Parametric Statistics

Sofia Triantafyllou

Fall 2021

E-mail: sof.triantafillou@gmail.com
Web: **https://polyhedron.math.uoc.gr/2122/moodle/enrol/index.php?id=15**
Class Hours: Tuesday, Friday 11.00-13.00
Office Hours: TBD
Office: B316
Class Room: TBD

This is a tentative syllabus for the course MEM-262: Parametric Statistics.

Course Description

This is a one-semester course covering the basics of parametric statistics. We will start with a quick introduction to probability theory, and then cover fundamental topics in mathematical statistics such as estimation, hypothesis testing, regression, and convergence. This course also includes an introduction to statistical programming, using R language.

Required Materials

The course notes are based on the following books:

(LW) All of Statistics, Larry Wasserman, 2004

(DGS) Probability and Statistics, Morris DeGroot and Mark Schervish (International Edition), 2014 (4th edition)

Course notes will be available on the class website. The material can also be found in Greek in the following books:

- Εισαγωγή στη Στατιστική ΜΕΡΟΣ Ι, Δαμιανού Χ., Κούτρας Μ.
- Στατιστική συμπερασματολογία, (Τόμοι Ι & ΙΙ), Ρούσσας Γεώργιος Γ., Σταματέλος Γεώργιος.

Prerequisites

Required preliminary math tools are elementary probability, calculus and basic linear algebra. There are no formal prequisites for this class, but I would advise that you have passed Calculus

I, Calculus II and Probability Theory. Statistics are the flip side of probability. In your probability course, you start with a distribution (say, a normal distribution with a mean μ and a variance σ^2) and predict features of future observations $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n})$; in parametric statistics we observe the data $\mathbf{x}$ and then try to guess the parameters of the distribution that generated them.

Course Objectives

By the end of the course, you should be familiar with the following statistical concepts:

- Probability basics: Probability is the way to quantify how likely something is to occur. We will talk about experiments, sample spaces, events, independence, and conditional probability.
- Random Variables and their distributions: Random variables are real-valued functions link sample spaces and events to data. They model unknown quantities (e.g. the height of a person) in statistical analyses. A random variable takes each one of its possible values with some probability. The collection of these probabilities is the distribution of the random variable. We will discuss random variables and their distributions. We will also discuss some summaries of the distribution, such as expectation and variance, and why they are useful. Finally. we will discuss some mathematical tools for describing the limiting behavior (convergence) of sequences of random variables. This helps us understand what we can expect when we gather more and more data.
- Statistical Inference: Statistical inference is the process of using data to infer (learn) the distribution that generated the data. Sets of distributions that can be described using a finite number of parameters are called parametric models. We will discuss the following types of parametric inference:
  - Point Estimation: Point estimation refers to providing a single "best guess" (estimate) for some quantity of interest, very often a parameter of a parametric model. We will discuss different methods for deriving these guesses: (maximum likelihood, method of moments), as well as their asymptotic properties. We will also discuss how we can choose the best estimator, using statistical decision theory, and how we can quantify uncertainty on the estimated quantities with confidence intervals.
  - Hypothesis testing: Hypothesis testing has to do with trying to decide whether a parameter lies on a specific space of the parameter space or not. For example, we may want to decide if a coin is fair after we have done a number of coin flips. We will discuss how we formalize and test statistical hypotheses, and some specific parametric hypothesis tests.

  We will also discuss the two dominant approaches in statistical inference, frequentist inference and Bayesian inference.
- Linear and Logistic regression: Sometimes, we may have observations in pairs, for example we may have heights of children and their fathers. We may want to predict one of the variables based on the other (e.g., predict the height of a child based on the height of their father). This is called regression. We will talk about linear and logistic regression, the general linear model, and analysis of variance.

Grading

Grading will be based on 2 midterm exams, each accounting for **30%** of the grade, a final exam, accounting for **30%**of the grade, and 1 homework assignment (due near the end of the semester), accounting for **20%** (Summing to 110%,you have a 10% bonus). To pass the course, your average grade on the three needs to be above 50/100. You can do the homework assignment in teams of up to 3 students.

Late assignments will be accepted for no penalty if a valid excuse is communicated to the instructor before the deadline. After the deadline, assignments will be accepted for a 50% deduction to the score up to 2 days after the deadline. After this any assignments handed in will be given 0.