Probability Theory and Statistics
Lecture 1

Bence Csonka

Budapest University of Technology and Economics
csonkab@edu.bme.hu

September 8

Information

Teacher: Bence Csonka

Email: csonkab@edu.bme.hu

Room: IE.2.17.3

Lectures:

Date	Classroom
Monday 12.15-13.45	IE220
Tuesday 10.15-11.45	IE220

Practices: Exercise classes are taught by Humara Khan.

Date	Classroom
Monday 14:15-16:00	IB140
Wednesday 14:15-16:00	E402
Friday 12:15-14:00	E406

Assessments dates

- Midterm: October 27, 8:00-10:00
- Retake midterm: November 10, 8:00-10:00
- Second retake midterm: December 15.
- A written final exam will be held during the exam period.

Midterms and exams

Midterms:

- There will be one written midterm exam consisting of 6 problems, each worth 20 points.
- You have 90 minutes for the midterm.
- Students receive the semester signature if they achieve at least 40 points in the midterm.
- One retake opportunity during the teaching period, and an additional paid retake opportunity.

Midterms and exams

Exams:

- There will be one written final exam consisting of 6 problems, each worth 20 points.
- You have 100 minutes for the exam.
- Students pass if they achieve at least 40 points on the exam.
- If you pass the written exam, you may take an optional oral exam to adjust your grade (it can raise or lower it)

Grades

The final score calculation:

$$\text{Final score} = 0.4 \cdot \text{Midterm score} + 0.6 \cdot \text{Exam score}.$$

Grades:

- $[40; 55)$: Satisfactory (pass)
- $[55; 70)$: Fair (average)
- $[70; 85)$: Good
- $[85, 100]$: Excellent

More information and lecture note

- Attendance at lectures and practices is **not** mandatory, but **recommended**!
- There are two lecture notes: Probability Theory, Statistics.
- **Lecture notes**, **slides**, **solutions of the exercises**, and more information about the course will be available at:

  www.cs.bme.hu/~bence.csonka/prob25

On the prerequisites

- A significant portion of random quantities are continuous $\rightarrow$ the use of tools from analysis is unavoidable.
- Up to the midterm, we will mainly need differentiation and integration of simple single-variable functions (polynomials, exponential functions, etc.), starting around weeks 4–5.
- Officially, the only prerequisite for taking Probability Theory is Analysis 1, but we will also need certain tools known from Analysis 2, such as partial derivatives and integration in two variables (over normal domains, sometimes changing to polar coordinates), starting around weeks 7–8.
- Most of the basic concepts of probability theory will first be introduced in the discrete case, in the hope that this will also help in understanding the continuous case.
- From previous SZIT courses, we will most often refer back to linear algebra, and occasionally to graph theory and algorithm theory as well.

Introduction

Probability theory first appeared in the study of games of chance.

Example: In a friendly gathering, we can bet on one of two options. After tossing a coin 10 times, we get 10 heads; or after rolling a die 4 times, we get four sixes. What is worth betting on?

Probability theory first appeared in the study of games of chance.
Example: In a friendly gathering, we can bet on one of two options. After tossing a coin 10 times, we get 10 heads; or after rolling a die 4 times, we get four sixes. What is worth betting on?

Correct answer: the coin.

Introduction

Probability theory first appeared in the study of games of chance.

Example: In a friendly gathering, we can bet on one of two options. After tossing a coin 10 times, we get 10 heads; or after rolling a die 4 times, we get four sixes. What is worth betting on?

Correct answer: the coin.

Modern applications include risk analysis, simulation of economic models, networks, and space exploration, among many others.

Motivation for formalization

- There is a family with two children (they are not twins). What is the probability that both children are all girls?

Motivation for formalization

- There is a family with two children (they are not twins). What is the probability that both children are all girls?
- Firstly, we have three outcomes: two boys; a boy and a girl; two girls. The number of favorable cases is 1 and the total number of cases is 3.

Motivation for formalization

- There is a family with two children (they are not twins). What is the probability that both children are all girls?
- Firstly, we have three outcomes: two boys; a boy and a girl; two girls. The number of favorable cases is 1 and the total number of cases is 3.
- The probability: $\frac{1}{3}$.

Motivation for formalization

- There is a family with two children (they are not twins). What is the probability that both children are all girls?
- Firstly, we have three outcomes: two boys; a boy and a girl; two girls. The number of favorable cases is 1 and the total number of cases is 3.
- The probability: $\frac{1}{3}$.
- We have the following outcomes: both children are boys, the first child is a boy and the second child is a girl, the first child is a girl and the second child is a boy, the children are girls.

Motivation for formalization

- There is a family with two children (they are not twins). What is the probability that both children are all girls?
- Firstly, we have three outcomes: two boys; a boy and a girl; two girls. The number of favorable cases is 1 and the total number of cases is 3.
- The probability: $\frac{1}{3}$.
- We have the following outcomes: both children are boys, the first child is a boy and the second child is a girl, the first child is a girl and the second child is a boy, the children are girls.
- Probability $\frac{1}{4}$

Motivation for formalization

- There is a family with two children (they are not twins). What is the probability that both children are all girls?
- Firstly, we have three outcomes: two boys; a boy and a girl; two girls. The number of favorable cases is 1 and the total number of cases is 3.
- The probability: $\frac{1}{3}$.
- We have the following outcomes: both children are boys, the first child is a boy and the second child is a girl, the first child is a girl and the second child is a boy, the children are girls.
- Probability $\frac{1}{4}$

What is the truth?

Randomness

Random:

does *not* mean that what happens has no cause, some of the causes are unknown to us.

- all conditions fixed $\Rightarrow$ deterministic evolution
- only part of the conditions considered $\Rightarrow$ non-deterministic $\Rightarrow$ random

Example: tossing a coin

Randomness

Random:
does *not* mean that what happens has no cause, some of the causes are unknown to us.

- all conditions fixed $\Rightarrow$ deterministic evolution
- only part of the conditions considered $\Rightarrow$ non-deterministic $\Rightarrow$ random

Example: tossing a coin

Random experiment: we know every possible outcome

Basic concepts and notation

Definition

- Ω sample space: the set of possible outcomes (nonempty)
- $\omega \in \Omega$ outcome: the elements of the sample space. The singleton $\{\omega\}$ is called an elementary event.
- $A \subseteq \Omega$ events: certain subsets of the sample space.

An event occurs if an elementary event that belongs to it is realized when the experiment is performed.
An event can be specified by listing its elementary outcomes, or by a logical expression.

Example: die roll — "we roll an even number"; tossing coin 2 times — "we toss two heads".

Basic concepts and notation

Definition

- Ω sample space: the set of possible outcomes (nonempty)
- $\omega \in \Omega$ outcome: the elements of the sample space. The singleton $\{\omega\}$ is called an elementary event.
- $A \subseteq \Omega$ events: certain subsets of the sample space.

An event occurs if an elementary event that belongs to it is realized when the experiment is performed.
An event can be specified by listing its elementary outcomes, or by a logical expression.

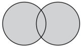
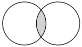
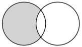
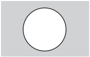
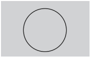
Example: die roll — "we roll an even number"; tossing coin 2 times — "we toss two heads".

If Ω is finite, in most examples we will treat *every* subset of Ω as an event. For more complex sample spaces (e.g. $\Omega = \mathbb{R}$) the situation is subtler $\rightarrow$ see soon.

Operations on events (set operations)

events $=$ sets $\Rightarrow$ operations $\Rightarrow$ set operations

Let A and B be events. Then

- $A \cup B$ (union): occurs if at least one of them occurs 
  ($A \cup B$ occurs $\Leftrightarrow$ A or B occurs)
- $A \cap B$ (intersection): occurs if both occur 
  ($A \cap B$ occurs $\Leftrightarrow$ A and B both occur)
- $A \setminus B$ (difference): occurs if A occurs but B does not 
- $\overline{A} = \Omega \setminus A$ (complement): occurs if A does not occur 
- Ω is the sure event 
- $\emptyset$ is the impossible event 

Properties of the operations

- $\overline{\overline{A}} = A$
- $A \cap A = A \cup A = A$
- $A \cap \Omega = A$, $A \cap \emptyset = \emptyset$
- $A \cup \Omega = \Omega$, $A \cup \emptyset = A$
- intersection is commutative: $A \cap B = B \cap A$ and associative: $A \cap (B \cap C) = (A \cap B) \cap C$; we write $A \cap B \cap C$
- similarly, union is commutative and associative
- intersection is distributive over union: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- conversely, union is distributive over intersection

De Morgan identities

- For two events: $\overline{A \cap B} = \overline{A} \cup \overline{B}$, $\quad \overline{A \cup B} = \overline{A} \cap \overline{B}$
- Also for a countable family of events: $\overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i}$ and $\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \overline{A_i}$

Operations on events — examples

Let A, B, C be three events. Express the following events:

- at least one event occurs: $A \cup B \cup C$
- A and B occur, but C does not: $(A \cap B) \setminus C = (A \cap B) \cap \overline{C}$
- all events occur: $A \cap B \cap C$
- none of the events occurs : $\overline{A} \cap \overline{B} \cap \overline{C} = \overline{A \cup B \cup C}$
- exactly one event occurs:

$$A \cup B \cup C \setminus ((A \cap B) \cup (A \cap C) \cup (B \cap C))$$

Further operations on events

Definition

- $A \subseteq B$ (subset): A implies B; the occurrence of A entails B
- $A \cap B = \emptyset$: A and B are mutually exclusive; they cannot occur simultaneously
- Let $A_1, A_2, \ldots$ be a sequence of events. We say $A_1, A_2, \ldots$ are pairwise disjoint if $A_i \cap A_j = \emptyset$ for all $i \neq j$. (For finitely many events the definition is analogous.)

We said events will be certain subsets of Ω.
Goal: to define "event" so that applying the above operations to events produces events again. This motivates the concept of an event algebra (σ-algebra).

σ-algebra

Let $\Omega \neq \emptyset$ be a fixed set.
Let $\mathcal{P}(\Omega)$ denote the set of all subsets of Ω (the power set of Ω).

Definition

A family $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is called a σ-algebra on Ω (or over Ω) if

- $\Omega \in \mathcal{F}$,
- $\forall A \subseteq \Omega$: $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$,
- and $\forall A_1, A_2, \ldots \subseteq \Omega$: if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Examples:

- the full power set, $\mathcal{F} = \mathcal{P}(\Omega)$,
- the *trivial σ-algebra*, $\mathcal{F} = \{\emptyset, \Omega\}$.

Both are σ-algebras for any $\Omega \neq \emptyset$.
If Ω is finite, then $\mathcal{P}(\Omega)$ is often a good modeling choice.

On choosing a σ-algebra (briefly)

If Ω is finite, then $\mathcal{F} = \mathcal{P}(\Omega)$ is often a good modeling choice.
Factors influencing the choice:

- geometric probability space (see next): Ω is a "nice" subset of the plane $\mathbb{R}^2$, e.g. a rectangle or disk, and probabilities are proportional to area. However, by known measure-theoretic results, "area" cannot be extended to *all* subsets of $\mathbb{R}^2$.
- observability affects what we regard as an event
- for random processes, what can be observed—and thus what we treat as events—may change over time (e.g. when observing a stock price).

Probability space

Definition

Let $\mathcal{F}$ be a σ-algebra on an arbitrary set Ω. A function $\mathbb{P}\colon \mathcal{F} \to [0,1]$ is a probability measure if

- $\mathbb{P}(\Omega) = 1$,
- and if $A_1, A_2, \dots$ is a sequence of events that are pairwise disjoint ($A_i \cap A_j = \emptyset$ for $i \neq j$), then $\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Definition

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space if $\Omega \neq \emptyset$ is a set, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure on $\mathcal{F}$.

Example: die roll

Classical (finite) model: Ω finite, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(A) = |A|/|\Omega| =$ "$\frac{\text{number of favorable cases}}{\text{number of total cases}}$" for all $A \subseteq \Omega$.

Properties of probability measures

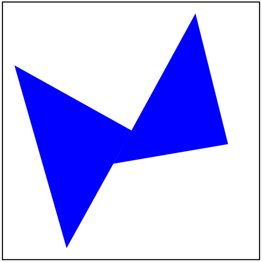
Statement

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space.*

- $A, B \in \mathcal{F}$ *with* $A \cap B = \emptyset$*, then* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$
- $\mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A)$
- *If* $B \subseteq A$*, then* $\mathbb{P}(B) \leq \mathbb{P}(A)$*.*

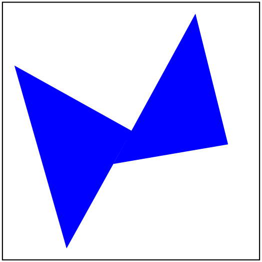
(Proof)

Geometric probability space



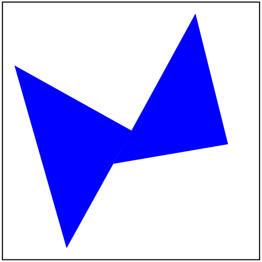
We choose a point "at random" from the white square. What is the probability that it falls into the blue region?

Geometric probability space



In a geometric probability space, Ω is a subset of $\mathbb{R}^2$ (here, the white square) and probability is proportional to area:
$\mathbb{P}$(a uniformly chosen point in Ω falls in the blue region)
$= \text{“}\frac{A_{\text{favorable}}}{A_{\text{total}}}\text{”} = \frac{A_{\text{blue}}}{A_\Omega}$.
$\mathbb{P}$(the chosen point lies in Ω) $= \frac{A_\Omega}{A_\Omega} = 1$.

Geometric probability space



The same formula holds with the blue region replaced by any two-dimensional shape (e.g. polygons, circles and other conic sections, and finite unions/intersections/differences of these shapes).
It also holds for finite or countably infinite subsets of the plane; these have area 0.

Example

Computational examples: We choose a point from a unit square with uniform probability.

Let $A :=$ the chosen point is closer to the middle point then $\frac{1}{2}$.

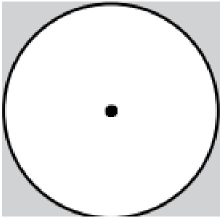
What is $\mathbb{P}(A)$?

Example

Computational examples: We choose a point from a unit square with uniform probability.

Let $A :=$ the chosen point is closer to the middle point then $\frac{1}{2}$.

What is $\mathbb{P}(A)$?



Answer: $\mathbb{P}(A) = \frac{A_{\text{circle}}}{A_{\text{square}}} = \frac{\pi/4}{1} = \frac{\pi}{4}$.

Poincaré formula for two and three events

We have seen: for mutually exclusive events A, B,
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
Question (familiar also from high school): If A and B are not mutually exclusive, how do we compute $\mathbb{P}(A \cup B)$?

Poincaré formula for two and three events

We have seen: for mutually exclusive events A, B,
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
Question (familiar also from high school): If A and B are not mutually exclusive, how do we compute $\mathbb{P}(A \cup B)$?

Statement (Poincaré formula for 2 events)

For any $A, B \in \mathcal{F}$,

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

(Proof)

What about the probability of the union of three events A, B, C?

Poincaré for 3 events

Statement (Poincaré formula for 3 events)

For any $A, B, C \in \mathcal{F}$,

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\ + \mathbb{P}(A \cap B \cap C).\end{aligned}$$

General Poincaré formula

Let $A_1, \ldots, A_n \in \mathcal{F}$ be events.

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) = ?$$

As in the case of two or three sets: „the probability of intersections of an odd number of sets enters with a positive sign, intersections of an even number with a negative sign."

General Poincaré formula

Formally: for a natural number n, let $[n] \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$. Thus $[0] = \emptyset$.

$$S_k := \sum_{\substack{I \subseteq [n] \\ |I| = k}} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right).$$

Theorem (Poincaré formula)

For arbitrary events $A_1, A_2, \ldots, A_n \in \mathcal{F}$,

$$\mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) = \sum_{k=1}^{n} (-1)^{k+1} S_k.$$

Bonferroni inequalities

We can obtain information about the probability of a union even if we only take the first few terms of the sum and neglect the rest (for example, because in practice they are very small and slow to compute):

Statement (Bonferroni inequalities)

Let $A_1, \ldots, A_n \in \mathcal{F}$ be events, and let $1 \leq m_1, m_2 \leq n$ be integers, where m_1 is odd and m_2 is even. Then

$$\mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) \leq \sum_{k=1}^{m_1} (-1)^{k+1} S_k \qquad \textit{and} \qquad \mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) \geq \sum_{k=1}^{m_2} (-1)^{k+1} S_k.$$

Bonferroni inequalities

We can obtain information about the probability of a union even if we only take the first few terms of the sum and neglect the rest (for example, because in practice they are very small and slow to compute):

Statement (Bonferroni inequalities)

Let $A_1, \ldots, A_n \in \mathcal{F}$ be events, and let $1 \leq m_1, m_2 \leq n$ be integers, where m_1 is odd and m_2 is even. Then

$$\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) \leq \sum_{k=1}^{m_1} (-1)^{k+1} S_k \qquad \text{and} \qquad \mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) \geq \sum_{k=1}^{m_2} (-1)^{k+1} S_k.$$

In other words: if we sum the terms of the Poincaré formula up to an odd index m_1, we always get an upper bound. If we sum up to an even index m_2, we always get a lower bound.

Example for Bonferroni

As an illustration: for $n = 3$, the Poincaré formula implies:

- $m_1 = 1$: $\mathbb{P}\Big(A_1 \cup A_2 \cup A_3\Big) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3)$,
- $m_2 = 2$: $\mathbb{P}\Big(A_1 \cup A_2 \cup A_3\Big) \geq$ $\mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3)$,
- $m_1 = 3$: we obtain the Poincaré formula for 3 events (with $\leq$ instead of $=$, which is of course also true).

Boole's inequality

We will not prove the Poincaré formula now. It can be shown by induction, for example. Later we will give a straightforward proof for the 3-event case, which can be generalized to *n* events.

The following corollary shows that the sum of probabilities is always an upper bound for the probability of the union – even if the events are not mutually exclusive:

Statement (Boole's inequality)

For $A_1, A_2, \ldots, A_n \in \mathcal{F}$*,*

$$\mathbb{P}\Big(\bigcup_{j=1}^{n} A_j\Big) \leq \sum_{j=1}^{n} \mathbb{P}(A_j), \qquad \mathbb{P}\Big(\bigcap_{j=1}^{n} A_j\Big) \geq 1 - \sum_{j=1}^{n} \mathbb{P}(\overline{A_j}).$$

The first statement will not be proved here. The second follows from the first and the De Morgan laws (applied to the events $\overline{A_j}$).