# NMAI059 Probability and statistics 1
# Class 2

Robert Šámal

# What we have learned

- definition of a probability space $(\Omega, \mathcal{F}, P)$: two axioms
- **naive** probability space: $\Omega$ finite, $\mathcal{F} = \mathcal{P}(\Omega)$,
  $P(A) := |A|/|\Omega|$
- **discrete** probability space: $\Omega = \{\omega_1, \omega_2, \dots\}$ , $\mathcal{F} = \mathcal{P}(\Omega)$,
  $\sum\limits_i p_i = 1$
  $P(A) := \sum\limits_{i:\omega_i \in A} p_i$
- **geometric** probability space:
  $\Omega \subseteq \mathbb{R}^d$ with a finite volume,
  $P(A) := V_d(A)/V_d(\Omega)$
- probability space **continuous with density**:
  $\Omega \subseteq \mathbb{R}^d$ with a function $f$, where $\int_\Omega f = 1$,
  $P(A) := \int_A f$

# What we have learned: Basic properties

In a probability space $(\Omega, \mathcal{F}, P)$ we have for $A, B \in \mathcal{F}$

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

- We define conditional probability (when $P(B) > 0$).

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

- $Q(A) = P(A \mid B)$ satisfies the axioms of probability

# Overview

Conditional probability

Discrete random variables

Examples of discrete r.v.'s

# 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*

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

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

- Ex.: we pick 3 cards from a deck of 52. What is P(no heart)?

# 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 – exhausting all possibilities

- Application 1. We have three coins: H+T, H+H, T+T, we choose from them at random. What is the probability that we toss a tail?

# Law of total probability – “wishful thinking”

- Application 2. Gambler’s ruin.
  We have $a$ CZK (crowns), our opponent $b$ CZK. We play repeatedly a fair game for 1 CZK, until someone loses all his/her money. What is the probability that we win?

# Bayes' rule

### Theorem

*Let $B_1, B_2, \dots$ 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 $P(B) > 0$.

Ex.: we toss a coin twice.

- $A = \{\omega \in \Omega : \omega_1 = H\} =$ "first toss was a head"
- $B = \{\omega \in \Omega : \omega_2 = H\} =$ "second toss was a head"
- $C = \{\omega \in \Omega : \omega_1 \neq \omega_2\} =$ "exactly one toss was a head"

# 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

Conditional probability

Discrete random variables

Examples of discrete r.v.'s

# Random variable

Often we are interested in a number given as a result of a random experiment.

- We throw a dart and measure the distance from the center of the dartboard.
- We roll a die untill we get a six, then count how many rolls it took.
- In a quicksort algorithm (with a random choice of pivot) we measure the number of operations.

## Definition

*Given a probability space $(\Omega, \mathcal{F}, P)$. We call a function $X : \Omega \to \mathbb{R}$ a discrete random variable, if $Im(X)$ (range of $X$) is a countable set and if for every real $x$ we have*

$$\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}.$$

# PMF

## Definition

*Probability mass function, pmf of a discrete random variable* $X$ *is a function* $p_X : \mathbb{R} \to [0,1]$ *such that*

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

- $\sum_{x \in Im(X)} p_X(x) =$ ?
- $S := Im(X) \qquad Q(A) := \sum_{x \in A} p_X(x)$
  $(S, \mathcal{P}(S), Q)$ is a discrete probability space.
- For $S = \{s_i : i \in I\}$ countable set of reals and $c_i \in [0,1]$ satisfying $\sum_{i \in I} c_i = 1$ there is a probability space and a discrete r.v. $X$ on it such that $p_X(s_i) = c_i$ for $i \in I$.

# Overview

Conditional probability

Discrete random variables

Examples of discrete r.v.'s

# Bernoulli/alternate distribution

- $X =$ number of tails in one toss of a coin (not necessary a fair one)
- We write $X \sim Bern(p)$. (Sometimes $Alt(p)$.)

- Given $p \in [0, 1]$.
- $p_X(1) = p$
- $p_X(0) = 1 - p$
- $p_X(k) = 0$ for $k \neq 0, 1$

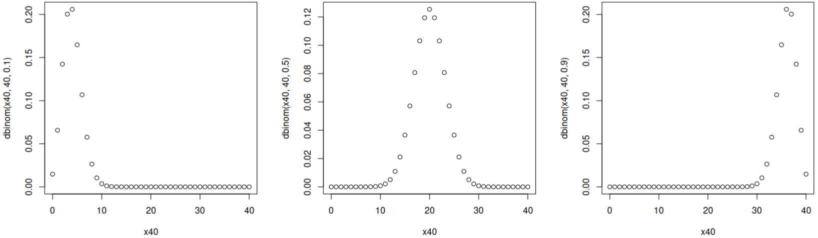
- For an event $A \in \mathcal{F}$ we define *indicator random variable* $I_A$:
- $I_A(\omega) = 1$ if $\omega \in A$, $I_A(\omega) = 0$ otherwise.
- $I_A \sim Bern(P(A))$

# Binomial distribution

- $X =$ number of tails in $n$ independent tosses of a loaded coin.
- We write $X \sim Bin(n, p)$.

- $X = \sum_{i=1}^{n} X_i$ for independent r.v.'s $X_1, \ldots, X_n \sim Bern(p)$.
- Given $p \in [0, 1]$.
- $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$

# Binomial distribution: pmf




**Generated by the following code in R**

```
x40 <- 0:40
plot(x40, dbinom(x40, 40, 0.1))
plot(x40, dbinom(x40, 40, 0.5))
plot(x40, dbinom(x40, 40, 0.9))
```