MTMS.01.099 Mathematical Statistics
LTMS.00.056 Mathematical Statistics (part I)
LTMS.00.057 Mathematical Statistics (part II)

Course introduction

Oleksandr Chepizhko
oleksandr.chepizhko@ut.ee

Essentials

Communication:

- Moodle page (grades, materials, announcements - source of final truth)
- Classroom or Zoom calls - lectures&lab sessions will be casted live and recorded
- email (see contacts in Moodle)
- Slack channel

Essentials

Communication:

- Moodle page (grades, materials, announcements - source of final truth)
- Classroom or Zoom calls - lectures&lab sessions will be casted live and recorded
- email (see contacts in Moodle)
- Slack channel

Course duration:

- weeks 1-16 (for 6 ECTS)
- weeks 1-8 (for 3 ECTS)
- 1 lecture (Tue) + 1 lab session (Wed/Thu) per week
- 6 ECTS Lab group 1 & 2 (Nicholas Lupul)
- 3 ECTS Lab group 1 & 2 (Anastassia Kolde)

How to pass this course? (1-2)

Tests:

- Tests during lab sessions (Wed/Thursdays)
  - 1st test week 9 (for 3 ECTS this is the final exam)
  - 2nd test week 14
- Max 20 points each, 10 points to pass the test
  - for 3 ECTS max 30 points with some extra exercises (and extra time)
- All tests have to be passed to get to the exam

How to pass this course? (1-2)

Tests:

- Tests during lab sessions (Wed/Thursdays)
  - 1st test week 9 (for 3 ECTS this is the final exam)
  - 2nd test week 14
- Max 20 points each, 10 points to pass the test
  - for 3 ECTS max 30 points with some extra exercises (and extra time)
- All tests have to be passed to get to the exam
- Total of 2 retake possibilities (e.g. 1st test can be retried 2x if 2nd test is passed with first try)
  - for 3 ECTS only one retake opportunity
- Test retakes will be on a separate agreed time
- If you pass a test, then no retest is allowed to improve results

How to pass this course? (2-2)

Homeworks:

- 4-5 homework assignments (3 for 3 ECTS)
- max 20 points in total (max 10 for 3 ECTS)
- Deadline 1 week

How to pass this course? (2-2)

Homeworks:

- 4-5 homework assignments (3 for 3 ECTS)
- max 20 points in total (max 10 for 3 ECTS)
- Deadline 1 week

Exam (6 ECTS):

- max 40 points, min 20 points to pass the exam.
- Exam times - 1x Dec, 2x Jan

How to pass this course? (2-2)

Homeworks:

- 4-5 homework assignments (3 for 3 ECTS)
- max 20 points in total (max 10 for 3 ECTS)
- Deadline 1 week

Exam (6 ECTS):

- max 40 points, min 20 points to pass the exam.
- Exam times - 1x Dec, 2x Jan

Exam (3 ECTS):

- max 30 points, min 15 points to pass the exam.
- Exam times - 1x Oct

How to pass this course? (2-2)

Homeworks:

- 4-5 homework assignments (3 for 3 ECTS)
- max 20 points in total (max 10 for 3 ECTS)
- Deadline 1 week

Exam (6 ECTS):

- max 40 points, min 20 points to pass the exam.
- Exam times - 1x Dec, 2x Jan

Exam (3 ECTS):

- max 30 points, min 15 points to pass the exam.
- Exam times - 1x Oct

Grade (6 ECTS): 20p HWs + 40p tests + 40p exam

Grade (3 ECTS): 10p HWs + 30p test

How to EXCEL this course?

Lectures: prepare for lectures!

- study slides before lecture
- watch or read helpful material (provided in Moodle)

How to EXCEL this course?

Lectures: prepare for lectures!

- study slides before lecture
- watch or read helpful material (provided in Moodle)

Lab sessions: solve exercises before lab

How to EXCEL this course?

Lectures: prepare for lectures!

- study slides before lecture
- watch or read helpful material (provided in Moodle)

Lab sessions: solve exercises before lab

Extra points:

- Answer quiz questions after lecture in Moodle on Week 2-15
- Helpful tips, material, correction of mistakes to the lecturer (at lecturer's discretion)

How to EXCEL this course?

Lectures: prepare for lectures!

- study slides before lecture
- watch or read helpful material (provided in Moodle)

Lab sessions: solve exercises before lab

Extra points:

- Answer quiz questions after lecture in Moodle on Week 2-15
- Helpful tips, material, correction of mistakes to the lecturer (at lecturer's discretion)

Tests & exam:

- solve extra exercises in the book by G. Blom or provided by lecturer
- look for material online
- discuss difficult topics with fellow students

Textbooks

In English:

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.

THE "HOW?" PART

- Classical probability theory.
- Random variables and probability distributions.
- Statistical inference: Population and sample.
- Main statics (mean, st. dev). Confidence interval.
- Hypothesis testing.
- Measures of dependence.

TEST 1

Course outline (2-2)

THE "WHY?" PART

- Estimation function. Properties of estimators.
- Methods for deriving estimates.
- Confidence interval (advanced)
- Hypothesis testing (advanced):
  - Test statistic, critical region, type I error, type II error.
  - Power function, independent samples, paired samples.
  - Chi-square test, F-test, p-value method, testing with confidence intervals.

TEST 2

- Measures of dependence (advanced).
- Introduction to linear regression.

EXAM

Preparations

$$e = 2.7\ 1828\ 1828\ 459045\ 23536$$

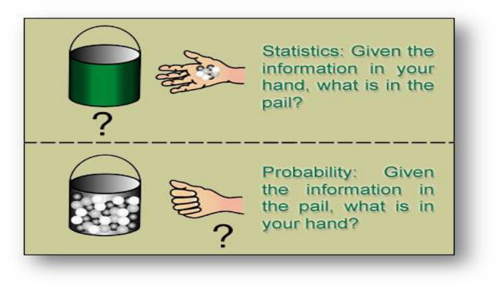
Why statistics?

Why statistics?

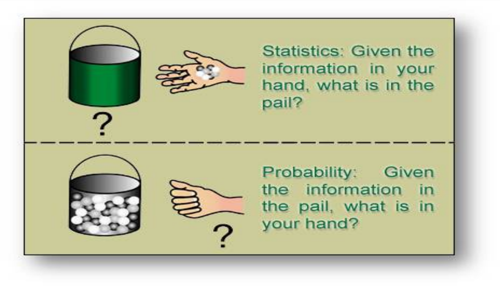
Nowadays statistics is used in every field of science and business.

- Is a new medicine more effective (safe) than the existing one?
- Is the product improvement better than current version?
- Studying the relation between salary increases and employee productivity.
- How big the reserves of an insurance company have to be?
- How accurate are the gallups and opinion polls?
- What will be the unemployment rate next year?
- Is current e-mail spam or not?
- ...

Probability theory vs Statistics



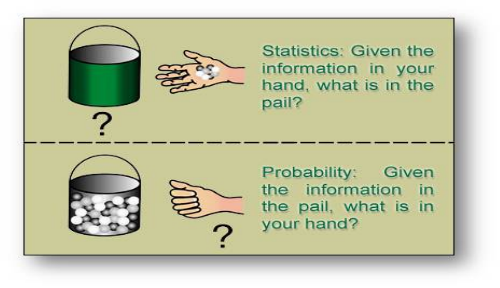
Probability theory vs Statistics



Probability - deals with predicting the likelihood of future events:

- What is the probability of getting 20 sixes when throwing a dice 100 times?

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?

Set theory - 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 Ω (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)

Events

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 Ω consists of these outcomes and can be written $\Omega = \{1, 2, ..., 6\}$.

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

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 $A = \{1, 2\}$.

If event B = ”Odd number results”, then $B = \{\ldots\}$?

Example

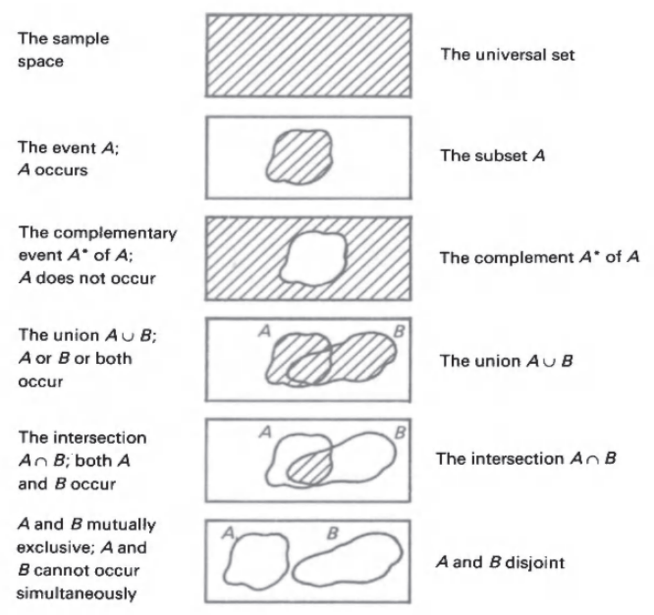
Let us throw a single dice. Assume that the dice has six sides. We introduce six outcomes, which we denote by $1, 2, ..., 6$. The sample space $\boldsymbol{\Omega}$ consists of these outcomes and can be written $\boldsymbol{\Omega} = \{1, 2, ..., 6\}$.

Example of an event is A = "number of points is at most 2". Using the symbols for the outcomes, we can write this event $A = \{1, 2\}$.

If event B = ”Odd number results”, then $B = \{\ldots\}$?

If event C = ”Prime number results”, then $C = \{\ldots\}$?

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 $A \cup B$; A or B or both occur
A
B
The union $A \cup B$
The intersection $A \cap B$; both A and B occur
A
B
The intersection $A \cap B$
A and B mutually exclusive; A and B cannot occur simultaneously
A
B
A and B disjoint

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 $\boldsymbol{\Omega}$ is the entire sample space, then $P(\boldsymbol{\Omega}) = 1$.

Axiom 3 (Addition Formula). If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.

The sample space $\boldsymbol{\Omega}$ 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

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

Translation: 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.

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".

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. Manufactured units are defective with probabilities $5\%, 4\%$ and 2%, respectively. The units are mixed and sent to the customers. Find the probability that a randomly chosen unit is defective.

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. Manufactured units are defective with probabilities $5\%, 4\%$ and 2%, respectively. 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"

Example

In a factory, units are manufactured by machines H_1, H_2, H_3 in the proportions 25: 35: 40. Manufactured units are defective with probabilities $5\%, 4\%$ and 2%, respectively. 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)

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)

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

Another Example; Drug Testing

True positive rate 90%, true negative rate 80% $\rightarrow$ means false positive rate 20%.

Assuming 5% prevalence, what is probability that random positive-tested person is really a drug user?

Another Example; Drug Testing

$$P(\text{user}|+) = \frac{P(+|\text{user})P(\text{user})}{P(+)} =$$

$$\frac{P(+|\text{user})P(\text{user})}{P(+|\text{user})P(\text{user}) + P(+|\text{non-user})P(\text{non-user})} =$$

$$\frac{0.90 \times 0.05}{0.90 \times 0.05 + 0.20 \times 0.95} =$$

$$\frac{0.045}{0.045 + 0.19} \approx 19\% \, .$$

Another Example; Drug Testing