# Parametric Statistics

## Lecture 1: Introduction, Probability Axioms

Sofia Triantafillou

sof.triantafillou@uoc.gr

University of Crete
Department of Mathematics and Applied Mathematics

# Course Info

E-mail: `sof.triantafillou@uoc.gr`
Web:
**https://polyhedron.math.uoc.gr/2223/moodle/course/view.php?id=29**
Class Hours: Monday, Wednensday 09.15-11.00
Office Hours: Wednesday 11.15-13.00
Recitations: TBD
Office: B316

# Course Materials

The course notes are based on the following books

- (DGS) Probability and Statistics, Morris DeGroot and Mark Schervish (International Edition), 2014 (4th edition)
- (JR) John A. Rice, Mathematical Statistics and Data Analysis, 3rd edition.
- (WM) R. E. Walpole, R.H Myers, S.L Myers, K. Ye, A. Τσακανίκας, Στατιστική και Πιθανότητες, Εκδόσεις Τζιόλα, 2019, Κωδικός Εύδοξου: 68374152
- (ΘΠΣ) Θέματα παραμετρικής στατιστικής συμπερασματολογίας, Σταύρος Κουρουκλής, Κωδικός Εύδοξου: 59303581

# Grading

- Midterm exams, 50%
- Final exam, 50%

The date of the midterm exam will be announced by the end of September

**No other arrangements will be made.**

September exam will be comprehensive and will count for 100% of the grade. I strongly advise participating in the midterm exam.

# How to succeed

## Prerequisites

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

## Study!

- Two to four hours every week throughout the semester.
- Try to recap for half an hour the day of the class
- Try to solve the exercises before the recitation.
- Tips for studying from the UNC: https://learningcenter.unc.edu/tips-and-tools/studying-101-study-smarter-not-harder/

# What is statistics

Statistics and probability are concerned with describing uncertainty. In both disciplines, we have some **experiments** that can have **multiple outcomes** that occur with some **probability**.

In your probability course, you start with a distribution (say, a normal distribution with mean $\mu = 10$ and a variance $\sigma^2 = 25$) and predict features of future observations $\mathrm{x} = (x_1, x_2, ..., x_n)$.

# What is statistics

In statistics we observe the data x and then try to understand the properties of the data-generating process.

In **parametric** statistics, we assume that the process can be described using a finite number of parameters $\theta$ and try to identify these parameters.

# Outline of the Course

- Probability Recap
- Random Variables & Their Distributions
- Large Random Samples
- Estimation, Sampling Distributions of Estimators
- Testing Hypotheses

We will discuss both the Bayesian approach and the frequentist approach.

# Example

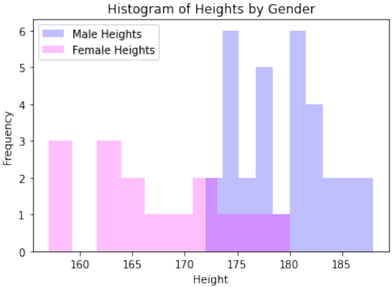
## Example

Assume we want to predict the height of the next student in the class.

Last year, we collected data from the students of parametric statistics course.
Students were given an empty card. They wrote their height and gender on the card (answering was optional).

# Let's do some statistical inference

| Height | Gender |
|---|---|
| 1.75 | F |
| ⋮ | ⋮ |
| 1.68 | M |




## Point estimation

Assuming the data follow a Normal $N(\mu, \sigma^2)$ distribution, what is the "best guess" for $\mu$ and $\sigma$?

## Hypothesis testing

Is the mean height of females (significantly) smaller than the height of males?

# Statistical inference using R

- Install R: https://cran.r-project.org/
- Install R-studio: https://www.rstudio.com/products/rstudio/download
- Download class_data_heights.xlsx from the class website
- Run the commands in Rscript0.R to see the summary of our data.

# Today

1.4 Set Theory
1.5 Definition of Probability
1.6 Finite Sample Spaces
1.7 Counting Methods
1.8 Combinatorial Methods

# What is probability?

## Probability

is a language for quantifying uncertainty. It is a way to quantify how likely something is to occur.

## Experiment

An experiment is any real or hypothetical process, in which the **possible outcomes** can be identified ahead of time. Events are **sets** of possible outcomes. Probability is then a way to describe how likely each event is.

## Example experiments

- Tossing a coin.
- Tossing a coin twice.
- Measuring the temperature in Heraklion.

# The sample space

- The sample space is $\Omega$ is the set of possible outcomes of an experiment.
- $\omega \in \Omega$ is are called **sample outcomes**, or **elements**.
- Subsets of $\Omega$ are called **events**.
- Elements of $\Omega$ must be
  1. mutually exclusive
  2. collectively exhaustive

# Discrete Sample Space

We toss a coin.

# Discrete Sample Space

We toss a coin.

Possible Outcomes: Heads (H), Tails (T)

Sample Space: $\{H, T\}$

Examples of Events: $H$, $T$

Probability of each event: $P(H) = P(T) = 0.5$

## Discrete Sample Space

We toss a coin two times.

# Discrete Sample Space

We toss a coin two times.

Possible Outcomes: Heads - Heads , Heads - Tails , Tails - Heads , Tails - Tails

Sample Space: $\{HH, HT, TH, TT\}$

Examples of Events: First toss is Heads

Probability of this event: $P(\text{First toss is heads}) = 0.5$

# Continuous Sample Space

Measuring the temperature in Heraklion.

# Continuous Sample Space

Measuring the temperature in Heraklion.

Possible Outcomes: ?

Sample Space: $(-\infty, \infty)$ OR $(-50, 100)$

Example events:

- temperature is $25^\circ C$
- temperature is at least $10^\circ C$ but less than $20^\circ C$: $A = [10, 20)$.

Probability of this event:

# Probability Measure /Probability Law

We want to assign a real number $\mathbb{P}(A)$ to every event $A$ which represents how likely event $A$ is to occur.

- Axiom 1: $\mathbb{P}(A) \geq 0$ for every $A$.
- Axiom 2: $\mathbb{P}(\Omega) = 1$.
- Axiom 3: for an infinite sequence $A_1, A_2, \ldots$ of disjoint events

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

What kind of mathematical object is $\mathbb{P}$?

# Probability Measure /Probability Law

We want to assign a real number $\mathbb{P}(A)$ to every event $A$ which represents how likely event $A$ is to occur.

- Axiom 1: $\mathbb{P}(A) \geq 0$ for every $A$.
- Axiom 2: $\mathbb{P}(\Omega) = 1$.
- Axiom 3: for an infinite sequence $A_1, A_2, \ldots$ of disjoint events

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

What kind of mathematical object is $\mathbb{P}$?

- What is the domain?
- What is the codomain?

# $\sigma$-algebra

## Why not define $\mathbb{P}$ for every possible subset of $\Omega$?

Let's try for $(-50, 100)$ assuming every temperature is equally likely to occur.

Remember the Axioms.

## Definition

A $\sigma$-algebra or a $\sigma$-field on $\Omega$ is a set of subsets $\mathcal{A}$ of $\Omega$ that sastisfy the following conditions:

- Contains both empty set $\emptyset$ and $\Omega$
- Closed under complement: for all $\alpha \in \mathcal{A}, \alpha^c \in \mathcal{A}$
- Closed under (countable) union: for all $\alpha, \beta \in \mathcal{A}, \alpha \cup \beta \in \mathcal{A}$
- Implies closed under (countable) intersection: $\alpha \cap \beta = (\alpha^c \cup \beta^c)^c$

The sets in $\mathcal{A}$ are said to be measurable. We call $(\Omega, \mathcal{A})$ a measurable space.

# Another reason for using $\sigma-$ algebras

- Reason 1: We want to be able to assign probabilities to all the events we are interested in without problems.
- Reason 2: We may be interested in a specific set of events (we do not need all events) or we may want to represent partial information.
- Example: Assume someone flips a coin twice, but you only observe the first flip. What is the sigma-algebra that can represent the information you have?

# Generating a $\sigma$-algebra

You can always construct the smallest $\sigma$-algebra for a set of subsets $\mathcal{M}$. This is called the $\sigma$-algebra generated by $\mathcal{M}$, and we denote it $\sigma(\mathcal{M})$.

## Example

$$\Omega = \{a, b, c, d\}, \mathcal{M} = \{\{a\}, \{b\}\}$$

$$\sigma(\mathcal{M}) = ?$$

This is easy in discrete spaces, and much more complicated in continuous sample spaces. The most common $\sigma$-algebra for continuous sample spaces is the Borel $\sigma$-algebra, which is generated by all open intervals (i.e., it is the smallest $\sigma$-algebra that contains all open intervals.

# Probability Measure

We want to assign a real number $\mathbb{P}(A)$ to every element $A$ of a $\sigma-$algebra $\mathcal{A}$ to $[0,1]$ such that $\mathbb{P}(A)$represents how likely event $A$ is to occur.

- Axiom 1: $\mathbb{P}(A) \geq 0$ for every $A$
- Axiom 2: $\mathbb{P}(\Omega) = 1$
- Axiom 3: for an infinite sequence $A_1, A_2, \ldots$ of disjoint events

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

The triplet $(\Omega, \mathcal{A}, \mathbb{P})$ is called a **probability space**.

## Properties of Probabilities (1)

Based on the axioms of probability, we can derive several properties:

- The probability of an impossible event is 0:

$$P(\emptyset) = 0$$

- Axiom 3 also holds for finite sequences of events:

$$P(\bigcup_{i=1}^{N} A_i) = \sum_{i=1}^{N} P(A_i)$$

- The probability of any event is no more than 1:

$$P(A) \leq 1$$

# Properties of Probabilities (2)

## Probability of the Union of Events

For any events $A$ and $B$,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

## The law of total probability

Let $B_1, \ldots, B_n$ be a partition of the sample space. Then for any event $A$,

$$P(A) = \sum_{i}^{N} P(A \cap B_i)$$

# Recap

- Probability spaces formalize uncertainty.
- Sample spaces describe possible outcomes of a random experiment.
- Events are subsets of the sample space.
- Probability measure defines how probable events are.
- We need to define them on sets of events that are *measurable*, i.e., $\sigma$-algebras.
- Probability measures need to satisfy the *axioms of probability.*

# Counting in a uniform (simple) probability space

In discrete sample spaces, if all outcomes of an experiments are considered equally likely, then for each event $A$,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

Example:

We roll a (fair) die twice. $\Omega$ has 36 elements:

$$11, 12, \ldots, 16$$

$$21, 22 \ldots, 26,$$

$$\vdots$$

$$61, 62 \ldots, 66$$

What is P(at least one 6)?

# Combinatorics

## Multiplication Rule

- $r$ steps.
- $n_r$ choices at each step.
- Then the number of choices are $n_1 \times n_2 \times \cdots \times n_r$.

## Permutations

Number of ways to arrange $n$ elements:

$$n(n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!}$$

## Combinations

Ways to chose $k$ elements from a collection of $n$ elements:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

# Let's practice

- Number of license plates with 3 letters (choose from 14 letters A,B, E, Z, H, I, K, M, N, O, P, T, Y, X) and 4 digits.
  - Without repetitions.
  - With repetitions.
- Suppose that 35 people are divided in a random manner into two teams in such a way that one team contains 10 people and the other team contains 25 people. What is the probability that two particular people $A$ and $B$ will be on the same team?

# Practice Exercises

| Section | Exercises |
|---|---|
| 1.4 | 1, 6, 8 |
| 1.5 | 3, 4, 8, 10, 14 |
| 1.6 | 1, 6 |
| 1.7 | 4, 5, 6 |
| 1.8 | 2, 3, 4, 12 |