# NMAI059 Probability and statistics 1
# Class 1

Robert Šámal

# Overview

Organization

Probability – An introduction

Conditional probability

Bonus

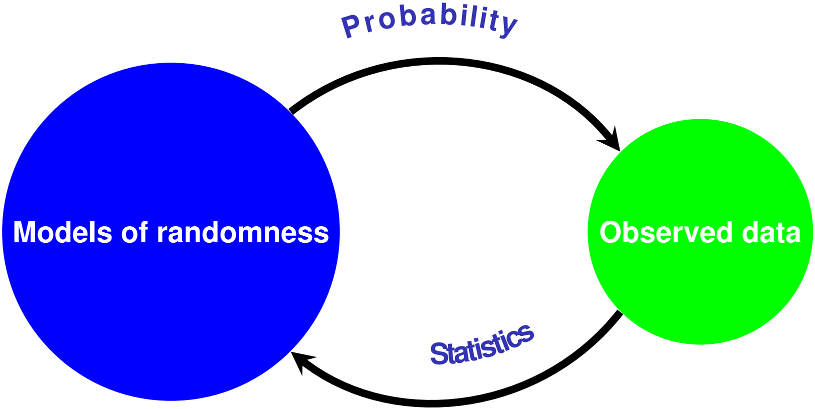
# Organization of the class

- Lectures in Zoom as long as needed (probably the whole semester). The lecture has its page in Moodle (link in SIS). Everything will be there. (Czech and English classes have different Moodle pages!)
- If there are no technical complications, the video recording of each lecture will be available (after logging in to SIS).
- If you mind being recorded, you can turn off your camera or ask questions in the chat instead of by audio.
- But I'll be happy if you turn on the camera to see how slowly/fast I'm talking, what surprised you, etc.
- Also use the Zoom functions: raise hand, slower/faster.
- We will use short polls during the lecture.
- Pdf version of the "board" will also be available – before the lecture, and after with my hand-written comments.
- The exam will ideally be a normal written exam with the possibility of an oral examination.
- There is also place in Moodle to discuss any issues. Alternatively, contact me by email.

# Organization of the tutorials

- Details provided by your TA

# Lecture overview



# Overview

Organization

Probability – An introduction

Conditional probability

Bonus

# A warm-up application

### Example

*Given two degree $d$ polynomials, $f(x), g(x)$, We want to find out, whether they are equal, as fast as possible.*

# Probability – intuition, definition

Some events we can’t/don’t want to describe causally:

- a die roll
- three die rolls, infinitely many die rolls
- throwing a dart on a dartboard
- number of emails received in a day
- running time of an algorithm (in a real computer)

---

Reasons:

- physical properties of nature
- complicated process (weather, medicine, gas molekules)
- unknown influences (other people, programs, weather conditions, . . . )
- randomized algorithms (polynomial identity testing, primality test, quicksort)
- random graphs (estimates of Ramsey numbers)

---

For description by probability theory we first choose a *sample space* $\Omega$. (An $\omega \in \Omega$ is usually called an elementary event.)

# Event space

Next we choose an *event space* $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ – set of events for which we want to measure probability.
Often $\mathcal{F} = \mathcal{P}(\Omega)$, this is possible when $\Omega$ is countable. But for instance for $\Omega = \mathbb{R}$ we must choose fewer sets to an event space.

## Definition

*$\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a sample space (also called $\sigma$-algebra), if*

- *$\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$,*
- *$A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$, and*
- *$A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

# Axioms of probability

### Definition

*$P : \mathcal{F} \to [0,1]$ is called a probability, if*

- *$P(\emptyset) = 0$, $P(\Omega) = 1$, and*
- *$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, for any sequence of pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{F}$.*

### Definition

*Probability space is a triple $(\Omega, \mathcal{F}, P)$ such that*

- *$\Omega \neq \emptyset$ is a set,*
- *$\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a sample space,*
- *$P$ is a probability.*

# Terminology

- *Odds* of an event $A$ is $O(A) = \frac{P(A)}{P(A^c)}$. E.g., having odds to win a race 1 to 2 means that the probability of win is $1/3$; odds for a six on a die is 1 to 5.

- We say $A$ occurs *almost surely (a.s.)* if $P(A) = 1$.

# Basic properties

### Theorem

*Given a probability space* $(\Omega, \mathcal{F}, P)$ *and* $A, B \in \mathcal{F}$ *we have*

1. $P(A) + P(A^c) = 1 \qquad (A^c = \Omega \setminus A)$
2. $A \subseteq B \Rightarrow P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ *(subaditivity, Boole inequality)*

# Examples of a probability space 1

- **Finite with a uniform probability**
  $\Omega$ is any finite set, $\mathcal{F} = \mathcal{P}(\Omega)$, $P(A) = |A|/|\Omega|$.
- **Discrete**
  $\Omega = \{\omega_1, \omega_2, \dots\}$ is any countable set. We are given $p_1, p_2, \dots \in [0, 1]$ that sum to 1.
  $P(A) = \sum_{i:\omega_i \in A} p_i$

# Examples of a probability space 2

- **Continuous**
  $\Omega \subseteq R^d$ for some $d$ (e.g. $\Omega$ open or closed)
  appropriate $\mathcal{F}$ (e.g. having all open and closed sets)
  $f : \Omega \to [0,1]$ is a function such that $\int_\Omega f(x)dx = 1$.
  $P(A) = \int_A f(x)dx$

  Special case $f(x) = 1/V_d(\Omega)$
  $P(A) = V_d(A)/V_d(\Omega)$,
  where $V_d(A) = \int_A 1$ is the $d$-dimensional volume of $A$.
- **Bernoulli cube – infinite repetition**
  $\Omega = S^{\mathbb{N}}$, here $S$ is discrete with probability $Q$,
  appropriate $\mathcal{F}$ (contains all sets of form
  $A = A_1 \times \cdots \times A_k \times S \times S \times \cdots$
  $P(A) = Q(A_1) \cdots Q(A_k)$

  Example: $\{0,1\}^{\mathbb{N}}$ infinite sequence of coin-tossing

# Non-examples

- **A random integer** can be chosen by several ways. In this class we will meat geometric and Poisson distribution. But we cannot require, that all integers are equally likely. (Why?) "A random integer is even with probability $1/2$." ???
- **A random real number** Again, there is no preferred way, how to define probability for $\Omega = \mathbb{R}$.
  For usual definitions, each real number has probability 0! Moreover, it is not possible to define the probability so, that it is translation-invariant, i.e., $P([0,1]) = P([1,2]) = \dots$
- **Random chord of a circle – Bertrand paradox**
  We choose a random chord of a given circle. What is the probability that it is longer than the side of an inscribed triangle?

# Overview

Organization

Probability – An introduction

Conditional probability

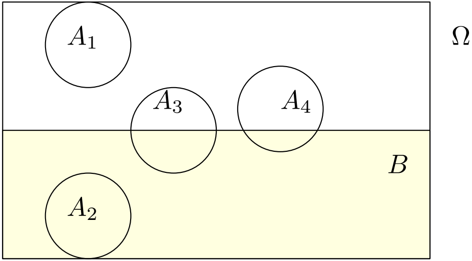
Bonus

# Conditional probability

### Definition

*Given $A, B \in \mathcal{F}$ with $P(B) > 0$, we define probability of $A$ given $B$ as*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$



- $Q(A) := P(A \mid B)$. Then $(\Omega, \mathcal{F}, Q)$ is a probability space.

# Chain rule

- $P(A \cap B) = P(B)P(A \mid B)$

## Theorem

*If* $A_1, \ldots, A_n \in \mathcal{F}$ *and* $P(A_1 \cap \cdots \cap A_n) > 0$*, then*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) =$$

$$P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \ldots P(A_n \mid \bigcap_{i=1}^{n-1} A_i)$$

# Law of total probability

## Definition

*Countable family of sets $B_1, B_2, \ldots \in \mathcal{F}$ is a partition of $\Omega$, if*

- $B_i \cap B_j = \emptyset$ *for* $i \neq j$ *and*
- $\bigcup_i B_i = \Omega$.

## Theorem

*If $B_1, B_2, \ldots$ is a partition of $\Omega$ and $A \in \mathcal{F}$, then*

$$P(A) = \sum_i P(A \mid B_i) P(B_i)$$

*(terms with $P(B_i) = 0$ are counted as 0).*

# Law of total probability

# Bayes' rule

### Theorem

*Let* $B_1, B_2, \ldots$ *be a partition of* $\Omega$*,* $A \in \mathcal{F}$ *and* $P(A), P(B_j) > 0$*. Then*

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{\sum_i P(A \mid B_i)P(B_i)}$$

*(terms with* $P(B_i) = 0$ *are counted as 0).*

# Bayes' rule

# Independent events

### Definition

*Events $A, B \in \mathcal{F}$ are independent if $P(A \cap B) = P(A)P(B)$.*

- Then we also have $P(A \mid B) = P(A)$, provided $P(B) > 0$.

# Mutually independent events

### Definition

*Events $\{A_i : i \in I\}$ are (mutually) independent if for every finite set $J \subseteq I$*

$$P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

*If the condition is true only for sets $J$ with $|J| = 2$, we call the collection $\{A_i\}$ pairwise independent.*

# Continuity of probability

## Theorem

*Suppose that events in $\mathcal{F}$ satisfy*

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

*and $A = \cup_{i=1}^{\infty} A_i$. Then we have*

$$P(A) = \lim_{i \to \infty} P(A_i).$$

- $A_n \subset \{H, T\}^{\mathbb{N}}$, $A_n =$ in the first $n$ tosses there was at least one tail.

# Overview

Organization

Probability – An introduction

Conditional probability

Bonus

# Borel-Cantelli lemma

### Theorem

*Suppose events* $A_1, A_2, \dots$ *satisfy* $P(A_i) = p_i > 0$ *for each* $i$*. Let* $None$ *be the event "none of events* $\{A_i\}$ *occured" and* $Inf$ *the event "infinitely many among* $\{A_i\}$ *occured".*

1. *If* $\sum_i p_i < \infty$*, then* $P(Inf) = 0$*.*
2. *If events* $A_1, A_2, \dots$ *are mutually independent and* $\sum_i p_i = \infty$*, then* $P(None) = 0$*,* $P(Inf) = 1$*.*