LES FACULTÉS
DE L'UNIVERSITÉ
CATHOLIQUE DE LILLE

Inference & Estimation theory

PARAMETER ESTIMATION & LEARNING

Baptiste Mokas

baptiste.mokas@gmail.com

weeki.io/dynamical-système

linktr.ee/baptistemokas

+33 7 69 08 54 19

KNOWLEDGE TREE

Part 1: Foundations and Principles of Parameter Estimation

1.1 Introduction to Parameter Estimation

1.2 Types and Methods of Estimation

- Estimators, Estimates, and their properties
- Point and Interval Estimation
- Confidence Intervals: Construction and Interpretation
- Exact and Asymptotic Confidence Regions and Intervals
- Types of parameters: Mean, variance, proportion, and rate

1.3 Parametric Families and Empirical Distributions

1.4 Criteria for Good Estimators

- Unbiasedness, consistency, and efficiency
- Mean squared error and its components

Part 2: Methods and Properties of Estimators

2.1 Detailed Exploration of Estimation Methods

- Method of Moments
- Maximum Likelihood Estimation (MLE) and its asymptotic analysis
- Bayesian Estimation / Bayes' Theorem
- Least Squares Estimation

2.2 Properties of Estimators

- Bias, Variance, and Efficiency
- Sufficiency, Completeness, and Exponential Families
- Cramér-Rao Inequality in different models

2.3 Advanced Topics

- Lehmann-Scheffé Theorem
- Score, Fisher Information, and Regular Models

4.1 Nonparametric Estimation Techniques

- Kernel Density Estimation
- Nonparametric Regression
- Bootstrap Resampling

Part 3: Practical Considerations and Challenges

3.1 Computational Aspects

- Numerical optimization techniques
- Challenges in high-dimensional parameter spaces

3.2 Robust Estimation and Handling Outliers

- Resistant measures
- Heavy-tailed distributions

3.3 Data Challenges

- Handling Missing Data: Mechanisms and Impacts
- Reporting and Interpretation: Clear communication and caveats

Part 4: Exercises, Advanced Topics, and Problem Solutions

4.1 Construction and Comparison of Estimators

- Efficient estimator construction
- Comparing Estimators using the Quadratic Risk Function and other metrics

4.2 Advanced Statistical Concepts

- Improved Rao-Blackwell Estimators
- Exhaustive Statistic and Complete Statistic

4.3 Exercises and Problem Solutions

- Set of exercises to test understanding
- Problems related to parameter estimation
- Solutions for exercises and problems

OUR STORY LINE

Probability and Statistics

- STEP -1 _ PROGRAM INTRODUCTION
- STEP 0 _ FOUNDATIONS
- STEP 1 _ THEORY OF SYSTEMS
- STEP 2 _ STOCHASTIC DYNAMICS & PROBABILITY
- STEP 3 _ DATA OBSERVATION
- STEP 4 _ INFERENCE & ESTIMATION THEORY
- STEP 5 _ LINEAR MODEL EXAMPLES
- STEP 6 _ OTHER MODEL EXAMPLES
- STEP 7 _ NON LINEAR MODELS

OUR STORY LINE

Probability and Statistics

STEP -1 _ PROGRAM INTRODUCTION
- -1 - PROGRAM INTRODUCTION

STEP 0 _ FOUNDATIONS
- 0.1 - ELEMENTS OF CALCULUS & TOOLS
- 0.2 - EPISTEMOLOGY & THEORY OF KNOWLEDGE

STEP 1 _ THEORY OF SYSTEMS
- 1.1 - DYNAMICAL SYSTEMS
- 1.2 - COMPLEX ADAPTIVE SYSTEMS

STEP 2 _ STOCHASTIC DYNAMICS & PROBABILITY
- 2.1 - MEASURE THEORY
- 2.2 - PROBABILITY THEORY
- 2.3 - USUAL PROBABILITY DISTRIBUTIONS
- 2.4 - ASYMPTOTIC STATISTICS
- 2.5 - STOCHASTIC PROCESS & TIME SERIES
- 2.6 - INFORMATION GEOMETRY

STEP 3 _ DATA OBSERVATION
- 3.1 - DESCRIPTIVE STATISTICS & DATAVIZUALISATION
- 3.2 - EXPLORATORY DATA ANALYSIS

STEP 4 _ INFERENCE & ESTIMATION THEORY
- 4.1 - PARAMETERS ESTIMATIONS & LEARNING
- 4.2 - EXPERIMENTAL DESIGN & HYPOTHESIS TESTING
- 4.4 - DECISION TREES & MODEL SELECTION
- 4.5 - BAYESIAN INFERENCES

STEP 5 _ LINEAR MODEL EXAMPLES
- 5.1 - SIMPLE LINEAR REGRESSION
- 5.2 - MULTIPLE LINEAR REGRESSION
- 5.3 - OTHER REGRESSIONS MODELS

STEP 6 _ OTHER MODEL EXAMPLES
- 6.1 - USUAL UNIVARIATE TESTING
- 6.2 - USUAL MULTIVARIATE TESTING
- 6.3 - NON PARAMETRIC STATISTICS

STEP 7 _ NON LINEAR MODELS
- 7.1 - PROBABILISTIC GRAPHICAL MODELS
- 7.2 - PERCOLATION THEORY
- 7.3 - SPATIAL STATISTICS
- 7.4 - EXTREM VALUE THEORY
- 7.5 - AGENT BASED MODELING
- 7.6 - NETWORK DYNAMICS

Role and importance of parameter estimation

Part 1: Foundations and Principles of Parameter Estimation

1.1 Introduction to Parameter Estimation

Parameters to be Estimated:
- θ : This is the vector of unknown parameters we aim to estimate. It represents the fundamental characteristics of the statistical or probabilistic model.

Observed Data:
- $\mathbf{X_1, X_2, \ldots, X_n}$: These are random variables comprising a random sample of size $\mathbf{n}$ drawn from the underlying population.

Observed Values:
- $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$: These are the actual values taken by the random variables in the observed sample.

Parameter Estimator:
- $\hat{\theta}$ (theta-hat): This is the parameter estimator, which is a statistic calculated from the observed data and is used to estimate the values of $\boldsymbol{\theta}$. It can be a single value or a set of values, depending on the type of estimation.

Parameter: A parameter (θ) is a fixed and unknown characteristic of a population or a statistical distribution.

Statistic: A statistic (S) is a measurable quantity calculated from a sample of data. It provides information about the sample and can vary from one sample to another.

Estimators, Estimates, and their properties

Part 1: Foundations and Principles of Parameter Estimation

1.2 Types and Methods of Estimation

Estimators:

- An estimator, denoted as θ (theta-hat), is a statistic used to estimate an unknown parameter θ.
- The mathematics behind an estimator involves a function, say g, applied to the observed data $X_1, X_2, \ldots, X_n$ Mathematically, an estimator can be represented as: $\theta = g(X_1, X_2, \ldots, x_n)$.
- Some commonly used estimators include the sample mean, sample variance, maximum likelihood estimator (MLE), and method of moments estimator (MME).
- The goal is to find an estimator that is unbiased, efficient, and consistent. In mathematical terms, we aim for $E(\theta) = \theta, \text{where } E(\theta)$ represents the expected value of the estimator, θ.

Estimates:

- An estimate, often denoted as θ (theta-hat), is the specific numerical value produced by an estimator after plugging in the observed data. It is an estimate of the true population parameter θ.
- For example, if the sample mean is used as an estimator for the population mean μ, the estimate is $\theta = \bar{x}$ (the sample mean).
- The estimate can vary from one sample to another, as it depends on the specific data observed in each case. The estimate provides a point estimate or a range of values (confidence interval) for the parameter θ.

Properties of Estimators:

- Unbiasedness: An estimator is unbiased if, on average, it gives the correct value of the parameter. Mathematically, $\text{E}(\theta) = \theta$.
- Efficiency: An efficient estimator has the smallest possible variance among all unbiased estimators for the parameter. In other words, it provides precise estimates.
- Consistency: A consistent estimator converges in probability to the true parameter value as the sample size increases. Mathematically, as n (sample size) goes to infinity, θ converges in probability to θ.
- Sufficiency: A sufficient estimator provides all the information in the data about the parameter θ.
- Minimum Variance: The minimum variance estimator is the one that has the smallest possible variance among all estimators. The Cramér-Rao lower bound sets the theoretical limit for the variance of unbiased estimators.

These properties are essential for evaluating the quality of estimators and their estimates, ensuring that they are reliable, accurate, and informative in statistical analysis.

Point and Interval Estimation

Part 1: Foundations and Principles of Parameter Estimation

1.2 Types and Methods of Estimation

Point Estimation:

- Definition: Point estimation involves using a statistic, denoted as θ (theta-hat), to estimate an unknown population parameter θ. Mathematically, θ is a single numerical value derived from the sample data, representing the best guess for θ.

Interval Estimation:

- Definition: Interval estimation, often expressed as a confidence interval, provides a range $[L, U]$ such that we are confident that the unknown parameter θ falls within this interval with a specified confidence level. Mathematically, $[L, U]$ is the interval within which θ is estimated to lie.

These definitions offer a succinct mathematical understanding of point estimation and nterval estimation in statistics.

Confidence Intervals: Construction and Interpretation

Construction of a Confidence Interval:

Select a Significance Level (a):

- Choose a significance level, a, which represents the probability of making a Type I error (rejecting a true null hypothesis). Common choices are 0.05 or 0.01 .

Determine the Distribution:

- Identify the probability distribution of the sample statistic that corresponds to the parameter you want to estimate. For example, if estimating the population mean, use the normal distribution for large samples or the t-distribution for small samples.

Calculate the Standard Error:

- Calculate the standard error (SE) of the sample statistic, which quantifies the variability of the estimator. For the population mean, $\mathrm{SE} = \sigma/\sqrt{\mathrm{n}}$, where σ is the population standard deviation, and n is the sample size.

Find the Critical Value(s):

- Based on the chosen a and the distribution (e.g., Z for normal or t for t-distribution), find the critical value(s) that corresponds to the desired level of confidence. Critical values are denoted as $\mathrm{z}^{\star}$ or $\mathrm{t}^{\star}$.

Compute the Margin of Error (ME):

- The margin of error (ME) is calculated as $\mathrm{ME} =$ critical value * standard error.

Construct the Confidence Interval:

- Finally, the confidence interval is constructed as (point estimate - ME, point estimate + ME). For estimating the population mean μ, the confidence interval is often written as: (x - ME, $\bar{x} + ME$).

Interpretation of a Confidence Interval:

The confidence interval provides a range of values within which the true population parameter is estimated to lie with a specific level of confidence, represented by $(1 - a) * 100\%$.

For example, a 95% confidence interval implies that, on average, if you were to take many random samples and compute confidence intervals, approximately 95% of those intervals would contain the true parameter.

If the interval does not contain a specific value (e.g., 0 for a difference in means), you can infer that it is statistically significant at the chosen significance level a.

The narrower the confidence interval, the more precise the estimate, and the wider it is, the less precise.

Exact and Asymptotic Confidence Regions and Intervals

Exact Confidence Regions and Intervals:

Exact Confidence Region: A precise region in which the parameter is estimated to lie with a certain level of confidence. Exact calculations depend on the specific data distribution.

Exact Confidence Interval: A precise one-dimensional interval estimate for a parameter. Construction depends on the data distribution.

Asymptotic Confidence Regions and Intervals:

Asymptotic Confidence Region: A region constructed using asymptotic approximation, often based on the Central Limit Theorem for large samples.

Asymptotic Confidence Interval: An interval constructed using asymptotic approximation, particularly useful for large samples.

Interpretation involves understanding the confidence level $(1 - a)$ and the likelihood of the parameter being within the region or interval. Exact methods are tailored to specific data distributions, while asymptotic methods are suitable for large samples.

Types of parameters: Mean, variance, proportion, and rate

Part 1: Foundations and Principles of Parameter Estimation

1.2 Types and Methods of Estimation

Population Mean (μ) :

- The population mean, often denoted as μ, represents the average or expected value of a random variable in a population.
- Equation:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Where:
- μ : Population mean
- N : Total number of observations in the population
- x_i : Individual data points in the population

Population Variance (σ^2) :

- The population variance, denoted as σ^2, quantifies the spread or dispersion of data points in a population.
- Equation:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

Where:
- σ^2 : Population variance
- N : Total number of observations in the population
- x_i : Individual data points in the population
- μ : Population mean

Types of parameters: Mean, variance, proportion, and rate

Part 1: Foundations and Principles of Parameter Estimation

1.2 Types and Methods of Estimation

Sample Mean $(\bar{x})$:

- The sample mean, denoted as $\bar{x}$, represents the average of observed values in a sample.
- Equation:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Where:

- $\bar{x}$: Sample mean
- n : Sample size
- x_i : Individual data points in the sample

Sample Variance (s^2) :

- The sample variance, denoted as s^2, quantifies the spread of data points in a sample.
- Equation:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Where:

- s^2 : Sample variance
- n : Sample size
- x_i : Individual data points in the sample
- $\bar{x}$: Sample mean

Types of parameters: Mean, variance, proportion, and rate

Part 1: Foundations and Principles of Parameter Estimation

1.2 Types and Methods of Estimation

Population Proportion (p):

- The population proportion, denoted as p, represents the proportion of a specific characteristic in a population.
- Equation:

$$p = \frac{\text{Number of individuals with the characteristic}}{\text{Total population size}}$$

Population Rate (λ) :

- The population rate, denoted as λ, is often used in the context of Poisson processes and represents the average number of events occurring in a given time interval.
- Equation:

$$\lambda = \frac{\text{Total number of events}}{\text{Total time interval}}$$

Sample Proportion ($\hat{p}$) :

- The sample proportion, denoted as $\hat{p}$, represents the proportion of a specific characteristic in a sample.
- Equation:

$$\hat{p} = \frac{\text{Number of individuals with the characteristic in the sample}}{n}$$

Where:
- $\hat{p}$: Sample proportion
- n : Sample size

Sample Rate ($\hat{\lambda}$) :

- The sample rate, denoted as $\hat{\lambda}$, is used in the context of a sample from a Poisson process and represents the average number of events occurring in a given time interval in the sample.
- Equation:

$$\hat{\lambda} = \frac{\text{Total number of events in the sample}}{\text{Total time interval in the sample}}$$

Empirical Distribution & Parametric Families

Part 1: Foundations and Principles of Parameter Estimation

1.3 Parametric Families and Empirical Distributions

Empirical Distribution:

Definition: The empirical distribution is a probability distribution derived directly from observed data. It assigns a probability mass to each unique data point based on its frequency of occurrence in the sample.

Mathematics:

- For a sample dataset with observed values $x_1, x_2, \ldots, x_n$, the probability mass function of the empirical distribution is given by:

$$P(X = x_i) = \frac{\text{Number of occurrences of } x_i}{n} \text{ for } i = 1, 2, \ldots, n$$

- Here, $P(X = x_i)$ represents the probability of observing x_i, and n is the sample size.

Parametric Families:

Definition: Parametric families are sets of probability distributions characterized by a finite number of parameters. These distributions are used to model data when certain assumptions about the data's underlying distribution are made.

Mathematics:

- A parametric distribution is typically defined by a probability density function (PDF) or a probability mass function (PMF) with specific mathematical expressions. For example, the normal distribution, a commonly used parametric family, has the PDF:

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- In this equation, μ and σ^2 are parameters that characterize the distribution.
- Parametric families often involve additional parameters for shape, location, and scale, and the choice of the family depends on the specific characteristics of the data and the underlying assumptions.

The empirical distribution provides a non-parametric way to estimate the distribution of data, directly from observed values, without making specific assumptions about the form of the distribution. In contrast, parametric families rely on specific mathematical expressions with parameters to describe the data distribution based on assumed characteristics.

Unbiasedness, consistency, and efficiency

Unbiasedness:

An estimator $\hat{\theta}$ is unbiased for a parameter θ if:

$$E(\hat{\theta}) = \theta$$

Consistency:

An estimator $\hat{\theta}$ is consistent for a parameter θ if, as the sample size n approaches infinity, the estimator approaches the true parameter in probability:

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \epsilon) = 1$$

Efficiency:

Efficiency (η) of an estimator is inversely proportional to its variance. An estimator is more efficient if it has a smaller variance and can be calculated as:

$$\eta = \frac{1}{\mathrm{Var}(\hat{\theta})}$$

- Smaller η values indicate higher efficiency.

In summary, unbiasedness is determined by comparing the expected value of the estimator to the true parameter. Consistency examines how the estimator behaves as the sample size increases. Efficiency is evaluated by comparing the variance of the estimator to the CramérRao lower bound, with smaller η values indicating higher efficiency.

Mean squared error and its components

Part 1: Foundations and Principles of Parameter Estimation

1.4 Criteria for Good Estimators

Mean Squared Error (MSE):

Definition: The Mean Squared Error (MSE) is a measure of the average squared difference between estimated values and the true values, often used to assess the quality of an estimator.

Mathematics:

- The MSE is calculated as the expected value of the squared difference between the estimator $(\hat{\theta})$ and the true parameter (θ) :

$$\text{MSE} = E\left((\hat{\theta} - \theta)^2\right)$$

Components of the MSE:

Bias:

- Definition: Bias measures the systematic error in an estimator, indicating how much the estimator tends to overestimate or underestimate the true parameter.
- Mathematics:
- The bias of an estimator $\hat{\theta}$ for a parameter θ is calculated as:

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Variance:

- Definition: Variance measures the spread or variability of the estimator's values around its expected value.
- Mathematics:
- The variance of an estimator $\hat{\theta}$ is calculated as:

$$\text{Var}(\hat{\theta}) = E\left((\hat{\theta} - E(\hat{\theta}))^2\right)$$

MSE as a Combination:

- The MSE can be expressed as the sum of the variance and the square of the bias:

$$\text{MSE} = \text{Var}(\hat{\theta}) + [\text{Bias}(\hat{\theta})]^2$$

In summary, the Mean Squared Error (MSE) quantifies the overall error of an estimator by considering its bias and variance. Bias measures systematic error, variance measures random error, and the MSE combines these two components to provide a comprehensive assessment of estimator quality.

Method of Moments

Method of Moments (MoM) Estimator:

Definition: The Method of Moments (MoM) is a method for estimating population parameters by equating sample moments to population moments. MoM estimators are obtained by setting equations involving moments equal to sample moments.

Mathematics:

- The MoM estimator $(\hat{\theta})$ for a parameter (θ) is obtained by solving equations that equate sample moments to population moments. These equations are typically set up based on the moments of the probability distribution being estimated.

Components of MoM Estimator:

Moment Equations:
- Definition: Moment equations are mathematical expressions that relate population moments (e.g., means, variances) to sample moments.
- Mathematics:
- Moment equations vary depending on the specific parameter and distribution being estimated. For example, for the mean (μ), the first moment equation is: Sample Mean = Population Mean
- Solving these equations for the unknown parameters yields the MoM estimators. Estimation Process:
- Definition: The estimation process involves solving the moment equations to obtain MoM estimators.
- Mathematics:
- The specific algebraic or numerical methods used to solve the moment equations depend on the complexity of the equations and the parameter being estimated.

Method of Moments

Example: Estimating the Population Mean (μ) using MoM
Suppose we have a random variable X with an unknown population mean (μ) that we want to estimate using the MoM.

Set Up the Moment Equation:
- For the population mean (μ), the first population moment is μ itself, i.e., $\mathrm{E}(X) = \mu$. Collect a Sample:
- Collect a random sample of n observations from the population: $x_1, x_2, \ldots, x_n$. Equate Sample Moment to

Population Moment:
- Equate the sample mean to the population mean:
Sample Mean $= \frac{1}{n} \sum_{i=1}^{n} x_i = \mu$

Solve for the Estimator:
- Solve for the unknown parameter (μ) by isolating it on one side of the equation:

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

Here, $\hat{\mu}$ represents the MoM estimator for the population mean (μ).

Interpretation:
- The MoM estimator $\hat{\mu}$ is the sample mean, which is calculated by taking the average of the observed data points.

In this example, the MoM estimator for the population mean (μ) is simply the sample mean ($\bar{x}$). The method of moments estimates the population parameter by equating the first moment of the population to the corresponding sample moment.

Maximum Likelihood Estimation (MLE) and its asymptotic analysis

Part 2: Methods and Properties of Estimators

2.1 Detailed Exploration of Estimation Methods

Maximum Likelihood Estimation (MLE) is a method used to estimate the parameters of a statistical model by maximizing the likelihood function. The likelihood function represents the probability of observing the given data under a particular set of parameter values. Here are the key steps and formulas for MLE:

Likelihood Function:

For a statistical model with a parameter (or a set of parameters) denoted as θ, the likelihood function $L(\theta)$ is defined as the probability of observing the given data x for the given parameter values θ.

$$L(\theta) = f(x \mid \theta)$$

- $L(\theta)$ represents the likelihood function.
- $f(x \mid \theta)$ is the probability density function (PDF) or probability mass function (PMF) of the data x given the parameter θ.

Log-Likelihood Function:

To simplify the calculations, we often work with the log-likelihood function, which is the natural logarithm of the likelihood function:

$\ln(L(\theta))$

- $\ln(L(\theta))$ represents the log-likelihood function.

Maximizing the Log-Likelihood:

To find the MLE, we maximize the log-likelihood function with respect to the parameter(s). This is done by taking the derivative of the log-likelihood with respect to the parameter(s) and setting it equal to zero.

$$\frac{d}{d\theta}\ln(L(\theta)) = 0$$

- This equation represents the condition for maximizing the log-likelihood.

Solve for the parameter(s) by finding the root of the equation. This is often done numerically because analytical solutions may not always exist.

Maximum Likelihood Estimation (MLE) and its asymptotic analysis

Part 2: Methods and Properties of Estimators

2.1 Detailed Exploration of Estimation Methods

MLE Estimators:

The solutions to the equation $\frac{d}{d\theta}\ln(L(\theta)) = 0$ are the MLE estimators for the parameter(s). These estimators, denoted as $\hat{\theta}$, represent the values of the parameter(s) that maximize the likelihood of the observed data.

$$\hat{\theta} = \arg\max \ln(L(\theta))$$

Interpretation:

The MLE estimators represent the parameter values that make the observed data the most probable under the assumed statistical model. MLE is a powerful and widely used method for parameter estimation in various fields of statistics and machine learning.

Maximum Likelihood Estimation (MLE) and its asymptotic analysis

Part 2: Methods and Properties of Estimators

2.1 Detailed Exploration of Estimation Methods

Example: Estimating the Probability Parameter (p) of a Binomial Distribution using MLE

Suppose you have observed a sequence of independent, identical trials, and you want to estimate the probability of success (p) for each trial using MLE.

Likelihood Function:

- For a binomial distribution, the likelihood function is given by:

$$L(p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Where:
- n is the number of trials.
- x is the number of successful trials.
- p is the probability of success.

Log-Likelihood Function:

- To simplify calculations, we often work with the log-likelihood function (log-likelihood), which is the natural logarithm of the likelihood function:

$$\ln(L(p)) = \ln\left(\binom{n}{x}\right) + x\ln(p) + (n-x)\ln(1-p)$$

Maximize the Log-Likelihood:

- To find the MLE for p, you need to maximize the log-likelihood function by taking its derivative with respect to p and setting it equal to zero. This can involve solving for p :

$$\frac{d}{dp}\ln(L(p)) = \frac{x}{p} - \frac{n-x}{1-p} = 0$$

- Solve for p to find the MLE of the probability p.

Interpretation:

- The MLE, $\hat{p}$, represents the estimated probability of success (p) for each trial in the binomial distribution based on the observed data.

In this example, the MLE is used to estimate the probability of success (p) in each trial of a binomial distribution. The MLE $\hat{p}$ is obtained by maximizing the log-likelihood function by taking its derivative and setting it equal to zero.

Bayesian Estimation

Bayesian Estimation is a statistical approach that combines prior knowledge (expressed as a probability distribution) with observed data to update our beliefs about an unknown parameter. In Bayesian estimation, the goal is to find the posterior probability distribution for the parameter of interest, given both the prior distribution and the likelihood function. Here are the key mathematical concepts and formulas in Bayesian estimation:

Key Terminology:

- θ : The parameter of interest that we want to estimate.
- X : The data or observations.
- $\pi(\theta)$: The prior probability distribution of the parameter θ.
- $f(X \mid \theta)$: The likelihood function representing the probability of observing the data X given the parameter θ.
- $\pi(\theta \mid X)$: The posterior probability distribution of the parameter θ given the data X.

Bayes' Theorem:

Bayesian estimation is based on Bayes' theorem, which relates the posterior distribution to the prior distribution and the likelihood function:

$$\pi(\theta \mid X) = \frac{f(X \mid \theta) \cdot \pi(\theta)}{f(X)}$$

Where:

- $\pi(\theta \mid X)$ is the posterior distribution.
- $f(X \mid \theta)$ is the likelihood function.
- $\pi(\theta)$ is the prior distribution.
- $f(X)$ is the marginal likelihood (also called the evidence).

Part 2: Methods and Properties of Estimators

2.1 Detailed Exploration of Estimation Methods

Marginal Likelihood (Evidence):

The marginal likelihood (evidence) is the probability of observing the data X integrated over all possible values of the parameter θ :

$$f(X) = \int f(X \mid \theta) \cdot \pi(\theta) d\theta$$

This term is often challenging to compute because it involves integrating over the entire parameter space.

Posterior Distribution:

The posterior distribution $\pi(\theta \mid X)$ represents our updated belief about the parameter θ after observing the data X. It is a probability distribution that summarizes the uncertainty about the parameter.
Estimating the Parameter:
To estimate the parameter θ in Bayesian estimation, various approaches can be used:

Maximum A Posteriori (MAP) Estimation:

- The MAP estimate is the mode of the posterior distribution, i.e., the most likely value of the parameter θ.
- It is obtained by finding the value of θ that maximizes the posterior distribution.

Posterior Mean:

- The posterior mean is the expected value of the parameter θ based on the posterior distribution.
- It is calculated by integrating θ with respect to the posterior distribution.

Credible Intervals:

- Bayesian estimation provides credible intervals (analogous to confidence intervals in frequentist statistics) that express the range within which the parameter θ is likely to fall with a specified probability.

Least Squares Estimation

Part 2: Methods and Properties of Estimators

2.1 Detailed Exploration of Estimation Methods

Least Squares Estimation is a method used to find the best-fitting model (usually a linear one) for a set of data points. It aims to minimize the sum of the squared differences between observed values and predicted values from the model. The most common application is linear regression.

Key Terminology:
y_i : Observed values in the dataset.
x_i : Corresponding independent variables or predictors.
$\hat{y}_i$: Predicted values from the model.
n : The number of data points.

Objective:
The goal in least squares estimation is to find model parameters $\beta_0, \beta_1, \ldots, \beta_k$ (for a linear model) such that the sum of squared residuals is minimized. The residual for each data point i is defined as the difference between the observed value y_i and the predicted value $\hat{y}_i$:

$$e_i = y_i - \hat{y}_i$$

Minimization Objective:
The objective is to minimize the sum of squared residuals, known as the "sum of squares of errors" (SSE):

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Least Squares Estimation

Part 2: Methods and Properties of Estimators

2.1 Detailed Exploration of Estimation Methods

Linear Regression:

In linear regression, the goal is to find the best-fitting linear model, usually expressed as:

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

Where β_0 and β_1 are the parameters to be estimated.

Minimizing SSE:

The least squares estimates of β_0 and β_1 are obtained by minimizing SSE. This is done by finding the values of β_0 and β_1 that satisfy the normal equations:

$$\frac{\partial SSE}{\partial \beta_0} = 0 \quad \text{and} \quad \frac{\partial SSE}{\partial \beta_1} = 0$$

Solving these equations leads to the following formulas for the least squares estimates:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Where:
- $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimates.
- $\bar{x}$ is the mean of the independent variable.
- $\bar{y}$ is the mean of the dependent variable.

These estimates are chosen to minimize the sum of squared residuals, making them the "best" linear fit for the data in the sense of minimizing the error.

Bias, Variance, and Efficiency

Part 2: Methods and Properties of Estimators

2.2 Properties of Estimators

Bias of an Estimator:

Definition: The bias of an estimator $\hat{\theta}$ for a parameter θ measures the systematic error in the estimation. A biased estimator is one whose expected value is different from the true parameter value.

Mathematics:

- The bias $(\text{Bias}(\hat{\theta}))$ is defined as the expected difference between the estimator and the true parameter:

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Variance of an Estimator:

Definition: The variance of an estimator $\hat{\theta}$ quantifies the spread or variability in its estimates. A low variance indicates that the estimator provides consistent and precise estimates.

Mathematics:

- The variance $(\text{Var}(\hat{\theta}))$ is calculated as the expected squared deviation of the estimator from its expected value:

$$\text{Var}(\hat{\theta}) = E\left[(\hat{\theta} - E(\hat{\theta}))^2\right]$$

Bias, Variance, and Efficiency

Part 2: Methods and Properties of Estimators

2.2 Properties of Estimators

Efficiency of an Estimator:
Definition: The efficiency of an estimator measures its precision relative to other estimators. A more efficient estimator has a smaller variance and provides more precise estimates.
Mathematics:
- The efficiency (η) is often defined as the reciprocal of the mean squared error (MSE) of the estimator:

$$\eta = \frac{1}{\mathrm{MSE}(\hat{\theta})}$$

- The MSE of the estimator is given by:

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Bias}(\hat{\theta})^2 + \mathrm{Var}(\hat{\theta})$$

- The efficiency (η) can also be expressed as the ratio of the efficiency of one estimator to another. For example, the efficiency of estimator A compared to estimator B:

$$\eta_A = \frac{\mathrm{Var}\left(\hat{\theta}_B\right)}{\mathrm{Var}\left(\hat{\theta}_A\right)}$$

In summary, the bias, variance, and efficiency of an estimator provide critical measures of its performance. Bias indicates the systematic error, variance quantifies the variability, and efficiency compares the precision of one estimator to another. An efficient estimator has a low variance and is less affected by bias.

Sufficiency, Completeness, and Exponential Families

Part 2: Methods and Properties of Estimators

2.2 Properties of Estimators

Cramér-Rao Inequality in different models

Part 2: Methods and Properties of Estimators

2.2 Properties of Estimators

Lehmann-Scheffé Theorem

Part 2: Methods and Properties of Estimators

2.3 Advanced Topics

Score, Fisher Information, and Regular Models

Part 2: Methods and Properties of Estimators

2.3 Advanced Topics

Numerical optimization techniques

Part 3: Practical Considerations and Challenges

3.1 Computational Aspects

Challenges in high-dimensional parameter spaces

Part 3: Practical Considerations and Challenges

3.1 Computational Aspects

Resistant measures

Part 3: Practical Considerations and Challenges

3.2 Robust Estimation and Handling Outliers

Heavy-tailed distributions

Part 3: Practical Considerations and Challenges

3.2 Robust Estimation and Handling Outliers

Handling Missing Data: Mechanisms and Impacts

Part 3: Practical Considerations and Challenges

3.3 Data Challenges

Reporting and Interpretation: Clear communication and caveats

Part 3: Practical Considerations and Challenges

3.3 Data Challenges

Efficient estimator construction

Part 4: Exercises, Advanced Topics, and Problem Solutions

4.1 Construction and Comparison of Estimators

Comparing Estimators using the Quadratic Risk Function and other metrics

Part 4: Exercises, Advanced Topics, and Problem Solutions

4.1 Construction and Comparison of Estimators

Improved Rao-Blackwell Estimators

Part 4: Exercises, Advanced Topics, and Problem Solutions

4.2 Advanced Statistical Concepts

Exhaustive Statistic and Complete Statistic

Part 4: Exercises, Advanced Topics, and Problem Solutions

4.2 Advanced Statistical Concepts

Set of exercises to test understanding

Part 4: Exercises, Advanced Topics, and Problem Solutions

Problems related to parameter estimation

Part 4: Exercises, Advanced Topics, and Problem Solutions

Solutions for exercises and problems

Part 4: Exercises, Advanced Topics, and Problem Solutions

4.3 Exercises and Problem Solutions

KEYWORDS (NEW)

Erreur de type I · Statistiques asymptotiques · Théorème de Pitman-Koopman-Darmois · Fonction de risque quadratique · Risque quadratique · Outlier · Règle de décision

Estimateur · Théorie de l'estimation · Méthode de la transformée inverse · Statistique inférentielle · Paramètre · Biais d'estimation · Biais systématique · Erreur d'estimation

Estimation · Intervalle de confiance · Estimation par méthode des moindres carrés · Modèle de mélange gaussien (GMM) · Estimation ponctuelle · Erreur de type II

Likelihood (Vraisemblance) · Intervales de confiances sur la moyenne · Processus stochastique bayésien · Information de Fisher · Estimateur linéaire · Niveau de confiance

Méthode des moments · Estimateur du maximum de vraisemblance (EMV) · Apprentissage de paramètres · Estimation par méthode des moments généralisés (GMM)

Estimateur consistant · Estimation par méthode de moindres carrés généralisés (GLS) · Erreur quadratique moyenne (EQM) · Région de confiance · Estimation par méthode de moindres carrés pondérés (WLS)

Estimateur convergent · Estimateur asymptotiquement normal · Estimateur efficace · Estimation par méthode de moindres carrés non linéaires (NLS) · Estimateur sans biais

Estimation d'un paramètre · Estimation asymptotiquement normale · Estimation bayésienne · Estimation de la vraisemblance · Estimation de paramètres · Estimation par intervalle de confiance

Use Case with python

Use Case: Parameter Estimation in Quality Control

In the context of the course on Parameter Estimation, covering topics on the role of parameter estimation, methods of estimation, properties of estimators, and practical considerations, let's explore a use case related to quality control in manufacturing. This use case involves applying various parameter estimation techniques to ensure product quality meets specific standards.

Description:

In this use case, we will focus on using parameter estimation techniques to maintain quality control in a manufacturing process. Specifically, we will estimate the parameters of a quality-related variable and determine whether the manufacturing process is within acceptable limits.

Key Components:

Introduction to Parameter Estimation: Understanding the importance of parameter estimation in quality control, distinguishing parameters from statistics, and defining criteria for good estimators.

Methods of Estimation: Utilizing various estimation methods, including Method of Moments, Maximum Likelihood Estimation (MLE), Bayesian Estimation, and Least Squares Estimation.

Properties and Comparisons of Estimators: Evaluating estimators in terms of bias, variance, efficiency, sufficiency, consistency, and relative efficiency.

Practical Considerations and Challenges: Addressing computational aspects of estimation, robust estimation to handle outliers, dealing with missing data in quality control, and ensuring clear reporting and interpretation of estimated parameters.

Python Code Example (Parameter Estimation in Quality Control):

```python
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

# Generate synthetic quality control data (e.g., product length measurements)
np.random.seed(42)
product_lengths = np.random.normal(10, 1, size=100)

# Parameter estimation using Maximum Likelihood Estimation (MLE)
mean_estimate = np.mean(product_lengths)
std_dev_estimate = np.std(product_lengths, ddof=1)  # Use Bessel's correction

# Confidence interval calculation
confidence_interval = stats.norm.interval(0.95, loc=mean_estimate, scale=std_dev_estimate / np.sqrt(len(product_lengths)))

# Visualize data distribution and confidence interval
plt.hist(product_lengths, bins=15, density=True, alpha=0.6, color='b', label='Product Lengths')
plt.axvline(mean_estimate, color='r', linestyle='--', label='Mean Estimate')
plt.fill_between(confidence_interval, 0, 0.4, color='g', alpha=0.3, label='95% Confidence Interval')
plt.xlabel('Product Length')
plt.ylabel('Probability Density')
plt.legend()
plt.title('Parameter Estimation in Quality Control')
plt.show()

print(f'Mean Estimate: {mean_estimate:.2f}')
print(f'Standard Deviation Estimate: {std_dev_estimate:.2f}')
print(f'95% Confidence Interval: ({confidence_interval[0]:.2f}, {confidence_interval[1]:.2f})')
```

In this code, we generate synthetic product length measurements and use Maximum Likelihood Estimation (MLE) to estimate the mean and standard deviation of the product lengths. We calculate a 95% confidence interval to determine whether the manufacturing process is within acceptable quality limits.

This use case demonstrates how parameter estimation techniques can be applied to maintain quality control and ensure that products meet specific quality standards in manufacturing.

TOP 20 REFERENCES

- Lehmann, E. L., & Casella, G. (1998). Theory of Point Estimation. Springer.
- Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). Introduction to the Theory of Statistics. McGraw-Hill.
- Cox, D. R., & Hinkley, D. V. (1974). Theoretical Statistics. CRC Press.
- Rao, C. R. (1973). Linear Statistical Inference and Its Applications. Wiley.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
- Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. Philosophical Transactions of the Royal Society of London.
- DeGroot, M. H., & Schervish, M. J. (2012). Probability and Statistics. Pearson.
- White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. Econometrica.
- Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. CRC Press.
- Wasserman, L. (2004). All of Nonparametric Statistics. Springer.
- Casella, G., & Berger, R. L. (2002). Statistical Inference. Duxbury Press.
- Huber, P. J. (2011). Robust Statistics. Wiley.
- Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. MIT Press.
- Tsybakov, A. B. (2009). Introduction to Nonparametric Estimation. Springer.
- Little, R. J., & Rubin, D. B. (2019). Statistical Analysis with Missing Data. Wiley.
- Kiefer, J., & Wolfowitz, J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. The Annals of Mathematical Statistics.
- Neyman, J., & Pearson, E. S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. Philosophical Transactions of the Royal Society of London.
- Zhang, C.-H. (2010). Nearly Unbiased Variable Selection under Minimax Concave Penalty. The Annals of Statistics.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature.

TEXTE DE DESCRIPTION DU COURS

Welcome to "Introduction to Parameter Estimation: Mastering Statistical Inference!" In this comprehensive course, you'll dive deep into the foundations and methods of parameter estimation, a crucial aspect of statistics that empowers you to draw meaningful conclusions from data. Whether you're a data scientist, researcher, or statistician, this course equips you with the skills to estimate population parameters, understand their properties, and navigate practical challenges.

"Introduction to Parameter Estimation" commences with a clear understanding of the "Role of Parameter Estimation in Statistics." Explore the definition and importance of parameter estimation and learn how to distinguish parameters from statistics. Dive into "Point and Interval Estimation," where you'll master point estimators, their properties, and the concept of confidence intervals.

Discover the diverse "Types of Parameters," including mean, variance, proportion, and rate. Understand the intricate relationship between sample statistics and population parameters. In "Criteria for Good Estimators," delve into unbiasedness, consistency, efficiency, and mean squared error, dissecting their components and implications.

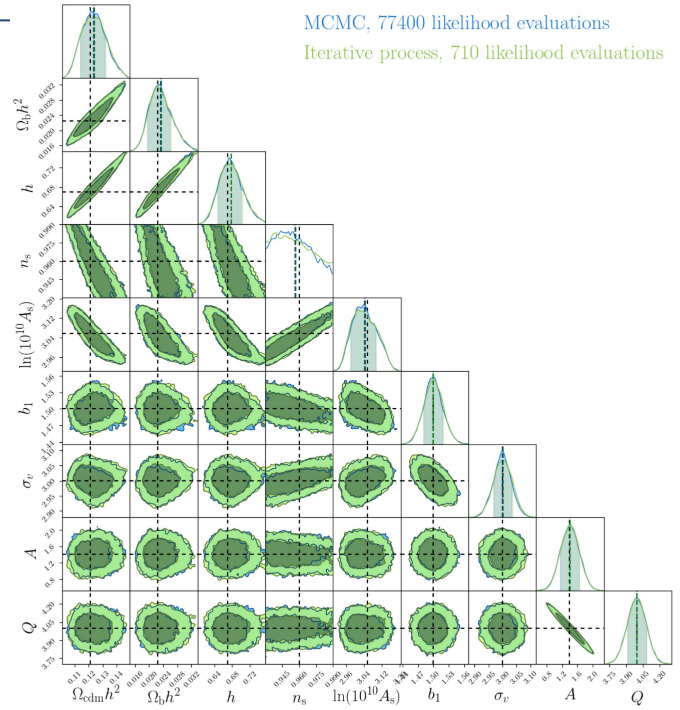
"Methods of Estimation" takes center stage, starting with the "Method of Moments." Understand the underlying theory and apply it in simple applications and examples. "Maximum Likelihood Estimation (MLE)" follows, revealing its principles, properties, and practical computation for common distributions. Explore the Bayesian perspective in "Bayesian Estimation," covering prior and posterior distributions, Bayes' estimators, and their properties. Wrap up with "Least Squares Estimation," where you'll grasp the concept of minimizing the sum of squared residuals and its application in simple linear regression models.

"Properties and Comparisons of Estimators" opens with "Bias and Variance," where you'll define bias and explore its implications, along with the trade-off between bias and variance. Investigate the concepts of "Efficiency and Sufficiency," touching on the Cramér-Rao Lower Bound and the principle of sufficiency's importance. Delve into "Consistency of Estimators," understanding convergence in probability and almost sure convergence, along with criteria for determining consistency. Wrap up this section by "Comparing Estimators," using relative efficiency and mean squared error as metrics for comparison.

"Practical Considerations and Challenges" explores the real-world aspects of parameter estimation. Dive into "Computational Aspects of Estimation," learning numerical optimization techniques and addressing challenges in high-dimensional parameter spaces. Discover "Robust Estimation," where you'll explore resistant measures and strategies for dealing with outliers and heavy-tailed distributions. "Handling Missing Data" sheds light on the mechanisms of missingness and the impact of missing data on parameter estimation. Finally, in "Reporting and Interpretation," you'll recognize the importance of clear communication of estimated parameters and understand the caveats and limitations in interpretation.

By the end of this course, you'll be well-versed in parameter estimation, equipped with a comprehensive toolkit to tackle statistical inference challenges and make informed decisions based on data.

DESIGN DE LA CARD



Author: Baptiste Mokas, Weeki

Course Name: Simple Linear Regression

#ParameterEstimation
#StatisticalInference
#StatisticalMethods