Tutoriat PS 1

Mihai Duzi, Cristian Lucan, Lumînăraru Ionuț

Contents

1 Introduction to probabilities and statistics

Terminology

Experiment: a repeatable procedure with well-defined possible outcomes.

Sample space: the set of all possible outcomes. Usually denoted by Ω.

Event: a subset of the sample space, usually denoted by A, B, C, etc.

Probability function: a function that gives the probability of each outcome, usually written as:

$$P(A) = \frac{\text{Number of favorable cases}}{\text{Number of possible cases}}.$$

This function satisfies:

$$0 \leq P(A) \leq 1 \quad \text{and} \quad P(\Omega) = 1.$$

In other words, 0 represents 0%, 1 represents 100%, and the total probability of all possible outcomes is 1.

Example: Tossing a fair coin three times

Experiment: We toss the coin three times and list the outcomes.

Sample space:

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Event: $A =$ getting heads exactly two times.

Probability function: each outcome is equally likely, with probability 1/8.

$$P(A) = \frac{3}{8}.$$

2 Use of sets in probabilities

Since we refer to "sets of outcomes" and "subsets of the sample space," probability theory heavily relies on **set theory**.

Let A and B be two events, and Ω be the sample space.

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

The probability of A or B occurring equals the sum of the individual probabilities minus the overlap (when both occur).

$$P(\overline{A}) = 1 - P(A)$$

The probability of A not occurring is 1 minus the probability of A.

$$P(A \cap B) = P(A) \times P(B)$$

This only holds if A and B are **independent events** (we'll define this soon).

De Morgan's Laws:

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

Cartesian product:

$$A \times B = \{(a, b) \mid a \in A,\, b \in B\}.$$

3 Independence, dependence and Bayes probabilities

3.1 Difference between dependent and independent events

Let $P(A|B)$ be the probability of A given that B has already occurred (the **conditional probability**).

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

This also implies:

$$P(A \cap B) = P(A|B) \times P(B) = P(B|A) \times P(A).$$

Intuition: The formula expresses the idea that the probability of both A and B happening equals the probability that B happens, multiplied by how likely A is given B has already happened. In other words, it breaks a complex event ("A and B") into two simpler stages.

If A and B are independent, then $P(A|B) = P(A)$. This makes sense because if B does not influence A, knowing that B occurred doesn't change A's probability. Substituting gives:

$$P(A \cap B) = P(A) \times P(B),$$

which is the defining property of independent events.

An good example of why the nicer independent event formula above does not always work is $P(A \cap A)$ which should be P(A), but with the independence formula we get something $<= P(A)$

Other Examples:

If I flip a coin, will the next flip be independent of the first one?

Let's say I have a bowl with 3 green balls and 5 red balls. Let'say I extract two balls one after another. Will the second extraction be independent of the first one?

3.2 Law of total probability

Let Ω be partitioned into three disjoint subsets B_1, B_2, B_3. Then the probability of A can be expressed as:

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3).$$

Substituting the intersection formula:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3).$$

This can be generalized to any number of disjoint subsets forming a partition of Ω.

Example:

Let's say we have bowl of black(B) and white(W) balls. Find the probability of the first ball being white(P(W1)) and the probability of the second one being black (P(B2)).

Notation and Terminology check:

Experiment: Extract to balls one after another
Sample Space: $\Omega = \{WW, WB, BW, BB\}$, where WW means we extracted two white balls, BB means we extract two black balls and so on.
Events: $W1 = \{WW, WB\}$, $W2 = \{WW, BW\}$, $B1 = \{BW, BB\}$, $B2 = \{WB, BB\}$
I will also give the following probabilities:
$P(W1 \cap W2) = \frac{1}{2}$
$P(W1 \cap B2) = \frac{1}{4}$
$P(B1 \cap W2) = \frac{1}{8}$
$P(B1 \cap B2) = \frac{1}{8}$

Solution:

We only have black and white balls => $W1 = \overline{B1}, W2 = \overline{B2}$
Apply law of total probability
$P(W1) = P(W1 \cap W2) + P(W1 \cap \overline{W2}) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$

3.3 Bayes' theorem

Sometimes we need to find $P(A|B)$ without direct access to $P(A \cap B)$, but knowing $P(B|A)$:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}.$$

Derivation:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(B|A) \times P(A).$$

Substitute this into the first equation to obtain Bayes' formula.

3.4 Example: Probability of disease

Let's say we got an infectious disease(what did you think of?), we know that 5 in 1000 people have it.
We also have a mechanism to test people to check whether they have the disease (positive result => disease). But the mechanism is not perfect:
We know that for 10 in 100 sick people our test gives a negative result and for 5 in 100 healthy people it gives a positive result.
You tested and got a positive result. Knowing that, what is the probability of you having the disease?

Notation:
A = person has disease event
$\overline{A}$ = person does not have disease
T+ = test result is positive
T- = test result is negative

$$P(A) = 0.005, \quad P(T^-|A) = 0.1, \quad P(T^+|\overline{A}) = 0.05.$$

Compute complementary probabilities:

$$P(T^+|A) = 1 - P(T^-|A) = 0.9, \quad P(T^-|\overline{A}) = 1 - 0.05 = 0.95.$$

Compute $P(T^+)$ using the law of total probability:

$$P(T^+) = P(T^+|A)P(A) + P(T^+|\overline{A})P(\overline{A}) = 0.9 \times 0.005 + 0.05 \times 0.995 = 0.05425.$$

Finally, apply Bayes' theorem:

$$P(A|T^+) = \frac{P(T^+|A)P(A)}{P(T^+)} = \frac{0.9 \times 0.005}{0.05425} \approx 0.083 \text{ or } 8.3\%.$$

Even with a "good" test, the probability of truly having the disease after a positive test is low — because the disease itself is rare.

Uses in AI:
The first model you will learn about in the next semester when doing machine learning is the Multinomial Naive Bayes classifier. You will also learn about the pretty complicated bayesian networks when doing knowledge representation.

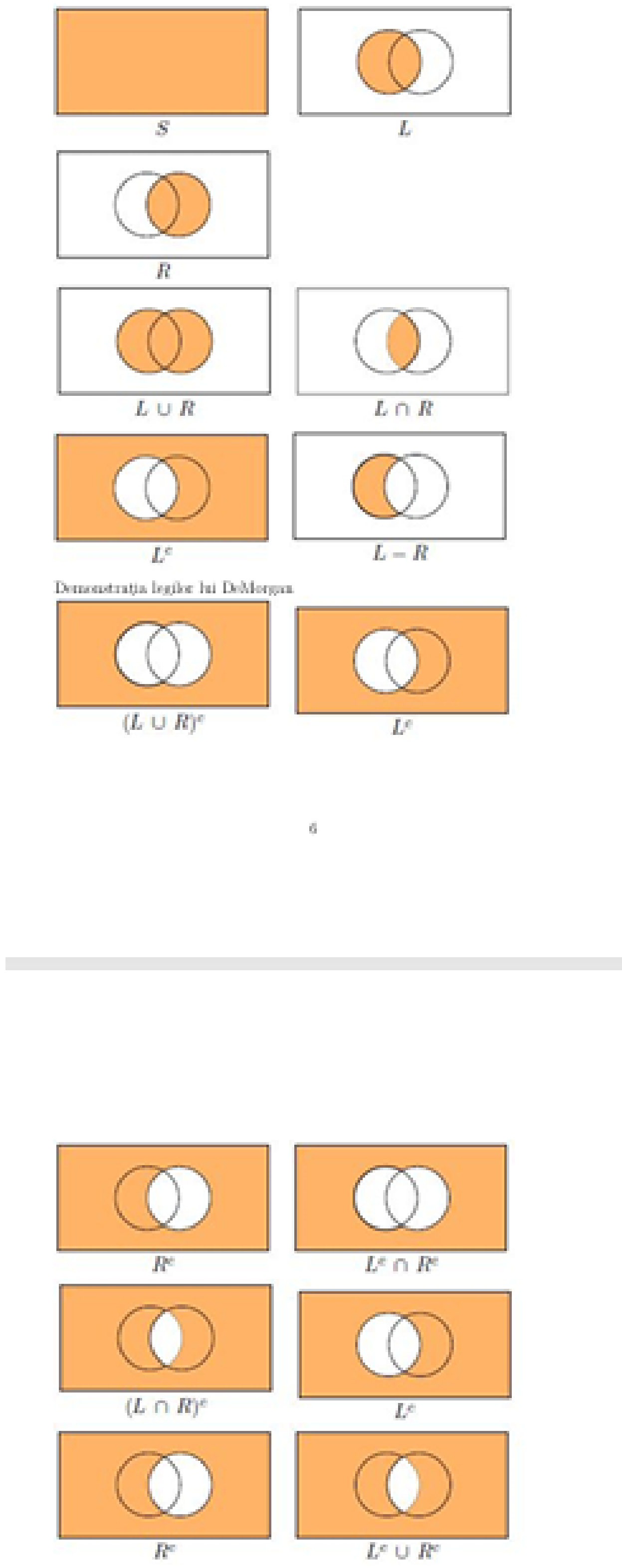

Figure 1: Set operations

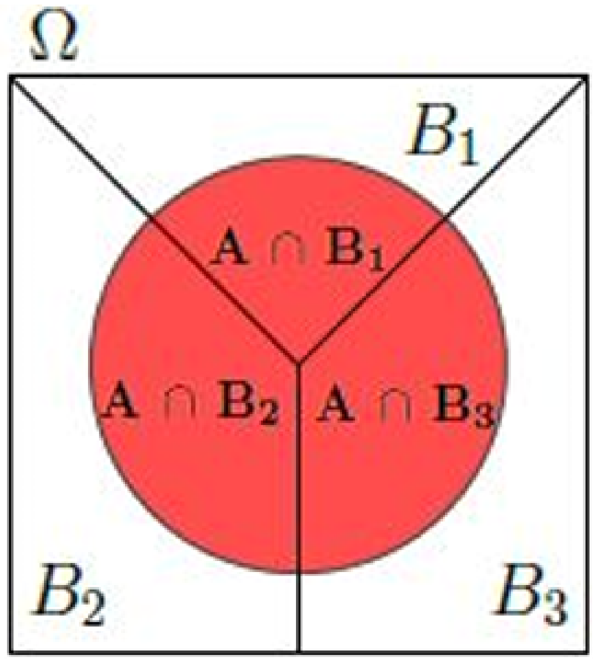

Figure 2: Total probability law illustration