximum likelihood estimator

Problem

Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Exp(\theta)$. Find the MLE of θ.

Solution

- $f(y_i \mid \theta) = \frac{1}{\theta} e^{-y_i/\theta}$
- $L(\theta) = L(y_1, y_2, ..., y_n|\theta) = \theta^{-n} e^{-(y_1+...+y_n)/\theta}$
- $\log L(\theta) = -n \log \theta - (y_1 + ... + y_n)/\theta$
- $\frac{d(\log L(\theta))}{d\theta} = -\frac{n}{\theta} + \frac{y_1+...+y_n}{\theta^2}$
- We need to solve for $\hat{\theta}$ such that $\frac{n}{\hat{\theta}} = \frac{y_1+...+y_n}{\hat{\theta}^2}$
- We then have $\hat{\theta} = \frac{y_1+...+y_n}{n}$.

Example of MLE with multiple parameters

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Find the MLE of μ and σ.

Example of MLE with multiple parameters

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Find the MLE of μ and σ.
- $f(y_i \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(y_i-\mu)^2}{2\sigma^2}]$

Example of MLE with multiple parameters

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Find the MLE of μ and σ.
- $f(y_i \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(y_i-\mu)^2}{2\sigma^2}]$
- $L(\lambda) = L(y_1, y_2, ..., y_n|\lambda) = (\frac{1}{2\pi\sigma^2})^{n/2} \exp[-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}]$

Example of MLE with multiple parameters

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Find the MLE of μ and σ.
- $f(y_i \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(y_i-\mu)^2}{2\sigma^2}]$
- $L(\lambda) = L(y_1, y_2, ..., y_n|\lambda) = (\frac{1}{2\pi\sigma^2})^{n/2} \exp[-\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2}]$
- $\log L(\lambda) = -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2.$

Example of MLE with multiple parameters

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Find the MLE of μ and σ.
- $f(y_i \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(y_i-\mu)^2}{2\sigma^2}]$
- $L(\lambda) = L(y_1, y_2, ..., y_n|\lambda) = (\frac{1}{2\pi\sigma^2})^{n/2} \exp[-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}]$
- $\log L(\lambda) = -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$.
- Taking derivatives:

$$\frac{\partial(\log L(\mu, \sigma^2))}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)$$

$$\frac{\partial(\log L(\mu, \sigma^2))}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2$$

Example of MLE with multiple parameters

- Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Find the MLE of μ and σ.
- $f(y_i \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(y_i-\mu)^2}{2\sigma^2}]$
- $L(\lambda) = L(y_1, y_2, ..., y_n|\lambda) = (\frac{1}{2\pi\sigma^2})^{n/2} \exp[-\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2}]$
- $\log L(\lambda) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2$.
- Taking derivatives:

$$\frac{\partial(\log L(\mu, \sigma^2))}{\partial\mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu)$$

$$\frac{\partial(\log L(\mu, \sigma^2))}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \mu)^2$$

- Setting these derivatives equal to zero, we get $\mu = \frac{1}{n} \sum_{i=1}^n y_i$ and $\hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2$.

Properties of MLEs

- We are often interested in estimating a function of a parameter $g(\theta)$
- The invariance property of MLEs states that if $\hat{\theta}$ is a MLE of θ, and $g(\cdot)$ is any function, then $g(\hat{\theta})$ is a MLE of $g(\theta)$.

Example

Problem

Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of $Var(\sum_{i=1}^n Y_i)$.

Solution

- $Var(\sum_{i=1}^n y_i) = nVar(y_1) = np(1-p)$

Example

Problem

Suppose that $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of $Var(\sum_{i=1}^n Y_i)$.

Solution

- $Var(\sum_{i=1}^n y_i) = nVar(y_1) = np(1-p)$
- The MLE of p is $\hat{p} = \frac{1}{n}\sum_{i=1}^n Y_i$

Example

Problem

Suppose that $Y_1, ..., Y_n \overset{i.i.d.}{\sim} Ber(p)$. Find the MLE of $Var(\sum_{i=1}^n Y_i)$.

Solution

- $Var(\sum_{i=1}^n y_i) = nVar(y_1) = np(1-p)$
- The MLE of p is $\hat{p} = \frac{1}{n}\sum_{i=1}^n Y_i$
- The MLE of $np(1-p)$ is $n\hat{p}(1-\hat{p}) = \sum_{i=1}^n Y_i(1 - \frac{1}{n}\sum_{i=1}^n Y_i)$

Asymptotic properties of estimators

- The asymptotic behavior of an estimator refers to its behavior as the sample size n increases towards ∞

Asymptotic properties of estimators

- The asymptotic behavior of an estimator refers to its behavior as the sample size n increases towards ∞
- Since an estimator $\hat{\theta}$ usually depends on n, we may denote a sequence of such estimators by $\hat{\theta}_n$

Asymptotic properties of estimators

- The asymptotic behavior of an estimator refers to its behavior as the sample size n increases towards ∞
- Since an estimator $\hat{\theta}$ usually depends on n, we may denote a sequence of such estimators by $\hat{\theta}_n$
- Intuitively, we would like our estimator to "get closer" to the target parameter as $n \to \infty$

Asymptotic properties of estimators

- The asymptotic behavior of an estimator refers to its behavior as the sample size n increases towards ∞
- Since an estimator $\hat{\theta}$ usually depends on n, we may denote a sequence of such estimators by $\hat{\theta}_n$
- Intuitively, we would like our estimator to "get closer" to the target parameter as $n \to \infty$
- Definition: An estimator $\hat{\theta}_n$ is a consistent estimator of θ if, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|\hat{\theta}_n - \theta| \leq \epsilon) = 1,$$

also denote as $\hat{\theta}_n \xrightarrow{p} \theta$ (converge in probability).

Asymptotic properties of MLE

Theorem

If $Y_1, ..., Y_n \overset{i.i.d.}{\sim} f_Y(y;\theta)$ and $\hat{\theta}$ is the MLE of θ, then assuming certain regularity conditions:

$$\hat{\theta}_n \xrightarrow{p} \theta \text{ as } n \to \infty.$$

$$\frac{\hat{\theta}_n - \theta}{\sigma_{\hat{\theta}}} \to N(0,1),$$

where $\sigma^2_{\hat{\theta}} = 1/(n\mathbb{E}[-\frac{\partial^2 \log f(Y;\theta)}{\partial \theta^2}])$.

Remark: this implies that $\hat{\theta}$ is consistent and asymptotically normal. The term $n\mathbb{E}[-\frac{\partial^2 \log f(Y;\theta)}{\partial \theta^2}]$ is called Fisher information.

CI based on MLE

We can then obtain the following approximate large-sample $100(1-\alpha)\%$ confidence interval for θ:

$$\hat{\theta}_n \pm z_{\alpha/2}\sqrt{1/(n\mathbb{E}[-\frac{\partial^2 \log f(Y;\theta)}{\partial\theta^2}])} \,|_{\theta=\hat{\theta}_n} .$$

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.
- We know the MLE is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.
- We know the MLE is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$
- Also, $f(Y \mid p) = p^Y(1-p)^{1-Y}$ for $Y = 0, 1$

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.
- We know the MLE is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$
- Also, $f(Y \mid p) = p^Y (1-p)^{1-Y}$ for $Y = 0, 1$
- $\log f(Y \mid p) = Y \log p + (1 - Y) \log(1-p)$

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.
- We know the MLE is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$
- Also, $f(Y \mid p) = p^Y(1-p)^{1-Y}$ for $Y = 0, 1$
- $\log f(Y \mid p) = Y \log p + (1-Y)\log(1-p)$
- $\frac{\partial \log f(Y|p)}{\partial p} = \frac{Y}{p} - \frac{1-Y}{1-p}$

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.
- We know the MLE is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$
- Also, $f(Y \mid p) = p^Y(1-p)^{1-Y}$ for $Y = 0, 1$
- $\log f(Y \mid p) = Y \log p + (1-Y)\log(1-p)$
- $\frac{\partial \log f(Y|p)}{\partial p} = \frac{Y}{p} - \frac{1-Y}{1-p}$
- $\frac{\partial^2 \log f(Y|p)}{\partial p^2} = -\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}$

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.
- We know the MLE is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$
- Also, $f(Y \mid p) = p^Y(1-p)^{1-Y}$ for $Y = 0, 1$
- $\log f(Y \mid p) = Y \log p + (1-Y)\log(1-p)$
- $\frac{\partial \log f(Y|p)}{\partial p} = \frac{Y}{p} - \frac{1-Y}{1-p}$
- $\frac{\partial^2 \log f(Y|p)}{\partial p^2} = -\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}$
- $\mathbb{E}[-\frac{\partial^2 \log f(Y|p)}{\partial p^2}] = \mathbb{E}[\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}] = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.
- We know the MLE is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$
- Also, $f(Y \mid p) = p^Y(1-p)^{1-Y}$ for $Y = 0, 1$
- $\log f(Y \mid p) = Y \log p + (1-Y)\log(1-p)$
- $\frac{\partial \log f(Y|p)}{\partial p} = \frac{Y}{p} - \frac{1-Y}{1-p}$
- $\frac{\partial^2 \log f(Y|p)}{\partial p^2} = -\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}$
- $\mathbb{E}[-\frac{\partial^2 \log f(Y|p)}{\partial p^2}] = \mathbb{E}[\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}] = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$
- $\hat{\theta}_n \pm z_{\alpha/2}\sqrt{1/(n\mathbb{E}[-\frac{\partial^2 \log f(Y;\theta)}{\partial \theta^2}])}\ |_{\theta=\hat{\theta}_n}$.

Example

- For random variable with a Bernoulli distribution with parameter p. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for p.
- We know the MLE is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n}$
- Also, $f(Y \mid p) = p^Y(1-p)^{1-Y}$ for $Y = 0, 1$
- $\log f(Y \mid p) = Y \log p + (1-Y)\log(1-p)$
- $\frac{\partial \log f(Y|p)}{\partial p} = \frac{Y}{p} - \frac{1-Y}{1-p}$
- $\frac{\partial^2 \log f(Y|p)}{\partial p^2} = -\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}$
- $\mathbb{E}[-\frac{\partial^2 \log f(Y|p)}{\partial p^2}] = \mathbb{E}[\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}] = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$
- $\hat{\theta}_n \pm z_{\alpha/2}\sqrt{1/(n\mathbb{E}[-\frac{\partial^2 \log f(Y;\theta)}{\partial \theta^2}])} \mid_{\theta=\hat{\theta}_n}$.
- The $100(1-\alpha)\%$ CI is $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.
- We know the MLE is $\hat{\theta} = \bar{Y}$

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.
- We know the MLE is $\hat{\theta} = \bar{Y}$
- $f(Y \mid \theta) = \frac{1}{\theta} e^{-Y/\theta}$ for $Y = 0, 1, 2, ...$

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.
- We know the MLE is $\hat{\theta} = \bar{Y}$
- $f(Y \mid \theta) = \frac{1}{\theta} e^{-Y/\theta}$ for $Y = 0, 1, 2, ...$
- $\log f(Y \mid \theta) = -\log \theta - Y/\theta$

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.
- We know the MLE is $\hat{\theta} = \bar{Y}$
- $f(Y \mid \theta) = \frac{1}{\theta} e^{-Y/\theta}$ for $Y = 0, 1, 2, ...$
- $\log f(Y \mid \theta) = -\log \theta - Y/\theta$
- $\frac{\partial \log f(Y|\theta)}{\partial \theta} = -\frac{1}{\theta} + \frac{Y}{\theta^2}$

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.
- We know the MLE is $\hat{\theta} = \bar{Y}$
- $f(Y \mid \theta) = \frac{1}{\theta} e^{-Y/\theta}$ for $Y = 0, 1, 2, ...$
- $\log f(Y \mid \theta) = -\log \theta - Y/\theta$
- $\frac{\partial \log f(Y|\theta)}{\partial \theta} = -\frac{1}{\theta} + \frac{Y}{\theta^2}$
- $\frac{\partial^2 \log f(Y|\theta)}{\partial \theta^2} = \frac{1}{\theta^2} - 2\frac{Y}{\theta^3}$

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.
- We know the MLE is $\hat{\theta} = \bar{Y}$
- $f(Y \mid \theta) = \frac{1}{\theta} e^{-Y/\theta}$ for $Y = 0, 1, 2, ...$
- $\log f(Y \mid \theta) = -\log\theta - Y/\theta$
- $\frac{\partial \log f(Y|\theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{Y}{\theta^2}$
- $\frac{\partial^2 \log f(Y|\theta)}{\partial\theta^2} = \frac{1}{\theta^2} - 2\frac{Y}{\theta^3}$
- $\mathbb{E}[-\frac{\partial^2 \log f(Y|\theta)}{\partial\theta^2}] = \frac{1}{\theta^2}$, since $\mathbb{E}[Y] = \theta$.

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.
- We know the MLE is $\hat{\theta} = \bar{Y}$
- $f(Y \mid \theta) = \frac{1}{\theta} e^{-Y/\theta}$ for $Y = 0, 1, 2, ...$
- $\log f(Y \mid \theta) = -\log \theta - Y/\theta$
- $\frac{\partial \log f(Y|\theta)}{\partial \theta} = -\frac{1}{\theta} + \frac{Y}{\theta^2}$
- $\frac{\partial^2 \log f(Y|\theta)}{\partial \theta^2} = \frac{1}{\theta^2} - 2\frac{Y}{\theta^3}$
- $\mathbb{E}[-\frac{\partial^2 \log f(Y|\theta)}{\partial \theta^2}] = \frac{1}{\theta^2}$, since $\mathbb{E}[Y] = \theta$.
- $\hat{\theta}_n \pm z_{\alpha/2}\sqrt{1/(n\mathbb{E}[-\frac{\partial^2 \log f(Y;\theta)}{\partial \theta^2}])} \mid_{\theta=\hat{\theta}_n}$.

Example

- For random variable with a Exponential distribution with parameter θ. If $Y_1, Y_2, ..., Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for θ.
- We know the MLE is $\hat{\theta} = \bar{Y}$
- $f(Y \mid \theta) = \frac{1}{\theta} e^{-Y/\theta}$ for $Y = 0, 1, 2, ...$
- $\log f(Y \mid \theta) = -\log\theta - Y/\theta$
- $\frac{\partial \log f(Y|\theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{Y}{\theta^2}$
- $\frac{\partial^2 \log f(Y|\theta)}{\partial\theta^2} = \frac{1}{\theta^2} - 2\frac{Y}{\theta^3}$
- $\mathbb{E}[-\frac{\partial^2 \log f(Y|\theta)}{\partial\theta^2}] = \frac{1}{\theta^2}$, since $\mathbb{E}[Y] = \theta$.
- $\hat{\theta}_n \pm z_{\alpha/2}\sqrt{1/(n\mathbb{E}[-\frac{\partial^2 \log f(Y;\theta)}{\partial\theta^2}])}\ |_{\theta=\hat{\theta}_n}$.
- The $100(1-\alpha)\%$ CI is $\bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\bar{Y}^2}{n}}$.

Summary

- General confidence intervals construction based on
  - pivotal quantities
  - MLE (likelihood function, Fisher information)

Homework

8.44, 8.102, 9.82 (b), 9.84 (a, d), 9.97, 9.102