Probability Theory

LECTURE 1

Course Information

- Lecturer: Oana Lang, e-mail: oana.lang@ubbcluj.ro
- Lectures: Thursday 2pm - 4pm between 27 February - 5 June 2025, Room 6/II, 1 M. Kogalniceanu, Level 2.
  Seminars/Tutorials:

Ziua	Orele	Frecventa	Sala	Anul	Formatia	Tipul	Cadrul didactic
Luni	16-18		A313	1 Inteligenta Artificiala in limba engleza	1011	Seminar	Conf. LISEI Hannelore
Joi	12-14		A312	1 Inteligenta Artificiala in limba engleza	1012	Seminar	Lect. LANG Oana
Joi	14-16		6/II	1 Inteligenta Artificiala in limba engleza	IA1	Curs	Lect. LANG Oana

Ziua	Orele	Frecventa	Sala	Anul	Formatia	Tipul	Cadrul didactic
Luni	8-10		A321	2 Matematica informatica - linia de studiu engleza	821	Seminar	Asist. MICU Tudor
Miercuri	8-10		A321	2 Matematica informatica - linia de studiu engleza	822	Seminar	Asist. MICU Tudor

- Examination:
  - 70% written examination in June
  - 30% Coursework - Deadline: 15 May 2025
  - extra 10% possible based on seminar activity - minimal requirement: 5 'stars' per term where 1 $\star$ is given for a demonstration at the board.
- Pre-requisites: basic concepts in set theory

Outline for today

- Why should we learn Probability Theory?
  - What is *uncertainty*?
- What is a *probability*? (classical definition)
- What is a *probability space?* (axiomatic definition)
  - Outcomes, events, sample space
- Rules of Probability
- Real-world applications
- Time for questions and discussion

Why should we learn Probability Theory?

Suppose you are planning to wait for a friend at the train station. The train is coming from Bucharest and it is **supposed to** arrive on time. All of a sudden, the website is updated and it is announced that the train left Sighișoara one hour behind the schedule.

Given the new information, are you ***sure*** *that the train will arrive on time in Cluj-Napoca?*

Why should we learn Probability Theory?

What is *Uncertainty*?

- Uncertainty surrounds us every day:
  - Before entering the lecture room, you couldn't be certain about the number of colleagues you are going to find in the room.
  - When you toss a coin, roll a die, or buy a lottery ticket.

  - *I wonder if I will risk and lose the bet?*
  - *Is that really **possible**?*
  - *According to meteorological forecasts it **might** rain tomorrow.*

  - Viruses can attack a computer system at *unpredictable/uncertain* times and can affect an *unpredictable/uncertain* number of files/directories.

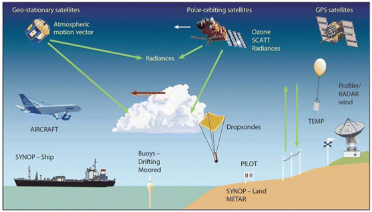
$\Rightarrow$ we are forced to make decision *under uncertainty.*

- *Uncertainty* refers to a condition when a situation/event cannot be predicted *for sure*, with no error. ([2])

We will use Probability Theory to measure and quantify uncertainty. This will allow us to make decisions under uncertainty.

Further Examples

- Medical diagnosis: the chance of having a particular medical condition given a positive test result is calculated using probabilities (tests for diseases).
- Weather forecasting: meteorologists use historical data and current conditions to estimate the probability of rain, storms, or other weather events occurring in a given location. They can predict, say 70% chance of rain, but they don't know the future situation *for sure*.



What is Probability Theory?

Probability Theory is a branch of mathematics that deals with the study of *random* phenomena.

random = happening by chance with no cause or reason; unpredictable, unintended
- *aleatorius* (lat.) = random
- *alea* (lat.) = dice (for a game); dice game

$\hookrightarrow$ One measures the chances of success or the risk of failure of events.

Random experiments, trials, events

The terms *trial* or *experiment* are used in probability theory to describe virtually any process or action whose *outcome* is not known in advance with certainty $\rightarrow$ it has a **random** behaviour.

The random event is the result of an experiment.

- roll two game dice
  $\rightarrow$ both dice show 1
- draw a playing card
  $\rightarrow$ a 3 was drawn
- lottery draw (6 out of 49)
  $\rightarrow$ the first number drawn is 23

Outcomes, events, and probabilities
Sample space

Intuitively: the probability of an event represents the *chance* that the event will happen.

- When you toss a (fair) coin: 50-50 chance of turning up heads or tails $\Rightarrow$ the probability of each side is equal to $\frac{1}{2}$.
- When making predictions (e.g. in forecasting): it is common to speak about probability as a *likelihood* e.g a company's profit is *likely* to rise this year.
- Two software companies are competing for an important contract. We know that company A is *twice as likely* to win the contract as company $B \Rightarrow$ the probability to win the contract for company A is equal to $\frac{2}{3}$, while this probability is just $\frac{1}{3}$ for company B.

Outcomes, events, and probabilities
Sample space

- Probabilities appear when we consider and *weight* possible results of some *experiments*. Some of these results are more *likely* than others.

- A collection of all elementary results, or **outcomes** of an experiment is called **sample space**.
- A set of outcomes is an **event**. $\Rightarrow$ events are subsets of the sample space.

Outcomes, events, and their probabilities
Sample space

- Example 1: A tossed die can produce 6 possible outcomes: $1, 2, 3, \ldots,$ or 6 dots. Each outcome is an event, in this case. We can observe also other events: an even/odd number of dots, a number of dots less than 4, etc.
- Example 2: Consider a football game between U Cluj and CFR Cluj. The sample space consists of 3 outcomes:

$$\Omega = \{\text{U Cluj wins, CFR Cluj wins, they tie}\}.$$

  If we combine these outcomes in all possible ways, we obtain $2^3 = 8$ events: U Cluj wins, loses, ties, gets at least a tie, gets at most a tie, no tie, gets *some* result, gets *no result*.
  - The event "some result" is the entire sample space $\Omega \Rightarrow$ it should have probability 1.
  - The event "no result" is empty (it does not contain any outcome) $\Rightarrow$ it has probability 0.

What is the *probability* of an event? (classical definition)

Definition 1

We consider an experiment which has finitely many equally probable outcomes. The probability that the event A occurs is

$$\mathbb{P}(A) = \frac{\text{the number of favorable outcomes for the occurrence of } A}{\text{number of all possible outcomes within the experiment}}.$$

Definition 2

Let A be a random event appearing in an experiment; the experiment is repeated n times (under the same given conditions) and denote by k_n how many times the event A appears; the *relative frequency of the event A* is the number

$$h_n(A) = \frac{k_n}{n}$$

and k_n is the *absolute frequency of the event A*.

- After repeating an experiment n times (n sufficiently large), under the same conditions, the relative frequency $h_n(A)$ of the event A is approximately equal to the probability $\mathbb{P}(A)$

$$h_n(A) \approx \mathbb{P}(A), \text{ if } n \to \infty.$$

$\Rightarrow$ In the long run, the probability of an event can be viewed as a *proportion* of times this event happens i.e. its relative frequency.

Dice Rolling - History

The correspondence between B. Pascal and P. Fermat, in which they investigated the dice rolling problem of the French nobleman and gambler Chevalier de Méré is famous:
It is said that de Méré had been betting that in four rolls of a die at least one six would turn up. He was consistently winning and to get more people to play, he changed the game bet: in 24 rolls of two dice, a pair of sixes would turn up. But with this second bet, de Méré lost and felt that 25 rolls were necessary to make the game favorable.
We will calculate and compare the probabilities of the following events:
A : we obtain at least one six in 4 rolls of a die;
B : we obtain at least one pair of sixes in 24 rolls of two dice;
C : we obtain at least one pair of sixes in 25 rolls of two dice.

Dice Rolling

For this problem it is easier to determine the probabilities of the contrary events $\bar{A}, \bar{B}$ and $\bar{C}$. The event $\bar{A}$ means no six is obtained in 4 rolls of a die.

$$\Longrightarrow \mathbb{P}(\bar{A}) = \frac{5^4}{6^4} \Longrightarrow \mathbb{P}(A) = 1 - \frac{5^4}{6^4} \approx 0.5177.$$

The event $\bar{B}$ means no pair of sixes is obtained in 24 rolls of two dice

$$\mathbb{P}(\bar{B}) = \frac{35^{24}}{36^{24}} \Longrightarrow \mathbb{P}(B) = 1 - \frac{35^{24}}{36^{24}} \approx 0.4914.$$

The event $\bar{C}$ means no pair of sixes is obtained in 25 rolls of two dice

$$\mathbb{P}(\bar{C}) = \frac{35^{25}}{36^{25}} \Longrightarrow \mathbb{P}(C) = 1 - \frac{35^{25}}{36^{25}} \approx 0.5055.$$

Comparing now the calculated probabilities we notice

$$\mathbb{P}(B) < \frac{1}{2} < \mathbb{P}(C) < \mathbb{P}(A).$$

Set operations in Probability

- Events are *sets* of outcomes $\Rightarrow$ in order to learn how to compute probabilities of events, we use set operations.

- The **complement** of an event A is an event that occurs every time when A does not occur. It consists of outcomes excluded from A. *Notation:* A^c or $\bar{A}$.
- A **union of events** $A, B, C, \ldots$ is an event which consists of all the outcomes in all these events. It occurs if any of $A, B, C, \ldots$ occurs.
- An **intersection** of events $A, B, C, \ldots$ is an event which consists of outcomes which are common in all these events. It occurs if each $A, B, C, \ldots$ occurs.
- A **difference** of events A and B consists of all outcomes included in A but excluded from B. It occurs when A occurs and B does not occur.

Set operations in Probability

- The events A and B are disjoint if their intersection is empty, that is

$$A \cap B = \emptyset.$$

- The events $A_1, A_2, A_3, \ldots$ are **mutually exclusive** or **pairwise disjoint** if any two of these events are disjoint, that is

$$A_i \cap A_j = \emptyset \quad \text{for any} \quad i \neq j.$$

- The events $A_1, A_2, A_3, \ldots$ are **exhaustive** if their union equals the whole sample space, that is

$$A_1 \cup A_2 \cup \ldots = \Omega.$$

What is a Probability Space? (axiomatic definition)

In probability theory, we use set notation to define key concepts:

- Sample Space (Ω): The sample space is the set of all possible outcomes of a random experiment.
- Events (in $\mathcal{F}$): Events are subsets of the sample space Ω. Formally, let $\mathcal{F}$ be a sigma-algebra over Ω. Any subset $A \in \mathcal{F}$ is considered an *event*.
- Probability Measure ($\mathbb{P}$): The probability measure $\mathbb{P}$ assigns a real number between 0 and 1 to each event in $\mathcal{F}$.
- Outcomes (ω): An outcome ω is a specific element in the sample space. For example, if the experiment is rolling a six-sided die, a possible outcome might be a specific number like 3 or 6.

What is a probability space?

Consider a non empty set Ω. A σ-field or σ-algebra $\mathcal{F}$ on Ω is a collection of subsets of Ω with the following properties:

1. The empty set $\emptyset$ belongs to $\mathcal{F}$.
2. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ where $A^c = \Omega \backslash A$.
3. Given a countable collection of sets $A_i, i = 1, 2, \ldots$ from $\mathcal{F}$, their union

$$\bigcup_{i=1}^{\infty} A_i$$

also belongs to $\mathcal{F}$.

- Note that if $A_i, i = 1, 2, \ldots$ is a countable collection of sets from $\mathcal{F}$, their intersection $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c$ also belongs to $\mathcal{F}$.

What is a probability space?

Let $\mathcal{F}$ be a σ-field on a non empty set Ω. A probability measure $\mathbb{P}$ on Ω is a function

$$\mathbb{P} : \mathcal{F} \to [0, 1]$$

such that:

1. $\mathbb{P}(\Omega) = 1$.
2. For any (infinite) sequence of disjoint sets $\{A_i\}_{i \geq 1}$, that is $A_i \cap A_j = \emptyset$, if $i \neq j$, then

$$\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}\left(A_i\right).$$

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

Rules of Probability

- Complement rule:

$$\mathbb{P}(A^c) = \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A).$$

- Probability of a union

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

  - For mutually exclusive events:

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

Rules of Probability

Theorem 3

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let* $A, B \in \mathcal{F}$*. Then the following properties are true:*

(1) $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$ *and* $0 \leq \mathbb{P}(A) \leq 1$*.*

(2) $\mathbb{P}(\emptyset) = 0$*.*

(3) $\mathbb{P}(A \backslash B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$*.*

(4) If $A \subseteq B$*, then* $\mathbb{P}(A) \leq \mathbb{P}(B)$*, i.e.* $\mathbb{P}$ *is monotone.*

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

Rules of Probability

Proof:

(1) Since A and $\bar{A}$ are disjoint and $A \cup \bar{A} = \Omega$, we have $\mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(\bar{A})$. Taking into account that $\mathbb{P}(\Omega) = 1$, it follows that $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$. But $\mathbb{P}(\bar{A}) \geq 0$ implies $\mathbb{P}(A) \leq 1$.

(2) We apply (1) for $A = \Omega$ and use $\mathbb{P}(\Omega) = 1$.

(3) In virtue of the equality $A = (A \cap B) \cup (A \backslash B)$ we have $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \backslash B)$.

(4) We have $A \cap B = A$, since $A \subseteq B$. Then by (1) and (3) it follows that $0 \leq \mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$.

(5) The following property holds $A \cup (\bar{A} \cap B) = A \cup B$, where the union on the left side of the equality is composed of disjoint sets. Thus,

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(\bar{A} \cap B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \end{aligned}$$

where we also used property (3).

Rules of Probability

Theorem 4

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If $(A_n)_{n\geq 1}$ is a sequence of events from $\mathcal{F}$. The inclusion-exclusion principle holds

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}\left(A_i\right) - \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{P}\left(A_i \cap A_j\right) + \cdots + (-1)^{n-1}\mathbb{P}\left(A_1 \cap \cdots \cap A_n\right)$$

for all $n \in \mathbb{N}^*$.

Rules of Probability

Proof:

We use the induction method. For $n = 2$ the property was proved in Theorem 3 - (5). Assuming the property is true for $n \in \mathbb{N}^*$, we prove that it is also true for $n+1$. By Theorem 3 - (5) we write

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbb{P}\left(A_{n+1}\right) - \mathbb{P}\left(\bigcup_{i=1}^{n} \left(A_i \cap A_{n+1}\right)\right)$$

and applying the property (inclusion-exclusion principle) for n sets we have

$$\mathbb{P}\left(\bigcup_{i=1}^{n} \left(A_i \cap A_{n+1}\right)\right)$$
$$= \sum_{i=1}^{n} \mathbb{P}\left(A_i \cap A_{n+1}\right) - \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{P}\left(A_i \cap A_j \cap A_{n+1}\right) + \cdots + (-1)^{n-1} \mathbb{P}\left(A_1 \cap \cdots \cap A_n \cap A_{n+1}\right).$$

Using these relations in (1) it follows that the inclusion-exclusion principle holds also for $n+1$ sets.

Rules of Probability

Definition 5

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

$(A_n)_{n\geq 1}$ is an *increasing sequence* of events from $\mathcal{F}$, if $A_n \subseteq A_{n+1}$ for each $n \in \mathbb{N}^*$.

$(A_n)_{n\geq 1}$ is a *decreasing sequence* of events from $\mathcal{F}$, if $A_{n+1} \subseteq A_n$ for each $n \in \mathbb{N}^*$.

Rules of Probability

Theorem 6

Let $(\Omega, \mathcal{F}, \mathcal{P})$ *be a probability space. Then the following properties are true:*

(1) If $(A_n)_{n\geq 1}$ *is an increasing sequence of events from* $\mathcal{F}$*, then*

$$\lim_{n\to\infty} \mathbb{P}\left(A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right).$$

(2) If $(A_n)_{n\geq 1}$ *is a decreasing sequence of events from* $\mathcal{F}$*, then*

$$\lim_{n\to\infty} \mathbb{P}\left(A_n\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Rules of Probability

Proof:

(1) We define in $\mathcal{F}$ the sequence $(B_n)_{n\geq 1}$ of events by

$$B_1 = A_1, \quad B_n = A_n \backslash A_{n-1} \text{ for } n \geq 2.$$

Since $(A_n)_{n\geq 1}$ is an increasing sequence of events, we have

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n \text{ and } B_i \cap B_j = \emptyset \text{ for } i \neq j \in \mathbb{N}^*.$$

We can write

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \mathbb{P}(A_1) + \sum_{n=1}^{\infty} \mathbb{P}(A_{n+1} \backslash A_n)$$
$$= \lim_{n\to\infty} \left(\mathbb{P}(A_1) + \mathbb{P}(A_2 \backslash A_1) + \ldots \mathbb{P}(A_n \backslash A_{n-1})\right).$$

By Theorem 3 we have

$$\mathbb{P}(A_{n+1} \backslash A_n) = \mathbb{P}(A_{n+1}) - \mathbb{P}(A_n) \qquad \text{for all } n \in \mathbb{N}^*,$$

because $A_n \subseteq A_{n+1}$. Thus,

$$\mathbb{P}(A_1) + \mathbb{P}(A_2 \backslash A_1) + \ldots \mathbb{P}(A_n \backslash A_{n-1}) = \mathbb{P}(A_n).$$

Rules of Probability

Proof:

Then by (3)

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}\left(A_n\right).$$

(2) We apply (1) for $B_n = \bar{A}_n$. So $\left(B_n\right)_{n\geq 1}$ becomes an increasing sequence of events from $\mathcal{F}$ and it holds by the previous result that

$$\lim_{n\to\infty} \mathbb{P}\left(B_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right).$$

Due to De Morgan's laws this equality implies

$$\lim_{n\to\infty} \left(1 - \mathbb{P}\left(A_n\right)\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \bar{A}_n\right) = \mathbb{P}\left(\overline{\bigcap_{n=1}^{\infty} A_n}\right) = 1 - \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Therefore, $\lim_{n\to\infty} \mathbb{P}\left(A_n\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$.

Rules of Probability

Remark:

The operations of union and intersection are commutative:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A;$$

associative:

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cup (B \cap C);$$

and distributive:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C),$$

It holds: $A \cup \bar{A} = S, A \cap \bar{A} = \emptyset$ and the laws of De Morgan:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B}.$$

Real-world Applications

- Networks: *Conditional probabilities* are used in network security for detecting anomalies or potential security threats. It helps assess the probability of a network event to be a security threat based on historical data.
- Machine Learning and AI: Probabilities are used in Bayesian machine learning, where models are updated based on new data. Bayesian networks use probability rules to model relationships between variables and make predictions based on observed data.

Real-world Applications

- Finance and Risk Management: Probability models are extensively used in finance for risk assessment, portfolio optimization, and pricing of financial derivatives. For example, the Black-Scholes model for option pricing involves probability distributions.
- Epidemiology: Probability models are essential in epidemiology to understand the spread of diseases, predict outbreaks, and assess the effectiveness of intervention strategies.

Real-world Applications

In all these applications, the use of probability theory provides a structured and quantitative framework to model uncertainty, make predictions, and guide decision-making processes. This mathematical foundation allows practitioners to assess risks, optimize processes, and make informed choices in the face of uncertainty.

LECTURE 2

Outline

- Basics of Probability - Quick Review
- Conditional Probability
- Independence

Why do we need conditional probabilities?
Quick example

Suppose you are planning to wait for a friend at the train station. The train is coming from Bucharest and it is supposed to arrive on time, with probability 70%. All of a sudden, the website is updated and it is announced that the train left Sighișoara one hour behind the schedule $\Rightarrow$ the probability for the train to arrive on time **becomes** 5%.

$\Rightarrow$ new information influences the probability of meeting your friend on time.

The updated probability is called **conditional probability**, where the new information (the train left Sighișoara late) is a condition.

Conditional Probability

Definition 7

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider two events $A, B \in \mathcal{F}$ such that $\mathbb{P}(B) \neq 0$. The conditional probability of A given B is defined by

$$\boxed{\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}} \tag{1}$$

- This represents *the probability that event A occurs, when event B is known to have occurred*.
- It can be rewritten to obtain a general formula for the probability of intersection of two events:

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

Conditional Probability

- Remark that, given the new information = occurence of the event B, we replace the (standard) formula for the *unconditional probability* of A

$$\mathbb{P}(A) = \frac{\text{number of outcomes in} \quad A}{\text{number of outcomes in} \quad \Omega}$$

by the *conditional probability* of A given B

$$\mathbb{P}(A \mid B) = \frac{\text{number of outcomes in} \quad A \cap B}{\text{number of outcomes in} \quad B}.$$

Conditional Probability

Example 1

Consider a deck of 52 cards. Let

A: the event of drawing a red card
B: the event of drawing a queen.

We want to find the conditional probability for

$\mathbb{P}(A|B)$: the probability of drawing a red card given that a queen is drawn. By definition

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

In our particular case:

$$\mathbb{P}(A \cap B) = \mathbb{P}(\text{drawing a red queen})$$

$$\mathbb{P}(B) = \mathbb{P}(\text{drawing a queen})$$

Note: if event A occurred i.e. the first card was red, then the probability of drawing a second red card becomes slightly smaller. We say that the two events are **dependent**.

Conditional Probability
Example 1

We have

$$\mathbb{P}(A \cap B) = \mathbb{P}(\text{drawing a red queen}) = \frac{2}{52}$$

$$\mathbb{P}(B) = \mathbb{P}(\text{drawing a queen}) = \frac{4}{52}$$

Substituting these values into the formula:

$$\mathbb{P}(A|B) = \frac{\frac{2}{52}}{\frac{4}{52}}$$

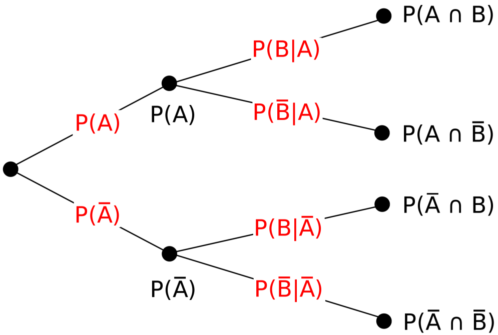
Simplifying:

$$\mathbb{P}(A|B) = \frac{2}{4} = \frac{1}{2}$$

Therefore $\mathbb{P}(A|B) = \frac{1}{2}$.

Conditional Probability

Visual interpretation



Tree diagram - conditional probabilities

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$$
$$\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(\bar{B} \mid A)\mathbb{P}(A)$$
$$\mathbb{P}(\bar{A} \cap B) = \mathbb{P}(B \mid \bar{A})\mathbb{P}(\bar{A})$$
$$\mathbb{P}(\bar{A} \cap \bar{B}) = \mathbb{P}(\bar{B} \mid \bar{A})\mathbb{P}(\bar{A})$$

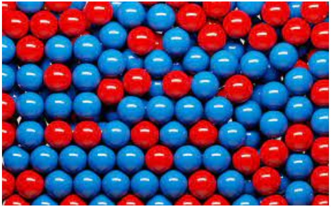
Conditional Probability

Example 2

An urn contains 4 blue marbles and 5 red marbles. Two marbles are successively drawn without replacement.

a) Knowing that the first marble is red, what is the probability that the second marble is blue?

b) What is the probability that both balls are red?



Conditional Probability

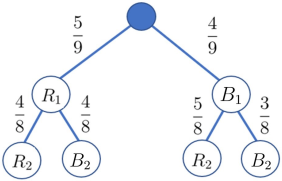
Example 2

For $i \in \{1, 2\}$ consider the events:

R_i : the i^{th} drawn marble is red
$B_i = \bar{R}_i$: the i^{th} drawn marble is blue

a) conditional probability: $\mathbb{P}(B_2 \mid R_1) = \frac{4}{8}$.

b)
$\mathbb{P}(R_1 \cap R_2) = \mathbb{P}(R_2 \mid R_1)\,\mathbb{P}(R_1) = \frac{4}{8} \cdot \frac{5}{9}$.



Independent events

Definition 8

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two events $A, B \in \mathcal{F}$. The events A and B are called independent if

$$\boxed{\mathbb{P}(A \mid B) = \mathbb{P}(A)} \tag{2}$$

that is *the occurrence of B does not affect the probability of A.*

$\Rightarrow$ Substituting this into

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A \mid B) \tag{3}$$

gives

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

$\Rightarrow$ *conditional probability* equals *unconditional probability* when the events are **independent**.

Independent events

Alternative definition

Definition 9

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The events $A, B \in \mathcal{F}$ are said to be independent events if

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)} \tag{4}$$

Remark:

- *Independent events* A and $B \Rightarrow$ the occurrence of the event A not affecting the occurrence of the event B and vice versa.
- *Dependent events* A and $B \Rightarrow$ the occurrence of the event A affecting the occurrence of the event B or vice versa.

Independent events

Examples:
(1) When you toss two coins, the outcome of one does not affect the other, therefore the events are independent.

(2) When you roll a die and toss a coin, the outcome of one does not affect the outcome of the other, therefore the events are independent.

(3) You draw a first card from a deck with 52 cards; denote the event B_1: the first drawn card is black; you draw a second card; denote the event B_2 : the second drawn card is black. By knowing that the first card was black, you made the probability of drawing a second black card slightly smaller. The two events B_1 and B_2 are dependent.

Independent events

Proposition 1

In a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ let $A, B \in \mathcal{F}$. Then the following assertions are equivalent:

(1) A and B are independent.

(2) $\bar{A}$ and B are independent.

(3) A and $\bar{B}$ are independent.

(4) $\bar{A}$ and $\bar{B}$ are independent.

Independent events

Proof.

In order to prove the equivalences we use the properties from Theorem 3 and definition 9.

(1)⇔(2):

$$\begin{aligned}\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B) &\Leftrightarrow \mathbb{P}(B)-\mathbb{P}(A\cap B)=\mathbb{P}(B)-\mathbb{P}(A)\mathbb{P}(B)\\ &\Leftrightarrow \mathbb{P}(B)-\mathbb{P}(A\cap B)=\mathbb{P}(B)\mathbb{P}(\bar{A})\\ &\Leftrightarrow \mathbb{P}(B\setminus A)=\mathbb{P}(B)\mathbb{P}(\bar{A}) \Leftrightarrow \mathbb{P}(\bar{A}\cap B)=\mathbb{P}(\bar{A})\mathbb{P}(B).\end{aligned}$$

(1)⇔(3):

$$\begin{aligned}\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B) &\Leftrightarrow \mathbb{P}(A)-\mathbb{P}(A\cap B)=\mathbb{P}(A)-\mathbb{P}(A)\mathbb{P}(B)\\ &\Leftrightarrow \mathbb{P}(A)-\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(\bar{B})\\ &\Leftrightarrow \mathbb{P}(A\setminus B)=\mathbb{P}(A)\mathbb{P}(\bar{B}) \Leftrightarrow \mathbb{P}(A\cap\bar{B})=\mathbb{P}(A)\mathbb{P}(\bar{B}).\end{aligned}$$

(1)⇔(4):

$$\begin{aligned}&\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B) \Leftrightarrow \mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A)\mathbb{P}(B)\\ &\Leftrightarrow 1-\mathbb{P}(A\cup B)=(1-\mathbb{P}(A))(1-\mathbb{P}(B)) \Longleftrightarrow 1-\mathbb{P}(A\cup B)=\mathbb{P}(\bar{A})\mathbb{P}(\bar{B})\\ &\Leftrightarrow \mathbb{P}(\bar{A}\cap\bar{B})=\mathbb{P}(\bar{A})\mathbb{P}(\bar{B}).\end{aligned}$$

Independent events

Definition 2

- *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Then* $A_1, ..., A_n \in \mathcal{F}$ *are* ***independent events*** *if*

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_m}) = \mathbb{P}(A_{i_1}) \cdot \ldots \cdot \mathbb{P}(A_{i_m})$$

for each subset $\{i_1, \ldots, i_m\} \subset \{1, ..., n\}$ *with* $2 \leq m \leq n$.

- $A_1, ..., A_n \in \mathcal{F}$ *are* ***pairwise independent events*** *if*

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \textit{for all} \quad i, j \in \{1, \ldots, n\}, i \neq j.$$

Remark:
The independence of two (or more) events means that the occurrence of one has no effect on the occurrence or not of the other(s). It can be extended to sets, (collections of) σ-algebras, or even *random variables*.

Independent events

Exercises:
1) What does it mean that three events A, B, C are independent?
2) How many conditions must be verified for n events $B_1, \ldots, B_n$ to be independent?
3) You roll a die twice. Consider the events:
A: the first number is 6; B: the second number is 5; C: the first number is 1;

- Are the following events independent or dependent?

a) A and B;
b) A and C;
c) B and C.

- Are the following events disjoint?

a) A and B;
b) A and C;
c) B and C.

Independent events

Answers:
1) $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$ and $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.
2) $C_n^2 + C_n^3 + ... + C_n^n = 2^n - C_n^1 - C_n^0 = 2^n - n - 1$ conditions.
3) • a) A and B are independent;
b) A and C are dependent;
c) B and C are independent.
• a) A and B are not disjoint (they both can appear at the same time within the experiment);
b) A and C are disjoint (they cannot appear at the same time within the experiment);
c) B and C are not disjoint.

LECTURE 3

Outline

- Independent events - continuation
- Total Probability Rule
- Bayes' Rule
- Applications

Independent events

Example

Let the four faces of a regular tetrahedron be painted as follows: one red, one blue, one green. The fourth face is painted with all three colours.
The tetrahedron is thrown, and the following events are considered:
R: the tetrahedron falls on a surface which contains red colour.
B: the tetrahedron falls on a surface which contains blue colour
G: the tetrahedron falls on a surface which contains green colour.
Are the 3 events R, B, G independent?

Independent events

Answer: R, B, G are not independent, because $\mathbb{P}(R \cap B \cap G) = \frac{1}{4} \neq \mathbb{P}(R)\mathbb{P}(B)\mathbb{P}(G) = \frac{1}{2^3}$. But R, B, G are pairwise independent, since it holds

$$\mathbb{P}(R \cap B) = \mathbb{P}(R)\mathbb{P}(B); \mathbb{P}(B \cap G) = \mathbb{P}(B)\mathbb{P}(G); \mathbb{P}(R \cap G) = \mathbb{P}(R)\mathbb{P}(G).$$

Independent events
Example 3

We have the following scenario:

- 90% of planes depart on time
- 80% of planes arrive on time
- 75% of planes depart on time **and** arrive on time

(a) Alex is meeting Sabrina. Sabrina's plane departed on time. What is the probability that Sabrina will arrive on time?

(b) Alex has met Sabrina, and she arrived on time. What is the probability that her plane departed on time?

(c) Are the events *departing on time* and *arriving on time* independent?

Independent events

Example 3

Let

$$A = \{ \text{ arriving on time } \}$$
$$D = \{ \text{ departing on time } \}.$$

We have:

$$\mathbb{P}(A) = 0.8, \quad \mathbb{P}(D) = 0.9, \quad \mathbb{P}(A \cap D) = 0.75.$$

Then

(a) $\mathbb{P}(A \mid D) = \frac{\mathbb{P}(A \cap D)}{\mathbb{P}(D)} = \frac{0.75}{0.9} = 0.8333.$

(b) $\mathbb{P}(D \mid A) = \frac{\mathbb{P}(A \cap D)}{\mathbb{P}(A)} = \frac{0.75}{0.8} = 0.9375.$

(c) Events are not independent, since

$$\mathbb{P}(A \mid D) \neq \mathbb{P}(A), \quad \mathbb{P}(D \mid A) \neq \mathbb{P}(D), \quad \mathbb{P}(A \cap D) \neq \mathbb{P}(A)\mathbb{P}(D).$$

Note that $\mathbb{P}(A \mid D) > \mathbb{P}(A)$ and $\mathbb{P}(D \mid A) > \mathbb{P}(D) \Rightarrow$ departing on time increases the probability of arriving on time, and vice versa.

In general: the two conditional probabilities $\mathbb{P}(A \mid B)$ and $\mathbb{P}(B \mid A)$ are not the same.

Total Probability Rule

- Consider a partition of the sample space Ω with mutually exclusive and exhaustive events $B_1, B_2, \ldots, B_k$. That is

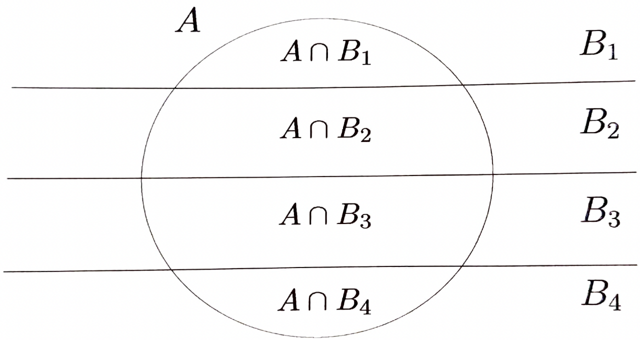
$$B_i \cap B_j = \emptyset \text{ for any } i \neq j \text{ and } B_1 \cup \ldots \cup B_k = \Omega.$$

- Assume that these events also partition the event A, that is

$$A = (A \cap B_1) \cup \ldots \cup (A \cap B_k)$$

and this is also a union of mutually exclusive events.

Total Probability Rule



Partition of the sample space Ω and the event A[1]

[1] M. Baron, *Probability and Statistics for Computer Scientists*

Total Probability Rule

Then

$$\mathbb{P}(A) = \sum_{j=1}^{k} \mathbb{P}\left(A \cap B_j\right)$$

$\Rightarrow$ The Total Probability Rule or The Law of Total Probability:

$$\boxed{\mathbb{P}(A) = \sum_{j=1}^{k} \mathbb{P}\left(A \mid B_j\right) \mathbb{P}(B_j)} \tag{5}$$

- Relates the unconditional probability of an event A with its conditional probabilities.
- It is used when it is easier to compute conditional probabilities of A given additional information.

Total Probability Rule

- Overall, probability $\mathbb{P}(A)$ is just a "weighted average" of the conditional probabilities $\mathbb{P}(A \mid B_j)$, where the "weights" are $\mathbb{P}(B_j)$.
- These "weights" add up to 1, since $B_1, \ldots, B_k$ are a partition of all the possible outcomes, whose total probability is 1.
- This means that the overall probability $\mathbb{P}(A)$ will always lie somewhere between the conditional probabilities $\mathbb{P}(A \mid B_j)$, with more weight given to the more probable scenarios.

Bayes' Rule

Consider the following scenario:

- There exists a test for a certain infection (say Covid19) which is:
  - 99% reliable for healthy patients
  - 95% reliable for infected patients
- Let us define the following events:
  V: the patient has the virus
  S: the test is positive

$\Rightarrow$ if the patient has the virus, the test will show this with probability $\mathbb{P}(S \mid V) = 0.95$.
$\Rightarrow$ if the patient does not have the virus, the test will show this with probability $\mathbb{P}(S^c \mid V^c) = 0.99$.

- Consider a patient whose test result is positive. Knowing that sometimes the test is wrong, the patient wants to now if she/he has indeed the virus, that is we need to find out $\mathbb{P}(V \mid S)$.
- Since $S \cap V = V \cap S$ we have $\mathbb{P}(V)\mathbb{P}(S \mid V) = \mathbb{P}(S)\mathbb{P}(V \mid S)$ and then

$$\mathbb{P}(V \mid S) = \frac{\mathbb{P}(S \mid V)\mathbb{P}(V)}{\mathbb{P}(S)}.$$

Bayes' Rule

- Describes the probability of an event B, based on prior knowledge of conditions that may be related to the event:

$$\boxed{\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}} \qquad (6)$$

- Bayes' Rule is to probability "what Pythagora's theorem is to geometry".
- It allows to update the probability of an event using new information $\rightarrow$ a method for adjusting or refining current predictions given new or additional evidence.
- Widely used today in machine learning.

Thomas Bayes (1701-1761)

Bayes' Rule

$$\boxed{\mathbb{P}\left(B \mid A\right) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}} \tag{7}$$

- Heuristics:

$$\text{Posterior} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Evidence}}$$

- Probability $\Rightarrow$ measures a "degree of belief". Bayes' theorem links the degree of belief in a proposition/system before and after accounting for *evidence* (information, direct observation).
- $\mathbb{P}(B) =$ *the prior* is the initial degree of belief in B before any updates (before any new information/observation).
- $\mathbb{P}\left(B \mid A\right) =$ *the posterior* is the degree of belief after incorporating information regarding the fact that A is true.
- $\frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$ incorporates the *evidence* A provides for B.

Bayes' and Total Probability Rule

- For $k = 2$ the total probability rule becomes

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$$

  which together with Bayes' rule gives

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)}.$$

- In general

$$\mathbb{P}(A) = \sum_{j=1}^{k} \mathbb{P}\left(A \mid B_j\right) \mathbb{P}(B_j)$$

  and then

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}{\sum_{j=1}^{k} \mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}.$$

Applications

In Filtering

$\Rightarrow$ all these are probabilistic tools which can be used to estimate the state of a dynamical system

- Conditional probability is closely linked to *conditional expectation*. These can be defined for *random variables* = some "functions" defined on our sample space.
- Random variables are sometimes solutions corresponding to sets of equations = a *mathematical model*, which can describe the current/past/future state of a dynamical system $\Rightarrow$ making inferences about such a solution (about its conditional expectation/probability) $\leftrightarrow$ making inferences about the state of the dynamical system itself.
- Explicit calculations can be performed only for simple cases. For more complicated (= realistic) models we can only *approximate* the conditional probability using specific *approximation methods*.

Applications

In Filtering

$$\boxed{\mathbb{P}\left(B \mid A\right) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}}$$

$$\text{Posterior} = \mathbb{P}\left(B \mid A\right) = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Evidence}}$$

$$\pi_{t-1}^{z_{0:t-1}} \xrightarrow[\substack{\text{model}\\ \text{forecast}\\ \text{prediction}}]{\mathcal{M}_t} \mathcal{M}_t\left(\pi_{t-1}^{z_{0:t-1}}\right) =: p_t \xrightarrow[\substack{\text{assimilation}\\ \text{analysis}\\ \text{update}}]{g_t^{z_t}\star} g_t^{z_t}\star p_t = \pi_t^{z_{0:t}}.$$

π_t is the *posterior distribution* which gives information about the state of the dynamical system at time t, taken into account the information provided by observations made up to time t (modelled via $z_{0:t}$).

LECTURE 4

Random Variables and Random Vectors

Are these useful?

Assume a spaceship is launched.

- The costs involved are estimated in millions of dollars/pounds/euros $\Rightarrow$ trying several times would be extremely expensive $\Rightarrow$ spaceship's performance is first *simulated* $\rightarrow$ this allows experts to *evaluate* the associated risks (reliability, safety, etc) in advance.

- *Computer simulations* are used to estimate quantities/costs/outputs for which the direct computation is too risky/difficult/expensive or even impossible.
- *Monte Carlo Methods* are used for computing probabilistic characteristics associated with such simulations: instead of generating the *real* process (spaceship launch, weather forecast etc) we only simulate some associated *random variables*. This suffices in order to infer the probabilistic behaviour of the system.

Random Variables and Random Vectors

- The implementation of Monte Carlo methods reduces to generating random variables from given *distributions*.

  $\Rightarrow$ random variables and vectors provide a rigorous framework for calculating probabilities, defining distributions, and formulating statistical hypotheses.

- In statistical applications, these concepts become the foundation for parameter estimation, hypothesis testing, and building predictive models.

- Random variables and random vectors: *discrete* or *continuous*.

Random Variables

Definition 3

*A **random variable** is a measurable function*

$$X : \Omega \to \mathbb{R}. \tag{8}$$

- *Measurable:*

$$X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$
$$\forall A \in \mathcal{B}(\mathbb{R}), \quad X^{-1}(A) \in \mathcal{F}$$

  where

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}.$$

- Its range can be the set of all real numbers, $(0, \infty)$, the integers $\mathbb{Z}$, an interval (a, b) etc.
- Once the experiment is completed and the outcome $\omega \in \Omega$ is known, the value $X(\omega)$ of the random variable becomes *determined*.
- A random variable is a quantity that depends on *chance*.

- The **Borel σ-algebra** $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing all open sets in $\mathbb{R}$. Explicitly, the Borel σ-algebra satisfies:

$$\mathcal{B}(\mathbb{R}) = \bigcap \{\mathcal{A} \mid \mathcal{A} \text{ is a } \sigma\text{-algebra containing all open sets in } \mathbb{R}\}$$

  meaning it is the intersection of all σ-algebras that contain the open sets.

- The **Borel σ-algebra** $\mathcal{B}(\mathbb{R})$ is generated by the collection of all open intervals of $\mathbb{R}$. Specifically, the collection of sets:

$$(-\infty, x] = \{y \in \mathbb{R} : y \leq x\}, \quad x \in \mathbb{R}$$

  is sufficient to generate $\mathcal{B}(\mathbb{R})$ through countable set operations (unions, intersections, complements).
- Any other set (e.g., an open or closed set) in $\mathbb{R}$ can be constructed using countable unions or intersections of sets of the form $(-\infty, x]$.

Even though open intervals (a, b) are part of $\mathcal{B}(\mathbb{R})$, the collection of sets

$$(-\infty, x] = \{y \in \mathbb{R} \mid y \leq x\}, \quad x \in \mathbb{R}$$

is also sufficient to generate $\mathcal{B}(\mathbb{R})$. Open sets can be approximated using $(-\infty, x]$. Specifically,

$$(a, b) = \bigcup_{n=1}^{\infty} \left((-\infty, b - \tfrac{1}{n}] \cap (-\infty, a + \tfrac{1}{n}]^c\right).$$

Since a σ-algebra is closed under countable unions, this shows that open sets can be built from $(-\infty, x]$. A closed interval $[a, b]$ can be expressed as an intersection of countably many sets of the form $(-\infty, x]$:

$$[a, b] = \bigcap_{n=1}^{\infty} \left((-\infty, b + \tfrac{1}{n}] \cap (-\infty, a - \tfrac{1}{n}]^c\right).$$

Since σ-algebras are closed under countable intersections, this means that closed intervals also belong to the Borel σ-algebra.

Random Variables

- Remarks:
  1. We can also say that $X : \Omega \to \mathbb{R}$ *is a random variable if*
  $$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \text{for each} \quad x \in \mathbb{R}. \tag{9}$$
  2. Notation:
  $$X^{-1}(A) = \{X \in A\} = \{\omega \in \Omega : X(\omega) \in A\} \quad \text{for each} \quad A \in \mathcal{B}(\mathbb{R}).$$
  3. *Indicator function* of the event $A \in \mathcal{F}$
  $$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \in \bar{A}. \end{cases}$$
  This is a discrete random variable (check the definition above).

Random Variables

Examples

E1 Throw two dice, the sum of the obtained numbers is a random variable
$X : \Omega \to \{2, 3, ..., 12\}$, where Ω contains all elementary events, i.e. the sample space is

$$\Omega = \big\{(\omega_i^1, \omega_j^2) : i, j \in \{1, 2, ..., 6\}\big\},$$

where (ω_i^1, ω_j^2) is the elementary event. The first die indicates the number i and the second die indicates the number j, for $i, j \in \{1, 2, ..., 6\} \implies X(\omega_i^1, \omega_j^2) = i + j$ for each $i, j \in \{1, 2, ..., 6\}$.
For example $\mathbb{P}(X = 5) = \frac{4}{36}$ (since $5 = 1 + 4 = 4 + 1 = 2 + 3 = 3 + 2$); $\mathbb{P}(X = 6) = \frac{5}{36}$, etc.

E2 A player throws two coins $\Rightarrow \Omega = \{(H, T), (H, H), (T, H), (T, T)\}$

The random variable X indicates how many times tail T has appeared: $\Rightarrow X : \Omega \to \{0, 1, 2\}$

$\Rightarrow \mathbb{P}(X = 0) = \mathbb{P}(X = 2) = \frac{1}{4}, \mathbb{P}(X = 1) = \frac{1}{2}$.

Random Variables

Examples

E3 ([2]) Consider an experiment in which we toss 3 fair coins and we count the number of heads. The same *model* suits the numbers of girls/boys in a family with 3 children, the number of 1's or 0's in a random binary string of 3 characters, and so on.

- Let X be the numbers of *heads* (girls, 1's, etc). Prior to the experiment, we don't know the value of X, all we know about it is that X is an integer between 0 and 3. We assume that each value is an event, and therefore we can compute the corresponding probabilities:

$$\mathbb{P}(X=0) = \mathbb{P}(3\text{ tails}) = \mathbb{P}(TTT) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}$$
$$\mathbb{P}(X=1) = \mathbb{P}(HTT) + \mathbb{P}(THT) + \mathbb{P}(TTH) = \frac{3}{8}$$

Random Variables

Examples

$$\mathbb{P}(X=2)=\mathbb{P}(HHT)+\mathbb{P}(HTH)+\mathbb{P}(THH)=\frac{3}{8}$$

$$\mathbb{P}(X=3)=\mathbb{P}(HHH)=\frac{1}{8}$$

Overall

x	$\boldsymbol{P}\{X=x\}$
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

The table contains everything we can know about the *random variable* X **before** the experiment. Before finding out the outcome $\omega \in \Omega$ we cannot say what is the actual value of X - all we can do is to list all *possible* values of X together with their corresponding probabilities.

Discrete and Continuous Random Variables

- ***Discrete***: X is said to be a discrete random variable if its range $X(\Omega)$ is a countable set i.e. $X(\Omega) = \{x_i : i \in I\}$ where $I \subseteq \mathbb{N}$ (set of indices) and $\mathbb{P}(X = x_i) > 0$ for each $i \in I$.
  - it has finitely many values $(x_1, \ldots, x_n)$ or countably infinitely many values $(x_1, \ldots, x_n, \ldots)$; the values can be listed
  - *numerical* random variables: the number of cars in a parking lot, the number of sixes in 100 dice rolls, the number of defective parts during a production, the number of items sold at a store on a certain day, the number of customers that enter a certain shop on a given day, the number of voters who showed up to the polls.
  - *categorical* random variables: weather forecast e.g. rainy, cloudy, foggy, clear $\rightarrow$ classification into categories.

Discrete and Continuous Random Variables

- ***Continuous***: the set of its possible values is uncountable. Its possible values comprise either a single interval from $\mathbb{R}$, or a union of disjoint intervals of $\mathbb{R}$.
  - The values cannot be listed.
  - E.g. the running time of a machine until first defection, software installation time, code execution time, the temperature in a certain city within a year, amount of rainfall in a certain city over a year, the speed of a car passing a speed camera, the time it takes to complete an exam for a 60 minute test.

Random Variables

Theoretical examples

- Let X and Y be random variables. Then $aX + b$ (where $a, b \in \mathbb{R}$), $|X|$, $\min\{X, Y\}$, $\max\{X, Y\}$, $X + Y$, $X - Y$ and $X \cdot Y$ are random variables. Moreover, if $Y(\omega) \neq 0$ for all $\omega \in \Omega$, then also $\dfrac{X}{Y}$ is a random variable.

- Let $(X_n)_{n\geq 1}$ be a sequence of random variables such that $\sup\limits_{n\geq 1} X_n$, $\inf\limits_{n\geq 1} X_n \in \mathbb{R}$ for each $\omega \in \Omega$. Then $\sup\limits_{n\geq 1} X_n$, $\inf\limits_{n\geq 1} X_n$, $\limsup\limits_{n\to\infty} X_n := \inf\limits_{n\geq 1}(\sup\limits_{k\geq n} X_k)$, $\liminf\limits_{n\to\infty} X_n := \sup\limits_{n\geq 1}(\inf\limits_{k\geq n} X_k)$are random variables.

Random Variables and Their Distributions

Definition 4

The collection of all probabilities related to X is the distribution of X.

Definition 5

The function $P : \mathbb{R} \to [0, 1]$ with

$$P(x) = \mathbb{P}(X = x)$$

is the probability mass function (pmf).

Definition 6

The probability function

$$\begin{aligned} &P_X : \mathcal{B}(\mathbb{R}) \to [0, 1] \\ &P_X(B) = \mathbb{P}(X \in B) \text{ for each } B \in \mathcal{B}(\mathbb{R}) \end{aligned}$$

is called the probability distribution function or the law of X.

Here $\mathbb{P}(X \in B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$.

Random Variables and Their Distributions

Proposition 7

Let X be a random variable. Then the mapping $P_X : \mathcal{B}(\mathbb{R}) \to [0,1]$ defined by

$$P_X(B) = \mathbb{P}(X \in B) \text{ for each } B \in \mathcal{B}(\mathbb{R}) \tag{10}$$

is a probability over $\mathcal{B}(\mathbb{R})$ i.e. $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ is a probability space.

Random Variables and Their Distributions

Proof.

(i) $P_X(\mathbb{R}) = \mathbb{P}(\omega \in \Omega : X(\omega) \in \mathbb{R}) = \mathbb{P}(\Omega) = 1.$
(ii) $P_X(B) \geq 0$ for all $B \in \mathcal{B}(\mathbb{R})$.
(iii) If $(B_n)_{n \geq 1}$ are pairwise disjoint events, then

$$
\begin{aligned}
P_X\left(\bigcup_{n=1}^{\infty} B_n\right) &= \mathbb{P}\left(\left\{\omega \in \Omega : X(\omega) \in \bigcup_{n=1}^{\infty} B_n\right\}\right) \\
&= \mathbb{P}\left(X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} X^{-1}(B_n)\right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left(X^{-1}(B_n)\right) = \sum_{n=1}^{\infty} P_X(B_n).
\end{aligned}
$$

Note that $\left(X^{-1}(B_n)\right)_{n \geq 1}$ are pairwise disjoint events, since $(B_n)_{n \geq 1}$ are pairwise disjoint events and X is measurable. □

Random Variables - Probability mass function (pmf)

If X is a discrete random variable with range $X(\Omega) = \{x_i : i \in I\}$, then P_X is completely determined by the values

$$\mathbb{P}\left(X = x_i\right) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) = x_i\}\right) \quad i \in I$$

i.e. by the probability mass function of the discrete random variable X.
We say that X has a *discrete distribution*. We write

$$X \sim \begin{pmatrix} x_1 & x_2 & \ldots & x_i & \ldots \\ p_1 & p_2 & \ldots & p_i & \ldots \end{pmatrix} = \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$$

where

$$I \subseteq \mathbb{N}, p_i := \mathbb{P}\left(X = x_i\right) > 0, i \in I, \sum_{i \in I} p_i = 1.$$

Random Variables - Cumulative Distribution Function (cdf)

Definition 8

The cumulative distribution function (cdf) of a random variable X is defined as $F : \mathbb{R} \to \mathbb{R}$

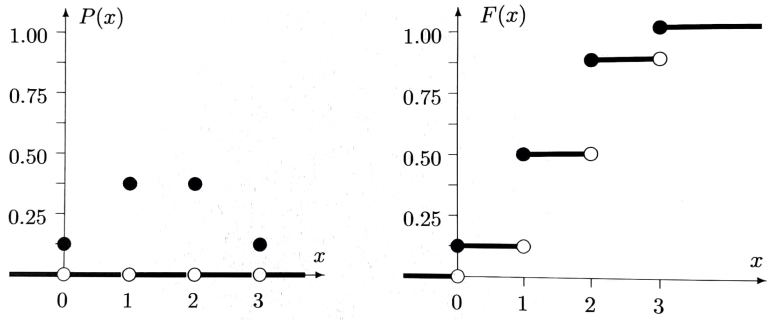
$$F(x) = \mathbb{P}(X \leq x).$$

- The set of possible values of X is called the support of the distribution F.
- In the discrete case:

$$F(x) = \mathbb{P}(X \leq x) = \sum_{y \leq x} P(y), \quad x \in \mathbb{R}$$

  and $F(x)$ has jumps of magnitude $P(x)$.
- In the continuous case $P(x) = 0$ so there are no jumps.

Random Variables - Discrete Case - Example



The probability mass function (pmf) $P(x)$ and the cumulative distribution function (cdf) $F(x)$ for Example E3. White circles denote points which are excluded. Picture from [2].

Random Variables - Discrete Case - Examples

Examples:

(1) If X is a discrete random variable and takes the values $\{x_i : i \in I\}$, then its distribution function is $F : \mathbb{R} \to \mathbb{R}$

$$F(x) = \mathbb{P}(X \le x) = \sum_{\substack{i \in I \\ x_i \le x}} \mathbb{P}\left(X \le x_i\right), x \in \mathbb{R}.$$

(2) Consider the discrete random variable X such that

$X \sim \begin{pmatrix} -1 & 1 & 2 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$. Then the cumulative distribution function of X is $F : \mathbb{R} \to \mathbb{R}$ given by

$$F(x) = \begin{cases} 0, & \text{if } x < -1 \\ 0.5, & \text{if } -1 \le x < 1 \\ 0.5 + 0.25 = 0.75, & \text{if } 1 \le x < 2 \\ 0.5 + 0.25 + 0.25 = 1, & \text{if } 2 \le x \end{cases}$$

Random Variables - Discrete Case

For every outcome ω, the discrete variable X takes one and only one value x. This makes the events $\{X = x\}$ disjoint and exhaustive, and therefore,

$$\sum_x P(x) = \sum_x \mathbb{P}(X = x) = 1.$$

Note that the cdf $F(x)$ is a non-decreasing function of x, always between 0 and 1, with

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to +\infty} F(x) = 1.$$

$F(x)$ is constant between any two subsequent values of X. It jumps by $P(x)$ at each possible value x of X.

Random Variables - Discrete Case

Remember that one way to compute the probability of an event is to add probabilities of all the outcomes corresponding to that particular event. Hence, for any set A,

$$\mathbb{P}(X \in A) = \sum_{x \in A} P(x)$$

If A is an interval, its probability can be computed directly from the cdf $F(x)$ that is

$$\mathbb{P}(a < X \leq b) = F(b) - F(a).$$

pmf vs. cdf

- Any pmf $P(x)$ can assign positive probabilities to a finite or countable set only. We need to have $\sum_x P(x) = 1 \Rightarrow$ we can have at most 2 values of x with $P(x) \geq 1/2$, at most 4 with $P(x) \geq 1/4$ etc. Therefore we can list all x for which $P(x) > 0 \rightarrow$ the set of x_i with $P(x_i) > 0$ cannot be an interval (uncountable), it has to be at most countable.
- For all continuous random variables, the probability mass function is always equal to zero i.e. $P(x) = 0$ for all $x \in \mathbb{R}$. We can instead use the *cumulative distribution function* (cdf).

Random Variables - Properties of the cdf

Theorem 9

The cumulative distribution function $F : \mathbb{R} \to \mathbb{R}$ *of a random variable* X *has the following properties:*
(1) $\mathbb{P}(a < X \leq b) = F(b) - F(a)$ *for* $a < b$*;*
(2) $\mathbb{P}(X = b) = F(b) - F(b - 0)$ *for* $b \in \mathbb{R}$*;*
(3) F *is monotonically increasing;*
(4) F *is right-continuous, i.e.* $F(b + 0) = \lim_{x \searrow b} F(x) = F(b)$ *for each* $b \in \mathbb{R}$*;*
(5) $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$*.*

Remark: The properties (3), (4) and (5) from Theorem 9 characterize a cumulative distribution function, i.e. if a function $F : \mathbb{R} \to [0, 1]$ has these properties, then there exists a probability space and a random variable X which has F as its cumulative distribution function.

References

1. H. Lisei, *Probability Theory - Lecture Notes* (2022-2023)
2. M. Baron, *Probability and Statistics for Computer Scientists* (2019)
3. H. Lisei, *Probability Theory* (2004)
4. Pictures from *google.com*
5. S. Chongchitnan, *Exploring University Mathematics with Python*. Python codes: here.
6. Bright side of Maths: here.