Probability Theory and Statistics
Lecture 2

Bence Csonka

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

September 9

Poincaré reminder

Statement (Poincaré formula for 2 events)

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

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

Poincaré reminder

Statement (Poincaré formula for 2 events)

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

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

Statement (Poincaré formula for 3 events)

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

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

Example for 2 events

In a class of 30 students, 12 attend football practice, 10 attend piano lessons, and 5 students attend both. What is the probability that a randomly chosen student attends piano lessons or football practice?

Example for 3 events

In a city, a survey was conducted about the means of transportation used by the inhabitants.

- 40% travel by bus,
- 30% by tram,
- 20% by metro.

The overlaps are:

- 15% use both bus and tram,
- 10% use both bus and metro,
- 8% use both tram and metro,
- 5% use all three.

What is the probability that a randomly chosen person uses at least one of these means of transportation?

Independence of Two Events

Definition

Events A and B are called *independent* if

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

- First example: we roll a fair die; let the outcome be x. Are the events $A = \{x \text{ is prime}\}$ and $B = \{x \text{ is even}\}$ independent?
- In the definition, the roles of A and B are interchangeable.
- Thus, independence of A and B means that the occurrence of A is not affected by B (and vice versa).
- Caution: independence *does not* mean that $A \cap B = \emptyset$ (quite the opposite in many examples).

Independence of Two Events

Definition

Events A and B are called *independent* if

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

- First example: we roll a fair die; let the outcome be x. Are the events $A = \{x \text{ is prime}\}$ and $B = \{x \text{ is even}\}$ independent?
- In the definition, the roles of A and B are interchangeable.
- Thus, independence of A and B means that the occurrence of A is not affected by B (and vice versa).
- Caution: independence *does not* mean that $A \cap B = \emptyset$ (quite the opposite in many examples).

Statement

If A and B are independent, then A and $\overline{B}$ are also independent.

Independence of Multiple Events

Recall: $[n] = \{1, 2, \ldots, n\}$.

Definition

Events $A_1, \ldots, A_n$ are (jointly) independent if for every $I \subseteq [n]$ we have

$$\mathbb{P}\Big(\bigcap_{i\in I} A_i\Big) = \prod_{i\in I} \mathbb{P}(A_i). \tag{1}$$

In words: the probability of the intersection of *any* subcollection equals the product of their probabilities.

Independence of Multiple Events

Recall: $[n] = \{1, 2, \ldots, n\}$.

Definition

Events $A_1, \ldots, A_n$ are (jointly) independent if for every $I \subseteq [n]$ we have

$$\mathbb{P}\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} \mathbb{P}(A_i). \tag{1}$$

In words: the probability of the intersection of *any* subcollection equals the product of their probabilities.

This stronger definition is **necessary**, because:

(1) does *not* follow from pairwise independence. If every distinct pair (A_i, A_j) is independent, we say $A_1, \ldots, A_n$ are pairwise independent, which is strictly weaker than joint independence.
(Joint independence $\Rightarrow$ pairwise independence, but not conversely.)

Conditional Probability

Example: roll a fair die, $\Omega = \{1, \ldots, 6\}$. We know:

$$\mathbb{P}(1) = \mathbb{P}(2) = \cdots = \mathbb{P}(6) = \frac{1}{6}.$$

Assume we roll without seeing the outcome, and someone informs us that the result is even.

How do probabilities change in light of this additional information?

Conditional Probability

Example: roll a fair die, $\Omega = \{1, \ldots, 6\}$. We know:

$$\mathbb{P}(1) = \mathbb{P}(2) = \cdots = \mathbb{P}(6) = \frac{1}{6}.$$

Assume we roll without seeing the outcome, and someone informs us that the result is even.

How do probabilities change in light of this additional information?

Introduce new probabilities for this situation, denoted by $\tilde{\mathbb{P}}$ (for now—soon we will adopt better notation). By intuition and by the axioms:

$$\tilde{\mathbb{P}}(2) = \tilde{\mathbb{P}}(4) = \tilde{\mathbb{P}}(6) = \frac{1}{3}, \qquad \tilde{\mathbb{P}}(1) = \tilde{\mathbb{P}}(3) = \tilde{\mathbb{P}}(5) = 0.$$

Conditional Probability

Definition

Let $A, B \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. The *conditional probability* of B given A is

$$\mathbb{P}(B \mid A) \stackrel{\text{def}}{=} \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

Read: "the probability of B given A".

Conditional Probability

Definition

Let $A, B \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. The *conditional probability* of B given A is

$$\mathbb{P}(B \mid A) \stackrel{\text{def}}{=} \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

Read: "the probability of B given A".

The formula is not symmetric: the roles of A and B differ. We require $\mathbb{P}(A) > 0$. We do *not* require $\mathbb{P}(B) > 0$. (If $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) = 0$, then indeed $\mathbb{P}(B \mid A) = 0$.)

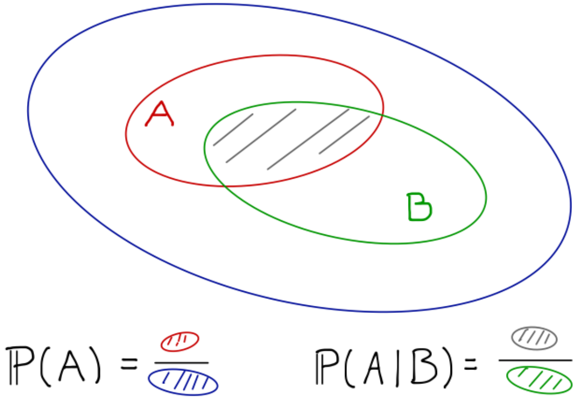
Independence and conditional probability

Statement

For events A, B with $\mathbb{P}(A) > 0$, A and B are independent iff $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.
(Intuition: knowing A does not change the probability of B.)

(Proof)

Conditional Probability



Conditional Probability

Statement

For every $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, the mapping

$$\mathbb{P}(\cdot \mid A) \colon \mathcal{F} \to [0,1], \qquad B \mapsto \mathbb{P}(B \mid A)$$

is a probability measure.

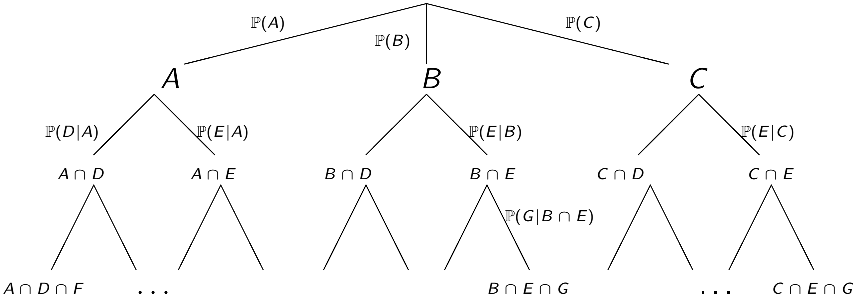
- (Proof)
- Why is this useful? Because any statement proved for $\mathbb{P}(\cdot)$ also holds with $\mathbb{P}(\cdot \mid A)$.

Multi-step Experiments

Random experiments executed in several stages can be represented by a tree:

- Nodes: the possible results at a given stage,
- Edges: the conditional probabilities of moving to the next result, given the current stage,
- Leaves: the final outcomes determined by the intermediate results.

Multi-step Experiments: Tree Representation



Multi-step Experiments: Multiplication Rule

Rearranging the definition of conditional probability yields

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

for $A, B \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. Hence:

Multiplication rule: In a multi-step experiment, the probability of a leaf (outcome) equals the product of the (conditional) probabilities along the path leading to that leaf.

Formally:

Statement (Multiplication rule)

) Let $A_1, \ldots, A_n \in \mathcal{F}$ with $\mathbb{P}(A_i) > 0$ for all i. Then

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}\Big(A_i \,\Big|\, \bigcap_{k=1}^{i-1} A_k\Big).$$

Law of Total Probability

Definition

A sequence $A_1, \ldots, A_n \in \mathcal{F}$ is a complete system of events (a partition) if

- the events are pairwise disjoint (i.e., for all distinct $i, j \in [n]$, $A_i \cap A_j = \emptyset$),
- and $\bigcup_{i=1}^n A_i = \Omega$.

Theorem (Law of Total Probability, LTP)

If $A_1, \ldots, A_n \in \mathcal{F}$ form a complete system and $\mathbb{P}(A_i) > 0$ for all i, then for any event B

$$\mathbb{P}(B) \ = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i).$$

Law of Total Probability

Theorem (Law of Total Probability, LTP)

If $A_1, \ldots, A_n \in \mathcal{F}$ form a complete system and $\mathbb{P}(A_i) > 0$ for all i, then for any event B

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i).$$

Remarks:

- The LTP remains valid if we weaken $A_i \cap A_j = \emptyset$ (for $i \neq j$) to $\mathbb{P}(A_i \cap A_j) = 0$, and $\bigcup_{i=1}^{n} A_i = \Omega$ to $\mathbb{P}(\bigcup_{i=1}^{n} A_i) = 1$. Later we will see these conditions are not fully equivalent to the strict version.
- The LTP also holds for countably many events $A_1, A_2, \ldots$ (require $\mathbb{P}(A_i) > 0$ for all i).

Multi-step Experiments: Sum Rule

An equivalent phrasing of the Law of Total Probability:

Sum rule: In a multi-step experiment, the probability of an event equals the sum (over all leaves consistent with the event) of the leaf probabilities computed via the multiplication rule.

Law of Total Probability

Theorem (Law of Total Probability, LTP)

If $A_1, \ldots, A_n \in \mathcal{F}$ form a complete system and $\mathbb{P}(A_i) > 0$ for all i, then for any event B

$$\mathbb{P}(B) \;=\; \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\, \mathbb{P}(A_i).$$

(Proof)

Bayes' Theorem

Suppose it is easy to compute $\mathbb{P}(B \mid A)$, but what we actually need is $\mathbb{P}(A \mid B)$.

Theorem (Simple Bayes' theorem)

Let $A, B \in \mathcal{F}$ with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. Then

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\overline{A})\mathbb{P}(\overline{A})}.$$

(Proof)

Bayes theorem

Combining with the LTP yields the general form.

Theorem (Bayes' theorem)

Let $B, A_1, \ldots, A_n \in \mathcal{F}$ *with* $\mathbb{P}(B) > 0$ *and* $\mathbb{P}(A_i) > 0$ *for all* i*, and suppose* $A_1, \ldots, A_n$ *form a complete system. Then*

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(B|A_1)\mathbb{P}(A_1)}{\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

(Proof)

Example for Bayes theorem

Pollsters measured the following:

- If a person is male, the probability that he has long hair is 0.05.
- If a person is female, the probability that she has long hair is 0.9.

Question: What is the probability that, when you see a person with long hair from behind, that person is a man?