

STAT604-MidTerm

Ashish Verma

2024-02-26

Problem 1

Solution 1

Likelihood Function

The probability density function (PDF) is given by:

$$f(x|\theta) = \frac{1}{2\theta^3} x^2 e^{-\frac{2}{\theta}x}$$

The likelihood function for a random sample x_1, x_2, \dots, x_n is:

$$L(\theta|x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{1}{2\theta^3} x_i^2 e^{-\frac{2}{\theta}x_i}$$

Taking the log-likelihood:

$$\ell(\theta|x_1, x_2, \dots, x_n) = \sum_{i=1}^n \left(-\log(2\theta^3) + 2\log(x_i) - \frac{2}{\theta}x_i \right)$$

The log-likelihood function is given by:

$$\ell(\theta|x_1, x_2, \dots, x_n) = -n\log(2\theta^3) + 2\sum_{i=1}^n \log(x_i) - \frac{2}{\theta}\sum_{i=1}^n x_i$$

Derivative with Respect to θ

Differentiating with respect to θ :

$$\frac{d}{d\theta}\ell(\theta|x_1, x_2, \dots, x_n) = -\frac{6n}{\theta} + \frac{4}{\theta^2}\sum_{i=1}^n x_i$$

Solve for θ

Setting the derivative equal to zero and solving for θ :

$$-6n\theta + 4 \sum_{i=1}^n x_i = 0$$

Multiplying through by θ^2 to clear the fraction:

$$-6n\theta^2 + 4 \sum_{i=1}^n x_i = 0$$

Solving for θ :

$$\theta_{\text{MLE}} = \frac{2}{3n} \sum_{i=1}^n x_i$$

Maximum Likelihood Estimate (MLE)

To find the MLE, we need to maximize the log-likelihood function.

```
# Given data
data <- c(7, 10, 7, 8, 9, 10, 9, 8, 10, 12, 14, 15, 12, 6, 3)

# Calculate MLE of theta
n <- length(data)
theta_mle <- (2 / (3 * n)) * sum(data)

# Print MLE
theta_mle
```

```
## [1] 6.222222
```

Problem 2

Solution 2 Set the first sample moment equal to the first population moment:

$$\bar{X} = \mu$$

Set the second sample moment around the mean equal to the second population moment around the mean:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{\sigma^2}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \sigma^2$$

First, let's substitute the given expressions for μ and σ^2 :

$$\bar{X} = \theta \cdot \frac{\pi}{2}$$

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \theta^2 \cdot \left(4 - \frac{\pi}{2}\right)$$

Now, let's solve these equations for θ .

For the second equation:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \theta^2 \cdot \left(4 - \frac{\pi}{2}\right)$$

Substitute the expression we found for θ :

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \left(\frac{\bar{X}}{\frac{\pi}{2}}\right)^2 \cdot \left(4 - \frac{\pi}{2}\right)$$

Simplify and solve for θ :

$$\theta = \frac{4 - \pi}{\frac{\pi}{2} \cdot \bar{X}}$$

Now, we have the method of moments estimator for θ :

$$\hat{\theta} = \frac{4 - \pi}{\frac{\pi}{2} \cdot \bar{X}}$$

Let's set up the equations:

1. The first sample moment around the mean (\bar{X}) is the mean (μ):

$$\bar{X} = \mu = \frac{\theta\pi}{2}$$

2. The second sample moment around the mean ($\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$) is the variance (σ^2):

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{\theta^2(4 - \pi)}{2}$$

Now, let's solve these equations for θ :

From the first equation:

$$\theta = \frac{2}{\pi} \bar{X}$$

Now substitute this into the second equation:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{2}{\pi} \bar{X}$$

Simplify and solve for θ :

$$\theta = \frac{4 - \pi}{\pi} \cdot \frac{1}{\bar{X}}$$

This is the method of moments estimator for θ . Now, let's check if it's unbiased:

$$\begin{aligned} E(\hat{\theta}) &= E\left(\frac{4 - \pi}{\pi} \cdot \frac{1}{\bar{X}}\right) \\ &= \frac{4 - \pi}{\pi} \cdot \frac{1}{\theta} \end{aligned}$$

Now, compare this with the true parameter θ :

$$\theta \neq \frac{4 - \pi}{\pi} \cdot \frac{1}{\theta}$$

Since $E(\hat{\theta})$ is not equal to θ , the method of moments estimator $\hat{\theta}$ is biased for the parameter θ .

Problem 3

Solution 3

```
# Given data
data <- c(3.8, 3.0, 4.0, 4.8, 3.0, 4.2, 3.5, 4.7, 4.4, 4.2, 4.3, 3.8, 3.3, 4.0, 3.8)

# Standard normal distribution approach
confidence_interval_z <- t.test(data, conf.level = 0.9)$conf.int

# Bootstrap resampling methods
set.seed(123) # Setting seed for reproducibility

# Standard normal bootstrap
bootstrap_samples_standard_normal <- replicate(1000, mean(sample(data, replace = TRUE)))
confidence_interval_bootstrap_standard_normal <- quantile(bootstrap_samples_standard_normal, c(0.05, 0.95))

# Basic bootstrap
bootstrap_samples_basic <- replicate(1000, mean(sample(data, replace = TRUE)))
confidence_interval_bootstrap_basic <- quantile(bootstrap_samples_basic, c(0.05, 0.95))

# Percentile bootstrap
bootstrap_samples_percentile <- replicate(1000, quantile(sample(data, replace = TRUE), c(0.025, 0.975)))
confidence_interval_bootstrap_percentile <- apply(bootstrap_samples_percentile, 1, mean)

# t bootstrap
t_statistic <- qt(c(0.025, 0.975), df = length(data) - 1)
bootstrap_samples_t <- replicate(1000, mean(sample(data, replace = TRUE)) + t_statistic * sd(sample(data, replace = TRUE)))
confidence_interval_bootstrap_t <- quantile(bootstrap_samples_t, c(0.05, 0.95))

# BCa bootstrap
library(boot)
bootstrap_bca <- boot(data, function(x, indices) mean(x[indices]), R = 1000)
```

```

confidence_interval_bootstrap_BCa <- boot.ci(bootstrap_bca, conf = 0.9, type = "bca")

# Print the results
cat("Confidence Interval (Standard Normal):", confidence_interval_z, "\n")

## Confidence Interval (Standard Normal): 3.670202 4.169798

cat("Confidence Interval (Bootstrap Standard Normal):", confidence_interval_bootstrap_standard_normal,

## Confidence Interval (Bootstrap Standard Normal): 3.68 4.146667

cat("Confidence Interval (Bootstrap Basic):", confidence_interval_bootstrap_basic, "\n")

## Confidence Interval (Bootstrap Basic): 3.68 4.146667

cat("Confidence Interval (Bootstrap Percentile):", confidence_interval_bootstrap_percentile, "\n")

## Confidence Interval (Bootstrap Percentile): 3.08309 4.678615

cat("Confidence Interval (Bootstrap t):", confidence_interval_bootstrap_t, "\n")

## Confidence Interval (Bootstrap t): 3.452491 4.403693

cat("Confidence Interval (Bootstrap BCa):", confidence_interval_bootstrap_BCa$bca[4:5], "\n")

## Confidence Interval (Bootstrap BCa): 3.693333 4.133333

```

Problem 4

Solution 4 Time magazine (July 15, 2002) quoted a poll of 1000 Americans in which only 5% said they were vegetarians. Let \hat{p} denote the sample proportion of vegetarians in the United States.

Construct a 98% confidence interval for the population proportion who are vegetarians.

The margin of error (E) for a confidence interval is given by:

$$E = Z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

where Z is the Z-score corresponding to the desired confidence level. For a 98% confidence interval, $Z = 2.33$ (approximately).

The confidence interval is then given by:

$$\text{Confidence Interval} = (\hat{p} - E, \hat{p} + E)$$

Substitute the given values to compute the interval.

```

# Given values
sample_proportion <- 0.05
confidence_level <- 0.98
Z_score <- qnorm((1 + confidence_level) / 2)

# Margin of error
margin_of_error <- Z_score * sqrt((sample_proportion * (1 - sample_proportion)) / 1000)

# Confidence interval
confidence_interval_lower <- sample_proportion - margin_of_error
confidence_interval_upper <- sample_proportion + margin_of_error

# Output
cat("98% Confidence Interval: (", confidence_interval_lower, ",", confidence_interval_upper, ")\n")

## 98% Confidence Interval: ( 0.03396675 , 0.06603325 )

```

Problem 5

Solution 5

```

# Load the necessary library for confidence intervals
library(stats)

# Load your dataset
alloy_data <- read.csv("C:\\Users\\Ashish\\Documents\\ODU\\STAT604\\MidTerm\\alloy.csv")

# Given confidence level
confidence_level <- 0.97

# Find a 97% confidence interval for the population standard deviation
sd_value <- sd(alloy_data$Tensile.Strength..MPa)
n <- length(alloy_data$Tensile.Strength..MPa)

# Calculate the standard error of the standard deviation
se_sd <- sd_value / sqrt(2 * (n - 1))

# Find the critical value for the t-distribution
critical_value <- qt((confidence_level + 1) / 2, df = n - 1)

# Calculate the margin of error
margin_of_error <- critical_value * se_sd

# Calculate the confidence interval for the population standard deviation
lower_bound_sd_ci <- sd_value - margin_of_error
upper_bound_sd_ci <- sd_value + margin_of_error

cat("97% CI for population standard deviation:", lower_bound_sd_ci, "to", upper_bound_sd_ci, "\n")

## 97% CI for population standard deviation: 143.9958 to 157.8537

```