EFOP-3.4.3-16-2016-00014

Probability

Handout

Prepared by

János Marcell BENKE, PhD

Methodological expert: Edit Gyáfrás

This teaching material has been made at the University of Szeged and supported by the European Union. Project identity number: EFOP-3.4.3-16-2016-00014

2020

Szegedi Tudományegyetem
Cím: 6720 Szeged, Dugonics tér 13.
`www.u-szeged.hu`
`www.szechenyi2020.hu`

Preface

Probability incorporates two courses: the related lecture and the seminar. The aim of these courses is to teach the meaning of the basic notions of probability theory as probability, random variables, probability distribution, expected value, standard deviation, variance, covariance, density function, normal distribution.

This handout presents the definitions and theorems applied during the semester and provides examples for better understanding and for practicing and also includes sample tests. The starred exercises are a slightly more difficult then the others. Problems like these will not appear in the tests.

Lecturer: János Marcell BENKE

Contents

Course information

Course title: PROBABILITY

Course code:
60A106 Lecture
60A107 Seminar

Credit: $3 + 2$

Type: lecture and seminar

Contact hours / week: $2 + 1$

Evaluation:
Lecture: exam mark (five-grade)
Seminar: practical course mark (five-grade)

Semester: 3^{rd}

Prerequisites: Calculus

Learning outcomes

(a) regarding knowledge, the student

- can use basic combinatorical counting methods as permutation, variation and combination to count the number of possible outcomes of an experiment
- is able to simplify complex events using operations on events
- knows the definition of probability and the heuristic of it as well
- can calculate probabilities on classical probability space
- can calculate probabilities of events connected to discrete random variable using its probability distribution
- can calculate expected value, standard deviation, variance of discrete random variable using its probability distribution
- understands the heuristical meaning of expectation and standard deviation
- understand the meaning of covariance and correlation
- is able to identify binomial, hypergeometric and geometric distribution
- understands the meaning of conditional probability and is able to calculate it
- can calculate probabilities of events connected to continuous random variable using its probability density function
- can calculate expected value, standard deviation, variance of continuous random variable using its probability density function
- can calculate probabilities of events connected to normally distributed random variable
- can calculate approximated probabilities using normal distribution

(b) regarding skills, the student

- can uncover facts and basic connections
- can draw conclusions and make critical observations along with preparatory suggestions using the theories and methods learned

(c) regarding attitude, the student

- behaves in a proactive, problem oriented way to facilitate quality work
- is open to new information, new professional knowledge and new methodologies

(d) regarding autonomy, the student

- conducts the tasks defined in his/her job description independently under general professional supervision
- takes responsibility for his/her analyses, conclusions and decisions

Requirements

During the semester there are 5 small tests and each small test is worth 5 points. There is no way to retake any of the small tests. Based on the collected points the following grade is given for the seminar:

1: 0-5 points,
2: 6-8 points,
3: 9-11 points,
4: 12-14 points,
5: 15-25 points.

The examination is a big test, which can be written in the exam period. It is worth 50 points. It is profitable to collect more points from the small test, because the points over 15 will be added to the big test points. Based on the collected points the following grade is given for the lecture:

1: 00-19 points,
2: 20-26 points,
3: 27-33 points,
4: 34-39 points,
5: 40-50 points.

Example: in the case of 20/25 small test points and 37/50 big test points, the seminar mark is 5, and the overall point is 5+37=42, so the lecture mark is 5 as well.

Course topics

Combinatorial counting methods, basic properties of probability, classical probability space, conditional probability. Discrete random variables, expectation, variance. Continuous random variables. Moments, skewness, curtosis, median and quantiles. Law of large numbers and central limit theorem.

1 Counting methods

When calculating probabilities we often have to rely on combinatorial methods to find the total number of possible outcomes of a random experiment. In this chapter we are going to refresh these methods, and introduce (standard) notations for them.

Definition 1.1 (Factorial)**.** The factorial of a non-negative integer n, denoted by $n!$, is the product of all positive integers less than or equal to n, that is

$$0! := 1, \qquad n! := \prod_{i=1}^{n} i = 1 \cdot 2 \cdot \ldots \cdot n, \quad n \geqslant 1.$$

Permutations of different objects

Proposition 1.2 (Permutations without repetition)**.** *Suppose we have n different objects, then we can arrange them in order in $n!$ different ways.*

Proof. We have n choices for the first element of the sequence, then we have only $n-1$ choices for the second element, because we are allowed to use the same object only once. For the third element we have $n-2$ choices, and so on. Multiplying these numbers together gives the desired result. □

Example 1.3. How many ways can we rearrange the letters of the word MATH?

Solution. Since the word MATH consists of 4 distinct letters, the number of permutations of these letters is $4! = 24$. It is such a small number, that we can list all those rearrangements:

MATH	MAHT	MTAH	MTHA	MHAT	MHTA
ATHM	ATMH	AHTM	AHMT	AMTH	AMHT
THAM	THMA	TAMH	TAHM	TMAH	THHA
HMAT	HMTA	HTAM	HTMA	HATM	HAMT

□

Example 1.4. A deck of French playing cards consists of 52 cards, one card for each of the possible rank and suit combinations, where there are 13 ranks namely 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King, Ace and four suits Clubs(♣), Diamonds(♢), Hearts(♡), Spades(♠). If we shuffle this deck of cards what is the number of possible outcomes?

Solution. Since each card in the deck is unique, we have 52 different objects, therefore the number of possible permutations is $52! \approx 8.065 \cdot 10^{67}$. That is a huge number with 68 digits. It would be futile to try to list all the possible outcomes. □

Permutations of not necessarily different objects

Our method fails to give the correct answer if the objects we aim to rearrange are not unique. In order to find all the rearrangements of the letters of the word FOOD we need to refine our way of calculating permutations.

Proposition 1.5 (Permutations with repetition)**.** *Suppose we have $\ell > 0$ types of objects. We take k_1 of object 1, k_2 of object 2, and so on k_ℓ of object ℓ, a total of $n = k_1 + \cdots + k_\ell$ objects. The number of permutations these objects have is*

$$\frac{n!}{k_1! k_2! \cdot \ldots \cdot k_\ell!}.$$

Note that if we only have 1 of each object then we get back the statement of Proposition 1.2, since $1! = 1$. We will not prove this proposition, however we will illustrate the idea of the proof in the following example.

Example 1.6. How many ways can we rearrange the letters of the word FOOD?

Solution. First we assume that we have a way to distinguish the 2 letters O, for example we can use a different colour for them. Then we have 4 different letters, and we already know the number of rearrangements is $4! = 24$. We list them to illustrate the point

DFOO	DFOO	DOFO	DOFO	DOOF	DOOF
FDOO	FDOO	FODO	FODO	FOOD	FOOD
ODFO	ODFO	ODFO	ODFO	ODOF	ODOF
OODF	OODF	OOFD	OOFD	OFOD	OFOD

The problem is, that we have counted some words more than once. In fact we have counted each word twice! The reason for this is that we have two choices when we select which letter O has red colour. Therefore the correct answer, the number of possible rearrangements of the letters FOOD is $4!/2 = 12$. Since $2! = 1 \cdot 2 = 2$ that is the same answer we get using Proposition 1.5. □

Example 1.7. How many ways can we rearrange the letters of the word MATHEMATICS?

Solution. The word mathematics consists of 11 letters, but the letters A, M, and T appear twice. Using Proposition 1.5 we get that the number of possible rearrangements are

$$\frac{11!}{2! \cdot 2! \cdot 2!} = 4989600.$$

□

Variations of objects

Previously we were limited by the number of available objects of each type. If we allow ourselves the option to use each object as many times as we wish, then the number of possible sequences we can make are infinite, however if we restrict ourselves to sequences of a given length the answer becomes finite.

Proposition 1.8 (Variations with repetition)**.** *Suppose we have $n > 0$ types of objects. If we are allowed to use each object multiple times, then the number of sequences of length $\ell > 0$ we can create using these objects is n^ℓ.*

Proof. We have n choices for the first element of the sequence, then we have the same n choices for the second element, because we are allowed to use the same object more than once. This gives us n choices for each of the ℓ elements of the sequence, multiplying them together gives the desired result. □

Example 1.9. How many 3 digit numbers can we make using the digits 1,4,7 if we can use them as many times as we want to?

Solution. The previous proposition gives us the answer, we have $n = 3$ different objects, and we want to count sequences of length $\ell = 3$ therefore the answer is $n^\ell = 3^3 = 27$. We can list the possible numbers

111	114	117	141	144	147	171	174	177
411	414	417	441	444	447	471	474	477
711	714	717	741	744	747	771	774	777

□

Example 1.10. We roll a dice twice, list all the possible outcomes!

Solution. A dice has 6 sides marked with numbers 1, 2, 3, 4, 5, 6, that is our $n = 6$ different objects. We roll the dice twice, that means we are looking for sequences of length $\ell = 2$. Therefore the number of these outcomes is $n^\ell = 6^2 = 36$. Here we list them

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

□

Note that we included both (1, 2) and (2, 1) in our list, that is because we care about the order in which these numbers appear. There are cases when we don't need information about the order. One such case is discussed in the next subsection.

Combination without repetition

So far we cared about the order of things, but this time we study questions where the order of things is irrelevant. For example in a game called *poker*, each player draws 5 cards from a shuffled deck of French playing cards and after some betting (which we ignore here, thus vastly simplifying the game) the player with the most powerful combination of cards wins. The combination called *royal flush* is a hand of 5 cards with the ranks 10, Jack, Queen, King and Ace with the same suit, it is the most powerful one as it beats all other combinations and ties with another royal flush. In order to study the number of different poker hands we have to introduce the binomial coefficient.

Definition 1.11. Let $n \in \mathbb{N}$ a positive integer, and $0 \leq k \leq n$ another integer, then the n choose k binomial coefficient is denoted by $\binom{n}{k}$ and can be calculated as

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Proposition 1.12 (Combination without repetition)**.** *A set with n unique elements has $\binom{n}{k}$ different subsets of size k, that is, if we have n different objects, then we can choose k of them in $\binom{n}{k}$ ways.*

Proof. We are going to use permutations to demonstrate this. Suppose we select the objects by taking a permutation of the n elements, and taking the first k objects of that ordering. Then we have $n!$ permutations, however not all of them produce a different selection, because the order of the first k and last $n-k$ elements don't matter. So we have counted each selection $k!(n-k)!$ times, dividing with this number gives the desired result. □

Example 1.13. Suppose we want to make pizza. Our basic pizza will have tomato sauce and cheese but we want additional toppings. How many different pizzas can we make using exactly two of the four available toppings: ham, pineapple, corn, mushroom?

Solution. The previous proposition gives us the correct result

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{24}{4} = 6.$$

In order to illustrate the idea of the proof, we are going to list the permutations and the pizzas we can create using them. Let H, P, C, and M denote ham, pineapple, corn and mushroom respectively. Then we can list all possible permutations and the pizzas we would make using those permutations as a basis

HPCM	HPMC	PHCM	PHMC	$\longrightarrow$	{ham, pineapple}
HCPM	HCMP	CHPM	CHMP	$\longrightarrow$	{ham, corn}
HMCP	HMPC	MHCP	MHPC	$\longrightarrow$	{ham, mushroom}
PCHM	PCMH	CPHM	CPMH	$\longrightarrow$	{pineapple, corn}
PMHC	PMCH	MPHC	MPCH	$\longrightarrow$	{pineapple, mushroom}
CMHP	CMPH	MCHP	MCPH	$\longrightarrow$	{corn, mushroom}

□

Example 1.14. How many poker hands are possible, that is how many ways can we draw 5 cards out of a deck of french playing cards?

Solution. Using the n choose k formula, all we have to do is substitute $n = 52$, the number of cards in the deck, and $k = 5$, the number of cards we are drawing. Then the answer we are looking for is

$$\binom{52}{5} = \frac{52!}{5!47!} = \frac{48 \cdot 49 \cdot 50 \cdot 51 \cdot 52}{5!} = 2598960.$$

□

Example 1.15. The most famous Hungarian lottery system is called *ötöslottó*, and it is a game of luck, where you can buy a ticket for some amount of money (at the time of writing this, it is 225 Hungarian forints) where you have to mark 5 numbers out of 90. At each weekend the company responsible for the lottery randomly generates 5 numbers out of 90, and gives out prizes for those who guessed at least 2 of them correctly on their ticket. While the prize for guessing only 2 numbers correctly is only a small amount of money, guessing all 5, thus hitting the jackpot makes the winner wealthy. How many ways can we fill out a lottery ticket, that is how many different tickets should we fill out, if we want to guarantee a jackpot?

Solution. Using the n choose k formula, all we have to do is substitute $n = 90$, the number of numbers on the ticket, and $k = 5$, the number of guesses we have to make. Then the answer we are looking for is

$$\binom{90}{5} = \frac{90!}{5!85!} = \frac{86 \cdot 87 \cdot 88 \cdot 89 \cdot 90}{5!} = 43949268.$$

□

This is not a winning strategy as the total price of those tickets would be much more than the jackpot of a lottery game. Even if the jackpot would be greater than the price, we would face at least two problems. In case of multiple winners the prize is split evenly and even if no one fills out all the possible tickets someone might get lucky and ruin our investment. Also filling out the tickets would be no small task, filling out 1 ticket per second (an unreasonable speed) it would take 43949268 seconds, roughly 1 years and 5 months of non-stop work.

Further readings:

- `https://en.wikipedia.org/wiki/Pascal%27s_triangle`
- `https://en.wikipedia.org/wiki/Poker`
- `https://en.wikipedia.org/wiki/Lottery`

1.1 Exercises

Problem 1.1. How many ways can we rearrange the letters of the word PROBLEM?

Problem 1.2. How many four digit numbers can we create using the digits 1,3,5,6, if we can only use each digit once? How many even four digit numbers can we create using the same digits? What is the answer to these questions if we can use the same digits more than once?

Problem 1.3. A group of friends Emma, Jennifer, Peter, Sam and Adam go to the cinema to watch the movie The Lion King. They have tickets for seat numbers 3 to 7 in the 8th row. How many ways can they be seated? How many ways can they be seated if Emma sits next to Sam?

Problem 1.4. How many ways can we rearrange the letters of the word EXERCISES?

Problem 1.5. How many four digit numbers can we create using the digits 1,2,2,6? How many odd four digit numbers can we create using the same digits?

Problem 1.6. We have 5 red, 3 green and 3 yellow balls. How many ways can we arrange them?

Problem 1.7. How many ways are there to order three scoops of ice cream in a cone if the shop sells 12 different flavours of ice cream? (Suppose that the order of scoops matter.)

Problem 1.8. How many possible outcomes does the experiment of rolling a dice 10 consecutive times have?

Problem 1.9. The local florist sells 5 different kinds of flowers. We'd like to surprise someone with a bouquet made from two different kinds of flowers. How many ways can we do that?

Problem 1.10. How many ways can we select a student council with 3 equal members out of a class of 31 students? How many can we select a student council consisting of a president, a secretary, and a spokesman with different responsibilities out of a class of 31? (Note that no student can hold more than one title of the student council.)

Problem 1.11. The deck of 52 French playing cards is the most common deck of playing cards used today. It includes thirteen ranks of each of the four French suits; clubs($\clubsuit$), diamonds($\diamondsuit$), hearts($\heartsuit$), spades($\spadesuit$). We shuffle the deck and draw 5 cards from it. How many possible 5 card hands can we get? (We don't care about the order of the cards drawn just the cards themselves.)

The final answers to these problems can be found in section 10.

2 Probability

Probability is the branch of mathematics that models randomness. Therefore we need a concept of randomness before we can talk about probabilities of events.

Sample space, events

Definition 2.1 (Random experiment)**.** By a random experiment we mean an experiment that has the following properties

- The possible outcomes are known.
- The outcome of the experiment is not predictable in advance.
- The experiment can be repeated under the same circumstances an arbitrary amount of times.

The first two requirements seem reasonable as we want to speculate about the future, for that we need to know what can happen, but we have to be uncertain as to what will happen. In order to make sense of the third requirement we are going to introduce events and their relative frequencies.

Definition 2.2 (Sample space)**.** The sample space is the set of all possible outcomes of a given random experiment. We denote it by Ω.

We introduce events in an intuitive way. An event is a statement about the outcome of a random experiment whose truth can be decided after carrying out the experiment. Since this representation is not unique (we are going to see this in the first example) we need a unique way to describe an event, and that is going to be the set of outcomes which satisfy the statement.

Definition 2.3 (Event)**.** Let Ω be the sample space of a random experiment, then the subsets of Ω are called events.

For absolute mathematical precision we would have to restrict our definition of events to certain subsets of the sample space, but that is only relevant in cases where the sample space is infinite. Since we almost always use a finite sample space during this subject we are going to omit this considerations.

Definition 2.4. Let Ω be any set. Then we call the set that contains all subsets of Ω the power set of Ω, and we denote it by 2^{Ω}.

Every event is an element of the power set of the sample space of the related random experiment. We say that the event $A \subset \Omega$ occurs or happens if after carrying out the experiment the outcome is in A.

Definition 2.5. There are two events corresponding to the always false and the always true statements that we give special names. We call the event represented by the set Ω the certain event, while we call the event corresponding to the empty set, $\emptyset$ the impossible event.

Example 2.6. Describe the random experiment related to tossing a coin once, and observing whether it lands on heads or tails! Find the set representation for the following statements!

- The coin lands on heads.
- The coin lands on tails.
- The coin does not land on tails.

Solution: The experiment has two possible outcomes, heads or tails, so

$$\Omega = \{heads, tails\}.$$

We can create a table listing all the possible outcomes in the columns and the statements in the rows, and write true or false if the corresponding statements is true or false for the give outcome.

	heads	*tails*
The coin lands on heads.	true	false
The coin lands on tails.	false	true
The coin does not land on tails.	true	false

From this table we can simply list for each statement the set of outcomes which has true assigned to them and we get the following.

The coin lands on heads.	$\longrightarrow$	$\{heads\}$
The coin lands on tails.	$\longrightarrow$	$\{tails\}$
The coin does not land on tails.	$\longrightarrow$	$\{heads\}$

As we can see the first and third statements are different, however they are true for the same outcomes, their set representation is unique. □

Example 2.7. Describe the sample space of the random experiment of rolling a dice! Find at least one statement that has the given set representation as an event for each of the following sets.

- $\{1\}$
- $\{1, 4\}$
- $\{2, 4, 6\}$
- $\{2, 3\}$

Solution: A dice has six sides marked with numbers 1, 2, 3, 4, 5, and finally 6. Therefore the set of all possible outcomes is

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

For each of the sets above we give a trivial representation as a statement, just listing the outcomes it contains and requiring the experiment in resulting one of these, and a non trivial one. The trivial ones,

$\{1\}$	$\longrightarrow$	The dice roll results in 1.
$\{1, 4\}$	$\longrightarrow$	The dice shows 1 or 4.
$\{2, 4, 6\}$	$\longrightarrow$	The outcome is 2 or 4 or 6.
$\{2, 3\}$	$\longrightarrow$	The number shown is either 2 or 3.

and the non trivial ones

$\{1\}$	$\longrightarrow$	The result is less than 2.
$\{1, 4\}$	$\longrightarrow$	The dice shows a square number.
$\{2, 4, 6\}$	$\longrightarrow$	The outcome isn't odd.
$\{2, 3\}$	$\longrightarrow$	The outcome is less than 4, but not equal to 1.

□

Operations on events

Since we represent events as sets, we can use set operations on them. For thr sake of completeness we define the set operations here.

Definition 2.8 (Operations on sets)**.** Let A and B arbitrary subsets of Ω. Then

- The complement of the set A is denoted by $\overline{A}$ (or A^c), it is the set of elements not contained in A, that is
$$x \in \overline{A} \Longleftrightarrow x \notin A.$$
- The union of sets A and B is denoted by $A \cup B$, it is the set of elements that are in either A or B, that is
$$x \in A \cup B \Longleftrightarrow x \in A \text{ or } x \in B.$$
- The intersection of sets A and B is denoted by $A \cap B$, it is the set of elements that are both in A and in B, that is
$$x \in A \cap B \Longleftrightarrow x \in A \text{ and } x \in B.$$
- The difference of sets A and B is denoted by $A \setminus B$, it is the set of elements that are in A, but not in B, that is
$$x \in A \setminus B \Longleftrightarrow x \in A \text{ and } x \notin B.$$
  Be aware that unlike the union and the intersection of sets this operation is not commutative, that is $A \setminus B$ can be different from $B \setminus A$. The complement is a special case of difference as $\overline{A} = \Omega \setminus A$.
- The symmetric difference of sets A and B is denoted by $A \Delta B$, it is the set of elements that are in exactly one of the sets A and B, that is
$$x \in A \Delta B \Longleftrightarrow (x \in A \text{ and } x \notin B) \text{ or } (x \notin A \text{ and } x \in B).$$

We can give probabilistic interpretations of these operations the following way. Let $A, B \subset \Omega$ be arbitrary events, then

- The event $\overline{A}$ happens if A doesn't.
- The event $A \cup B$ happens if A or B happens.
- The event $A \cap B$ happens, if A and B both happen.
- The event $A \setminus B$ happens if A happens, but B doesn't.
- The event $A \Delta B$ happens if exactly one of A and B happens.

Definition 2.9 (Relations of events)**.** We can define relations on events based on how they are related as sets.

- We say that the events A and B are mutually exclusive (or disjoint) is $A \cap B = \emptyset$.
- We say that the event A is a consequence of the event B if $B \subset A$.

Proposition 2.10. *If the events A and B are mutually exclusive then they cannot happen at the same time.*

Proposition 2.11. *If the event A is a consequence of B, then if B happens so does A.*

Relative frequency of events

The heuristic meaning of the probability of an event can be reached through the relative frequency. For instance, why can we make fair decisions by tossing a coin? Because we know that in half of the cases we get heads and in the other half of the cases we get tails.

Definition 2.12 (Relative frequency)**.** Let Ω be the sample space of a random experiment, and $A \subset \Omega$ an arbitrary event. Repeat the experiment $n \in \mathbb{N}$ times and let $k_n(A)$ denote the number of times the event A has occurred. Then the relative frequency of the event A after n repetitions is

$$r_n(A) = \frac{k_n(A)}{n} = \frac{\text{occurences of } A}{\text{total number of repetitions}}.$$

We'd like to interpret the probability of the event A as the limit of the relative frequency $r_n(A)$ as n, the number of repetitions tends to infinity. However there is a problem, as the relative frequency is a random quantity. Take for example these two sequences of random coin tosses

sequence #1:	*tails,*	*tails,*	*heads,*	*tails,*	*heads,*	*heads,*	...
sequence #2:	*heads,*	*heads,*	*tails,*	*tails,*	*tails,*	*tails,*	...

they result in two different sequence of relative frequencies for the event that the coin lands on *heads*

#1:	0,	0,	$\frac{1}{3} \approx 0.33$,	$\frac{1}{4} = 0.25$,	$\frac{2}{5} = 0.4$,	$\frac{3}{6} = 0.5$,	...
#2:	1,	1,	$\frac{2}{3} \approx 0.67$,	$\frac{2}{4} = 0.5$,	$\frac{2}{5} = 0.4$,	$\frac{2}{6} \approx 0.33$,	...

It is possible that these two sequences have the same limit, but we cannot prove it yet, so we cannot build a definition of probability upon that. Instead we are going to prove a few properties of relative frequency and then require those properties in the definition of probability.

Proposition 2.13 (Properties of relative frequency)**.** *Let Ω be the sample space of a random experiment, and $A, B \subset \Omega$ arbitrary events. Then*

(i) the relative frequency is nonnegative, that is $r_n(A) \geq 0$,

(ii) the relative frequency of the certain event is always 1, that is $r_n(\Omega) = 1$,

(iii) the relative frequency of unions of mutually exclusive events adds up, that is

$$A \cap B \Longrightarrow r_n(A \cup B) = r_n(A) + r_n(B).$$

There are other properties of relative frequency however these three will be enough to define probability.

Probability space

Definition 2.14 (Probability Space)**.** Let $(\Omega, \mathcal{A})$ denote the sample space and the set of events for a random experiment. The function $\mathrm{P} : \mathcal{A} \to \mathbb{R}$ is called *probability* if

(i) $\mathrm{P}(A) \geq 0$, for all $A \in \mathcal{A}$,

(ii) $\mathrm{P}(\Omega) = 1$,

(iii) for any pairwise mutually exclusive events $A_1, A_2, \dots \in \mathcal{A}$

$$\mathrm{P}\left(A_1 \cup A_2 \cup \dots\right) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \dots.$$

Whenever we examine a random experiment we work with a triplet $(\Omega, \mathcal{A}, \mathrm{P})$ that is related to the experiment, and we call it the related *probability space*.

The third criteria for P is called the σ-additivity of the probability. Sometimes the probability P is called probability measure. The reason is we can imagine probability as a quantity like length, area, volume, time, and so on, which we can measure. The basic concept of any measurement is that we can measure any object with splitting it into smaller pieces, and the measure of the original object equals the sum of the measure of the pieces. This property is the σ-additivity. The σ prefix means that this property is valid not only for a finite number of pieces but for countable infinite as well.

Theorem 2.15 (Properties of probability)**.** *Let* $(\Omega, \mathcal{A}, \mathrm{P})$ *be a probability space, also let* $A, B, A_1, A_2, \dots, A_n \in \mathcal{A}$*. Then*

(i) $\mathrm{P}(\emptyset) = 0$,

(ii) $\mathrm{P}(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} \mathrm{P}(A_k)$ *provided that* $A_i \cap A_j = \emptyset$ *if* $i \neq j$, $i, j = 1, \dots, n$ *(finite additivity of the probability),*

(iii) if $B \subset A$, *then* $\mathrm{P}(A \setminus B) = \mathrm{P}(A) - \mathrm{P}(B)$ *and* $\mathrm{P}(B) \leq \mathrm{P}(A)$ *(monotonicity of probability),*

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

(v) $\mathrm{P}(A) \leq 1$,

(vi) $\mathrm{P}(A \cup B) \leqslant \mathrm{P}(A) + \mathrm{P}(B)$ *(subadditivity of probability),*

(vii) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$,

(viii) $\mathrm{P}(A_1 \cup A_2 \cup A_3) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) - \mathrm{P}(A_1 \cap A_2) - \mathrm{P}(A_1 \cap A_3) - \mathrm{P}(A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_2 \cap A_3)$.

We can solve a lot of problems with a special, important probability space.

Definition 2.16 (Classical probability space)**.** It is a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ such that

- Ω is a finite, non-empty set,
- $\mathcal{A} := 2^{\Omega}$,
- $\mathrm{P} : \mathcal{A} \to \mathbb{R}$ such that

$$\mathrm{P}(A) := \frac{|A|}{|\Omega|} = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}}, \qquad A \in \mathcal{A}.$$

In a classical probability space, for every $\omega \in \Omega$, we have $\mathrm{P}(\{\omega\}) = \frac{1}{|\Omega|}$. If we model a random experiment using a classical probability space, then it is often useful to use the tools of counting methods to find the number of favorable outcomes and thus the probability of an event.

Example 2.17. Tossing a fair coin twice, what is the probability of getting one heads and one tails?

Solution: By distinguishing the two coins, the sample space is

$$\Omega = \{(\text{H,H}), (\text{H,T}), (\text{T,H}), (\text{T,T})\},$$

where each of the four outcomes has the same probability. So the probability of the event

$$A = \{(\text{H,T}), (\text{T,H})\}$$

in question is

$$\mathrm{P}(A) = \frac{2}{4} = \frac{1}{2}.$$

□

Example 2.18 (Birthday problem)**.** In a set of n randomly chosen people, what is the probability that there are at least two people who have birthdays on the same day? (We exclude leap years and we suppose that each day of the year is equally probable for a birthday.

Solution: If $n > 365$, then, by the pigeonhole principle, the event A in question is the certain event, so it has probability 1.
If $n \leqslant 365$, then

$$\mathrm{P}(A) = 1 - \mathrm{P}(\overline{A}) = 1 - \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}$$

$$= 1 - \frac{365!}{(365-n)! \cdot 365^n} \approx \begin{cases} 0.284 & \text{if } n = 16, \\ 0.476 & \text{if } n = 22, \\ 0.507 & \text{if } n = 23, \\ 0.891 & \text{if } n = 40, \\ 0.970 & \text{if } n = 50, \\ 0.990 & \text{if } n = 57. \end{cases}$$

□

Independence

We have explored relations of events, namely two events being exclusive or one being a consequence of the other. This time our aim is to define a relation of events by having their probabilities satisfy an equation.

Definition 2.19 (Independence of two events)**.** We say the events A and B are *independent* if

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

Independence is a useful notion if we describe (independent) repetitions of random experiments.

Example 2.20. Tossing a fair coin twice, what is the probability of getting two heads.

Solution: The first solution is the following. As in Example 2.17 we can use the sample space

$$\Omega = \{(\mathrm{H,H}), (\mathrm{H,T}), (\mathrm{T,H}), (\mathrm{T,T})\}$$

to model this random experiment, which is a classical probability space, hence the probability of the event

$$A = \{(H, H)\}$$

in question is

$$\mathrm{P}(A) = \frac{1}{4}.$$

However, we can find this probability using independence. Let A_1 and A_2 are the events that we get heads with the first and with the second tossing, respectively. We can assume that these two events are independent. And we know the probabilities

$$\mathrm{P}(A_1) = \mathrm{P}(A_2) = \frac{1}{2},$$

hence using the definition of independence, we get

$$\mathrm{P}(A) = \mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\,\mathrm{P}(A_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

□

Even in the simplest case of tossing a coin twice we have events that are not the product of two events related to the single coin tossing experiment. For example if

$$A = \text{"At least one coin toss result is heads."}$$

then A can't be written as a product. However we can write A as a union of mutually exclusive products, since

$A =$ "The first toss is heads, but the second is tails."
or "The first toss result is tails, but the second is heads."
or "Both the first and second toss is heads."

hence using the additivity and independence we get

$$\begin{aligned}
\mathrm{P}(A) &= \mathrm{P}\left((A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2) \cup (A_1 \cap A_2)\right) \\
&= \mathrm{P}(A_1 \cap \overline{A_2})) + \mathrm{P}(\overline{A_1} \cap A_2) + \mathrm{P}(A_1 \cap A_2) \\
&= \mathrm{P}(A_1)\,\mathrm{P}(\overline{A_2}) + \mathrm{P}(\overline{A_1})\,\mathrm{P}(A_2) + \mathrm{P}(A_1)\,\mathrm{P}(A_2) \\
&= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}.
\end{aligned}$$

Indeed, we can solve this problem in this simple experiment easier using

$$A = \{(T, H), (H, T), (H, H)\}$$

hence $\mathrm{P}(A) = \frac{3}{4}$, however in many more complex cases, the previous idea is more effective.

We can generalise independence to any number of events.

Definition 2.21 (Independence of n events)**.** We say the events $A_1, A_2, \ldots, A_n$ are independent if for any $1 \le k \le n$ and $1 \le i_1 < \ldots i_k \le n$ we have

$$\mathrm{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathrm{P}(A_{i_1}) \ldots \mathrm{P}(A_{i_k}).$$

We will learn another concept connected to the notion of indepence when we learn about conditional probability.

Further readings:

- `https://en.wikipedia.org/wiki/Birthday_problem`
- `https://en.wikipedia.org/wiki/Andrey_Kolmogorov`
- `https://en.wikipedia.org/wiki/Coin_flipping`

2.1 Exercises

Problem 2.1. We roll a dice 10 times, and we get the following result:
$\{2, 5, 6, 2, 1, 3, 4, 6, 1, 1\}$.
Let A be the event of rolling a prime number.

1. What is the relative frequency of the event A?
2. What is the probability of the event A?

Problem 2.2. We toss a coin 5 times. What are the probabilities of the following events?

A : The first coin lands on heads.

B : The first and last toss are the same.

C : We find an even number of heads.

D : We find exactly 3 heads.

E : We find more heads than tails.

F : No two consecutive tosses have the same result.

Problem 2.3. Find the answers for the previous problem with the modification that we toss the coin 3 times in stead of 5.

Problem 2.4. We roll a dice three times. What are the probabilities of the following events?

A : All dice shows 6.

B : The number 2 isn't rolled.

C : Each number is different.

D : At least two numbers are the same.

E : All numbers are odd.

F : The number 5 appears at least once.

Problem 2.5. Find the answers for the previous problem with the modification that we roll the dice 4 times in stead of 3.

Problem 2.6. We shuffle a deck of 52 French playing cards and draw the top 5 card. Find the probability of the following events.

A : We draw no cards with the suit spades.

B : We only draw cards with the suit spades.

C : We draw at least one card with the suit spades.

D : We have the Queen of Hearts in our hand.

E : We have all the kings.

F : We have exactly 2 kings and 2 queens.

The deck of French cards is made up of thirteen ranks (2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King, Ace) of each of the four French suits; clubs, diamonds, hearts and spades.

Problem 2.7. We shuffle a deck of 52 French playing cards and draw the top 5 card. Find the probability of the following poker hands.

A : Royal flush (10, J, Q, K, A of the same suit)

B : Straight flush (5 cards of the same suit in increasing order)

C : Four of a kind (4 of the same rank of cards)

D : Full house (3 of a kind and a pair)

E : Flush (5 cards of the same suit, not in order)

F : Straight (5 cards in increasing order, not of the same suit)

G : Three of a kind (3 of the same rank cards)

H : Two pairs

I : A pair

J : High card (none of the above)

Problem 2.8. The class has 43 members. 11 of these people will go to the wine festival this week, and 30 people will write a perfect test next week. 7 people belong to the each of these groups. We choose somebody randomly. What are the probabilities of the following events?

A : Go to the wine festival, but write the test not perfectly.

B : Go to the wine festival or write the test perfectly. (Both of these events can occur.)

C : Go to the wine festival or write the test perfectly. (Both of these events can not occur.)

Problem 2.9. Let's consider the probability space related to rolling two dice. What is the probability of the following events?

(a) Both numbers are 6.

(b) The first dice shows 1 and the second dice shows 2.

(c) We see the numbers 1 and 2.

The final answers to these problems can be found in section 10.

3 Discrete random variables I. - distribution

So far we always constructed an entire probability space to solve the exercises, but it is not always necessary. In this chapter we introduce random variables, which are numbers assigned to each outcome of a given random experiment, therefore producing a random number. Their power lies in the fact that we do not have to construct an entire probability space to describe them, we need only something called the distribution of the random variable.

Definition 3.1 (Discrete random variable)**.** Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space related to a random experiment, and $X : \Omega \to \mathbb{Z}$ an integer-valued function on the sample space. If $\{\omega \in \Omega : X(\omega) = k\} \in \mathcal{A}$ for all $k \in \mathbb{Z}$ the X is a *discrete random variable*.

Namely, a random variable is discrete if the possible values of it are integer numbers. The word *discrete* comes from the fact that the cardinality of the set of the integer numbers ($\mathbb{Z}$) is countable.

The condition $\{\omega \in \Omega : X(\omega) = k\} \in \mathcal{A}$ for all $k \in \mathbb{Z}$ in the definition above implies that we can investigate the probabilities like

$$\mathrm{P}(\{\omega \in \Omega : X(\omega) = k\}), \quad \mathrm{P}(\{\omega \in \Omega : X(\omega) \geq k\}, \quad \mathrm{P}(\{\omega \in \Omega : k \leq X(\omega) \leq l\}).$$

For example the last one is the probability of the event that the random variable X is between the integers k and l. Usually we use the following shorter notations

$$\mathrm{P}(X = k), \quad \mathrm{P}(X \geq k), \quad \mathrm{P}(k \leq X \leq l).$$

Definition 3.2 (Range)**.** The set of the possible values of a random variable is called the *range* of the random variable.

Proposition 3.3 (Discrete random variables on finite probability spaces)**.** *If* $|\Omega| < \infty$ *and* $\mathcal{A} = 2^\Omega$ *then any integer-valued function from* Ω *to* $\mathbb{Z}$ *is a discrete random variable.*

We can define an object which can help us calculate the probabilities connected to discrete random variables.

Definition 3.4 (Distribution of a discrete random variable)**.** By the *(probability) distribution* of a discrete random variable X we mean the probabilities

$$p_k := \mathrm{P}(X = k), \quad k \in \mathbb{Z}.$$

Theorem 3.5 (Properties of the distribution of a discrete random variable)**.** *For any discrete random variable* X *with distribution* p_k *the following are valid.*

(i) $p_k \geq 0$ *for all* $k \in \mathbb{Z}$,

(ii) $\sum_{k \in \mathbb{Z}} p_k = 1$.

Due to this theorem we can imagine the probability distribution as a mass distribution. We have unit mass (e.g. 1 kg sugar cubes) and we distribute this mass on the possible values, and the amount of mass in each value represents the probability that the random variable equals to this value.

Example 3.6. Let's consider the probability space related to rolling two dices. Then all of the following are discrete random variables:

- $X :=$ the number shown on the first dice,
- $Y :=$ the number shown on the second dice,
- $X + Y$,
- $\max\{X, Y\}$, $\min\{X, Y\}$,
- $(X - Y)^2$.

Example 3.7. What is the distribution of $Z := X + Y$ in the previous example?

Answer: Since the smallest number on a dice is 1 and we roll twice the sum is at least 2, and the greatest number on a dice is 6 therefore the sum is at most 12,

$$Z \in \{2, 3, \ldots, 11, 12\}.$$

We are going to list all the possible values of the sum, the probability of each value and the outcomes that produce that value

k	$\mathrm{P}(Z = k)$	outcomes
2	$\frac{1}{36} \approx 0.028$	$(1,1)$
3	$\frac{2}{36} \approx 0.056$	$(1,2), (2,1)$
4	$\frac{3}{36} \approx 0.083$	$(1,3), (2,2), (3,1)$
5	$\frac{4}{36} \approx 0.111$	$(1,4), (2,3), (3,2), (4,1)$
6	$\frac{5}{36} \approx 0.139$	$(1,5), (2,4), (3,3), (4,2), (5,1)$
7	$\frac{6}{36} \approx 0.166$	$(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)$
8	$\frac{5}{36} \approx 0.139$	$(2,6), (3,5), (4,4), (5,3), (6,2)$
9	$\frac{4}{36} \approx 0.111$	$(3,6), (4,5), (5,4), (6,3)$
10	$\frac{3}{36} \approx 0.083$	$(4,6), (5,5), (6,4)$
11	$\frac{2}{36} \approx 0.056$	$(5,6), (6,5)$
12	$\frac{1}{36} \approx 0.028$	$(6,6)$

In the case, when we have a finite number of possible values, we can represent the distribution in a table:

k	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

or graphically in a probability histogram (Figure 1).

In general we can add the distribution as a function of k. In this case the function p_k is sometimes called by the probability mass function.

$$p_k = \begin{cases} \frac{k-1}{36}, & k = 2, \ldots, 7, \\ \frac{13-k}{36}, & k = 8, \ldots, 12. \end{cases}$$

□

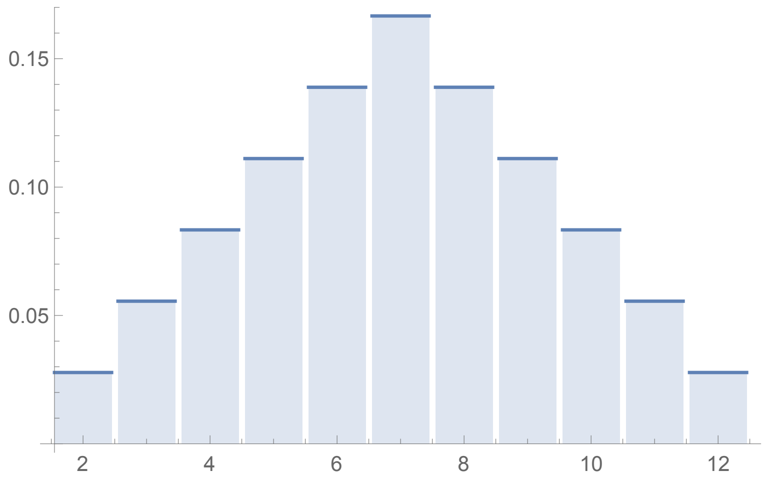


Figure 1: Probability distribution of the sum of two dice rollings.

Proposition 3.8 (Calculating probabilities using the distribution). *Let X be a discrete random variable with distribution p_k, and let $a < b$ are integers. Then*

$$P(a \leq X \leq b) = \sum_{k=a}^{b} p_k,$$
$$P(a \leq X) = \sum_{k=a}^{\infty} p_k,$$
$$P(X \leq b) = \sum_{k=-\infty}^{b} p_k.$$

Definition 3.9 (Independence of discrete random variables). The discrete random variables X and Y are called independent if for any $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$, the events $\{X = x\}$ and $\{Y = y\}$ are independent.

Further readings:

- `https://en.wikipedia.org/wiki/Cardinality`
- `https://www.wolframalpha.com/input/?i=histogram`

3.1 Exercises

Problem 3.1. Let's consider the probability space related to rolling two dice. Find the distribution of the following discrete random variables.

(a) $X :=$ the number shown on the first dice,

(b) $Y :=$ the number shown on the second dice,

(c) $X + Y$,

(d) $\max\{X, Y\}$,

(e) $\min\{X, Y\}$,

(f) $(X - Y)^2$

Problem 3.2. We take a dice and change the numbers 2 and 3 to show 5. What is the distribution of a number generated by rolling this modified dice?

Problem 3.3. We toss a coin 20 times. What is the distribution of the number of tails shown?

Problem 3.4. Three friends, Chandler, Joey and Ross order 3 different pizzas. When the pizzas are delivered, they are handed out randomly between the three Friends. Denote by X the number of the Friends, who get the pizza they want. What is the distribution of X?

Problem 3.5. A bag contains 5 green and 7 yellow balls. We pull a ball out of the bag, note its colour and put them back. We repeat this process 5 times. Find the distribution of the number of yellow balls drawn. Would it change anything if we didn't put the balls back in to the bag?

Problem 3.6. We toss a coin. If the result is heads, then we toss the coin once more, else we toss the coin two more times. Denote by X the number of heads shown. What is the distribution of X?

Problem 3.7. There are 3 machines in a factory, which are working at a given time with probability 0.5, 0.6 and 0.7, respectively. Denote by X the number of working machines. What is the distribution of X?

Problem 3.8. We play the following game. We roll the dice, and if the result is greater than 3, we win HUF 1,000. Furthermore, we can roll the dice again. If the result of the second rolling is greater than 4, we win HUF 2,000, additionally, and we can roll again. If the result of the third rolling is 6, we win HUF 6,000, additionally, and the game is over. Denote by X the amount of money earned. What is the distribution of X?

The final answers to these problems can be found in section 10.

4 Discrete random variables II. - expectation, notable distributions

In this part we define some special or notable distributions. After that we investigate a crucial object called expectation.

Notable discrete distributions

Definition 4.1 (Bernoulli distribution)**.** Let us consider a random experiment, let A be an event whose probability is $\mathrm{P}(A) = p$, and let

$$I_A := \begin{cases} 1, & \text{if } A \text{ happens,} \\ 0, & \text{otherwise.} \end{cases}$$

Then I_A is the indicator of the event A, it has range $\{0, 1\}$, and

$$p_0 = \mathrm{P}(I_A = 0) = 1 - p, \quad p_1 = \mathrm{P}(I_A = 1) = p.$$

Introduce the notation $I_A \sim \mathrm{Bernoulli}(p)$ for the *Bernoulli distribution* with parameter p.

Example 4.2 (Motivational example: number of successes out of a fixed number of repetitions)**.** Let's consider the random experiment of rolling a fair dice. Let $n \in \mathbb{N}$ be fixed, and repeat the experiment n times independently. Let X denote the number of dices showing 3 or 5. Find the distribution of this discrete random variable.

Answer: The variable X can take the following values: $0, 1, \ldots, n-1, n$. Its distribution is

$$\mathrm{P}(X = k) = \frac{\binom{n}{k} 2^k 4^{n-k}}{6^n} = \binom{n}{k} \left(\frac{2}{6}\right)^k \left(\frac{4}{6}\right)^{n-k} = \binom{n}{k} \left(\frac{2}{6}\right)^k \left(1 - \frac{2}{6}\right)^{n-k}$$

for any $k \in \{0, 1, \ldots, n\}$. □

Definition 4.3 (Binomial distribution)**.** Let us consider a random experiment, and let A be an event whose probability is $\mathrm{P}(A) = p$. Repeat the random experiment n times independently, then let X be the number of times the event A occured. Then X has *binomial distribution* with parameters n and p and

$$p_k = \mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0, 1, \ldots, n\}.$$

Introduce the notation $X \sim \mathrm{binom}(n, p)$ for the binomial distribution with parameters n and p.

Hence we can observe, that the binomial distribution describes the probability of k successes in n draws *with* replacement, or *sampling with replacement*. We are drawing with replacement, so it is good to see, that the draws are independent. Another remark is the connection between the binomial and the Bernoulli distribution. We can see, that if X is a binomial distributed variable with parameters n and p, than it has the following representation

$$X = I_1 + \cdots + I_n,$$

where I_i is the indicator of the event that the ith draw is a success, hence $I_i \sim \mathrm{Bernoulli}(p)$, $i = 1, \ldots, n$, and these random variables are independent.

Example 4.4. We have to take a test that consists of 10 questions, where you have to choose between 4 possible answers. Only one of the answers is correct. Suppose we fill out the test randomly. We get the best grade if we answer at least 9 questions correctly, what is the probability of that?

Answer: Let X denote the number of correct answers. Then X has the binomial distribution with parameters $n = 10$ and $p = 0.25$, that is

$$P(X = k) = \binom{10}{k} 0.25^k 0.75^{10-k}, \qquad k \in \{0, 1, \dots, 10\}.$$

Using this formula we get

$$\begin{aligned} P(\text{we get the best grade}) &= P(X \geq 9) = P(X = 9) + P(X = 10) \\ &= \binom{10}{9} 0.25^9 0.75 + \binom{10}{10} 0.25^{10} 0.75^0 \\ &\approx 0.0000296. \end{aligned}$$

□

Example 4.5 (Motivational example: sampling without replacement)**.** Let's consider the following random experiment. We have a bag with 5 green and 7 red balls and pull 3 balls out. Let X denote the number of green balls drawn. Find the distribution of this discrete random variable.

Answer: The variable X can take the following values: $0, 1, 2, 3$. Its distribution is

$$P(X = k) = \frac{\binom{5}{k}\binom{7}{3-k}}{\binom{12}{3}}, \qquad k \in \{0, 1, 2, 3\}.$$

□

Definition 4.6 (Hypergeometric distribution)**.** Let us consider the following random experiment. We have a bag with N balls, K green and $N - K$ red balls and pull n balls out. Let X denote the number of green balls drawn. Then X has *hypergeometric distribution* with parameters N, K, n and

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \qquad k \in \{0, 1, \dots, n\}.$$

Introduce the notation $X \sim \text{hypergeo}(N, K, n)$ for the hypergeometric distribution with parameters N, K, n.

That is, the hypergeometric distribution describes the probability of k successes in n draws *without* replacement, or *sampling without replacement*. The draws are not independent in this case. However, we have the representation

$$X = I_1 + \cdots + I_n,$$

where I_i is the indicator of the event that the ith draw is a success, hence $I_i \sim \text{Bernoulli}(p_i)$, $i = 1, \dots, n$, with some parameters p_i.

Example 4.7. We fill out a lottery ticket (5-of-90 lottery). Let X denote the number of correctly guessed numbers. Find the distribution of this discrete random variable. What is the probability of winning some money?

Answer: X has hypergeometric distribution with parameters $N = 90$, $K = 5$ and $n = 5$, that is

$$\mathrm{P}(X = k) = \frac{\binom{5}{k}\binom{85}{5-k}}{\binom{90}{5}}, \qquad k \in \{0, 1, 2, 3, 4, 5\}.$$

$$\begin{aligned}\mathrm{P}(\text{winnig some money}) &= \mathrm{P}(X \geq 2) = 1 - \mathrm{P}(X < 2) = 1 - \big(\mathrm{P}(X = 0) + \mathrm{P}(X = 1)\big) \\ &= 1 - \left(\frac{\binom{5}{0}\binom{85}{5}}{\binom{90}{5}} + \frac{\binom{5}{1}\binom{85}{4}}{\binom{90}{5}}\right) \approx 1 - (0.7464 + 0.2304) = 0.0232.\end{aligned}$$

□

Example 4.8 (Motivational example: number of repetitions until the first success)**.** Let's consider the random experiment of rolling a fair dice. Let X denote the number of times we have to roll the dice until we see 3 or 5 as the result. Find the distribution of this discrete random variable.

Answer: The random variable X can take any positive integer as a value. Its distribution is

$$\mathrm{P}(X = k) = \frac{4^{k-1}}{6^k} = \left(\frac{4}{6}\right)^{k-1}\frac{2}{6}$$

for any $k \in \mathbb{N} = \{1, 2, 3, \ldots\}$. □

Definition 4.9 (Geometric distribution)**.** Let us consider a random experiment, and let A be an event whose probability is $\mathrm{P}(A) = p$. Repeat the random experiment until the first occurrence of A and let X be the number of repetitions necessary.
Then X has *geometric distribution* with parameter p and

$$p_k := \mathrm{P}(X = k) = (1 - p)^{k-1}p, \qquad k \in \mathbb{N} = \{1, 2, 3, \ldots\}.$$

Introduce the notation $X \sim \mathrm{geom}(p)$ for the geometric distribution with parameter p.

This is our first distribution, which has infinite many possible values. We can check that this distribution is well-defined, namely the equation

$$\sum_{k=1}^{\infty} p_k = 1$$

holds or not? (See, Theorem 3.5.) To answer this question we need the following results.

Proposition 4.10 (Geometric series)**.** *For any $p \in (0, 1)$, we have*

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1 - p}.$$

If we use this result, we get

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty}(1 - p)^{k-1}p = p\sum_{k=1}^{\infty}(1 - p)^{k-1} = p\sum_{l=0}^{\infty}(1 - p)^l = p\frac{1}{1 - (1 - p)} = 1,$$

hence the geometric distribution is a well-defined distribution. □

Example 4.11. We play darts. The chance of us hitting the bullseye (the center of the target, that is worth 50 points) is 5%. We keep trying until we hit it. What is the probability of us succeeding in at most 2 tries?

Answer: Let X denote the number of attempts we need. Then X has a geometric distribution with parameter $p = 0.05$, that is

$$\mathrm{P}(X = k) = 0.95^{k-1}0.05, \quad k \in \{1, 2, \dots\}.$$

Using this formula we get

$$\begin{aligned}\mathrm{P}(\text{we need at most 2 tries}) &= \mathrm{P}(X \le 2) \\ &= \mathrm{P}(X = 1) + \mathrm{P}(X = 2) \\ &= 0.05 + 0.95 \cdot 0.05 \\ &= 0.0975.\end{aligned}$$

□

Expectation of a discrete random variable

In this part our aim is to introduce our first descriptive quantity about a random variable, namely the *expectation or mean*. We will do so by examining a motivating example.

Example 4.12 (Motivational example)**.** Alice and Bob play the following game. Alice rolls a dice and Bob pays Alice X\$, where X is the number shown on the dice. How much should Alice pay Bob for a chance to play this game?

Answer: In each round Alice gets somewhere between 1\$ and 6\$ from Bob. Clearly if Alice pays less than 1\$ per game, then she wins some amount of money each round, while if she pays more than 6\$, then she loses some money every round. So the fair price of this game is somewhere between 1\$ and 6\$.

We can apply the following idea: let Alice play n games and find her average gain, if this average has a limit as $n \to \infty$ then let that be the fair price of the game. Let $X, X_1, X_2, \dots, X_n$ denote independent dice rolls. Then Alice's average gain after n games is

$$A_n := \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We can regroup these games by the amount of money Bob has paid in them and get that Alice's average gain is

$$A_n = \frac{1}{n}\sum_{k=1}^{6} k k_n(X = k) = \sum_{k=1}^{6} k\frac{k_n(X = k)}{n} = \sum_{k=1}^{6} k r_n(X = k),$$

where $r_n(X = k)$ is the relative frequency of the event, that the dice shows k. We would like to interpret the probability of an event by the limit of relative frequencies so if A_n has a limit then it is

$$\sum_{k=1}^{6} k\,\mathrm{P}(X = k) = \sum_{k=1}^{6} k\frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.5.$$

So the fair price of this game is 3.5\$ for each round. We will define the expected value based on this fair price approach. □

Definition 4.13 (Expectation of a discrete random variable)**.** Let X be a discrete random variable with distribution $p_k = \mathrm{P}(X = k)$. If the sum

$$\mathrm{E}(X) := \sum_{k\in\mathbb{Z}} k\ p_k$$

exists, then we call it the *expectation (or mean)* of X.

There are random variables for which the expectation is not defined, since it can happen that $\sum_{k\in\mathbb{Z}} |k|\, p_k = \infty$. However, if a discrete random variable X is bounded, then its expectation always exists and is finite.

Example 4.14 (Dice roll)**.** Let X denote the result of a fair dice roll. Find $\mathrm{E}(X)$.

Answer: The distribution of X is

k	1	2	3	4	5	6
$\mathrm{P}(X = k)$	1/6	1/6	1/6	1/6	1/6	1/6

Hence the expectation of X is

$$\mathrm{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

Indeed, this is the same as in the motivational example above. □

The expectation of a random variable is a number that indicates the expected or averaged value of the random variable. It means that if we take a lot of independent copies of the random variable, then the average of these numbers oscillates around some number, which is the expectation. This is the so–called law of large numbers.

Theorem 4.15 (Kolmogorov's strong law of large numbers (1933))**.** *Let $X_1, X_2, \dots$ be independent and identically distributed random variables whose first absolute moment is finite, that is, $\mathrm{E}(|X_1|) < \infty$. Then*

$$\mathrm{P}\left(\lim_{n\to\infty} \frac{X_1 + \dots + X_n}{n} = \mathrm{E}(X)\right) = 1.$$

Thus the law of large numbers means that the average of the independent results of some experiment equals to the expectation. That is the heuristic meaning of the expectation. Furthermore, we can use the law of large numbers to prove our initial goal, that is to show that the relative frequencies of an event converge to the probability of the same event.

Theorem 4.16 (Law of large numbers and relative frequency)**.** *The relative frequency of an event converges to the probability of the event in question with probability 1.*

Proof. Indeed, let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $A \in \mathcal{A}$ an event whose probability is $\mathrm{P}(A) = p \in [0, 1]$. Repeat the experiment, related to the event A, n times independently and let

$$I_k := \begin{cases} 1, & \text{if } A \text{ happens during the } k\text{-th repetition,} \\ 0, & \text{otherwise,} \end{cases}$$

where $k \in \{1, \dots, n\}$. Then the relative frequency of the event A after n repetitions is the average of the random variables $I_1, \dots, I_n$, and, by the strong law of large numbers

$$r_n(A) = \frac{I_1 + \dots + I_n}{n} \to \mathrm{E}(I) = p, \quad n \to \infty$$

holds with probability 1. □

Let X be the gain of a game (the amount of money we win). Denote by C the entry fee of the game. In this case the profit is $X - C$. We have 3 cases.

- If $C = \mathrm{E}(X)$, then the game is fair in the sense that the long-run averaged profit is 0. Hence we can say that $\mathrm{E}(X)$ is the fair price of the game.
- If $C > \mathrm{E}(X)$, then the game is unfair and it is not favorable to us, because the long-run averaged profit is negative. We should not play this game.
- If $C < \mathrm{E}(X)$, then the game is unfair but it is favorable to us, the long-run averaged profit is positive. We should play it.

Proposition 4.17 (Properties of expectation)**.** *Let X and Y be arbitrary random variables on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ whose expectation exists. Then*

(i) The expectation is linear, that is for any constants $a, b \in \mathbb{R}$ we have

$$\mathrm{E}(aX + bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y).$$

(ii) If the random variables X and Y are independent, then

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y).$$

(iii) Let $g : \mathbb{R} \to \mathbb{R}$ be a function. Then

$$\mathrm{E}(g(X)) = \sum_{k \in \mathbb{Z}} g(k)\,\mathrm{P}(X = k)$$

Investigate the expectation of the notable distribution which we already learned.

Theorem 4.18 (Expectation of Bernoulli distribution)**.** *Let $X \sim$ Bernoulli(p), then*

$$\mathrm{E}(X) = p.$$

Proof: We can use the definition of expectation and get

$$\mathrm{E}(X) = 0 \cdot \mathrm{P}(X = 0) + 1 \cdot \mathrm{P}(X = 1) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

□

Theorem 4.19 (Expectation of binomial distribution)**.** *Let $X \sim$ binom(n, p), then*

$$\mathrm{E}(X) = np.$$

Proof: We can use the definition of expectation and get

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=0}^{n} k \cdot \mathrm{P}(X = k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} = np \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \\
&= np \sum_{j=0}^{m} \mathrm{P}(Y = j) = np,
\end{aligned}$$

where $j = k - 1, m = n - 1$ and $Y \sim \mathrm{binom}(m, p)$.

However, there is an alternative way to find the expectation of a binomial distribution. Let $X \sim \text{binom}(n, p)$, then the distribution of X is the same as the distribution of

$$Y_1 + \cdots + Y_n,$$

where $Y_i \sim$ Bernoulli(p), i=1,...,n.

Using the linearity of the expectation and the expectation of the Bernoulli distribution, we get

$$\mathrm{E}(X) = \mathrm{E}(Y_1 + \cdots + Y_n) = \mathrm{E}(Y_1) + \cdots + \mathrm{E}(Y_n) = np.$$

□

Theorem 4.20 (Expectation of geometric distribution). *Let $X \sim$ geom(p), then*

$$\mathrm{E}(X) = \frac{1}{p}.$$

Proof: We can use the definition of expectation and get the result directly. However, as in the case of binomial, there is an alternative way again to find the expectation of the geometric distribution. We have to ask the question what if we fail or succeed on the first trial?

We succeed with probability p and if we do then $X = 1$. If we fail (with probability $1 - p$), then we can denote the remaining trials until the first success by Y. Note that Y has the same distribution as X and therefore has the same expectation. We arrive at the following equation

$$\begin{aligned} \mathrm{E}(X) &= p\,\mathrm{E}(1) + (1-p)\,\mathrm{E}(1+Y) = p + (1-p)\,\mathrm{E}(1+X) \\ &= p + (1-p)(1+\mathrm{E}(X)) = 1 + (1-p)\,\mathrm{E}(X) \end{aligned}$$

Hence

$$\begin{aligned} \mathrm{E}(X) - (1-p)\,\mathrm{E}(X) &= 1 \\ \mathrm{E}(X) &= \frac{1}{p}. \end{aligned}$$

For the precise proof, see the remark after Proposition 6.12

□

Theorem 4.21 (Expectation of hypergeometric distribution). *Let $X \sim$ hypergeo(N, K, n), then*

$$\mathrm{E}(X) = n\frac{K}{N}.$$

Proof: For the proof, as in the binomial and the geometric case as well, we have two possibilities. One can use some combinatorial identities, and then after some tedious calculations, the result can be derived.

The other way is similar as in the binomial case. The distribution of X is the same as the distribution of

$$Y_1 + \cdots + Y_n,$$

where $Y_i \sim$ Bernoulli(p_i), i=1,...,n, namely Y_i is the result of the ith trial. It can be shown (we will see it when we learn about conditional probability, see Example 6.10), that the distributions of Y_i are the same, and $p_i = \frac{K}{N}$. That is, using the linearity of the expectation and the expectation of the Bernoulli distribution, we get

$$\mathrm{E}(X) = \mathrm{E}(Y_1 + \cdots + Y_n) = \mathrm{E}(Y_1) + \cdots + \mathrm{E}(Y_n) = n\frac{K}{N}.$$

□

Further readings:

- https://en.wikipedia.org/wiki/List_of_probability_distributions
- https://en.wikipedia.org/wiki/Expected_value
- https://en.wikipedia.org/wiki/Law_of_large_numbers
- https://en.wikipedia.org/wiki/Geometric_series

4.1 Exercises

Problem 4.1. We get 6 lottery tickets as a birthday present; each ticket has a 40% probability of winning. What is the probability that we will have exactly 4 winning tickets? What is the expected number of winning tickets?

Problem 4.2. At a fair, we can play the following game: we throw a coin until we obtain heads, then we get 100 Ft times the number of throws. How much should we pay to play this game?

Problem 4.3. At a driving test, we pass with a probability of 15% each time. Each test costs HUF 10,000. What is the probability that we will pass the test exactly on the fifth try? What is the expected cost of the tests if we keep trying until we obtain a driver's license?

Problem 4.4. On a city road there are 5 traffic lights. If we have to stop for a light, we lose 10 seconds. Supposing that the lamps operate independently of each other and that there is a 60% chance of having to stop for a light, what is the expected amount of delay on this road? What is the probability that we will be delayed exactly 30 seconds?

Problem 4.5. We throw two dices simultaneously. If the sum of the numbers is 3, we get HUF 100, if the sum is 7, we get HUF 30. How much should we pay to play this game?

Problem 4.6. In the 5-of-90 lottery the winnings are: 500,000,000 for 5 hits, 2,000,000 for 4 hits, 300,000 for 3 hits and 2,000 for 2 hits. What is the expectation of our winnings?

Problem 4.7. We can play the following game. We roll with a dice once and we can find the amount of money we win in this table.

result	1	2	3	4	5	6
prize	0	0	0	250	250	1000

What is the fair price for this game?

Problem 4.8. We can play the following game. We roll with a dice three times and we win HUF 54,000 if we roll only sixes, otherwise we win nothing. What is the fair price for this game?

Problem 4.9. In a video game there is a very difficult map. We only have 0.17 probability of completing the map successfully. If we fail the map, we can try again as many times as we wish. Each attempt takes 10 minutes. What is the expected number of attempts needed to complete the map? What is the expected time to finish the map? We only have an hour to play, what is the chance of success during that time?

Problem 4.10. We run a cinema, tonight is the premier of the new Star Wars movie, and we sold all 100 tickets. However we have a problem. We only have enough popcorn for 35 servings. Assume that each person buys popcorn for the movie with a probability 0.2 independently of each other. What is the probability that everyone gets popcorn who wants to buy it?

Problem 4.11. We toss a fair coin 3 times. If all tosses result in the same outcome, then we have to pay HUF 32 , if we get exactly 2 heads, we win HUF 64 and finally if we get exactly 2 tails we win HUF 16. What is the fair price of this game? How would this price change if we would exchange our fair coin with a biased one, that lands on heads only 1/4 of the time?

Problem 4.12. * A drunk sailor comes out of a pub. He is so drunk that every minute he picks a random direction (up the street or down the street) with equal probability. What is the probability that he will be back at the pub after 10 minutes? What is the probability of coming back after 20 minutes? What if he prefers to go up the street with probability 2/3?

Problem 4.13. * (Coupon collector's problem) The Leays company comes up with the following promotion. They put a card with one of the following colours into every bag of chips: red, yellow, blue, purple, green. Anyone who collects a card of each colour gets a mug with the Leays logo on it for free. The company hires us to investigate the effects of this promotion. What is the expected number of chips someone has to buy to collect one of each card? (Assume that a bag of chips contains any of the coloured cards with equal probability.)

The final answers to these problems can be found in section 10.

5 Discrete random variables III. - variance, covariance, correlation

Variance

Variance is the second descriptive quantity we introduce about a random variable. The expectation described a long term average behaviour of a random variable. But two random variables can have different distributions and still have the same expectation. We saw that a dice roll has expectation 3.5, let's construct another random variable with expectation 3.5.

Example 5.1 (Motivational example)**.** Take a fair coin, write 3 on one side and 4 on the other. Toss the coin once and let Y denote the number shown. Find the expectation of Y.

Answer: Let's find the distribution of Y first. Clearly Y has only 2 possible values 3 and 4 and they have the same probability so

$$\mathrm{P}(Y = 3) = \mathrm{P}(Y = 4) = \frac{1}{2}.$$

Then the expectation of Y is

$$\mathrm{E}(Y) = 3\,\mathrm{P}(Y = 3) + 4\,\mathrm{P}(Y = 4) = \frac{3}{2} + \frac{4}{2} = \frac{7}{2} = 3.5.$$

□

So a dice roll and our modified coin toss has the same expectation. However these random variables behave differently, the actual result of a dice roll can be further away from its expectation then our coin toss. We want variance to describe how much a random variable deviates from the expectation on average. However if we take the average difference from the expectation we get

$$\mathrm{E}\Big(X - \mathrm{E}(X)\Big) = 0$$

for all random variable X whose expectation exists. To get a meaningful quantity we square a difference and define the variance as the average squared difference from the expectation.

Definition 5.2 (Variance and standard deviation)**.** Let X be a random variable and suppose that $\mathrm{E}(X)$ exists and is finite. Then the *variance* of X is defined by

$$\mathrm{Var}(X) := \mathrm{E}\Big((X - \mathrm{E}(X))^2\Big).$$

The *standard deviation* of X is defined by $\mathsf{D}(X) := \sqrt{\mathrm{Var}(X)}$.

Using this definition we can calculate the variance of the dice roll and the modified coin toss. We expect that the variance of the dice roll will be greater than the modified coin toss. Indeed,

p_k	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
X	1	2	3	4	5	6
$X - 3.5$	-2.5	-1.5	-0.5	0.5	1.5	2.5
$(X - 3.5)^2$	6.25	2.25	0.25	0.25	2.25	6.25

p_k	$\frac{1}{2}$	$\frac{1}{2}$
Y	3	4
$Y - 3.5$	-0.5	0.5
$(Y - 3.5)^2$	0.25	0.25

hence

$$\mathrm{Var}(X) = \mathrm{E}\left((X-3.5)^2\right) = 6.25 \cdot \frac{1}{6} + \cdots + 6.25 \cdot \frac{1}{6} = 2.92,$$
$$\mathrm{Var}(Y) = \mathrm{E}\left((Y-3.5)^2\right) = 0.25 \cdot \frac{1}{2} + 0.25 \cdot \frac{1}{2} = 0.25.$$

Proposition 5.3 (Properties of variance)**.** *Let* X *and* Y *be random variables such that their variances exist and are finite. Then*

(i) $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$,

(ii) for any constants $c, d \in \mathbb{R}$, $\mathrm{Var}(cX+d) = c^2\,\mathrm{Var}(X)$,

(iii) $\mathrm{Var}(X) \geq 0$, *and* $\mathrm{Var}(X) = 0$ *if and only if* $\mathrm{P}(X = \mathrm{E}(X)) = 1$,

(iv) if X *and* Y *are independent, then* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Soon, we will discuss the variance of sums of random variables that are dependent, and a new concept, the covariance of random variables will be introduced.

Using the first property above and the properties of the expectation we can calculate variance simpler.

Example 5.4 (Variance of dice roll)**.** Let X denote the result of a fair dice roll. Find $\mathrm{Var}(X)$!

Answer: Recall that the expectation of X is $\mathrm{E}(X) = 3.5$ and the distribution of X is

k	1	2	3	4	5	6
$\mathrm{P}(X=k)$	1/6	1/6	1/6	1/6	1/6	1/6

The expectation of X^2 (also called the second moment of X) is

$$\mathrm{E}(X^2) = \frac{1}{6}\sum_{k=1}^{6} k^2 = \frac{91}{6}.$$

Then the variance of X is

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{91}{6} - (3.5)^2 = \frac{35}{12} \approx 2.92.$$

□

One can calculate the variances of the four notable discrete distributions that we have discussed.

Proposition 5.5 (Variance of the notable discrete distributions)**.**

(i) If $X \sim$ *Bernoulli*(p), *then* $\mathrm{Var}(X) = p(1-p)$.

(ii) If $X \sim$ *binom*(n,p), *then* $\mathrm{Var}(X) = np(1-p)$.

(iii) If $X \sim$ *hypergeo*(N,K,n), *then*
$$\mathrm{Var}(X) = n\frac{K}{N}\left(1-\frac{K}{N}\right)\left(1-\frac{n-1}{N-1}\right).$$

(iv) If $X \sim$ geom(p), then

$$\mathrm{Var}(X) = \frac{1-p}{p^2}.$$

If we want to attribute an informal meaning to variance as we did for the expectation with the fair price example, then we might think of it as a measure of risk. A high variance means that the random variable is capable of producing values far from the expectation with positive probability.

This is not a perfect measure of risk because it counts both big gains and big losses as a risk factor, while usually we only want to avoid big losses. There exists more refined measures of risk such as 'value at risk (VaR)' and 'expected shortfall (ES)', however we do not use them in this introductory material.

Example 5.6. We play a game, in which we roll a fair dice. If we get 1, then the game is over, our score is 1. Otherwise, we can decide to roll again or stop. Our score will be the last result of the rolling. How should you play this game to maximize your expected score?

Denote by X_i be the score using the strategy i. Let the strategy A is the following: roll only once. Then $\mathrm{E}(X_A) = 3.5$.

Let the strategy B is the following: if we roll 4,5 or 6, then we stop, if we roll 2 or 3, then roll again. Then

$$X_B \in \{1, 4, 5, 6\}$$

$$\begin{aligned} \mathrm{P}(X_B = 1) &= \mathrm{P}(\text{roll 1 in the 1th round}) + \mathrm{P}(\text{roll 1 in the 2nd round}) \\ &\quad + \mathrm{P}(\text{roll 1 in the 3rd round}) + \dots \\ &= \frac{1}{6} + \frac{2}{6} \cdot \frac{1}{6} + \left(\frac{2}{6}\right)^2 \frac{1}{6} + \dots = \frac{1}{6}\left(\sum_{i=0}^{\infty}\left(\frac{2}{6}\right)^i\right) \\ &= \frac{1}{6}\left(\frac{1}{1-\frac{2}{6}}\right) = \frac{1}{4}. \end{aligned}$$

Similarly, one can get

$$\mathrm{P}(X_B = 1) = \mathrm{P}(X_B = 4) = \mathrm{P}(X_B = 5) = \mathrm{P}(X_B = 6) = \frac{1}{4},$$

hence

$$\mathrm{E}(X_B) = \frac{1+4+5+6}{4} = 4.$$

Let the strategy C is the following: if we roll 3,4,5,6, then we stop, if we roll 2, then roll again.

$$X_C \in \{1, 3, 4, 5, 6\}$$

$$\begin{aligned} \mathrm{P}(X_C = 1) &= \mathrm{P}(\text{roll 1 in the 1th round}) + \mathrm{P}(\text{roll 1 in the 2nd round}) \\ &\quad + \mathrm{P}(\text{roll 1 in the 3rd round}) + \dots \\ &= \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \left(\frac{1}{6}\right)^2 \frac{1}{6} + \dots = \frac{1}{6}\left(\sum_{i=0}^{\infty}\left(\frac{1}{6}\right)^i\right) \\ &= \frac{1}{6}\left(\frac{1}{1-\frac{1}{6}}\right) = \frac{1}{5}. \end{aligned}$$

Similarly, one can get

$$P(X_C = 1) = P(X_C = 3) = P(X_C = 4) = P(X_C = 5) = P(X_C = 6) = \frac{1}{5},$$

hence

$$E(X_C) = \frac{1 + 3 + 4 + 5 + 6}{5} = 4.$$

So we have found that we have the same expectation for strategy B and C. The question is which should we choose. Naturally, one should choose the strategy with less risk. In general if we have two games (or strategies) with the same expected gain, then we choose the game with less risk. We use the variance to measure the risk, hence we say that the game with less variance has less risk.

We can calculate, that

$$\text{Var}(X_B) = 3.5 \qquad \text{Var}(X_C) = 1.4$$

thus we should choose the strategy C, because it is less risky then strategy B. □

Covariance, correlation

In this section we introduce a new object, which describes the dependence structure between two random variables.

Example 5.7 (Motivational example: resource management)**.** Suppose we assign values to the random variables X and Y based on a fair dice roll in the following way

dice roll	1	2	3	4	5	6
X	2	4	4	5	5	10
Y	9	6	6	4	4	1

Let X and Y represent the rewards of two betting games (related to the same fair dice roll).

- If you have enough money to bet on a single game, then which one (X or Y) should you bet on?
- If you have enough money to bet on two games, then which one (two X-s, or two Y-s, or one X and one Y) should you bet on?

We choose a strategy which has a higher expected reward, and in case of two strategies with the same expected rewards, we choose the one with less variance (risk).

Answer: Let's start by calculating the expectations (fair prices) for the games:

$$E(X) = \frac{2 + 4 + 4 + 5 + 5 + 10}{6} = \frac{30}{6} = 5,$$
$$E(Y) = \frac{9 + 6 + 6 + 4 + 4 + 1}{6} = \frac{30}{6} = 5.$$

So the two games have the same expected rewards. We should choose the one with less risk.

The second moments are

$$\mathrm{E}(X^2) = \frac{2^2+4^2+4^2+5^2+5^2+10^2}{6} = \frac{186}{6} = 31,$$
$$\mathrm{E}(Y^2) = \frac{9^2+6^2+6^2+4^2+4^2+1^2}{6} = \frac{186}{6} = 31.$$

Therefore $\mathrm{Var}(X) = \mathrm{Var}(Y) = 6$, that is, each game has the same risk. So, if we are playing only one game, it doesn't matter which one we choose.

Let's examine the case when we have enough money to bet on two games. We have 3 options: two X-s, or two Y-s, or one X and one Y. The rewards are the following:

dice roll	1	2	3	4	5	6
$2X$	4	8	8	10	10	20
$2Y$	18	12	12	8	8	2
$X+Y$	11	10	10	9	9	11

It is easy to see, that $\mathrm{E}(2X) = \mathrm{E}(2Y) = \mathrm{E}(X+Y) = 10$, and that $\mathrm{Var}(2X) = \mathrm{Var}(2Y) = 24$. The remaining question is what is the variance of $X+Y$, namely can we reduce the risk by splitting our bets?

$$\mathrm{E}\left((X+Y)^2\right) = \frac{11^2+10^2+10^2+9^2+9^2+11^2}{6} = \frac{604}{6} = \frac{302}{3}.$$

So the variance of $X+Y$ is

$$\mathrm{Var}(X+Y) = \mathrm{E}\left((X+Y)^2\right) - \mathrm{E}\left((X+Y)\right)^2 = \frac{2}{3} \approx 0.667.$$

Splitting our bets produces lower risk than doubling down on either game. It is because the random variables X and Y are not independent. Indeed, if they were independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 12$ would hold. They depend on each other in such a way that when X is big then Y is small and the other way around. □

Note also that $\mathrm{Var}(X+Y) - (\mathrm{Var}(X) + \mathrm{Var}(Y)) = -\frac{34}{3}$. This difference is related to the *covariance* of X and Y that we introduce now.

Definition 5.8 (Covariance)**.** Let X and Y be two random variables on the same probability space such that $\mathrm{Var}(X) < \infty$ and $\mathrm{Var}(Y) < \infty$. Then the *covariance* of X and Y is

$$\mathrm{Cov}(X,Y) := \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))).$$

Proposition 5.9 (Properties of the covariance)**.** *If X, Y, and Z are random variables such that $\mathrm{Var}(X) < \infty$, $\mathrm{Var}(Y) < \infty$ and $\mathrm{Var}(Z) < \infty$, and $a, b \in \mathbb{R}$ are constants, then*

(i) $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$,

(ii) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$,

(iii) $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$,

(iv) $\mathrm{Cov}(aX+bY, Z) = a\,\mathrm{Cov}(X,Z) + b\,\mathrm{Cov}(Y,Z)$,

(v) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$,

(vi) $\mathrm{Var}(aX+bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y)$.

Using property (iii), one can calculate correlation simpler.

Example 5.10. Let us consider our previous example on resource management, and determine the covariance of X and Y.

Answer: We know the distribution of XY:

dice roll	1	2	3	4	5	6
XY	18	24	24	20	20	10

which yields that

$$\mathrm{E}(XY) = \frac{18+24+24+20+20+10}{6} = \frac{116}{6},$$

and hence

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = \frac{116}{6} - 5\cdot 5 = -\frac{34}{6} = -\frac{17}{3}.$$

□

The covariance of two random variable measures the level and the direction of the dependence of the two variable. To interpret this kind of measure, we introduce another related notion, the *correlation*.

Definition 5.11 (Correlation)**.** Let X and Y be two random variables. Then the *correlation* of X and Y is

$$\mathrm{corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathrm{corr}(X,Y)}{\mathsf{D}(X)\mathsf{D}(Y)}.$$

This quantity sometimes called by Pearson's correlation coefficient.

Covariance and correlation are very similar. One can say that correlation is a normalized version of covariance. For instance, if we work with random variables with unit variance, then $\mathrm{Cov}(X,Y) = \mathrm{corr}(X,Y)$. Due to the similarity, we have similar properties. One important property, which is the difference of covariance and correlation (and this is the reason for using correlation instead of covariance) is that

$$-1 \le \mathrm{corr}(X,Y) \le 1.$$

Proposition 5.12 (Properties of correlation)**.** *Let X, Y, and Z be random variables and a, b real numbers. Then*

(i) $\mathrm{Corr}(X,X) = 1$,

(ii) $\mathrm{Corr}(X,Y) = \mathrm{Corr}(Y,X)$,

(iii) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathsf{D}(X)\mathsf{D}(Y)\,\mathrm{Corr}(X,Y)$,

(iv) $\mathrm{Var}(aX+bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\mathsf{D}(X)\mathsf{D}(Y)\,\mathrm{corr}(X,Y)$.

We have a special case of dependence when there are not any dependence between the variables, namely if they are independent. We investigate this case now.

Definition 5.13 (Uncorrelated variables)**.** If $\mathrm{corr}(X,Y) = 0$, then we say that the random variables X and Y are *uncorrelated*.

Theorem 5.14 (Connection between independence and uncorrelation)**.** *If X and Y are independent, then they are uncorrelated, but the converse does not hold in general.*

Proof: By the definitions, we can see that X and Y are uncorrelated if and only if $\mathrm{Cov}(X,Y)=0$. If X and Y are independent, then using the properties of covariance and expectation, we get

$$\mathrm{corr}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)=\mathrm{E}(X)\,\mathrm{E}(Y)-\mathrm{E}(X)\,\mathrm{E}(Y)=0,$$

thus X and Y are uncorrelated.

Too see that two uncorrelated variables are not necessary independent, we add a counterexample. Let (X,Y) be a random vector uniformly distributed on the points $(-1,0)$, $(0,-1)$, $(0,1)$ and $(1,0)$, i.e.,

$$\mathrm{P}(X\!=\!-1,Y\!=\!0)=\mathrm{P}(X\!=\!0,Y\!=\!-1)=\mathrm{P}(X\!=\!0,Y\!=\!1)=\mathrm{P}(X\!=\!1,Y\!=\!0)=\frac{1}{4}.$$

Then $\mathrm{E}(X)=\mathrm{E}(Y)=0$ and $\mathrm{E}(XY)=0$ yielding that

$$\mathrm{Cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)=0,$$

i.e., X and Y uncorrelated, but

$$\mathrm{P}(X=-1)=\mathrm{P}(X=1)=\frac{1}{4},\quad \mathrm{P}(X=0)=\frac{1}{2},$$
$$\mathrm{P}(Y=-1)=\mathrm{P}(Y=1)=\frac{1}{4},\quad \mathrm{P}(Y=0)=\frac{1}{2},$$

hence X and Y are not independent, since for example

$$\mathrm{P}(X=1,\,Y=0)=\frac{1}{4},\qquad \mathrm{P}(X=1)\,\mathrm{P}(Y=0)=\frac{1}{8}.$$

□

To understand why measures correlation the level and the direction of the linear dependence of the related random variables, we investigate the following problem.

Example 5.15 (The Best Linear Predictor)**.** What linear function of X is closest to Y in the sense of minimizing the mean square error (second moment of the error)?

Thus we can imagine, that we can only observe the variable X, and using this observation, we have to estimate (or predict) the variable Y. Hence the task is the following: find the value of a and b to minimize $\mathrm{E}\left((Y-(aX+b))^2\right)$.

In the simplest case, if $\mathrm{E}(X)=\mathrm{E}(Y)=0$ and $\mathrm{Var}(X)=\mathrm{Var}(Y)=1$, we get

$$a=\mathrm{corr}(X,Y)\qquad b=0,$$

hence the best linear predictor of Y given X is $\mathrm{corr}(X,Y)X$.

Furthermore, the mean square error of the best linear predictor is

$$\mathrm{E}\left((Y-\mathrm{corr}(X,Y)X)^2\right)=1-(\mathrm{corr}(X,Y))^2.$$

In general one can show that the best linear predictor of Y given X is

$$\mathrm{E}(Y)+\frac{\mathsf{D}(Y)}{\mathsf{D}(X)}\,\mathrm{corr}(X,Y)(X-\mathrm{E}(X)),$$

and the mean square error of it is

$$\mathrm{Var}(Y)\left(1-(\mathrm{corr}(X,Y))^2\right).$$

Proposition 5.16. *With the correlation we can derive the direction of the linear dependence of the related random variables, namely*

(i) if $\operatorname{corr}(X, Y) = 0$ *(uncorrelated case), then does not exist linear dependence between* X *and* Y.

(ii) If $\operatorname{corr}(X, Y) > 0$ *(positively correlated case), then we have positive linear dependence between* X *and* Y*. If* X *is bigger, then* Y *is bigger too, and the other way around.*

(iii) If $\operatorname{corr}(X, Y) < 0$ *(negatively correlated case), then we have negative linear dependence between* X *and* Y*. If* X *is bigger, then* Y *is smaller, and the other way around.*

Proposition 5.17. *With the correlation we can derive the level of the linear dependence of the related random variables as well, namely*

(i) if $\operatorname{corr}(X, Y)$ *is closer and closer to 1 or -1, then the error is closer and closer to 0,*

(ii) if $\operatorname{corr}(X, Y) = 1$*, then we* $P(Y = aX + b) = 1$ *with some constant* $a > 0$ *and* b*,*

(iii) if $\operatorname{corr}(X, Y) = -1$*, then we* $P(Y = aX + b) = 1$ *with some constant* $a < 0$ *and* b*.*

A case study: Mean-Variance portfolio analysis

The task is the following. We have a large capital, 1 million dollars that we want to invest into stocks. On the market there are two kinds of stocks available let's label them A and B, and assume that they both cost 10$. How should you invest your money? What is the portfolio with the largest expected return? What is the portfolio with the lowest risk? Given a maximum acceptable level of risk what is the highest expected return we can reach?

To solve these kind of problems, we can use *Mean-Variance portfolio analysis*, which can be used for more complex markets as well.

On this simple market we can describe a portfolio, π_c by a constant $c \in [0, 1]$ that denotes the fraction of capital invested into stock A, then the rest is invested into stock B. For example if

$$\begin{aligned}
c =& 1, \text{ then } \pi_1 \text{ contains 100 000 pieces of stock } A \\
c =& 0.5, \text{ then } \pi_{0.5} \text{ contains 50 000 pieces of stock } A \\
& \quad \text{and 50 000 pieces of stock } B \\
c =& 0.9, \text{ then } \pi_{0.9} \text{ contains 90 000 pieces of stock } A \\
& \quad \text{and 10 000 pieces of stock } B \\
c =& 0, \text{ then } \pi_0 \text{ contains 100 000 pieces of stock } B
\end{aligned}$$

Let S_A and S_B denote the future price of stock A and B respectively. Suppose they have the following distribution

k	8$	12$	16$
$P(S_A = k)$	0.2	0.6	0.2

k	6$	12$	20$
$P(S_B = k)$	0.3	0.4	0.3

We can find that

$$\begin{aligned}
E(S_A) &= 8 \cdot 0.2 + 12 \cdot 0.6 + 16 \cdot 0.2 = 12 \\
E(S_B) &= 6 \cdot 0.3 + 12 \cdot 0.4 + 20 \cdot 0.3 = 12.6
\end{aligned}$$

and

$$\begin{aligned}\operatorname{Var}(S_A) &= \mathrm{E}(S_A^2) - (\mathrm{E}(S_A))^2 = 6.4\\ \operatorname{Var}(S_B) &= \mathrm{E}(S_B^2) - (\mathrm{E}(S_B))^2 = 29.64\end{aligned}$$

We can see that stock B has better expected return than A, but it also has larger risk (variance).

Let X_c denote the gain, then by the linearity of the expectation

$$\begin{aligned}\mathrm{E}(X_c) &= 100\ 000\, \mathrm{E}\big(cS_A + (1-c)S_B\big) - 1\ 000\ 000\\ &= 100\ 000(12c + 12.6(1-c)) - 1\ 000\ 000\\ &= 260\ 000 - 60\ 000c\end{aligned}$$

We can plot the expected gain as a function of c (see, Figure 2). We can see that to achieve

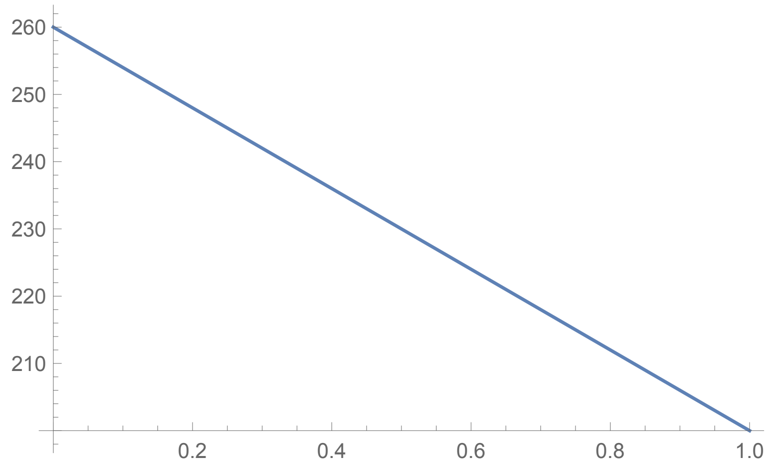


Figure 2: The expected gain of portfolio π_c in1000\$

the most expected return (\$260 000) we would have to invest all of our capital into stock B. However in this case the risk involved is

$$\operatorname{Var}(X_0) = \operatorname{Var}(100\ 000 S_B - 1\ 000\ 000) = 2.96 \times 10^{11},$$

while investing all our capital into stock A yields a lower expected return, \$200 000, but also a lower risk

$$\operatorname{Var}(X_1) = \operatorname{Var}(100\ 000 S_A - 1\ 000\ 000) = 6.4 \times 10^{10},$$

At this point we don't have enough information to find the portfolio with the lowest risk, that would require some description of how the prices of these two stocks are related to each other.

Case 1. Independent companies (uncorrelated case).

Let's assume that the companies' performance does not affect each other, that is the random variables S_A and S_B are independent (enough to assume to be uncorrelated). By the properties of variance we have

$$\begin{aligned}\operatorname{Var}(X_c) &= \operatorname{Var}\big(100\ 000(cS_A + (1-c)S_B) - 1\ 000\ 000\big)\\ &= 100\ 000^2\big(c^2 \operatorname{Var}(S_A) + (1-c)^2 \operatorname{Var}(S_B)\big)\\ &= 10^{10}\big(29.64 - 59.28c + 36.04c^2\big).\end{aligned}$$

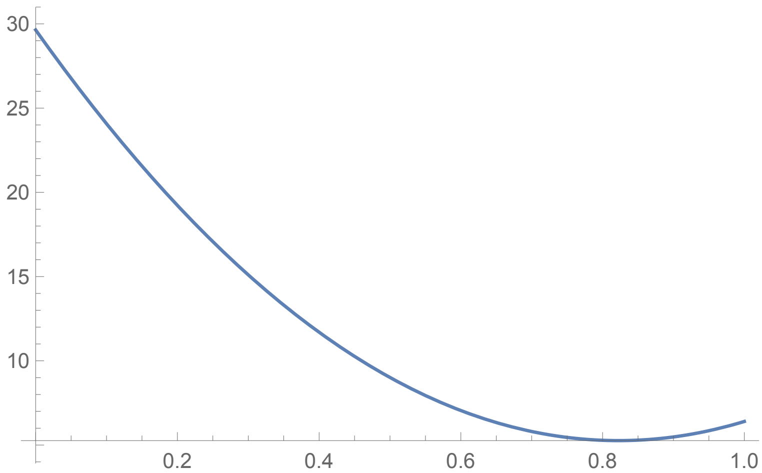


Figure 3: The risk of portfolio π_c in Case 1

Leaving out the constant multiplier 10^{10} we can plot the variance (see, Figure 3)

We can find the portfolio with minimal risk by finding the minima of the this function. It is at $c = 0.82242$ and then the expected gain and variance is

$$\mathrm{E}(X_{0.82242}) = 210\ 655, \qquad \mathrm{Var}(X_{0.82242}) = 5.26349 \times 10^{10}.$$

It is interesting to note that the least risky strategy is not the one where we invest all our capital into the least risky stock.

Case 2. Competing companies (negatively correlated case).

In this case we assume that the two stocks are related to two companies that compete for market share in the same sector. One could imagine Apple and Samsung both competing for larger smart phone sales. In this case if the value of a stock rises then the value of the other stock should decrease.

We assume that S_A and S_B are not indepenedent, and $\mathrm{corr}(S_A, S_B) = -0.81$. The expected gain is the same:

$$\begin{aligned}\mathrm{E}(X_c) &= 100\ 000\,\mathrm{E}\big(cS_A + (1-c)S_B\big) - 1\ 000\ 000 \\ &= 260\ 000 - 60\ 000c.\end{aligned}$$

However the variance changes:

$$\begin{aligned}\mathrm{Var}(X_c) &= \mathrm{Var}\big(100\ 000(cS_A + (1-c)S_B) - 1\ 000\ 000\big) \\ &= 10^{10}\,\mathrm{Var}(cS_A + (1-c)S_B) \\ &= 10^{10}\left(\mathrm{Var}(cS_A) + \mathrm{Var}((1-c)S_B) + 2\,\mathrm{Cov}(cS_A, (1-c)S_B)\right).\end{aligned}$$

By the properties of covariance we get

$$\mathrm{Var}(X_c) = 10^{10}(29.64 - 81.68c + 58.44c^2),$$

hence we can find the portfolio with minimal risk by finding the minima of the above function. It is at $c = 0.6988$ and then the expected gain and variance is

$$\mathrm{E}(X_{0.6988}) = 218\ 070, \qquad \mathrm{Var}(X_{0.6988}) = 1.09952 \times 10^{10}.$$

Case 3. Cooperating companies (positively correlated case).

In this case we assume that the two stocks are related to two companies that are in the same sector but instead of competing they are complementing each other. Continuing our example with smart phones if Apple sales more smart phones then those companies that produce applications for the iPhone are benefit from this. In this case if the value of a stock rises then the value of the other stock should rise too.

We assume that S_A and S_B are not independent, and $\mathrm{corr}(S_A, S_B) = 0.81$. The expected gain is the same:

$$\begin{aligned} \mathrm{E}(X_c) &= 100\ 000\,\mathrm{E}\big(cS_A + (1-c)S_B\big) - 1\ 000\ 000 \\ &= 260\ 000 - 60\ 000c \end{aligned}$$

By the properties of covariance we get

$$\mathrm{Var}(X_c) = 10^{10}(29.64 - 36.88c + 13.64c^2)$$

We can find the portfolio with minimal risk by finding the minima of the this function. It is at $c = 1$ and then the expected gain and variance is

$$\mathrm{E}(X_1) = 200\ 000, \qquad \mathrm{Var}(X_1) = 6.4 \times 10^{10}$$

In this case the safest portfolio is the one where we spend all of our capital on the least risky stock.

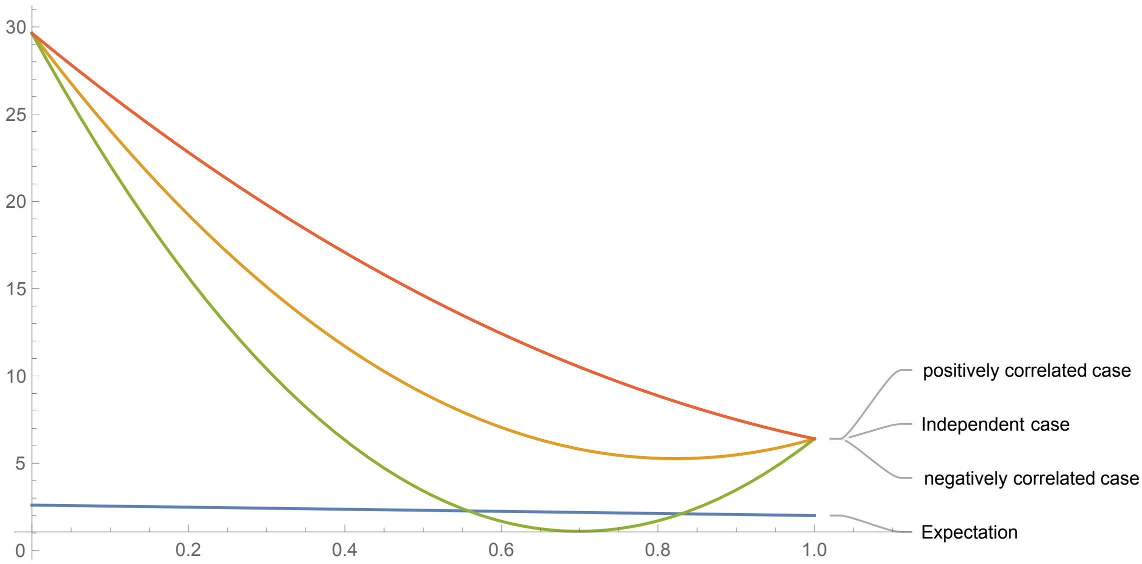


Figure 4: The expected gain and the variances in the three cases.

Finally, we can answer the last question as well, given a maximum acceptable level of risk what is the highest expected return we can reach? For example in the Case 2 (negatively correlated case):

If our maximum acceptable level of risk is 2, namely we are looking for a portfolio with variance 2×10^{10}, then we get $c = 0.574705$, hence $\mathrm{E}(X_{0.574705}) = 225\ 518$ (see, Figure 5).

Further readings:

- `https://en.wikipedia.org/wiki/Value_at_risk`

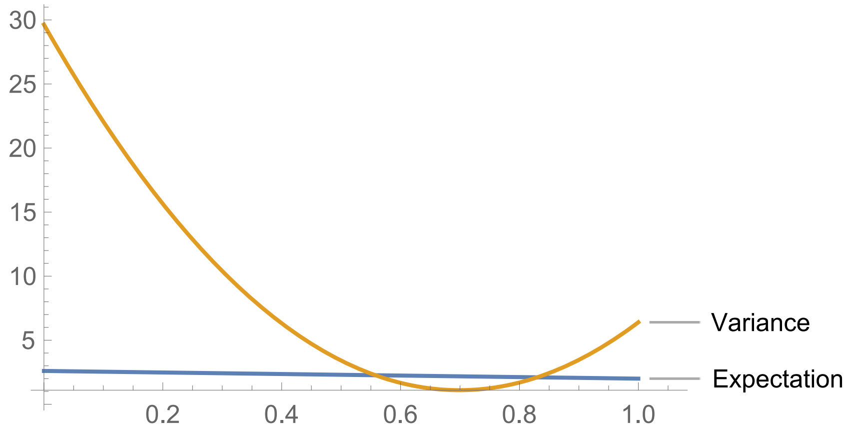


Figure 5: The expected gain and the variance in Case 2.

- https://en.wikipedia.org/wiki/Expected_shortfall
- https://en.wikipedia.org/wiki/Modern_portfolio_theory
- https://www.buzzfeednews.com/article/kjh2110/the-10-most-bizarre-correlations

5.1 Exercises

Problem 5.1. In a car factory they produce 1000 cars each week. They test the cars before shipping them out to the car dealers, 2% of the cars fail the test and are never shipped out. Find the expected number and the variance of faulty cars.

Problem 5.2. We keep rolling a 7 sided dice until we roll either 7 or a number less than or equal to 2. Find the expectation and variance of the number of trials necessary.

Problem 5.3. In a casino we can choose between the following two games. We roll two fair dices. In the first game we win HUF 18,000, if we roll two sixes. Otherwise we do not win anything. In the second game we win HUF 3,000, if we roll the same numbers, otherwise we do not win anything. Which game should be preferred?

Problem 5.4. We play a game in which we roll a fair dice. If we get 1, then the game is over, our score is 1. Otherwise, we can decide to roll again or stop. Our score will be the result of the last rolling. How should you play this game to maximize your expected score?

Problem 5.5. Suppose that the probabilities are 0.4, 0.3, 0.2, and 0.1, respectively, that 0, 1, 2, or 3 power failures will strike a certain subdivision in any given year. Find the expectation and variance of the random variable X representing the number of power failures striking this subdivision.

Problem 5.6. Let X be a discrete random variable such that $\mathrm{E}((X-1)^2) = 10$ and $\mathrm{E}((X-2)^2) = 5$. Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

Problem 5.7. Let X and Y represent the results of two independent dice rolling. Find the variance of the random variables $3X - Y$, and $X + 5Y - 5$.

Problem 5.8. Suppose we assign values to the random variables X and Y based on a fair dice roll in the following ways. Find the covariance and the correlation of X and Y in each cases.

Case (a)						
dice roll	1	2	3	4	5	6
X	1	2	3	4	5	6
Y	6	5	4	3	2	1

Case (b)						
dice roll	1	2	3	4	5	6
X	2	2	4	4	6	6
Y	1	3	5	7	9	11

Case (c)						
dice roll	1	2	3	4	5	6
X	1	1	3	3	3	6
Y	5	5	4	3	2	2

Case (d)						
dice roll	1	2	3	4	5	6
X	3	9	7	5	4	10
Y	1	7	5	4	10	2

Problem 5.9. * We roll two dice. Denote the results by Z_1 and Z_2 . Determine the covariance and the correlation of the random variables $X = Z_1 + Z_2$ and $Y = Z_1 \cdot Z_2$.

Problem 5.10. We investigate a company's profit in a month, which is the excess of revenue over cost. We assume that the revenue and the cost are two random variables. The expected revenue is HUF 120 million with standard deviation HUF 30 million. The expected cost is HUF 80 million with standard deviation HUF 20 million.

(a) What is the expectation and the standard deviation of the profit in the case when the revenue and the cost are independent?

(b) What is the expectation and the standard deviation of the profit in the case when the revenue and the cost are not independent and their correlation is 0.8?

Problem 5.11. * There are two stocks that you can buy for HUF 1,000 Ft each. A year later the first stock can have a price of HUF 800, HUF 1,200 or HUF 1,600 with probability 0.2, 0.6 and 0.2 respectively. The second stock can have a price of HUF 600, HUF 1,200 or HUF 2,000 with probability 0.3, 0.4 and 0.3 respectively. Let π_α denote the future value of the portfolio where we buy $\alpha \in [0, 1]$ from the first stock and $1 - \alpha$ from the second.

(a) Assume that the two stock prices are independent and find the expected return and risk of π_α.

(b) Assume that the prices are distributed in the following way

ω	1	2	3	4	5	6	7	8	9	10
$X_1(\omega)$	800	800	1200	1200	1200	1200	1200	1200	1600	1600
$Z_1(\omega)$	600	600	600	1200	1200	1200	1200	2000	2000	2000

Find the expected return and risk of π_α.

(c) Assume that the prices are distributed in the following way

ω	1	2	3	4	5	6	7	8	9	10
$X_2(\omega)$	800	800	1200	1200	1200	1200	1200	1200	1600	1600
$Z_2(\omega)$	2000	2000	2000	1200	1200	1200	1200	600	600	600

Find the expected return and risk of π_α.

The final answers to these problems can be found in section 10.

6 Conditional probability

In some cases, we have a background information about the outcome of a random experiment. In this situation the probability of an event can change.

Example 6.1 (Motivational example)**.** Someone rolls a dice. What is the probability that the outcome is odd if the only information we have is

(i) the outcome is a prime number, or

(ii) the outcome is less then 5?

Answer: The heuristic answers are the following. In the first case, we know that the outcome cannot be 1, 4 or 6, so the probability is 2/3. And in the second case, we know that the outcome must be 1,2,3 or 4, so the probability of getting an odd number is 1/2. □

This idea, which we used is the base of the following definition.

Definition 6.2 (Conditional probability)**.** The *conditional probability of the event* A *given the event* B (i.e., if we know that the event B has occurred):

$$\mathrm{P}(A \mid B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)},$$

provided that $\mathrm{P}(B) > 0$.

We can interpret conditional probability as a fraction (see, Figure 6)

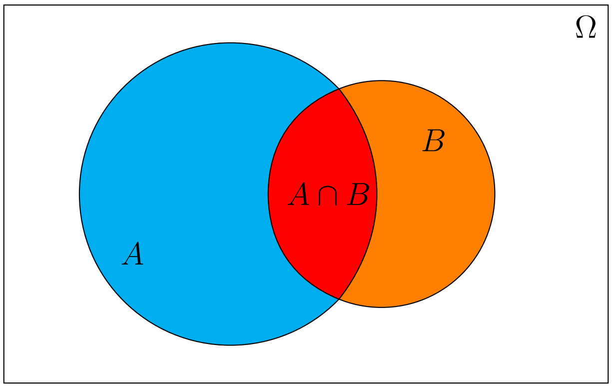


Figure 6: Conditional probability is a fraction.

Proposition 6.3 (Properties of the conditional probability)**.** *Let event* B *be fixed. The conditional probability* $\mathrm{P}(A \mid B)$ *is a probability. Consequently all the properties of the usual probability are valid in the conditional case.*

Namely, e.g.:

(i) $0 \le \mathrm{P}(A \mid B) \le 1$,

(ii) $\mathrm{P}(\overline{A} \mid B) = 1 - \mathrm{P}(A \mid B)$.

Theorem 6.4 (Connection with the independence)**.** *The following are equivalent.*

(i) A *and* B *are independent.*

(ii) $\mathrm{P}(A \mid B) = \mathrm{P}(A)$.

(iii) $\mathrm{P}(B \mid A) = \mathrm{P}(B)$.

Theorem 6.5 (Chain rule (product rule))**.** *For any event* $A_1, \ldots, A_n$

$$\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathrm{P}(A_1)\ \mathrm{P}(A_2 \mid A_1)\ \mathrm{P}(A_3 \mid A_1 \cap A_2) \cdots \mathrm{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}),$$

provided that $\mathrm{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$.

Proof: The right-hand side takes the form

$$\mathrm{P}(A_1)\, \frac{\mathrm{P}(A_1 \cap A_2)}{\mathrm{P}(A_1)}\, \frac{\mathrm{P}(A_1 \cap A_2 \cap A_3)}{\mathrm{P}(A_1 \cap A_2)} \cdots \frac{\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1} \cap A_n)}{\mathrm{P}(A_1 \cap A_2 \cdots \cap A_{n-1})}.$$

□

Example 6.6. A bag contains 5 green and 7 yellow balls. We pull a ball 3 times without replacement. What is the probability that the first ball is green, the second ball is yellow and the third ball is green?

Answer:
A_1 = the 1th ball is green.
A_2 = the 2nd ball is yellow.
A_3 = the 3rd ball is green.

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\,\mathrm{P}(A_2 \mid A_1)\,\mathrm{P}(A_3 \mid A_1 \cap A_2) = \frac{5}{12} \cdot \frac{7}{11} \cdot \frac{4}{10}.$$

□

Theorem 6.7 (Bayes formula)**.** *If* A *and* B *are events such that* $\mathrm{P}(A) > 0$ *and* $\mathrm{P}(B) > 0$, *then*

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

Proof: $\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$ and $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B \mid A)$. □

Definition 6.8 (Partition of Ω)**.** It is a countable decomposition of Ω into pairwise disjoint events,
i.e., it is a finite or infinite set of events $\{B_1, B_2, \ldots\}$ such that
they are pairwise disjoint and their union is Ω, i.e.,

$$B_i \cap B_j = \emptyset \quad \text{if} \quad i \neq j, \qquad \text{and} \qquad \bigcup_i B_i = \Omega.$$

An important remark is that if $\{B_1, B_2, \ldots\}$ is a partition of Ω, then exactly one of these events occurs. The simplest partition is an event with its complement.

Theorem 6.9 (Law of total probability)**.** *If* $\{B_1, B_2, \ldots\}$ *is a partition of* Ω *such that* $\mathrm{P}(B_i) > 0,\ i \in \mathbb{N}$, *then for any event* A, *we have*

$$\mathrm{P}(A) = \sum_i \mathrm{P}(A \mid B_i) \cdot \mathrm{P}(B_i).$$

Proof: Using the σ-additivity, we get

$$P(A) = \sum_i P(A \cap B_i),$$

because the events $A \cap B_i$ are disjoint. Then using the chain rule, we get the right-hand side. □

We have already used the Law of total probability, when we gave the expectation of the hypergeometric distribution.

Example 6.10. Lets investigate the case of the hypergeometric distribution, namely sampling without replacement. Take a bag with K green balls and $N - K$ red balls. We pull 2 balls out without replacement. The question is what is the probability that the second ball is green?

Answer: Let A be the event that the first ball is green and let B be the event that the second ball is green. Then $\{A, \overline{A}\}$ is a partition. Using the Law of total probability, we get

$$\begin{aligned} P(B) = P(B \mid A)\,P(A) + P(B \mid \overline{A})\,P(\overline{A}) &= \frac{K-1}{N-1} \cdot \frac{K}{N} + \frac{K}{N-1} \cdot \frac{N-K}{N} \\ &= \frac{K^2 - K + KN}{N(N-1)} = \frac{K(N-1)}{N(N-1)} = \frac{K}{N}. \end{aligned}$$

The same can be shown for the further pulls as well. □

Definition 6.11 (Conditional expectation)**.** Let A be an event with positive probability, and X be a discrete random variable. Then the *conditinal expectation of X given the event A* is

$$E(X \mid A) = \sum_k k\,P(X = k \mid A).$$

Proposition 6.12 (Law of total expecation)**.** *If $\{A_1, A_2, \ldots\}$ is a partition of Ω, then for any discrete random variable X, we have*

$$E(X) = \sum_i E(X \mid A_i) \cdot P(A_i).$$

We have already used the Law of total expectation when we gave the expectation of the geometric distribution. Indeed, let X be a geometric distributed random variable with parameter p, thus the number of trials needed until the first success. Let A be the event that the first trial is successful. Then $\{A, \overline{A}\}$ is a partition. Using the Law of total expectation, we get

$$E(X) = E(X \mid A)\,P(A) + E(X \mid \overline{A})\,P(\overline{A}) = 1 \cdot p + E(X \mid \overline{A})(1 - p).$$

Finally, as we have already discussed, the remaining trials until the first success, in the case when the first trial is failure, has the same distribution as X, hence $E(X \mid \overline{A}) = 1 + E(X)$.

Further readings:

- https://en.wikipedia.org/wiki/Law_of_total_probability
- https://en.wikipedia.org/wiki/Memorylessness
- https://en.wikipedia.org/wiki/Markov_chain

6.1 Exercises

Problem 6.1. In an exam you have to speak about 1 topic out of 10 possible topics. There are 4 easy and 6 difficult topics. One day there are 3 students, and they pull 1 topic with replacement.

(a) What is the probability that everybody pulls an easy topic?

(b) What is the probability that the first student pulls an easy and the third one pulls a difficult topic?

(c) What is the probability that exactly 2 students pull an easy topic?

(d) In the case when they pull the topics without replacement, which student has the greatest probability of pulling an easy topic?

Problem 6.2. There is a city, where the number of men and women are the same. The probability that a man is color-blind is 5%, and 2.5% for a women.

(a) What is the probability that a randomly chosen person is color-blind?

(b) What is the probability that a randomly chosen color-blind person is man?

Problem 6.3. You have to write a test, where for each question there are 3 possible answers, but only one is good. Assume that you know the proper answer with probability p. If you don't know the right answer, you randomly choose one of the answers.

(a) What is the probability that you choose the right answer?

(b) During the checking, the teacher sees a good answer. What is the probability that you knew it?

Problem 6.4. There is a packaging factory, where apples are packed. There are 4 producers who deliver apples to the factory. The fractions of the apples delivered by the producers are 10%, 30 %, 40% and 20%. We sort the apples to 2 class, first-class and second-class. For each producer 40%, 50%, 20% and 100% of the delivered quantity is first class.

(a) What is the probability that a randomly chosen apple is first class?

(b) What is the probability that a randomly chosen apple is delivered by the first producer, if we know that the apple is second-class?

Problem 6.5. There is a serious sickness. 1% of the people suffer from this disease. We have a test for it. The test has 99% confidence, which means that if the patient is sick, then the test will be positive with probability 99%, and if the patient is not sick, then the test will be negative with probability 99%. Assume that you test yourself and the result is positive. What is the probability that you are sick, indeed?

Problem 6.6. * We play the following game. We roll a dice and then toss as many coins as the result was at the dice rolling. We get as many points as the number of heads we get. What is the expected number of gained points?

Problem 6.7. * (Randomized response) We have to make a query, but there is a embarrassing question in it which might not be answered honestly even if the query is anonymous. For example have you ever make a crime, or do you sing in the shower.

The idea is the following. We change the question for this one: Toss a coin twice, and answer the question depending on the result:

two heads	Is that true that you sing in the shower?
anything else	Is that true that you do not sing in the shower?

Because the result of the coin tossing is secret for us, we cannot know the right answer to the original question.

Then we have 2000 people who answered this query, and we see 875 answer *yes*. Can we derive the fraction of people who sing in the shower?

The final answers to these problems can be found in section 10.

7 Continuous random variables

Distribution function

So far we only discussed random variables with countable (finite or countable infinite) many possible values, and we called it discrete random variables. In this section we introduce another kind of random variables, which are called *continuous random variables*. The most important feature is that the set of possible values of a random variable is not countable. Usually it is an interval.

Example 7.1 (Motivational example)**.** Choose a (real) number randomly in the interval $[0, 1]$ and denote it by X. What is the probability of X is less then $2/3$?

The problem is that X is not a discrete random variable, because the number of its possible values is not countable. (The cardinality of the interval $[0, 1]$ is not countable, but continuum.) Hence, the probability distribution cannot be defined, and so we cannot calculate probabilities like in the discrete case.

However, we can calculate the probability in question by introducing a special probability space.

Definition 7.2 (Geometric probability on an interval)**.** We choose a number in the interval $[a, b]$. Denote the interval $[a, b]$ by Ω. Then the probability space $(\Omega, \mathcal{A}, \mathrm{P})$ can be defined, where the probability of choosing from a subset $A \subset [a, b]$ is the length of A divided by the length of $[a, b]$, which is $b - a$, hence

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{b-a},$$

where $|A|$ is the length of A. In this case $(\Omega, \mathcal{A}, \mathrm{P})$ called by a *geometric probability space*.

In some sense the geometric probability space is the continuous version of the classical probability space, because we can calculate probability with the classical formula. However, in this case, we should not count the cases as on a classical probability space, but *measure* it by measuring the length in question.

Definition 7.3 (Uniform distribution)**.** If we denote by X the randomly chosen number on the interval $[a, b]$, then we said that X has *uniform distribution* on the interval $[a, b]$.

Answer to Example 7.1: X is a random variable on the geometric probability space $(\Omega, \mathcal{A}, \mathrm{P})$, where $\Omega = [0, 1]$, and if we denote by A the event of X is less then $2/3$, then we get

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{2/3}{1} = \frac{2}{3}.$$

□

Example 7.4 (Motivational example)**.** Choose a number randomly in the interval $[0, 1]$ and denote it by X. What is the probability of X is equal to $1/2$?

Answer: X is a random variable on the geometric probability space $(\Omega, \mathcal{A}, \mathrm{P})$, where $\Omega = [0, 1]$, and if we denote by B the event of X is equal to $1/2$, then we get

$$\mathrm{P}(B) = \frac{|B|}{|\Omega|} = \frac{0}{1} = 0.$$

□

It is important to see that event B is an example for an event with probability 0, which is not the impossible event. This fact can be weird, but this is a common feature of continuous random variables.

The same proof implies that for any arbitrary number $z \in \mathbb{R}$, we get

$$\mathrm{P}(X = z) = 0,$$

hence the distribution of X cannot be defined like in the discrete case. Instead of $\mathrm{P}(X = z)$, calculate the probability $\mathrm{P}(X \leq z)$. This idea is the base of the following general definition of a random variable and the so-called *distribution function*.

Definition 7.5 (Random variables and distribution function)**.** If $(\Omega, \mathcal{A}, \mathrm{P})$ is a probability space, then a function $X : \Omega \to \mathbb{R}$ is called a *random variable*, if for all $z \in \mathbb{R}$, we have $\{\omega \in \Omega : X(\omega) \leq z\} \in \mathcal{A}$. Then the function $F_X : \mathbb{R} \to [0, 1]$,

$$F_X(z) := \mathrm{P}(X \leq z), \quad z \in \mathbb{R},$$

is called the *(cumulative) distribution function* of X.

Using the distribution function, we can calculate probabilities in general. The following result can be used in the continuous case.

Proposition 7.6 (Calculation of probabilities of a continuous random variable using the distribution function)**.** *If X is a continuous random variable, then for any constants a and b, we have*

$$\mathrm{P}(a \leq X \leq b) = F_X(b) - F_X(a), \qquad \mathrm{P}(a \leq X) = 1 - F_X(a), \qquad \mathrm{P}(X \leq b) = F_X(b).$$

The distribution function is well defined in the discrete case as well, but we do not discussed it in this course.

Example 7.7. Choose a number randomly in the interval $[0, 1]$ and denote it by X, namely X is a uniformly distributed random variable on $[0, 1]$. What is the distribution function of X?

Answer: X is a random variable on the geometric probability space $(\Omega, \mathcal{A}, \mathrm{P})$, where $\Omega = [0, 1]$, so the distribution function of X is

$$F_X(z) = \mathrm{P}(X \leq z) = \begin{cases} 0, & z < 0, \\ z, & 0 \leq z \leq 1, \\ 1, & 1 < z. \end{cases}$$

We can draw this function as well, see Figure 7.

□

Density function

In the continuous case there is an other crucial object, which is called the *density function*.

Definition 7.8 (Continuous random variables and density function)**.** If $(\Omega, \mathcal{A}, \mathrm{P})$ is a probability space, $X : \Omega \to \mathbb{R}$ is a random variable and there exists a function $f_X : \mathbb{R} \to [0, \infty)$ such that

$$F_X(z) = \int_{-\infty}^{z} f_X(t)\,\mathrm{d}t, \qquad x \in \mathbb{R},$$

then f_X is called the *density function* of X and we say that the random variable X is *(absolutely) continuous*.

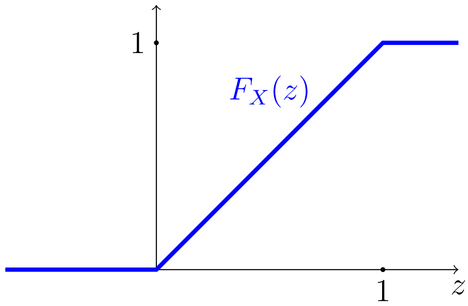


Figure 7: The distribution function of the uniform distribution on $[0, 1]$.

Due to this definition we can see that we can calculate probabilities using the density function as well. The probability of belonging to an interval is equal to the area under the density function on the interval in question.

Proposition 7.9 (Calculation of probabilities using density function)**.** *If X is a continuous random variable, then for any constants a and b, we have*

$$\mathrm{P}(a \le X \le b) = \int_a^b f_X(t)\mathrm{d}t, \qquad \mathrm{P}(a \le X) = \int_a^\infty f_X(t)\mathrm{d}t, \qquad \mathrm{P}(X \le b) = \int_{-\infty}^b f_X(t)\mathrm{d}t.$$

The definition of continuous random variables and density function seems to be complicated. It is useful to investigate the analogical connection between the discrete and continuous case. The density function is the continuous version of the probability distribution. Recall Proposition 3.8. We know that for any discrete random variable X with distribution $p_k = \mathrm{P}(X = k)$

$$\mathrm{P}(a \le X \le b) = \sum_{k=a}^{b} p_k.$$

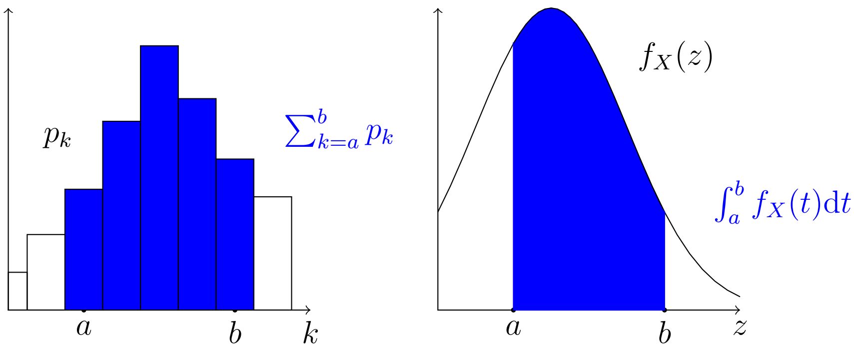


Figure 8: Analogical connection between discrete and continuous case.

Furthermore recall Theorem 3.5. We know that for any discrete random variable X with distribution p_k the following are valid.

(i) $p_k \ge 0$ for all $k \in \mathbb{Z}$,

(ii) $\sum_{k\in\mathbb{Z}} p_k = 1$.

Theorem 7.10 (Properties of density function)**.** *For any continuous random variable X with density function f the following are valid.*

(i) $f(t) \geq 0$ *for all* $t \in \mathbb{R}$,

(ii) $\int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1$.

Due to these results we can imagine the density function as a mass distribution. We have unit mass (e.g. 1 kg sugar) and we distribute this mass continuously on the interval of the possible values, and the amount of mass in each interval represents the probability that the random variable is in this interval.

Proposition 7.11 (Connection between distribution and density function)**.** *If X is a continuous random variable, then*

$$F_X' = f_X.$$

Example 7.12. What is the density function of X, which is a uniformly distributed random variable on $[0,1]$?

Answer: We know the distribution function of X

$$F_X(z) = \mathrm{P}(X \leq z) = \begin{cases} 0, & z < 0, \\ z, & 0 \leq z \leq 1, \\ 1, & 1 < z. \end{cases}$$

Hence the density function f_X is the derivative of F, namely

$$f_X(t) = \begin{cases} 1, & 0 \leq t \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

We can draw this function as well, see Figure 9.

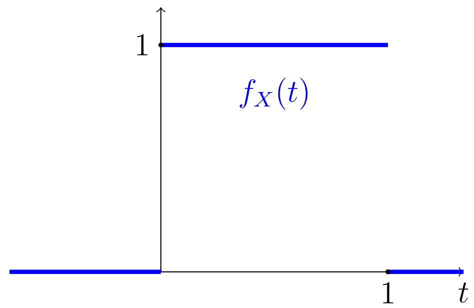


Figure 9: The density function of the uniform distribution on $[0,1]$.

□

Expectation, variance

The expectation can be defined in the same way as in the discrete case. Indeed, the expectation of a continuous random variable is a weighted average of the possible values. Due to the fact that the range is typically an interval, the average is an integral-average, and the weighting is based on the density function.

Definition 7.13 (Expectation)**.** If $X : \Omega \to \mathbb{R}$ is a continuous random variable with density function $f_X : \mathbb{R} \to [0,\infty)$, then the quantity

$$\mathrm{E}(X) := \int_{-\infty}^{\infty} t\, f_X(t)\,\mathrm{d}t$$

is called the *expectation* of X, provided that $\int_{-\infty}^{\infty} |t| f_X(t)\,\mathrm{d}t < \infty$.

After the definition, all of the results and heuristics learned in the discrete case are valid.

Proposition 7.14 (Expectation of a function of a continuous random variable)**.** *Let* $g : \mathbb{R} \to \mathbb{R}$ *be a function. If* $X : \Omega \to \mathbb{R}$ *is a continuous random variable with density function* f_X*, then*

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t)\,\mathrm{d}t.$$

Proposition 7.15 (Linearity of expectation)**.** *Let* X *and* Y *be random variables whose expectations exist and finite, further let* $a, b \in \mathbb{R}$ *be arbitrary constants. Then*

$$\mathrm{E}(aX + bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y).$$

Definition 7.16 (Independence of random variables)**.** The random variables X and Y are called *independent* if for any $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$, the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent.

Proposition 7.17 (Expectation of products of independent random variables)**.** *Let* X *and* Y *be independent random variables whose expectations exist and is finite. Then*

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y).$$

The definition and the properties of the variance are exactly the same as in the discrete case.

Definition 7.18 (Variance)**.** Let X be a random variable such that $\mathrm{E}(X)$ exists and finite. Then the *variance* of X is defined by

$$\mathrm{Var}(X) := \mathrm{E}\left((X - \mathrm{E}(X))^2\right).$$

Proposition 7.19 (Properties of variance)**.** *Let* X *and* Y *be random variables such that their variances exist and are finite. Then*

(i) $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$*,*

(ii) for any constants $c, d \in \mathbb{R}$*,* $\mathrm{Var}(cX + d) = c^2\,\mathrm{Var}(X)$*,*

(iii) $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X, Y)$*.*

Finally, a summary is listed to emphasize to analogical connection between the discrete and the continuous case, see Figure 10.

Further readings:

- `https://en.wikipedia.org/wiki/List_of_probability_distributions`
- `https://en.wikipedia.org/wiki/Integral`
- `https://en.wikipedia.org/wiki/Bertrand_paradox_(probability)`
- `https://en.wikipedia.org/wiki/Non-measurable_set`

If X is a discrete random variable with distribution $p_k,\ k \in \mathbb{Z}$:	If X is a continuous random variable with density function $f(t),\ t \in \mathbb{R}$:
$p_k \geq 0$ and $\sum_k p_k = 1$.	$f(t) \geq 0$ and $\int_{-\infty}^{\infty} f(t)\mathrm{d}t = 1$.
Range: $X \in \{k \in \mathbb{Z} : p_k > 0\}$.	Range: $X \in \{t \in \mathbb{R} : f(t) > 0\}$.
$\mathrm{P}(a \leq X \leq b) = \sum_{k=a}^{b} p_k$.	$\mathrm{P}(a \leq X \leq b) = \int_a^b f(t)\mathrm{d}t$.
$\mathrm{E}(X) = \sum_k k\ p_k$.	$\mathrm{E}(X) = \int_{-\infty}^{\infty} t\ f(t)\mathrm{d}t$.
$\mathrm{E}(X^2) = \sum_k k^2\ p_k$.	$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} t^2\ f(t)\mathrm{d}t$.

Figure 10: Summary of the analogical connection between the discrete and the continuous case.

7.1 Exercises

Problem 7.1. Consider an investment with initial capital \$900. The future value of this investment is random, denote it by X in thousand of dollars. We know the density function of X, which is the following

$$f(t) = \begin{cases} ct, & 0 \leq t \leq 1, \\ c(2-t), & 1 \leq t \leq 2, \\ 0, & \text{otherwise,} \end{cases}$$

where c is an unknown parameter.

(a) Determine the value of c, such that f is a density function. Plot the density function.

(b) What are the possible values of X?

(c) What is the probability that the future value is at least \$1500?

(d) What is the expectation of X? What is the standard deviation of X?

(e) Based on these results, is this investment valuable or not?

(f) * Determine the distribution function of X, and plot it.

(g) * Determine the value of t_0, such that $\mathrm{P}(X \geq t_0) = 0.9$.

Problem 7.2. Consider the previous problem, but with the following density function:

(a)

$$f(t) = \begin{cases} c, & 0 \leq t \leq 2, \\ 0, & \text{otherwise,} \end{cases}$$

(b)

$$f(t) = \begin{cases} ct, & 0 \leq t \leq 2, \\ 0, & \text{otherwise,} \end{cases}$$

(c)

$$f(t) = \begin{cases} c, & 0 \leq t \leq 1, \\ -ct + 2c, & 1 \leq t \leq 2, \\ 0, & \text{otherwise,} \end{cases}$$

(d)

$$f(t) = \begin{cases} 2c, & 0 \le t \le 1, \\ c, & 1 \le t \le 2, \\ 0, & \text{otherwise,} \end{cases}$$

(e)

$$f(t) = \begin{cases} ct^2, & 0 \le t \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

Problem 7.3. * Let's denote by X a continuous random variable with uniform distribution on the interval $[a, b]$. It means that X is randomly chosen number in the interval $[a, b]$.

(a) Determine the distribution function and plot it.

(b) Determine the density function and plot it.

(c) Calculate the expectation.

Problem 7.4. We investigate the corn production in Hungary. We assume that the production has uniform distribution on the interval $[3.5, 5.5]$ in million tons. If the production is greater then the market price is lower, so we assume that if the production is x, then the market price of a ton of corn is $100 - 10x$ in thousand of forints.

(a) Denote by X the corn production in million tons. What is the density function of X? What is the probability that X is greater then 5 million tons. What is the expectation of X? *Determine the value of t_0, such that $\mathrm{P}(X \ge t_0) = 0.9$.

(b) Denote by Y the market price of a ton of corn in thousand forints. What is the probability that Y is less than 50 thousand forints. What is the expectation of Y?

(c) * Denote by Z the value of the corn production in billion forints. What is the expectation of Z?

Problem 7.5. Let's investigate the arrivals of the customers in a shop. Denote by X the time between two arrivals. We know that the expectation of X is 6 minutes, and assume that X has *exponential distribution with parameter* λ, which means that

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \ge 0, \\ 0, & t < 0. \end{cases}$$

(a) Determine the value of the parameter λ, such that the expectation of X is 6 minutes.

(b) What is the probability that X is greater then 10 minutes?

(c) What is the probability that X is between 5 and 10 minutes?

(d) *What is the probability that the next customer comes after at least 10 minutes, if we know that the previous one came before 1 hour?

The final answers to these problems can be found in section 10.

8 Normal distribution

There is a notable continuous distribution, which plays a crucial rule in the theory and in applications of probability.

Definition 8.1 (Normal distribution)**.** If X is a continuous random variable with a density function

$$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}, \qquad t \in \mathbb{R},$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, then we call X *a normally distributed* (or Gaussian) random variable with parameters (μ, σ^2). Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proposition 8.2 (The meaning of the parameters)**.** *If* $X \sim \mathcal{N}(\mu, \sigma^2)$*, then*

$$\mathrm{E}(X) = \mu, \qquad \mathrm{Var}(X) = \sigma^2.$$

The meaning of the parameters can be seen by investigating the density function. The parameter μ is connected to a linear transformation of the density function. The parameter σ describes the shape of the density function, see Figure 11.

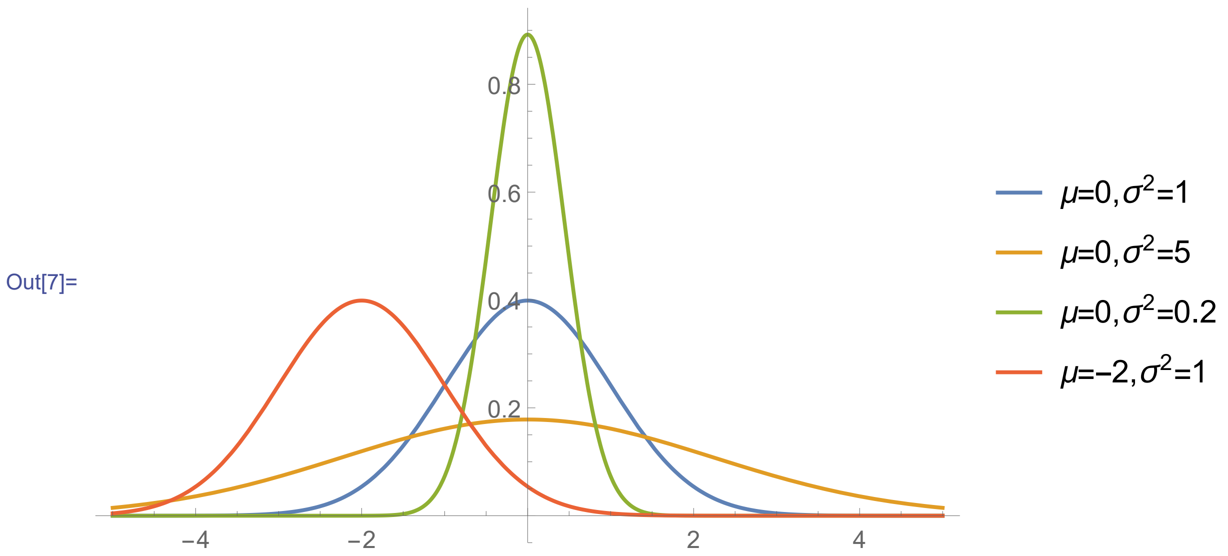


Figure 11: The meaning of the parameters.

We have a special element in the family of normal distributions.

Definition 8.3 (The standard normal distribution)**.** In the case of $\mu = 0$ and $\sigma = 1$, that is if

$$\varphi(t) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{t^2}{2}}, \qquad t \in \mathbb{R},$$

Z has *standard normal distribution.*

The density function of the standard normal distribution sometimes called by the *bell curve*, see Figure 12, and it is traditionally denoted by φ.

If we are working with normal distributions, we have a problem with calculating probabilities. The problem is that the integral

$$\mathrm{P}(a \le Z \le b) = \int_a^b \varphi(t)\mathrm{d}t = \int_a^b \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$$

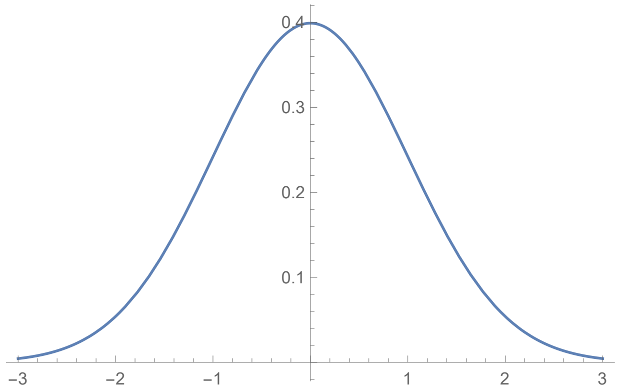


Figure 12: The bell curve.

can not be calculated explicitly. Instead of the density function, we use the approximated values of the distribution function. The distribution function of the standard normal distribution is conventionally denoted by $\varPhi$,

$$\varPhi(z) = \mathrm{P}(Z \leq z), \qquad z \in \mathbb{R}.$$

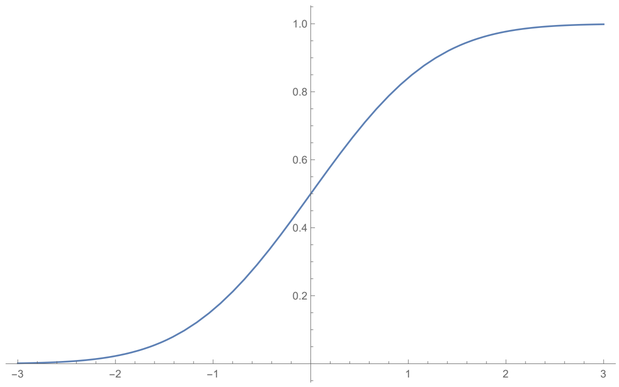


Figure 13: The distribution function $\varPhi$ of the standard normal distribution.

Proposition 8.4 (Properties of $\varPhi$)**.** *The function $\varPhi$ is a continuous and strictly monotone increasing, and*

$$\varPhi(-z) = 1 - \varPhi(z), \qquad z \in \mathbb{R}.$$

Proposition 8.5 (Calculating probabilities with the distribution function)**.**

$$\mathrm{P}(a \leq Z \leq b) = \varPhi(b) - \varPhi(a), \qquad a, b \in \mathbb{R},$$

$$\mathrm{P}(Z \leq b) = \varPhi(b), \qquad b \in \mathbb{R},$$

$$\mathrm{P}(a \leq Z) = 1 - \varPhi(a), \qquad a \in \mathbb{R}.$$

We have already discussed how we can calculate probabilities connected to the standard normal distribution. In the general case, we use the method of *standardization*.

Proposition 8.6 (Standardization)**.** *Let be $X \sim \mathcal{N}(\mu, \sigma^2)$. Then the random variable*

$$Z := \frac{X - \mu}{\sigma}$$

is the standardization of X, which has standard normal distribution.

Hence

$$\mathrm{P}(a \le X \le b) = \mathrm{P}\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Further readings:

- `https://en.wikipedia.org/wiki/Normal_distribution`
- `https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss`
- `https://en.wikipedia.org/wiki/Gaussian_integral`

8.1 Exercises

Problem 8.1. There is a machine in the milk factory which fills the 1 litre milk box automatically. This machine does not work perfectly, the amount of the milk in a milk box has normal distribution with an expected value of 1000 ml and a standard deviation of 10 ml.

(a) What is the probability that the amount of the milk in a milk box is greater then 1010 ml?

(b) What is the probability that the amount of the milk in a milk box differs only with up to 20 ml?

(c) Give an interval (the interval, which is around the expectation symmetrically) such that the amount of the milk in a milk box is in this interval with probability 95%. Namely, find the value of d such that $\mathrm{P}(\mu - d \leq X \leq \mu + d) = 0.95$.

Problem 8.2. The intelligence quotient (IQ) is a score derived from several tests designed to assess human intelligence. These tets are constructed to assess IQ in a normal distribution with an expected value of 100 and a standard derivation of 15.

(a) What is the probability that the IQ is between 90 and 120?

(b) What is the probability that the IQ is at least 131?

(c) Give an interval (the interval, which is around the expectation symmetrically) such that the IQ is in this interval with probability 95%.

Problem 8.3. Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.

Problem 8.4. Given a normal random variable with $\mu = 40$ and $\sigma = 6$, find the value of x that has

(a) $\mathrm{P}(X < x) = 0.45$;

(b) $\mathrm{P}(X > x) = 0.14$.

Problem 8.5. A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

Problem 8.6. In an industrial process, the diameter of a ball bearing is an important measurement. The buyer sets specifications for the diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean $\mu = 3.0$ cm and standard deviation $\sigma = 0.005$ cm. On average, how many manufactured ball bearings will be scrapped?

Problem 8.7. A lawyer commutes daily from his suburban home to his midtown office. The average time for a one-way trip is 24 minutes, with a standard deviation of 3.8 minutes. Assume the distribution of trip times to be normally distributed.

(a) What is the probability that a trip will take at least half an hour?

(b) If the office opens at 9:00 A.M. and the lawyer leaves his house at 8:45 A.M. daily, what percentage of the time is he late for work?

(c) If he leaves the house at 8:35 A.M. and coffee is served at the office from 8:50 A.M. until 9:00 A.M., what is the probability that he misses coffee?

(d) Find the length of time above which we find the slowest 15% of the trips.

(e) *Find the probability that 2 of the next 3 trips will take at least half an hour.

The final answers to these problems can be found in section 10.

9 Approximation to normal distribution

It is mentioned before that normal distribution plays an important role in the theory and in the applications of probability as well. The reason is that in many situations when the outcome of a random experiment depends on a lot of small random effects, basically the outcome is equal to the sum of these random variables, then the distribution of the result is close to a normal distribution. This means that we can approximate the original, typically unknown distribution with a normal distribution, and the approximation is better if there are more summands. This approach is covered by the de–Moivre–Laplace and the central limit theorem which are presented in this part.

Example 9.1 (Motivational example)**.** We toss a coin 10 times. What is the probability that the number of heads will be between 4 and 6?

Answer: We can solve this problem. Denote by X the number of heads, then $X \sim \text{binom}(n, p)$, with $n = 10$ and $p = 1/2$, hence we get

$$\begin{aligned} \mathrm{P}(4 \le X \le 6) &= \sum_{k=4}^{6} \mathrm{P}(X = k) = \mathrm{P}(X = 4) + \mathrm{P}(X = 5) + \mathrm{P}(X = 6) \\ &= \binom{10}{4}(1/2)^{10} + \binom{10}{5}(1/2)^{10} + \binom{10}{6}(1/2)^{10} = 0.6563. \end{aligned}$$

□

Example 9.2 (Motivational example)**.** We toss a coin 1000 times. What is the probability that the number of heads will be between 480 and 520?

Answer 1.: We can solve this problem in the same way as before. Indeed, if $X =$ number of heads, then $X \sim \text{binom}(n, p)$, with $n = 1000$ and $p = 1/2$, hence we get

$$\mathrm{P}(480 \le X \le 520) = \sum_{k=480}^{520} \mathrm{P}(X = k) = \sum_{k=480}^{520} \binom{1000}{k}(1/2)^{1000}.$$

□

However, this is difficult to calculate and impossible to solve with a simple calculator. Using a computer one can derive the exact answer, which is 0.8052. It would be good if we could calculate or approximate at least this probability in an easier way. The next result can help us to solve this problem.

Theorem 9.3 (De Moivre–Laplace theorem)**.** *Let $S_n \sim$ binom(n, p). Then we have for all $a, b \in \mathbb{R} \cup \{\pm\infty\}$, $a < b$*

$$\lim_{n\to\infty} \mathrm{P}(a \le S_n \le b) = \mathrm{P}(a \le X \le b),$$

where $X \sim \mathcal{N}(\mu, \sigma^2)$, with

$$\mu = \mathrm{E}(S_n) = np \qquad \textit{and} \qquad \sigma = \mathsf{D}(S_n) = \sqrt{np(1-p)}.$$

As a consequence, we can give an approximate answer for probabilities $\mathrm{P}(a \le S_n \le b)$ using the de-Moivre–Laplace theorem, if n is large enough (e.g. if $n > 100$ is a good rule of thumb).

Answer 2.: Denote by $S_n =$ the number of heads, then $S_n \sim \text{binom}(n, p)$, with $n = 1000$ and $p = 1/2$, hence using the de-Moivre–Laplace theorem, because n is large enough ($n > 100$), we get

$$\mathrm{P}(480 \leq S_n \leq 520) \approx \mathrm{P}(480 \leq X \leq 520),$$

where $X \sim \mathcal{N}(\mu, \sigma^2)$, with $\mu = \mathrm{E}(S_n) = np = 500$ and $\mathsf{D}(S_n) = \sqrt{np(1-p)} = 15.81$.

$$\mathrm{P}(480 \leq X \leq 520) = \mathrm{P}(-1.27 \leq Z \leq 1.27) = 2\Phi(1.27) - 1 = 0.796.$$

□

We can see, that the approximated solution (0.796) is close to the real one (0.8052). To illustrate the accuracy of the approximation, we plot the binomial distribution with parameter $n = 10$ and $n = 1000$, and the density function of the corresponding normal distribution, see Figure 14 and Figure 15.

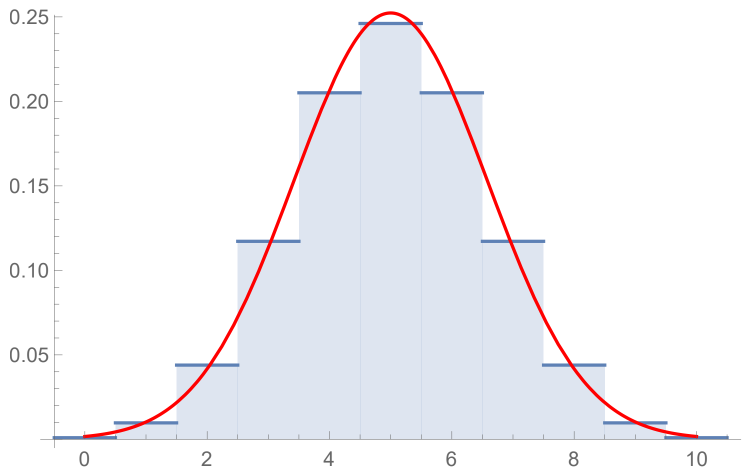


Figure 14: The distribution of $\text{binom}(n, p)$ with $n = 10$ and $p = 1/2$ (blue) and the density function of $\mathcal{N}(\mu, \sigma^2)$ with $\mu = 5$ and $\sigma = 1.581$ (red).

The de-Moivre–Laplace theorem is a special case of the following so-called *central limit theorem*, which can be used in a more general case.

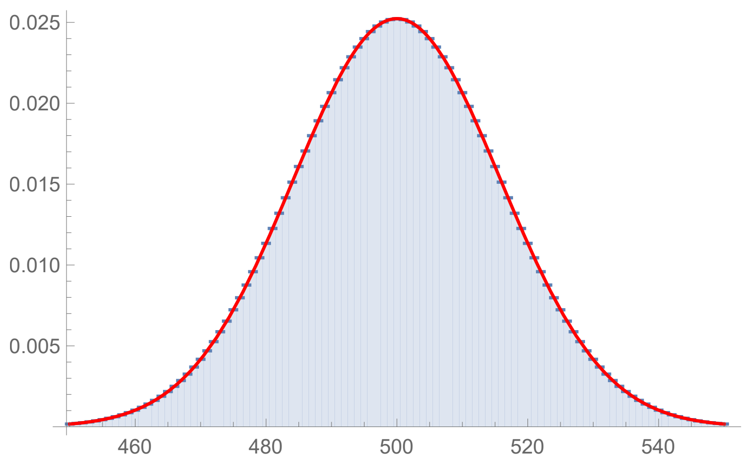


Figure 15: The distribution of $\text{binom}(n, p)$ with $n = 1000$ and $p = 1/2$ (blue) and the density function of $\mathcal{N}(\mu, \sigma^2)$ with $\mu = 500$ and $\sigma = 15.81$ (red).

Theorem 9.4 (Central limit theorem)**.** *Let be $X_1, X_2, \ldots$ a sequence of independent and identically distributed random variables with finite standard deviation and let $S_n := X_1 + \cdots + X_n$. Then we have for all $a, b \in \mathbb{R} \cup \{\pm\infty\}$, $a < b$*

$$\lim_{n\to\infty} \mathrm{P}(a \le S_n \le b) = \mathrm{P}(a \le X \le b),$$

where $X \sim \mathcal{N}(\mu, \sigma^2)$, with

$$\mu = \mathrm{E}(S_n) \qquad \textit{and} \qquad \sigma = \mathsf{D}(S_n).$$

Using the properties of the expectation and the variance, we can express the above quantities with the common expectation $\mathrm{E}(X)$ and common standard deviation $\mathsf{D}(X)$. Indeed,

$$\mathrm{E}(S_n) = \mathrm{E}(X_1 + \cdots + X_n) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = n\,\mathrm{E}(X),$$

and

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\,\mathrm{Var}(X),$$
$$\mathsf{D}(S_n) = \sqrt{\mathrm{Var}(S_n)} = \sqrt{n\,\mathrm{Var}(X)} = \sqrt{n}\mathsf{D}(X).$$

As a consequence, we can give an approximate answer for probabilities $\mathrm{P}(a \le S_n \le b)$ using the central limit theorem, if S_n is a sum of independent and identically distributed random variables and if n is large enough (e.g. if $n > 100$ is a good rule of thumb).

The de–Moivre–Laplace theorem is a special case of the central limit theorem if the common distribution of X_i is Bernoulli and hence S_n is a binomial distributed variable. The strength of the central limit theorem is that we do not have to know the distribution of the summands to approximate probabilities connected to the sum. It is enough to know the expectation and the standard deviation. To illustrate this result, we plot the case if X has uniform distribution on the interval $[0, 1]$ with $n = 2$ and $n = 4$, see Figure 16 and Figure 17.

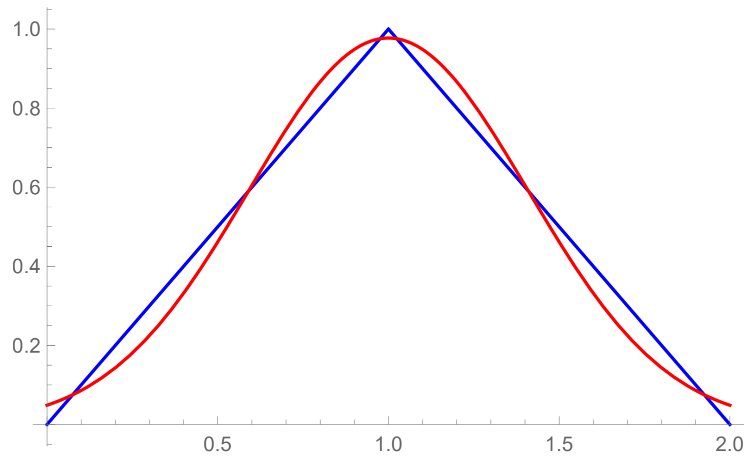


Figure 16: The density function of S_2 (blue) and the corresponding normal distribution (red).

Example 9.5. Consider a certain type of insurance at an insurance company. We know that the expected value of the amount of loss is \$300 and the standard deviation is \$50, but we do not know the distribution of it. If the number of contracts with loss is 100, what is the approximate probability that the total amount of losses is between 29 and 31 thousand of dollars?

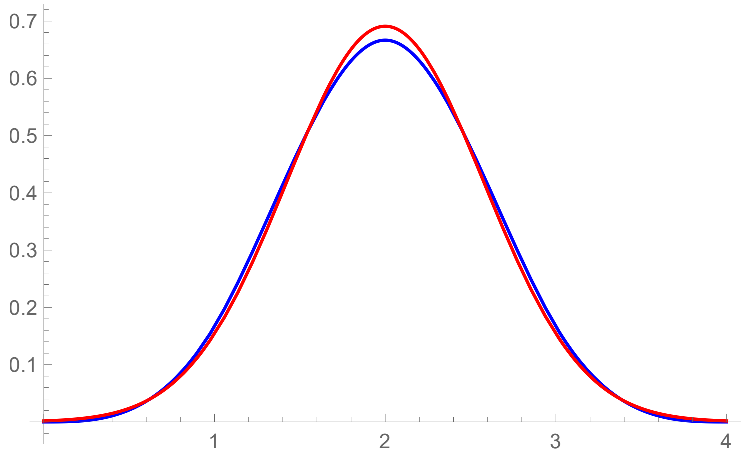


Figure 17: The density function of S_4 (blue) and the corresponding normal distribution (red).

Answer: Denote by X the amount of a single loss. Then $\mathrm{E}(X) = 300$, $\mathsf{D}(X) = 50$ and $S_{100} = X_1 + \ldots X_{100}$. Hence using the central limit theorem, we get

$$\mathrm{P}(29000 \leq S_{100} \leq 31000) \approx \mathrm{P}(29000 \leq Y \leq 31000),$$

where $Y \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = n\,\mathrm{E}(X) = 100 \cdot 300 = 30000$ and $\sigma = \sqrt{n}\mathsf{D}(X) = 10 \cdot 50 = 500$. Finally, this probability can be calculated after standardization, so

$$\mathrm{P}(29000 \leq Y \leq 31000) = 0.9545.$$

□

Further readings:

- https://en.wikipedia.org/wiki/Bean_machine
- https://en.wikipedia.org/wiki/Central_limit_theorem

9.1 Exercises

Problem 9.1. We roll a fair dice 200 times. What is the approximate probability that the number of sixes is between 30 and 40?

Problem 9.2. There are 490 students on the microeconomics lecture. Every student visits the lecture with probability 5/7 independent of each other. Find the approximate probability that the number of attendees on a given day is between 338 and 362.

Problem 9.3. In a hotel there are 600 guests, but because of the fire alarm we have to evacuate the building. The hotel manager ask nearby hotels for the number of rooms they could provide for the night. Hotel A has 375, while hotel B has 255 open rooms. The manager has no time to find a room for everyone, so he suggests that guests go to whichever hotel they like more. Given that each guest chooses hotel A with probability 0.6, find the approximate probability that everyone can find a room in the first hotel they visit.

Problem 9.4. We want to optimize train travel between Chicago and Los Angeles. We want to offer two trains departing form two different stations in Chicago. We think that 1000 people would want to use our trains and each of them would choose between the two options with equal probability. Choose the carrying capacity k for the trains in a way that the approximate probability of a traveller missing the train because there is no seat available is less than 0.01.

Problem 9.5. An insurance company has 10 000 contracts. Each of the contract is associated with a loss with probability 1%, independently in a certain year. Denoted by Z the number of contracts with loss. What is the approximate probability that Z is between 85 and 115? Find the value of t such that $\mathrm{P}(Z \geq t) = 0.1$.

Problem 9.6. There is an elevator at the dorm with a maximum capacity of 800 kg. What is the approximate probability that 10 people cannot use this lift, if we know that the weight of a person has expectation 80 kg and standard deviation 15 kg?

Problem 9.7. * A statistician wishes to examine p, the ratio of smokers in the population of Budapest. She devises the following method: choose n person to ask about their smoking habits with everyone being equally likely to be selected, then use $p' = k/n$ as an estimate of p where k is the amount of smokers among the survey participants. Find a lower bound for n such that the estimate p' is at most 0.005 off with probability at least 0.95.

The final answers to these problems can be found in section 10.

10 Final answers and solutions to the exercises

Final answers to the Exercises 1.1

Problem 1.1. $7!$

Problem 1.2. $4!;\ \ 3!;\ \ 4^4;\ \ 4^3$

Problem 1.3. $5!;\ \ 2 \cdot 4!$

Problem 1.4. $\frac{9!}{3! \cdot 2!}$

Problem 1.5. $\frac{4!}{2!};\ \ \frac{3!}{2!}$

Problem 1.6. $\frac{11!}{5! \cdot 3! \cdot 3!}$

Problem 1.7. 12^3

Problem 1.8. 6^{10}

Problem 1.9. $\binom{5}{2}$

Problem 1.10. $\binom{31}{3};\ \ 31 \cdot 30 \cdot 29$

Problem 1.11. $\binom{52}{5}$

Final answers to the Exercises 2.1

Problem 2.1. 1. $\frac{4}{10}$; 2. $\frac{3}{6}$

Problem 2.2. A: $\frac{2^4}{2^5}$; B: $\frac{2 \cdot 2^3}{2^5}$; C: $\frac{\binom{5}{2}+\binom{5}{4}}{2^5}$; D: $\frac{\binom{5}{3}}{2^5}$; E: $\frac{\binom{5}{5}+\binom{5}{4}+\binom{5}{3}}{2^5}$; F: $\frac{2}{2^5}$

Problem 2.3. A: $\frac{2^2}{2^3}$; B: $\frac{2 \cdot 2}{2^3}$; C: $\frac{\binom{3}{2}}{2^3}$; D: $\frac{\binom{3}{3}}{2^3}$; E: $\frac{\binom{3}{3}+\binom{3}{2}}{2^3}$; F: $\frac{2}{2^3}$

Problem 2.4. A: $\frac{1}{6^3}$; B: $\frac{5^3}{6^3}$; C: $\frac{6 \cdot 5 \cdot 4}{6^3}$; D: $1 - \frac{6 \cdot 5 \cdot 4}{6^3}$; E: $\frac{3^3}{6^3}$; F: $1 - \frac{5^3}{6^3}$

Problem 2.5. A: $\frac{1}{6^4}$; B: $\frac{5^4}{6^4}$; C: $\frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4}$; D: $1 - \frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4}$; E: $\frac{3^4}{6^4}$; F: $1 - \frac{5^4}{6^4}$

Problem 2.6. A: $\frac{\binom{39}{5}}{\binom{52}{5}}$; B: $\frac{\binom{13}{5}}{\binom{52}{5}}$; C: $1 - \frac{\binom{39}{5}}{\binom{52}{5}}$; D: $\frac{\binom{51}{4}}{\binom{52}{5}}$; E: $\frac{\binom{48}{1}}{\binom{52}{5}}$; F: $\frac{\binom{4}{2} \cdot \binom{4}{2} \cdot \binom{44}{1}}{\binom{52}{5}}$

Problem 2.7. A: $\frac{4}{\binom{52}{5}}$; B: $\frac{4 \cdot 10 - 4}{\binom{52}{5}}$; C: $\frac{13 \cdot 48}{\binom{52}{5}}$; D: $\frac{13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}}{\binom{52}{5}}$; E: $\frac{4 \cdot \binom{13}{5} - 40}{\binom{52}{5}}$; F: $\frac{10 \cdot 4^5 - 40}{\binom{52}{5}}$; G: $\frac{13 \cdot \binom{4}{3} \cdot \binom{12}{2} \cdot 4^2}{\binom{52}{5}}$; H $\frac{\binom{13}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot 11 \cdot 4}{\binom{52}{5}}$; I: $\frac{13 \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3}{\binom{52}{5}}$; J: $\frac{\left(\binom{13}{5} - 10\right) \cdot \left(4^5 - 4\right)}{\binom{52}{5}}$

Problem 2.8. A: $\frac{4}{43}$; B: $\frac{34}{43}$; C: $\frac{27}{43}$

Problem 2.9. (a) $\frac{1}{36}$; (b) $\frac{1}{36}$; (c) $\frac{2}{36}$

Final answers to the Exercises 3.1

Problem 3.1. (a) $P(X = k)$, $k \in \{1, \ldots, 6\}$, namely the discrete uniform distribution on the set $\{1, \ldots, 6\}$.
(b) $P(Y = k)$, $k \in \{1, \ldots, 6\}$, namely the discrete uniform distribution on the set $\{1, \ldots, 6\}$.
(c) $Z = X + Y$

k	2	3	4	5	6	7	8	9	10	11	12
$P(Z = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

(d) $Z = \max\{X, Y\}$

k	1	2	3	4	5	6
$P(Z = k)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

(e) $Z = \min\{X, Y\}$

k	1	2	3	4	5	6
$P(Z = k)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

(f) $Z = (X - Y)^2$

k	0	1	4	9	16	25
$P(Z = k)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

Problem 3.2. Denote by X the number generated by rolling this modified dice.

k	1	4	5	6
$P(X = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{3}{6}$	$\frac{1}{6}$

Problem 3.3. Denote by X the number of tails shown. Then $X \sim \text{binom}(20, 0.5)$, namely

$$P(X = k) = \frac{\binom{20}{k}}{2^{20}}, \qquad k \in \{0, 1, \ldots, 20\}.$$

Problem 3.4. This can be described by a classical probability space, where $|\Omega| = 3! = 6$. So the distribution of X is the following.

k	0	1	3
$P(X = k)$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{1}{6}$

Problem 3.5. Denote by X the number of yellow balls drawn. In the first case, when we put the balls back: $X \sim \text{binom}\left(5, \frac{7}{12}\right)$.

In the second case, when we don't put the balls back: $X \sim \text{hypergeo}(13, 7, 5)$.

Problem 3.6.

k	0	1	2
$P(X = k)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{3}{8}$

Problem 3.7.

k	0	1	2	3
$P(X = k)$	0.06	0.29	0.44	0.21

Problem 3.8.

k	0	1000	3000	9000
$P(X = k)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{5}{36}$	$\frac{1}{36}$

Final answers to the Exercises 4.1

Problem 4.1. Denote by X the number of winning tickets. Then $X \sim \text{binom}(6, 0.4)$.

$$P(X = 4) = \binom{6}{4} 0.4^4 \cdot 0.6^2 \qquad E(X) = 2.4$$

Problem 4.2. Denote by X the number of throws. Then $X \sim \text{geo}(0.5)$, so $E(X) = 2$. Denote by Z the amount of earnings. Then we should pay $E(Z)$ Ft, which is the fair price of the game. Using that $Z = 100X$, we get

$$E(Z) = E(100X) = 100\, E(X) = 200.$$

Problem 4.3. Denote by X the number of tests needed. Then $X \sim \text{geo}(0.15)$.

$$P(X = 5) = (1 - 0.15)^4 \cdot 0.15 = 0.85^4 \cdot 0.15 \approx 0.0783.$$

Denote by Z the total cost of the tests. Then $Z = 10000X$, so the expected cost is

$$E(Z) = E(10000X) = 10000\, E(X) = 10000 \cdot \frac{100}{15} \approx 66667.$$

Problem 4.4. Denote by X the number of stop for a light. Then $X \sim \text{binom}(5, 0.6)$. Denote by Z the amount of delay in seconds. Then $Z = 10X$, so

$$E(Z) = E(10X) = 10\, E(X) = 10 \cdot 3 = 30.$$

$$P(Z = 30) = P(10X = 30) = P(X = 3) = \binom{5}{3} 0.6^3 \cdot 0.4^2.$$

Problem 4.5. Denote by Z the amount of earnings. Then the distribution of Z is the following.

k	100	30	0
$P(Z = k)$	$\frac{2}{36}$	$\frac{6}{36}$	$\frac{28}{36}$

Then we should pay $E(Z)$ Ft, which is the fair price of the game.

$$E(Z) = 100 \cdot \frac{2}{36} + 30 \cdot \frac{6}{36} \approx 10.56.$$

Problem 4.6. Denote by Z the amount of earnings. Then the distribution of Z is the following.

k	500,000,000	2,000,000	300,000	2,000	0
$P(Z = k)$	$\frac{1}{\binom{90}{5}}$	$\frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}}$	$\frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}}$	$\frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}}$	$\frac{\binom{85}{5}}{\binom{90}{5}}$

$$E(Z) = 500,000,000 \cdot \frac{1}{\binom{90}{5}} + \cdots + 2,000 \cdot \frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}} \approx 319.36.$$

Problem 4.7. Denote by Z the amount of earnings. Then the fair price of this game is $E(Z)$.

$$E(Z) = 250 \cdot \frac{2}{6} + 1000 \cdot \frac{1}{6} = 250.$$

Problem 4.8. Denote by Z the amount of earnings. Then the fair price of this game is $E(Z)$.

$$E(Z) = 54,000 \cdot \frac{1}{6^3} = 250.$$

Problem 4.9. Denote by X the number of attempts needed. Then $X \sim \text{geo}(0.17)$, so $\mathrm{E}(X) = \frac{100}{17} \approx 5.88$.

Denote by Z the amount of time to finish the map (in minutes). Then $Z = 10X$, so

$$\mathrm{E}(Z) = \mathrm{E}(10X) = 10\,\mathrm{E}(X) = \frac{1000}{17} \approx 58.8.$$

$$\mathrm{P}(Z \le 60) = \mathrm{P}(10X \le 60) = \mathrm{P}(X \le 6) = \sum_{k=1}^{6} \mathrm{P}(X = k) = \sum_{k=1}^{6} 0.83^{k-1} \cdot 0.17 \approx 0.67.$$

Problem 4.10. Denote by X the number of popcorns needed. Then $X \sim \text{binom}(100, 0.2)$.

$$\mathrm{P}(X \le 35) = \sum_{k=0}^{35} \mathrm{P}(X = k) = \sum_{k=0}^{35} \binom{100}{k} 0.2^k \cdot 0.8^{100-k} \approx 0.99,$$

so we have not a real problem. :)

Problem 4.11. Denote by Z the amount of earnings. Then the distribution of Z is the following.

k	-32	64	16
$\mathrm{P}(Z = k)$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{3}{8}$

$$\mathrm{E}(Z) = (-32) \cdot \frac{2}{8} + 64 \cdot \frac{3}{8} + 16 \cdot \frac{3}{8} = 22.$$

In the biased coin case the distribution of Z is changed.

k	-32	64	16
$\mathrm{P}(Z = k)$	$0.25^3 \cdot (0.75)^3$	$\binom{3}{2} 0.25^2 \cdot 0.75$	$\binom{3}{1} 0.25 \cdot 0.75^2$

$$\mathrm{E}(Z) = 15.54.$$

Problem 4.12. Denote by X the number of upwards steps in the first 10 minutes. Then $X \sim \text{binom}(10, 0.5)$.

$$\mathrm{P}(\text{he will be back at the pub after 10 minutes}) = \mathrm{P}(X = 5) = \binom{10}{5} 0.5^5 \cdot 0.5^5 \approx 0.2461.$$

Denote by Y the number of upwards steps in the first 20 minutes. Then $Y \sim \text{binom}(20, 0.5)$.

$$\mathrm{P}(\text{he will be back at the pub after 20 minutes}) = \mathrm{P}(Y = 10) = \binom{20}{10} 0.5^{10} \cdot 0.5^{10} \approx 0.1762.$$

In the case when he prefers to go up the street with probability 2/3, then $X \sim \text{binom}(10, 2/3)$, and $Y \sim \text{binom}(20, 2/3)$.

$$\mathrm{P}(\text{he will be back at the pub after 10 minutes}) = \mathrm{P}(X = 5) = \binom{10}{5} (2/3)^5 \cdot (1/3)^5 \approx 0.1332.$$

$$\mathrm{P}(\text{he will be back at the pub after 20 minutes}) = \mathrm{P}(Y = 10) = \binom{20}{10} (2/3)^{10} \cdot (1/3)^{10} \approx 0.0516.$$

Further reading: `https://en.wikipedia.org/wiki/Random_walk`

Problem 4.13. `https://en.wikipedia.org/wiki/Coupon_collector%27s_problem`

Final answers to the Exercises 5.1

Problem 5.1. Denote by X the number of faulty cars. Then $X \sim \text{binom}(n, p)$ with $n = 1000$ and $p = 0.02$.

$$\mathrm{E}(X) = np = 1000 \cdot 0.02 = 20 \qquad \mathrm{Var}(X) = np(1-p) = 1000 \cdot 0.02(1-0.02) = 19.6$$

Problem 5.2. Denote by X the number of trials necessary. Then $X \sim \text{geo}(p)$ with $p = 3/7$.

$$\mathrm{E}(X) = \frac{1}{p} = \frac{7}{3} \qquad \mathrm{Var}(X) = \frac{1-p}{p^2} = \frac{28}{9}$$

Problem 5.3. Denote by Z_1 the amount of earnings in the first game, and Z_2 in the second game. Then

$$\mathrm{E}(Z_1) = 18000\,\mathrm{P}(Z_1 = 18000) = 18000 \cdot \frac{1}{36} = 500,$$
$$\mathrm{E}(Z_2) = 3000\,\mathrm{P}(Z_2 = 3000) = 3000 \cdot \frac{6}{36} = 500.$$

Hence the expectation are the same. To choose between the games, we should calculate the variances (risk).

$$\mathrm{E}(Z_1^2) = 9\ 000\ 000, \qquad \mathrm{E}(Z_2^2) = 1\ 500\ 000,$$

$$\mathrm{Var}(Z_1) = \mathrm{E}(Z_1^2) - \mathrm{E}^2(Z_1) = 8\ 750\ 000, \qquad \mathrm{Var}(Z_2) = 1\ 250\ 000.$$

In the first game we have more risk than in the second one, thus we should play the second game instead of the first one.

Problem 5.4. Denote by Z_A our score using the following strategy. If we roll 3,4,5,6, then we stop, if we roll 2, then roll again.

$$Z_A \in \{1, 3, 4, 5, 6\}$$

$$\begin{aligned}
\mathrm{P}(Z_A = 1) &= \mathrm{P}(\text{roll 1 in the 1th round}) + \mathrm{P}(\text{roll 1 in the 2nd round}) \\
&\quad + \mathrm{P}(\text{roll 1 in the 3rd round}) + \dots \\
&= \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \left(\frac{1}{6}\right)^2 \frac{1}{6} + \dots = \frac{1}{6} \left(\sum_{i=0}^{\infty} \left(\frac{1}{6}\right)^i \right) \\
&= \frac{1}{6} \left(\frac{1}{1 - \frac{1}{6}} \right) = \frac{1}{5}.
\end{aligned}$$

Similarly, one can get

$$\mathrm{P}(Z_A = 1) = \mathrm{P}(Z_A = 3) = \mathrm{P}(Z_A = 4) = \mathrm{P}(Z_A = 5) = \mathrm{P}(Z_A = 6) = \frac{1}{5},$$

hence

$$\mathrm{E}(Z_A) = \frac{1 + 3 + 4 + 5 + 6}{5} = 4.$$

Denote by Z_B our score using the following strategy. If we roll 4,5 or 6, then we stop, if we roll 2 or 3, then roll again.

$$Z_B \in \{1, 4, 5, 6\}$$

$$
\begin{aligned}
\mathrm{P}(Z_B = 1) &= \mathrm{P}(\text{roll 1 in the 1th round}) + \mathrm{P}(\text{roll 1 in the 2nd round}) \\
&\quad + \mathrm{P}(\text{roll 1 in the 3rd round}) + \dots \\
&= \frac{1}{6} + \frac{2}{6} \cdot \frac{1}{6} + \left(\frac{2}{6}\right)^2 \frac{1}{6} + \dots = \frac{1}{6}\left(\sum_{i=0}^{\infty} \left(\frac{2}{6}\right)^i\right) \\
&= \frac{1}{6}\left(\frac{1}{1-\frac{2}{6}}\right) = \frac{1}{4}.
\end{aligned}
$$

Similarly, one can get

$$
\mathrm{P}(Z_B = 1) = \mathrm{P}(Z_B = 4) = \mathrm{P}(Z_B = 5) = \mathrm{P}(Z_B = 6) = \frac{1}{4},
$$

hence

$$
\mathrm{E}(Z_B) = \frac{1+4+5+6}{4} = 4.
$$

Thus, we have derived, that the strategy A and B has the same expected score. To choose between the strategies, we should calculate the variances (risk).

$$
\mathrm{Var}(Z_A) = 1.4, \qquad \mathrm{Var}(Z_B) = 3.5,
$$

so we should choose the strategy A, because it is less risky then strategy B.

Problem 5.5.

$$
\mathrm{E}(X) = 1, \qquad \mathrm{E}(X^2) = 2, \qquad \mathrm{Var}(X) = 1.
$$

Problem 5.6.

$$
\mathrm{E}((X-1)^2) = \mathrm{E}(X^2) - 2\,\mathrm{E}(X) + 1, \qquad \mathrm{E}((X-2)^2) = \mathrm{E}(X^2) - 4\,\mathrm{E}(X) + 4,
$$

hence we have to solve this system

$$
\begin{aligned}
\mathrm{E}(X^2) - 2\,\mathrm{E}(X) - 9 &= 0 \\
\mathrm{E}(X^2) - 4\,\mathrm{E}(X) - 1 &= 0.
\end{aligned}
$$

We get $\mathrm{E}(X) = 4$ and $\mathrm{E}(X^2) = 17$, which implies $\mathrm{Var}(X) = 1$.

Problem 5.7. We known, that $\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{35}{12} \approx 2.92$. Using the properties of the variance, we get

$$
\mathrm{Var}(3X - Y) = \mathrm{Var}(3X) + \mathrm{Var}(-Y) = 9\,\mathrm{Var}(X) + \mathrm{Var}(Y) \approx 29.17.
$$

$$
\mathrm{Var}(X + 5Y - 5) = \mathrm{Var}(X + 5Y) = \mathrm{Var}(X) + \mathrm{Var}(5Y) = \mathrm{Var}(X) + 25\,\mathrm{Var}(Y) \approx 75.92.
$$

Problem 5.8. We can use the identity $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$.

Case (a).

$$
\begin{gathered}
\mathrm{E}(X) = 3.5, \quad \mathrm{E}(Y) = 3.5, \\
\mathrm{E}(XY) = \frac{1\cdot 6 + 2\cdot 5 + 3\cdot 4 + 4\cdot 3 + 5\cdot 2 + 6\cdot 1}{6} = 9.33 \\
\mathrm{Cov}(X,Y) = -2.92.
\end{gathered}
$$

Furthermore

$$
\mathsf{D}(X) = 1.71, \quad \mathsf{D}(Y) = 1.71,
$$

hence using the definition of the correlation, we get

$$\mathrm{corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathsf{D}(X)\mathsf{D}(Y)} = -1,$$

so we have deterministic negative linear dependence between X and Y. Indeed, $Y = 7 - X$.

Case (b). $\mathrm{Cov}(X,Y) = 5.33$ and $\mathrm{corr}(X,Y) = 0.96$, hence we have found a strong positive linear dependence between X and Y.

Case (c). $\mathrm{Cov}(X,Y) = -1.75$ and $\mathrm{corr}(X,Y) = -0.83$, hence we have found a strong negative linear dependence between X and Y.

Case (d). $\mathrm{Cov}(X,Y) = -0.44$ and $\mathrm{corr}(X,Y) = -0.057$, hence we have found a weak negative linear dependence between X and Y.

Problem 5.9.

$$\mathrm{Var}(X) = 5.84 \qquad \mathrm{Var}(Y) = 80.07 \qquad \mathrm{Cov}(X,Y) = 20.42 \qquad \mathrm{corr}(X,Y) = 0.95$$

Problem 5.10. Denote by R and by C the revenue and the cost of the company. Then we know that

$$\mathrm{E}(R) = 120 \qquad \mathsf{D}(R) = 30 \qquad \mathrm{Var}(R) = 30^2 = 900$$

$$\mathrm{E}(C) = 80 \qquad \mathsf{D}(C) = 20 \qquad \mathrm{Var}(C) = 20^2 = 400$$

and denote by P the profit, hence $P = R - C$.

In both cases, (a) and (b), the expectation will be the same.

$$\mathrm{E}(P) = \mathrm{E}(R - C) = \mathrm{E}(R) - \mathrm{E}(C) = 120 - 80 = 40$$

The variance will be different.

(a) The case of independence.

$$\mathrm{Var}(P) = \mathrm{Var}(R - C) = \mathrm{Var}(R) + \mathrm{Var}(C) = 900 + 400 = 1300$$

(b) The case of dependence. We know that $\mathrm{corr}(R,C) = 0.8$.

$$\begin{aligned}\mathrm{Var}(P) &= \mathrm{Var}(R - C) = \mathrm{Var}(R) + \mathrm{Var}(C) - 2\,\mathrm{Cov}(R,C) \\ &= 900 + 400 - 2\,\mathrm{corr}(R,C)\mathsf{D}(R)\mathsf{D}(C) = 1300 - 2 \cdot 0.8 \cdot 30 \cdot 20 = 1300 - 960 = 340\end{aligned}$$

Problem 5.11. This is connected to the so-called Mean-Variance portfolio analysis, see the case study.

Final answers to the Exercises 6.1

Problem 6.1. Denote by A_i the event that the ith student pulls an easy topic, $i = 1, 2, 3$.

a)

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\,\mathrm{P}(A_2)\,\mathrm{P}(A_3) = \left(\frac{4}{10}\right)^3$$

because the events A_1, A_2 and A_3 are independent. We know that, because they pull the topics with replacement.

b)

$$\mathrm{P}(A_1 \cap \overline{A_3}) = \mathrm{P}(A_1)(1 - \mathrm{P}(A_3)) = \frac{4}{10} \cdot \frac{6}{10}$$

c) Denote by X the number of pulled easy topics. Then $X \sim \mathrm{binom}(n, p)$ with $n = 3$ and $p = 4/10$.

$$\mathrm{P}(X = 2) = \binom{3}{2} \cdot \left(\frac{4}{10}\right)^2 \cdot \frac{6}{10}$$

d) They pull the topics now without replacement, so the events A_1, A_2 and A_3 are not independent anymore. $\mathrm{P}(A_1) = 4/10$. Further, $\mathrm{P}(A_2)$ can be calculated by the law of total probability.

$$\mathrm{P}(A_2) = \mathrm{P}(A_2 \,|\, A_1)\,\mathrm{P}(A_1) + \mathrm{P}(A_2 \,|\, \overline{A_1})\,\mathrm{P}(\overline{A_1}) = \frac{3}{9} \cdot \frac{4}{10} + \frac{4}{9} \cdot \frac{6}{10} = \frac{4}{10}$$

The same can be derived for $\mathrm{P}(A_3)$, so the order doesn't matter to getting an easy topic.

Problem 6.2. Denote by A the event that a person is color-blind, and denote by B that a person is a man. Then we know that

$$\mathrm{P}(B) = \mathrm{P}(\overline{B}) = 0.5 \qquad \mathrm{P}(A \,|\, B) = 0.05 \quad \mathrm{P}(A \,|\, \overline{B}) = 0.0.25$$

a) By the law of total probability, we get

$$\mathrm{P}(A) = \mathrm{P}(A \,|\, B)\,\mathrm{P}(B) + \mathrm{P}(A \,|\, \overline{B})\,\mathrm{P}(\overline{B}) = 0.05 \cdot 0.5 + 0.025 \cdot 0.5 = 0.0375$$

b) By the Bayes formula, we get

$$\mathrm{P}(B \,|\, A) = \frac{\mathrm{P}(A \,|\, B)\,\mathrm{P}(B)}{\mathrm{P}(A)} = \frac{0.05 \cdot 0.5}{0.0375} = 0.67$$

Problem 6.3. Denote by A the event that we know the right answer, and denote by B that we choose the right answer. Then we know that

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

$$\mathrm{P}(B \,|\, A) = 1 \qquad \mathrm{P}(B \,|\, \overline{A}) = \frac{1}{3}$$

a) By the law of total probability, we get

$$\mathrm{P}(B) = \mathrm{P}(B \,|\, A)\,\mathrm{P}(A) + \mathrm{P}(B \,|\, \overline{A})\,\mathrm{P}(\overline{A}) = 1 \cdot p + \frac{1}{3}(1 - p) = \frac{1 + 2p}{3}$$

b) By the Bayes formula, we get

$$\mathrm{P}(A \,|\, B) = \frac{\mathrm{P}(B \,|\, A)\,\mathrm{P}(A)}{\mathrm{P}(B)} = \frac{1 \cdot p}{\frac{1+2p}{3}} = \frac{3p}{1 + 2p}$$

Problem 6.4. Denote by B_i the event that the apple delivered by the ith producer, $i = 1, 2, 3, 4$, and denote by A that the apple is first class. Then we know that

$$\mathrm{P}(B_1) = 0.1 \quad \mathrm{P}(B_2) = 0.3 \quad \mathrm{P}(B_3) = 0.4 \quad \mathrm{P}(B_4) = 0.2$$
$$\mathrm{P}(A \,|\, B_1) = 0.4 \quad \mathrm{P}(A \,|\, B_2) = 0.5 \quad \mathrm{P}(A \,|\, B_3) = 0.2 \quad \mathrm{P}(A \,|\, B_4) = 1$$

a) By the law of total probability, we get

$$\mathrm{P}(A) = \sum_{i=1}^{4} \mathrm{P}(A \,|\, B_i)\,\mathrm{P}(B_i) = 0.47$$

b) By the Bayes formula, we get

$$\mathrm{P}(B_1 \,|\, \overline{A}) = \frac{\mathrm{P}(\overline{A} \,|\, B_1)\,\mathrm{P}(B_1)}{\mathrm{P}(\overline{A})} = \frac{(1-0.4)0.1}{1-0.47} = 0.11$$

Problem 6.5. Denote by A the event that you are sick, and denote by B that the result of the test is positive. Then we know that

$$\mathrm{P}(A) = 0.01 \qquad \mathrm{P}(\overline{A}) = 0.99$$
$$\mathrm{P}(B \,|\, A) = 0.99 \qquad \mathrm{P}(\overline{B} \,|\, \overline{A}) = 0.99$$

By the Bayes formula and the law of total probability, we get

$$\begin{aligned}\mathrm{P}(A \,|\, B) &= \frac{\mathrm{P}(B \,|\, A)\,\mathrm{P}(A)}{\mathrm{P}(B)} = \frac{\mathrm{P}(B \,|\, A)\,\mathrm{P}(A)}{\mathrm{P}(B \,|\, A)\,\mathrm{P}(A) + \mathrm{P}(B \,|\, \overline{A})\,\mathrm{P}(\overline{A})} \\ &= \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + (1-0.99) \cdot 0.99} = 0.5\end{aligned}$$

Problem 6.6. Denote by Z the random variable of number of gained points. Then the question is the value of $\mathrm{E}(Z)$. Denote by A_i the event that we roll i with the dice. Then $A_1, \ldots, A_6$ is a partition, and $\mathrm{P}(A_i) = \frac{1}{6}$ for any $i = 1, \ldots 6$.

$$\mathrm{E}(Z \,|\, A_i) = \frac{i}{2},$$

because the conditional distribution of Z given A_i is binomial$(i, \frac{1}{2})$. Hence, we can use the Law of total expectation

$$\mathrm{E}(Z) = \sum_{i=1}^{6} \mathrm{E}(Z \,|\, A_i)\,\mathrm{P}(A_i) = \sum_{i=1}^{6} \frac{i}{2} \cdot \frac{1}{6} = \frac{7}{4}.$$

Problem 6.7. Denote by A the event that somebody sings in the shower. Then the question is $\mathrm{P}(A)$. Denote by B the event that tossing two heads with the coins, and denote by C the event that somebody gives the answer *yes* to the question. Then we know

$$\mathrm{P}(B) = \frac{1}{4}, \qquad \mathrm{P}(C) = \frac{875}{2000}.$$

Furthermore, we known that

$$\mathrm{P}(C \,|\, B) = \mathrm{P}(A), \qquad \mathrm{P}(C \,|\, \overline{B}) = \mathrm{P}(\overline{A}).$$

Using the Law of total probability and the information above, we get

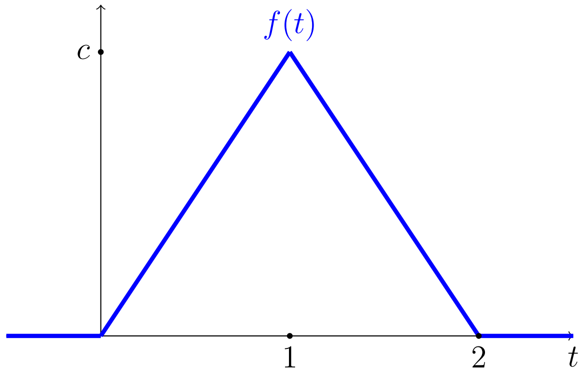
$$\begin{aligned}\mathrm{P}(C) &= \mathrm{P}(C \,|\, B)\,\mathrm{P}(B) + \mathrm{P}(C \,|\, \overline{B})\,\mathrm{P}(\overline{B}) \\ &= \mathrm{P}(A)\,\mathrm{P}(B) + \mathrm{P}(\overline{A})\,\mathrm{P}(\overline{B}),\end{aligned}$$

thus we get an equation for $\mathrm{P}(A)$. After ordering and substituting, we get

$$\mathrm{P}(A) = \frac{\mathrm{P}(C) - (1 - \mathrm{P}(B))}{2\,\mathrm{P}(B) - 1} = \frac{5}{8}.$$

Final answers to the Exercises 7.1

Problem 7.1. (a) f is a density function if and only if $f(t) \geq 0$ for all $t \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.



Using the formula for the area of a triangle (or calculating the integral), we get

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = c.$$

So we derive that $c = 1$, thus

$$f(t) = \begin{cases} t, & 0 \leq t \leq 1, \\ 2 - t, & 1 \leq t \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

(b) The possible values of a continuous random variable are the numbers t, for which $f(t) > 0$.

Hence now $X \in [0, 2]$.

(c)

$$\mathrm{P}(X \geq 1.5) = \int_{1.5}^{\infty} f(x)\mathrm{d}x = \frac{1}{8}.$$

(d) $\mathrm{E}(X) = 1$, $\mathsf{D}(X) = 1/\sqrt{6}$.

(e) The expected income is \$1000, the initial capital \$900, so the expected profit is \$100, positive, hence this investment is valuable.

(f) We known that $F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$.

If $t < 0$, then $F(t) = 0$, and if $t > 2$, then $F(t) = 1$.

Further if $t \in [0, 1]$, then

$$F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x = \int_{0}^{t} x\mathrm{d}x = \frac{t^2}{2},$$

and if $t \in [1, 2]$, then

$$F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x = \int_{0}^{1} x\mathrm{d}x + \int_{1}^{t} (2 - x)\mathrm{d}x = -\frac{t^2}{2} + 2t - 1.$$

So finally,

$$F(z) = \begin{cases} 0, & t < 0, \\ \frac{t^2}{2}, & 0 \leq t \leq 1, \\ -\frac{t^2}{2} + 2t - 1, & 1 \leq t \leq 2, \\ 1, & t > 2, \end{cases}$$

(g)

$$\begin{aligned}
\mathrm{P}(X \geq t_0) &= 0.9 \\
1 - \mathrm{P}(X \leq t_0) &= 0.9 \\
1 - F(t_0) &= 0.9 \\
F(t_0) &= 0.1 \\
\frac{t_0^2}{2} &= 0.1 \\
t_0 &= \sqrt{0.2} \approx 0.447.
\end{aligned}$$

Problem 7.2. (a)

(a) $c = 1/2$. (b) $X \in [0,2]$. (c) $\mathrm{P}(X \leq 1500) = 0.25$. (d) $\mathrm{E}(X) = 1$, $\mathsf{D}(X) = 1/\sqrt{3}$. (e) Valuable.

(b)

(a) $c = 1/2$. (b) $X \in [0,2]$. (c) $\mathrm{P}(X \leq 1500) = 7/16$. (d) $\mathrm{E}(X) = 4/3$, $\mathsf{D}(X) = \sqrt{2}/3$. (e) Valuable.

(c)

(a) $c = 2/3$. (b) $X \in [0,2]$. (c) $\mathrm{P}(X \leq 1500) = 1/12$. (d) $\mathrm{E}(X) = 7/9$, $\mathsf{D}(X) = \frac{\sqrt{\frac{37}{2}}}{9}$. (e) Not valuable.

(d)

(a) $c = 1/3$. (b) $X \in [0,2]$. (c) $\mathrm{P}(X \leq 1500) = 1/6$. (d) $\mathrm{E}(X) = 5/6$, $\mathsf{D}(X) = \frac{\sqrt{11}}{6}$. (e) Not valuable.

(e)

(a) $c = 3/8$. (b) $X \in [0,2]$. (c) $\mathrm{P}(X \leq 1500) = 37/64$. (d) $\mathrm{E}(X) = 3/2$, $\mathsf{D}(X) = \frac{\sqrt{\frac{3}{5}}}{2}$. (e) Valuable.

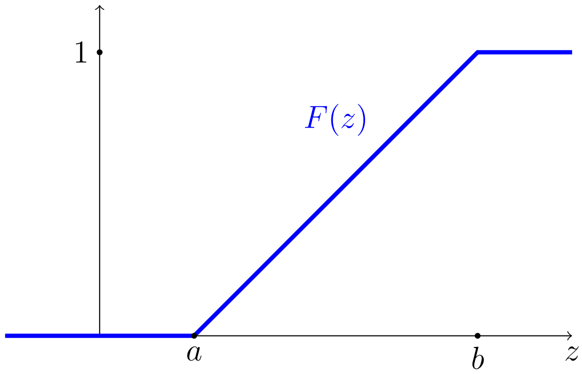
Problem 7.3. This problem is just for proving that the density and the distribution function and the expectation of the uniform distribution are really the same as in the formula sheet.

(a) The distribution function is $F(z) = \mathrm{P}(X \leq z)$ by definition.
If $z < a$, then $\mathrm{P}(X \leq z) = 0$, so $F(z) = 0$.
If $z > b$, then $\mathrm{P}(X \leq z) = 1$, so $F(z) = 1$.
If $z \in [a,b]$, then because of it it a geometric probability, we get

$$F(z) = \mathrm{P}(X \leq z) = \frac{z-a}{b-a}.$$

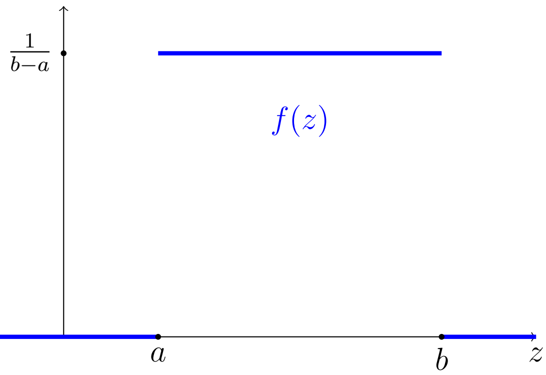
So finally, we get

$$F(z) = \begin{cases} 0, & z < a, \\ \frac{z-a}{b-a}, & a \leq z \leq b, \\ 1, & z > b, \end{cases}$$



(b) We know that $F' = f$, where f is the density function. Hence we get

$$f(z) = \begin{cases} \frac{1}{b-a}, & a \le z \le b, \\ 0, & \text{otherwise}, \end{cases}$$



Problem 7.4. (a) X has uniform distribution on $[3.5, 5.5]$, so

$$f(t) = \begin{cases} 0.5, & 3.5 \le t \le 5.5, \\ 0, & \text{otherwise}. \end{cases}$$

$$\mathrm{P}(X > 5) = \int_5^5 .5 f(x)\mathrm{d}x = 0.25.$$

$$\begin{aligned} \mathrm{P}(X \ge t_0) &= 0.9 \\ (5.5 - t_0)0.5 &= 0.9 \\ t_0 &= 3.7. \end{aligned}$$

$$\mathrm{E}(X) = \frac{a+b}{2} = \frac{3.5+5.5}{2} = 4.5.$$

(b) $Y = 100 - 10X$, so

$$\mathrm{P}(Y < 50) = \mathrm{P}(100 - 10X < 50) = \mathrm{P}(5 < X) = 0.25.$$

$$\mathrm{E}(Y) = \mathrm{E}(100 - 10X) = 100 - 10\,\mathrm{E}(X) = 100 - 10 \cdot 4.5 = 55.$$

(c) $Z = XY$, so we get

$$\begin{aligned} \mathrm{E}(Z) = \mathrm{E}(XY) &= \mathrm{E}(X(100 - 10X)) = \mathrm{E}(100X - 10X^2) = 100\,\mathrm{E}(X) - 10\,\mathrm{E}(X^2) \\ &= 100 \cdot 4.5 - 10 \cdot 20.58 = 244.2. \end{aligned}$$

Problem 7.5. Proof of that the density function of the exponential distribution is a real density function:

We need $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^{\infty} \lambda \mathrm{e}^{-\lambda x}\mathrm{d}x = \lambda \int_0^{\infty} \mathrm{e}^{-\lambda x}\mathrm{d}x = \lambda \left[-\frac{\mathrm{e}^{-\lambda x}}{\lambda}\right]_{x=0}^{\infty} = \lambda \left(\lim_{t\to\infty} -\frac{\mathrm{e}^{-\lambda x}}{\lambda} + \frac{1}{\lambda}\right) = 1.$$

So we get that f is a density function for all $\lambda > 0$. This is the so-called *exponential distribution* with parameter λ.

(a) $\mathrm{E}(X) = 1/\lambda$, hence $\lambda = 1/6$.

(b)

$$\mathrm{P}(X > 10) = \int_{10}^{\infty} \frac{1}{6}\mathrm{e}^{-\frac{1}{6}x}\mathrm{d}x = \mathrm{e}^{-\frac{10}{6}} \approx 0.19.$$

(c)

$$\mathrm{P}(5 < X < 10) = \int_5^{10} \frac{1}{6}\mathrm{e}^{-\frac{1}{6}x}\mathrm{d}x = \mathrm{e}^{-\frac{10}{6}} + \mathrm{e}^{-\frac{5}{6}} \approx 0.25.$$

Or if you calculate the distribution function F, which is

$$F(t) = \begin{cases} 1 - \mathrm{e}^{-\lambda t}, & t \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

then you can calculate the probabilities with it as well:

$$\mathrm{P}(X > 10) = 1 - F(10) = \mathrm{e}^{-\frac{10}{6}},$$

$$\mathrm{P}(5 < X < 10) = F(10) - F(5) = 1 - \mathrm{e}^{-\frac{10}{6}} - (1 - \mathrm{e}^{-\frac{5}{6}}) = \mathrm{e}^{-\frac{10}{6}} + \mathrm{e}^{-\frac{5}{6}}.$$

Final answers to the Exercises 8.1

Problem 8.1. X = the amount of milk in a milk box in ml. $X \sim \mathcal{N}(\mu, \sigma^2)$, with $\mu = 1000$ and $\sigma = 10$. (a)

$$P(X > 1010) = P\left(\frac{X-\mu}{\sigma} > \frac{1010-1000}{10}\right) = P(Z > 1)$$

with $Z = \frac{X-\mu}{\sigma}$ which has standard normal distribution, hence

$$P(Z > 1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587.$$

(b)

$$\begin{aligned} P(980 < X < 1020) &= P\left(\frac{980-1000}{10} < \frac{X-\mu}{\sigma} < \frac{1020-1000}{10}\right) \\ &= P(-2 < Z < 2) = \Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(2)) = \\ &= 2\Phi(2) - 1 = 2 \cdot 0.9772 - 1 = 0.9544. \end{aligned}$$

(c)

$$\begin{aligned} P(\mu - d \le X \le \mu + d) &= 0.95 \\ P\left(\frac{1000-d-1000}{10} < \frac{X-\mu}{\sigma} < \frac{1000+d-1000}{10}\right) &= 0.95 \\ P\left(\frac{d}{10} < Z < \frac{d}{10}\right) &= 0.95 \\ \Phi\left(\frac{d}{10}\right) - \Phi\left(-\frac{d}{10}\right) &= 0.95 \\ \Phi\left(\frac{d}{10}\right) - \left(1 - \Phi\left(\frac{d}{10}\right)\right) &= 0.95 \\ 2\Phi\left(\frac{d}{10}\right) - 1 &= 0.95 \\ \Phi\left(\frac{d}{10}\right) &= 0.975 \\ \frac{d}{10} &= 1.96 \\ d &= 19.6. \end{aligned}$$

Hence we get that the interval is $[980.4, 1019.6]$, namely

$$P(980.4 \le X \le 1019.6) = 0.95.$$

Problem 8.2. (a) 0.6568.
(b) 0.0192.
(c) $d = 29.4$, so the interval is $[70.6, 129.4]$.

Problem 8.3. 0.5764.

Problem 8.4. (a) $x = 39.28$.
(b) $x = 46.48$.

Problem 8.5. 0.0808.

Problem 8.6.

$$P(X < 2.99 \text{ cm or } X > 3.01 \text{ cm}) = 0.0456.$$

Thus on average, 4.56% of the manufactured ball bearings will be scrapped.

Problem 8.7. X = time for a one-way trip in minutes. $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = 24$ and $\sigma = 3.8$.

(a) $P(X > 30) = 0.0571$.
(b) $P(X > 15) = 0.9911$.
(c) $P(X < 15 \text{ or } X > 25) = 0.4051$.
(d) $P(X > x) = 0.15 \Rightarrow x = 27.94$.
(e) Y=number of trips of the next 3 trips will take at least 30 minutes. Then $Y \sim$ binom(3,p), with $p = P(X > 30) = 0.0571$.

$$P(Y = 2) = \binom{3}{2} p^2(1-p) = 0.0092.$$

Final answers to the Exercises 9.1

Problem 9.1. X = number of sixes, $X \sim \text{binom}(n, p)$ with $n = 200$ and $p = 1/6$. Using the de-Moivre–Laplace theorem, we get

$$P(30 \le X \le 40) \approx P(30 \le Y \le 40),$$

where $Y \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = E(X) = np = 33.33$ and $\sigma = \mathsf{D}(X) = \sqrt{np(1-p)} = 5.27$. Hence this probability can be calculated after standardization, so

$$P(30 \le Y \le 40) = 0.6335.$$

Problem 9.2. X = number of attendees, $X \sim \text{binom}(n, p)$ with $n = 490$ and $p = 5/7$. Using the de-Moivre–Laplace theorem, we get

$$P(338 \le X \le 362) \approx P(338 \le Y \le 362),$$

where $Y \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = E(X) = np = 350$ and $\sigma = \mathsf{D}(X) = \sqrt{np(1-p)} = 10$. Hence this probability can be calculated after standardization, so

$$P(338 \le Y \le 362) = 0.7699.$$

Problem 9.3. X = number of people who choose the hotel A, $X \sim \text{binom}(n, p)$ with $n = 600$ and $p = 0.6$. Of course, the number of people who choose the hotel B is $600 - X$, so we have to calculate the probability

$$P(X \le 375 \text{ and } 600 - X \le 255) = P((X \le 375 \text{ and } 345 \le X) = P(345 \le X \le 375).$$

Using the de-Moivre–Laplace theorem, we get

$$P(345 \le X \le 375) \approx P(345 \le Y \le 375),$$

where $Y \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = E(X) = np = 360$ and $\sigma = \mathsf{D}(X) = \sqrt{np(1-p)} = 12$. Hence this probability can be calculated after standardization, so

$$P(345 \le Y \le 375) = 0.7887.$$

Problem 9.4. X = number of people who choose the train A, $X \sim \text{binom}(n, p)$ with $n = 1000$ and $p = 0.5$. Of course, the number of people who choose the train B is $1000 - X$,so we have to calculate

$$\mathrm{P}(X \leq k \text{ and } 1000 - X \leq k) = \mathrm{P}(1000 - k \leq X \leq k)$$

such that the above probability is equal to 0.99. Using the de-Moivre–Laplace theorem, we get

$$\mathrm{P}(1000 - k \leq X \leq k) \approx \mathrm{P}(1000 - k \leq Y \leq k),$$

where $Y \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = \mathrm{E}(X) = np = 500$ and $\sigma = \mathsf{D}(X) = \sqrt{np(1-p)} = 15.81$. Hence this probability can be calculated after standardization, so

$$\mathrm{P}(1000 - k \leq Y \leq k) = \Phi\left(\frac{k - 500}{15.81}\right)$$

Finally, we have to solve the equation

$$\Phi\left(\frac{k - 500}{15.81}\right) = 0.99,$$

which implies that $k = 540.63$, so at least 541 seats are needed.

Problem 9.5. $\mathrm{P}(85 \leq Z \leq 155) \approx 0.8683$, $t = 112.75$.

Problem 9.6. X = the weight of a person. $\mathrm{E}(X) = 80$, $\mathsf{D}(X) = 15$. $S_{10} = X_1 + \ldots X_{10}$. Then using the central limit theorem ($n = 10$, so the approximation may be not so good), we get

$$\mathrm{P}(S_{10} > 800) \approx \mathrm{P}(Y > 800),$$

where $Y \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = n\,\mathrm{E}(X) = 10 \cdot 80 = 800$ and $\sigma = \sqrt{n}\mathsf{D}(X) = \sqrt{10} \cdot 15 = 47.43$. Hence this probability can be calculated after standardization, so

$$\mathrm{P}(Y > 800) = 0.5.$$