Module 3. Estimation Techniques

(use Markdown Preview Github Styling on vscode if necessary)

Estimation in statistics involves inferring the characteristics of a population based on a sample. Various tools and techniques are employed to perform these estimations, each suited to different types of data and estimation goals. Below is a detailed overview of the key tools and techniques used for statistical estimation

1. Point Estimation

- Objective: Provide a single "best guess" value for a population
- parameter. Common Tools:
  - **Sample Mean ($\bar{x}$)**: Used to estimate the population mean.
  - **Sample Proportion ($\hat{p}$)**: Used to estimate the population proportion.
  - **Sample Variance (s^2)**: Used to estimate the population variance.
  - **Sample Standard Deviation (σ)**: Used to estimate the population standard deviation.
- Techniques:
  - **Method of Moments**: Estimates parameters by equating sample moments to population moments.
  - **Least Squares Estimation**: Often used in regression, it minimizes the sum of the squared differences between observed and predicted values.

Example:

Suppose you have the following sample data on the number of sales day for five days: **50,55,60,65,70.**

Sample Mean ($\bar{x}$)**= $\frac{50+55+60+65+70}{5} = 60$

Sample Proportion ($\hat{p}$)**= $\frac{(50-60)^2+(55-60)^2+(60-60)^2+(65-60)^2+(70-60)^2}{(5-1)} = 62.5$

critical value =

$$t_{\text{crit, two-tailed}} = t_{1-\frac{\alpha}{2}, df}$$

$$t_{\text{crit, one-tailed}} = t_{1-\alpha, df}$$

$$t_{\text{crit}} = t_{0.975, 4} \approx 2.776$$

R code:

```r
# Sample data
data <- c(50, 55, 60, 65, 70)
# Sample mean
mean(data)
# Sample variance
var(data)
# Sample standard deviation
sd(data)
```

2. Interval Estimation

- Objective: Provide a range of values (interval) within which the population parameter islikely to lie, along with a confidence level.

Common Tools:

1. **Confidence Interval (CI)** - An interval estimate for a population parameter, e.g., a 95% CI for the population mean.
2. **Prediction Interval** - An interval estimate for a single future observation.
3. **Tolerance Interval** - An interval within which a specified proportion of the population is expected to fall.

Techniques:

1. **Normal Distribution (Z-distribution)** - Used for large sample sizes or when the population standard deviation is known.
2. **T-Distribution** - Used for small sample sizes when the population standard deviation is unknown.
3. **Bootstrap Methods** - Non-parametric method that uses resampling with replacement to estimate the sampling distribution of a statistic.

Example:

$$
\begin{aligned}
\bar{X} \pm t_{a/2,n-1} * \frac{s}{\sqrt{n}} \\
\text{Using } t_{0.025,4} = 2.776 : CI \\
60 \pm 2.776 * \frac{8}{\sqrt{5}} \\
60 \pm 9.9 \\
\underline{[50.1, 69.9]}
\end{aligned}
$$

Where:

confidence level$(a) = 95\% = \frac{0.05}{2} = 0.025$

$60 \rightarrow$ mean

7.906 (8) → sample SD

0.025 → half of significance level (/ 2)

4 → degrees of freedom (n-1)

2.776 → critical t value

R code:

```
# Sample data
data <- c(50, 55, 60, 65, 70)
# Confidence interval for mean
t.test(data)$conf.int
```

3. Maximum Likelihood Estimation (MLE):

- **Objective** - Estimate the parameters of a probability distribution by maximizing the likelihood function.

Common Tools:

1. Likelihood Function- Represents the probability of observing the given sample data as a function of the parameters.
2. Log-Likelihood- The natural logarithm of the likelihood function, often used because it simplifies the math.

Techniques:

1. **Differentiation and Optimization** - The likelihood function is differentiated with respect to the parameters, and the resulting equations are solved to find the maximum likelihood estimates (MLEs).
2. **Numerical Methods** - Used when analytic solutions are difficult, such as the Newton-Raphson method.

Example 1. For a dataset of coin flips with heads(H) occurring 40 times out of 100 flips, the likelihood function for the probability p of heads is: L(p)=p40 x (1-p)60 Maximizing the likelihood: log L(p)=40 log p + 60 log (1-p) Differentiating and solving: $\hat{p} = \frac{40}{100} = 0.4$

R Code:

```
# Coin flip data (40 heads out of 100 flips)
library(stats4)

coin_likelihood <- function(p) {
return(-(40 * log(p) + 60 * log(1 - p))) # Negative Log Likelihood
}

mle(coin_likelihood, start = list(p = 0.5))
```

Example 1: Estimating the Mean of a Normal Distribution (Known Variance):

Problem:

Given the data: $X_1, X_2, ..., X_n = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\ exp - \frac{(X_i - \mu)^2}{2\sigma^2}$

The likelihood function is: $Log\ L(\mu) = -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$

Differentiating with respect μ and setting it to 0: $\frac{d}{du}\ log\ L(\mu) = -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)$

or just:

$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$

R code:

```
# sample data
data <- c(4.5, 5.1, 4.9, 5.2, 5.0)

# Mle for mean
mean(data)
```

Example 2. Estimating Mean and Variance of a Normal Distribution:

problem:

Given the data $X_1, X_2 ......, X_n\ N(\mu, \sigma)^2$ estimate both the population mean μ and variance σ

Paper and pen solution:

Given the data: $L(\mu, \sigma^2; X_1, X_2, ..., X_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\ exp - \frac{(X_i - \mu)^2}{2\sigma^2}$

The likelihood function is: $Log\ L(\mu, \sigma^2) = -\frac{n}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$

maximizing the respect to μ: $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Maximixing the respect of σ^2: $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$

or just: $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$.

R code:

```
# Sample data
data <- c(4.5, 5.1, 4.9, 5.2, 5.0)

# MLE for mean
mle_mean <- mean(data)

# MLE for variance (divide by n instead of n-1)
mle_variance <- var(data)  * (length(data)-1)/length(data)
```

```
# print out the result
mle_mean
mle_variance
```

In latex/paper:

$$\text{using the formula for} \hat{\mu}:$$
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$\text{for } \sigma^2:$$
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

$$\text{data: [4.5, 5.1, 4.9, 5.2, 5.0]}$$
$$\text{First step is to get the mean:}$$
$$\hat{\mu} = \frac{4.5 + 5.1 + 4.9 + 5.2 + 5.0}{5}$$
$$\hat{\mu} = \frac{24.7}{5}$$
$$\underline{\hat{\mu} = 4.94}$$

$$\text{now for } \sigma^2:$$
$$\sigma^2 = \frac{(4.5 - 4.94)^2 + (5.1 - 4.94)^2 + (4.9 - 4.94)^2 + (5.2 - 4.94)^2 + (5.0 - 4.94)^2}{5}$$
$$\sigma^2 = \frac{0.1936 + 0.0256 + 0.0016 + 0.0676 + 0.0036}{5}$$
$$\sigma^2 = \frac{0.292}{5}$$
$$\sigma^2 = 0.0584$$

Example 3. Estimating the Variance (Known Mean)

Given the data $X_1, X_2......, X_n$ $N(\mu, \sigma^2)$

Pen and Paper solution:

The likelihood function is: $Log\ L(\mu) = -\frac{1}{n}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$

Differentiating with the respect to σ^2 : $log\ L(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$ solving for σ^2 : $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$.

or just:

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

R code:

```
data <- c(4.5, 5.1, 4.9, 5.2, 5.0)
mu <- 5

# MLE for variance (known mean)
mle_variance <- mean((data - mu)^2)
mle_variance
```

manual:

$$\text{using the formula}$$
$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

$$data = [4.5, 5.1, 4.9, 5.2, 5.0], \quad \mu = 5, \quad n = 5$$
$$\hat{\sigma^2} = \frac{(4.5-5)^2 + (5.1-5)^2 + (4.9-5)^2 + (5.2-5)^2 + (5.0-5)^2}{5}$$
$$\hat{\sigma^2} = \frac{0.31}{5}$$
$$\hat{\sigma^2} = 0.062$$

Example 4. Estimating Parameters for a Normal Distribution with Log-Likelihood

Problem:

Estimate μ and σ^2 for a normal distribution using the log likelihood function.

pen and paper solution:

The log likelihood function for $X_1, X_2, ..., X_n\ \ N(\mu, \sigma^2)$: $log\ \ L(\mu, \sigma^2) = -\frac{n}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{2}(X_i - \mu)^2$

Taking partial derivatives and solving yields the MLEs for μ and σ^2.

R code:

```
library(stats4)

data <- c(4.5, 5.1, 4.9, 5.2, 5.0)

log_likelihood  <- function(mu, sigma) {
  -sum(dnorm(data, mean = mu, sd = sigma, log = TRUE))
}
```

```r
# MLE Estimation
mle(log_likelihood, start = list(mu = mean(data), sigma = sd(data)))
```

fixed code:

```r
library(stats4)

data <- c(4.5, 5.1, 4.9, 5.2, 5.0)

# Log-likelihood with positive sigma ensured
log_likelihood  <- function(mu, log_sigma) {
  sigma <- exp(log_sigma)  # ensure sigma > 0
  -sum(dnorm(data, mean = mu, sd = sigma, log = TRUE))
}

# MLE Estimation
mle(log_likelihood, start = list(mu = mean(data), log_sigma = log(sd(data))))
```

manual:

$$\text{using the formula for} \hat{\mu}:$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\text{for } \sigma^2:$$

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

$$\text{data: } [4.5, 5.1, 4.9, 5.2, 5.0]$$

$$\text{First step is to get the mean:}$$

$$\hat{\mu} = \frac{4.5 + 5.1 + 4.9 + 5.2 + 5.0}{5}$$

$$\hat{\mu} = \frac{24.7}{5}$$

$$\underline{\hat{\mu} = 4.94}$$

now for σ^2 :

$$\sigma^2 = \frac{(4.5-4.94)^2 + (5.1-4.94)^2 + (4.9-4.94)^2 + (5.2-4.94)^2 + (5.0-4.94)^2}{5}$$

$$\sigma^2 = \frac{0.1936 + 0.0256 + 0.0016 + 0.0676 + 0.0036}{5}$$

$$\sigma^2 = \frac{0.292}{5}$$

$$\sigma^2 = 0.0584$$

$$\text{For log } \hat{\sigma} : \hat{\sigma} = \sqrt{\hat{\sigma}^2}, \widehat{\log \sigma} = \ln(\hat{\sigma})$$

$$\hat{\sigma} = \sqrt{0.0584}$$

$$\hat{\sigma} \approx 0.2415$$

$$\widehat{\log \sigma} = \ln(0.2415)$$

$$\widehat{\log \sigma} \approx -1.4209$$

Example 5. MLE for Normal Distribution (Multivariate Case)

Problem:

Given a bivariate normal distribution $X \sim N(\mu, \sum)$, where $\mu = (\mu_1, \mu_2)$ and $\sum$ is the covariance matrix, estimate μ_1, μ_2 *and* $\sum$

Pen and paper solution:

the log-likelihood for a multivariate normal distribution is: $log\ L(\mu, \sum) = -\frac{n}{2} log|\sum| - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^T \sum^{-1}(X_i - \mu)$

the mle estimates for μ are $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$

the mle estimate for $\sum$ is $\hat{\sum} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)(X_i - \hat{\mu})^T$

R code:

```
# Sample bivariate data
data <- matrix(c(4.5, 5.1, 4.9, 5.2), ncol = 2)

# MLE for mean and covariance matrix
mle_mean <- colMeans(data)
mle_covariance <- cov(data)
mle_mean
mle_covariance
```

Fixed for MLE code:

```
# Sample bivariate data
data <- matrix(c(4.5, 5.1, 4.9, 5.2), ncol = 2)
```

```
# MLE for mean and covariance matrix
mle_mean <- colMeans(data)
mle_covariance <- cov(data) * (nrow(data)-1)/mle_mean
mle_covariance
nrow(data)
```

manual on paper example:

$$Data = \begin{bmatrix} 4.5 & 5.1 \\ 4.9 & 5.2 \end{bmatrix}$$

Computing for mean:

$$\begin{aligned} \text{Column 1: } & 4.5, 4.9 \\ \bar{x}_1 = & \frac{4.5+4.9}{2} \\ barx_1 = & \frac{9.4}{2} \\ & 4.7 \\ \text{column 2: } & 5.1, 5.2 \\ \bar{x}_2 = & \frac{5.1+5.2}{2} \\ \bar{x}_2 = & \frac{10.3}{2} \\ & 5.15 \\ \text{Result: } \hat{\mu} = & \begin{bmatrix} 4.7 \\ 5.15 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} & \text{subtract each mean for each row} \\ x_1 - \hat{\mu} = & \begin{bmatrix} 4.5 - 4.7 \\ 5.1 - 5.15 \end{bmatrix} = \begin{bmatrix} -0.2 \\ -0.05 \end{bmatrix} \\ x_2 - \hat{\mu} = & \begin{bmatrix} 4.9 - 4.7 \\ 5.2 - 5.15 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.05 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} & \text{Compute outer products} \\ (x_1 - \hat{\mu})(x_1 - \hat{\mu})^T = & \begin{bmatrix} -0.2 \\ -0.05 \end{bmatrix} \begin{bmatrix} -0.2 & -0.05 \end{bmatrix} = \begin{bmatrix} 0.04 & 0.01 \\ 0.01 & 0.0025 \end{bmatrix} \\ (x_2 - \hat{\mu})(x_2 - \hat{\mu})^T = & \begin{bmatrix} 0.2 \\ 0.05 \end{bmatrix} \begin{bmatrix} 0.2 & 0.05 \end{bmatrix} = \begin{bmatrix} 0.04 & 0.01 \\ 0.01 & 0.0025 \end{bmatrix} \end{aligned}$$

$$\hat{\Sigma} = \frac{1}{2} \left(\begin{bmatrix} 0.04 & 0.01 \\ 0.01 & 0.0025 \end{bmatrix} + \begin{bmatrix} 0.04 & 0.01 \\ 0.01 & 0 \end{bmatrix} \right) = \frac{1}{2} \begin{bmatrix} 0.08 & 0.02 \\ 0.02 & 0.005 \end{bmatrix} = \begin{bmatrix} 0.04 & 0.01 \\ 0.01 & 0.0025 \end{bmatrix}$$

MLE covariance matrix is:

$$\hat{\sum} = \begin{bmatrix} 0.04 & 0.01 \\ 0.01 & 0.0025 \end{bmatrix}$$

4. Bayesian Estimation

Objective- Estimate parameters by combining prior information with data, resulting in a posterior distribution.

Common Tools::

1. Prior Distribution- Represents the initial beliefs about the parameters before observing the data.
2. Likelihood Function- Represents the probability of the observed data given the parameters.
3. Posterior Distribution- Combines the prior distribution and likelihood to provide updated beliefs after observing the data.

Techniques:

1. Bayes' Theorem- Used to update the prior distribution in light of the observed data to obtain the posterior distribution.
2. Markov Chain Monte Carlo (MCMC)- A computational technique to sample from the posterior distribution when it cannot be calculated directly.

5. Resampling Methods

Objective: Estimate the sampling distribution of a statistic by repeatedly sampling from the data.

Common Tools:

1. Bootstrap- Involves sampling with replacement from the original sample to create "bootstrap samples" and then estimating the statistic of interest across these samples.
2. Jackknife- Involves systematically leaving out one observation at a time from the sample set and recalculating the estimate.

Techniques:

1. Bootstrap Confidence Intervals- Construct confidence intervals by using the distribution of bootstrap samples.
2. Jackknife Estimation- Provides an estimate of bias and variance by systematically recalculating the statistic with one observation omitted at a time.

6. Bayesian Methods

Objective- Estimate population parameters by incorporating prior knowledge or beliefs with the observed data.

Common Tools:

1. Prior Distributions- These reflect what is known about a parameter before observing the data.
2. Posterior Distributions- These combine the prior distribution with the likelihood of the observed data to update the belief about the parameter.
3. Bayesian Inference- Draws conclusions about parameters based on the posterior distribution.

Techniques:

1. Bayes' Theorem: Used to update the probability estimate for a hypothesis as more evidence or information becomes available.
2. MCMC (Markov Chain Monte Carlo): A class of algorithms used to sample from the posterior distribution when it is complex and not analytically tractable.

7. Non-Parametric Methods

Objective- Estimate population parameters without assuming a specific parametric form for the distribution.

Common Tools:

1. Kernel Density Estimation: Estimates the probability density function of a random variable.
2. Empirical Cumulative Distribution Function (ECDF): Estimates the cumulative distribution function without assuming a specific distribution.

Techniques:

1. Bootstrap Methods- Resampling technique that does not assume any particular distribution.
2. Rank-Based Methods- Estimation methods that rely on the ranks of data rather than their actual values.

8. Moment Estimation

Objective- Estimate parameters by equating sample moments (mean, variance, etc.) with population moments.

Common Tools:

1. Sample Moments- Calculated directly from the data.
2. Population Moments- Defined by the distribution and its parameters.

Techniques:

1. Method of Moments- Set sample moments equal to theoretical moments and solve for the parameters.

9. Monte Carlo Methods

Objective: Use random sampling and statistical modeling to estimate mathematical or physical systems.

Common Tools:

1. Monte Carlo Simulation- Uses random sampling to obtain numerical results, often used to estimate complex functions or distributions.

Techniques:

1. Simulating Random Variables: Generates random variables according to a specified distribution.
2. Estimating Probabilities: Uses repeated random sampling to estimate the probability of an event.

10. Goodness-of-Fit Tests

Objective- Evaluate how well a statistical model fits the observed data.

Common Tools:

1. Chi-Square Test- Compares observed and expected frequencies.
2. Kolmogorov-Smirnov Test- Compares the empirical distribution function with a specified distribution.

Techniques:

1. Chi-Square Goodness-of-Fit- Tests whether a sample comes from a population with a specific distribution.
2. Kolmogorov-Smirnov Test- Non-parametric test to compare the observed data distribution with a theoretical one.

Summary:

Estimation in statistics encompasses a variety of tools and techniques, each suited to different types of data and estimation needs. The choice of tool and technique depends on the nature of the data, the underlying assumptions, and the specific goals of the analysis. Mastery of these techniques is important for making accurate and reliable inferences in statistical practice.