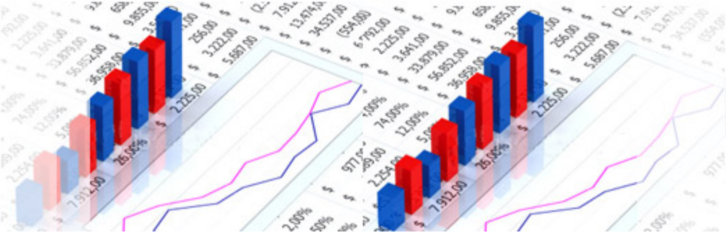
MTMS.01.099 Mathematical Statistics

Lecture 1. Introduction

Tõnu Kollo

tonu.kollo@ut.ee



Fall 2017

Textbooks

Gunnar Blom, *Probability and Statistics: Theory and Applications*. Springer, New York.
Several editions

In Estonian:

Kalev Pärna, *Tõenäosusteooria algkursus*. Tartu, 2013
Imbi Traat, *Matemaatilise statistika põhikursus*. Tartu, 2006

Course outline

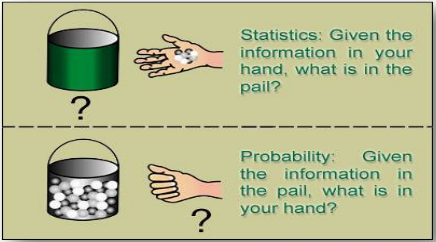
- Weeks 1-16
- 2 Tests: max 20 points each, one has to get min 10 points to pass the test
  - 1.test week 8
  - 2.test week 15
- Homeworks
  - 3-4 homework assignments, 10 points in total.
- Written exam: max 50 points, min 26 points are needed to pass the exam
  - Exam times: to be announced
- *Extra comments
  - For admittance to exam both tests have to be passed.
  - One week to submit the solutions of homework.
  - If you fail with a test, there will be a possibility to repeat it once.

Why statistics?

Nowadays statistics is used in every field of science.

- Is a new medicine more effective than the existing one?
- Comparing the effectiveness of three cleaning products in removing four different types of stains.
- Studying the relation between salary increases and employee productivity.
- How big have to be reserves of a insurance company?
- How accurate are the gallups and opinion polls?
- What will be the unemployment rate next year?
- How to improve quality of product?
- ...

Probability theory vs Statistics



Probability:

- deals with predicting the likelihood of future events
- What is the probability of getting 20 sixes when throwing a dice 100 times?

Statistics:

- involves the analysis of the frequency of past events
- A dice is thrown 100 times. How (and how precisely) we can estimate the probability to get six?

Events

Outcome - the result of a random trial

- denoted by $u_1, u_2, \ldots$

Sample space - the set of all possible outcomes

- denoted by $\boldsymbol{\Omega}$ (omega)

Event - a collection of outcomes

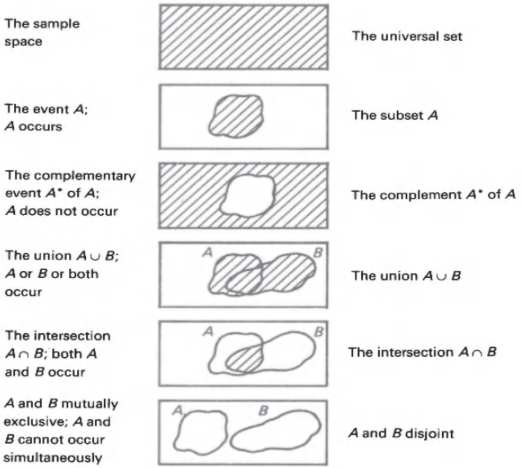
- denoted by capital letters $A, B, C, \ldots$

Thus an event is a subset of the sample space (or possibly the whole sample space)

Example

Let us throw a single dice. Assume that the dice has six sides. We introduce six outcomes, which we denote by $1, 2, ..., 6$. The sample space $\boldsymbol{\Omega}$ consists of these outcomes and can be written $\boldsymbol{\Omega} = \{1, 2, ..., 6\}$. Example of an event is A = "number of points is at most 2". Using the symbols for the outcomes, we can write this event $A = \{1, 2\}$.

Events


The sample space
The universal set
The event A; A occurs
The subset A
The complementary event A* of A; A does not occur
The complement A* of A
The union A ∪ B; A or B or both occur
A
B
The union A ∪ B
The intersection A ∩ B; both A and B occur
A
B
The intersection A ∩ B
A and B mutually exclusive; A and B cannot occur simultaneously
A
B
A and B disjoint

Probability

Definition

The following axioms for the probabilities $P(\cdot)$ should be fulfilled:
Axiom 1. If A is any event, then $0 \leq P(A) \leq 1$.
Axiom 2. If Ω is the entire sample space, then $P(\Omega) = 1$.
Axiom 3 (Addition Formula). If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.
The sample space Ω and the probabilities $P(\cdot)$ together constitute a *probability space*.

Conditional Probability

Definition

The expression

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) \neq 0$$

is called the **conditional probability** of B given A.

The formula can also be expressed in the form

$$P(A \cap B) = P(A)P(B|A)$$

or in words: *The probability that two events both occur is equal to the probability that one of them occurs, multiplied by the conditional probability that the other occurs, given that the first one has occurred.*

Conditional Probability. Example

Example

In a batch of 50 units there are 5 defectives. A unit is selected at random, and thereafter one more from the remaining ones. Find the probability that both are defective.

Let A be the event that "the first unit is defective" and B the event that "the second unit is defective".

It is seen that $P(A) = 5/50$. If A occurs, there remain 49 units, 4 of which are defective. Hence we conclude that $P(B|A) = 4/49$, and it follows:

$$P(A \cap B) = \frac{5}{50} \cdot \frac{4}{49} = \frac{2}{245}$$

Total Probability Theorem

Theorem

If the events $H_1, H_2, \ldots, H_n$ are mutually exclusive, have positive probabilities, and together fill Ω completely, any event A satisfies the formula

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i).$$

Total Probability Theorem. Example

Example

In a factory, units are manufactured by machines H_1, H_2, H_3 in the proportions 25: 35: 40. The percentages $5\%, 4\%$ and 2%, respectively, of the manufactured units are defective. The units are mixed and sent to the customers. Find the probability that a randomly chosen unit is defective.
Taking, H_i = "unit produced by machine H_i" and A = "unit is defective", we find

$$P(A) = 0.25 \cdot 0.05 + 0.35 \cdot 0.04 + 0.40 \cdot 0.02 = 0.0345.$$

Bayes' Theorem

Theorem

Under the same conditions as in Total Probability Theorem

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^{n} P(H_j)P(A|H_j)}$$

Example

Suppose that a customer discovers that a certain unit is defective. What is the probability that it has been manufactured by machine H_1?(A = defective unit)

$$P(H_1|A) = \frac{0.25 \cdot 0.05}{0.25 \cdot 0.05 + 0.35 \cdot 0.04 + 0.40 \cdot 0.02} = 0.36.$$