Probability theory

Prof.dr.hab. Viorel Bostan

Technical University of Moldova

viorel.bostan@adm.utm.md

Lecture 1

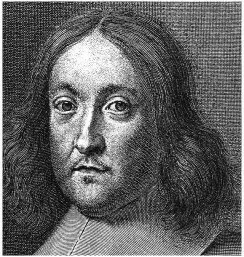
Lecture 1 Outline

- General concept of probability;
- Sample space;
- Probability distribution;
- Axioms of probability;
- Properties of probability;
- Tables and trees to represent outcomes;
- Simple examples.

Founding fathers of probability

Probability theory began in 17th century France in the correspondence of two great French mathematicians, Blaise Pascal and Pierre de Fermat.

Blaise Pascal
(1623–1662)

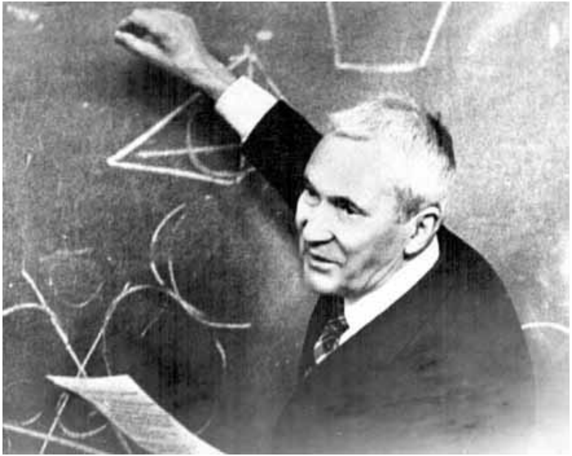


Pierre Fermat
(1607–1665)

They influenced such researchers as Huygens, Bernoulli and DeMoivre in establishing a mathematical theory of probability.

More bits of history

A.N.Kolmogorov in 1920 developed the theory based on Lebesgue spaces.



Andrey Nikolaevich Kolmogorov (1903–1987)

Why probability is important for Computer Science?

- Many algorithms rely on randomization.
- Many aspects of computer systems are designed around probabilistic assumptions and analysis:
  - memory management;
  - branch prediction;
  - packet routing;
  - load balancing.
- Probability comes up in:
  - information theory;
  - cryptography;
  - artificial intelligence;
  - big data analysis;
  - bio-informatics;
  - optimization theory;
  - game theory.

Why probability is important for CS?

- Beyond these engineering applications, an understanding of probability gives insight into many daily problems, such as:
  - polling/surveys;
  - weather prediction;
  - DNA testing;
  - data analysis;
  - actuarial sciences (insurance);
  - risk assessment;
  - financial investing;
  - disease prevention;
  - pharmaceutics;
  - gambling.

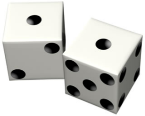
Machine Learning is based on Probability!

So probability is good stuff!

Probably!

Back to history

Consider a dice game that played an important role in the historical development of probability.



Caravaggio, “The Cardsharps” (1594)

Chevalier de Mere

Back to history

It is said that de Mere had been betting that, in **four rolls of a die**, at least one six would turn up. He was winning consistently.

In order to get more people to play, he changed the game to bet that, **in** 24 **rolls of two dice**, a pair of sixes would turn.

It is claimed that de Mere was loosing money in long term with 24 and felt that 25 rolls were necessary to make the game favorable.

So, he asked Blaise Pascal to explain him that "mystery".

Pascal send an **e-mail** to Fermat and started the discussion.

The probability that no 6 turns up on the first toss is $\frac{5}{6}$.

The probability that no 6 turns up on either of the first two tosses is

$$\left(\frac{5}{6}\right)^2.$$

Back to history

Reasoning in the same way, the probability that no 6 turns up on any of the first four tosses is

$$\left(\frac{5}{6}\right)^4 .$$

Thus, the probability of at least one 6 in the first four tosses is

$$1 - \left(\frac{5}{6}\right)^4 = 0.518.$$

So, in more than 51% of the cases de Mere was winning!

Similarly, for the second bet, with 24 rolls, the probability that de Mere wins is

$$1 - \left(\frac{35}{36}\right)^{24} = 0.491$$

and for 25 rolls it is

$$1 - \left(\frac{35}{36}\right)^{25} = 0.506.$$

What is a probability problem?

- Probability theory is a mathematical framework for reasoning about **uncertainty**.
- Every probability problem involves some sort of **randomized** experiment, process, or game.
- A **randomized** experiment/process is an experiment/process whose outcome nobody knows at the beginning.
- Each such problem involves two distinct challenges:
  1. How do we model the situation mathematically?
     - List of all possible outcomes (sample space);
     - Describe event or events of interest;
     - Probability law: describes our beliefs about which outcomes are more likely to occur than others.
     - Probability laws have to obey certain basic properties (axioms).
  2. How do we solve the resulting mathematical problem?
- Probability problems are classified upon the number of possible outcomes into
  - **Discrete** probability problem;
  - **Continuous** probability problem.

Discrete Probability. Outcomes

Consider chance experiments with a finite number of possible outcomes: $\omega_1, \omega_2, \dots, \omega_n$.

Example

Roll a die (which is an experiment) and the possible outcomes are: $\{1, 2, 3, 4, 5, 6\}$ corresponding to the side that turns up.

Example

Toss a coin (another experiment) with possible outcomes: H (heads) and T (tails).

Example

Outcomes of two tosses of a coin (experiment) are: $\{HH, HT, TH, TT\}$.

Example

Rolling a pair of dice (experiment) will have the possible outcomes of the form $\big\{(1,1), (1,2), (1,3), \dots (3,3), (3,4), \dots (5,6), (6,6)\big\}$.

Discrete Probability. Sample space

Definition

Suppose we have an experiment whose outcome depends on chance. The set of all possible outcomes of a chance experiment is called **sample space** of the experiment. Usually it is denoted by Ω.

Sample space must be

- Collectively exhaustive;
- Mutually exclusive.

Model requirement: to be at the “right” detalization. We have some freedom about the details of how we're going to describe sample space. And the question is:
How much detail are we going to include?

Definition

The elements of a sample space are called **outcomes**.
Any subset of a sample space is defined to be an **event**.

Probabilities are assigned to events.

Frequently we might refer **numerically** to an outcome of an experiment.

Example

Consider the mathematical expression which gives the sum of three rolls of a six-sided die. To do this, we could let X_1, X_2 and X_3 represent the values of the outcomes of the three rolls, and then we could write the expression

$$Y = X_1 + X_2 + X_3$$

for the sum of the three rolls. The X_i's and Y are so-called **random variables**.

Definition

A **random variable** is the expression whose value is (or depends on) the outcome of a particular experiment.

Just as in the case of other types of variables in mathematics, random variables can take on different values.

Example

A die is rolled once. We let X denote the outcome of this experiment. Then the sample space for this experiment is the 6−element set

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

where each outcome i, for $i = 1, \dots, 6$, corresponds to the number of dots on the face which turns up. The event

$$E = \{2, 4, 6\}$$

corresponds to the statement that the result of the roll is an even number.
The event E can also be described by saying that X is even.

Definition

If the sample space is either finite or *countably infinite*, the experiment, the sample space and the associated random variables are said to be **discrete**.

Probability Axioms

1. How do we model the situation mathematically?
   - List of all possible outcomes (sample space);
   - Describe event or events of interest;
   - Probability law: describes our beliefs about which outcomes are more likely to occur than others.
   - Probability laws have to obey certain basic properties (axioms).

Next we shall assign probabilities.

Usual convention is that probabilities are numbers between 0 and 1.

Axioms

1. **Nonnegativity**: $P(E) \geqslant 0$ for any event E.
2. **Normalization**: $P(\Omega) = 1$.
3. **Additivity**: If $E \cap F = \emptyset$, then $P(E \cup F) = P(E) + P(F)$.

Probability Axioms

Axioms

1. **Nonnegativity**: $P(E) \geqslant 0$ for any event E.
2. **Normalization**: $P(\Omega) = 1$.
3. **Additivity**: If $E \cap F = \emptyset$, then $P(E \cup F) = P(E) + P(F)$.

$$1 = P(\Omega) = P(E \cup E^c) = P(E) + P(E^c) \geqslant P(E).$$

Therefore, for any event E we have $0 \leqslant P(E) \leqslant 1$. From the same relation we get

$$P(E^c) = 1 - P(E).$$

Note that the third axiom needs some strengthening!

$$\begin{aligned} P(E \cup F \cup G) &= P((E \cup F) \cup G) = P(E \cup F) + P(G) \\ &= P(E) + P(F) + P(G). \end{aligned}$$

for any disjoint sets E, F and G.

Probability Axioms

Theorem

If $A_1, \ldots, A_n$ are pairwise disjoint subsets of Ω (i.e., no two of the A_i have an element in common), then

$$P(A_1 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n),$$

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

Suppose we have a probabilistic experiment with discrete sample space Ω and an event $E = \{\omega_1, \omega_2, \ldots, \omega_k\}$, which is a finite subset of Ω, $E \subset \Omega$. What is $P(E)$?

$$\begin{aligned} P(E) &= P(\{\omega_1, \omega_2, \ldots, \omega_k\}) \\ &= P(\{\omega_1\} \cup \{\omega_2\} \cup \ldots \{\omega_k\}) \\ &= P(\{\omega_1\}) + P(\{\omega_2\}) + \cdots + P(\{\omega_k\}) \\ &= P(\omega_1) + P(\omega_2) + \cdots + P(\omega_k). \end{aligned}$$

Definition

Consider an experiment with finite sample space Ω and let X be a random variable associated with this experiment. A **distribution function** (also called a probability mass function, or pmf) for X is a real-valued function m whose domain is Ω and which satisfies:

$$1. \quad m(\omega) \geq 0, \quad \mathit{for} \quad \mathit{all} \quad \omega \in \Omega,$$

$$2. \quad \sum_{\omega \in \Omega} m(\omega) = 1.$$

For any subset E of Ω, define **the probability of** E to be the number $P(E)$ given by

$$P(E) = \sum_{\omega \in E} m(\omega).$$

Example

For rolling a die,

$$m(i) = \frac{1}{6}, \quad i = 1, 2, \ldots, 6.$$

Clearly, it is a distribution function. Then

$$\begin{aligned} P(E) &= P(\{2,4,6\}) \\ &= \sum_{\omega \in \{2,4,6\}} m(\omega) \\ &= m(2) + m(4) + m(6) \\ &= \frac{3}{6} = \frac{1}{2}. \end{aligned}$$

Consider rolling a pair of dice. Sample space is

$$\Omega = \{(i,j) \mid 1 \leqslant i \leqslant 6, 1 \leqslant j \leqslant 6, i,j \in \mathbb{Z}\}.$$

Let X be the result of the first roll, and Y the result of the second. Both are random variables associated to this experiment. We can represent the outcomes in the table form.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$X =$ First roll

$Y =$ Second roll

What probability law we should assign?
Every possible outcome is equally likely, thus we will assign to every outcome the same probability $\frac{1}{36}$: $m(\omega) = \frac{1}{36}$ for any $\omega \in \Omega$.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$X =$ First roll

$Y =$ Second roll

- $P\big((X, Y) = (1, 3) \text{ or } (X, Y) = (4, 5)\big)$?
- $P\big(X = 2\big)$?
- $P\big(Y \text{ is even}\big)$?
- $P\big(X + Y \geqslant 9\big)$?
- $P\big(\min(X, Y) = 3\big)$?

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

X = First roll

Y = Second roll

- $P\big((X,Y)=(1,3) \text{ or } (X,Y)=(4,5)\big) = 1/36 + 1/36 = 2/36.$
- $P\big(X=2\big) = 6/36.$
- $P\big(Y \text{ is even}\big) = 18/36.$
- $P\big(X+Y \geqslant 9\big) = 10/36.$
- $P\big(\min(X,Y)=3\big) = 7/36.$

Example

Consider tossing a coin twice. There are several ways to record the outcomes of this experiment:

1. Record the two tosses, in the order they occurred:

$$\Omega_1 = \{HH, HT, TH, TT\}.$$

2. Record the outcomes by simply recording the number of heads that appeared:

$$\Omega_2 = \{0, 1, 2\}.$$

3. Record the two outcomes, without regard to the order in which they occurred:

$$\Omega_3 = \{HH, HT, TT\}.$$

Example (Contd.)

Assume that all four outcomes are equally likely, and define the distribution function $m(\omega)$ by

$$m(HH) = m(HT) = m(TH) = m(TT) = \frac{1}{4}.$$

Let $E = \{HH, HT, TH\}$. Then

$$P(E) = m(HH) + m(HT) + m(TH) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

Similarly, if $F = \{HH, HT\}$ then

$$P(F) = m(HH) + m(HT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Properties of probability

Theorem

Let $A_1, \ldots, A_n$ be pairwise disjoint events with $\Omega = A_1 \cup \ldots \cup A_n$ and let E be any event. Then

$$P(E) = \sum_{i=1}^{n} P(E \cap A_i).$$

Corollary

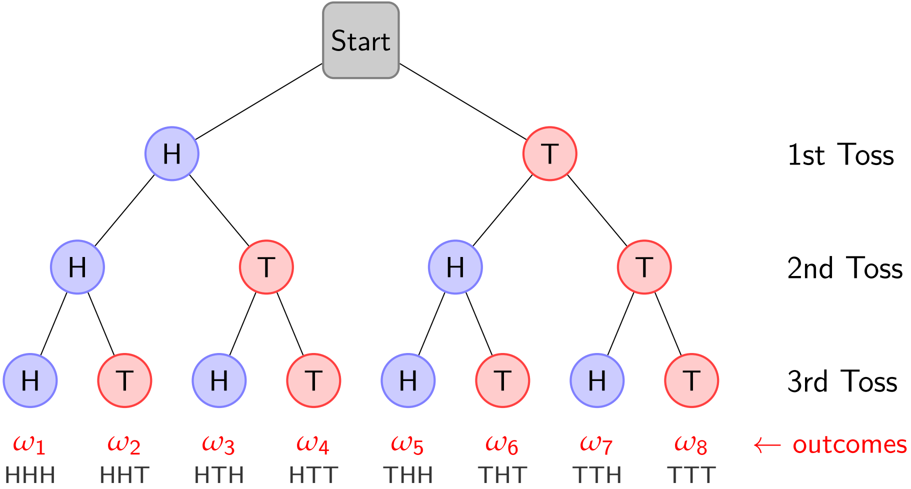
For any two events E and F, $P(E) = P(E \cap F) + P(E \cap F^c)$.

Theorem

If E and F are any two events, then

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

A very useful tool in solving discrete probability problems are tree diagrams.
Consider three tosses of a coin.



Example (Contd.)

Sample space is

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_8\}$$
$$= \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

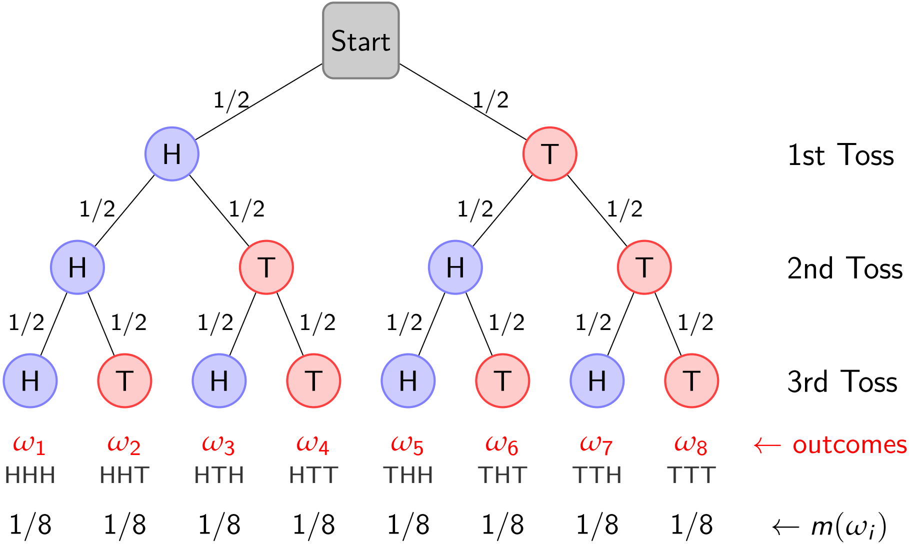
All 8 outputs are equally likely to occur.

Let E be the event that at least one head will occur,

and F be the event that exactly one pair of either heads or tails will happen.

Compute $P(E)$ and $P(F)$.

Tree diagrams



Example (Example contd.)

Let E be the event that at least one head turns up.

Then E^c is the event that no heads turn up. This event occurs for only one outcome ω_8.

Thus, $E^c = \{TTT\}$ and therefore $P(E^c) = m(TTT) = \frac{1}{8}$.

$$P(E) = 1 - P(E^c) = 1 - \frac{1}{8} = \frac{7}{8}.$$

Let F be the event that either a pair of heads or a pair of tails turn up.

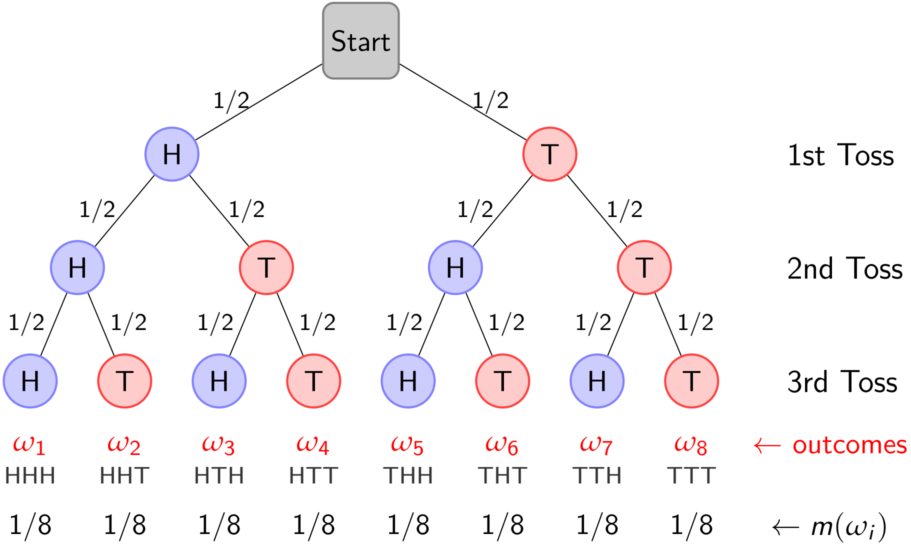
Let F_1 be the event that exactly one pair of heads turns up,

and F_2 consists of exactly a pair of tails. Then $F = F_1 \cup F_2$ and use formula:

$$\begin{aligned} P(F) &= P(F_1) + P(F_2) - P(F_1 \cap F_2), \\ &= P(\{HHT, HTH, THH\}) + P(\{TTH, THT, HTT\}) - P(\emptyset), \\ &= \frac{3}{8} + \frac{3}{8} - 0 = \frac{3}{4}. \end{aligned}$$

Tree diagrams

Compute the probability that either the first outcome is head or the second outcome is a tail.



Example (Contd.)

Probability that either 1st outcome is head or 2nd outcome is tail.

Let $A = \{\text{1st outcome is head}\} = \{\omega_1, \omega_2, \omega_3, \omega_4\}$,

and $B = \{\text{2nd outcome is tail}\} = \{\omega_3, \omega_4, \omega_7, \omega_8\}$.

By looking at the paths in the tree, we see that

$$\begin{aligned} P(A) = P(B) &= \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} = \tfrac{1}{2}, \\ A \cap B &= \{\omega_3, \omega_4\}, \\ P(A \cap B) &= \tfrac{1}{8} + \tfrac{1}{8} = \tfrac{1}{4}, \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \tfrac{1}{2} + \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{3}{4}. \end{aligned}$$

Basic steps in solving discrete probability problems

Distinguish 4 basic steps in the solution process of discrete probability problems:

1. Find the Sample Space Ω.
2. Define Events of Interest.
3. Determine Outcome Probabilities (find distribution function $m(\omega)$).
4. Compute Event Probabilities.

Lecture 1 Summary

- General concept of probability;
- Sample space;
- Probability distribution;
- Axioms of probability;
- Properties of probability;
- Tables and trees to represent outcomes;
- Simple examples.

Probability Joke

The probability of someone watching you
is proportional to...

the stupidity of your action.