BIE-PST – Probability and Statistics

Lecture 1: Basic notions of probability

Winter semester 2024/2025

Lecturer:
Francesco Dolce

Department of Applied Mathematics
Faculty of Information Technology
Czech Technical University in Prague

Table of contents

Course organization

Evaluation

Requirements for passing the course

- **Tutorials:**
  - there will be 6 small tests, each for 6 points, the 5 best results will count – 30p
  - homework assignment – 10p
  - needed at least 20p from 40p possible.
- **Exam:**
  - compulsory written exam max 60p – at least 30p needed
  - points from exam and tutorials will be added
  - optional theoretical exam – max 5p
  - taking the theoretical exam is possible only after successfully passing the written exam.

Bibliography

English books:

- D. P. Bertsekas & J. N. Tsitsiklis: *Introduction to Probability, Athena Scientific* MIT (2008)
- G. R. Grimmett & D. R. Stirzaker: *Probability and Random Processes*, Oxford University Press (2001)
- Ch. M. Grinstead & J. L. Snell: *Introduction to Probability*, AMS (1997) – (online)

1 Basic notions of probability

1.1 Motivation

Probability and statistics

Goal: Achieve better understanding of the world in situations where *randomness* is involved.

What do we understand as *random*? In real life we often encounter processes (experiments, tests, natural phenomena, ...), for which *we are not sure*, how they will end and which result will occur.

Exact prediction may not be possible because they are either too complex or we do not have all necessary information available.

Usually we say that the result is unpredictable or *random* and is given by *chance*.

Probability theory vs mathematical statistics

Probability theory quantifies the unpredictability from a mathematical point of view.

- Outcomes of an experiment are assigned *probability*, giving the ideal proportion of cases when the outcome will occur.
- Starting from simple models, complex problems may be solved.
- E.g., if we know that a coin is balanced, we can compute what is the probability of getting 100× Heads out of 1000 tosses.

Mathematical statistics estimates the unpredictability using experimental data.

- Outcomes of a repeated experiment are used for estimation.
- Probabilistic models are suggested and verified.
- E.g., if we get only 100× Heads out of 1000 tosses, is it enough evidence to say that a coin is not balanced?

1.1.1 Classical definition of probability

- Finite number n of *mutually different* results (outcomes) of some experiment.
- We suppose that each outcome has *the same probability* of occurring.
- If exactly m of the outcomes satisfy realization of the event A (e.g., 6 rolled two times in 4 rolls) then we define the *probability of the event A* as

$$\mathrm{P}(A) = \frac{m}{n} = \frac{\text{number of favorable outcomes}}{\text{number of all outcomes}}.$$

Imperfection of the definition:

- What to do if the die is unbalanced?
- What to do if there are infinitely many possible results?

Example 1.1 (– Toss with two coins)**.** *What is the probability that at least one head appears?*
There are three outcomes with at least one heads. In total there are 4 possibilities. Thus:

$$\mathrm{P}(\text{at least one heads}) = \frac{3}{4}.$$

Example 1.2 (– Rolling a six-sided dice)**.** *What is the probability of rolling an even number?*
There are three even numbers (2,4,6). Totally there are six possibilities. Thus:

$$\mathrm{P}(\text{even}) = \frac{3}{6} = \frac{1}{2}.$$

✓ *Recall yourself the basic combinatorics!*

1.1.2 Recap of Combinatorics

Consider a set of n elements, where $n \in \mathbb{N}$.

- The number of ways to order these elements (*permutations*) is $n!$
- The number of ways to select k elements without repetitions when the order is important (*variations*) is $\frac{n!}{(n-k)!}$.
- The number of ways to select k elements with repetitions when the order is important (*variations with repetition*) is n^k.
- The number of ways to select k elements without repetitions when the order is not important (*combinations*) is $\binom{n}{k}$.
- The number of ways to select k elements with repetitions when the order is not important (*combinantions with repetition*) is $\binom{n+k-1}{k}$.

1.1.3 Geometric definition of probability

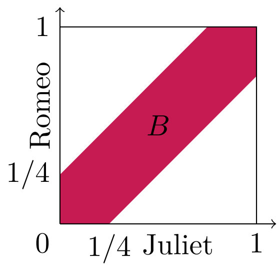
- The outcomes appear in some geometric object of a *finite size* (length, area, capacity) S.
- Each outcome (point) has *equal probability* of occurring.
- If S_B stands for the size of the set of outcomes satisfying the realization of event B, then we define *probability of the event B* as

$$\mathrm{P}(B) = \frac{S_B}{S} = \frac{\text{size of the favorable outcomes set}}{\text{size of the all outcomes set}}.$$

Imperfection of the definition:

- How to introduce unequal distribution of probability?
- What to do with objects of infinite size?

Example 1.3 (– Romeo and Juliet)**.** Romeo and Juliet have to meet at a secret place between midday and 1 p.m. Each of them arrives at a random moment between midday and 1 p.m., but will wait for only 15 minutes. If the partner does not arrive, the waiting person will leave. *What is the probability that the two people will actually meet?*



The total area is 1. The coloured area corresponds to the successful meeting.

$$\mathrm{P}(B) = \frac{1 - (3/4) \cdot (3/4)}{1} = \frac{7}{16}.$$

2 Axiomatic definition of probability

2.1 Sample space

The two above definitions can solve many basic situations. However, we want to establish a general theoretical background which can be easily extended.

For a proper and general definition of probability we need to correctly define *events* – sets to which we will assign a probability.

Definition 2.1. The set of all possible outcomes of an experiment is called the *sample space* and is denoted by Ω.

An arbitrary possible result $\omega \in \Omega$ is called an *outcome* (elementary event).

The outcomes in Ω should always be *mutually exclusive* and *exhaustive.*
The first step always consists of:
a decision which possibilities we can observe and distinguish. This determines Ω.

Outcomes in Ω should always be
mutually exclusive and *exhaustive.*

Mutually exclusive: consider, e.g., outcomes: on a die we rolled 1 *or* 2, 1 *or* 3,. . . ? If 1 is rolled, it is not clear which outcome it should be!

Exhaustive: each result of the experiment should be interpretable as an outcome.

For tossing a coin we actually might have $\Omega = \{H, T, E\}$, where E denotes the result when the coin stops at the edge.

The Sample space should be detailed enough to distinguish between different results but it should ignore unimportant details.

Examples 2.2 (– sample spaces)**.**

- Toss with a coin: $\Omega = \{\mathrm{H}, \mathrm{T}\}$
- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Toss with two coins: $\Omega = \{\mathrm{H}, \mathrm{T}\} \times \{\mathrm{H}, \mathrm{T}\} \equiv \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H}), (\mathrm{T}, \mathrm{T})\}$
- Height of the missile above the earth surface: $\Omega = [0, +\infty)$
- Random text in email in `UTF-32` encoding (constant length) of maximal length 1MB. The maximal number of characters in the message is

$$\frac{1\ \mathrm{MB}}{32\ \mathrm{bits}} = \frac{2^{20}\ \mathrm{bytes}}{4\ \mathrm{bytes}} = 2^{18} = 262144.$$

  Thus: $\Omega = \{(x_1, x_2, \ldots, x_{2^{18}}) \,|\, x_i \in \mathbb{N}, 0 \le x_i < 2^{32},\ \text{for all } i\}$

Examples 2.3 (– sample spaces)**.**

- Series of n rolls with a die: $\Omega = \{1, 2, 3, 4, 5, 6\}^n$

- Series of n rolls with a die, where we are interested only in numbers of appearance of each side:
$$\Omega = \left\{ (k_1, k_2, k_3, k_4, k_5, k_6) \in \mathbb{Z}_+^6 : \sum_{l=1}^{6} k_l = n \right\}.$$

- Throwing darts into $T \subset \mathbb{R}^2$: $\Omega = T \cup \{*\}$, where$\{*\}$ is a one point set representing the outcome "the dart does not reach the target".

  If the target is divided into, say, 5 strips and we are interested only in which strip was reached: $\Omega = \{1, 2, 3, 4, 5, *\}$.

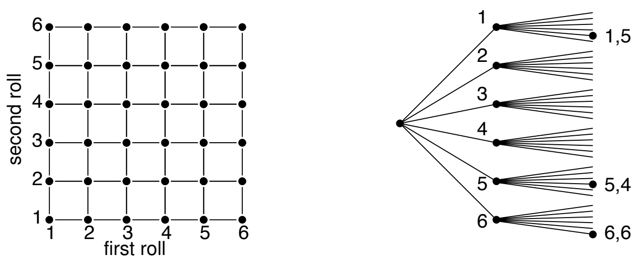
- Tossing a coin until first Tails appears: countable space $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$, where ω_i means that the first $i-1$ tosses are Heads and the i-th toss is Tail.

- Tossing a coin infinitely many times: countable space $\Omega = \{H, T\}^{\mathbb{N}}$.

Visualization of an outcome of a series of experiments

Two rolls of a die:

A *coordinate* description or a sketch in the form of a *tree* where each sequence of results of particular rolls corresponds to a single leaf that is uniquely determined by the path from the root to the leaf (in the illustration, only 3 leaves corresponding to 3 outcomes are explicitly marked).



2.2 Random events

An *event* A is some collection of outcomes – i.e., a subset $A \subset \Omega$.

Example 2.4 (– Toss with two coins)**.** *Express the event "at least one Head appears" as a set.*

The sample space is $\Omega = \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H}), (\mathrm{T}, \mathrm{T})\}$.

The event A denoting "at least one Head appears" is:
$$A = \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H})\} \subset \Omega.$$

Example 2.5 (– Rolling a die (6-sided))**.** *Express the event "an even number appears" as a set.*

Even numbers are 2,4, and 6. The event A denoting that "an even number appears" is
$$A = \{2, 4, 6\} \subset \Omega.$$

Operations with events

It is possible to apply all classical set operations on events (events are sets).
In probability theory, a specific terminology is used for this operations:

- A^c complement of A – *no outcome in A occurs*
- $A \cap B$ intersection of A and B – *both A and B occur*
- $A \cup B$ union A and B – *either A or B or both*
- $A \setminus B$ difference A and B – *A but not B*
- $A \subset B$ subset – *if A then B*
- $\emptyset$ empty set – *impossible event*
- Ω collection of objects – *whole sample space*
- ω member of Ω – *outcome, elementary event*

Structure of events

It is *superfluous* and sometimes impossible to *demand all subsets* of Ω to be events.

We want all above mentioned set operations to be meaningful even for countable repetitions of any set of events. It can be shown that it is enough to take the events as elements of some *σ-algebra $\mathcal{F}$*:

Definition 2.6. A system $\mathcal{F}$ containing subsets of Ω is called a *σ-algebra* if the following conditions hold:

i) $\emptyset \in \mathcal{F}$ – contains the impossible event

ii) if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_i A_i \in \mathcal{F}$ – contains any countable union of events

iii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ – contains any complementary event

Clearly: $\Omega \in \mathcal{F}$ – contains the sample space itself.
When specifying a probability model we always consider a couple $(\Omega, \mathcal{F})$, called the *measurable (observatory) space.*

The choice of $\mathcal{F}$ informs us which events can be "observed" and have their probability measured. The subsets of Ω which are not contained in $\mathcal{F}$ are not events and we cannot measure their probability.

Examples 2.7 (– possible σ-algebras).

- $\mathcal{F} = \{\emptyset, \Omega\}$ is a σ-algebra.
- For an arbitrary $A \subset \Omega$, $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ is a σ-algebra.
- $\mathcal{F} = 2^\Omega$ – all subsets create a σ-algebra;
- Borel σ-algebra on $\mathbb{R}$ – smallest σ algebra containing all open intervals.

2.3 Probability measure

Definition 2.8. A *probability measure* P on $(\Omega, \mathcal{F})$ is a function $\mathrm{P} : \mathcal{F} \to \mathbb{R}$ satisfying:

i) *non-negativity:* for all $A \in \mathcal{F}$ it holds that $\mathrm{P}(A) \geq 0$

ii) *normalization:* $\mathrm{P}(\Omega) = 1$,

iii) *σ−additivity:* if $A_1, A_2, \ldots \in \mathcal{F}$ is a collection of disjoint events (i.e., if $A_i \cap A_j = \emptyset$ for $\forall\, i, j$ with $i \neq j$), then

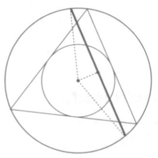
$$\mathrm{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} \mathrm{P}(A_i).$$

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

The probability can also be given as a percentage between 0% and 100%.

The choice of P determines what we understand as *"random"*. Vague assignment can lead to "paradoxes".

Example 2.9 (– Bertrand paradox (Joseph Bertrand, 1889))**.**



What is the probability that a *randomly placed chord* χ on the unit circle will be longer than an the side ℓ of an equilateral triangle in the unit circle? I.e., what is $\mathrm{P}(A)$, where $A = \{|\chi| > \ell\}$.
It depends on what we understand as "random":

Option 1: We choose randomly uniformly the *centre* of χ:

$\Omega_1 = \{x \in \mathbb{R}^2 \,:\, |x| < 1\}, \quad A_1 = \{x \in \Omega_1 \,:\, |x| < 1/2\}, \quad \mathrm{P}_1(A_1) = \dfrac{\pi(1/2)^2}{\pi 1^2} = \dfrac{1}{4}.$

Option 2: We choose randomly uniformly the *angle* and the direction (irrelevant thanks to the rotation symmetry) of chord χ observed from the circle centre:

$\Omega_2 = (0, \pi], \quad A_2 = (2\pi/3, \pi], \quad \mathrm{P}_2(A_2) = \dfrac{\pi/3}{\pi} = \dfrac{1}{3}.$

Option 3: We choose randomly uniformly the *distance of the chord χ from the circle centre* and (again irrelevant) the direction:

$\Omega_3 = [0, 1), \quad A_3 = [0, 1/2), \quad \mathrm{P}_3(A_3) = \dfrac{1}{2}.$

Intermezzo

How to establish $\mathcal{F}$?

- Finite or countable Ω:
  - We can take $\mathcal{F}$ as all subsets of Ω. ($\mathcal{F} = 2^{\Omega}$, i.e., power set of Ω)
  - Events are arbitrary subsets of Ω.

- For an uncountable Ω:
  - It is not possible to assign a positive probability to each $A \subset \Omega$, because then we would have $\mathrm{P}(\Omega) = \infty$.

- If $\Omega \subset \mathbb{R}^d$ is some subinterval of $\mathbb{R}^d$ (e.g., $[0, +\infty)$) we can take $\mathcal{F}$ as the Borel σ-algebra on Ω.
- Except for an at most a countable subset, each singular point must have a zero probability.

Definition of probability for "classical" settings
Definition of probability for uniform distribution of finite number of outcomes:
If Ω is finite with equally likely realizations:

$$\mathrm{P}(A) = \frac{\#A}{\#\Omega} = \frac{\text{number of favorable outcomes}}{\text{number of all outcomes}}.$$

Geometric definition of probability:
Let Ω be any arbitrary space with a measure μ, i.e., we can measure size (area, capacity, etc.), and $0 < \mu(\Omega) < +\infty$. For any event $A \subset \Omega$ we define:

$$\mathrm{P}(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{\text{size of } A}{\text{size of } \Omega}.$$

It can be easily verified that both approaches satisfy the formal definition of probability as stated above.

2.4 Properties of probability

Theorem 2.10. *Let A and B be events on a probability space with measure* P*. Then it holds that:*

i) $\mathrm{P}(\emptyset) = 0$

ii) If A and B are disjoint, then $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$

iii) $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$

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

v) if $A \subset B$, then $\mathrm{P}(A) \leq \mathrm{P}(B)$ – monotonicity

Consequences:

- $0 \leq \mathrm{P}(A) \leq 1$ – from *v)*
- $\mathrm{P}(A \cup B) \leq \mathrm{P}(A) + \mathrm{P}(B)$ – from *iv)*

Proof. i) We create a sequence of disjoint events: $A_i = \emptyset$ for all $i \in \mathbb{N}$. From property *iii) of probability measure* we have

$$\mathrm{P}(\emptyset) = \mathrm{P}\left(\bigcup_{i=1}^{+\infty} \emptyset\right) = \sum_{i=1}^{+\infty} \mathrm{P}(\emptyset),$$

it can be fulfilled only for $\mathrm{P}(\emptyset) = 0$.

ii) We create sequence of disjoint events: $A_1 = A$, $A_2 = B$ and $A_i = \emptyset$ for $i > 2$. From properties *i)* and *iii)* of probability measure we have

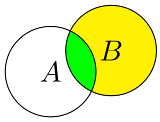
$$\mathrm{P}(A \cup B) = \mathrm{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} \mathrm{P}(A_i) = \mathrm{P}(A) + \mathrm{P}(B).$$

iii) $1 = \mathrm{P}(\Omega) = \mathrm{P}(A \cup A^c) = \mathrm{P}(A) + \mathrm{P}(A^c)$. Thus $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$.

□

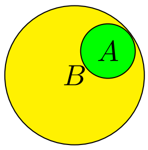
Proof. iv) The set $A \cup B$ can be written as the disjoint union $A \cup (B \setminus A)$.
From *ii)* it follows that $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B \setminus A)$.
Since $\mathrm{P}(B) = \mathrm{P}(B \setminus A) + \mathrm{P}(A \cap B)$,
we finally have:

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



v) If $A \subset B$, then $A \cap B = A$.

$$\mathrm{P}(B) = \mathrm{P}(B \setminus A) + \mathrm{P}(A \cap B) = \mathrm{P}(B \setminus A) + \mathrm{P}(A) \geq \mathrm{P}(A)$$



□

Theorem 2.11. *Let $A_1, A_2, \ldots$ be events on a probability space with measure* P*. Then it holds that:*

i) $\sigma-$sub additivity:

$$\mathrm{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) \leq \sum_{i=1}^{+\infty} \mathrm{P}(A_i).$$

ii) Inclusion – exclusion principle:

$$\mathrm{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\substack{J \subset \{1,2,\ldots,n\} \\ J \neq \emptyset}} (-1)^{|J|-1}\, \mathrm{P}\left(\bigcap_{i \in J} A_i\right).$$

For 3 events: $\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(A \cap B) - \mathrm{P}(A \cap C) - \mathrm{P}(B \cap C) + \mathrm{P}(A \cap B \cap C)$

Proof. i) The set $\bigcup_{i=1}^{+\infty} A_i$ can be written as a disjoint union $\bigcup_{i=1}^{+\infty} \left(A_i \setminus \bigcup_{k=1}^{i-1} A_k\right)$. From property *iii)* of probability measure we have

$$\mathrm{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) = \mathrm{P}\left(\bigcup_{i=1}^{+\infty} \left(A_i \setminus \bigcup_{k=1}^{i-1} A_k\right)\right) = \sum_{i=1}^{+\infty} \mathrm{P}\left(A_i \setminus \bigcup_{k=1}^{i-1} A_k\right) \leq \sum_{i=1}^{+\infty} \mathrm{P}(A_i).$$

ii) We prove the statement for 3 events A, B, C.

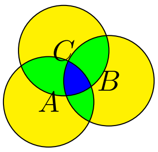
The set $A \cup B \cup C$ can be written as a disjoint union $A \cup ((B \cup C) \setminus A)$.
From point *ii)* of the previous theorem it follows that

$$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A) + \mathrm{P}((B \cap C) \setminus A).$$

Since $\mathrm{P}(B \cup C) = \mathrm{P}((B \cup C) \setminus A) + \mathrm{P}(A \cap (B \cup C))$, we have:

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

and again, applying point *ii)* of the previous theorem to $\mathrm{P}(B \cup C)$ and to $\mathrm{P}((A \cap B) \cup (A \cap C))$ $\quad \mathrm{P}(B \cup C) = \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(B \cap C)$
$\mathrm{P}((A \cap B) \cup (A \cap C)) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap C) - \mathrm{P}(A \cap B \cap A \cap C)$.



finally we have:
$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(A \cap B) - \mathrm{P}(A \cap C) - \mathrm{P}(B \cap C) + \mathrm{P}(A \cap B \cap C)$.
□

2.5 Continuity of probability

Theorem 2.12. *Let $A_1, A_2, \ldots$ be a sequence of events* increasing *in the sense of inclusion, i.e., such that $A_1 \subset A_2 \subset A_3 \subset \ldots$. If we denote*

$$A = \bigcup_{i=1}^{+\infty} A_i,$$

then $\mathrm{P}(A) = \lim_{n \to +\infty} \mathrm{P}(A_n)$.

Similarly, let $B_1, B_2, \ldots$ be a sequence of events decreasing *in the sense of inclusion, i.e., such that $B_1 \supset B_2 \supset B_3 \supset \cdots$. If we denote*

$$B = \bigcap_{i=1}^{+\infty} B_i,$$

then $\mathrm{P}(B) = \lim_{n \to +\infty} \mathrm{P}(B_n)$.

Proof. We prove the first part of the statement: The set A can be written as the disjoint union $A = \bigcup_{i=1}^{+\infty} (A_i \setminus A_{i-1})$ Then it holds:

$$\mathrm{P}(A) = \mathrm{P}\left(\bigcup_{i=1}^{+\infty} (A_i \setminus A_{i-1})\right) = \sum_{i=1}^{+\infty} \mathrm{P}(A_i \setminus A_{i-1}) = \lim_{n \to \infty} \sum_{i=1}^{n} \mathrm{P}(A_i \setminus A_{i-1})$$
$$= \lim_{n \to \infty} \mathrm{P}(A_n)$$

We prove the second part of the statement by means of the De Morgan rules and the proof of the first part:

$$\mathrm{P}(B) = \mathrm{P}\left(\bigcap_{i=1}^{+\infty} B_i\right) = \mathrm{P}\left(\left(\bigcup_{i=1}^{+\infty} B_i^c\right)^c\right) = 1 - \mathrm{P}\left(\bigcup_{i=1}^{+\infty} B_i^c\right)$$

The sets B_i^C satisfy the assumptions of the first part of the Theorem and thus:

$$\mathrm{P}(B) = 1 - \lim_{n\to\infty} \mathrm{P}(B_n^c) = \lim_{n\to\infty} (1 - \mathrm{P}(B_n^c)) = \lim_{n\to\infty} \mathrm{P}(B_n).$$

□