Upper level undergraduate probability with actuarial and financial applications

Richard F. Bass

Patricia Alonso Ruiz, Fabrice Baudoin, Maria Gordina, Phanuel Mariano, Oleksii Mostovyi, Ambar Sengupta, Alexander Teplyaev, Emiliano Valdez

Acknowledgments. The authors are very grateful to Joe Chen, Tom Laetsch, Tom Roby, Linda Westrick for their important contributions to this project, and to all the University of Connecticut students in Math 3160 for their tireless work in the course.

We gratefully acknowledge the generous support from the *University of Connecticut Open and Affordable Initiative*, the *Davis Educational Foundation*, the *UConn Co-op Bookstore*, and the *University of Connecticut Libraries*.

Preface

This textbook has been created as a part of the *University of Connecticut Open and Affordable Initiative*, which in turn was a response to the Connecticut State Legislature Special Act No. 15-18 (House Bill 6117), *An Act Concerning the Use of Digital Open-Source Textbooks in Higher Education.* At the University of Connecticut this initiative was supported by the *UConn Bookstore* and the *University of Connecticut Libraries.* Generous external support was provided by the *Davis Educational Foundation.*

Even before this initiative, our department had a number of freely available and internal resources for Math 3160, our basic undergraduate probability course. This included lecture notes prepared by Richard Bass, the Board of Trustees Distinguished Professor of Mathematics. Therefore, it was natural to extend the lecture notes into a complete textbook for the course. Two aspects of the courses were taken into account. On the one hand, the course is taken by many students who are interested in financial and actuarial careers. On the other hand, this course has multivariable calculus as a prerequisite, which is not common for most of the undergraduate probability courses taught at other US universities. The 2018 edition of the textbook has 4 parts divided into 15 chapters. The first 3 parts consist of required material for Math 3160, and the 4th part contains optional material for this course.

This textbook has been used in classrooms during 3 semesters at UConn, and received overwhelmingly positive feedback from students. However, we are still working on improving the text, and will be grateful for comments and suggestions.

May 2018.

Contents

Part 1

Discrete Random Variables

CHAPTER 1

Combinatorics

1.1. Basic counting principle and combinatorics

1.1.1. Basic counting principle. The first basic counting principle is to multiply. Namely, if there are n possible outcomes of doing something and m outcomes of doing another thing, then there are $m \cdot n$ possible outcomes of performing both actions.

Basic counting principle

Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes, and if for each outcome of experiment 1 there are n possible outcomes of experiment 2, then there are $m \cdot n$ possible outcomes of the two experiments together.

Example 1.1. Suppose we have 4 shirts of 4 different colors and 3 pants of different colors. How many different outfits are there? For each shirt there are 3 different colors of pants, so altogether there are $4 \times 3 = 12$ possibilities.

Example 1.2. How many different license plate numbers with 3 letters followed by 3 numbers are possible?

The English alphabet has 26 different letters, therefore there are 26 possibilities for the first place, 26 for the second, 26 for the third, 10 for the fourth, 10 for the fifth, and 10 for the sixth. We multiply to get $(26)^3(10)^3$.

1.1.2. Permutations. How many ways can one arrange letters a, b, c? We can list all possibilities, namely,

$$abc \quad acb \quad bac \quad bca \quad cab \quad cba.$$

There are 3 possibilities for the first position. Once we have chosen the letter in the first position, there are 2 possibilities for the second position, and once we have chosen the first two letters, there is only 1 choice left for the third. So there are $3 \times 2 \times 1 = 6 = 3!$ arrangements. In general, if there are n distinct letters, there are $n!$ different arrangements of these letters.

Example 1.3. What is the number of possible batting orders (in baseball) with 9 players? Applying the formula for the number of permutations we get $9! = 362880$.

 Typesetting date: September 2, 2020

Example 1.4. How many ways can one arrange 4 math books, 3 chemistry books, 2 physics books, and 1 biology book on a bookshelf so that all the math books are together, all the chemistry books are together, and all the physics books are together?

We can arrange the math books in 4! ways, the chemistry books in 3! ways, the physics books in 2! ways, and the biology book in $1! = 1$ way. But we also have to decide which set of books go on the left, which next, and so on. That is the same as the number of ways of arranging of four objects (such as the letters M, C, P, B), and there are 4! ways of doing that. We multiply to get the answer $4! \cdot (4! \cdot 3! \cdot 2! \cdot 1!) = 6912$.

In permutations the order **does** matter as is illustrated by the next example.

Example 1.5. How many ways can one arrange the letters a, a, b, c? Let us label them first as A, a, b, c. There are $4! = 24$ ways to arrange these letters. But we have repeats: we could have Aa or aA which are the same. So we have a repeat for each possibility, and so the answer should be $4!/2! = 12$.

If there were 3 as, 4 bs, and 2 cs, we would have

$$\frac{9!}{3!4!2!} = 1260.$$

What we just did is called finding the ***number of permutations***. These are permutations of a given set of objects (elements) unlike the example with the licence plate numbers where we could choose the same letter as many times as we wished.

Permutations

The number of permutations of n objects is equal to

$$n! := 1 \cdot ... \cdot n,$$

with the usual convention $0! = 1$.

1.1.3. Combinations.

Now let us look at what are known as ***combinations***.

Example 1.6. How many ways can we choose 3 letters out of 5? If the letters are a, b, c, d, e and order matters, then there would be 5 choices for the first position, 4 for the second, and 3 for the third, for a total of $5 \times 4 \times 3$. Suppose now the letters selected were a, b, c. If order does not matter, in our counting we will have the letters a, b, c six times, because there are 3! ways of arranging three letters. The same is true for any choice of three letters. So we should have $5 \times 4 \times 3/3!$. We can rewrite this as

$$\frac{5 \cdot 4 \cdot 3}{3!} = \frac{5!}{3!2!} = 10$$

This is often written as $\binom{5}{3}$, read "5 *choose* 3". Sometimes this is written $C_{5,3}$ or ${}_5C_3$.

Combinations (binomial coefficients)

The number of different groups of k objects chosen from a total of n objects is equal to

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

Note that this is true when the order of selection is irrelevant, and if the order of selection is relevant, then there are

$$n \cdot (n-1) \cdot ... \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

ways of choosing k objects out of n.

Example 1.7. How many ways can one choose a committee of 3 out of 10 people? Applying the formula of the number of combinations we get $\binom{10}{3} = 120$.

Example 1.8. Suppose there are 8 men and 8 women. How many ways can we choose a committee that has 2 men and 2 women? We can choose 2 men in $\binom{8}{2}$ ways and 2 women in $\binom{8}{2}$ ways. The number of possible committees is then the product

$$\binom{8}{2} \cdot \binom{8}{2} = 28 \cdot 28 = 784.$$

Example 1.9. Suppose one has 9 people and one wants to divide them into one committee of 3, one committee of 4, and the last one of 2. There are $\binom{9}{3}$ ways of choosing the first committee. Once that is done, there are 6 people left and there are $\binom{6}{4}$ ways of choosing the second committee. Once that is done, the remainder must go in the third committee. So the answer is

$$\frac{9!}{3!6!}\frac{6!}{4!2!} = \frac{9!}{3!4!2!}.$$

Example 1.10. For any $k \leqslant n$ we have that

choosing k objects is the same as rejecting $n-k$ objects,

$$\binom{n}{k} = \binom{n}{n-k}.$$

Indeed, the left-hand side gives the number of different groups of k objects chosen from a total of n objects which is the same to choose $n-k$ objects not to be in the group of k objects which is the number on the right-hand side.

Combinations (multinomial coefficients)

The number of ways to divide n objects into one group of n_1 objects, one group of n_2, ..., and a rth group of n_r objects, where $n = n_1 + \cdots + n_r$, is equal to

$$\binom{n}{n_1, ..., n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}.$$

Example 1.11. Suppose we have 4 Americans and 6 Canadians.

(a) How many ways can we arrange them in a line?

(b) How many ways if all the Americans have to stand together?

(c) How many ways if not all the Americans are together?

(d) Suppose you want to choose a committee of 3, which will be all Americans or all Canadians. How many ways can this be done?

(e) How many ways for a committee of 3 that is not all Americans or all Canadians?

For (a) we can simply use the number of arrangements of 10 elements, that is, 10!.

For (b) we can consider the Americans as one group (element) and each Canadian as a distinct group (6 elements); this gives 7 distinct groups (elements) to be arranged, which can be done in 7! ways. Once we have these seven groups arranged, we can arrange the Americans within their group in 4! ways, so we get 4!7! by the basic counting principle.

In (c) the answer is the answer to (a) minus the answer to (b): $10! - 4!7!$

For (d) we can choose a committee of 3 Americans in $\binom{4}{3}$ ways and a committee of 3 Canadians in $\binom{6}{3}$ ways, so the answer is $\binom{4}{3} + \binom{6}{3}$.

Finally for (e) we can choose a committee of 3 out of 10 in $\binom{10}{3}$ ways, so the answer is $\binom{10}{3} - \binom{4}{3} - \binom{6}{3}$.

Finally, we consider three interrelated examples.

Example 1.12. First, suppose one has 8 copies of o and two copies of $|$. How many ways can one arrange these symbols in order? There are 10 spots, and we want to select 8 of them in which we place the os. So we have $\binom{10}{8}$.

Example 1.13. Next, suppose one has 8 indistinguishable balls. How many ways can one put them in 3 boxes? Let us use sequences of os and $|$ s to represent an arrangement of balls in these 3 boxes; any such sequence that has $|$ at each side, 2 other $|$ s, and 8 os represents a way of arranging balls into boxes. For example, if one has

$$|\quad o \quad o \quad | \quad o \quad o \quad o \quad | \quad o \quad o \quad o \quad |,$$

this would represent 2 balls in the first box, 3 in the second, and 3 in the third. Altogether there are $8 + 4$ symbols, the first is a $|$ as is the last, so there are 10 symbols that can be either | or *o*. Also, 8 of them must be *o*. How many ways out of 10 spaces can one pick 8 of them into which to put a *o*? We just did that, so the answer is $\binom{10}{8}$.

Example 1.14. Now, to finish, suppose we have \$8,000 to invest in 3 mutual funds. Each mutual fund required you to make investments in increments of \$1,000. How many ways can we do this? This is the same as putting 8 indistinguishable balls in 3 boxes, and we know the answer is $\binom{10}{8}$.

1.2. Further examples and explanations

1.2.1. Generalized counting principle. Here we expand on the basic counting principle formulated in Section 1.1.1. One can visualize this principle by using the box method below. Suppose we have two experiments to be performed, namely, one experiment can result in n outcomes, and the second experiment can result in m outcomes. Each box represents the number of possible outcomes in that experiment.

Experiment 1	Experiment 2	=	Experiment 1 and 2 together
m	**n**	=	**mn**

Example 1.15. There are 20 teachers and 100 students in a school. How many ways can we pick a teacher and student of the year? Using the box method we get $20 \times 100 = 2000$.

Generalized counting principle

Suppose that k experiments are to be performed, with the number of possible outcomes being n_i for the ith experiment. Then there are

$$n_1 \cdot ... \cdot n_k$$

possible outcomes of all k experiments together.

Example 1.16. A college planning committee consists of 3 freshmen, 4 sophomores, 5 juniors, and 2 seniors. A subcommittee of 4 consists of 1 person from each class. How many choices are possible? The counting principle or the box method gives $3 \times 4 \times 5 \times 2 = 120$.

Example 1.17 (Example 1.2 revisited)**.** Recall that for 6-place license plates, with the first three places occupied by letters and the last three by numbers, we have $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10$ choices. What if *no* repetition is allowed? We can use the counting principle or the box method to get $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8$.

Example 1.18. How many functions defined on k points are possible if each function can take values as either 0 or 1. The counting principle or the box method on the $1, \ldots, k$ points gives us 2^k possible functions. This is the generalized counting principle with $n_1 = n_2 = ... = n_k = 2$.

1.2.2. Permutations. Now we give more examples on permutations, and we start with a general results on the number of possible permutations, see also Combinations (multinomial coefficients) on page 6.

Permutations revisited

The number of different permutations of n objects of which n_1 are alike, n_2 are alike, ..., n_r are alike is equal to

$$\frac{n!}{n_1! \cdots n_r!}.$$

Example 1.19. How many ways can one arrange 5 math books, 6 chemistry books, 7 physics books, and 8 biology books on a bookshelf so that all the math books are together, all the chemistry books are together, and all the physics books are together.

We can arrange the math books in 5! ways, the chemistry in 6! ways, the physics in 7! ways, and biology books in 8! ways. We also have to decide which set of books go on the left, which next, and so on. That is the same as the number of ways of arranging the letters M,C,P, and B, and there are 4! ways of doing that. So the total is $4! \cdot (5! \cdot 6! \cdot 7! \cdot 8!)$ ways.

Now consider a couple of examples with *repetitions*.

Example 1.20. How many ways can one arrange the letters a, a, b, b, c, c?

Let us first re-label the letters by A, a, B, b, C, c. Then there are $6! = 720$ ways to arrange these letters. But we have repeats (for example, Aa or aA) which produce the same arrangement for the original letters. So dividing by the number of repeats for A, a, B, b and C, c, so the answer is

$$\frac{6!}{(2!)^3} = 90.$$

Example 1.21. How many different letter arrangements can be formed from the word PEPPER?

There are three copies of P and two copies of E, and one of R. So the answer is

$$\frac{6!}{3!2!1!} = 60.$$

Example 1.22. Suppose there are 4 Czech tennis players, 4 U.S. players, and 3 Russian players, in how many ways could they be arranged, if we do not distinguish players from the same country? By the formula above we get $\frac{11!}{4!4!3!}$.

1.2.3. Combinations.

Below are more examples on combinations.

Example 1.23. Suppose there are 9 men and 8 women. How many ways can we choose a committee that has 2 men and 3 women?

We can choose 2 men in $\binom{9}{2}$ ways and 3 women in $\binom{8}{3}$ ways. The number of committees is then the product

$$\binom{9}{2} \cdot \binom{8}{3}.$$

Example 1.24. Suppose somebody has n friends, of whom k are to be invited to a meeting.

(1) How many choices do exist for such a meeting if two of the friends will not attend together?
(2) How many choices do exist if two of the friends will only attend together?

We use a similar reasoning for both questions.

(1) We can divide all possible groups into two (disjoint) parts: one is for groups of friends none of which are these two, and another which includes exactly one of these two friends. There are $\binom{n-2}{k}$ groups in the first part, and $\binom{n-2}{k-1}$ in the second. For the latter we also need to account for a choice of one out of these two incompatible friends. So altogether we have

$$\binom{n-2}{k} + \binom{2}{1} \cdot \binom{n-2}{k-1}$$

(2) Again, we split all possible groups into two parts: one for groups which have none of the two inseparable friends, and the other for groups which include both of these two friends. Then

$$\binom{n-2}{k} + 1 \cdot 1 \cdot \binom{n-2}{k-2}.$$

Theorem 1.1 The binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

PROOF. We give two proofs.

First proof: let us expand the left-hand side $(x+y) \cdot \ldots \cdot (x+y)$. This is the sum of 2^n terms, and each term has n factors. For now we keep each product in the order we expanded the left-hand side, therefore we have all possible (finite) sequences of variables x and y, with the total power being n. We would like to collect all the terms having the same number of xs and ys.

Counting all the terms having k copies of x and $n-k$ copies of n is the same as asking in a sequence of n positions, how many ways can one choose k of them in which to put x. The answer is $\binom{n}{k}$ which gives the coefficient for $x^k y^{n-k}$. To illustrate it we take $k=2$ and $n=3$, then all possible terms are

$$x \cdot x \cdot y \qquad x \cdot y \cdot x \qquad y \cdot x \cdot x$$

Second proof: we will use (mathematical) induction on n. For $n = 1$ we have that the left-hand side is $x + y$, and the right-hand side

$$\sum_{k=0}^{1} \binom{1}{k} x^k y^{1-k} = \binom{1}{0} x^0 y^{1-0} + \binom{1}{1} x^1 y^{1-1}$$
$$= y + x = x + y,$$

so the statement holds for $n = 1$. Suppose now that the statement holds for $n = N$, we would like to show it for $n = N + 1$.

$$(x+y)^{N+1} = (x+y)(x+y)^N = (x+y)\sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k}$$
$$= x\sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k} + y\sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k}$$
$$= \sum_{k=0}^{N} \binom{N}{k} x^{k+1} y^{N-k} + \sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k+1}$$
$$= \sum_{k=1}^{N+1} \binom{N}{k-1} x^k y^{N-k+1} + \sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k+1},$$

where we replaced k by $k - 1$ in the first sum. Then we see that

$$(x+y)^{N+1} = \binom{N}{N} x^{N+1} y^0 + \sum_{k=1}^{N} \left(\binom{N}{k-1} + \binom{N}{k} \right) x^k y^{N-k+1} + \binom{N}{0} x^0 y^{N+1}$$
$$= x^{N+1} + \sum_{k=1}^{N} \left(\binom{N}{k-1} + \binom{N}{k} \right) x^k y^{N-k+1} + y^{N+1} = \sum_{k=0}^{N+1} \binom{N+1}{k}.$$

Here we used Example 1.26. □

Example 1.25. We can use combinatorics to show that

$$\tbinom{10}{4} = \tbinom{9}{3} + \tbinom{9}{4}$$

without evaluating these expressions explicitly.

Indeed, the left-hand side represents the number of committees consisting of 4 people out of the group of 10 people. Now we would like to represent the right-hand side. Let's say Tom Brady is one these ten people, and he might be in one of these committees and he is special, so we want to know when he will be there or not. When he is in the committee of 4, then there are $1 \cdot \binom{9}{3}$ number of ways of having a committee with Tom Brady as a member, while $\binom{9}{4}$ is the number of committees that do not have Tom Brady as a member. Adding it up gives us the number of committees of 4 people chosen out of the 10.

$$\binom{0}{0}$$
$$\binom{1}{0} \quad \binom{1}{1}$$
$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$
$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$
$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$
$$\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5}$$
$$\binom{6}{0} \quad \binom{6}{1} \quad \binom{6}{2} \quad \binom{6}{3} \quad \binom{6}{4} \quad \binom{6}{5} \quad \binom{6}{6}$$
$$\binom{7}{0} \quad \binom{7}{1} \quad \binom{7}{2} \quad \binom{7}{3} \quad \binom{7}{4} \quad \binom{7}{5} \quad \binom{7}{6} \quad \binom{7}{7}$$

Pascal's triangle

Example 1.26. The more general identity is

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

which can be proven either using the same argument or a formula for binomial coefficients.

Example 1.27. Expand $(x+y)^3$. This can be done by applying Theorem 1.1 (The binomial theorem) to get $(x+y)^3 = y^3 + 3xy^2 + 3x^2y + x^3$.

1.2.4. Multinomial coefficients.

Example 1.28. Suppose we are to assign 10 police officers: 6 patrols, 2 in station, 2 in schools. Then there are $\frac{10!}{6!2!2!}$ different assignments.

Example 1.29. We have 10 flags: 5 of them are blue, 3 are red, and 2 are yellow. These flags are indistinguishable, except for their color. Then there are $\frac{10!}{5!3!2!}$ many different ways we can order them on a flag pole.

Example 1.30 (Example 1.13 revisited)**.** Suppose one has n indistinguishable balls. How many ways can one put them in k boxes, assuming $n > k$?

As in Example 1.13 we use sequences of os and $|$ s to represent each an arrangement of balls in boxes; any such sequence that has $|$ at each side, $k-1$ copies of $|$ s, and n copies of os. How many different ways can we arrange this, if we have to start with $|$ and end with $|$? Between these, we are only arranging $n+k-1$ symbols, of which only n are os. So the question can be re-formulated as this: how many ways out of $n+k-1$ *spaces* can one pick n of *them* into which to put an o? This gives $\binom{n+k-1}{n}$. Note that this counts all possible ways including the ones when some of the boxes can be empty.

Suppose now we want to distribute n balls in k boxes so that none of the boxes are empty. Then we can line up n balls represented by os, instead of putting them in boxes we can place $|$ s in spaces between them. Note that we should have a $|$ on each side, as all balls have to be put to a box. So we are left with $k-1$ copies of $|$ s to be placed among n balls. This means that we have $n-1$ places, and we need to pick $k-1$ out of these to place $|$ s. So we can reformulate the problem as choose $k-1$ places out of $n-1$, and so the answer is $\binom{n-1}{k-1}$.

We can check that for $n=3$ and $k=2$ we indeed have 4 ways of distributing three balls in two boxes, and only two ways if every box has to have at least one ball.

1.3. Exercises

Exercise 1.1. Suppose a license plate must consist of 7 numbers or letters. How many license plates are there if

(A) there can only be letters?
(B) the first three places are numbers and the last four are letters?
(C) the first three places are numbers and the last four are letters, but there can not be any repetitions in the same license plate?

Exercise 1.2. A school of 50 students has awards for the top math, English, history and science student in the school

(A) How many ways can these awards be given if each student can only win one award?
(B) How many ways can these awards be given if students can win multiple awards?

Exercise 1.3. A password can be made up of any 4 digit combination.

(A) How many different passwords are possible?
(B) How many are possible if all the digits are odd?
(C) How many can be made in which all digits are different or all digits are the same?

Exercise 1.4. There is a school class of 25 people made up of 11 guys and 14 girls.

(A) How many ways are there to make a committee of 5 people?
(B) How many ways are there to pick a committee of all girls?
(C) How many ways are there to pick a committee of 3 girls and 2 guys?

Exercise 1.5. If a student council contains 10 people, how many ways are there to elect a president, a vice president, and a 3 person prom committee from the group of 10 students?

Exercise 1.6. Suppose you are organizing your textbooks on a book shelf. You have three chemistry books, 5 math books, 2 history books and 3 English books.

(A) How many ways can you order the textbooks if you must have math books first, English books second, chemistry third, and history fourth?
(B) How many ways can you order the books if each subject must be ordered together?

Exercise 1.7. If you buy a Powerball lottery ticket, you can choose 5 numbers between 1 and 59 (picked on white balls) and one number between 1 and 35 (picked on a red ball). How many ways can you

(A) win the jackpot (guess all the numbers correctly)?
(B) match all the white balls but not the red ball?
(C) match exactly 3 white balls and the red ball?
(D) match at least 3 white balls and the red ball?

Exercise 1.8. A couple wants to invite their friends to be in their wedding party. The groom has 8 possible groomsmen and the bride has 11 possible bridesmaids. The wedding party will consist of 5 groomsmen and 5 bridesmaids.

(A) How many wedding party's are possible?
(B) Suppose that two of the possible groomsmen are feuding and will only accept an invitation if the other one is not going. How many wedding parties are possible?
(C) Suppose that two of the possible bridesmaids are feuding and will only accept an invitation if the other one is not going. How many wedding parties are possible?
(D) Suppose that one possible groomsman and one possible bridesmaid refuse to serve together. How many wedding parties are possible?

Exercise 1.9. There are 52 cards in a standard deck of playing cards. The poker hand consists of five cards. How many poker hands are there?

Exercise 1.10. There are 30 people in a communications class. Each student must interview one another for a class project. How many total interviews will there be?

Exercise 1.11. Suppose a college basketball tournament consists of 64 teams playing head to head in a knockout style tournament. There are 6 rounds, the round of 64, round of 32, round of 16, round of 8, the final four teams, and the finals. Suppose you are filling out a bracket, such as this, which specifies which teams will win each game in each round.

How many possible brackets can you make?

Exercise 1.12. We need to choose a group of 3 women and 3 men out of 5 women and 6 men. In how many ways can we do it if 2 of the men refuse to be chosen together?

Exercise 1.13. Find the coefficient in front of x^4 in the expansion of $(2x^2 + 3y)^4$.

Exercise 1.14. In how many ways can you choose 2 or less (maybe none!) toppings for your ice-cream sundae if 6 different toppings are available? (You can use combinations here, but you do not have to. Next, try to find a general formula to compute in how many ways you can choose k or less toppings if n different toppings are available

Exercise* 1.1. Use the binomial theorem to show that

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n,$$
$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0.$$

Exercise* 1.2. Prove the multinomial theorem

$$(x_1 + ... + x_k)^n = \sum_{\substack{(n_1,\cdots,n_k)\\ n_1+\cdots+n_k=n}} \binom{n}{n_1,...,n_k} x_1^{n_1} \cdot ... \cdot x_k^{n_k}.$$

Exercise* 1.3. Show that there are $\binom{n-1}{k-1}$ distinct positive integer-valued vectors $(x_1, ..., x_k)$ satisfying

$$x_1 + ... + x_k = n, x_i > 0 \text{ for all } i = 1, ..., k.$$

Exercise* 1.4. Show that there are $\binom{n+k-1}{k-1}$ distinct non-positive integer-valued vectors $(x_1, ..., x_k)$ satisfying

$$x_1 + ... + x_k = n, x_i \geqslant 0 \text{ for all } i = 1, ..., k.$$

Exercise* 1.5. Consider a smooth function of n variables. How many different partial derivatives of order k does f possess?

1.4. Selected solutions

Solution to Exercise 1.1(A): in each of the seven places we can put any of the 26 letters giving

$$26^7$$

possible letter combinations.

Solution to Exercise 1.1(B): in each of the first three places we can place any of the 10 digits, and in each of the last four places we can put any of the 26 letters giving a total of

$$10^3 \cdot 26^4.$$

Solution to Exercise 1.1(C): if we can not repeat a letter or a number on a license plate, then the number of license plates becomes

$$(10 \cdot 9 \cdot 8) \cdot (\cdot 26 \cdot 25 \cdot 24 \cdot 23)\,.$$

Solution to Exercise 1.2(A):

$$50 \cdot 49 \cdot 48 \cdot 47$$

Solution to Exercise 1.2(B):

$$50^4$$

Solution to Exercise 1.3(A):

$$10^4$$

Solution to Exercise 1.3(B):

$$5^4$$

Solution to Exercise 1.3(C):

$$10 \cdot 9 \cdot 8 \cdot 7 + 10$$

Solution to Exercise 1.4(A):

$$\binom{25}{5}$$

Solution to Exercise 1.4(B):

$$\binom{14}{5}$$

Solution to Exercise 1.4(C):

$$\binom{14}{3} \cdot \binom{11}{2}$$

Solution to Exercise 1.5:

$$10 \cdot 9 \cdot \binom{8}{3}$$

Solution to Exercise 1.6(A):

$$5!3!3!2!$$

Solution to Exercise 1.6(B):

$$4!\,(5!3!3!2!)$$

Solution to Exercise 1.7(A):

$$1$$

Solution to Exercise 1.7(B):

$$1 \cdot 34$$

Solution to Exercise 1.7(C):

$$\binom{5}{3} \cdot \binom{54}{2} \cdot \binom{1}{1}$$

Solution to Exercise 1.7(D):

$$\binom{5}{3} \cdot \binom{54}{2} \cdot \binom{1}{1} + \binom{5}{4} \cdot \binom{54}{1} \cdot \binom{1}{1} + 1$$

Solution to Exercise 1.8(A):

$$\binom{8}{5} \cdot \binom{11}{5}$$

Solution to Exercise 1.8(B):

$$\binom{6}{5} \cdot \binom{11}{5} + \binom{2}{1} \cdot \binom{6}{4} \cdot \binom{11}{5}$$

Solution to Exercise 1.8(C):

$$\binom{8}{5} \cdot \binom{9}{5} + \binom{8}{5} \cdot \binom{2}{1} \cdot \binom{9}{4}$$

Solution to Exercise 1.8(D):

$$\binom{7}{5} \cdot \binom{10}{5} + 1 \cdot \binom{7}{4} \cdot \binom{10}{5} + \binom{7}{5} \cdot 1 \cdot \binom{10}{4}$$

Solution to Exercise 1.9:

$$\binom{52}{5}$$

Solution to Exercise 1.10:

$$\binom{30}{2}$$

Solution to Exercise 1.11: First notice that the 64 teams play 63 total games: 32 games in the first round, 16 in the second round, 8 in the 3rd round, 4 in the regional finals, 2 in the final four, and then the national championship game. That is, $32+16+8+4+2+1=63$. Since there are 63 games to be played, and you have two choices at each stage in your bracket, there are 2^{63} different ways to fill out the bracket. That is,

$$2^{63} = 9,223,372,036,854,775,808.$$

Solution to Exercise* 1.1: use the binomial formula

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

with $x = y = 1$ to see

$$2^n = (1+1)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot 1^k \cdot 1^{n-k} = \sum_{k=0}^{n} \binom{n}{k},$$

and with $x = -1$, $y = 1$

$$0 = (-1+1)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot (-1)^k \cdot (1)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (-1)^k .$$

Solution to Exercise* 1.2: we can prove the statement using mathematical induction on k. For $k = 1$ we have

$$(x_1)^n = \sum_{n_1=n} \binom{n}{n_1} x_1 = x_1^n,$$

which is true; for $k = 2$ we have

$$(x_1 + x_2)^n = \sum_{\substack{(n_1,n_2)\\ n_1+n_2=n}} \binom{n}{n_1,n_2} x_1^{n_1} \cdot x_2^{n_2} = \sum_{n_1=0}^{n} \binom{n}{n_1} x_1^{n_1} \cdot x_2^{n-n_1},$$

which is the binomial formula itself. Now suppose the multinomial formula holds for $k = K$ (*induction hypothesis*), that is,

$$(x_1 + ... + x_K)^n = \sum_{\substack{(n_1,\cdots,n_K)\\ n_1+\cdots+n_K=n}} \binom{n}{n_1,...,n_K} \cdot x_1^{n_1} \cdot ... \cdot x_K^{n_K},$$

and we need to show

$$(x_1 + ... + x_{K+1})^n = \sum_{\substack{(n_1,\cdots,n_{K+1})\\ n_1+\cdots+n_{K+1}=n}} \binom{n}{n_1,...,n_{K+1}} \cdot x_1^{n_1} \cdot ... \cdot x_{K+1}^{n_{K+1}}.$$

Denote

$$y_1 = x_1, .., y_{K-1} := x_{K-1}, y_K := x_K + x_{K+1},$$

then by the induction hypothesis

$$(x_1 + ... + x_{K+1})^n = (y_1 + ... + y_K)^n = \sum_{\substack{(n_1,\cdots,n_K)\\ n_1+\cdots+n_K=n}} \binom{n}{n_1,...,n_K} \cdot y_1^{n_1} \cdot ... \cdot y_{K-1}^{n_{K-1}} \cdot y_K^{n_K}$$

$$= \sum_{\substack{(n_1,\cdots,n_K)\\ n_1+\cdots+n_K=n}} \binom{n}{n_1,...,n_K} \cdot x_1^{n_1} \cdot ... \cdot x_{K-1}^{n_{K-1}} \cdot (x_K + x_{K+1})^{n_K} .$$

By the binomial formula

$$(x_K + x_{K+1})^{n_K} = \sum_{m=1}^{n_K} \binom{n_K}{m} \cdot x_K^m \cdot x_{K+1}^{n_K-m},$$

therefore

$$(x_1 + ... + x_{K+1})^n = \sum_{\substack{(n_1,\cdots,n_K) \\ n_1+\cdots+n_K=n}} \binom{n}{n_1,\dots,n_K} \cdot x_1^{n_1} \cdot \dots \cdot x_{K-1}^{n_{K-1}} \cdot \sum_{m=1}^{n_K} \binom{n_K}{m} \cdot x_K^m \cdot x_{K+1}^{n_K - m}.$$

It is easy to see (using the definition of multinomial coefficients) that

$$\binom{n}{n_1,\dots,n_K}\binom{n_K}{m} = \binom{n}{n_1,\dots,n_K,m}, n_1 + ... + n_K + m = n.$$

Indeed,

$$\binom{n}{n_1,\dots,n_K}\binom{n_K}{m} = \frac{n!}{n_1!n_2! \cdot \dots \cdot n_{K-1}! \cdot n_K!} \frac{n_K!}{m!\,(n_K - m)!}$$
$$= \frac{n!}{n_1!n_2! \cdot \dots \cdot n_{K-1}! \cdot m!\,(n_K - m)!} = \binom{n}{n_1,\dots,n_K,m}.$$

Thus

$$(x_1 + ... + x_{K+1})^n = \sum_{\substack{(n_1,\cdots,n_K) \\ n_1+\cdots+n_K=n}} \sum_{m=1}^{n_K} \binom{n}{n_1,\dots,n_K,m} \cdot x_1^{n_1} \cdot \dots \cdot x_{K-1}^{n_{K-1}} \cdot \cdot x_K^m \cdot x_{K+1}^{n_K - m}.$$

Note that $n_K = m + (n_K - m)$, so if we denote $m_1 := n_1, m_2 := n_2, ..., m_{K-1} := n_{K-1}, m_K := m, m_{K+1} := n_K - m$ then we see that

$$(x_1 + ... + x_{K+1})^n = \sum_{\substack{(m_1,\cdots,m_K,m_{K+1}) \\ m_1+\cdots+m_{K+1}=n}} \binom{n}{m_1,\dots,m_K,m_{K+1}} \cdot x_1^{m_1} \cdot \dots \cdot x_{K-1}^{m_{K-1}} \cdot x_K^{m_K} \cdot x_{K+1}^{m_{K+1}}$$

which is what we wanted to show.

Solution to Exercise* 1.3: this is the same problem as dividing n indistinguishable balls into k boxes in such a way that each box has at least one ball. To do so, you can select $k-1$ of the $n-1$ spaces between the objects. There are $\binom{n-1}{k-1}$ possible selections that is equal to the number of possible positive integer solutions to the equation.

Solution to Exercise* 1.4: define $y_i := x_i + 1$ and apply the previous problem.

Solution to Exercise* 1.5: the same answer as in the previous problem.

CHAPTER 2

The probability set-up

2.1. Basic theory of probability

We will have a *sample space*, denoted by S (sometimes Ω) that consists of all possible outcomes. For example, if we roll two dice, the sample space would be all possible pairs made up of the numbers one through six. An *event* is a subset of S.

Another example is to toss a coin 2 times, and let

$$S = \{HH, HT, TH, TT\};$$

or to let S be the possible orders in which 5 horses finish in a horse race; or S the possible prices of some stock at closing time today; or $S = [0, \infty)$; the age at which someone dies; or S the points in a circle, the possible places a dart can hit. We should also keep in mind that the same setting can be described using different sample set. For example, in two solutions in Example 1.30 we used two different sample sets.

2.1.1. Sets. We start by describing elementary operations on sets. By a *set* we mean a collection of distinct objects called *elements of the set*, and we consider a set as an object in its own right.

Set operations

Suppose S is a set. We say that $A \subset S$, that is, A is a *subset* of S if every element in A is contained in S;
$A \cup B$ is the *union* of sets $A \subset S$ and $B \subset S$ and denotes the points of S that are in A or B or both;
$A \cap B$ is the *intersection* of sets $A \subset S$ and $B \subset S$ and is the set of points that are in both A and B;
$\emptyset$ denotes the *empty set*;
A^c is the *complement* of A, that is, the points in S that are not in A.

We extend this definition to have $\cup_{i=1}^n A_i$ is the union of sets $A_1, \cdots, A_n$, and similarly $\cap_{i=1}^n A_i$.

An exercise is to show that

De Morgan's laws

$$\left(\bigcup_{i=1}^n A_i\right)^c = \bigcap_{i=1}^n A_i^c \quad \text{and} \quad \left(\bigcap_{i=1}^n A_i\right)^c = \bigcup_{i=1}^n A_i^c.$$

 Typesetting date: September 2, 2020

We will also need the notion of *pairwise disjoint sets* $\{E_i\}_{i=1}^{\infty}$ which means that $E_i \cap E_j = \emptyset$ unless $i = j$.

There are no restrictions on the sample space S. The collection of events, $\mathcal{F}$, is assumed to be a σ-field, which means that it satisfies the following.

Definition (σ-field)

A collection $\mathcal{F}$ of sets in S is called a *σ-field* if
(i) Both $\emptyset, S$ are in $\mathcal{F}$,
(ii) if A is in $\mathcal{F}$, then A^c is in $\mathcal{F}$,
(iii) if $A_1, A_2, \ldots$ are in $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i$ and $\bigcap_{i=1}^{\infty} A_i$ are in $\mathcal{F}$.

Typically we will take $\mathcal{F}$ to be all subsets of S, and so (i)-(iii) are automatically satisfied. The only times we won't have $\mathcal{F}$ be all subsets is for technical reasons or when we talk about conditional expectation.

2.1.2. Probability axioms. So now we have a sample space S, a σ-field $\mathcal{F}$, and we need to talk about what a probability is.

Probability axioms

(1) $0 \leqslant \mathbb{P}(E) \leqslant 1$ for all events $E \in \mathcal{F}$.
(2) $\mathbb{P}(S) = 1$.
(3) If the $\{E_i\}_{i=1}^{\infty}, E_i \in \mathcal{F}$ are pairwise disjoint, $\mathbb{P}(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$.

Note that probabilities are probabilities of subsets of S, not of points of S. However, it is common to write $\mathbb{P}(x)$ for $\mathbb{P}(\{x\})$.

Intuitively, the probability of E should be the number of times E occurs in n experiments, taking a limit as n tends to infinity. This is hard to use. It is better to start with these axioms, and then to prove that the probability of E is the limit as we hoped.

Below are some easy consequences of the probability axioms.

Proposition 2.1 (Properties of probability)

(1) $\mathbb{P}(\emptyset) = 0$.
(2) If $E_1, \ldots, E_n \in \mathcal{F}$ are pairwise disjoint, then $\mathbb{P}(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} \mathbb{P}(E_i)$.
(3) $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$ for any $E \in \mathcal{F}$.
(4) If $E \subset F$, then $\mathbb{P}(E) \leqslant \mathbb{P}(F)$, for any $E, F \in \mathcal{F}$.
(5) for any $E, F \in \mathcal{F}$

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F). \tag{2.1.1}$$

The last property is sometimes called the *inclusion-exclusion identity*.

PROOF. To show (1), choose $E_i = \emptyset$ for each i. These are clearly disjoint, so $\mathbb{P}(\emptyset) = \mathbb{P}(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mathbb{P}(E_i) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$. If $\mathbb{P}(\emptyset)$ were strictly positive, then the last term would be infinity, contradicting the fact that probabilities are between 0 and 1. So the probability of $\emptyset$ must be zero.

Part (2) follows if we let $E_{n+1} = E_{n+2} = \cdots = \emptyset$. Then $\{E_i\}_{i=1}^{\infty}, E_i \in \mathcal{F}$ are still pairwise disjoint, and $\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{n} E_i$, and by (1) we have

$$\sum_{i=1}^{\infty} \mathbb{P}(E_i) = \sum_{i=1}^{n} \mathbb{P}(E_i).$$

To prove (3), use $S = E \cup E^c$. By (2), $\mathbb{P}(S) = \mathbb{P}(E) + \mathbb{P}(E^c)$. By axiom (2), $\mathbb{P}(S) = 1$, so (1) follows.

To prove (4), write $F = E \cup (F \cap E^c)$, so $\mathbb{P}(F) = \mathbb{P}(E) + \mathbb{P}(F \cap E^c) \geqslant \mathbb{P}(E)$ by (2) and axiom (1).

Similarly, to prove (5), we have $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(E^c \cap F)$ and $\mathbb{P}(F) = \mathbb{P}(E \cap F) + \mathbb{P}(E^c \cap F)$. Solving the second equation for $\mathbb{P}(E^c \cap F)$ and substituting in the first gives the desired result. □

It is common for a probability space to consist of finitely many points, all with equally likely probabilities. For example, in tossing a fair coin, we have $S = \{H, T\}$, with $\mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2}$. Similarly, in rolling a fair die, the probability space consists of $\{1, 2, 3, 4, 5, 6\}$, each point having probability $\frac{1}{6}$.

Example 2.1. What is the probability that if we roll 2 dice, the sum is 7?

There are 36 possibilities, of which 6 have a sum of 7: $(1, 6)$, $(2, 5)$, $(3, 4)$, $(4, 3)$, $(5, 2)$, $(6, 1)$. Since they are all equally likely, the probability is $\frac{6}{36} = \frac{1}{6}$.

Example 2.2. What is the probability that in a poker hand (5 cards out of 52) we get exactly 4 of a kind?

We have four suits: clubs, diamonds, hearts and spades. Each suit includes an ace, a king, queen and jack, and ranks two through ten.

For example, the probability of 4 aces and 1 king is

$$\frac{\binom{4}{4}\binom{4}{1}}{\binom{52}{5}}.$$

The probability of 4 jacks and one 3 is the same. There are 13 ways to pick the rank that we have 4 of, and then 12 ways to pick the rank we have one of, so the answer is

$$13 \cdot 12 \cdot \frac{\binom{4}{4}\binom{4}{1}}{\binom{52}{5}}.$$

Example 2.3. What is the probability that in a poker hand we get exactly3 of a kind (and the other two cards are of different ranks)?

For example, the probability of 3 aces, 1 king and 1 queen is

$$\frac{\binom{4}{3}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}}.$$

We have 13 choices for the rank we have 3 of, and $\binom{12}{2}$ choices for the other two ranks, so the answer is

$$13\binom{12}{2}\frac{\binom{4}{3}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}}.$$

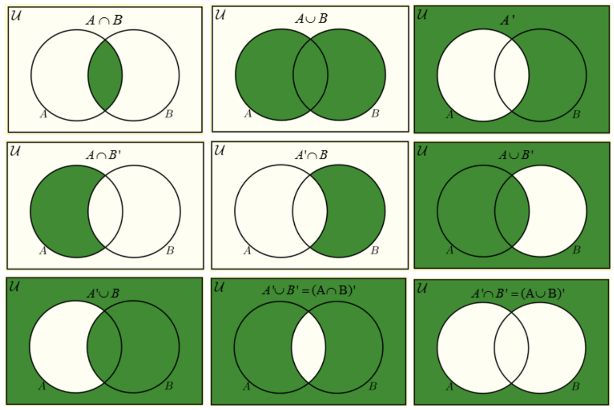
Example 2.4. In a class of 30 people, what is the probability everyone has a different birthday? (We assume each day is equally likely.)

We assume that it is not a leap year. Let the first person have a birthday on some day. The probability that the second person has a different birthday will be $\frac{364}{365}$. The probability that the third person has a different birthday from the first two people is $\frac{363}{365}$. So the answer is

$$\frac{364}{365}\cdot\frac{363}{365}\cdot\ldots\cdot\frac{336}{365}.$$

2.2. Further examples and applications

2.2.1. Sets revisited. A visual way to represent set operations is given by the Venn diagrams.



A picture of Venn diagrams from

`http://www.onlinemathlearning.com/shading-venn-diagrams.html`

Example 2.5. Roll two dice. We can describe the sample set S as *ordered pairs* of numbers $1, 2, ..., 6$, that is, S has 36 elements. Examples of events are

$$\begin{aligned}
&E = \text{ the two dice come up equal and even } = \{(2,2),(4,4),(6,6)\},\\
&F = \text{ the sum of the two dice is } 8 = \{(2,6),(3,5),(4,4),(5,3),(6,2)\},\\
&E \cup F = \{(2,2),(2,6),(3,5),(4,4),(5,3),(6,2),(6,6)\},\\
&E \cap F = \{(4,4)\},\\
&F^c = \text{ all 31 pairs that do not include } \{(2,6),(3,5),(4,4),(5,3),(6,2)\}.
\end{aligned}$$

Example 2.6. Let $S = [0, \infty)$ be the space of all possible ages at which someone can die. Possible events are

$A =$ person dies before reaching $30 = [0, 30)$.
$A^c = [30, \infty) =$ person dies after turning 30.
$A \cup A^c = S$,
$B =$ a person lives either less than 15 or more than 45 years $= (15, 45]$.

2.2.2. Axioms of probability revisited.

Example 2.7 (Coin tosses)**.** In this case $S = \{H, T\}$, where H stands for heads, and T stands for tails. We say that a coin is fair if we toss a coin with each side being equally likely, that is,

$$\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}.$$

We may write $\mathbb{P}(H) = \mathbb{P}(T) = \frac{1}{2}$. However, if the coin is biased, then still $S = \{H, T\}$ but each side can be assigned a different probability, for instance

$$\mathbb{P}(H) = \frac{2}{3}, \mathbb{P}(T) = \frac{1}{3}.$$

Example 2.8. Rolling a fair die, the probability of getting an even number is

$$\mathbb{P}(\{\text{even}\}) = \mathbb{P}(2) + \mathbb{P}(4) + \mathbb{P}(6) = \frac{1}{2}.$$

Let us see how we can use properties of probability in Proposition 2.1 to solve problems.

Example 2.9. UConn Basketball is playing Kentucky this year and from past experience the following is known:

- a home game has 0.5 chance of winning;
- an away game has 0.4 chance of winning;
- there is a 0.3 chance that UConn wins both games.

What is probability that UConn loses both games?

Let us denote by A_1 the event of a *home game win*, and by A_2 an *away game win*. Then, from the past experience we know that $\mathbb{P}(A_1) = 0.5$, $\mathbb{P}(A_2) = 0.4$ and $\mathbb{P}(A_1 \cap A_2) = 0.3$. Notice that the event *UConn loses both games* can be expressed as $A_1^c \cap A_2^c$. Thus we want to find $\mathbb{P}(A_1^c \cap A_2^c)$. Use De Morgan's laws and (3) in Proposition 2.1 we have

$$\mathbb{P}(A_1^c \cap A_2^c) = \mathbb{P}((A_1 \cup A_2)^c) = 1 - \mathbb{P}(A_1 \cup A_2).$$

The inclusion-exclusion identity (2.1.1) tells us

$$\mathbb{P}(A_1 \cup A_2) = 0.5 + 0.4 - 0.3 = 0.6,$$

and hence $\mathbb{P}(A_1^c \cap A_2^c) = 1 - 0.6 = 0.4$.

The inclusion-exclusion identity is actually true for any finite number of events. To illustrate this, we give next the formula in the case of three events.

Proposition 2.2 (Inclusion-exclusion identity)

For any three events $A, B, C \in \mathcal{F}$ in the sample state S

$$\begin{aligned} \mathbb{P}(A \cup B \cup C) = {} & \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ & - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C). \end{aligned} \tag{2.2.1}$$

2.2.3. Uniform discrete distribution. If in an experiment the probability space consists of finitely many points, all with equally likely probabilities, the probability of any given event has the following simple expression.

Uniform discrete distribution

The probability of an event $E \in \mathcal{F}$ in the sample state S is given by

$$\mathbb{P}(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S}.$$

To show this formula rigorously we can start by considering an event E consisting of exactly one element, and use axioms of probability to see that this formula holds for such an E. Then we can represent any event E as a disjoint union of one-element events to prove the statement.

Example 2.10. A committee of 5 people is to be selected from a group of 6 men and 9 women. What is probability that it consists of 3 men and 2 women?

In this case, in counting the ways to select a group with 3 men and 2 women the order is irrelevant. We have

$$\mathbb{P}(E) = \frac{\text{the number of groups with 3 men and 2 women}}{\text{the number of groups of 5}} = \frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}} = \frac{240}{1001}.$$

Many experiments can be modeled by considering a set of balls from which some will be withdrawn. There are two basic ways of withdrawing, namely *with* or *without* replacement.

Example 2.11. Three balls are randomly withdrawn without replacement from a bowl containing 6 white and 5 black balls. What is the probability that one ball is white and the other two are black?

This is a good example of the situation where a choice of the sample space might be different.

First solution: we model the experiment so that the order in which the balls are drawn is important. That is, we can describe the sample state S as *ordered* triples of letters W and B. Then

$$\begin{aligned} P(E) &= \frac{WBB + BWB + BBW}{11 \cdot 10 \cdot 9} \\ &= \frac{6 \cdot 5 \cdot 4 + 5 \cdot 6 \cdot 4 + 5 \cdot 4 \cdot 6}{990} = \frac{120 + 120 + 120}{990} = \frac{4}{11}. \end{aligned}$$

Second solution: we model the experiment so that the order in which the balls are drawn is *not* important. In this case

$$P(E) = \frac{(\text{one ball is white})\,(\text{two balls are black})}{\binom{11}{3}} = \frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{4}{11}.$$

2.3. Exercises

Exercise 2.1. Consider a box that contains 3 balls: 1 red, 1 green, and 1 yellow.

(A) Consider an experiment that consists of taking 1 ball from the box, placing it back in the box, and then drawing a second ball from the box. List all possible outcomes.
(B) Repeat the experiment but now, after drawing the first ball, the second ball is drawn from the box without replacing the first. List all possible outcomes.

Exercise 2.2. Suppose that A and B are pairwise disjoint events for which $\mathbb{P}(A) = 0.2$ and $\mathbb{P}(B) = 0.4$.

(A) What is the probability that B occurs but A does not?
(B) What is the probability that neither A nor B occurs?

Exercise 2.3. Forty percent of the students at a certain college are members neither of an academic club nor of a Greek organization. Fifty percent are members of an academic club and thirty percent are members of a Greek organization. What is the probability that a randomly chosen student is

(A) member of an academic club or a Greek organization?
(B) member of an academic club and of a Greek organization?

Exercise 2.4. In a city, 60% of the households subscribe to newspaper A, 50% to newspaper B, 40% to newspaper C, 30% to A and B, 20% to B and C, and 10% to A and C. None subscribe to all three.

(A) What percentage subscribe to exactly one newspaper?(Hint: Draw a Venn diagram)
(B) What percentage subscribe to at most one newspaper?

Exercise 2.5. There are 52 cards in a standard deck of playing cards. There are 4 *suits*: hearts, spades, diamonds, and clubs ($\heartsuit\spadesuit\diamondsuit\clubsuit$). Hearts and diamonds are red while spades and clubs are black. In each suit there are 13 *ranks*: the numbers $2, 3 \ldots, 10$, the three face cards, Jack, Queen, King, and the Ace. Note that Ace is not a face card. A *poker hand* consists of five cards. Find the probability of randomly drawing the following poker hands.

(A) All 5 cards are red?
(B) Exactly two 10s and exactly three Aces?
(C) all 5 cards are either face cards or no-face cards?

Exercise 2.6. Find the probability of randomly drawing the following poker hands.

(A) A *one pair*, which consists of two cards of the same rank and three other distinct ranks. (e.g. 22Q59)
(B) A *two pair*, which consists of two cards of the same rank, two cards of another rank, and another card of yet another rank. (e.g.JJ779)
(C) A *three of a kind*, which consists of a three cards of the same rank, and two others of distinct rank (e.g. 4449K).

(D) A *flush*, which consists of all five cards of the same suit (e.g. HHHH, SSSS, DDDD, or CCCC).
(E) A *full house*, which consists of a two pair and a three of a kind (e.g. 88844). (Hint: Note that 88844 is a different hand than a 44488.)

Exercise 2.7. Suppose a standard deck of cards is modified with the additional rank of *Super King* and the additional suit of *Swords* so now each card has one of 14 ranks and one of 5 suits. What is the probability of

(A) selecting the *Super King of Swords*?
(B) getting a six card hand with exactly three pairs (two cards of one rank and two cards of another rank and two cards of yet another rank, e.g. 7,7,2,2,J,J) ?
(C) getting a six card hand which consists of three cards of the same rank, two cards of another rank, and another card of yet another rank. (e.g. 3,3,3,A,A,7)?

Exercise 2.8. A pair of fair dice is rolled. What is the probability that the first die lands on a strictly higher value than the second die?

Exercise 2.9. In a seminar attended by 8 students, what is the probability that at least two of them have birthday in the same month?

Exercise 2.10. Nine balls are randomly withdrawn without replacement from an urn that contains 10 blue, 12 red, and 15 green balls. What is the probability that

(A) 2 blue, 5 red, and 2 green balls are withdrawn?
(B) at least 2 blue balls are withdrawn?

Exercise 2.11. Suppose 4 valedictorians from different high schools are accepted to the 8 Ivy League universities. What is the probability that each of them chooses to go to a different Ivy League university?

Exercise 2.12. Two dice are thrown. Let E be the event that the sum of the dice is even, and F be the event that at least one of the dice lands on 2. Describe EF and $E \bigcup F$.

Exercise 2.13. If there are 8 people in a room, what is the probability that no two of them celebrate their birthday in the same month?

Exercise 2.14. Box I contains 3 red and 2 black balls. Box II contains 2 red and 8 black balls. A coin is tossed. If H, then a ball from box I is chosen; if T, then from from box II.

(1) What is the probability that a red ball is chosen?
(2) Suppose now the person tossing the coin does not reveal if it has turned H or T. If a red ball was chosen, what is the probability that it was box I (that is, H)?

Exercise* 2.1. Prove Proposition 2.2 by grouping $A \cup B \cup C$ as $A \cup (B \cup C)$ and using the Equation (2.1.1) for two sets.

2.4. Selected solutions

Solution to Exercise 2.1(A): Since every marble can be drawn first and every marble can be drawn second, there are $3^2 = 9$ possibilities: RR, RG, RB, GR, GG, GB, BR, BG, and BB (we let the first letter of the color of the drawn marble represent the draw).

Solution to Exercise 2.1(B): In this case, the color of the second marble cannot match the color of the rest, so there are 6 possibilities: RG, RB, GR, GB, BR, and BG.

Solution to Exercise 2.2(A): Since $A \cap B = \emptyset$, $B \subseteq A^c$ hence $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) = 0.4$.

Solution to Exercise 2.2(B): By De Morgan's laws and property (3) of Proposition 2.1,

$$\mathbb{P}(A^c \cap B^c) = \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B) = 1 - (\mathbb{P}(A) + \mathbb{P}(B)) = 0.4.$$

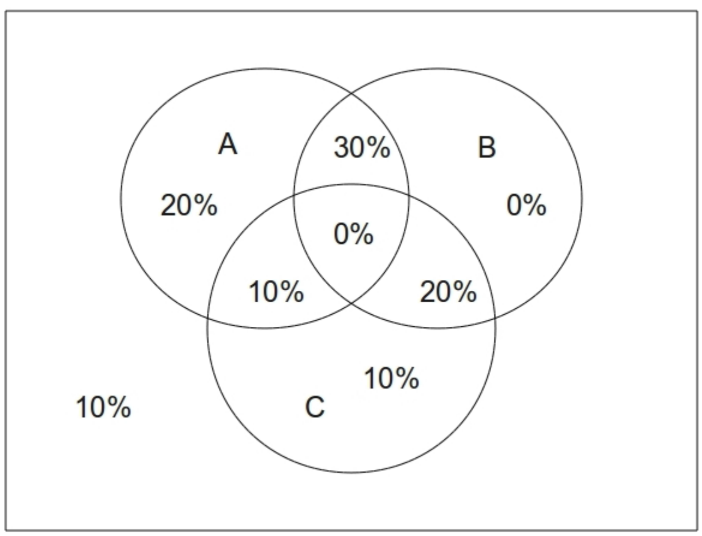
Solution to Exercise 2.3(A): $\mathbb{P}(A \cup B) = 1 - .4 = .6$

Solution to Exercise 2.3(B): Notice that

$$0.6 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.5 + 0.3 - \mathbb{P}(A \cap B)$$

Thus, $\mathbb{P}(A \cap B) = 0.2$.

Solution to Exercise 2.4(A): We use these percentages to produce the Venn diagram below:



This tells us that 30% of households subscribe to exactly one paper.

Solution to Exercise 2.4(B): The Venn diagram tells us that $100\% - (10\% + 20\% + 30\%) = 40\%$ of the households subscribe to at most one paper.

Solution to Exercise 2.5(A): $\dfrac{\binom{26}{5}}{\binom{52}{5}}$.

Solution to Exercise 2.5(B): $\dfrac{\binom{4}{2} \cdot \binom{4}{3}}{\binom{52}{5}}$

Solution to Exercise 2.5(C): $\frac{\binom{12}{5}}{\binom{52}{5}} + \frac{\binom{40}{5}}{\binom{52}{5}}$

Solution to Exercise 2.6(A): $\frac{13\cdot\binom{4}{2}\cdot\binom{12}{3}\cdot\binom{4}{1}\cdot\binom{4}{1}\cdot\binom{4}{1})}{\binom{52}{5}}$

Solution to Exercise 2.6(B): $\frac{\binom{13}{2}\cdot\binom{4}{2}\cdot\binom{4}{2}\cdot\binom{44}{1}}{\binom{52}{5}}$

Solution to Exercise 2.6(C): $\frac{13\cdot\binom{4}{3}\cdot\binom{12}{2}\cdot\binom{4}{1}\cdot\binom{4}{1}}{\binom{52}{5}}$

Solution to Exercise 2.6(D): $\frac{4\cdot\binom{13}{5}}{\binom{52}{5}}$

Solution to Exercise 2.6(E): $\frac{13\cdot12\cdot\binom{4}{3}\cdot\binom{4}{2}}{\binom{52}{5}}$

Solution to Exercise 2.7(A): $\frac{1}{70}$

Solution to Exercise 2.7(B): $\frac{\binom{14}{3}\cdot\binom{5}{2}\cdot\binom{5}{2}\cdot\binom{5}{2}}{\binom{70}{6}}$

Solution to Exercise 2.7(C): $\frac{14\cdot\binom{5}{3}\cdot13\cdot\binom{5}{2}\cdot12\cdot\binom{5}{1}}{\binom{70}{6}}$

Solution to Exercise 2.8: we can simply list all possibilities

$$\begin{array}{rl} (6,1)\,,\ (6,2)\,,\ (6,3),\ (6,4)\,,\ (6,5) & \text{5 possibilities} \\ (5,1)\,,\ (5,2),\ (5,3)\,,\ (5,4) & \text{4 possibilities} \\ (4,1)\,,\ (4,2),\ (4,3) & \text{3 possibilities} \\ (3,1)\,,\ (3,2) & \text{2 possibilities} \\ (2,1) & \text{1 possibility} \\ & = \text{15 possibilities in total} \end{array}$$

Thus the probability is $\frac{15}{36}$.

Solution to Exercise 2.9:

$$1 - \frac{12\cdot11\cdot10\cdot9\cdot8\cdot7\cdot6\cdot5}{12^8}$$

Solution to Exercise 2.10(A): $\frac{\binom{10}{2}\cdot\binom{12}{5}\cdot\binom{15}{2}}{\binom{37}{9}}$

Solution to Exercise 2.10(B):

$$1 - \frac{\binom{27}{9}}{\binom{37}{9}} - \frac{\binom{10}{1}\cdot\binom{27}{8}}{\binom{37}{9}}$$

Solution to Exercise 2.11:

$$\frac{8\cdot7\cdot6\cdot5}{8^4}$$

Solution to Exercise* 2.1: to prove Proposition 2.2 we will use Equation (2.1.1) several times, as well as a distribution law for sets. First, we apply Equation (2.1.1) to two sets A and $B \cup C$ to see that

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

We can now apply Equation (2.1.1) to the sets B and C to see that

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C) - \mathbb{P}(A \cap (B \cup C)). \tag{2.4.1}$$

Then the distribution law for sets says

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

which we can see by using the Venn diagrams. Now we can apply Equation (2.1.1) to the sets $(A \cap B)$ and $(A \cap C)$ to see that

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

Finally observe that

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

so

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

Use this in Equation (2.4.1) to finish the proof.

CHAPTER 3

Independence

3.1. Independent events

Suppose we have a probability space of a sample space S, σ-field $\mathcal{F}$ and probability $\mathbb{P}$ defined on $\mathcal{F}$.

Definition (Independence)

We say that $E, F \in \mathcal{F}$ are *independent events* if

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\,\mathbb{P}(F)$$

Example 3.1. Suppose you flip two coins. The outcome of heads on the second is independent of the outcome of tails on the first. To be more precise, if A is tails for the first coin and B is heads for the second, and we assume we have fair coins (although this is not necessary), we have $\mathbb{P}(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A)\mathbb{P}(B)$.

Example 3.2. Suppose you draw a card from an ordinary deck. Let E be you drew an ace, F be that you drew a spade. Here $\frac{1}{52} = \mathbb{P}(E \cap F) = \frac{1}{13} \cdot \frac{1}{4} = \mathbb{P}(E) \cap \mathbb{P}(F)$.

Proposition 3.1

If E and F are independent, then E and F^c are independent.

Proof.

$$\begin{aligned}\mathbb{P}(E \cap F^c) &= \mathbb{P}(E) - \mathbb{P}(E \cap F) = \mathbb{P}(E) - \mathbb{P}(E)\mathbb{P}(F)\\ &= \mathbb{P}(E)[1 - \mathbb{P}(F)] = \mathbb{P}(E)\mathbb{P}(F^c).\end{aligned}$$

□

The concept of independence can be generalized to any number of events as follows.

Definition (Jointly independent events)

Let $A_1, \dots, A_n \subset S$ be a collection of n events. We say that they are *jointly (mutually) independent* if for all possible subcollections $i_1, \dots, i_r \in \{1, \dots, n\}$, $1 \leqslant r \leqslant n$, it holds that

$$\mathbb{P}\left(\bigcap_{k=1}^{r} A_{i_k}\right) = \prod_{k=1}^{r} \mathbb{P}(A_{i_k}).$$

For example, for three events, E, F, and G, they are independent if E and F are independent, E and G are independent, F and G are independent, and $\mathbb{P}(E \cap F \cap G) = \mathbb{P}(E)\mathbb{P}(F)\mathbb{P}(G)$.

Example 3.3 (Pairwise but not jointly independent events)**.** Throw two fair dice. Consider three events

$$\begin{aligned}
E &:= \{\text{the sum of the points is } 7\}, \\
F &:= \{\text{the first die rolled } 3\}, \\
G &:= \{\text{the second die rolled } 4\}.
\end{aligned}$$

The sample space consists of 36 elements $(i, j), i, j = 1, ..., 6$, each having the same probability. Then

$$\begin{aligned}
E &:= \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}, \\
F &:= \{(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \}, \\
G &:= \{(1,4), (2,4), (3,4), (4,4), (5,4), (6,4)\}, \\
E \cap F &= E \cap G = F \cap G = E \cap F \cap G = \{(3,4)\}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{P}(E) = \mathbb{P}(F) = \mathbb{P}(F) = \frac{1}{6}, \\
&\mathbb{P}(E \cap F) = \mathbb{P}(F \cap G) = \mathbb{P}(E \cap G) = \mathbb{P}(E \cap F \cap G) = \frac{1}{36},
\end{aligned}$$

so E, F, G are pairwise disjoint, but they are **not** jointly independent.

Example 3.4. What is the probability that exactly 3 threes will show if you roll 10 dice?

The probability that the 1st, 2nd, and 4th dice will show a three and the other 7 will not is $\frac{1}{6}^3 \frac{5}{6}^7$. Independence is used here: the probability is $\frac{1}{6}\frac{1}{6}\frac{5}{6}\frac{1}{6}\frac{5}{6} \cdots \frac{5}{6}$. The probability that the 4th, 5th, and 6th dice will show a three and the other 7 will not is the same thing. So to answer our original question, we take $\frac{1}{6}^3 \frac{5}{6}^7$ and multiply it by the number of ways of choosing 3 dice out of 10 to be the ones showing threes. There are $\binom{10}{3}$ ways of doing that.

This is a particular example of what are known as Bernoulli trials or the binomial distribution.

Proposition 3.2 (Bernoulli trials: binomial distribution)

If an experiment with probability of success p is repeated n times independently, the probability of having k successes for any $0 \leqslant k \leqslant n$ is given by

$$\mathbb{P}\{k \text{ successes in } n \text{ independent trials}\} = \binom{n}{k} p^k \left(1-p\right)^{n-k}$$

Proof. The probability there are k successes is the number of ways of putting k objects in n slots (which is $\binom{n}{k}$) times the probability that there will be k successes and $n-k$ failures in exactly a given order. So the probability is $\binom{n}{k} p^k (1-p)^{n-k}$. □

The name *binomial* for this distribution comes from the simple observation that if we denote by

$$E_k := \{k \text{ successes in } n \text{ independent trials}\}.$$

Then $S = \bigcup_{k=0}^{n} E_k$ is the disjoint decomposition of the sample space, so

$$\mathbb{P}(S) = \mathbb{P}\left(\bigcup_{k=0}^{n} E_k\right) = \sum_{k=0}^{n} \mathbb{P}(E_k) = \sum_{k=0}^{n} \binom{n}{k} p^k \left(1-p\right)^{n-k} = \left(p + (1-p)\right)^n = 1$$

by the binomial formula. This shows we indeed have the second axiom of probability for the binomial distribution. We denote by $\operatorname{Binom}(n, p)$ the binomial distribution with parameters n and p.

3.2. Further examples and explanations

3.2.1. Examples.

Example 3.5. A card is drawn from an ordinary deck of cards (52 cards). Consider the events $F :=$ *a face is drawn*, $R :=$ *a red color is drawn.*

These are independent events because, for one card, being a face does not affect it being red: there are 12 faces, 26 red cards, and 6 cards that are red and faces. Thus,

$$\mathbb{P}(F)\,\mathbb{P}(R) = \frac{12}{52}\cdot\frac{26}{52} = \frac{3}{26},$$
$$\mathbb{P}(F\cap R) = \frac{6}{52} = \frac{3}{26}.$$

Example 3.6. Suppose that two unfair coins are flipped: the first coin has the heads probability 0.5001 and the second has heads probability 0.5002. The events $A_T =:=$ *the first coin lands tails*, $B_H :=$ *the second coin lands heads* are independent. Why? The sample space $S = \{\text{HH, HT, TH, TT}\}$ has 4 elements, all of them of different probabilities, given as products. The events correspond to $A_T = \{\text{TH,TT}\}$ and $B_H = \{\text{HH,TH}\}$ respectively, and the computation of the probabilities is given by

$$\mathbb{P}(A_T\cap B_H) = 0.4999\cdot 0.5002 = \mathbb{P}(A_T)\,\mathbb{P}(B_H).$$

Example 3.7. An urn contains 10 balls, 4 red and 6 blue. A second urn contains 16 red balls and an unknown number of blue balls. A single ball is drawn from each urn and the probability that both balls are the same color is 0.44. How many blue balls are there in the second urn?

Let us define the events $R_i :=$ *a red ball is drawn from urn i*, $B_i :=$ *a blue ball is drawn from urn i*, and let x denote the (unknown) number of blue balls in urn 2, so that the second urn has $16+x$ balls in total. Using the fact that the events $R_1\cap R_2$ and $B_1\cap B_2$ are independent (check this!), we have

$$\begin{aligned}0.44 &= \mathbb{P}\Big((R_1\cap R_2)\bigcup(B_1\cap B_2)\Big) = \mathbb{P}(R_1\cap R_2) + \mathbb{P}(B_1\cap B_2)\\ &= \mathbb{P}(R_1)\,\mathbb{P}(R_2) + \mathbb{P}(B_1)\,\mathbb{P}(B_2)\\ &= \frac{4}{10}\frac{16}{x+16} + \frac{6}{10}\frac{x}{x+16}.\end{aligned}$$

Solving this equation for x we get $x = 4$.

3.2.2. Bernoulli trials. Recall that successive independent repetitions of an experiment that results in a success with some probability p and a failure with probability $1-p$ are called *Bernoulli trials*, and the distribution is given in Proposition 3.2. Sometimes we can view an experiment as the successive repetition of a *simpler* one. For instance, rolling 10 dice can be seen as rolling one single die ten times, each time independently of the other.

Example 3.8. Suppose that we roll 10 dice. What is the probability that at most 4 of them land a two?

We can regard this experiment as consequently rolling one single die. One possibility is that the first, second, third, and tenth trial land a two, while the rest land something else. Since each trial is independent, the probability of this event will be

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

Note that the probability that the 10th, 9th, 8th, and 7th dice land a two and the other 6 do not is the same as the previous one. To answer our original question, we thus need to consider the number of ways of choosing 0, 1, 2, 3 or 4 trials out of 10 to be the ones showing a two. This means

$$\mathbb{P}(\text{exactly 0 dice land a two})=\binom{10}{0}\cdot\left(\frac{1}{6}\right)^0\cdot\left(\frac{5}{6}\right)^{10}=\left(\frac{5}{6}\right)^{10}.$$

$$\mathbb{P}(\text{exactly 1 dice lands a two})=\binom{10}{1}\cdot\left(\frac{1}{6}\right)\cdot\left(\frac{5}{6}\right)^{9}.$$

$$\mathbb{P}(\text{exactly 2 dice land a two})=\binom{10}{2}\cdot\left(\frac{1}{6}\right)^2\cdot\left(\frac{5}{6}\right)^{8}.$$

$$\mathbb{P}(\text{exactly 3 dice land a two})=\binom{10}{3}\cdot\left(\frac{1}{6}\right)^3\cdot\left(\frac{5}{6}\right)^{7}.$$

$$\mathbb{P}(\text{exactly 4 dice land a two})=\binom{10}{4}\cdot\left(\frac{1}{6}\right)^4\cdot\left(\frac{5}{6}\right)^{6}.$$

The answer to the question is the sum of these five numbers.

3.3. Exercises

Exercise 3.1. Let A and B be two independent events such that $\mathbb{P}(A \cup B) = 0.64$ and $\mathbb{P}(A) = 0.4$. What is $\mathbb{P}(B)$?

Exercise 3.2. In a class, there are 4 male math majors, 6 female math majors, and 6 male actuarial science majors. How many actuarial science females must be present in the class if sex and major are independent when choosing a student selected at random?

Exercise 3.3. Following Proposition on 3.1, prove that E and F are independent if and only if E and F^c are independent.

Exercise 3.4. Suppose we toss a fair coin twice, and let E be the event that both tosses give the same outcome, F that the first toss is a heads, and G is that the second toss is heads. Show that E, F and G are pairwise independent, but not jointly independent.

Exercise 3.5. Two dice are simultaneously rolled. For each pair of events defined below, compute if they are independent or not.

(a) $A_1 = \{\text{the sum is 7}\}$, $B_1 = \{\text{the first die lands a 3}\}$.
(b) $A_2 = \{\text{the sum is 9}\}$, $B_2 = \{\text{the second die lands a 3}\}$.
(c) $A_3 = \{\text{the sum is 9}\}$, $B_3 = \{\text{the first die lands even}\}$.
(d) $A_4 = \{\text{the sum is 9}\}$, $B_4 = \{\text{the first die is less than the second}\}$.
(e) $A_5 = \{\text{two dice are equal}\}$, $B_5 = \{\text{the sum is 8}\}$.
(f) $A_6 = \{\text{two dice are equal}\}$, $B_6 = \{\text{the first die lands even}\}$.
(g) $A_7 = \{\text{two dice are not equal}\}$, $B_7 = \{\text{the first die is less than the second}\}$.

Exercise 3.6. Are the events A_1, B_1 and B_3 from Exercise 3.5 independent?

Exercise 3.7. A hockey team has 0.45 chances of losing a game. Assuming that each game is independent from the other, what is the probability that the team loses 3 of the next upcoming 5 games?

Exercise 3.8. You make successive independent flips of a coin that lands on heads with probability p. What is the probability that the 3rd heads appears on the 7th flip?

Hint: express your answers in terms of p; do not assume $p = 1/2$.

Exercise 3.9. Suppose you toss a fair coin repeatedly and independently. If it comes up heads, you win a dollar, and if it comes up tails, you lose a dollar. Suppose you start with $\$M$. What is the probability you will get up to $\$N$ before you go broke? Give the answer in terms of M and N, assuming $0 < M < N$.

Exercise 3.10. Suppose that we roll n dice. What is the probability that at most k of them land a two?

3.4. Selected solutions

Solution to Exercise 3.1: Using independence we have $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)$ and substituting we have

$$0.64 = 0.4 + \mathbb{P}(B) - 0.4\mathbb{P}(B).$$

Solving for $\mathbb{P}(B)$ we have $\mathbb{P}(B) = 0.4$.

Solution to Exercise 3.2: Let x denote the number of actuarial sciences females. Then

$$\begin{aligned}\mathbb{P}(\text{male} \cap \text{math}) &= \frac{4}{16+x},\\ \mathbb{P}(\text{male}) &= \frac{10}{16+x}\\ \mathbb{P}(\text{math}) &= \frac{10}{16+x}.\end{aligned}$$

Then using independence $\mathbb{P}(\text{male} \cap \text{math}) = \mathbb{P}(\text{male})\,\mathbb{P}(\text{math})$ so that

$$\frac{4}{16+x} = \frac{10^2}{(16+x)^2} \implies 4 = \frac{100}{16+x}$$

and solving for x we have $x = 9$.

Solution to Exercise 3.3: Proposition 3.1 tells us that if E and F are independent, then E and F^c are independent. Let us now assume that E and F^c are independent. We can apply Proposition 3.1 and say that E and $(F^c)^c$ are independent. Since $(F^c)^c = F$ (draw a Venn diagram), the assertion is proved.

Solution to Exercise 3.4:

Solution to Exercise 3.7: These are Bernoulli trials. Each game is a trial and the probability of loosing is $p = 0.45$. Using Proposition 3.2 with $k = 3$ and $n = 5$ we have

$$\mathbb{P}(3 \text{ loses in } 5 \text{ trials}) = \binom{5}{3} 0.45^3 \cdot 0.55^2.$$

Solution to Exercise 3.8: The 3rd head appearing on the 7th flip means that exactly two heads during the previous 6 flips appear and the 7th is heads. Since the flips are independent we have that the probability we search is

$$\mathbb{P}(\textit{2 heads in 6 trials } \textbf{AND} \textit{ heads in the 7th flip}) = \mathbb{P}(\textit{2 heads in 6 trials})\mathbb{P}(H).$$

Using Bernoulli trials, $\mathbb{P}(\textit{2 heads in 6 trials}) = \binom{6}{2} p^2(1-p)^4$ and therefore the total probability is

$$\binom{6}{2} p^2(1-p)^4 \cdot p = \binom{6}{2} p^3(1-p)^4.$$

Solution to Exercise 3.10:

$$\sum_{r=0}^{k} \binom{n}{r} \cdot \left(\frac{1}{6}\right)^r \cdot \left(\frac{5}{6}\right)^{n-r}.$$

CHAPTER 4

Conditional probability

4.1. Definition, Bayes' Rule and examples

Suppose there are 200 men, of which 100 are smokers, and 100 women, of which 20 are smokers. What is the probability that a person chosen at random will be a smoker? The answer is 120/300. Now, let us ask, what is the probability that a person chosen at random is a smoker given that the person is a women? One would expect the answer to be 20/100 and it is.

What we have computed is

$$\frac{\text{number of women smokers}}{\text{number of women}} = \frac{\text{number of women smokers }/300}{\text{number of women }/300},$$

which is the same as the probability that a person chosen at random is a woman and a smoker divided by the probability that a person chosen at random is a woman.

With this in mind, we give the following definition.

Definition 4.1 (Conditional probability)

If $\mathbb{P}(F) > 0$, we define *the probability of* E *given* F as

$$\mathbb{P}(E \mid F) := \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}.$$

Note $\mathbb{P}(E \cap F) = \mathbb{P}(E \mid F)\mathbb{P}(F)$.

Example 4.1. Suppose you roll two dice. What is the probability the sum is 8?

There are five ways this can happen $\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$, so the probability is 5/36. Let us call this event A. What is the probability that the sum is 8 given that the first die shows a 3? Let B be the event that the first die shows a 3. Then $\mathbb{P}(A \cap B)$ is the probability that the first die shows a 3 and the sum is 8, or 1/36. $\mathbb{P}(B) = 1/6$, so $\mathbb{P}(A \mid B) = \frac{1/36}{1/6} = 1/6$.

Example 4.2. Suppose a box has 3 red marbles and 2 black ones. We select 2 marbles. What is the probability that second marble is red given that the first one is red?

Let A be the event the second marble is red, and B the event that the first one is red. $\mathbb{P}(B) = 3/5$, while $\mathbb{P}(A \cap B)$ is the probability both are red, or is the probability that we chose 2 red out of 3 and 0 black out of 2. Then $\mathbb{P}(A \cap B) = \binom{3}{2}\binom{2}{0}/\binom{5}{2}$, and so $\mathbb{P}(A \mid B) = \frac{3/10}{3/5} = 1/2$.

Example 4.3. A family has 2 children. Given that one of the children is a boy, what is the probability that the other child is also a boy?

Let B be the event that one child is a boy, and A the event that both children are boys. The possibilities are bb, bg, gb, gg, each with probability 1/4. $\mathbb{P}(A \cap B) = \mathbb{P}(bb) = 1/4$ and $\mathbb{P}(B) = \mathbb{P}(bb, bg, gb) = 3/4$. So the answer is $\frac{1/4}{3/4} = 1/3$.

Example 4.4. Suppose the test for HIV is 99% accurate in both directions and 0.3% of the population is HIV positive. If someone tests positive, what is the probability they actually are HIV positive?

Let D is the event that a person is HIV positive, and T is the event that the person tests positive.

$$\mathbb{P}(D \mid T) = \frac{\mathbb{P}(D \cap T)}{\mathbb{P}(T)} = \frac{(0.99)(0.003)}{(0.99)(0.003) + (0.01)(0.997)} \approx 23\%.$$

A short reason why this surprising result holds is that the error in the test is much greater than the percentage of people with HIV. A little longer answer is to suppose that we have 1000 people. On average, 3 of them will be HIV positive and 10 will test positive. So the chances that someone has HIV given that the person tests positive is approximately 3/10. The reason that it is not exactly 0.3 is that there is some chance someone who is positive will test negative.

Suppose you know $\mathbb{P}(E \mid F)$ and you want to find $\mathbb{P}(F \mid E)$. Recall that

$$\mathbb{P}(E \cap F) = \mathbb{P}(E \mid F)\mathbb{P}(F),$$

and so

$$\mathbb{P}(F \mid E) = \frac{\mathbb{P}(F \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E \mid F)\mathbb{P}(F)}{\mathbb{P}(E)}$$

Example 4.5. Suppose 36% of families own a dog, 30% of families own a cat, and 22% of the families that have a dog also have a cat. A family is chosen at random and found to have a cat. What is the probability they also own a dog?

Let D be the families that own a dog, and C the families that own a cat. We are given $\mathbb{P}(D) = 0.36, \mathbb{P}(C) = 0.30, \mathbb{P}(C \mid D) = 0.22$. We want to know $\mathbb{P}(D \mid C)$. We know $\mathbb{P}(D \mid C) = \mathbb{P}(D \cap C)/\mathbb{P}(C)$. To find the numerator, we use $\mathbb{P}(D \cap C) = \mathbb{P}(C \mid D)\mathbb{P}(D) = (0.22)(0.36) = 0.0792$. So $\mathbb{P}(D \mid C) = 0.0792/0.3 = 0.264 = 26.4\%$.

Example 4.6. Suppose 30% of the women in a class received an A on the test and 25% of the men received an A. The class is 60% women. Given that a person chosen at random received an A, what is the probability this person is a women?

Let A be the event of receiving an A, W be the event of being a woman, and M the event of being a man. We are given $\mathbb{P}(A \mid W) = 0.30, \mathbb{P}(A \mid M) = 0.25, \mathbb{P}(W) = 0.60$ and we want $\mathbb{P}(W \mid A)$. From the definition

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

As in the previous example,

$$\mathbb{P}(W \cap A) = \mathbb{P}(A \mid W)\mathbb{P}(W) = (0.30)(0.60) = 0.18.$$

To find $\mathbb{P}(A)$, we write

$$\mathbb{P}(A) = \mathbb{P}(W \cap A) + \mathbb{P}(M \cap A).$$

Since the class is 40% men,

$$\mathbb{P}(M \cap A) = \mathbb{P}(A \mid M)\mathbb{P}(M) = (0.25)(0.40) = 0.10.$$

So

$$\mathbb{P}(A) = \mathbb{P}(W \cap A) + \mathbb{P}(M \cap A) = 0.18 + 0.10 = 0.28.$$

Finally,

$$\mathbb{P}(W \mid A) = \frac{\mathbb{P}(W \cap A)}{\mathbb{P}(A)} = \frac{0.18}{0.28}.$$

Proposition 4.1 (Bayes' rule)

If $\mathbb{P}(E) > 0$, then

$$\mathbb{P}(F \mid E) = \frac{\mathbb{P}(E \mid F)\mathbb{P}(F)}{\mathbb{P}(E \mid F)\mathbb{P}(F) + \mathbb{P}(E \mid F^c)\mathbb{P}(F^c)}.$$

PROOF. We use the definition of conditional probability and the fact that

$$\mathbb{P}(E) = \mathbb{P}((E \cap F) \cup (E \cap F^c)) = \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c).$$

This will be called the *law of total probability* and will be discussed again in Proposition 4.4 in more generality. Then

$$\mathbb{P}(F \mid E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E \mid F)\mathbb{P}(F)}{\mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c)}$$
$$= \frac{\mathbb{P}(E \mid F)\mathbb{P}(F)}{\mathbb{P}(E \mid F)\mathbb{P}(F) + \mathbb{P}(E \mid F^c)\mathbb{P}(F^c)}.$$

□

Here is another example related to conditional probability, although this is not an example of Bayes' rule. This is known as the *Monty Hall problem* after the host of the TV show in the 60s called *Let's Make a Deal*.

Example 4.7. There are three doors, behind one a nice car, behind each of the other two a goat eating a bale of straw. You choose a door. Then Monty Hall opens one of the other doors, which shows a bale of straw. He gives you the opportunity of switching to the remaining door. Should you do it?

Let us suppose you choose door 1, since the same analysis applies whichever door you chose. The *first strategy* is to stick with door 1. With probability $1/3$ you chose the car. Monty Hall shows you one of the other doors, but that doesn't change your probability of winning.

The *second strategy* is to change. Let us say the car is behind door 1, which happens with probability $1/3$. Monty Hall shows you one of the other doors, say door 2. There will be a goat, so you switch to door 3, and lose. The same argument applies if he shows you door 3. Suppose the car is behind door 2. He will show you door 3, since he doesn't want to give away the car. You switch to door 2 and win. This happens with probability $1/3$. The same argument applies if the car is behind door 3.

So you win with probability $2/3$ and lose with probability $1/3$. Thus strategy 2 is much superior.

A problem that comes up in actuarial science frequently is *gambler's ruin*.

Example 4.8 (Gambler's ruin). Suppose you play the game by tossing a fair coin repeatedly and independently. If it comes up heads, you win a dollar, and if it comes up tails, you lose a dollar. Suppose you start with \$50. What's the probability you will get to \$200 without first getting ruined (running out of money)?

It is easier to solve a slightly harder problem. The game can be described as having probability $1/2$ of winning 1 dollar and a probability $1/2$ of losing 1 dollar. A player begins with a given number of dollars, and intends to play the game repeatedly until the player either goes broke or increases his holdings to N dollars.

For any given amount n of current holdings, the conditional probability of reaching N dollars before going broke is independent of how we acquired the n dollars, so there is a unique probability $\mathbb{P}(N \mid n)$ of reaching N on the condition that we currently hold n dollars. Of course, for any finite N we see that $\mathbb{P}(N \mid 0) = 0$ and $\mathbb{P}(N \mid N) = 1$. The problem is to determine the values of $\mathbb{P}(N \mid n)$ for n between 0 and N.

We are considering this setting for $N = 200$, and we would like to find $\mathbb{P}(200 \mid 50)$. Denote $y(n) := \mathbb{P}(200 \mid n)$, which is the probability you get to 200 without first getting ruined if you start with n dollars. We saw that $y(0) = 0$ and $y(200) = 1$. Suppose the player has n dollars at the moment, the next round will leave the player with either $n+1$ or $n-1$ dollars, both with probability $1/2$. Thus the current probability of winning is the same as a weighted average of the probabilities of winning in player's two possible *next* states. So we have

$$y(n) = \tfrac{1}{2}y(n+1) + \tfrac{1}{2}y(n-1).$$

Multiplying by 2, and subtracting $y(n)+y(n-1)$ from each side, we have

$$y(n+1)-y(n)=y(n)-y(n-1).$$

This says that slopes of the graph of $y(n)$ on the adjacent intervals are constant (remember that x must be an integer). In other words, the graph of $y(n)$ must be a line. Since $y(0)=0$ and $y(200)=1$, we have $y(n)=n/200$, and therefore $y(50)=1/4$.

Another way to see what the function $y(n)$ is to use the telescoping sum as follows

$$\begin{aligned} y(n) &= y(n)-y(0)=(y(n)-y(n-1))+...+(y(1)-y(0)) \\ &= n(y(1)-y(0))=ny(1). \end{aligned} \tag{4.1.1}$$

since the all these differences are the same, and $y(0)=0$. To find $y(1)$ we can use the fact that $y(200)=1$, so $y(1)=1/200$, and therefore $y(n)=n/200$ and $y(50)=1/4$.

Example 4.9. Suppose we are in the same situation, but you are allowed to go arbitrarily far in debt. Let $z(n)$ be the probability you ever get to \$200 if you start with n dollars. What is a formula for $z(n)$?

Just as above, we see that z satisfies the recursive equation

$$z(n)=\tfrac{1}{2}z(n+1)+\tfrac{1}{2}z(n-1).$$

What we need to determine now are boundary conditions. Now that the gambler can go to debt, the condition that if we start with 0 we never get to \$200, that is, probability of getting \$200 is 0, is **not** true. Following Equation 4.1.1 with $z(0)\neq 0$ we see that

$$\begin{aligned} z(n)-z(0) &= (z(n)-z(n-1))+...+(z(1)-z(0)) \\ &= n(z(1)-z(0)), \end{aligned}$$

therefore

$$z(n)=n(z(1)-z(0))+z(0).$$

If we denote $a:=z(1)-z(0)$ and $b:=z(0)$ we see that as a function of n we have

$$z(n)=an+b.$$

We would like to find a and b now. Recall that this function is probability, so for any n we have $0\leqslant z(n)\leqslant 1$. This is possible only if $a=0$, that is,

$$z(1)=z(0),$$

so

$$z(n)=z(0)$$

for any n. We know that $z(200)=1$, therefore

$$z(n)=1 \text{ for all } n.$$

In other words, one is certain to get to \$200 eventually (provided, of course, that one is allowed to go into debt).

4.2. Further examples and applications

4.2.1. Examples, basic properties, multiplication rule, law of total probability.

Example 4.10. Landon is 80% sure he forgot his textbook either at the Union or in Monteith. He is 40% sure that the book is at the union, and 40% sure that it is in Monteith. Given that Landon already went to Monteith and noticed his textbook is not there, what is the probability that it is at the Union?

We denote by U the event that *textbook is at the Union*, and by M the event that *textbook is in Monteith*, notice that $U \subseteq M^c$ and hence $U \cap M^c = U$. Thus,

$$\mathbb{P}(U \mid M^c) = \frac{\mathbb{P}(U \cap M^c)}{\mathbb{P}(M^c)} = \frac{\mathbb{P}(U)}{1 - \mathbb{P}(M)} = \frac{4/10}{6/10} = \frac{2}{3}.$$

Example 4.11. Sarah and Bob draw 13 cards each from a standard deck of 52. Given that Sarah has exactly two aces, what is the probability that Bob has exactly one ace?

Let A be the event that *Sarah has two aces*, and let B be the event that *Bob has exactly one ace*. In order to compute $\mathbb{P}(B \mid A)$, we need to calculate $\mathbb{P}(A)$ and $\mathbb{P}(A \cap B)$. On the one hand, Sarah could have any of $\binom{52}{13}$ possible hands. Of these hands, $\binom{4}{2} \cdot \binom{48}{11}$ will have exactly two aces so that

$$\mathbb{P}(A) = \frac{\binom{4}{2} \cdot \binom{48}{11}}{\binom{52}{13}}.$$

On the other hand, the number of ways in which Sarah can pick a hand and Bob another (different) is $\binom{52}{13} \cdot \binom{39}{13}$. The the number of ways in which A and B can simultaneously occur is $\binom{4}{2} \cdot \binom{48}{11} \cdot \binom{2}{1} \cdot \binom{37}{12}$ and hence

$$\mathbb{P}(A \cap B) = \frac{\binom{4}{2} \cdot \binom{48}{11} \cdot \binom{2}{1} \cdot \binom{37}{12}}{\binom{52}{13} \cdot \binom{39}{13}}.$$

Applying the definition of conditional probability we finally get

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\binom{4}{2} \cdot \binom{48}{11} \cdot \binom{2}{1} \cdot \binom{37}{12} \Big/ \binom{52}{13} \cdot \binom{39}{13}}{\binom{4}{2} \cdot \binom{48}{11} \Big/ \binom{52}{13}} = \frac{2 \cdot \binom{37}{12}}{\binom{39}{13}}$$

Example 4.12. A total of 500 married couples are polled about their salaries with the following results

	husband makes less than \$25K	husband makes more than \$25K
wife makes less than \$25K	212	198
wife makes more than \$25K	36	54

(a) We can find the probability that a husband earns less than \$25K as follows.

$$\mathbb{P}(\text{husband makes } < \$25\text{K}) = \frac{212}{500} + \frac{36}{500} = \frac{248}{500} = 0.496.$$

(b) Now we find the probability that a wife earns more than \$25K, given that the husband earns as that much as well.

$$\mathbb{P}\left(\text{wife makes } > \$25\text{K} \mid \text{husband makes } > \$25\text{K}\right) = \frac{54/500}{(198+54)/500} = \frac{54}{252} = 0.214$$

(c) Finally we find the probability that a wife earns more than \$25K, given that the husband makes less than \$ 25K.

$$\mathbb{P}\left(\text{wife makes } > \$25\text{K} \mid \text{husband makes } < \$25\text{K}\right) = \frac{36/500}{248/500} = 0.145.$$

From the definition of conditional probability we can deduce some useful relations.

Proposition 4.2

Let $E, F \in \mathcal{F}$ be events with $\mathbb{P}(E), \mathbb{P}(F) > 0$. Then

(i) $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F \mid E)$,
(ii) $\mathbb{P}(E) = \mathbb{P}(E \mid F)\mathbb{P}(F) + \mathbb{P}(E \mid F^c)\mathbb{P}(F^c)$,
(iii) $\mathbb{P}(E^c \mid F) = 1 - \mathbb{P}(E \mid F)$.

Proof. We already saw (i) which is a rewriting of the definition of conditional probability $\mathbb{P}(F \mid E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)}$. Let us prove (ii): we can write E as the union of the pairwise disjoint sets $E \cap F$ and $E \cap F^c$. Using (i) we have

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) \\ &= \mathbb{P}(E \mid F)\mathbb{P}(F) + \mathbb{P}(E \mid F^c)\mathbb{P}(F^c).\end{aligned}$$

Finally, writing $F = E$ in the previous equation and since $\mathbb{P}(E \mid E^c) = 0$, we obtain (iii). ☐

Example 4.13. Phan wants to take either a Biology course or a Chemistry course. His adviser estimates that the probability of scoring an A in Biology is $\frac{4}{5}$, while the probability of scoring an A in Chemistry is $\frac{1}{7}$. If Phan decides randomly, by a coin toss, which course to take, what is his probability of scoring an A in Chemistry?

We denote by B the event that *Phan takes Biology*, and by C the event that *Phan takes Chemistry*, and by $A =$ the event that the *score is an A*. Then, since $\mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ we have

$$\mathbb{P}(A \cap C) = \mathbb{P}(C)\mathbb{P}(A \mid C) = \frac{1}{2} \cdot \frac{1}{7} = \frac{1}{14}.$$

The identity $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F \mid E)$ from Proposition 4.2(i) can be generalized to any number of events in what is sometimes called the *multiplication rule*.

Proposition 4.3 (Multiplication rule)

Let $E_1, E_2, \ldots, E_n \in \mathcal{F}$ be events. Then

$$\begin{aligned}&\mathbb{P}\left(E_1 \cap E_2 \cap \ldots \cap E_n\right)\\ &= \mathbb{P}\left(E_1\right)\mathbb{P}\left(E_2 \mid E_1\right)\mathbb{P}\left(E_3 \mid E_1 \cap E_2\right)\cdots\mathbb{P}\left(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}\right).\end{aligned}$$

Example 4.14. An urn has 5 blue balls and 8 red balls. Each ball that is selected is returned to the urn along with an additional ball of the same color. Suppose that 3 balls are drawn in this way.

(a) What is the probability that the three balls are blue?
In this case, we can define the sequence of events $B_1, B_2, B_3, \ldots$, where B_i is the event that *the ith ball drawn is blue*. Applying the multiplication rule yields

$$\mathbb{P}(B_1 \cap B_2 \cap B_3) = \mathbb{P}(B_1)\mathbb{P}(B_2 \mid B_1)\mathbb{P}(B_3 \mid B_1 \cap B_2) = \frac{5}{13}\frac{6}{14}\frac{7}{15}.$$

(b) What is the probability that only 1 ball is blue?
Denote by $R_i =$ the event that *the ith ball drawn is red*, we have

$$\mathbb{P}\left(\text{only 1 blue ball}\right) = \mathbb{P}(B_1 \cap R_2 \cap R_3) + \mathbb{P}(R_1 \cap B_2 \cap R_3) + \mathbb{P}(R_1 \cap R_2 \cap B_3) = 3\frac{5 \cdot 8 \cdot 9}{13 \cdot 14 \cdot 15}.$$

Also the identity (ii) in Proposition 4.2 can be generalized by partitioning the sample space S into several pairwise disjoint sets $F_1, \ldots, F_n$ (instead of simply F and F^c).

Proposition 4.4 (Law of total probability)

Let $F_1, \ldots, F_n \subseteq S$ be mutually exclusive and exhaustive events, i.e. $S = \bigcup_{i=1}^n F_i$. Then, for any event $E \in \mathcal{F}$ it holds that

$$\mathbb{P}\left(E\right) = \sum_{i=1}^{n} \mathbb{P}\left(E \mid F_i\right)\mathbb{P}\left(F_i\right).$$

4.2.2. Generalized Bayes' rule. The following example describes the type of problems treated in this section.

Example 4.15. An insurance company classifies insured policyholders into *accident prone* or *non-accident prone*. Their current risk model works with the following probabilities.

The probability that an *accident prone* insured has an accident within a year is 0.4
The probability that a *non-accident prone* insured has an accident within a year is 0.2.

If 30% of the population is *accident prone*,

(a) what is the probability that a policy holder will have an accident within a year?

Denote by $A_1 =$ the event that *a policy holder will have an accident within a year*, and denote by $A =$ the event that *a policy holder is accident prone*. Applying Proposition 4.2(ii) we have

$$\begin{aligned}\mathbb{P}(A_1) &= \mathbb{P}(A_1 \mid A)\,\mathbb{P}(A) + \mathbb{P}(A_1 \mid A^c)\,(1 - \mathbb{P}(A)) \\ &= 0.4 \cdot 0.3 + 0.2(1 - 0.3) = 0.26\end{aligned}$$

(b) Suppose now that the policy holder has had accident within one year. What is the probability that he or she is accident prone?

Use Bayes' formula to see that

$$\mathbb{P}(A \mid A_1) = \frac{\mathbb{P}(A \cap A_1)}{\mathbb{P}(A_1)} = \frac{\mathbb{P}(A)\,\mathbb{P}(A_1 \mid A)}{0.26} = \frac{0.3 \cdot 0.4}{0.26} = \frac{6}{14}.$$

Using the *law of total probability* from Proposition 4.4 one can generalize Bayes' rule, which appeared in Proposition 4.1.

Proposition 4.5 (Generalized Bayes' rule)

Let $F_1, \ldots, F_n \subseteq S$ be mutually exclusive and exhaustive events, i.e. $S = \bigcup_{i=1}^n F_i$. Then, for any event $E \subseteq S$ and any $j = 1, \ldots, n$ it holds that

$$\mathbb{P}(F_j \mid E) = \frac{\mathbb{P}(E \mid F_j)\,\mathbb{P}(F_j)}{\sum_{i=1}^n \mathbb{P}(E \mid F_i)\,\mathbb{P}(F_i)}$$

Example 4.16. Suppose a factory has machines I, II, and III that produce *iSung* phones. The factory's record shows that

Machines I, II and III produce, respectively, 2%, 1%, and 3% defective *iSung*s.
Out of the total production, machines I, II, and III produce, respectively, 35%, 25% and 40% of all *iSung*s.

An *iSung* is selected at random from the factory.

(a) What is probability that the *iSung* selected is defective?

By the law of total probability,

$$\begin{aligned}\mathbb{P}(D) &= \mathbb{P}(I)\,\mathbb{P}(D \mid I) + \mathbb{P}(II)\,\mathbb{P}(D \mid II) + P(III)\,\mathbb{P}(D \mid III) \\ &= 0.35 \cdot 0.02 + 0.25 \cdot 0.01 + 0.4 \cdot 0.03 = \frac{215}{10,000}.\end{aligned}$$

(b) Given that the *iSung* is defective, what is the conditional probability that it was produced by machine III?

Applying Bayes' rule,

$$\mathbb{P}(III \mid D) = \frac{\mathbb{P}(III)\,\mathbb{P}(D \mid III)}{\mathbb{P}(D)} = \frac{0.4 \cdot 0.03}{215/10,000} = \frac{120}{215}.$$

Example 4.17. In a multiple choice test, a student either knows the answer to a question or she/he will randomly guess it. If each question has m possible answers and the student knows the answer to a question with probability p, what is the probability that the student actually knows the answer to a question, given that he/she answers correctly?

Denote by K the event that *a student knows the answer*, and by C the event that *a student answers correctly*. Applying Bayes' rule we have

$$\mathbb{P}(K \mid C) = \frac{\mathbb{P}(C \mid K)\,\mathbb{P}(K)}{\mathbb{P}(C \mid K)\,\mathbb{P}(K) + \mathbb{P}(C \mid K^c)\,\mathbb{P}(K^c)} = \frac{1 \cdot p}{1 \cdot p + \frac{1}{m}(1-p)} = \frac{mp}{1+(m-1)p}.$$

4.3. Exercises

Exercise 4.1. Two dice are rolled. Consider the events $A = \{\text{sum of two dice equals 3}\}$, $B = \{\text{sum of two dice equals 7 }\}$, and $C = \{\text{at least one of the dice shows a 1}\}$.

(a) What is $\mathbb{P}(A \mid C)$?
(b) What is $\mathbb{P}(B \mid C)$?
(c) Are A and C independent? What about B and C?

Exercise 4.2. Suppose you roll two standard, fair, 6-sided dice. What is the probability that the sum is at least 9 given that you rolled at least one 6?

Exercise 4.3. A box contains 1 green ball and 1 red ball, and a second box contains 2 green and 3 red balls. First a box is chosen and afterwards a ball withdrawn from the chosen box. Both boxes are equally likely to be chosen. Given that a green ball has been withdrawn, what is the probability that the first box was chosen?

Exercise 4.4. Suppose that 60% of UConn students will be at random exposed to the flu. If you are exposed and did not get a flu shot, then the probability that you will get the flu *(after being exposed)* is 80%. If you did get a flu shot, then the probability that you will get the flu *(after being exposed)* is only 15%.

(a) What is the probability that a person who got a flu shot will get the flu?
(b) What is the probability that a person who did not get a flu shot will get the flu?

Exercise 4.5. Color blindness is a sex-linked condition, and 5% of men and 0.25% of women are color blind. The population of the United States is 51% female. What is the probability that a color-blind American is a man?

Exercise 4.6. Two factories supply light bulbs to the market. Bulbs from factory X work for over 5000 hours in 99% of cases, whereas bulbs from factory Y work for over 5000 hours in 95% of cases. It is known that factory X supplies 60% of the total bulbs available in the market.

(a) What is the probability that a purchased bulb will work for longer than 5000 hours?
(b) Given that a light bulb works for more than 5000 hours, what is the probability that it was supplied by factory Y?
(c) Given that a light bulb work does not work for more than 5000 hours, what is the probability that it was supplied by factory X?

Exercise 4.7. A factory production line is manufacturing bolts using three machines, A, B and C. Of the total output, machine A is responsible for 25%, machine B for 35% and machine C for the rest. It is known from previous experience with the machines that 5% of the output from machine A is defective, 4% from machine B and 2% from machine C. A bolt is chosen at random from the production line and found to be defective. What is the probability that it came from Machine A?

Exercise 4.8. A multiple choice exam has 4 choices for each question. The student has studied enough so that the probability they will know the answer to a question is 0.5, the probability that the student will be able to eliminate one choice is 0.25, otherwise all 4 choices seem equally plausible. If they know the answer they will get the question correct. If not they have to guess from the 3 or 4 choices. As the teacher you would like the test to measure what the student knows, and not how well they can guess. If the student answers a question correctly what is the probability that they actually know the answer?

Exercise 4.9. A blood test indicates the presence of Amyotrophic lateral sclerosis (ALS) 95% of the time when ALS is actually present. The same test indicates the presence of ALS 0.5% of the time when the disease is not actually present. One percent of the population actually has ALS. Calculate the probability that a person actually has ALS given that the test indicates the presence of ALS.

Exercise 4.10. A survey conducted in a college found that 40% of the students watch show A and 17% of the students who follow show A, also watch show B. In addition, 20% of the students watch show B.

(1) What is the probability that a randomly chosen student follows both shows?
(2) What is the conditional probability that the student follows show A given that she/he follows show B?

Exercise 4.11. Use Bayes' formula to solve the following problem. An airport has problems with birds. If the weather is sunny, the probability that there are birds on the runway is 1/2; if it is cloudy, but dry, the probability is 1/3; and if it is raining, then the probability is 1/4. The probability of each type of the weather is 1/3. Given that the birds are on the runaway, what is the probability

(1) that the weather is sunny?
(2) that the weather is cloudy (dry or rainy)?

Exercise 4.12. Suppose you toss a fair coin repeatedly and independently. If it comes up heads, you win a dollar, and if it comes up tails, you lose a dollar. Suppose you start with \$20. What is the probability you will get to \$150 before you go broke? (See Example 4.8 for a solution).

Exercise* 4.1. Suppose we play gambler's ruin game in Example 4.8 not with a fair coin, but rather in such a way that you win a dollar with probability p, and you lose a dollar with probability $1-p$, $0<p<1$. Find the probability of reaching N dollars before going broke if we start with n dollars.

Exercise* 4.2. Suppose F is an event, and define $\mathbb{P}_F(E) := \mathbb{P}(E \mid F)$. Show that the conditional probability $\mathbb{P}_F$ is a probability function, that is, it satisfies the axioms of probability.

Exercise* 4.3. Show directly that Proposition 2.1 holds for the conditional probability $\mathbb{P}_F$. In particular, for any events E and F

$$\mathbb{E}\left(E^c \mid F\right) = 1 - \mathbb{E}\left(E \mid F\right).$$

4.4. Selected solutions

Solution to Exercise 4.1(A): Note that the sample space is $S = \{(i,j) \mid i,j = 1,2,3,4,5,6\}$ with each outcome equally likely. Then

$$\begin{aligned} A &= \{(1,2),(2,1)\} \\ B &= \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} \\ C &= \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\} \end{aligned}$$

Then

$$\mathbb{P}(A \mid C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{2/36}{11/36} = \frac{2}{11}.$$

Solution to Exercise 4.1(B):

$$\mathbb{P}(B \mid C) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \frac{2/36}{11/36} = \frac{2}{11}.$$

Solution to Exercise 4.1(C): Note that $\mathbb{P}(A) = 2/36 \neq \mathbb{P}(A \mid C)$, so they are not independent. Similarly, $\mathbb{P}(B) = 6/36 \neq \mathbb{P}(B \mid C)$, so they are not independent.

Solution to Exercise 4.2: denote by E the event that *there is at least one 6* and by F the event that *the sum is at least 9*. We want to find $\mathbb{P}(F \mid E)$. Begin by noting that there are 36 possible rolls of these two dice and all of them are equally likely. We can see that 11 different rolls of these two dice will result in at least one 6, so $\mathbb{P}(E) = \frac{11}{36}$. There are 7 different rolls that will result in at least one 6 and a sum of at least 9. They are $\{(6,3),(6,4),(6,5),(6,6),(3,6),(4,6),(5,6)\}$, so $\mathbb{P}(E \cap F) = \frac{7}{36}$. This tells us that

$$\mathbb{P}(F \mid E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)} = \frac{7/36}{11/36} = \frac{7}{11}.$$

Solution to Exercise 4.3: denote by B_i the event that *the ith box is chosen*. Since both are equally likely, $\mathbb{P}(B_1) = \mathbb{P}(B_2) = \frac{1}{2}$. In addition, we know that $\mathbb{P}(G \mid B_1) = \frac{1}{2}$ and $\mathbb{P}(G \mid B_2) = \frac{2}{5}$. Applying Bayes' rule yields

$$\mathbb{P}(B_1 \mid G) = \frac{\mathbb{P}(G \mid B_1)\mathbb{P}(B_1)}{\mathbb{P}(G \mid B_1)\mathbb{P}(B_1) + \mathbb{P}(G \mid B_2)\mathbb{P}(B_2)} = \frac{1/4}{1/4 + 1/5} = \frac{5}{9}.$$

Solution to Exercise 4.4(A): Suppose we look at students who have gotten the flu shot. Denote by E the event that *a student is exposed to the flu*, and by F the event that *a student gets the flu*. We know that $\mathbb{P}(E) = 0.6$ and $\mathbb{P}(F \mid E) = 0.15$. This means that $\mathbb{P}(E \cap F) = (0.6)(0.15) = 0.09$, and it is clear that $\mathbb{P}(E^c \cap F) = 0$. Since $\mathbb{P}(F) = \mathbb{P}(E \cap F) + \mathbb{P}(E^c \cap F)$, we see that $\mathbb{P}(F) = 0.09$.

Solution to Exercise 4.4(B): Suppose we look at students who have not gotten the flu shot. Let E be the event that *a student is exposed to the flu*, and let F be the event that *a student gets the flu*. We know that $\mathbb{P}(E) = 0.6$ and $\mathbb{P}(F \mid E) = 0.8$. This means that $\mathbb{P}(E \cap F) = (0.6)(0.8) = 0.48$, and it is clear that $\mathbb{P}(E^c \cap F) = 0$. Since $\mathbb{P}(F) = \mathbb{P}(E \cap F) + \mathbb{P}(E^c \cap F)$, we see that $\mathbb{P}(F) = 0.48$.

Solution to Exercise 4.5: denote by M the event *an American is a man*, by C the event *an American is color blind.* Then

$$\mathbb{P}(M \mid C) = \frac{\mathbb{P}(C \mid M)\,\mathbb{P}(M)}{\mathbb{P}(C \mid M)\,\mathbb{P}(M) + \mathbb{P}(C \mid M^c)\,\mathbb{P}(M^c)} = \frac{(0.05)\,(0.49)}{(0.05)\,(0.49) + (0.0025)\,(0.51)} \approx 0.9505.$$

Solution to Exercise 4.6(A): let H be the event *a bulb works over* 5000 *hours*, X be the event that *a bulb comes from factory* X, and Y be the event *a bulb comes from factory* Y. Then by the *law of total probability*

$$\begin{aligned}\mathbb{P}(H) &= \mathbb{P}(H \mid X)\,\mathbb{P}(X) + \mathbb{P}(H \mid Y)\,\mathbb{P}(Y)\\ &= (0.99)\,(0.6) + (0.95)\,(0.4)\\ &= 0.974.\end{aligned}$$

Solution to Exercise 4.6(B): By Part (a) we have

$$\begin{aligned}\mathbb{P}(Y \mid H) &= \frac{\mathbb{P}(H \mid Y)\,\mathbb{P}(Y)}{\mathbb{P}(H)}\\ &= \frac{(0.95)\,(0.4)}{0.974} \approx 0.39.\end{aligned}$$

Solution to Exercise 4.6(C): We again use the result from Part (a)

$$\begin{aligned}\mathbb{P}(X \mid H^c) &= \frac{\mathbb{P}(H^c \mid X)\,\mathbb{P}(X)}{\mathbb{P}(H^c)} = \frac{\mathbb{P}(H^c \mid X)\,\mathbb{P}(X)}{1 - \mathbb{P}(H)}\\ &= \frac{(1-0.99)\,(0.6)}{1-0.974} = \frac{(0.01)\,(0.6)}{0.026}\\ &\approx 0.23\end{aligned}$$

Solution to Exercise 4.7: denote by D the event that *a bolt is defective*, A the event that *a bolt is from machine* A, by B the event that *a bolt is from machine* C. Then by Bayes' theorem

$$\mathbb{P}(A \mid D) = \frac{\mathbb{P}(D \mid A)\,\mathbb{P}(A)}{\mathbb{P}(D \mid A)\,\mathbb{P}(A) + \mathbb{P}(D \mid B)\,\mathbb{P}(B) + \mathbb{P}(D \mid C)\,\mathbb{P}(C)} = \frac{(0.05)\,(0.25)}{(0.05)\,(0.25) + (0.04)\,(0.35) + (0.02)\,(0.4)} = 0.362.$$

Solution to Exercise 4.8: Let C be the event *a student gives the correct answer*, K be the event *a student knows the correct answer*, E be the event *a student can eliminate one incorrect answer*, and G be the event *a student have to guess an answer*. Using Bayes' theorem we have

$$\mathbb{P}(K \mid C) = \frac{\mathbb{P}(C \mid K)\mathbb{P}(K)}{\mathbb{P}(C)} =$$
$$\frac{\mathbb{P}(C \mid K)\mathbb{P}(K)}{\mathbb{P}(C \mid K)\mathbb{P}(K) + \mathbb{P}(C \mid E)\mathbb{P}(E) + \mathbb{P}(C \mid G)\mathbb{P}(G)} =$$
$$\frac{1 \cdot \frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4}} = \frac{24}{31} \approx .774,$$

that is, approximately 77.4% of the students know the answer if they give the correct answer.

Solution to Exercise 4.9: Let + denote the event that *a test result is positive*, and by D the event that *the disease is present*. Then

$$\begin{aligned}\mathbb{P}(D \mid +) &= \frac{\mathbb{P}(+ \mid D)\, P(D)}{P(+ \mid D)\, P(D) + P(+ \mid D^c)P(D^c)} \\ &= \frac{(0.95)\,(0.01)}{(0.95)\,(0.01) + (0.005)\,(0.99)} = 0.657.\end{aligned}$$

Solution to Exercise 4.2: it is clear that $\mathbb{P}_F(E) = \mathbb{P}(E \mid F)$ is between 0 and 1 since the right-hand side of the identity defining $\mathbb{P}_F$ is. To see the second axiom, observe that

$$\mathbb{P}_F(S) = \mathbb{P}(S \mid F) = \frac{\mathbb{P}(S \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F)}{\mathbb{P}(F)} = 1.$$

Now take $\{E_i\}_{i=1}^{\infty}, E_i \in \mathcal{F}$ to be pairwise disjoint, then

$$\begin{aligned}\mathbb{P}_F\left(\bigcup_{i=1}^{\infty} E_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i \mid F\right) = \frac{\mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} E_i\right) \cap F\right)}{\mathbb{P}(F)} \\ &= \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} (E_i \cap F)\right)}{\mathbb{P}(F)} = \frac{\sum_{i=1}^{\infty} \mathbb{P}(E_i \cap F)}{\mathbb{P}(F)} \\ &= \sum_{i=1}^{\infty} \frac{\mathbb{P}(E_i \cap F)}{\mathbb{P}(F)} = \sum_{i=1}^{\infty} \mathbb{P}_F(E_i).\end{aligned}$$

In this we used the distribution law for sets $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ and the fact that $\{E_i \cap F\}_{i=1}^{\infty}$ are pairwise disjoint as well.

CHAPTER 5

Discrete random variables

5.1. Definition, properties, expectation, moments

As before, suppose S is a sample space.

Definition 5.1 (Random variable)

A *random variable* is a real-valued function on S. Random variables are usually denoted by $X, Y, Z, \ldots$. A *discrete random variable* is one that can take on only countably many values.

Example 5.1. If one rolls a die, let X denote the outcome, i.e. taking values $1, 2, 3, 4, 5, 6$.

Example 5.2. If one rolls a die, let Y be 1 if an odd number is showing, and 0 if an even number is showing.

Example 5.3. If one tosses 10 coins, let X be the number of heads showing.

Example 5.4. In n trials, let X be the number of successes.

Definition (PMF or density of a random variable)

For a discrete random variable X, we define the *probability mass function (PMF)* or the *density of* X by

$$p_X(x) := \mathbb{P}(X = x),$$

where $\mathbb{P}(X = x)$ is a standard abbreviation for

$$\mathbb{P}(X = x) = \mathbb{P}\left(X^{-1}(x)\right).$$

Note that the pre-image $X^{-1}(x)$ is the event $\{\omega \in S : X(\omega) = x\}$.

Suppose X is a discrete random variable taking on values $\{x_i\}_{i\in\mathbb{N}}$, then

$$\sum_{i\in\mathbb{N}} p_X(x_i) = \mathbb{P}(S) = 1.$$

Let X be the number showing if we roll a die. The *expected number* to show up on a roll of a die should be $1 \cdot \mathbb{P}(X = 1) + 2 \cdot \mathbb{P}(X = 2) + \cdots + 6 \cdot \mathbb{P}(X = 6) = 3.5$. More generally, we define

Definition 5.2 (Expectation of a discrete random variable)

For a discrete random variable X we define the *expected value* or *expectation* or *mean* of X as

$$\mathbb{E}X := \sum_{\{x: p_X(x)>0\}} x p_X(x)$$

provided this sum converges absolutely. In this case we say that the expectation of X is well-defined.

We need absolute convergence of the sum so that the expectation does not depend on the order in which we take the sum to define it. We know from calculus that we need to be careful about the sums of conditionally convergent series, though in most of the examples we deal with this will not be a problem. Note that $p_X(x)$ is nonnegative for all x, but x itself can be negative or positive, so in general the terms in the sum might have different signs.

Example 5.5. If we toss a coin and X is 1 if we have heads and 0 if we have tails, what is the expectation of X?

We start with the mass function for X

$$p_X(x) = \begin{cases} \frac{1}{2}, & x = 1 \\ \frac{1}{2}, & x = 0 \\ 0, & \text{all other values of } x. \end{cases}$$

Hence $\mathbb{E}X = (1)(\frac{1}{2}) + (0)(\frac{1}{2}) = \frac{1}{2}$.

Example 5.6. Suppose $X = 0$ with probability $\frac{1}{2}$, 1 with probability $\frac{1}{4}$, 2 with probability $\frac{1}{8}$, and more generally n with probability $1/2^{n+1}$. This is an example where X can take infinitely many values (although still countably many values). What is the expectation of X?

Here $p_X(n) = 1/2^{n+1}$ if n is a nonnegative integer and 0 otherwise. So

$$\mathbb{E}X = (0)\tfrac{1}{2} + (1)\tfrac{1}{4} + (2)\tfrac{1}{8} + (3)\tfrac{1}{16} + \cdots .$$

This turns out to sum to 1. To see this, recall the formula for a geometric series

$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}.$$

If we differentiate this, we get

$$1 + 2x + 3x^2 + \cdots = \frac{1}{(1-x)^2}.$$

We have

$$\begin{aligned}\mathbb{E}X &= 1(\tfrac{1}{4}) + 2(\tfrac{1}{8}) + 3(\tfrac{1}{16}) + \cdots \\ &= \tfrac{1}{4}\Big[1 + 2(\tfrac{1}{2}) + 3(\tfrac{1}{4}) + \cdots\Big] \\ &= \tfrac{1}{4}\frac{1}{(1-\frac{1}{2})^2} = 1.\end{aligned}$$

Example 5.7. Suppose we roll a fair die. If 1 or 2 is showing, let $X = 3$; if a 3 or 4 is showing, let $X = 4$, and if a 5 or 6 is showing, let $X = 10$. What is $\mathbb{E}X$?

We have $\mathbb{P}(X = 3) = \mathbb{P}(X = 4) = \mathbb{P}(X = 10) = \frac{1}{3}$, so

$$\mathbb{E}X = \sum x\mathbb{P}(X = x) = (3)(\tfrac{1}{3}) + (4)(\tfrac{1}{3}) + (10)(\tfrac{1}{3}) = \tfrac{17}{3}.$$

Example 5.8. Consider a discrete random variable taking only positive integers as values with $\mathbb{P}(X = n) = \frac{1}{n(n+1)}$. What is the expectation $\mathbb{E}X$?

First observe that this is indeed a probability since we can use telescoping partial sums to show that

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$$

Then

$$\mathbb{E}X = \sum_{n=1}^{\infty} n \cdot \mathbb{P}(X = n) = \sum_{n=1}^{\infty} n \cdot \frac{1}{n(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n+1} = +\infty,$$

so the expectation of X is infinite.

If we list all possible values of a discrete random variable X as $\{x_i\}_{i\in\mathbb{N}}$, then we can write

$$\mathbb{E}X = \sum_{\{x:p_X(x)>0\}} xp_X(x) = \sum_{i=1}^{\infty} x_i p_X(x_i).$$

We would like to show that the expectation is linear, that is, $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$.

We start by showing that we can write the expectation of a discrete random variable in a slightly different form. Note that in our definition of the expectation we first list all possible values of X and weights with probability that X attains these values. That is, we look at the range of X. Below we instead look at the domain of X and list all possible outcomes.

Proposition 5.1

If X is a random variable on a *finite* sample space S, then

$$\mathbb{E}X = \sum_{\omega\in S} X(\omega)\,\mathbb{P}(\{\omega\}).$$

Proof. For each $i \in \mathbb{N}$ we denote by S_i the event $\{\omega \in S : X(\omega) = x_i\}$. Then $\{S_i\}_{i\in\mathbb{N}}$ is a partition of the space S into disjoint sets. Note that since S is finite, then each set S_i is finite too, moreover, we only have a finite number of sets S_i which are non-empty.

$$\begin{aligned}\mathbb{E}X &= \sum_{i=1}^{\infty} x_i p(x_i) = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i) = \sum_{i=1}^{\infty} x_i \left(\sum_{\omega\in S_i} \mathbb{P}(\{\omega\})\right) \\ &= \sum_{i=1}^{\infty}\left(\sum_{\omega\in S_i} x_i \mathbb{P}(\{\omega\})\right) = \sum_{i=1}^{\infty}\sum_{\omega\in S_i} X(\omega)\mathbb{P}(\{\omega\}) \\ &= \sum_{\omega\in S} X(\omega)\mathbb{P}(\{\omega\}),\end{aligned}$$

where we used properties of sets $\{S_i\}_{i=1}^{\infty}$. □

Proposition 5.1 is true even if S is *countable* as long as $\mathbb{E}X$ is well-defined. First, observe that if S is countable then the random variable X is necessarily discrete. Where do we need to use the assumption that all sums converge absolutely? Note that the identity

$$\sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i) = \sum_{\omega\in S} X(\omega)\mathbb{P}(\{\omega\})$$

is a re-arrangement in the first sum, which we can do as long as the sums (series) converge absolutely. Note that if either the number of values of X or the sample space S is finite, we can use this argument.

Proposition 5.1 can be used to prove linearity of the expectation.

Theorem 5.1 (Linearity of expectation)

If X and Y are discrete random variables defined on the same sample space S and $a \in \mathbb{R}$, then

(i) $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$,
(ii) $\mathbb{E}[aX] = a\mathbb{E}X$,

as long as all expectations are well-defined.

Proof. Consider a random variable $Z := X + Y$ which is a discrete random variable on the sample space S. We use $\mathbb{P}(X = x, Y = y)$ to denote the probability of the event

$$\{\omega \in S : X(\omega) = x\} \cap \{\omega \in S : Y(\omega) = y\}.$$

Denote by $\{x_i\}_{i\in\mathbb{N}}$ the values that X is taking, and by $\{y_j\}_{j\in\mathbb{N}}$ the values that Y is taking. Denote by $\{z_k\}_{k\in\mathbb{N}}$ the values that Z is taking. Since we assume that all random variables have well-defined expectations, we can interchange the order of summations freely. Then by the *law of total probability* (Proposition 4.4) twice we have

$$
\begin{aligned}
\mathbb{E}Z &= \sum_{k=1}^{\infty} z_k \mathbb{P}(Z = z_k) = \sum_{k=1}^{\infty} \left(\sum_{i=1}^{\infty} z_k \mathbb{P}(Z = z_k, X = x_i) \right) \\
&= \sum_{k=1}^{\infty} \left(\sum_{i=1}^{\infty} z_k \mathbb{P}(X = x_i, Y = z_k - x_i) \right) \\
&= \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} z_k \mathbb{P}(X = x_i, Y = z_k - x_i, Y = y_j).
\end{aligned}
$$

Now $\mathbb{P}(X = x_i, Y = z_k - x_i, Y = y_j)$ will be 0, unless $z_k - x_i = y_j$. For each pair (i, j), this will be non-zero for only one value k, since the z_k are all different. Therefore, for each i and j

$$
\begin{aligned}
&\sum_{k=1}^{\infty} z_k \mathbb{P}(X = x_i, Y = z_k - x_i, Y = y_j) \\
&\qquad = \sum_{k=1}^{\infty} (x_i + y_j) \mathbb{P}(X = x_i, Y = z_k - x_i, Y = y_j) \\
&\qquad = (x_i + y_j) \mathbb{P}(X = x_i, Y = y_j).
\end{aligned}
$$

Substituting this to the above sum we see that

$$
\begin{aligned}
\mathbb{E}Z &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (x_i + y_j) \mathbb{P}(X = x_i, Y = y_j) \\
&\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i \mathbb{P}(X = x_i, Y = y_j) + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} y_j \mathbb{P}(X = x_i, Y = y_j) \\
&= \sum_{i=1}^{\infty} x_i \left(\sum_{j=1}^{\infty} \mathbb{P}(X = x_i, Y = y_j) \right) + \sum_{j=1}^{\infty} y_j \left(\sum_{i=1}^{\infty} \mathbb{P}(X = x_i, Y = y_j) \right) \\
&= \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i) + \sum_{j=1}^{\infty} y_j \mathbb{P}(Y = y_j) = \mathbb{E}X + \mathbb{E}Y,
\end{aligned}
$$

where we used the *law of total probability* (Proposition 4.4) again.

Note that if we have a countable sample space all these sums converge absolutely and so we can justify writing this similarly to Proposition 5.1 as

$$
\begin{aligned}
\mathbb{E}\left[X + Y\right] &= \sum_{\omega \in S} \left(X(\omega) + Y(\omega)\right) \mathbb{P}\left(\omega\right) \\
&= \sum_{\omega \in S} \left(X(\omega)\mathbb{P}\left(\omega\right) + Y(\omega)\mathbb{P}\left(\omega\right)\right) \\
&= \sum_{\omega \in S} X(\omega)\mathbb{P}\left(\omega\right) + \sum_{\omega \in S} Y(\omega)\mathbb{P}\left(\omega\right) \\
&= \mathbb{E}X + \mathbb{E}Y.
\end{aligned}
$$

For $a \in \mathbb{R}$ we have

$$\mathbb{E}[aX] = \sum_{\omega \in S} (aX(\omega)) \mathbb{P}(\omega) = a \sum_{\omega \in S} X(\omega) \mathbb{P}(\omega) = a\mathbb{E}X$$

since these sums converge absolutely as long as $\mathbb{E}X$ is well-defined. □

Using induction on the number of random variables linearity holds for a collection of random variables $X_1, X_2, \ldots, X_n$.

Corollary

If $X_1, X_2, \ldots, X_n$ are random variables, then

$$\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}X_1 + \mathbb{E}X_2 + \cdots + \mathbb{E}X_n.$$

Example 5.9. Suppose we roll a die and let X be the value that is showing. We want to find the expectation $\mathbb{E}X^2$ (second moment).

Let $Y = X^2$, so that $\mathbb{P}(Y = 1) = \frac{1}{6}$, $\mathbb{P}(Y = 4) = \frac{1}{6}$ etc. and

$$\mathbb{E}X^2 = \mathbb{E}Y = (1)\tfrac{1}{6} + (4)\tfrac{1}{6} + \cdots + (36)\tfrac{1}{6}.$$

We can also write this as

$$\mathbb{E}X^2 = (1^2)\tfrac{1}{6} + (2^2)\tfrac{1}{6} + \cdots + (6^2)\tfrac{1}{6},$$

which suggests that a formula for $\mathbb{E}X^2$ is $\sum_x x^2 \mathbb{P}(X = x)$. This turns out to be correct.

The only possibility where things could go wrong is if more than one value of X leads to the same value of X^2. For example, suppose $\mathbb{P}(X = -2) = \frac{1}{8}, \mathbb{P}(X = -1) = \frac{1}{4}, \mathbb{P}(X = 1) = \frac{3}{8}, \mathbb{P}(X = 2) = \frac{1}{4}$. Then if $Y = X^2$, $\mathbb{P}(Y = 1) = \frac{5}{8}$ and $\mathbb{P}(Y = 4) = \frac{3}{8}$. Then

$$\mathbb{E}X^2 = (1)\tfrac{5}{8} + (4)\tfrac{3}{8} = (-1)^2\tfrac{1}{4} + (1)^2\tfrac{3}{8} + (-2)^2\tfrac{1}{8} + (2)^2\tfrac{1}{4}.$$

But even in this case $\mathbb{E}X^2 = \sum_x x^2 \mathbb{P}(X = x)$.

Theorem 5.2

For a discrete random variable X taking values $\{x_i\}_{i=1}^{\infty}$ and a real-valued function g defined on this set, we have

$$\mathbb{E}g(X) = \sum_{i=1}^{\infty} g(x_i) \mathbb{P}(X = x_i) = \sum_{i=1}^{\infty} g(x_i) p(x_i).$$

Proof. Let $Y := g(X)$, then

$$\begin{aligned}\mathbb{E}Y &= \sum_y y \mathbb{P}(Y = y) = \sum_y y \sum_{\{x : g(x) = y\}} \mathbb{P}(X = x) \\ &= \sum_x g(x) \mathbb{P}(X = x).\end{aligned}$$

□

Example 5.10. As before we see that $\mathbb{E}X^2 = \sum x^2 p_X(x)$. Also if $g(x) \equiv c$ is a constant function, then we see that the expectation of a constant is this constant

$$\mathbb{E}g(X) = \sum_{i=1}^{\infty} cp(x_i) = c\sum_{i=1}^{\infty} p(x_i) = c \cdot 1 = c.$$

Definition (Moments)

$\mathbb{E}X^n$ is called the *nth moment* of a random variable X. If $M := \mathbb{E}X$ is well defined, then

$$\operatorname{Var}(X) = \mathbb{E}(X - M)^2$$

is called the *variance* of X. The square root of $\operatorname{Var}(X)$ is called the *standard deviation* of X

$$\operatorname{SD}(X) := \sqrt{\operatorname{Var}(X)}.$$

By Theorem 5.2 we know that the nth moment can be calculated by

$$\mathbb{E}X^n = \sum_{x: p_X(x) > 0} x^n p_X(x).$$

The variance measures how much spread there is about the expected value.

Example 5.11. We toss a fair coin and let $X = 1$ if we get heads, $X = -1$ if we get tails. Then $\mathbb{E}X = 0$, so $X - \mathbb{E}X = X$, and then $\operatorname{Var} X = \mathbb{E}X^2 = (1)^2 \frac{1}{2} + (-1)^2 \frac{1}{2} = 1$.

Example 5.12. We roll a die and let X be the value that shows. We have previously calculated $\mathbb{E}X = \frac{7}{2}$. So $X - \mathbb{E}X$ equals

$$-\frac{5}{2}, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \frac{5}{2},$$

each with probability $\frac{1}{6}$. So

$$\operatorname{Var} X = (-\tfrac{5}{2})^2 \tfrac{1}{6} + (-\tfrac{3}{2})^2 \tfrac{1}{6} + (-\tfrac{1}{2})^2 \tfrac{1}{6} + (\tfrac{1}{2})^2 \tfrac{1}{6} + (\tfrac{3}{2})^2 \tfrac{1}{6} + (\tfrac{5}{2})^2 \tfrac{1}{6} = \tfrac{35}{12}.$$

Using the fact that the expectation of a constant is the constant we get an alternate expression for the variance.

Proposition 5.2 (Variance)

Suppose X is a random variable with finite first and second moments. Then

$$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Proof. Denote $M := \mathbb{E}X$, then

$$\begin{aligned}\operatorname{Var} X &= \mathbb{E}X^2 - 2\mathbb{E}(XM) + \mathbb{E}(M^2)\\ &= \mathbb{E}X^2 - 2M^2 + M^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2.\end{aligned}$$

□

5.2. Further examples and applications

5.2.1. Discrete random variables. Recall that we defined a discrete random variable in Definition 5.1 as the one taking countably many values. A *random variable* is a function $X : S \longrightarrow \mathbb{R}$, and we can think of it as a numerical value that is random. When we perform an experiment, many times we are interested in some quantity (a function) related to the outcome, instead of the outcome itself. That means we want to attach a numerical value to each outcome. Below are more examples of such variables.

Example 5.13. Toss a coin and define

$$X = \begin{cases} 1 & \text{if outcome is heads (H)} \\ 0 & \text{if outcome is tails (T).} \end{cases}$$

As a random variable, $X(H) = 1$ and $X(T) = 0$. Note that we can perform computations on real numbers but directly not on the sample space $S = \{H, T\}$. This shows the need to covert outcomes to numerical values.

Example 5.14. Let X be the amount of liability (damages) a driver causes in a year. In this case, X can be any dollar amount. Thus X can attain any value in $[0, \infty)$.

Example 5.15. Toss a coin 3 times. Let X be the number of heads that appear, so that X can take the values $0, 1, 2, 3$. What are the associated probabilities to each value?

$$\begin{aligned} \mathbb{P}(X = 0) =& \mathbb{P}((T,T,T)) = \frac{1}{2^3} = \frac{1}{8}, \\ \mathbb{P}(X = 1) =& \mathbb{P}((T,T,H),(T,H,T),(H,T,T)) = \frac{3}{8}, \\ \mathbb{P}(X = 2) =& \mathbb{P}((T,H,H),(H,H,T),(H,T,H)) = \frac{3}{8}, \\ \mathbb{P}(X = 3) =& \mathbb{P}((H,H,H)) = \frac{1}{8}. \end{aligned}$$

Example 5.16. Toss a coin n times. Let X be the number of heads that occur. This random variable can take the values $0, 1, 2, \dots, n$. From the binomial formula we see that

$$\mathbb{P}(X = k) = \frac{1}{2^n}\binom{n}{k}.$$

Example 5.17. Suppose we toss a fair coin, and we let X be 1 if we have H and X be 0 if we have T. The probability mass function of this random variable is

$$p_X(x) = \begin{cases} \frac{1}{2} & x = 0 \\ \frac{1}{2} & x = 1, \\ 0 & \text{otherwise .} \end{cases}$$

Often the probability mass function (PMF) will already be given and we can then use it to compute probabilities.

Example 5.18. The PMF of a random variable X taking values in $\mathbb{N} \cup \{0\}$ is given by

$$p_X(i) = e^{-\lambda}\frac{\lambda^i}{i!}, i = 0, 1, 2, \ldots,$$

where λ is a positive real number.

(a) Find $\mathbb{P}(X = 0)$. By definition of the PMF we have

$$\mathbb{P}(X = 0) = p_X(0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-\lambda}.$$

(b) Find $\mathbb{P}(X > 2)$. Note that

$$\begin{aligned}\mathbb{P}(X > 2) &= 1 - \mathbb{P}(X \leqslant 2) \\ &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) \\ &= 1 - p_X(0) - p_X(1) - p_X(2) \\ &= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}.\end{aligned}$$

5.2.2. Expectation. We defined the expectation in Definition 5.2 in the case when X is a discrete random variable X taking values $\{x_i\}_{i \in \mathbb{N}}$. Then for a random variable X with the PMF $p_X(x)$ the expectation is given by

$$\mathbb{E}[X] = \sum_{x: p_(x) > 0} x p_X(x) = \sum_{i=1}^{\infty} x_i p_X(x_i).$$

Example 5.19. Suppose again that we have a coin, and let $X(H) = 0$ and $X(T) = 1$. What is $\mathbb{E}X$ if the coin is not necessarily fair?

$$\mathbb{E}X = 0 \cdot p_X(0) + 1 \cdot p_X(1) = \mathbb{P}(T).$$

Example 5.20. Let X be the outcome when we roll a fair die. What is $\mathbb{E}X$?

$$\mathbb{E}X = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2} = 3.5.$$

Note that in the last example X can never be 3.5. This means that the expectation may not be a value attained by X. It serves the purpose of giving an average value for X.

Example 5.21. Let X be the number of insurance claims a person makes in a year. Assume that X can take the values $0, 1, 2, 3 \ldots$ with $\mathbb{P}(X = 0) = \frac{2}{3}$, $\mathbb{P}(X = 1) = \frac{2}{9}, \ldots, \mathbb{P}(X = n) = \frac{2}{3^{n+1}}$. Find the expected number of claims this person makes in a year.

Note that X has infinite but countable number of values, hence it is a discrete random variable. We have that $p_X(i) = \frac{2}{3^{i+1}}$. We compute using the definition of expectation,

$$\begin{aligned}
\mathbb{E}X &= 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) + \cdots \\
&= 0 \cdot \frac{2}{3} + 1\frac{2}{3^2} + 2\frac{2}{3^3} + 3\frac{2}{3^4} + \cdots \\
&= \frac{2}{3^2}\left(1 + 2\frac{1}{3} + 3\frac{1}{3^2} + 4\frac{1}{3^3} + \cdots\right) \\
&= \frac{2}{9}\left(1 + 2x + 3x^2 + \cdots\right), \text{ where } x = \frac{1}{3} \\
&= \frac{2}{9}\frac{1}{(1-x)^2} = \frac{2}{9\left(1-\frac{1}{3}\right)^2} = \frac{2}{2^2} = \frac{1}{2}.
\end{aligned}$$

Example 5.22. Let $S = \{1, 2, 3, 4, 5, 6\}$ and assume that $X(1) = X(2) = 1$, $X(3) = X(4) = 3$, and $X(5) = X(6) = 5$.

(1) Using the initial definition, the random variable X takes the values $1, 3, 5$ and $p_X(1) = p_X(3) = p_X(5) = \frac{1}{3}$. Then

$$\mathbb{E}X = 1 \cdot \frac{1}{3} + 3\frac{1}{3} + 5\frac{1}{3} = \frac{9}{3} = 3.$$

(2) Using the equivalent definition, we list all of $S = \{1, 2, 3, 4, 5, 6\}$ and then

$$\mathbb{E}X = X(1)\mathbb{P}(\{1\}) + \cdots + X(6) \cdot \mathbb{P}(\{6\}) = 1\frac{1}{6} + 1\frac{1}{6} + 3\frac{1}{6} + 3\frac{1}{6} + 5\frac{1}{6} + 1\frac{1}{6} = 3.$$

5.2.3. The cumulative distribution function (CDF). We implicitly used this characterization of a random variable, and now we define it.

Definition 5.3 (Cumulative distribution function)

Let X be a random variable. The *cumulative distribution function* (CDF) or the *distribution function* of X is defined as

$$F_X(x) := \mathbb{P}(X \leqslant x),$$

for any $x \in \mathbb{R}$.

Note that if X is discrete and p_X is its PMF , then

$$F(x_0) = \sum_{x \leqslant x_0} p_X(x).$$

Example 5.23. Suppose that X has the following PMF

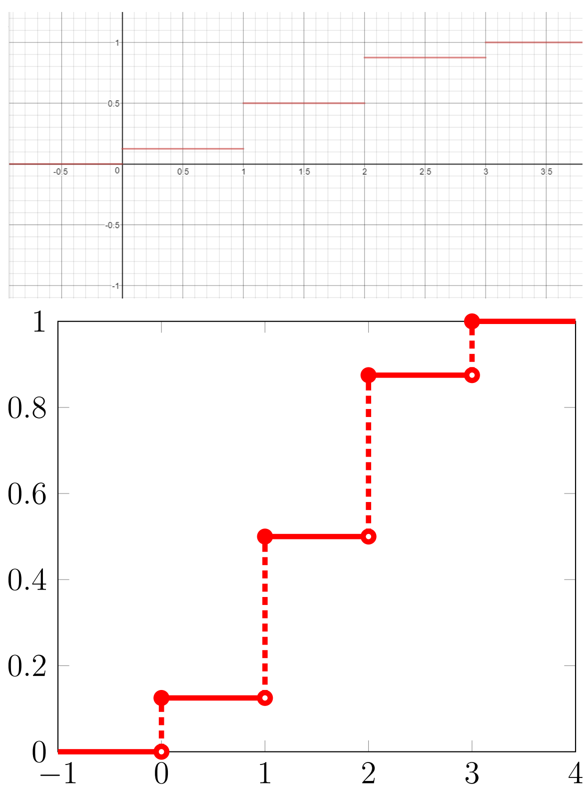
$$
\begin{aligned}
p_X(0) &= \mathbb{P}(X=0) = \frac{1}{8} \\
p_X(1) &= \mathbb{P}(X=1) = \frac{3}{8} \\
p_X(2) &= \mathbb{P}(X=2) = \frac{3}{8} \\
p_X(3) &= \mathbb{P}(X=3) = \frac{1}{8}.
\end{aligned}
$$

Find the CDF for X and plot the graph of the CDF.

This is done by summing up the probabilities up to the value of x we get the following

$$
F_X(x) = \begin{cases} 0 & -\infty < x < 0, \\ \frac{1}{8} & 0 \leqslant x < 1, \\ \frac{4}{8} & 1 \leqslant x < 2, \\ \frac{7}{8} & 2 \leqslant x < 3, \\ 1 & 3 \leqslant x < \infty. \end{cases}
$$

This is a step function shown in the Figure below on page 70.



Above are two graphs for $F_X(x)$ in Example 5.23.

Proposition 5.3 (Properties of cumulative distribution functions(CDF))

1. F is nondecreasing , that is, if $x < y$, then $F(x) \leqslant F(y)$.
2. $\lim_{x\to\infty} F(x) = 1$.
3. $\lim_{x\to-\infty} F(x) = 0$.
4. F is right continuous, that is, $\lim\limits_{u\downarrow x} F_X(u) = F_X(x)$, where $u \downarrow x$ means that u approaches x from above (from the right).

Example 5.24. Let X have distribution

$$F_X(x) = \begin{cases} 0 & x < 0, \\ \frac{x}{2} & 0 \leqslant x < 1, \\ \frac{2}{3} & 1 \leqslant x < 2, \\ \frac{11}{12} & 2 \leqslant x < 3, \\ 1 & 3 \leqslant x. \end{cases}$$

(a) We can find $\mathbb{P}(X < 3)$ by

$$\mathbb{P}(X < 3) = \lim_{n\to\infty} \mathbb{P}\left(X \leqslant 3 - \frac{1}{n}\right) = \lim_{n\to\infty} F_X\left(3 - \frac{1}{n}\right) = \frac{11}{12}.$$

(b) Now we find$\mathbb{P}(X = 1)$.

$$\mathbb{P}(X = 1) = \mathbb{P}(X \leqslant 1) - \mathbb{P}(X < 1) = F_X(1) - \lim_{x\to 1} \frac{x}{2} = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}.$$

(c) Find $\mathbb{P}(2 < X \leqslant 4)$.

$$\mathbb{P}(2 < X \leqslant 4) = F_X(4) - F_X(2) = \frac{1}{12}.$$

5.2.4. Expectation of a function of a random variable. Given a random variable X we would like to compute the expected value of expressions such as X^2, e^X or $\sin X$. How can we do this?

Example 5.25. Let X be a random variable whose PMF is given by

$$\mathbb{P}(X = -1) = 0.2,$$
$$\mathbb{P}(X = 0) = 0.5,$$
$$\mathbb{P}(X = 1) = 0.3.$$

Let $Y = X^2$, then find $\mathbb{E}[Y]$.

Note that Y takes the values 0^2, $(-1)^2$ and 1^2, which reduces to 0 or 1. Also notice that $p_Y(1) = 0.2 + 0.3 = 0.5$ and $p_Y(0) = 0.5$. Thus, $\mathbb{E}[Y] = 0 \cdot 0.5 + 1 \cdot 0.5 = 0.5$.

Note that $\mathbb{E}X^2 = 0.5$. While $(\mathbb{E}X)^2 = 0.01$ since $\mathbb{E}X = 0.3 - 0.2 = 0.1$. Thus in general

$$\mathbb{E}X^2 \neq (\mathbb{E}X)^2.$$

In general, there is a formula for $g(X)$ where g is function that uses the fact that $g(X)$ will be $g(x)$ for some x such that $X = x$. We recall Theorem 5.2. If X is a discrete distribution that takes the values $x_i, i \geq 1$ with probability $p_X(x_i)$, respectively, then for any real valued function g we have that

$$\mathbb{E}\left[g\left(X\right)\right] = \sum_{i=1}^{\infty} g\left(x_i\right) p_X(x_i).$$

Note that

$$\mathbb{E}X^2 = \sum_{i=1}^{\infty} x_i^2 p_X(x_i)$$

will be useful.

Example 5.26. Let us revisit the previous example. Let X denote a random variable such that

$$\begin{aligned} \mathbb{P}\left(X = -1\right) &= 0.2 \\ \mathbb{P}\left(X = 0\right) &= 0.5 \\ \mathbb{P}\left(X = 1\right) &= 0.3. \end{aligned}$$

Let $Y = X^2$, then find $\mathbb{E}Y$.

$$\mathbb{E}Y = \mathbb{E}X^2 = \sum_{i=1}^{\infty} x_i^2 p_X(x_i) = \left(-1\right)^2 (0.2) + 0^2(0.5) + 1^2(0.3) = 0.5$$

5.2.5. Variance. The variance of a random variable is a measure of how spread out the values of X are. The expectation of a random variable is quantity that help us differentiate between random variables, but it does not tell us how spread out its values are. For example, consider

$$\begin{aligned} X &= 0 \text{ with probability } 1 \\ Y &= \begin{cases} -1 & p = \frac{1}{2} \\ 1 & p = \frac{1}{2} \end{cases} \\ Z &= \begin{cases} -100 & p = \frac{1}{2} \\ 100 & p = \frac{1}{2} \end{cases}. \end{aligned}$$

What are the expected values? The are $0, 0$ and 0. But there is much greater spread in Z than Y and Y than X. Thus expectation is not enough to detect spread, or variation.

Example 5.27. Find $\mathrm{Var}(X)$ if X represents the outcome when a fair die is rolled.

Recall that we showed Equation 5.2 to find the variance

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

Previously we found that $\mathbb{E}X = \frac{7}{2}$. Thus we only need to find the second moment

$$\mathbb{E}X^2 = 1^2\left(\frac{1}{6}\right) + \cdots + 6^2\frac{1}{6} = \frac{91}{6}.$$

Using our formula we have that

$$\operatorname{Var}(X) = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Another useful formula is the following.

Proposition 5.4

For any constants $a, b \in \mathbb{R}$ we have that $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$.

Proof. By Equation 5.2 and linearity of expectation

$$\begin{aligned}
\operatorname{Var}(aX+b) &= \mathbb{E}\left(aX+b\right)^2 - \left(\mathbb{E}\left(aX+b\right)\right)^2 \\
&= \mathbb{E}\left(a^2X^2 + 2abX + b^2\right) - \left(a\mathbb{E}X + b\right)^2 \\
&= a^2\mathbb{E}X^2 + 2ab\mathbb{E}X + b^2 - a^2\left(\mathbb{E}X\right)^2 - 2ab\mathbb{E}X - b^2 \\
&= a^2\mathbb{E}X^2 - a^2\left(\mathbb{E}X\right)^2 = a^2\operatorname{Var}(X).
\end{aligned}$$

□

5.3. Exercises

Exercise 5.1. Three balls are randomly chosen with replacement from an urn containing 5 blue, 4 red, and 2 yellow balls. Let X denote the number of red balls chosen.

(a) What are the possible values of X?
(b) What are the probabilities associated to each value?

Exercise 5.2. Two cards are chosen from a standard deck of 52 cards. Suppose that you win \$2 for each heart selected, and lose \$1 for each spade selected. Other suits (clubs or diamonds) bring neither win nor loss. Let X denote your winnings. Determine the probability mass function of X.

Exercise 5.3. A financial regulator from the FED will evaluate two banks this week. For each evaluation, the regulator will choose with equal probability between two different stress tests. Failing under test one costs a bank 10K fee, whereas failing test 2 costs 5K. The probability that the first bank fails any test is 0.4. Independently, the second bank will fail any test with 0.5 probability. Let X denote the total amount of fees the regulator can obtain after having evaluated both banks. Determine the cumulative distribution function of X.

Exercise 5.4. Five buses carry students from Hartford to campus. Each bus carries, respectively, 50, 55, 60, 65, and 70 students. One of these students and one bus driver are picked at random.

(a) What is the expected number of students sitting in the same bus that carries the randomly selected student?
(b) Let Y be the number of students in the same bus as the randomly selected driver. Is $\mathbb{E}[Y]$ larger than the expectation obtained in the previous question?

Exercise 5.5. Two balls are chosen randomly from an urn containing 8 white balls, 4 black, and 2 orange balls. Suppose that we win \$2 for each black ball selected and we lose \$1 for each white ball selected. Let X denote our winnings.

(A) What are the possible values of X?
(B) What are the probabilities associated to each value?

Exercise 5.6. A card is drawn at random from a standard deck of playing cards. If it is a heart, you win \$1. If it is a diamond, you have to pay \$2. If it is any other card, you win \$3. What is the expected value of your winnings?

Exercise 5.7. The game of roulette consists of a small ball and a wheel with 38 numbered pockets around the edge that includes the numbers $1-36, 0$ and 00. As the wheel is spun, the ball bounces around randomly until it settles down in one of the pockets.

(A) Suppose you bet \$1 on a single number and random variable X represents the (monetary) outcome (the money you win or lose). If the bet wins, the payoff is \$35 and you get your money back. If you lose the bet then you lose your \$1. What is the expected profit on a 1 dollar bet?

(B) Suppose you bet \$1 on the numbers $1 - 18$ and random variable X represents the (monetary) outcome (the money you win or lose). If the bet wins, the payoff is \$1 and you get your money back. If you lose the bet then you lose your \$1. What is the expected profit on a 1 dollar bet ?

Exercise 5.8. An insurance company finds that Mark has a 8% chance of getting into a car accident in the next year. If Mark has any kind of accident then the company guarantees to pay him \$10,000. The company has decided to charge Mark a \$200 premium for this one year insurance policy.

(A) Let X be the amount profit or loss from this insurance policy in the next year for the insurance company. Find $\mathbb{E}X$, the expected return for the Insurance company? Should the insurance company charge more or less on its premium?

(B) What amount should the insurance company charge Mark in order to guarantee an expected return of \$100?

Exercise 5.9. A random variable X has the following probability mass function: $p_X(0) = \frac{1}{3}$, $p_X(1) = \frac{1}{6}$, $p_X(2) = \frac{1}{4}$, $p_X(3) = \frac{1}{4}$. Find its expected value, variance, and standard deviation, and plot its CDF.

Exercise 5.10. Suppose X is a random variable such that $\mathbb{E}[X] = 50$ and $\operatorname{Var}(X) = 12$. Calculate the following quantities.

(A) $\mathbb{E}[X^2]$,
(B) $\mathbb{E}[3X + 2]$,
(C) $\mathbb{E}\left[(X+2)^2\right]$,
(D) $\operatorname{Var}[-X]$,
(E) $\operatorname{SD}(2X)$.

Exercise 5.11. Does there exist a random variable X such that $\mathbb{E}[X] = 4$ and $\mathbb{E}[X^2] = 10$? Why or why not? (Hint: look at its variance)

Exercise 5.12. A box contains 25 peppers of which 5 are red and 20 green. Four peppers are randomly picked from the box. What is the expected number of red peppers in this sample of four?

5.4. Selected solutions

Solution to Exercise 5.1:

(a) X can take the values $0, 1, 2$ and 3.
(b) Since balls are withdrawn with replacement, we can think of *choosing red* as a success and apply Bernoulli trials with $p = \mathbb{P}(\text{red}) = \frac{4}{11}$. Then, for each $k = 0, 1, 2, 3$ we have

$$\mathbb{P}(X = k) = \binom{3}{k}\left(\frac{4}{11}\right)^k \cdot \left(\frac{7}{11}\right)^{3-k}.$$

Solution to Exercise 5.2: The random variable X can take the values $-2, -1, 0, 1, 2, 4$. Moreover,

$$\begin{aligned}
\mathbb{P}(X = -2) &= \mathbb{P}(2\spadesuit) = \frac{\binom{13}{2}}{\binom{52}{2}}, \\
\mathbb{P}(X = -1) &= \mathbb{P}(1\spadesuit \text{ and } 1(\diamondsuit \text{ or } \clubsuit)) = \frac{13 \cdot 26}{\binom{52}{2}}, \\
\mathbb{P}(X = 0) &= \mathbb{P}(2(\diamondsuit \text{ or } \clubsuit)) = \frac{\binom{26}{2}}{\binom{52}{2}}, \\
\mathbb{P}(X = 1) &= \mathbb{P}(1\heartsuit \text{ and } 1\spadesuit) = \frac{13 \cdot 13}{\binom{52}{2}}, \\
\mathbb{P}(X = 2) &= \mathbb{P}(1\heartsuit \text{ and } 1(\diamondsuit \text{ or } \clubsuit)) = \mathbb{P}(X = -1), \\
\mathbb{P}(X = 4) &= \mathbb{P}(2\heartsuit) = \mathbb{P}(X = -2).
\end{aligned}$$

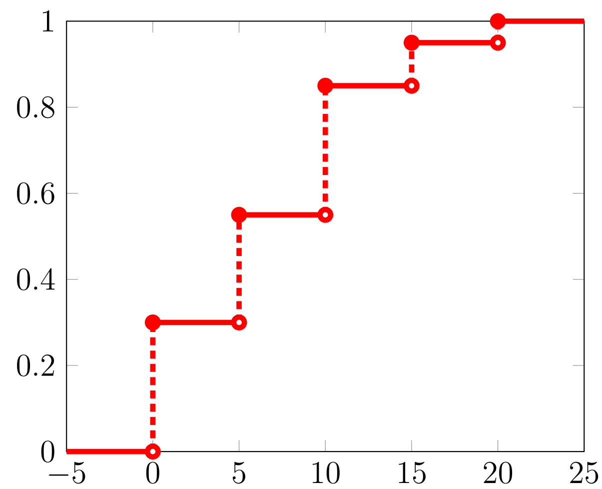
Thus the probability mass function is given by $p_X(x) = \mathbb{P}(X = x)$ for $x = -2, -1, 0, 1, 2, 4$ and $p_X(x) = 0$ otherwise.

Solution to Exercise 5.3: The random variable X can take the values $0, 5, 10, 15$ and 20 depending on which test was applied to each bank, and if the bank fails the evaluation or not. Denote by B_i the event that the *ith bank fails* and by T_i the event that *test i applied*. Then

$$\begin{aligned}
&\mathbb{P}(T_1) = \mathbb{P}(T_2) = 0.5, \mathbb{P}(B_1) = \mathbb{P}(B_1 \mid T_1) = \mathbb{P}(B_1 \mid T_2) = 0.4 \\
&\mathbb{P}(B_2) = \mathbb{P}(B_2 \mid T_1) = \mathbb{P}(B_2 \mid T_2) = 0.5.
\end{aligned}$$

Since banks and tests are independent we have

$$\begin{aligned}
\mathbb{P}(X = 0) &= \mathbb{P}(B_1^c \cap B_2^c) = \mathbb{P}(B_1^c) \cdot \mathbb{P}(B_2^c) = 0.6 \cdot 0.5 = 0.3, \\
\mathbb{P}(X = 5) &= \mathbb{P}(B_1)\mathbb{P}(T_1)\mathbb{P}(B_2^c) + \mathbb{P}(B_1^c)\mathbb{P}(B_2)\mathbb{P}(T_2) = 0.25, \\
\mathbb{P}(X = 10) &= \mathbb{P}(B_1)\mathbb{P}(T_1)\mathbb{P}(B_2^c) + \mathbb{P}(B_1)\mathbb{P}(T_2)\mathbb{P}(B_2)\mathbb{P}(T_2) + \mathbb{P}(B_1^c)\mathbb{P}(B_2)\mathbb{P}(T_1) = 0.3 \\
\mathbb{P}(X = 15) &= \mathbb{P}(B_1)\mathbb{P}(T_1)\mathbb{P}(B_2)\mathbb{P}(T_2) + \mathbb{P}(B_1)\mathbb{P}(T_2)\mathbb{P}(B_2)\mathbb{P}(T_1) = 0.1 \\
\mathbb{P}(X = 20) &= \mathbb{P}(B_1)\mathbb{P}(T_1)\mathbb{P}(B_2)\mathbb{P}(T_1) = 0.05.
\end{aligned}$$



The graph of the probability distribution function for Exercise 5.3

The probability distribution function is given by

$$F_X(x) = \begin{cases} 0 & x < 0, \\ 0.3 & 0 \leqslant x < 5, \\ 0.55 & 5 \leqslant x < 10, \\ 0.85 & 10 \leqslant x < 15, \\ 0.95 & 15 \leqslant x < 20, \\ 1 & x \geqslant 20. \end{cases}$$

Solution to Exercise 5.4: Let X denote the number of students in the bus that carries the randomly selected student.

(a) In total there are 300 students, hence $\mathbb{P}(X = 50) = \frac{50}{300}$, $\mathbb{P}(X = 55) = \frac{55}{300}$, $\mathbb{P}(X = 60) = \frac{60}{300}$, $\mathbb{P}(X = 65) = \frac{65}{300}$ and $\mathbb{P}(X = 70) = \frac{70}{300}$. The expected value of X is thus

$$\mathbb{E}[X] = 50\frac{50}{300} + 55\frac{55}{300} + 60\frac{60}{300} + 65\frac{65}{300} + 70\frac{70}{300} \approx 60.8333.$$

(b) In this case, the probability of choosing a bus driver is $\frac{1}{5}$, so that

$$\mathbb{E}[Y] = \frac{1}{5}(50 + 55 + 60 + 65 + 70) = 60$$

which is slightly less than the previous one.

Solution to Exercise 5.5(A): Note that $X = -2, -1, -0, 1, 2, 4$.

Solution to Exercise 5.5(B): below is the list of all probabilities.

$$\mathbb{P}(X=4)=\mathbb{P}(\{BB\})=\frac{\binom{4}{2}}{\binom{14}{2}}=\frac{6}{91},$$
$$\mathbb{P}(X=0)=\mathbb{P}(\{OO\})=\frac{\binom{2}{2}}{\binom{14}{2}}=\frac{1}{91}$$
$$\mathbb{P}(X=2)=\mathbb{P}(\{BO\})=\frac{\binom{4}{1}\binom{2}{1}}{\binom{14}{2}}=\frac{8}{91},$$
$$\mathbb{P}(X=-1)=\mathbb{P}(\{WO\})=\frac{\binom{8}{1}\binom{2}{1}}{\binom{14}{2}}=\frac{16}{91},$$
$$\mathbb{P}(X=1)=\mathbb{P}(\{BW\})=\frac{\binom{4}{1}\binom{8}{1}}{\binom{14}{2}}=\frac{32}{91},$$
$$\mathbb{P}(X=-2)=\mathbb{P}(\{WW\})=\frac{\binom{8}{2}}{\binom{14}{2}}=\frac{28}{91}$$

Solution to Exercise 5.6:

$$\mathbb{E}X=1\cdot\frac{1}{4}+(-2)\frac{1}{4}+3\cdot\frac{1}{2}=\frac{5}{4}$$

Solution to Exercise 5.7(A): The expected profit is $\mathbb{E}X=35\cdot\left(\frac{1}{38}\right)-1\cdot\frac{37}{38}=-\$0.0526$.

Solution to Exercise 5.7(B): If you will then your profit will be \$1. If you lose then you lose your \$1 bet. The expected profit is $\mathbb{E}X=1\cdot\left(\frac{18}{38}\right)-1\cdot\frac{20}{38}=-\$0.0526$.

Solution to Exercise 5.8(A): If Mark has no accident then the company makes a profit of 200 dollars. If Mark has an accident they have to pay him 10,000 dollars, but regardless they received 200 dollars from him as an yearly premium. We have

$$\mathbb{E}X=(200-10,000)\cdot(0.08)+200\cdot(0.92)=-600.$$

On average the company will lose \$600 dollars. Thus the company should charge more.

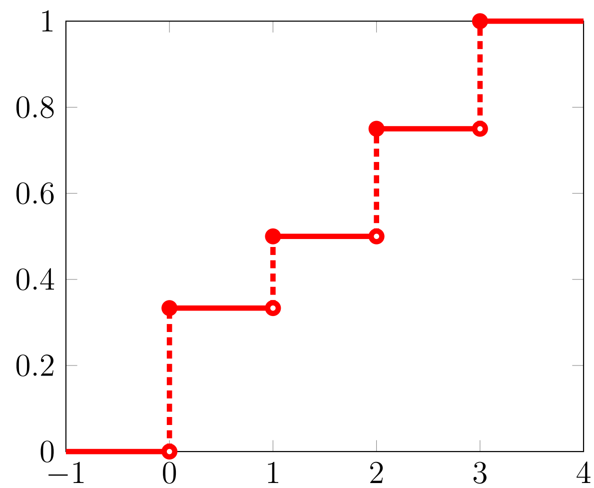
Solution to Exercise 5.8(B): Let P be the premium. Then in order to guarantee an expected return of 100 then

$$100=\mathbb{E}X=(P-10,000)\cdot(0.08)+P\cdot(0.92)$$

and solving for P we get $P=\$900$.

Solution to Exercise 5.9: we start with the expectation

$$\mathbb{E}X=0\cdot\frac{1}{3}+1\cdot\frac{1}{6}+2\cdot\frac{1}{4}+3\cdot\frac{1}{4}=\frac{34}{24}.$$



The plot of the CDF for Exercise 5.9

Now to find the variance we have

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 \\ &= 0^2 \cdot \frac{1}{3} - 1^2\frac{1}{6} + 2^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{1}{4} - \left(\frac{34}{24}\right)^2 \\ &= \frac{82}{24} - \frac{34^2}{24^2} = \frac{812}{24^2}.\end{aligned}$$

Taking the square root gives us

$$\mathrm{SD}(X) = \frac{2\sqrt{203}}{24}.$$

Solution to Exercise 5.10(A): Since $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = 12$ then

$$\mathbb{E}\left[X^2\right] = \mathrm{Var}(X) + (\mathbb{E}X)^2 = 12 + 50^2 = 2512.$$

Solution to Exercise 5.10(B):

$$\mathbb{E}[3X+2] = 3\mathbb{E}[X] + \mathbb{E}[2] = 3 \cdot 50 + 2 = 152.$$

Solution to Exercise 5.10(C):

$$\mathbb{E}\left[(X+2)^2\right] = \mathbb{E}\left[X^2\right] + 4\mathbb{E}[X] + 4 = 2512 + 4 \cdot 50 + 4 = 2716.$$

Solution to Exercise 5.10(D):

$$\mathrm{Var}[-X] = (-1)^2\,\mathrm{Var}(X) = 12$$

Solution to Exercise 5.10(E):

$$\mathrm{SD}(2X) = \sqrt{\mathrm{Var}(2X)} = \sqrt{2^2\,\mathrm{Var}(X)} = \sqrt{48} = 2\sqrt{12}.$$

Solution to Exercise 5.11: Using the hint let's compute the variance of this random variable which would be $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = 10 - 4^2 = -6$. But we know a random variable cannot have a negative variance. Thus no such a random variable exists.

CHAPTER 6

Some discrete distributions

6.1. Examples: Bernoulli, binomial, Poisson, geometric distributions

Bernoulli distribution

A random variable X such that $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$ is said to be a *Bernoulli random variable* with parameter p. Note $\mathbb{E}X = p$ and $\mathbb{E}X^2 = p$, so $\operatorname{Var} X = p - p^2 = p(1-p)$.

We denote such a random variable by $X \sim \operatorname{Bern}(p)$.

Binomial distribution

A random variable X has a *binomial distribution* with parameters n and p if $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

We denote such a random variable by $X \sim \operatorname{Binom}(n, p)$.

The number of successes in n Bernoulli trials is a binomial random variable. After some cumbersome calculations one can derive $\mathbb{E}X = np$. An easier way is to realize that if X is binomial, then $X = Y_1 + \cdots + Y_n$, where the Y_i are independent Bernoulli variables, so $\mathbb{E}X = \mathbb{E}Y_1 + \cdots + \mathbb{E}Y_n = np$.

We have not defined yet what it means for random variables to be independent, but here we mean that the events such as $(Y_i = 1)$ are independent.

Proposition 6.1

Suppose $X := Y_1 + \cdots + Y_n$, where $\{Y_i\}_{i=1}^n$ are independent Bernoulli random variables with parameter p, then

$$\mathbb{E}X = np, \operatorname{Var} X = np(1-p).$$

Proof. First we use the definition of expectation to see that

$$\mathbb{E}X = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^{n} i \binom{n}{i} p^i (1-p)^{n-i}.$$

Then

$$\begin{aligned}\mathbb{E}X &= \sum_{i=1}^{n} i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\ &= np \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!((n-1)-(i-1))!} p^{i-1} (1-p)^{(n-1)-(i-1)} \\ &= np \sum_{i=0}^{n-1} \frac{(n-1)!}{i!((n-1)-i)!} p^i (1-p)^{(n-1)-i} \\ &= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{(n-1)-i} = np,\end{aligned}$$

where we used the Binomial Theorem (Theorem 1.1).

To get the variance of X, we first observe that

$$\mathbb{E}X^2 = \sum_{i=1}^{n} \mathbb{E}Y_i^2 + \sum_{i \neq j} \mathbb{E}Y_i Y_j.$$

Now

$$\begin{aligned}\mathbb{E}Y_i Y_j &= 1 \cdot \mathbb{P}(Y_i Y_j = 1) + 0 \cdot \mathbb{P}(Y_i Y_j = 0) \\ &= \mathbb{P}(Y_i = 1, Y_j = 1) = \mathbb{P}(Y_i = 1)\mathbb{P}(Y_j = 1) = p^2\end{aligned}$$

using independence of random variables $\{Y_i\}_{i=1}^n$. Expanding $(Y_1 + \cdots + Y_n)^2$ yields n^2 terms, of which n are of the form Y_k^2. So we have $n^2 - n$ terms of the form $Y_i Y_j$ with $i \neq j$. Hence

$$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = np + (n^2 - n)p^2 - (np)^2 = np(1-p).$$

□

Later we will see that the variance of the sum of independent random variables is the sum of the variances, so we could quickly get $\operatorname{Var} X = np(1-p)$. Alternatively, one can compute $\mathbb{E}(X^2) - \mathbb{E}X = \mathbb{E}(X(X-1))$ using binomial coefficients and derive the variance of X from that.

Poisson distribution

A random variable X has the *Poisson distribution* with parameter λ if

$$\mathbb{P}(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}.$$

We denote such a random variable by $X \sim \operatorname{Pois}(\lambda)$. Note that

$$\sum_{i=0}^{\infty} \lambda^i / i! = e^{\lambda},$$

so the probabilities add up to one.

Proposition 6.2

Suppose X is a Poisson random variable with parameter λ, then

$$\mathbb{E}X = \lambda,$$
$$\operatorname{Var} X = \lambda.$$

PROOF. We start with the expectation

$$\mathbb{E}X = \sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda.$$

Similarly one can show that

$$\begin{aligned}\mathbb{E}(X^2) - \mathbb{E}X = \mathbb{E}X(X-1) &= \sum_{i=0}^{\infty} i(i-1) e^{-\lambda} \frac{\lambda^i}{i!} \\ &= \lambda^2 e^{-\lambda} \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} \\ &= \lambda^2,\end{aligned}$$

so $\mathbb{E}X^2 = \mathbb{E}(X^2 - X) + EX = \lambda^2 + \lambda$, and hence $\operatorname{Var} X = \lambda$. □

Example 6.1. Suppose on average there are 5 homicides per month in a given city. What is the probability there will be at most 1 in a certain month?

If X denotes the number of homicides, then we are given that $\mathbb{E}X = 5$. Since the expectation for a Poisson is λ, then $\lambda = 5$. Therefore $\mathbb{P}(X=0) + \mathbb{P}(X=1) = e^{-5} + 5e^{-5}$.

Example 6.2. Suppose on average there is one large earthquake per year in California. What is the probability that next year there will be exactly 2 large earthquakes? $\lambda = \mathbb{E}X = 1$, so $\mathbb{P}(X=2) = \frac{e^{-1}}{2}$.

We have the following proposition connecting binomial and Poisson distributions.

Proposition 6.3 (Binomial approximation of Poisson distribution)

If X_n is a binomial random variable with parameters n and p_n and $np_n \to \lambda$, then $\mathbb{P}(X_n = i) \to \mathbb{P}(Y = i)$, where Y is Poisson with parameter λ.

6.1 (Approximation of Poisson by binomials)

Note that by setting

$$p_n := \lambda/n \quad \text{for} \quad n > \lambda$$

we can approximate the Poisson distribution with parameter λ by binomial distributions with parameters n and p_n.

This proposition shows that the Poisson distribution models binomials when the probability of a success is small. The number of misprints on a page, the number of automobile accidents, the number of people entering a store, etc. can all be modeled by a Poisson distribution.

Proof. For simplicity, let us suppose that $\lambda = np_n$ for $n > \lambda$. In the general case we can use $\lambda_n = np_n \xrightarrow[n\to\infty]{} \lambda$. We write

$$\begin{aligned}\mathbb{P}(X_n = i) &= \frac{n!}{i!(n-i)!} p_n^i (1-p_n)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}.\end{aligned}$$

Observe that the following three limits exist

$$\begin{aligned}&\frac{n(n-1)\cdots(n-i+1)}{n^i} \xrightarrow[n\to\infty]{} 1, \\ &(1-\lambda/n)^i \xrightarrow[n\to\infty]{} 1, \\ &(1-\lambda/n)^n \xrightarrow[n\to\infty]{} e^{-\lambda},\end{aligned}$$

which completes the proof. □

In Section 2.2.3 we considered **discrete uniform distributions** with $\mathbb{P}(X = k) = \frac{1}{n}$ for $k = 1, 2, \ldots, n$. This is the distribution of the number showing on a die (with $n = 6$), for example.

Geometric distribution

A random variable X has the geometric distribution with parameter p, $0 < p < 1$, if

$$\mathbb{P}(X = i) = (1-p)^{i-1} p \quad \text{for} \quad i = 1, 2, \ldots.$$

Using a geometric series sum formula we see that

$$\sum_{i=1}^{\infty} \mathbb{P}(X = i) = \sum_{i=1}^{\infty} (1-p)^{i-1} p = \frac{1}{1-(1-p)} p = 1.$$

In Bernoulli trials, if we let X be the first time we have a success, then X will be a geometric random variable. For example, if we toss a coin over and over and X is the first time we get a heads, then X will have a geometric distribution. To see this, to have the first success occur on the k^{th} trial, we have to have $k-1$ failures in the first $k-1$ trials and then a success. The probability of that is $(1-p)^{k-1}p$.

Proposition 6.4

If X is a geometric random variable with parameter p, $0 < p < 1$, then

$$\begin{aligned}\mathbb{E}X &= \frac{1}{p},\\ \operatorname{Var} X &= \frac{1-p}{p^2},\\ F_X(k) &= \mathbb{P}(X \leqslant k) = 1-(1-p)^k.\end{aligned}$$

Proof. We will use

$$\frac{1}{(1-r)^2} = \sum_{n=0}^{\infty} n r^{n-1}$$

which we can show by differentiating the formula for geometric series $1/(1-r) = \sum_{n=0}^{\infty} r^n$. Then

$$\mathbb{E}X = \sum_{i=1}^{\infty} i \cdot \mathbb{P}(X=i) = \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} p = \frac{1}{(1-(1-p))^2} \cdot p = \frac{1}{p}.$$

Then the variance

$$\operatorname{Var} X = \mathbb{E}(X-\mathbb{E}X)^2 = \mathbb{E}\left(X-\frac{1}{p}\right)^2 = \sum_{i=1}^{\infty}\left(i-\frac{1}{p}\right)^2 \cdot \mathbb{P}(X=i)$$

To find the variance we will use another sum. First

$$\frac{r}{(1-r)^2} = \sum_{n=0}^{\infty} n r^n,$$

which we can differentiate to see that

$$\frac{1+r}{(1-r)^3} = \sum_{n=1}^{\infty} n^2 r^{n-1}.$$

Then

$$\mathbb{E}X^2 = \sum_{i=1}^{\infty} i^2 \cdot \mathbb{P}(X=i) = \sum_{i=1}^{\infty} i^2 \cdot (1-p)^{i-1} p = \frac{(1+(1-p))}{(1-(1-p))^3} \cdot p = \frac{2-p}{p^2}.$$

Thus

$$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

The cumulative distribution function (CDF) can be found by using the geometric series sum formula

$$1-F_X(k) = \mathbb{P}(X>k) = \sum_{i=k+1}^{\infty} \mathbb{P}(X=i) = \sum_{i=k+1}^{\infty} (1-p)^{i-1} p = \frac{(1-p)^k}{1-(1-p)} p = (1-p)^k.$$

□

Negative binomial distribution

A random variable X has *negative binomial distribution* with parameters r and p if

$$\mathbb{P}(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \ldots.$$

A negative binomial represents the number of trials until r successes. To get the above formula, to have the r^{th} success in the n^{th} trial, we must exactly have $r-1$ successes in the first $n-1$ trials and then a success in the n^{th} trial.

Hypergeometric distribution

A random variable X has hypergeometric distribution with parameters m, n and N if

$$\mathbb{P}(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}.$$

This comes up in sampling without replacement: if there are N balls, of which m are one color and the other $N-m$ are another, and we choose n balls at random without replacement, then X represents the probability of having i balls of the first color.

Another model where the hypergeometric distribution comes up is the probability of a success changes on each draw, since each draw decreases the population, in other words, when we consider sampling without replacement from a finite population). Then N is the population size, m is the number of success states in the population, n is the number of draws, that is, quantity drawn in each trial, i is the number of observed successes.

6.2. Further examples and applications

6.2.1. Bernoulli and binomial random variables.

Example 6.3. A company prices its hurricane insurance using the following assumptions:

(i) In any calendar year, there can be at most one hurricane.
(ii) In any calendar year, the probability of a hurricane is 0.05.
(iii) The numbers of hurricanes in different calendar years are mutually independent.

Using the company's assumptions, find the probability that there are fewer than 3 hurricanes in a 20-year period.

Denote by X the number of hurricanes in a 20-year period. From the assumptions we see that $X \sim \operatorname{Binom}(20, 0.05)$, therefore

$$\begin{aligned}
\mathbb{P}(X < 3) &= \mathbb{P}(X \leqslant 2) \\
&= \binom{20}{0}(0.05)^0(0.95)^{20} + \binom{20}{1}(0.05)^1(0.95)^{19} + \binom{20}{2}(0.05)^2(0.95)^{18} \\
&= 0.9245.
\end{aligned}$$

Example 6.4. Phan has a 0.6 probability of making a free throw. Suppose each free throw is independent of the other. If he attempts 10 free throws, what is the probability that he makes at least 2 of them?

If $X \sim \operatorname{Binom}(10, 0.6)$, then

$$\begin{aligned}
\mathbb{P}(X \geqslant 2) &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \\
&= 1 - \binom{10}{0}(0.6)^0(0.4)^{10} - \binom{10}{1}(0.6)^1(0.4)^9 \\
&= 0.998.
\end{aligned}$$

6.2.2. The Poisson distribution. Recall that a Poisson distribution models well events that have a low probability and the number of trials is high. For example, the probability of a misprint is small and the number of words in a page is usually a relatively large number compared to the number of misprints.

(1) The number of misprints on a random page of a book.
(2) The number of people in community that survive to age 100.
(3) The number of telephone numbers that are dialed in an average day.
(4) The number of customers entering post office on an average day.

Example 6.5. Levi receives an average of two texts every 3 minutes. If we assume that the number of texts is Poisson distributed, what is the probability that he receives five or more texts in a 9-minute period?

Let X be the number of texts in a 9−minute period. Then $\lambda = 3 \cdot 2 = 6$ and

$$\begin{aligned}\mathbb{P}(X \geqslant 5) &= 1 - \mathbb{P}(X \leqslant 4) \\ &= 1 - \sum_{n=0}^{4} \frac{6^n e^{-6}}{n!} \\ &= 1 - 0.285 = 0.715.\end{aligned}$$

Example 6.6. Let X_1, ..., X_k be independent Poisson random variables, each with expectation λ. What is the distribution of the random variable $Y := X_1 + ... + X_k$?

The distribution of Y is Poisson with the expectation $\lambda = k\lambda$. To show this, we use Proposition 6.3 and (6.1) to choose $n = mk$ Bernoulli random variables with parameter $p_n = k\lambda_1/n = \lambda_1/m = \lambda/n$ to approximation the Poisson random variables. If we sum them all together, the limit as $n \to \infty$ gives us a Poisson distribution with expectation $\lim\limits_{n\to\infty} np_n = \lambda$. However, we can re-arrange the same $n = mk$ Bernoulli random variables in k groups, each group having m Bernoulli random variables. Then the limit gives us the distribution of $X_1 + ... + X_k$. This argument can be made rigorous, but this is beyond the scope of this course. Note that we do not show that the we have convergence in distribution.

Example 6.7. Let $X_1, \ldots, X_k$ be independent Poisson random variables, each with expectation $\lambda_1, \ldots, \lambda_k$, respectively. What is the distribution of the random variable $Y = X_1 + ... + X_k$?

The distribution of Y is Poisson with expectation $\lambda = \lambda_1 + ... + \lambda_k$. To show this, we again use Proposition 6.3 and (6.1) with parameter $p_n = \lambda/n$. If n is large, we can separate these n Bernoulli random variables in k groups, each having $n_i \approx \lambda_i n/\lambda$ Bernoulli random variables. The result follows if $\lim\limits_{n\to\infty} n_i/n = \lambda_i$ for each $i = 1, ..., k$.

This entire set-up, which is quite common, involves what is called *independent identically distributed Bernoulli random variables* (i.i.d. Bernoulli r.v.).

Example 6.8. Can we use binomial approximation to find the mean and the variance of a Poisson random variable?

Yes, and this is really simple. Recall again from Proposition 6.3 and (6.1) that we can approximate Poisson Y with parameter λ by a binomial random variable $\text{Binom}(n, p_n)$, where $p_n = \lambda/n$. Each such a binomial random variable is a sum on n independent Bernoulli random variables with parameter p_n. Therefore

$$\begin{aligned}\mathbb{E}Y &= \lim_{n\to\infty} np_n = \lim_{n\to\infty} n\frac{\lambda}{n} = \lambda, \\ \operatorname{Var}(Y) &= \lim_{n\to\infty} np_n(1-p_n) = \lim_{n\to\infty} n\frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right) = \lambda.\end{aligned}$$

6.2.3. Table of distributions. The following table summarizes the discrete distributions we have seen in this chapter. Here $\mathbb{N}$ stands for the set of positive integers, and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ is the set of nonnegative integers.

Name	Notation	Parameters	PMF ($k \in \mathbb{N}_0$)	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
Bernoulli	$\mathrm{Bern}(p)$	$p \in [0,1]$	$\binom{1}{k} p^k (1-p)^{1-k}$	p	$p(1-p)$
Binomial	$\mathrm{Binom}(n,p)$	$n \in \mathbb{N}$ $p \in [0,1]$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Poisson	$\mathrm{Pois}(\lambda)$	$\lambda > 0$	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ
Geometric	$\mathrm{Geo}(p)$	$p \in (0,1)$	$\begin{cases} (1-p)^{k-1} p, & \text{for } k \geqslant 1, \\ 0, & \text{else.} \end{cases}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial	$\mathrm{NBin}(r,p)$	$r \in \mathbb{N}$ $p \in (0,1)$	$\begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r}, & \text{if } k \geq r, \\ 0, & \text{else.} \end{cases}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hyper-geometric	$\mathrm{Hyp}(N,m,n)$	$N \in \mathbb{N}_0$ $n, m \in \mathbb{N}_0$	$\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$	$\frac{nm}{N}$	$\frac{nm(N-n)}{N(N-1)}(1-\frac{m}{N})$

6.3. Exercises

Exercise 6.1. A UConn student claims that she can distinguish Dairy Bar ice cream from Friendly's ice cream. As a test, she is given ten samples of ice cream (each sample is either from the Dairy Bar or Friendly's) and asked to identify each one. She is right eight times. What is the probability that she would be right exactly eight times if she guessed randomly for each sample?

Exercise 6.2. A Pharmaceutical company conducted a study on a new drug that is supposed to treat patients suffering from a certain disease. The study concluded that the drug did not help 25% of those who participated in the study. What is the probability that of 6 randomly selected patients, 4 will recover?

Exercise 6.3. 20% of all students are left-handed. A class of size 20 meets in a room with 18 right-handed desks and 5 left-handed desks. What is the probability that every student will have a suitable desk?

Exercise 6.4. A ball is drawn from an urn containing 4 blue and 5 red balls. After the ball is drawn, it is replaced and another ball is drawn. Suppose this process is done 7 times.

(a) What is the probability that exactly 2 red balls were drawn in the 7 draws?
(b) What is the probability that at least 3 blue balls were drawn in the 7 draws?

Exercise 6.5. The expected number of typos on a page of the new Harry Potter book is 0.2. What is the probability that the next page you read contains

(a) 0 typos?
(b) 2 or more typos?
(c) Explain what assumptions you used.

Exercise 6.6. The monthly average number of car crashes in Storrs, CT is 3.5. What is the probability that there will be

(a) at least 2 accidents in the next month?
(b) at most 1 accident in the next month?
(c) Explain what assumptions you used.

Exercise 6.7. Suppose that, some time in a distant future, the average number of burglaries in New York City in a week is 2.2. Approximate the probability that there will be

(a) no burglaries in the next week;
(b) at least 2 burglaries in the next week.

Exercise 6.8. The number of accidents per working week in a particular shipyard is Poisson distributed with mean 0.5. Find the probability that:

(a) In a particular week there will be at least 2 accidents.

(b) In a particular two week period there will be exactly 5 accidents.
(c) In a particular month (i.e. 4 week period) there will be exactly 2 accidents.

Exercise 6.9. Jennifer is baking cookies. She mixes 400 raisins and 600 chocolate chips into her cookie dough and ends up with 500 cookies.

(a) Find the probability that a randomly picked cookie will have three raisins in it.
(b) Find the probability that a randomly picked cookie will have at least one chocolate chip in it.
(c) Find the probability that a randomly picked cookie will have no more than two bits in it (a bit is either a raisin or a chocolate chip).

Exercise 6.10. A roulette wheel has 38 numbers on it: the numbers 0 through 36 and a 00. Suppose that Lauren always bets that the outcome will be a number between 1 and 18 (including 1 and 18).

(a) What is the probability that Lauren will lose her first 6 bets.
(b) What is the probability that Lauren will first win on her sixth bet?

Exercise 6.11. In the US, albinism occurs in about one in 17,000 births. Estimate the probabilities no albino person, of at least one, or more than one albino at a football game with 5,000 attendants. Use the Poisson approximation to the binomial to estimate the probability.

Exercise 6.12. An egg carton contains 20 eggs, of which 3 have a double yolk. To make a pancake, 5 eggs from the carton are picked at random. What is the probability that at least 2 of them have a double yolk?

Exercise 6.13. Around 30,000 couples married this year in CT. Approximate the probability that at least in one of these couples

(a) both partners have birthday on January 1st.
(b) both partners celebrate birthday in the same month.

Exercise 6.14. A telecommunications company has discovered that users are three times as likely to make two-minute calls as to make four-minute calls. The length of a typical call (in minutes) has a Poisson distribution. Find the expected length (in minutes) of a typical call.

6.4. Selected solutions

Solution to Exercise 6.1: This should be modeled using a binomial random variable X, since there is a sequence of trials with the same probability of success in each one. If she guesses randomly for each sample, the probability that she will be right each time is $\frac{1}{2}$. Therefore,

$$\mathbb{P}(X=8) = \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 = \frac{45}{2^{10}}.$$

Solution to Exercise 6.2: $\binom{6}{4}(0.75)^4(0.25)^2$

Solution to Exercise 6.3: For each student to have the kind of desk he or she prefers, there must be no more than 18 right-handed students and no more than 5 left-handed students, so the number of left-handed students must be between 2 and 5 (inclusive). This means that we want the probability that there will be 2, 3, 4, or 5 left-handed students. We use the binomial distribution and get

$$\sum_{i=2}^{5}\binom{20}{i}\left(\frac{1}{5}\right)^i\left(\frac{4}{5}\right)^{20-i}.$$

Solution to Exercise 6.4(A):

$$\binom{7}{2}\left(\frac{5}{9}\right)^2\left(\frac{4}{9}\right)^5$$

Solution to Exercise 6.4(B):

$$\begin{aligned}&\mathbb{P}(X \geqslant 3) = 1 - \mathbb{P}(X \leqslant 2)\\ &= 1 - \binom{7}{0}\left(\frac{4}{9}\right)^0\left(\frac{5}{9}\right)^7 - \binom{7}{1}\left(\frac{4}{9}\right)^1\left(\frac{5}{9}\right)^6 - \binom{7}{2}\left(\frac{4}{9}\right)^2\left(\frac{5}{9}\right)^5\end{aligned}$$

Solution to Exercise 6.5(A): $e^{-0.2}$

Solution to Exercise 6.5(B): $1 - e^{-0.2} - 0.2e^{-0.2} = 1 - 1.2e^{-0.2}$.

Solution to Exercise 6.5(C): Since each word has a small probability of being a typo, the number of typos should be approximately Poisson distributed.

Solution to Exercise 6.6(A): $1 - e^{-3.5} - 3.5e^{-3.5} = 1 - 4.5e^{-3.5}$

Solution to Exercise 6.6(B): $4.5e^{-3.5}$

Solution to Exercise 6.6(C): Since each accident has a small probability it seems reasonable to suppose that the number of car accidents is approximately Poisson distributed.

Solution to Exercise 6.7(A): $e^{-2.2}$

Solution to Exercise 6.7(B): $1 - e^{-2.2} - 2.2e^{-2.2} = 1 - 3.2e^{-2.2}$.

Solution to Exercise 6.8(A): We have

$$\mathbb{P}(X \geqslant 2) = 1 - \mathbb{P}(X \leqslant 1) = 1 - e^{-0.5}\frac{(0.5)^0}{0!} - e^{-0.5}\frac{(0.5)^1}{1!}.$$

Solution to Exercise 6.8(B): In two weeks the average number of accidents will be $\lambda = 0.5 + 0.5 = 1$. Then $\mathbb{P}(X = 5) = e^{-1}\frac{1^5}{5!}$.

Solution to Exercise 6.8(C): In a 4 week period the average number of accidents will be $\lambda = 4 \cdot (0.5) = 2$. Then $\mathbb{P}(X = 2) = e^{-2}\frac{2^2}{2!}$.

Solution to Exercise 6.9(A): This calls for a Poisson random variable R. The average number of raisins per cookie is 0.8, so we take this as our λ . We are asking for $\mathbb{P}(R = 3)$, which is $e^{-0.8}\frac{(0.8)^3}{3!} \approx 0.0383$.

Solution to Exercise 6.9(B): This calls for a Poisson random variable C. The average number of chocolate chips per cookie is 1.2, so we take this as our λ. We are asking for $\mathbb{P}(C \geqslant 1)$, which is $1 - \mathbb{P}(C = 0) = 1 - e^{-1.2}\frac{(1.2)^0}{0!} \approx 0.6988$.

Solution to Exercise 6.9(C): This calls for a Poisson random variable B. The average number of bits per cookie is $0.8 + 1.2 = 2$, so we take this as our λ. We are asking for $\mathbb{P}(B \leqslant 2)$, which is $\mathbb{P}(B = 0) + \mathbb{P}(B = 1) + \mathbb{P}(B = 2) = e^{-2}\frac{2^0}{0!} + e^{-2}\frac{2^1}{1!} + e^{-2}\frac{2^2}{2!} \approx .6767$.

Solution to Exercise 6.10(A): $\left(1 - \frac{18}{38}\right)^6$

Solution to Exercise 6.10(B): $\left(1 - \frac{18}{38}\right)^5 \frac{18}{38}$

Solution to Exercise 6.11 Let X denote the number of albinos at the game. We have that $X \sim \text{Binom}(5000, p)$ with $p = 1/17000 \approx 0.00029$. The binomial distribution gives us

$\mathbb{P}(X = 0) = \left(\frac{16999}{17000}\right)^{5000} \approx 0.745$ $\qquad$ $\mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0) = 1 - \left(\frac{16999}{17000}\right)^{5000} \approx 0.255$

$\mathbb{P}(X > 1) = \mathbb{P}(X \geqslant 1) - \mathbb{P}(X = 1) =$

$= 1 - \left(\frac{16999}{17000}\right)^{5000} - \binom{5000}{1}\left(\frac{16999}{17000}\right)^{4999}\left(\frac{1}{17000}\right)^1 \approx 0.035633$

Approximating the distribution of X by a Poisson with parameter $\lambda = \frac{5000}{17000} = \frac{5}{17}$ gives

$\mathbb{P}(Y = 0) = \exp\left(-\frac{5}{17}\right) \approx 0.745$ $\qquad$ $\mathbb{P}(Y \geqslant 1) = 1 - \mathbb{P}(Y = 0) = 1 - \exp\left(-\frac{5}{17}\right) \approx 0.255$

$\mathbb{P}(Y > 1) = \mathbb{P}(Y \geqslant 1) - \mathbb{P}(Y = 1) = 1 - \exp\left(-\frac{5}{17}\right) - \exp\left(-\frac{5}{17}\right)\frac{5}{17} \approx 0.035638$

Solution to Exercise 6.12: Let X be the random variable that denotes the number of eggs with double yolk in the set of chosen 5. Then $X \sim \text{Hyp}(20, 3, 5)$ and we have that

$$\mathbb{P}(X \geqslant 2) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) = \frac{\binom{3}{2} \cdot \binom{17}{3}}{\binom{20}{5}} + \frac{\binom{3}{3} \cdot \binom{17}{2}}{\binom{20}{5}}.$$

Solution to Exercise 6.13: We will use Poisson approximation.

(a) The probability that both partners have birthday on January 1st is $p = \frac{1}{365^2}$. If X denotes the number of married couples where this is the case, we can approximate the distribution of X by a Poisson with parameter $\lambda = 30,000 \cdot 365^{-2} \approx 0.2251$. Hence, $\mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-0.2251}$.

(b) In this case, the probability of both partners celebrating birthday in the same month is $1/12$ and therefore we approximate the distribution by a Poisson with parameter $\lambda = 30,000/12 = 2500$. Thus, $\mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-2500}$.

Solution to Exercise 6.14: Let X denote the duration (in minutes) of a call. By assumption, $X \sim \text{Pois}(\lambda)$ for some parameter $\lambda > 0$, so that the expected duration of a call is $\mathbb{E}[X] = \lambda$. In addition, we know that $\mathbb{P}(X = 2) = 3\mathbb{P}(X = 4)$, which means

$$e^{-\lambda}\frac{\lambda^2}{2!} = 3e^{-\lambda}\frac{\lambda^4}{4!}.$$

From here we deduce that $\lambda^2 = 4$ and hence $\mathbb{E}[X] = \lambda = 2$.

Part 2

Continuous random variables

CHAPTER 7

Continuous distributions

7.1. Basic theory

7.1.1. Definition, PDF, CDF. We start with the definition a continuous random variable.

Definition (Continuous random variables)

A random variable X is said to have a *continuous distribution* if there exists a non-negative function $f = f_X$ such that

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f(x)dx$$

for every a and b. The function f is called the *density function* for X or the *PDF* for X.

More precisely, such an X is said to have an *absolutely continuous distribution*. Note that $\int_{-\infty}^{\infty} f(x)dx = \mathbb{P}(-\infty < X < \infty) = 1$. In particular, $\mathbb{P}(X = a) = \int_a^a f(x)dx = 0$ for every a.

Example 7.1. Suppose we are given that $f(x) = c/x^3$ for $x \geqslant 1$ and 0 otherwise. Since $\int_{-\infty}^{\infty} f(x)dx = 1$ and

$$c\int_{-\infty}^{\infty} f(x)dx = c\int_1^{\infty} \frac{1}{x^3}dx = \frac{c}{2},$$

we have $c = 2$.

PMF or PDF?

Probability mass function (PMF) and (probability) density function (PDF) are two names for the same notion in the case of discrete random variables. We say *PDF* or simply a *density function* for a general random variable, and we use *PMF* only for discrete random variables.

Definition (Cumulative distribution function (CDF))

The *distribution function* of X is defined as

$$F(y) = F_X(y) := \mathbb{P}(-\infty < X \leqslant y) = \int_{-\infty}^{y} f(x)dx.$$

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

We can define CDF for any random variable, not just continuous ones, by setting $F(y) := \mathbb{P}(X \leqslant y)$. Recall that we introduced it in Definition 5.3 for discrete random variables. In that case it is not particularly useful, although it does serve to unify discrete and continuous random variables. In the continuous case, the fundamental theorem of calculus tells us, provided f satisfies some conditions, that

$$f(y) = F'(y).$$

By analogy with the discrete case, we define the expectation of a continuous random variable.

7.1.2. Expectation, discrete approximation to continuous random variables.

Definition (Expectation)

For a continuous random variable X with the density function f we define its *expectation* by

$$\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx$$

if this integral is absolutely convergent. In this case we call X *integrable*.

Recall that this integral is absolutely convergent if

$$\int_{-\infty}^{\infty} |x|f(x)dx < \infty.$$

In the example above,

$$\mathbb{E}X = \int_1^{\infty} x\frac{2}{x^3}dx = 2\int_1^{\infty} x^{-2}dx = 2.$$

Later in Example 10.1 we will see that a continuous random variable with Cauchy distribution has infinite expectation.

Proposition 7.1 (Discrete approximation to continuous random variables)

Suppose X is a nonnegative continuous random variable with a finite expectation. Then there is a sequence of discrete random variables $\{X_n\}_{n=1}^{\infty}$ such that

$$\mathbb{E}X_n \xrightarrow[n\to\infty]{} \mathbb{E}X.$$

PROOF. First observe that if a continuous random variable X is nonnegative, then its density $f(x) = 0$ $x < 0$. In particular, $F(y) = 0$ for $y \leqslant 0$, thought the latter is not needed for our proof. Thus for such a random variable

$$\mathbb{E}X = \int_0^{\infty} xf(x)dx.$$

Suppose $n \in \mathbb{N}$, then we define $X_n(\omega)$ to be $k/2^n$ if $k/2^n \leqslant X(\omega) < (k+1)/2^n$, for $k \in \mathbb{N}\cup\{0\}$. This means that we are approximating X from below by the largest multiple of 2^{-n} that is still below the value of X. Each X_n is discrete, and X_n increase to X for each $\omega \in S$.

Consider the sequence $\{\mathbb{E}X_n\}_{n=1}^{\infty}$. This sequence is an increasing sequence of positive numbers, and therefore it has a limit, possibly infinite. We want to show that it is finite and it is equal to $\mathbb{E}X$.

We have

$$\begin{aligned}\mathbb{E}X_n &= \sum_{k=1}^{\infty} \frac{k}{2^n}\mathbb{P}\left(X_n = \frac{k}{2^n}\right)\\ &= \sum_{k=1}^{\infty} \frac{k}{2^n}\mathbb{P}\left(\frac{k}{2^n} \leqslant X < \frac{k+1}{2^n}\right)\\ &= \sum_{k=1}^{\infty} \frac{k}{2^n}\int_{k/2^n}^{(k+1)/2^n} f(x)dx\\ &= \sum_{k=1}^{\infty} \int_{k/2^n}^{(k+1)/2^n} \frac{k}{2^n} f(x)dx.\end{aligned}$$

If $x \in [k/2^n, (k+1)/2^n)$, then x differs from $k/2^n$ by at most $1/2^n$, and therefore

$$\begin{aligned}0 &\leqslant \int_{k/2^n}^{(k+1)/2^n} xf(x)dx - \int_{k/2^n}^{(k+1)/2^n} \frac{k}{2^n} f(x)dx\\ &= \int_{k/2^n}^{(k+1)/2^n} \left(x - \frac{k}{2^n}\right) f(x)dx \leqslant \frac{1}{2^n}\int_{k/2^n}^{(k+1)/2^n} f(x)dx\end{aligned}$$

Note that

$$\sum_{k=1}^{\infty} \int_{k/2^n}^{(k+1)/2^n} xf(x)dx = \int_0^{\infty} xf(x)dx$$

and

$$\sum_{k=1}^{\infty} \frac{1}{2^n}\int_{k/2^n}^{(k+1)/2^n} f(x)dx = \frac{1}{2^n}\sum_{k=1}^{\infty} \int_{k/2^n}^{(k+1)/2^n} f(x)dx = \frac{1}{2^n}\int_0^{\infty} f(x)dx = \frac{1}{2^n}.$$

Therefore

$$
\begin{aligned}
0 \leqslant \mathbb{E}X - \mathbb{E}X_n &= \int_0^\infty x f(x)dx - \sum_{k=1}^\infty \int_{k/2^n}^{(k+1)/2^n} \frac{k}{2^n} f(x)dx \\
&= \sum_{k=1}^\infty \int_{k/2^n}^{(k+1)/2^n} x f(x)dx - \sum_{k=1}^\infty \int_{k/2^n}^{(k+1)/2^n} \frac{k}{2^n} f(x)dx \\
&= \sum_{k=1}^\infty \left(\int_{k/2^n}^{(k+1)/2^n} x f(x)dx - \int_{k/2^n}^{(k+1)/2^n} \frac{k}{2^n} f(x)dx \right) \\
&\leqslant \sum_{k=1}^\infty \frac{1}{2^n} \int_{k/2^n}^{(k+1)/2^n} f(x)dx = \frac{1}{2^n} \xrightarrow[n\to 0]{} 0.
\end{aligned}
$$

□

We will not prove the following, but it is an interesting exercise: if X_m is *any* sequence of discrete random variables that increase up to X, then $\lim_{m\to\infty} \mathbb{E}X_m$ will have the same value $\mathbb{E}X$.

This fact is useful to show linearity, if X and Y are positive random variables with finite expectations, then we can take X_m discrete increasing up to X and Y_m discrete increasing up to Y. Then $X_m + Y_m$ is discrete and increases up to $X + Y$, so we have

$$
\begin{aligned}
\mathbb{E}(X+Y) &= \lim_{m\to\infty} \mathbb{E}(X_m + Y_m) \\
&= \lim_{m\to\infty} \mathbb{E}X_m + \lim_{m\to\infty} \mathbb{E}Y_m = \mathbb{E}X + \mathbb{E}Y.
\end{aligned}
$$

Note that we can not easily use the approximations to X, Y and $X + Y$ we used in the previous proof to use in this argument, since $X_m + Y_m$ might not be an approximation of the same kind.

If X is not necessarily positive, we can show a similar result; we will not do the details.

Similarly to the discrete case, we have

Proposition 7.2

Suppose X is a continuous random variable with density f_X and g is a real-valued function, then

$$
\mathbb{E}g(X) = \int_{-\infty}^{\infty} g(x) f(x) dx
$$

as long as the expectation of the random variable $g(X)$ makes sense.

As in the discrete case, this allows us to define moments, and in particular the *variance*

$$
\operatorname{Var} X := \mathbb{E}[X - \mathbb{E}X]^2.
$$

As an example of these calculations, let us look at the uniform distribution.

Uniform distribution

We say that a random variable X has a *uniform distribution* on $[a, b]$ if $f_X(x) = \frac{1}{b-a}$ if $a \leqslant x \leqslant b$ and 0 otherwise.

To calculate the expectation of X

$$\begin{aligned}\mathbb{E}X &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \int_a^b x \, dx \\ &= \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}.\end{aligned}$$

This is what one would expect. To calculate the variance, we first calculate

$$\mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{a^2 + ab + b^2}{3}.$$

We then do some algebra to obtain

$$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{(b-a)^2}{12}.$$

7.2. Further examples and applications

Example 7.2. Suppose X has the following p.d.f.

$$f(x)=\begin{cases}\frac{2}{x^3} & x\geqslant 1\\ 0 & x<1.\end{cases}$$

Find the CDF of X, that is, find $F_X(x)$. Use the CDF to find $\mathbb{P}\left(3\leqslant X\leqslant 4\right)$.

We have $F_X(x)=0$ if $x\leqslant 1$ and will need to compute

$$F_X(x)=\mathbb{P}\left(X\leqslant x\right)=\int_1^x\frac{2}{y^3}dy=1-\frac{1}{x^2}$$

when $x\geqslant 1$. We can use this formula to find the following probability

$$\begin{aligned}\mathbb{P}\left(3\leqslant X\leqslant 4\right)&=\mathbb{P}\left(X\leqslant 4\right)-\mathbb{P}\left(X<3\right)\\ &=F_X(4)-F_X(3)=\left(1-\frac{1}{4^2}\right)-\left(1-\frac{1}{3^2}\right)=\frac{7}{144}.\end{aligned}$$

Example 7.3. Suppose X has density

$$f_X(x)=\begin{cases}2x & 0\leqslant x\leqslant 1\\ 0 & \text{otherwise}\end{cases}.$$

We can find $\mathbb{E}X$ by

$$\mathbb{E}X=\int xf_X(x)dx=\int_0^1 x\cdot 2x\,dx=\frac{2}{3}.$$

Example 7.4. The density of X is given by

$$f_X(x)=\begin{cases}\frac{1}{2} & \text{if } 0\leqslant x\leqslant 2\\ 0 & \text{otherwise}\end{cases}.$$

We can use Proposition 7.2 with $g(x)=e^x$ to find $\mathbb{E}e^X$.

$$\mathbb{E}e^X=\int_0^2 e^x\cdot\frac{1}{2}dx=\frac{1}{2}\left(e^2-1\right).$$

Example 7.5. Suppose X has density

$$f(x)=\begin{cases}2x & 0\leqslant x\leqslant 1\\ 0 & \text{otherwise}\end{cases}.$$

Find $\mathrm{Var}(X)$. In Example 7.3 we found $\mathbb{E}\left[X\right]=\frac{2}{3}$. Now

$$\mathbb{E}\left[X^2\right]=\int_0^1 x^2\cdot 2xdx=2\int_0^1 x^3dx=\frac{1}{2}.$$

Thus

$$\mathrm{Var}(X)=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}.$$

Example 7.6. Suppose X has density

$$f(x) = \begin{cases} ax + b & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}.$$

and that $\mathbb{E}\left[X^2\right] = \frac{1}{6}$. Find the values of a and b.

We need to use the fact that $\int_{-\infty}^{\infty} f(x)dx = 1$ and $\mathbb{E}\left[X^2\right] = \frac{1}{6}$. The first one gives us

$$1 = \int_0^1 (ax + b)\, dx = \frac{a}{2} + b,$$

and the second one gives us

$$\frac{1}{6} = \int_0^1 x^2 (ax + b)\, dx = \frac{a}{4} + \frac{b}{3}.$$

Solving these equations gives us

$$a = -2, \text{ and } b = 2.$$

7.3. Exercises

Exercise 7.1. Let X be a random variable with probability density function

$$f(x) = \begin{cases} cx\,(5-x) & 0 \leqslant x \leqslant 5, \\ 0 & \text{otherwise.} \end{cases}$$

(A) What is the value of c?
(B) What is the cumulative distribution function of X? That is, find $F_X(x) = \mathbb{P}\,(X \leqslant x)$.
(C) Use your answer in part (b) to find $\mathbb{P}\,(2 \leqslant X \leq 3)$.
(D) What is $\mathbb{E}\,[X]$?
(E) What is $\mathrm{Var}(X)$?

Exercise 7.2. UConn students have designed the new U-phone. They have determined that the lifetime of a U-Phone is given by the random variable X (measured in hours), with probability density function

$$f(x) = \begin{cases} \frac{10}{x^2} & x \geqslant 10, \\ 0 & x \leq 10. \end{cases}$$

(A) Find the probability that the u-phone will last more than 20 hours.
(B) What is the cumulative distribution function of X? That is, find $F_X(x) = \mathbb{P}\,(X \leqslant x)$.
(C) Use part (b) to help you find $\mathbb{P}\,(X \geqslant 35)$?

Exercise 7.3. Suppose the random variable X has a density function

$$f(x) = \begin{cases} \frac{2}{x^2} & x > 2, \\ 0 & x \leqslant 2. \end{cases}$$

Compute $\mathbb{E}\,[X]$.

Exercise 7.4. An insurance company insures a large number of homes. The insured value, X, of a randomly selected home is assumed to follow a distribution with density function

$$f(x) = \begin{cases} \frac{3}{x^4} & x > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Given that a randomly selected home is insured for at least 1.5, calculate the probability that it is insured for less than 2.

Exercise 7.5. The density function of X is given by

$$f(x) = \begin{cases} a + bx^2 & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbb{E}\,[X] = \frac{7}{10}$, find the values of a and b.

Exercise 7.6. Let X be a random variable with density function

$$f(x) = \begin{cases} \frac{1}{a-1} & 1 < x < a, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that $\mathbb{E}[X] = 6 \operatorname{Var}(X)$. Find the value of a.

Exercise 7.7. Suppose you order a pizza from your favorite pizzeria at 7:00 pm, knowing that the time it takes for your pizza to be ready is uniformly distributed between 7:00 pm and 7:30 pm.

(A) What is the probability that you will have to wait longer than 10 minutes for your pizza?
(B) If at 7:15pm, the pizza has not yet arrived, what is the probability that you will have to wait at least an additional 10 minutes?

Exercise 7.8. The grade of deterioration X of a machine part has a continuous distribution on the interval $(0, 10)$ with probability density function $f_X(x)$, where $f_X(x)$ is proportional to $\frac{x}{5}$ on the interval. The reparation costs of this part are modeled by a random variable Y that is given by $Y = 3X^2$. Compute the expected cost of reparation of the machine part.

Exercise 7.9. A bus arrives at some (random) time uniformly distributed between $10:00$ and $10:20$, and you arrive at a bus stop at $10:05$.

(A) What is the probability that you have to wait at least 5 minutes until the bus comes?
(B) What is the probability that you have to wait at least 5 minutes, given that when you arrive today to the station the bus was not there yet (you are lucky today)?

Exercise* 7.1. For a continuous random variable X with finite first and second moments prove that

$$\begin{aligned} \mathbb{E}(aX + b) &= a\mathbb{E}X + b, \\ \operatorname{Var}(aX + b) &= a^2 \operatorname{Var} X. \end{aligned}$$

for any $a, b \in \mathbb{R}$.

Exercise* 7.2. Let X be a continuous random variable with probability density function

$$f_X(x) = \frac{1}{4} x e^{-\frac{x}{2}} \mathbb{1}_{[0,\infty)}(x),$$

where the indicator function is defined as

$$\mathbb{1}_{[0,\infty)}(x) = \begin{cases} 1, & 0 \leqslant x < \infty; \\ \\ 0, & \text{otherwise.} \end{cases}$$

Check that f_X is a valid probability density function, and find $\mathbb{E}(X)$ if it exists.

Exercise* 7.3. Let X be a continuous random variable with probability density function

$$f_X(x) = \frac{4\ln x}{x^3}\mathbb{1}_{[1,\infty)}(x),$$

where the indicator function is defined as

$$\mathbb{1}_{[1,\infty)}(x) = \begin{cases} 1, & 1 \leqslant x < \infty; \\ 0, & \text{otherwise}. \end{cases}$$

Check that f_X is a valid probability density function, and find $\mathbb{E}(X)$ if it exists.

7.4. Selected solutions

Solution to Exercise 7.1(A): We must have that $\int_{-\infty}^{\infty} f(x)dx = 1$, thus

$$1 = \int_0^5 cx(5-x)dx = \left[c\left(\frac{5x^2}{2} - \frac{x^3}{3}\right)\right]_0^5$$

and so we must have that $c = 6/125$.

Solution to Exercise 7.1(B): We have that

$$\begin{aligned} F_X(x) = \mathbb{P}(X \leqslant x) &= \int_{-\infty}^{x} f(y)dy \\ &= \int_0^x \frac{6}{125} y\,(5-y)\,dx = \frac{6}{125}\left[\left(\frac{5y^2}{2} - \frac{y^3}{3}\right)\right]_0^x \\ &= \frac{6}{125}\left(\frac{5x^2}{2} - \frac{x^3}{3}\right). \end{aligned}$$

Solution to Exercise 7.1(C): We have

$$\begin{aligned} \mathbb{P}(2 \leqslant X \leqslant 3) &= \mathbb{P}(X \leqslant 3) - \mathbb{P}(X < 2) \\ &= \frac{6}{125}\left(\frac{5 \cdot 3^2}{2} - \frac{3^3}{3}\right) - \frac{6}{125}\left(\frac{5 \cdot 2^2}{2} - \frac{2^3}{3}\right) \\ &= 0.296. \end{aligned}$$

Solution to Exercise 7.1(D): we have

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^5 x \cdot \frac{6}{125} x(5-x)dx \\ &= 2.5. \end{aligned}$$

Solution to Exercise 7.1(E): We need to first compute

$$\begin{aligned} \mathbb{E}\left[X^2\right] &= \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^5 x^2 \cdot \frac{6}{125} x(5-x)dx \\ &= 7.5. \end{aligned}$$

Then

$$\operatorname{Var}(X) = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2 = 7.5 - (2.5)^2 = 1.25.$$

Solution to Exercise 7.2(A): We have

$$\int_{20}^{\infty} \frac{10}{x^2} dx = \frac{1}{2}.$$

Solution to Exercise 7.2(B): We have

$$F(x) = \mathbb{P}(X \leqslant x) = \int_{10}^{x} \frac{10}{y^2} dy = 1 - \frac{10}{x}$$

for $x > 10$, and $F(x) = 0$ for $x < 10$.

Solution to Exercise 7.2(C): We have

$$\begin{aligned}\mathbb{P}(X \geqslant 35) &= 1 - \mathbb{P}(X < 35) = 1 - F_X(35) \\ &= 1 - \left(1 - \frac{10}{35}\right) = \frac{10}{35}.\end{aligned}$$

Solution to Exercise 7.3: $+\infty$

Solution to Exercise 7.4: $\frac{37}{64}$.

Solution to Exercise 7.5: we need to use the fact that $\int_{-\infty}^{\infty} f(x)dx = 1$ and $\mathbb{E}[X] = \frac{7}{10}$. The first one gives us

$$1 = \int_0^1 \left(a + bx^2\right) dx = a + \frac{b}{3}$$

and the second one gives

$$\frac{7}{10} = \int_0^1 x\left(a + bx^2\right) dx = \frac{a}{2} + \frac{b}{4}.$$

Solving these equations gives

$$a = \frac{1}{5}, \text{ and } b = \frac{12}{5}.$$

Solution to Exercise 7.6: Note that

$$\mathbb{E}X = \int_1^a \frac{x}{a-1} dx = \frac{1}{2}a + \frac{1}{2}.$$

Also

$$\operatorname{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

then we need

$$\mathbb{E}X^2 = \int_1^a \frac{x^2}{a-1} dx = \frac{1}{3}a^2 + \frac{1}{3}a + \frac{1}{3}.$$

Then

$$\begin{aligned}\operatorname{Var}(X) &= \left(\frac{1}{3}a^2 + \frac{1}{3}a + \frac{1}{3}\right) - \left(\frac{1}{2}a + \frac{1}{2}\right)^2 \\ &= \frac{1}{12}a^2 - \frac{1}{6}a + \frac{1}{12}.\end{aligned}$$

Then, using $\mathbb{E}[X] = 6\operatorname{Var}(X)$, we simplify and get $\frac{1}{2}a^2 - \frac{3}{2}a = 0$, which gives us $a = 3$.

Another way to solve this problem is to note that, for the uniform distribution on $[a, b]$, the mean is $\frac{a+b}{2}$ and the variance is $\frac{(a-b)^2}{12}$. This gives us an equation $6\frac{(a-1)^2}{12} = \frac{a+1}{2}$. Hence $(a-1)^2 = a + 1$, which implies $a = 3$.

Solution to Exercise 7.7(A): Note that X is uniformly distributed over $(0, 30)$. Then

$$\mathbb{P}(X > 10) = \frac{2}{3}.$$

Solution to Exercise 7.7(B): Note that X is uniformly distributed over $(0, 30)$. Then

$$\mathbb{P}(X > 25 \mid X > 15) = \frac{\mathbb{P}(X > 25)}{\mathbb{P}(X > 15)} = \frac{5/30}{15/30} = 1/3.$$

Solution to Exercise 7.8: First of all we need to find the PDF of X. So far we know that

$$f(x)=\begin{cases}\frac{cx}{5} & 0\leqslant x\leqslant 10,\\ 0 & \text{otherwise.}\end{cases}$$

Since

$$\int_0^{10} c\frac{x}{5}dx=10c,$$

we have $c=\frac{1}{10}$. Now, applying Proposition 7.2 we get

$$\mathbb{E}Y=\int_0^{10}\frac{3}{50}x^3dx=150.$$

Solution to Exercise 7.9(A): The probability that you have to wait at least 5 minutes until the bus comes is $\frac{1}{2}$. Note that with probability $\frac{1}{4}$ you have to wait less than 5 minutes, and with probability $\frac{1}{4}$ you already missed the bus.

Solution to Exercise 7.9(B): The conditional probability is $\frac{2}{3}$.

CHAPTER 8

Normal distribution

8.1. Standard and general normal distributions

Definition (Standard normal distribution)

A continuous random variable is a *standard normal* (written $\mathcal{N}(0,1)$) if it has density

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

A synonym for normal is Gaussian. The first thing to do is to show that this is a (probability) density.

Theorem

$f_Z(x)$ is a valid PDF, that is, it is a nonnegative function such that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1.$$

Suppose $Z \sim \mathcal{N}(0,1)$. Then

$$\mathbb{E}Z = 0,$$
$$\operatorname{Var} Z = 1.$$

PROOF. Let $I = \int_0^\infty e^{-x^2/2} dx$. Then

$$I^2 = \int_0^\infty \int_0^\infty e^{-x^2/2} e^{-y^2/2} dx\, dy.$$

Changing to polar coordinates,

$$I^2 = \int_0^{\pi/2} \int_0^\infty r e^{-r^2/2} dr = \pi/2.$$

So $I = \sqrt{\pi/2}$, hence $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$ as it should.
Note

$$\int_{-\infty}^{\infty} x e^{-x^2/2} dx = 0$$

by symmetry, so $\mathbb{E}Z = 0$. For the variance of Z, we use integration by parts as follows.

$$\mathbb{E}Z^2 = \frac{1}{\sqrt{2\pi}} \int x^2 e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int x \cdot x e^{-x^2/2} dx.$$

The integral is equal to

$$-xe^{-x^2/2}\Big]_{-\infty}^{\infty} + \int e^{-x^2/2}dx = \sqrt{2\pi}.$$

Therefore $\operatorname{Var} Z = \mathbb{E}Z^2 = 1$. □

Note that these integrals are improper, so our arguments are somewhat informal, but they are easy to make rigorous.

Definition (General normal distribution)

We say X is a $\mathcal{N}(\mu, \sigma^2)$ if $X = \sigma Z + \mu$, where $Z \sim \mathcal{N}(0, 1)$.

Proposition 8.1 (General normal density)

A random variable with a general normal distribution $\mathcal{N}(\mu, \sigma^2)$ is a continuous random variable with density

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

PROOF. Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$. We see that if $\sigma > 0$, then

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leqslant x) = \mathbb{P}(\mu + \sigma Z \leqslant x) \\ &= \mathbb{P}\left(Z \leqslant \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right), \end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$. A similar calculation holds if $\sigma < 0$.

Then by the chain rule X has density

$$f_X(x) = F_X'(x) = F_Z'\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

□

Proposition 8.2 (Properties of general normal variables)

For any $X \sim \mathcal{N}(\mu, \sigma)$

$$\begin{aligned} \mathbb{E}X &= \mu, \\ \operatorname{Var} X &= \sigma^2. \end{aligned}$$

For any $a, b \in \mathbb{R}$ the random variable $aX + b$ is a normal variable.

PROOF. By definition of X we have $X = \sigma Z + \mu$, where $Z \sim \mathcal{N}(0, 1)$, and so by linearity of the expectation we have $\mathbb{E}X = \mu + \mathbb{E}Z = \mu$. Recall that by Exercise* 7.1 we have $\operatorname{Var} X = \sigma^2 \operatorname{Var} Z$, so

$$\operatorname{Var} X = \sigma^2.$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$, then $Y = a(\mu + \sigma Z) + b = (a\mu + b) + (a\sigma)Z$, or Y is $\mathcal{N}(a\mu + b, a^2\sigma^2)$. In particular, if $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Z = (X - \mu)/\sigma$, then Z is $\mathcal{N}(0, 1)$. $\square$

The distribution function of a standard random variable $\mathcal{N}(0, 1)$ is often denoted as $\Phi(x)$, so that

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

Tables of $\Phi(x)$ are often given only for $x > 0$. One can use the symmetry of the density function to see that

$$\Phi(-x) = 1 - \Phi(x),$$

this follows from

$$\begin{aligned}
\Phi(-x) = \mathbb{P}(Z \leqslant -x) &= \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
&= \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \mathbb{P}(Z \geqslant x) \\
&= 1 - \mathbb{P}(Z < x) = 1 - \Phi(x).
\end{aligned}$$

Table 1. Cumulative distribution function for the standard normal variable

x	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999
3.1	0.99903	0.99906	0.9991	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.9994	0.99942	0.99944	0.99946	0.99948	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.9996	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.9997	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.9998	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.9999	0.9999	0.9999	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

Example 8.1. Find $\mathbb{P}(1 \leqslant X \leqslant 4)$ if X is $\mathcal{N}(2, 25)$.

Write $X = 2 + 5Z$, so

$$\begin{aligned}\mathbb{P}(1 \leqslant X \leqslant 4) &= \mathbb{P}(1 \leqslant 2 + 5Z \leqslant 4) \\ &= \mathbb{P}(-1 \leqslant 5Z \leqslant 2) = \mathbb{P}(-0.2 \leqslant Z \leqslant 0.4) \\ &= \mathbb{P}(Z \leqslant 0.4) - \mathbb{P}(Z \leqslant -0.2) \\ &= \Phi(0.4) - \Phi(-0.2) = 0.6554 - (1 - \Phi(0.2)) \\ &= 0.6554 - (1 - 0.5793)\,.\end{aligned}$$

Example 8.2. Find c such that $\mathbb{P}(|Z| \geqslant c) = 0.05$.

By symmetry we want to find c such that $\mathbb{P}(Z \geqslant c) = 0.025$ or $\Phi(c) = \mathbb{P}(Z \leqslant c) = 0.975$. From the table (Table 1) we see $c = 1.96 \approx 2$. This is the origin of the idea that the 95% significance level is ± 2 standard deviations from the mean.

Proposition 8.3

For a standard normal random variable Z we have the following bound. For $x > 0$

$$\mathbb{P}(Z \geqslant x) = 1 - \Phi(x) \leqslant \frac{1}{\sqrt{2\pi}\, x} e^{-x^2/2}.$$

PROOF. If $y \geqslant x > 0$, then $y/x \geqslant 1$, and therefore

$$\begin{aligned}\mathbb{P}(Z \geqslant x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\, dy \\ &\leqslant \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{y}{x} e^{-y^2/2}\, dy = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.\end{aligned}$$

□

This is a good estimate when x is large. In particular, for x large,

$$\mathbb{P}(Z \geqslant x) = 1 - \Phi(x) \leqslant e^{-x^2/2}.$$

8.2. Further examples and applications

Example 8.3. Suppose X is normal with mean 6. If $\mathbb{P}(X > 16) = 0.0228$, then what is the standard deviation of X?

Recall that we saw in Proposition 8.2 that $\frac{X-\mu}{\sigma} = Z \sim \mathcal{N}(0,1)$ and then

$$\begin{aligned}
\mathbb{P}(X > 16) = 0.0228 &\iff \mathbb{P}\left(\frac{X-6}{\sigma} > \frac{16-6}{\sigma}\right) = 0.0228 \\
&\iff \mathbb{P}\left(Z > \frac{10}{\sigma}\right) = 0.0228 \\
&\iff 1 - \mathbb{P}\left(Z \leqslant \frac{10}{\sigma}\right) = 0.0228 \\
&\iff 1 - \Phi\left(\frac{10}{\sigma}\right) = 0.0228 \\
&\iff \Phi\left(\frac{10}{\sigma}\right) = 0.9772.
\end{aligned}$$

Using the standard normal table we see that $\Phi(2) = 0.9772$, thus we have that

$$2 = \frac{10}{\sigma}$$

and hence $\sigma = 5$.

8.3. Exercises

Exercise 8.1. Suppose X is a normally distributed random variable with $\mu = 10$ and $\sigma^2 = 36$. Find (a) $\mathbb{P}(X > 5)$, (b) $\mathbb{P}(4 < X < 16)$, (c) $\mathbb{P}(X < 8)$.

Exercise 8.2. The height of maple trees at age 10 are estimated to be normally distributed with mean 200 cm and variance 64 cm. What is the probability a maple tree at age 10 grows more than 210cm?

Exercise 8.3. The peak temperature T, in degrees Fahrenheit, on a July day in Antarctica is a Normal random variable with a variance of 225. With probability .5, the temperature T exceeds 10 degrees.

(a) What is $\mathbb{P}(T > 32)$, the probability the temperature is above freezing?
(b) What is $\mathbb{P}(T < 0)$?

Exercise 8.4. The salaries of UConn professors is approximately normally distributed. Suppose you know that 33 percent of professors earn less than \$80,000. Also 33 percent earn more than \$120,000.

(a) What is the probability that a UConn professor makes more than \$100,000?
(b) What is the probability that a UConn professor makes between \$70,000 and \$80,000?

Exercise 8.5. Suppose X is a normal random variable with mean 5. If $\mathbb{P}(X > 0) = 0.8888$, approximately what is $\operatorname{Var}(X)$?

Exercise 8.6. The shoe size of a UConn basketball player is normally distributed with mean 12 inches and variance 4 inches. Ten percent of all UConn basketball players have a shoe size greater than c inches. Find the value of c.

Exercise 8.7. The length of the forearm of a UConn football player is normally distributed with mean 12 inches. If ten percent of the football team players have a forearm whose length is greater than 12.5 inches, find out the approximate standard deviation of the forearm length of a UConn football player.

Exercise 8.8. Companies C and A earn each an annual profit that is normally distributed with the same positive mean μ. The standard deviation of C's annual profit is one third of its mean. In a certain year, the probability that A makes a loss (i.e. a negative profit) is 0.8 times the probability that C does. Assuming that A's annual profit has a standard deviation of 10, compute (approximately) the standard deviation of C's annual profit.

Exercise 8.9. Let $Z \sim \mathcal{N}(0, 1)$, that is, a standard normal random variable. Find probability density for $X = Z^2$. *Hint:* first find the (cumulative) distribution function $F_X(x) = \mathbb{P}(X \leqslant x)$ in terms of $\Phi(x) = F_Z(x)$. Then use the fact that the probability density function can be found by $f_X(x) = F_X'(x)$, and use the known density function for Z.

8.4. Selected solutions

Solution to Exercise 8.1(A):

$$\begin{aligned}\mathbb{P}(X>5) &= \mathbb{P}\left(Z>\frac{5-10}{6}\right)=\mathbb{P}\left(Z>-\frac{5}{6}\right)\\ &= \mathbb{P}\left(Z<\frac{5}{6}\right)=\Phi\left(\frac{5}{6}\right)\approx\Phi(0.83)\approx 0.797\end{aligned}$$

Solution to Exercise 8.1(B): $2\Phi(1)-1=0.6827$

Solution to Exercise 8.1(C): $1-\Phi(0.3333)=0.3695$

Solution to Exercise 8.2: We have $\mu=200$ and $\sigma=\sqrt{64}=8$. Then

$$\begin{aligned}\mathbb{P}(X>210) &= \mathbb{P}\left(Z>\frac{210-200}{8}\right)=\mathbb{P}(Z>1.25)\\ &= 1-\Phi(1.25)=0.1056.\end{aligned}$$

Solution to Exercise 8.3(A): We have $\sigma=\sqrt{225}=15$. Since $\mathbb{P}(X>10)=0.5$ then we must have that $\mu=10$ since the PDF of the normal distribution is symmetric. Then

$$\begin{aligned}\mathbb{P}(T>32) &= \mathbb{P}\left(Z>\frac{32-10}{15}\right)\\ &= 1-\Phi(1.47)=0.0708.\end{aligned}$$

Solution to Exercise 8.3(B): We have $\mathbb{P}(T<0)=\Phi(-0.67)=1-\Phi(0.67)=0.2514$.

Solution to Exercise 8.4(A): First we need to figure out what μ and σ are. Note that

$$\begin{aligned}\mathbb{P}(X\leqslant 80,000)=0.33 &\iff \mathbb{P}\left(Z<\frac{80,000-\mu}{\sigma}\right)=0.33\\ &\iff \Phi\left(\frac{80,000-\mu}{\sigma}\right)=0.33\end{aligned}$$

and since $\Phi(0.44)=0.67$ then $\Phi(-0.44)=0.33$. Then we must have

$$\frac{80,000-\mu}{\sigma}=-0.44.$$

Similarly, since

$$\begin{aligned}\mathbb{P}(X>120,000)=0.33 &\iff 1-\mathbb{P}(X\leqslant 120,000)=0.33\\ &\iff 1-\Phi\left(\frac{120,000-\mu}{\sigma}\right)=0.33\\ &\iff \Phi\left(\frac{120,000-\mu}{\sigma}\right)=0.67\end{aligned}$$

Now again since $\Phi(0.44)=0.67$ then

$$\frac{120,000-\mu}{\sigma}=0.44.$$

Solving the equations

$$\frac{80,000-\mu}{\sigma}=-0.44 \text{ and } \frac{120,000-\mu}{\sigma}=0.44,$$

as a system for μ and σ we have that

$$\mu=100,000 \text{ and } \sigma\approx 45,454.5.$$

Then

$$\mathbb{P}(X>100,000)=0.5.$$

Solution to Exercise 8.4(B): We have

$$\mathbb{P}(70,000<X<80,000)\approx 0.0753.$$

Solution to Exercise 8.5: Since $\mathbb{P}(X>0)=.8888$, then

$$\begin{aligned}
\mathbb{P}(X>0)=0.8888 &\iff \mathbb{P}\left(Z>\frac{0-5}{\sigma}\right)=0.8888\\
&\iff 1-\mathbb{P}\left(Z\leqslant -\frac{5}{\sigma}\right)=0.8888\\
&\iff 1-\Phi\left(-\frac{5}{\sigma}\right)=0.8888\\
&\iff 1-\left(1-\Phi\left(\frac{5}{\sigma}\right)\right)=0.8888\\
&\iff \Phi\left(\frac{5}{\sigma}\right)=0.8888.
\end{aligned}$$

Using the table we see that $\Phi(1.22)=0.8888$, thus we must have that

$$\frac{5}{\sigma}=1.22$$

and solving this gets us $\sigma=4.098$, hence $\sigma^2\approx 16.8$.

Solution to Exercise 8.6: Note that

$$\begin{aligned}
\mathbb{P}(X>c)=0.10 &\iff \mathbb{P}\left(Z>\frac{c-12}{2}\right)=0.10\\
&\iff 1-\mathbb{P}\left(Z\leqslant\frac{c-12}{2}\right)=0.10\\
&\iff \mathbb{P}\left(Z\leqslant\frac{c-12}{2}\right)=0.9\\
&\iff \Phi\left(\frac{c-12}{2}\right)=0.9
\end{aligned}$$

Using the table we see that $\Phi(1.28)=0.90$, thus we must have that

$$\frac{c-12}{2}=1.28$$

and solving this gets us $c=14.56$.

Solution to Exercise 8.7: Let X denote the forearm length of a UConn football player and let σ denote its standard deviation. From the problem we know that

$$\mathbb{P}(X > 12.5) = \mathbb{P}\left(\frac{X-12}{\sigma} > \frac{0.5}{\sigma}\right) = 1 - \Phi\left(\frac{0.5}{\sigma}\right) = 0.1.$$

From the table we get $\frac{0.5}{\sigma} \approx 1.29$ hence $\sigma \approx 0.39$.

Solution to Exercise 8.8: Let A and C denote the respective annual profits, and μ their mean. Form the problem we know $\mathbb{P}(A < 0) = 0.8\mathbb{P}(C < 0)$ and $\sigma_A = \mu/3$. Since they are normal distributed, $\Phi\left(\frac{-\mu}{10}\right) = 0.8\Phi(-3)$ which implies

$$\Phi\left(\frac{\mu}{10}\right) = 0.2 + 0.8\Phi(3) \approx 0.998.$$

From the table we thus get $\mu/10 \approx 2.88$ and hence the standard deviation of C is $\mu/3 \approx 9.6$.

Solution to Exercise 8.9: see Example 10.2.

CHAPTER 9

Normal approximation to the binomial

A special case of the *central limit theorem* is the following statement.

Theorem 9.1 (Normal approximation to the binomial distribution)

If S_n is a binomial variable with parameters n and p, Binom (n, p), then

$$\mathbb{P}\left(a \leqslant \frac{S_n - np}{\sqrt{np(1-p)}} \leqslant b\right) \xrightarrow[n\to\infty]{} \mathbb{P}(a \leqslant Z \leqslant b),$$

as $n \to \infty$, where $Z \sim \mathcal{N}(0, 1)$.

This approximation is good if $np(1-p) \geqslant 10$ and gets better the larger this quantity gets. This means that if either p or $1-p$ is small, then this is valid for large n. Recall that by Proposition 6.1 np is the same as $\mathbb{E}S_n$ and $np(1-p)$ is the same as $\operatorname{Var} S_n$. So the ratio is equal to $(S_n - \mathbb{E}S_n)/\sqrt{\operatorname{Var} S_n}$, and this ratio has mean 0 and variance 1, the same as a standard $\mathcal{N}(0, 1)$.

Note that here p stays fixed as $n \to \infty$, unlike in the case of the Poisson approximation, as we described in Proposition 6.3.

Sketch of the proof. This is usually not covered in this course, so we only explain one (of many) ways to show why this holds. We would like to compare the distribution of S_n with the distribution of the normal variable $X \sim \mathcal{N}\left(np, \sqrt{np\,(1-p)}\right)$. The random variable X has the density

$$\frac{1}{\sqrt{2\pi np\,(1-p)}} e^{-\frac{(x-np)^2}{2np(1-p)}}.$$

The idea behind this proof is that we are interested in approximating the binomial distribution by the normal distribution in the region where the binomial distribution differs significantly from zero, that is, in the region around the mean np. We consider $\mathbb{P}(S_n = k)$, and we assume that k does not deviate too much from np. We measure deviations by some small number of standard deviations, which is $\sqrt{np\,(1-p)}$. Therefore we see that $k - np$ should be of order $\sqrt{n}$. This is not much of a restriction since once k deviates from np by many standard deviations, $\mathbb{P}(S_n = k)$ becomes very small and can be approximated by zero. In what follows we assume that k and $n-k$ of order n.

We use Stirling's formula is the following form

$$m! \sim \sqrt{2\pi m} e^{-m} m^m,$$

where by $\sim$ we mean that the two quantities are asymptotically equal, that is,their ratio tends to 1 as $m \to \infty$. Then for large n, k and $n-k$

$$
\begin{aligned}
\mathbb{P}\left(S_n = k\right) &= \frac{n!}{k!\,(n-k)!} p^k \left(1-p\right)^{n-k} \\
&\sim \frac{\sqrt{2\pi n} e^{-n} n^n}{\sqrt{2\pi k} e^{-k} k^k \sqrt{2\pi\left(n-k\right)} e^{-(n-k)} \left(n-k\right)^{n-k}} p^k \left(1-p\right)^{n-k} \\
&= \left(\frac{p}{k}\right)^k \left(\frac{1-p}{n-k}\right)^{n-k} n^n \sqrt{\frac{n}{2\pi k\left(n-k\right)}} = \left(\frac{np}{k}\right)^k \left(\frac{n\left(1-p\right)}{n-k}\right)^{n-k} \sqrt{\frac{n}{2\pi k\left(n-k\right)}}.
\end{aligned}
$$

Now we can use identities

$$
\begin{aligned}
\ln\left(\frac{np}{k}\right) &= -\ln\left(1+\frac{k-np}{np}\right), \\
\ln\left(\frac{n\left(1-p\right)}{n-k}\right) &= -\ln\left(1-\frac{k-np}{n\left(1-p\right)}\right).
\end{aligned}
$$

Then we can use $\ln\left(1+y\right) \sim y - \frac{y^2}{2} + \frac{y^3}{3}, y \to 0$ to see that

$$
\begin{aligned}
&\ln\left(\left(\frac{np}{k}\right)^k \left(\frac{n\left(1-p\right)}{n-k}\right)^{n-k}\right) = k\ln\left(\frac{np}{k}\right) + \left(n-k\right)\ln\left(\frac{n\left(1-p\right)}{n-k}\right) \\
&\sim k\left(-\frac{k-np}{np} + \frac{1}{2}\left(\frac{k-np}{np}\right)^2 - \frac{1}{3}\left(\frac{k-np}{np}\right)^3\right) \\
&+ \left(n-k\right)\left(\frac{k-np}{n\left(1-p\right)} + \frac{1}{2}\left(\frac{k-np}{n\left(1-p\right)}\right)^2 + \frac{1}{3}\left(\frac{k-np}{n\left(1-p\right)}\right)^3\right) \\
&\sim -\frac{\left(k-np\right)^2}{2np\left(1-p\right)}.
\end{aligned}
$$

Thus

$$
\left(\frac{np}{k}\right)^k \left(\frac{n\left(1-p\right)}{n-k}\right)^{n-k} \sim e^{-\frac{(k-np)^2}{2np(1-p)}}.
$$

Now we use our assumption that $k-np$ should be of order $\sqrt{n}$ to see that

$$
\begin{aligned}
k - np &\approx \sqrt{n}, \\
n - k &\approx n\left(1-p\right) - \sqrt{n}, \\
k\left(n-k\right) &\approx n^2 p\left(1-p\right),
\end{aligned}
$$

so

$$\sqrt{\frac{n}{2\pi k\,(n-k)}} \sim \frac{1}{\sqrt{2\pi np\,(1-p)}}.$$

□

Example 9.1. Suppose a fair coin is tossed 100 times. What is the probability there will be more than 60 heads?

First observe that $np = 50$ and $\sqrt{np(1-p)} = 5$. Then we have

$$\mathbb{P}(S_n \geqslant 60) = \mathbb{P}((S_n - 50)/5 \geqslant 2) \approx \mathbb{P}(Z \geqslant 2) \approx 0.0228.$$

Example 9.2. Suppose a die is rolled 180 times. What is the probability a 3 will be showing more than 50 times?

Here $p = \frac{1}{6}$, so $np = 30$ and $\sqrt{np(1-p)} = 5$. Then $\mathbb{P}(S_n > 50) \approx \mathbb{P}(Z > 4)$, which is less than $e^{-4^2/2}$.

Example 9.3. Suppose a drug is supposed to be 75% effective. It is tested on 100 people. What is the probability more than 70 people will be helped?

Here S_n is the number of successes, $n = 100$, and $p = 0.75$. We have

$$\begin{aligned}\mathbb{P}(S_n \geqslant 70) &= \mathbb{P}((S_n - 75)/\sqrt{300/16} \geqslant -1.154) \\ &\approx \mathbb{P}(Z \geqslant -1.154) \approx 0.87,\end{aligned}$$

where the last number came from table 1.

When $b - a$ is small, there is a correction that makes things more accurate, namely replace a by $a - \frac{1}{2}$ and b by $b + \frac{1}{2}$. This correction never hurts and is sometime necessary. For example, in tossing a coin 100 times, there is positive probability that there are exactly 50 heads, while without the correction, the answer given by the normal approximation would be 0.

Example 9.4. We toss a coin 100 times. What is the probability of getting 49, 50, or 51 heads?

We write $\mathbb{P}(49 \leqslant S_n \leqslant 51) = \mathbb{P}(48.5 \leqslant S_n \leqslant 51.5)$ and then continue as above.

In this case we again have

$$\begin{aligned}p &= 0.5, \\ \mu &= np = 50, \\ \sigma^2 &= np(1-p) = 25, \\ \sigma &= \sqrt{np(1-p)} = 5.\end{aligned}$$

The normal approximation can be done in three different ways:

$$\mathbb{P}(49 \leqslant S_n \leqslant 51) \approx \mathbb{P}(49 \leqslant 50 + 5Z \leqslant 51) = \Phi(0.2) - \Phi(-0.2) = 2\Phi(0.2) - 1 \approx 0.15852$$

or

$$\mathbb{P}(48 < S_n < 52) \approx \mathbb{P}(48 < 50 + 5Z < 52) = \Phi(0.4) - \Phi(-0.4) = 2\Phi(0.4) - 1 \approx 0.31084$$

or

$$\mathbb{P}(48.5 < S_n < 51.5) \approx \mathbb{P}(48.5 < 50{+}5Z < 51.5) = \Phi(0.3){-}\Phi(-0.3) = 2\Phi(0.3){-}1 \approx 0.23582$$

Here all three answers are approximate, and the third one, **0.23582**, is the most accurate among these three. We also can compute the precise answer using the binomial formula:

$$\mathbb{P}(49 \leqslant S_n \leqslant 51) = \sum_{k=49}^{51} \binom{100}{k} \left(\frac{1}{2}\right)^{100} = \frac{37339688790147532337148742857}{158456325028528675187087900672}$$
$$\approx 0.2356465655973331958...$$

In addition we can obtain the following normal approximations

$$\mathbb{P}(S_n = 49) \approx \mathbb{P}(48.5 \leqslant 50 + 5Z \leqslant 49.5) = \Phi(-0.1) - \Phi(-0.3) = \Phi(0.3) - \Phi(0.1) \approx 0.07808$$
$$\mathbb{P}(S_n = 50) \approx \mathbb{P}(49.5 \leqslant 50 + 5Z \leqslant 50.5) = \Phi(0.1) - \Phi(-0.1) = 2\Phi(0.1) - 1 \approx 0.07966$$
$$\mathbb{P}(S_n = 51) \approx \mathbb{P}(50.5 \leqslant 50 + 5Z \leqslant 51.5) = \Phi(0.3) - \Phi(0.1) \approx 0.07808$$

Finally, notice that

$$0.07808 + 0.07966 + 0.07808 = 0.23582$$

which is the approximate value for $\mathbb{P}(49 \leqslant S_n \leqslant 51) \approx \mathbb{P}(48.5 < 50 + 5Z < 51.5)$.

Continuity correction

If a continuous distribution such as the normal distribution is used to approximate a discrete one such as the binomial distribution, a *continuity correction* should be used.

For example, if X is a binomial random variable that represents the number of successes in n independent trials with the probability of success in any trial p, and Y is a normal random variable with the same mean and the same variance as X. Then for any integer k we have that $\mathbb{P}(X \leqslant k)$ is well approximated by $\mathbb{P}(Y \leqslant k)$ if $np(1-p)$ is not too small. It is better approximated by $\mathbb{P}(Y \leqslant k + 1/2)$ as explained at the end of this section. The role of $1/2$ is clear if we start by looking at the normal distribution first, and seeing how we use it to approximate the binomial distribution.

The fact that this approximation is better based on a couple of considerations. One is that a discrete random variable can only take on only discrete values such as integers, while a continuous random variable used to approximate it can take on any values within an interval around these specified values. Hence, when using the normal distribution to approximate the binomial, more accurate approximations are likely to be obtained if a continuity correction is used.

The second reason is that a continuous distribution such as the normal, the probability of taking on a particular value of a random variable is zero. On the other hand, when the normal approximation is used to approximate a discrete distribution, a continuity correction can be employed so that we can approximate the probability of a specific value of the discrete distribution.

For example, if we want to approximate $\mathbb{P}(3 \leqslant X \leqslant 5) = \mathbb{P}(X = 3 \text{ or } X = 4 \text{ or } X = 5)$ by a normal distribution, it would be a *bad* approximation to use $\mathbb{P}(Y = 3 \text{ or } Y = 4 \text{ or } Y = 5)$ as the probability of Y taking on 3, 4 and 5 is 0. We can use *continuity correction* to see that

$$\mathbb{P}(3 \leqslant X \leqslant 5) = \mathbb{P}(2.5 \leqslant X \leqslant 5.5)$$

and *then* use the normal approximation by $\mathbb{P}(2.5 \leqslant Y \leqslant 5.5)$.

Below is a table on how to use the continuity correction for normal approximation to a binomial.

Binomial	Normal
If $\mathbb{P}(X = n)$	use $\mathbb{P}(n - 0.5 < X < n + 0.5)$
If $\mathbb{P}(X > n)$	use $\mathbb{P}(X > n + 0.5)$
If $\mathbb{P}(X \leqslant n)$	use $\mathbb{P}(X < n + 0.5)$
If $\mathbb{P}(X < n)$	use $\mathbb{P}(X < n - 0.5)$
If $\mathbb{P}(X \geqslant n)$	use $\mathbb{P}(X > n - 0.5)$

9.1. Exercises

Exercise 9.1. Suppose that we roll 2 dice 180 times. Let E be the event that we roll two fives no more than once.

(a) Find the exact probability of E.
(b) Approximate $\mathbb{P}(E)$ using the normal distribution.
(c) Approximate $\mathbb{P}(E)$ using the Poisson distribution.

Exercise 9.2. About 10% of the population is left-handed. Use the normal distribution to approximate the probability that in a class of 150 students,

(a) at least 25 of them are left-handed.
(b) between 15 and 20 are left-handed.

Exercise 9.3. A teacher purchases a box with 50 markers of colors selected at random. The probability that marker is black is 0.6, independent of all other markers. Knowing that the probability of there being more than N black markers is greater than 0.2 and the probability of there being more than N + 1 black markers is less than 0.2, use the normal approximation to calculate N.

9.2. Selected solutions

Solution to Exercise 9.1(A): The probability of rolling two fives in a particular roll is $\frac{1}{36}$, so the probability that we roll two fives no more than once in 180 rolls is

$$p = \begin{pmatrix} 180 \\ 0 \end{pmatrix} \left(\frac{35}{36}\right)^{180} + \begin{pmatrix} 180 \\ 1 \end{pmatrix} \left(\frac{1}{36}\right) \left(\frac{35}{36}\right)^{179} \approx .0386.$$

Solution to Exercise 9.1(B): we are interested in the number of successes to be 0 or 1, that is, $\mathbb{P}(0 \leqslant S_{180} \leqslant 1)$. Since the binomial is integer-valued, we apply the continuity correction and calculate $\mathbb{P}(-0.5 \leqslant S_{180} \leqslant 1.5)$ instead. We find that the expected value is $\mu = 180 \cdot p = 5$ and the standard deviation is $\sigma = \sqrt{180p(1-p)} \approx 2.205$. Now, as always, we convert this question to a question about the standard normal random variable Z,

$$\begin{aligned} \mathbb{P}(-0.5 \leqslant S_{180} \leqslant 1.5) = \mathbb{P}\left(\frac{-0.5-5}{2.205} \leqslant Z \leqslant \frac{1.5-5}{2.205}\right) &= \mathbb{P}(-2.49 < Z < -1.59) \\ &= (1 - \Phi(1.59)) - (1 - \Phi(2.49)) \\ &= (1 - 0.9441) - (1 - 0.9936) = 0.0495. \end{aligned}$$

Solution to Exercise 9.1(C): We use $\lambda = np = 5$ (note that we found this already in (b)!). Now we see that

$$\mathbb{P}(E) \approx e^{-5}\frac{5^0}{0!} + e^{-5}\frac{5^1}{1!} \approx 0.0404.$$

Solution to Exercise 9.2: Let X denote the number of left-handed students in the class. We use Theorem 9.1 with $X \sim \text{Binom}(150, 0.1)$ below. Note that $np = 15$.

(a) $\mathbb{P}(X \geqslant 25) = \mathbb{P}\left(\frac{X-15}{\sqrt{13.5}} \geqslant \frac{10}{\sqrt{13.5}}\right) \approx 1 - \Phi(2.72) \approx 0.00364$. Note that in this approximation we implicitly use that the tail of this probability distribution is small, and so instead of a two-sided interval we just used a one-sided interval.

We can see that the result is really close to the two-sided estimates as follows.

$$\begin{aligned} &\mathbb{P}(150 \geqslant X \geqslant 25) = \mathbb{P}\left(\frac{135}{\sqrt{13.5}} \geqslant \frac{X-15}{\sqrt{13.5}} \geqslant \frac{10}{\sqrt{13.5}}\right) \\ &\approx \Phi(36.74) - \Phi(2.72) \approx 0.00364. \end{aligned}$$

Finally, with the *continuity correction* the solution is

$$\begin{aligned} &\mathbb{P}(150 \geqslant X \geqslant 25) = \mathbb{P}(150.5 \geqslant X \geqslant 24.5) \\ &= \mathbb{P}\left(\frac{135.5}{\sqrt{13.5}} \geqslant \frac{X-15}{\sqrt{13.5}} \geqslant \frac{9.5}{\sqrt{13.5}}\right) \approx \Phi(36.87) - \Phi(2.59) \approx 0.00480. \end{aligned}$$

(2) Similarly to the first part

$$\begin{aligned}\mathbb{P}(15 \leqslant X \leqslant 20) &= \mathbb{P}(14.5 < X < 20.5)\\ &= \Phi\left(\frac{5.5}{\sqrt{13.5}}\right) - \Phi\left(\frac{-0.5}{\sqrt{13.5}}\right) \approx \Phi(1.5) - 1 + \Phi(0.14) \approx 0.4889.\end{aligned}$$

Solution to Exercise 9.3: Let X denote the number of black markers. Since $X \sim \text{Binom}(50, 0.6)$ we have

$$\mathbb{P}(X > N) \approx 1 - \Phi\left(\frac{N-30}{2\sqrt{3}}\right) > 0.2 \text{ and } \mathbb{P}(X > N+1) \approx 1 - \Phi\left(\frac{N-29}{2\sqrt{3}}\right) < 0.2.$$

From this we deduce that $N \leqslant 32.909$ and $N \geqslant 31.944$ so that $N = 32$.

CHAPTER 10

Some continuous distributions

10.1. Examples of continuous random variables

We look at some other continuous random variables besides normals.

Uniform distribution

A continuous random variable has *uniform distribution* if its density is $f(x) = 1/(b-a)$ if $a \leqslant x \leqslant b$ and 0 otherwise.

For a random variable X with uniform distribution its expectation is

$$\mathbb{E}X = \frac{1}{b-a}\int_a^b x\,dx = \frac{a+b}{2}.$$

Exponential distribution

A continuous random variable has *exponential distribution* with parameter $\lambda > 0$ if its density is $f(x) = \lambda e^{-\lambda x}$ if $x \geqslant 0$ and 0 otherwise.

Suppose X is a random variable with an exponential distribution with parameter λ. Then we have

$$\mathbb{P}(X > a) = \int_a^\infty \lambda e^{-\lambda x} dx = e^{-\lambda a}, \tag{10.1.1}$$

$$F_X(a) = 1 - \mathbb{P}(X > a) = 1 - e^{-\lambda a},$$

and we can use integration by parts to see that $\mathbb{E}X = 1/\lambda$, $\operatorname{Var} X = 1/\lambda^2$. Examples where an exponential random variable is a good model is the length of a telephone call, the length of time before someone arrives at a bank, the length of time before a light bulb burns out.

Exponentials are memoryless, that is,

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s),$$

or given that the light bulb has burned 5 hours, the probability it will burn 2 more hours is the same as the probability a new light bulb will burn 2 hours. Here is how we can prove this

$$\begin{aligned}\mathbb{P}(X > s+t \mid X > t) &= \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} \\ &= \mathbb{P}(X > s),\end{aligned}$$

where we used Equation (10.1.1) for $a = t$ and $a = s + t$.

Gamma distribution

A continuous random variable has a *Gamma distribution* with parameters α and θ if its density is

$$f(x) = \frac{\alpha e^{-\alpha x}(\alpha x)^{\theta-1}}{\Gamma(\theta)}$$

if $x \geqslant 0$ and 0 otherwise. Here $\Gamma(\theta) = \int_0^\infty e^{-y} y^{\theta-1} dy$ is the gamma function. We denote such a distribution by $\Gamma(\alpha, \theta)$.

Note that $\Gamma(1) = \int_0^\infty e^{-y} dy = 1$, and using induction on n and integration by parts one can see that

$$\Gamma(n) = (n-1)!$$

so we say that gamma function interpolates the factorial.

While an exponential random variable is used to model the time for something to occur, a gamma random variable is the time for θ events to occur. A gamma random variable, $\Gamma\left(\frac{1}{2}, \frac{n}{2}\right)$, with parameters $\frac{1}{2}$ and $\frac{n}{2}$ is known as a χ^2_n, a chi-squared random variable with n degrees of freedom. Recall that in Exercise 8.9 we had a different description of a χ^2 random variable, namely, Z^2 with $Z \sim \mathcal{N}(0,1)$. Gamma and χ^2 random variables come up frequently in statistics.

Beta distribution

A continuous random variable has a *Beta distribution* if its density is

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1,$$

where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}$.

This is also a distribution that appears often in statistics.

Cauchy distribution

A continuous random variable has a *Cauchy distribution* if its density is

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x-\theta)^2}.$$

What is interesting about the Cauchy random variable is that it does not have a finite mean, that is, $\mathbb{E}|X| = \infty$.

Densities of functions of continuous random variables

Often it is important to be able to compute the density of $Y = g(X)$, where X is a continuous random variable. This is explained later in Theorem 11.1.

Let us give a couple of examples.

Example 10.1 (Log of a uniform random variable). If X is uniform on the interval $[0,1]$ and $Y = -\log X$, so $Y > 0$. If $x > 0$, then

$$\begin{aligned} F_Y(x) &= \mathbb{P}(Y \leqslant x) = \mathbb{P}(-\log X \leqslant x) \\ &= \mathbb{P}(\log X \geqslant -x) = \mathbb{P}(X \geqslant e^{-x}) = 1 - \mathbb{P}(X \leqslant e^{-x}) \\ &= 1 - F_X(e^{-x}). \end{aligned}$$

Taking the derivative we see that

$$f_Y(x) = \frac{d}{dx} F_Y(x) = -f_X(e^{-x})(-e^{-x}),$$

using the chain rule. Since $f_X(x) = 1$, $x \in [0,1]$, this gives $f_Y(x) = e^{-x}$, or Y is exponential with parameter 1.

Example 10.2 (χ^2, revisited). As in Exercise 8.9 we consider $Y = Z^2$, where $Z \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned} F_Y(x) &= \mathbb{P}(Y \leqslant x) = \mathbb{P}(Z^2 \leqslant x) = \mathbb{P}(-\sqrt{x} \leqslant Z \leqslant \sqrt{x}) \\ &= \mathbb{P}(Z \leqslant \sqrt{x}) - \mathbb{P}(Z \leqslant -\sqrt{x}) = F_Z(\sqrt{x}) - F_Z(-\sqrt{x}). \end{aligned}$$

Taking the derivative and using the chain rule we see

$$f_Y(x) = \frac{d}{dx} F_Y(x) = f_Z(\sqrt{x}) \left(\frac{1}{2\sqrt{x}} \right) - f_Z(-\sqrt{x}) \left(-\frac{1}{2\sqrt{x}} \right).$$

Recall that $f_Z(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ and doing some algebra, we end up with

$$f_Y(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2},$$

which is $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. As we pointed out before, this is also a χ^2 distributed random variable with one degree of freedom.

Example 10.3 (Tangent of a uniform random variable). Suppose X is a uniform random variable on $[-\pi/2, \pi/2]$ and $Y = \tan X$. Then

$$F_Y(x) = \mathbb{P}(X \leqslant \tan^{-1} x) = F_X(\tan^{-1} x),$$

and taking the derivative yields

$$f_Y(x) = f_X(\tan^{-1} x)\frac{1}{1+x^2} = \frac{1}{\pi}\frac{1}{1+x^2},$$

which is a Cauchy distribution.

10.2. Further examples and applications

Example 10.4. Suppose that the length of a phone call in minutes is an exponential random variable with average length 10 minutes.

(1) What is the probability of your phone call being more than 10 minutes?
Here $\lambda = \frac{1}{10}$, thus

$$\mathbb{P}(X > 10) = e^{-\left(\frac{1}{10}\right)10} = e^{-1} \approx 0.368.$$

(2) Between 10 and 20 minutes?
We have that

$$\mathbb{P}(10 < X < 20) = F(20) - F(10) = e^{-1} - e^{-2} \approx 0.233.$$

Example 10.5. Suppose the life of an Uphone has exponential distribution with mean life of 4 years. Let X denote the life of an Uphone (or time until it dies). Given that the Uphone has lasted 3 years, what is the probability that it will 5 more years.

In this case $\lambda = \frac{1}{4}$.

$$\mathbb{P}\left(X > 5 + 3 \mid X > 3\right) = \frac{\mathbb{P}\left(X > 8\right)}{\mathbb{P}\left(X > 3\right)}$$

$$= \frac{e^{-\frac{1}{4}\cdot 8}}{e^{-\frac{1}{4}\cdot 3}} = e^{-\frac{1}{4}\cdot 5} = \mathbb{P}\left(X > 5\right).$$

Recall that an exponential random variable is memoryless, so our answer is consistent with this property of X.

10.3. Exercises

Exercise 10.1. Suppose that the time required to replace a car's windshield can be represented by an exponentially distributed random variable with parameter $\lambda = \frac{1}{2}$.

(a) What is the probability that it will take at least 3 hours to replace a windshield?
(b) What is the probability that it will take at least 5 hours to replace a windshield given that it hasn't been finished after 2 hours?

Exercise 10.2. The number of years a Uphone functions is exponentially distributed with parameter $\lambda = \frac{1}{8}$. If Pat buys a used Uphone, what is the probability that it will be working after an additional 8 years?

Exercise 10.3. Suppose that the time (in minutes) required to check out a book at the library can be represented by an exponentially distributed random variable with parameter $\lambda = \frac{2}{11}$.

(a) What is the probability that it will take at least 5 minutes to check out a book?
(b) What is the probability that it will take at least 11 minutes to check out a book given that you have already waited for 6 minutes?

Exercise 10.4. Let X be an exponential random variable with mean $\mathbb{E}X = 1$. Define a new random variable $Y = e^X$. Find the PDF of Y, $f_Y(y)$.

Exercise 10.5. Suppose that X has an exponential distribution with parameter $\lambda = 1$. For any $c > 0$ show that $Y = \frac{X}{c}$ is exponential with parameter $\lambda = c$.

Exercise 10.6. Let X be a uniform random variable over $[0, 1]$. Define a new random variable $Y = e^X$. Find the probability density function of Y, $f_Y(y)$.

Exercise 10.7. An insurance company insures a large number of homes. The insured value, X, of a randomly selected home is assumed to follow a distribution with density function

$$f_X(x) = \begin{cases} \frac{8}{x^3} & x > 2, \\ 0 & \text{otherwise}. \end{cases}$$

(A) Given that a randomly selected home is insured for at most 4, calculate the probability that it is insured for less than 3.
(B) Given that a randomly selected home is insured for at least 3, calculate the probability that it is insured for less than 4.

Exercise 10.8. A hospital is to be located along a road of infinite length. If population density is exponentially distributed along the road, where should the station be located to minimize the expected distance to travel to the hospital? That is, find an a to minimize $\mathbb{E}|X - a|$, where X is exponential with rate λ.

10.4. Selected solutions

Solution to Exercise 10.1(A): We have

$$\begin{aligned}\mathbb{P}(X>3)&=1-\mathbb{P}(0<X<3)\\&=1-\int_0^3\frac{1}{2}e^{-\frac{x}{2}}dx=e^{-\frac{3}{2}}\approx 0.2231.\end{aligned}$$

Solution to Exercise 10.1(B): There are two ways to do this. The longer one is to explicitly find $\mathbb{P}(X>5\mid X>2)$. The shorter one is to remember that the exponential distribution is memoryless and to observe that $\mathbb{P}(X>t+3\mid X>t)=P(X>3)$, so the answer is the same as the answer to part (a).

Solution to Exercise 10.2: e^{-1}

Solution to Exercise 10.3(A): Recall that by Equation 10.1.1 $\mathbb{P}(X>a)=e^{-\lambda a}$, and therefore

$$\mathbb{P}(X>5)=e^{-\frac{10}{11}}.$$

Solution to Exercise 10.3(B): We use the memoryless property

$$\begin{aligned}\mathbb{P}(X>11\mid X>6)&=\mathbb{P}(X>6+5\mid X>6)\\&=\mathbb{P}(X>5)=e^{-\frac{10}{11}}.\end{aligned}$$

Solution to Exercise 10.4: Since $\mathbb{E}[X]=1/\lambda=1$ then we know that $\lambda=1$. Then its PDF and CDF are

$$\begin{aligned}f_X(x)&=e^{-x},x\geqslant 0\\F_X(x)&=1-e^{-x},x\geqslant 0.\end{aligned}$$

Thus

$$F_Y(y)=\mathbb{P}(Y\leqslant y)=\mathbb{P}\left(e^X\leqslant y\right)=\mathbb{P}(X\leqslant\log y)=F_X(\log y),$$

and so

$$F_Y(y)=1-e^{-\log y}=1-\frac{1}{y},\ \text{when}\ \log(y)\geqslant 0,$$

taking derivative we get

$$f_Y(y)=\frac{dF_y(y)}{dy}=\frac{1}{y^2}\ \text{when}\ y\geqslant 1.$$

Solution to Exercise 10.5: Since X is exponential with parameter 1, then its PDF and CDF are

$$\begin{aligned}f_X(x)&=e^{-x},x\geqslant 0\\F_X(x)&=1-e^{-x},x\geqslant 0.\end{aligned}$$

Thus

$$F_Y(y)=\mathbb{P}(Y\leqslant y)=\mathbb{P}\left(\frac{X}{c}\leqslant y\right)=\mathbb{P}(X\leqslant cy)=F_X(cy),$$

and so

$$F_Y(y)=1-e^{-cy},\ \text{when}\ cy\geqslant 0,$$

taking derivatives we get

$$f_Y(y) = \frac{dF_y(y)}{dy} = ce^{-cy}, \text{ when } y \geqslant 0.$$

Note that this is the PDF of an exponential with parameter $\lambda = c$.

Solution to Exercise 10.6: Since X is uniform over $[0,1]$, then its PDF and CDF are

$$f_X(x) = 1, 0 \leqslant x < 1$$
$$F_X(x) = x, 0 \leqslant x < 1.$$

Thus

$$F_y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}\left(e^X \leqslant y\right) = \mathbb{P}(X \leqslant \log y) = F_X(\log y)$$

and so

$$F_Y(y) = \log y, \text{ when } 0 \leqslant \log y < 1.$$

taking derivatives we get

$$f_Y(y) = \frac{dF_y(y)}{dy} = \frac{1}{y}, \text{ when } 1 < y < e^1.$$

Solution to Exercise 10.7(A): Using the definition of conditional probability

$$\mathbb{P}(X < 3 \mid X < 4) = \frac{\mathbb{P}(X < 3)}{\mathbb{P}(X < 4)}.$$

Since

$$\mathbb{P}(X < 4) = \int_2^4 \frac{8}{x^3} dx = -\frac{4}{x^2}\Big|_2^4 = -\frac{1}{4} + 1 = \frac{3}{4}$$

and

$$\mathbb{P}(X < 3) = \mathbb{P}(2 < X < 3) = \int_2^3 \frac{8}{x^3} dx = -\frac{4}{x^2}\Big|_2^3 = -\frac{4}{9} + 1 = \frac{5}{9}.$$

Thus the answer is

$$\frac{\frac{5}{9}}{\frac{3}{4}} = \frac{20}{27} \approx 0.74074074$$

Solution to Exercise 10.7(B): Using the definition of conditional probability

$$\mathbb{P}(X < 4 \mid X > 3) = \frac{\mathbb{P}(3 < X < 4)}{\mathbb{P}(X > 3)}.$$

Since

$$\mathbb{P}(X > 3) = \int_3^\infty \frac{8}{x^3} dx = -\frac{4}{x^2}\Big|_3^\infty = \frac{4}{9},$$
$$\mathbb{P}(3 < X < 4) = \int_3^4 \frac{8}{x^3} dx = -\frac{4}{x^2}\Big|_3^4 = \frac{7}{36}.$$

Thus the answer is

$$\frac{\frac{7}{36}}{\frac{4}{9}} = \frac{7}{16} \approx 0.4375$$

Part 3

Multivariate discrete and continuous random variables

CHAPTER 11

Multivariate distributions

11.1. Joint distribution for discrete and continuous random variables

We are often interested in considering several random variables that might be related to each other. For example, we can be interested in several characteristics of a randomly chosen object, such as gene mutations and a certain disease for a person, different kinds of preventive measures for an infection etc. Each of these characteristics can be thought of as a random variable, and we are interested in their dependence. In this chapter we d study joint distributions of several random variables.

Consider collections of random variables $(X_1, X_2, \ldots, X_n)$, which are known as random vectors. We start by looking at two random variables, though the approach can be easily extended to more variables.

Joint PMF for discrete random variables

The *joint probability mass function* of two discrete random variables X and Y is defined as

$$p_{XY}(x, y) = \mathbb{P}(X = x, Y = y).$$

Recall that here the comma means *and*, or the intersection of two events. If X takes values $\{x_i\}_{i=1}^{\infty}$ and Y takes values $\{y_j\}_{j=1}^{\infty}$, then the range of (X, Y) as a map from the probability space $(S, \mathcal{F}, \mathbb{P})$ to the set $\{(x_i, y_j)\}_{i,j=1}^{\infty}$. Note that p_{XY} is indeed a probability mass function as

$$\sum_{i,j=1}^{\infty} p_{XY}(x_i, y_j) = 1.$$

Definition (Joint PDF for continuous random variables)

Two random variables X and Y are *jointly continuous* if there exists a nonnegative function $f_{XY} : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$, such that, for any set $A \subseteq \mathbb{R} \times \mathbb{R}$, we have

$$\mathbb{P}((X, Y) \in A) = \iint_A f_{XY}(x, y)\, dxdy.$$

The function $f_{XY}(x, y)$ is called the *joint probability density function* (PDF) of X and Y.

In particular, for $A = \{(x, y) : a \leqslant x \leqslant b, c \leqslant y \leqslant d\}$ we have

$$\mathbb{P}(a \leqslant X \leqslant b, c \leqslant Y \leqslant d) = \int_a^b \int_c^d f_{XY}(x,y) dy\, dx.$$

Another example is $A = \{(x,y) : x < y\}$, then

$$\mathbb{P}(X < Y) = \iint_{\{x<y\}} f_{X,Y}(x,y) dy\, dx.$$

Example 11.1. If the density $f_{X,Y}(x,y) = ce^{-x}e^{-2y}$ for $0 < x < \infty$ and $x < y < \infty$, what is c?

We use the fact that a density must integrate to 1. So

$$\int_0^\infty \int_x^\infty ce^{-x}e^{-2y} dy\, dx = 1.$$

Recalling multivariable calculus, this double integral is equal to

$$\int_0^\infty ce^{-x}\tfrac{1}{2}e^{-2x} dx = \frac{c}{6},$$

so $c = 6$.

The multivariate distribution function (CDF) of (X,Y) is defined by $F_{X,Y}(x,y) = \mathbb{P}(X \leqslant x, Y \leqslant y)$. In the continuous case, this is

$$F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(x,y) dy\, dx,$$

and so we have

$$f(x,y) = \frac{\partial^2 F}{\partial x \partial y}(x,y).$$

The extension to n random variables is exactly similar.

Marginal PDFs

Suppose $f_{X,Y}(x,y)$ is a joint PDF of X and Y, then the *marginal densities* of X and of Y are given by

$$f_X(x) = \int_{-\infty}^\infty f_{X,Y}(x,y) dy, \quad f_Y(y) = \int_{-\infty}^\infty f_{X,Y}(x,y) dx.$$

Example 11.2 (Binomial distribution as a joint distribution)**.** If we have a binomial with parameters n and p, this can be thought of as the number of successes in n trials, and

$$\mathbb{P}(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

If we let $k_1 = k$, $k_2 = n - k$, $p_1 = p$, and $p_2 = 1 - p$, this can be rewritten as

$$\frac{n!}{k_1!k_2!}p_1^{k_1}p_2^{k_2},$$

as long as $n = k_1 + k_2$. Thus this is the probability of k_1 successes and k_2 failures, where the probabilities of success and failure are p_1 and p_2, resp.

Example 11.3 (Multivariate binomial distribution)**.** A multivariate random vector is $(X_1, \ldots, X_r)$ with

$$\mathbb{P}(X_1 = n_1, \ldots, X_r = n_r) = \frac{n!}{n_1! \cdots n_r!}p_1^{n_1} \cdots p_r^{n_r},$$

where $n_1 + \cdots + n_r = n$ and $p_1 + \cdots p_r = 1$. Thus this generalizes the binomial to more than 2 categories.

11.2. Independent random variables

Now we describe the concept of *independence* for random variables.

Definition (Independent random variables)

Two discrete random variables X and Y are *independent* if $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ for all x and y.
Two continuous random variables X and Y are *independent* if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

for all pairs of subsets A, B of the real line $\mathbb{R}$.

Recall that the left-hand side should be understood as

$$\mathbb{P}\left(\{\omega : X(\omega) \text{ is in } A \text{ and } Y(\omega) \text{ is in } B\}\right)$$

and similarly for the right-hand side.

In the discrete case, if we have independence, then in terms of PMFs we have

$$\begin{aligned} p_{X,Y}(x, y) &= \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &= p_X(x)p_Y(y). \end{aligned}$$

In other words, the joint probability mass faunction $p_{X,Y}$ factors.

In the continuous case for any a, b, c, d we have

$$\begin{aligned} \int_a^b \int_c^d f_{X,Y}(x, y)dy\,dx &= \mathbb{P}(a \leqslant X \leqslant b, c \leqslant Y \leqslant d) \\ &= \mathbb{P}(a \leqslant X \leqslant b)\mathbb{P}(c \leqslant Y \leqslant d) \\ &= \int_a^b f_X(x)dx \int_c^d f_Y(y)dy \\ &= \int_a^b \int_c^d f_X(x)f_Y(y)\,dy\,dx. \end{aligned}$$

One can conclude from this by taking partial derivatives that the joint density function factors into the product of the density functions. Going the other way, one can also see that if the joint density factors, then one has independence of random variables.

11.1 (The joint density function factors for independent random variables)

Jointly continuous random variables X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

Example 11.4 (Buffon's needle problem)**.** Suppose one has a floor made out of wood planks and one drops a needle onto it. What is the probability the needle crosses one of the cracks? Suppose the needle is of length L and the wood planks are D across.

Let X be the distance from the midpoint of the needle to the nearest crack and let Θ be the angle the needle makes with the vertical. Then X and Θ are independent random variables. X is uniform on $[0, D/2]$ and Θ is uniform on $[0, \pi/2]$. A little geometry shows that the needle will cross a crack if $L/2 > X/\cos\Theta$. We have $f_{X,\Theta} = \frac{4}{\pi D}$ and so we have to integrate this constant over the set where $X < L\cos\Theta/2$ and $0 \leqslant \Theta \leqslant \pi/2$ and $0 \leqslant X \leqslant D/2$. The integral is

$$\int_0^{\pi/2}\int_0^{L\cos\theta/2} \frac{4}{\pi D} dx\, d\theta = \frac{2L}{\pi D}.$$

Proposition 11.1 (Sums of independent random variables)

Suppose X and Y are independent continuous random variables, then the density of $X+Y$ is given by the following *convolution formula*

$$f_{X+Y}(a) = \int f_X(a-y)f_Y(y)dy.$$

Proof. If X and Y are independent, then the joint probability density function factors, and therefore

$$\begin{aligned}
\mathbb{P}(X+Y \leqslant a) &= \iint_{\{x+y\leqslant a\}} f_{X,Y}(x,y)dx\,dy \\
&= \iint_{\{x+y\leqslant a\}} f_X(x)f_Y(y)dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)dx\,dy \\
&= \int F_X(a-y)f_Y(y)dy.
\end{aligned}$$

Differentiating with respect to a, we have the convolution formula for the density of $X+Y$ as follows

$$f_{X+Y}(a) = \int f_X(a-y)f_Y(y)dy.$$

□

There are a number of cases where this is interesting.

Example 11.5. If X is a gamma random variable with parameters s and λ and Y is a gamma random variable with parameters t and λ, then a straightforward integration shows that $X+Y$ is a gamma with parameters $s+t$ and λ. In particular, the sum of n independent exponential random variables with parameter λ is a gamma random variable with parameters n and λ.

Example 11.6. If $Z \sim \mathcal{N}(0,1)$, then $F_{Z^2}(y) = \mathbb{P}(Z^2 \leqslant y) = \mathbb{P}(-\sqrt{y} \leqslant Z \leqslant \sqrt{y}) = F_Z(\sqrt{y}) - F_Z(-\sqrt{y})$. Differentiating shows that $f_{Z^2}(y) = ce^{-y/2}(y/2)^{(1/2)-1}$, or Z^2 is a gamma random variable with parameters $\frac{1}{2}$ and $\frac{1}{2}$. So using the previous example, if Z_i are independent $\mathcal{N}(0,1)$ random variables, then $\sum_{i=1}^n Z_i^2$ is a gamma random variable with parameters $n/2$ and $\frac{1}{2}$, i.e., a χ_n^2.

Example 11.7. If X_i is a $\mathcal{N}(\mu_i, \sigma_i^2)$ and the X_i are independent, then some lengthy calculations show that $\sum_{i=1}^n X_i$ is a $\mathcal{N}(\sum \mu_i, \sum \sigma_i^2)$.

Example 11.8. The analogue for discrete random variables is easier. If X and Y take only nonnegative integer values, we have

$$\begin{aligned}\mathbb{P}(X+Y=r) &= \sum_{k=0}^{r} \mathbb{P}(X=k, Y=r-k) \\ &= \sum_{k=0}^{r} \mathbb{P}(X=k)\mathbb{P}(Y=r-k).\end{aligned}$$

In the case where X is a Poisson random variable with parameter λ and Y is a Poisson random variable with parameter μ, we see that $X+Y$ is a Poisson random variable with parameter $\lambda+\mu$. To check this, use the above formula to get

$$\begin{aligned}\mathbb{P}(X+Y=r) &= \sum_{k=0}^{r} \mathbb{P}(X=k)\mathbb{P}(Y=r-k) \\ &= \sum_{k=0}^{r} e^{-\lambda}\frac{\lambda^k}{k!} e^{-\mu}\frac{\mu^{r-k}}{(r-k)!} \\ &= e^{-(\lambda+\mu)}\frac{1}{r!}\sum_{k=0}^{r}\binom{r}{k}\lambda^k\mu^{r-k} \\ &= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^r}{r!}\end{aligned}$$

using the binomial theorem.

Note that it is not always the case that the sum of two independent random variables will be a random variable of the same type.

Example 11.9. If X and Y are independent normals, then $-Y$ is also a normal with $\mathbb{E}(-Y) = -\mathbb{E}Y$ and $\operatorname{Var}(-Y) = (-1)^2 \operatorname{Var} Y = \operatorname{Var} Y$, and so $X - Y$ is also normal.

11.3. Conditioning for random variables

We now extend the notion of conditioning an event on another event to random variables. Suppose we have observed the value of a random variable Y, and we need to update the density function of another random variable X whose value we are still to observe. For this we use the conditional density function of X given Y.

We start by considering the case when X and Y are discrete random variables.

Definition (Conditional discrete random variable)

The *conditional random variable* $X \mid Y = y$ called X *given* $Y = y$, a discrete random variable with the probability mass function

$$\mathbb{P}\left((X \mid Y = y) = x\right) = \mathbb{P}\left(X = x \mid Y = y\right)$$

for values x of X.

In terms of probability mass functions we have

$$\begin{aligned} p_{(X|Y=y)}(x) &= \mathbb{P}_{X|Y}(x \mid y) = \mathbb{P}(X = x \mid Y = y) \\ &= \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{XY}(x, y)}{p_Y(y)} \end{aligned}$$

wherever $p_Y(y) \neq 0$.

Analogously, we define conditional random variables in the continuous case.

Definition (Conditional continuous random variable)

Suppose X and Y are jointly continuous, the *conditional probability density function (PDF)* of X given Y is given by

$$f_{X|Y=y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

11.4. Functions of random variables

First we formalize what we saw in the one-dimensional case. First we recall that $g(x)$ is called a *strictly increasing function* if for any $x_1 < x_2$, then $g(x_1) < g(x_2)$. Similarly we can define a *strictly decreasing function*. We also say g is a *strictly monotone function* on $[a, b]$ if it is either strictly increasing or strictly decreasing function on this interval.

Finally, we will use the fact that for a strictly monotone function g we have that for any point y in the range of the function there is a *unique* x such that $g(x) = y$. That is, we have a well-defined g^{-1} on the range of function g.

Theorem 11.1 (PDF for a function of a random variable)

Suppose $g(x)$ is differentiable and $g(x)$ is a strictly monotone function, and X is a continuous random variable. Then $Y = g(X)$ is a continuous random variable with the probability density function (PDF)

$$f_Y(y) = \begin{cases} \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|} = \frac{f_X(x)}{|g'(x)|}, & \text{if } y \text{ is in the range of the function } g \\ 0, & \text{if } y \text{ is } \textbf{not} \text{ in the range of the function } g \end{cases}, \tag{11.4.1}$$

where x is in the range of the random variable X.

Proof. Without loss of generality we can assume that g is strictly increasing. If X has a density f_X and $Y = g(X)$, then

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leqslant y) = \mathbb{P}(g(X) \leqslant y) \\ &= \mathbb{P}(X \leqslant g^{-1}(y)) = F_X(g^{-1}(y)). \end{aligned}$$

Taking the derivative, using the chain rule, and recalling that the derivative of $g^{-1}(y)$ is given by

$$(g^{-1}(y))' = \frac{1}{g'(x)} = \frac{1}{g'(g^{-1}(y))}.$$

Here we use that $y = g(x)$, $x = g^{-1}(y)$, and the assumption that $g(x)$ is an increasing function. □

The higher-dimensional case is very analogous. Note that the function below is assumed to be one-to-one, and therefore it is invertible. Note that in the one-dimensional case strictly monotone functions are one-to-one.

Theorem (PDF for a function of two random variables)

Suppose X and Y are two jointly continuous random variables. Let $(U, V) = g(X, Y) = (g_1(X, Y), g_2(X, Y))$, where $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a continuous one-to-one function with continuous partial derivatives. Denote by $h = g^{-1}$, so $h(U, V) = (h_1(U, V), h_2(U, V)) = (X, Y)$. Then U and V are jointly continuous and their joint PDF, $f_{UV}(u, v)$, is defined on the range of (U, V) and is given by

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))\,|J_h|,$$

where J_h is the Jacobian of the map h, that is,

$$J_h = \det \begin{pmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{pmatrix}.$$

PROOF. The proof is based on the change of variables theorem from multivariable calculus, and it is analogous to (11.4.1). Note that to reconcile this formula with some of the applications we might find the following property of the Jacobian for a map and its inverse useful

$$J_{g^{-1}} \circ g = (J_g)^{-1},$$

so

$$J_h \circ g = (J_g)^{-1}.$$

□

Example 11.10. Suppose X_1 is $\mathcal{N}(0,1)$, X_2 is $\mathcal{N}(0,4)$, and X_1 and X_2 are independent. Let $Y_1 = 2X_1+X_2, Y_2 = X_1-3X_2$. Then $y_1 = g_1(x_1,x_2) = 2x_1+x_2, y_2 = g_2(x_1,x_2) = x_1-3x_2$, so

$$J = \begin{pmatrix} 2 & 1 \\ 1 & -3 \end{pmatrix} = -7.$$

In general, J might depend on x, and hence on y. Some algebra leads to $x_1 = \frac{3}{7}y_1 + \frac{1}{7}y_2$, $x_2 = \frac{1}{7}y_1 - \frac{2}{7}y_2$. Since X_1 and X_2 are independent,

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}}e^{-x_1^2/2}\frac{1}{\sqrt{8\pi}}e^{-x_2^2/8}.$$

Therefore

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{\sqrt{2\pi}}e^{-(\frac{3}{7}y_1+\frac{1}{7}y_2)^2/2}\frac{1}{\sqrt{8\pi}}e^{-(\frac{1}{7}y_1-\frac{2}{7}y_2)^2/8}\frac{1}{7}.$$

11.5. Further examples and applications

Example 11.11. Suppose we roll two dice with sides $1, 1, 2, 2, 3, 3$. Let X be the largest value obtained on any of the two dice. Let $Y =$ be the sum of the two dice. Find the joint PMF of X and Y.

First we make a table of all the possible *outcomes*. Note that individually, $X = 1, 2, 3$ and $Y = 2, 3, 4, 5, 6$. The table for possible outcomes of (X, Y) jointly is given by the following table.

outcome	1	2	3
1	$(X = 1, Y = 2) = (1, 2)$	$(2, 3)$	$(3, 4)$
2	$(2, 3)$	$(2, 4)$	$(3, 5)$
3	$(3, 4)$	$(3, 5)$	$(3, 6)$

Using this table we have that the PMF is given by

$X\backslash Y$	2	3	4	5	6
1	$\mathbb{P}(X = 1, Y = 2) = \frac{1}{9}$	0	0	0	0
2	0	$\frac{2}{9}$	$\frac{1}{9}$	0	0
3	0	0	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Example 11.12. Let X, Y have joint PDF

$$f(x, y) = \begin{cases} ce^{-x}e^{-2y} & , 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}.$$

(a) Find c that makes this a valid PDF. The region that we integrate over is the first quadrant, therefore

$$\begin{aligned} 1 = & \int_0^\infty \int_0^\infty ce^{-x}e^{-2y} dx dy = c \int_0^\infty e^{-2y} \left[-e^{-x}\right]_0^\infty dy \\ = & c \int_0^\infty e^{-2y} dy = c - \frac{1}{2} e^{-2y} \Big|_0^\infty = \frac{c}{2}, \end{aligned}$$

and so $c = 2$.

(b) Find $\mathbb{P}(X < Y)$. We start by describing the region corresponding to this event, namely, $D = \{(x, y) \mid 0 < x < y, 0 < y < \infty\}$ and set up the double integral for the probability of this event.

$$\mathbb{P}(X < Y) = \int\int_D f(x,y)dA$$
$$= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}dxdy = \frac{1}{3}.$$

(c) Set up the double integral representing $\mathbb{P}(X > 1, Y < 1)$.

$$\mathbb{P}(X > 1, Y < 1) = \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}dxdy = (1 - e^{-2})e^{-1}.$$

(d) Find the marginal $f_X(x)$. We have

$$f_X(x) = \int_{-\infty}^\infty f(x,y)dy = \int_0^\infty 2e^{-x}e^{-2y}dy$$
$$= 2e^{-x}\left(\frac{-e^{-2y}}{2}\bigg|_0^\infty\right) = 2e^{-x}\left(0 + \frac{1}{2}\right)$$
$$= e^{-x}.$$

Example 11.13. Let X, Y be a random variable with the joint PDF

$$f_{XY}(x,y) = 6e^{-2x}e^{-3y} \quad 0 < x < \infty, 0 < y < \infty.$$

Are X, Y independent?

Using (11.1) we can find f_X and f_Y and see if f_{XY} factors, that is, $f_{XY} = f_X f_Y$. First

$$f_X(x) = \int_0^\infty 6e^{-2x}e^{-3y}dy = 2e^{-2x},$$
$$f_Y(y) = \int_0^\infty 6e^{-2x}e^{-3y}dx = 3e^{-3y}.$$

which are both exponential. Thus $f_{XY} = f_X f_Y$, therefore yes, X and Y are independent!

Example 11.14. Let X, Y have the joint PDF

$$f_{X,Y}(x,y) = x + y, \quad 0 < x < 1, 0 < y < 1$$

Are X, Y independent?

Note that there is no way to factor $x + y = f_X(x)f_Y(y)$, hence they can not be independent.

11.6. More than two random variables

All the concepts and some of the techniques we introduced for two random variables can be extended to more than two random variables. For example, we can define joint PMF, PDF, CDF, independence for three or more random variables. While many of the explicit expressions can be less tractable, the case of normal variables is tractable. First we comment on independence of several random variables.

Recall that we distinguished between jointly independent events (Definition 3.1) and pairwise independent events.

Definition

A set of n random variables $\{X_1, \ldots, X_n\}$ is *pairwise independent* if every pair of random variables is independent.

As for events, if the set of random variables is pairwise independent, it is not necessarily mutually independent as defined next.

Definition

A set of n random variables $\{X_1, \ldots, X_n\}$ is *mutually independent* if for any sequence of numbers $\{x_1, \ldots, x_n\}$, the events $\{X_1 \leqslant x_1\}, \ldots, \{X_n \leqslant x_n\}$ are mutually independent events.

This definition is equivalent to the following condition on the joint cumulative distribution function $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$. Namely, a set of n random variables $\{X_1, \ldots, X_n\}$ is mutually independent if

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n) \tag{11.6.1}$$

for all $x_1, \ldots, x_n$. Note that we do not need require that the probability distribution factorizes for all possible subsets as in the case for n events. This is not required because Equation 11.6.1 implies factorization for any subset of $1, ..., n$.

The following statement will be easier to prove later once we have the appropriate mathematical tools.

Proposition 11.2

If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ are *independent* for $1 \leqslant i \leqslant n$ then

$$X_1 + \cdots + X_n \sim \mathcal{N}\left(\mu_1 + \cdots \mu_n, \sigma_1^2 + \cdots + \sigma_n^2\right).$$

In particular if $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ then $X + Y \sim \mathcal{N}\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\right)$ and $X - Y \sim \mathcal{N}\left(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2\right)$. In general for two independent Gaussian X and Y we have $cX + dY \sim \mathcal{N}\left(c\mu_x + d\mu_y, c\sigma_x^2 + d\sigma_y^2\right)$.

Example 11.15. Suppose $T \sim \mathcal{N}(95, 25)$ and $H \sim \mathcal{N}(65, 36)$ represents the grades of T. and H. in their Probability class.

(a) What is the probability that their average grades will be less than 90?
(b) What is the probability that H. will have scored higher than T.?
(c) Answer question (b) if $T \sim \mathcal{N}(90, 64)$ and $H \sim \mathcal{N}(70, 225)$.

In (a) we can use Proposition 11.2 $T+H \sim \mathcal{N}(160, 61)$. Thus

$$\begin{aligned}\mathbb{P}\left(\frac{T+H}{2} \leqslant 90\right) &= \mathbb{P}(T+H \leqslant 180) \\ &= \mathbb{P}\left(Z \leqslant \frac{180-160}{\sqrt{61}}\right) = \Phi\left(\frac{180-160}{\sqrt{61}}\right) \\ &\approx \Phi(2.56) \approx 0.9961\end{aligned}$$

For (b) we can use $H-T \sim \mathcal{N}(-30, 61)$ to find

$$\begin{aligned}\mathbb{P}(H>T) &= \mathbb{P}(H-T>0) \\ &= 1-\mathbb{P}(H-T<0) \\ &= 1-\mathbb{P}\left(Z \leq \frac{0-(-30)}{\sqrt{61}}\right) \\ &\approx 1-\Phi(3.84) \approx 0.00006.\end{aligned}$$

In (c) we can use Proposition 11.2 to see that $T-H \sim \mathcal{N}(-20, 289)$ and so

$$\begin{aligned}\mathbb{P}(H>T) &= \mathbb{P}(H-T>0) \\ &= 1-\mathbb{P}(H-T<0) \\ &= 1-\mathbb{P}\left(Z \leqslant \frac{0-(-20)}{17}\right) \\ &\approx 1-\Phi(1.18) \approx 0.11900\end{aligned}$$

Example 11.16. Suppose X_1, X_2 have joint distribution

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} x_1 + \frac{3}{2}(x_2)^2 & 0 \leqslant x_1 \leqslant 1, 0 \leqslant x_2 \leqslant 1 \\ 0 & \text{otherwise} \end{cases}.$$

Find the joint PDF of $Y_1 = X_1 + X_2$ and $Y_2 = X_1^2$.

We use Theorem 11.4 here.

First we find the Jacobian of the following map

$$\begin{aligned} y_1 &= g_1(x_1, x_2) = x_1 + x_2, \\ y_2 &= g_2(x_1, x_2) = x_1^2. \end{aligned}$$

So

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 2x_1 & 0 \end{vmatrix} = -2x_1$$

Secondly, we need to invert the map g, that is, solve for x_1, x_2 in terms of y_1, y_2

$$\begin{aligned} x_1 &= \sqrt{y_2}, \\ x_2 &= y_1 - \sqrt{y_2}. \end{aligned}$$

The joint PDF of Y_1, Y_2 then is given by

$$
\begin{aligned}
f_{Y_1,Y_2}\left(y_1,y_2\right) &= f_{X_1,X_2}\left(x_1,x_2\right)\left|J\left(x_1,x_2\right)\right|^{-1} \\
&= f_{X_1,X_2}\left(\sqrt{y_2}, y_1-\sqrt{y_2}\right)\frac{1}{2x_1} \\
&= \begin{cases} \frac{1}{2\sqrt{y_2}}\left[\left(\sqrt{y_2}+\frac{3}{2}\left(y_1-\sqrt{y_2}\right)^2\right), & 0 \leqslant y_2 \leqslant 1,\ 0 \leqslant y_1-\sqrt{y_2} \leqslant 1 \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

11.7. Exercises

Exercise 11.1. Suppose that 2 balls are chosen without replacement from an urn consisting of 5 white and 8 red balls. Let X equal 1 if the first ball selected is white and zero otherwise. Let Y equal 1 if the second ball selected is white and zero otherwise.

(A) Find the probability mass function of X, Y.
(B) Find $\mathbb{E}(XY)$.
(C) Is it true that $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$?
(D) Are X, Y independent?

Exercise 11.2. Suppose you roll two fair dice. Find the probability mass function of X and Y, where X is the largest value obtained on any die, and Y is the sum of the values.

Exercise 11.3. Suppose the joint density function of X and Y is $f(x,y) = \frac{1}{4}$ for $0 < x < 2$ and $0 < y < 2$.

(A) Find $\mathbb{P}\left(\frac{1}{2} < X < 1, \frac{2}{3} < Y < \frac{4}{3}\right)$.
(B) Find $\mathbb{P}(XY < 2)$.
(C) Find the marginal distributions $f_X(x)$ and $f_Y(y)$.

Exercise 11.4. The joint probability density function of X and Y is given by

$$f(x,y) = e^{-(x+y)}, 0 \leqslant x < \infty, 0 \leqslant y < \infty.$$

Find $\mathbb{P}(X < Y)$.

Exercise 11.5. Suppose X and Y are independent random variables and that X is exponential with $\lambda = \frac{1}{4}$ and Y is uniform on $(2,5)$. Calculate the probability that $2X + Y < 8$.

Exercise 11.6. Consider X and Y given by the joint density

$$f(x,y) = \begin{cases} 10x^2y & 0 \leqslant y \leqslant x \leqslant 1 \\ 0 & \text{otherwise .} \end{cases}$$

(A) Find the marginal PDFs, $f_X(x)$ and $f_Y(x)$.
(B) Are X and Y independent random variables?
(C) Find $\mathbb{P}\left(Y \leqslant \frac{X}{2}\right)$.
(D) Find $\mathbb{P}\left(Y \leqslant \frac{X}{4} \mid Y \leqslant \frac{X}{2}\right)$.
(E) Find $\mathbb{E}[X]$.

Exercise 11.7. Consider X and Y given by the joint density

$$f(x,y) = \begin{cases} 4xy & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise.} \end{cases}$$

(A) Find the marginal PDFs, f_X and f_Y.
(B) Are X and Y independent?
(C) Find $\mathbb{E}Y$.

Exercise 11.8. Consider X, Y given by the joint PDF

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y) & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent random variables?

Exercise 11.9. Suppose that gross weekly ticket sales for UConn basketball games are normally distributed with mean \$2,200,000 and standard deviation \$230,000. What is the probability that the total gross ticket sales over the next two weeks exceeds \$4,600,000?

Exercise 11.10. Suppose the joint density function of the random variable X_1 and X_2 is

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y_1 = 2X_1 + X_2$ and $Y_2 = X_1 - 3X_2$. What is the joint density function of Y_1 and Y_2?

Exercise 11.11. Suppose the joint density function of the random variable X_1 and X_2 is

$$f(x_1, x_2) = \begin{cases} \frac{3}{2}(x_1^2 + x_2^2) & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y_1 = X_1 - 2x_2$ and $Y_2 = 2X_1 + 3X_2$. What is the joint density function of Y_1 and Y_2?

Exercise 11.12. We roll two dice. Let X be the minimum of the two numbers that appear, and let Y be the maximum. Find the joint probability mass function of (X, Y), that is, $P(X = i, Y = j)$:

$i \backslash j$	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Find the marginal probability mass functions of X and Y. Finally, find the conditional probability mass function of X given that $Y = 5$, that is, $\mathbb{P}(X = i | Y = 5)$, for $i = 1, \ldots, 6$.

Exercise* 11.1. Let X and Y be independent exponential random variables with parameters λ_X and λ_Y respectively. Find the cumulative distribution function and the probability density function of $U = X/Y$.

Exercise* 11.2. Let X and Y be random variables uniformly distributed in the triangle $\{(x, y) : x > 0, y > 0, x + y < 1\}$. Find the cumulative distribution function and the probability density function of $U = X/Y$.

11.8. Selected solutions

Solution to Exercise 11.1(A): We have

$$\begin{aligned}
p(0,0) &= \mathbb{P}(X=0,Y=0) = \mathbb{P}(RR) = \frac{8\cdot 7}{13\cdot 12} = \frac{14}{39},\\
p(1.0) &= \mathbb{P}(X=1,Y=0) = \mathbb{P}(WR) = \frac{5\cdot 8}{13\cdot 12} = \frac{10}{39},\\
p(0,1) &= \mathbb{P}(X=0,Y=1) = \mathbb{P}(RW) = \frac{8\cdot 5}{13\cdot 12} = \frac{10}{39},\\
p(1,1) &= \mathbb{P}(X=1,Y=1) = \mathbb{P}(WW) = \frac{5\cdot 4}{13\cdot 12} = \frac{5}{39}.
\end{aligned}$$

Solution to Exercise 11.1(B):

$$\mathbb{E}(XY) = \mathbb{P}(X=1,Y=1) = \frac{5}{39} \approx 0.1282$$

Solution to Exercise 11.1(C): Not true because

$$(\mathbb{E}X)(\mathbb{E}Y) = \mathbb{P}(X=1)\,\mathbb{P}(Y=1) = \left(\frac{5}{13}\right)^2 = \frac{25}{169} \approx 0.1479$$

Solution to Exercise 11.1(D): X and Y are not independent because

$$\mathbb{P}(X=1,Y=1) = \frac{5}{39} \neq \mathbb{P}(X=1)\,\mathbb{P}(Y=1) = \left(\frac{5}{13}\right)^2.$$

Solution to Exercise 11.2: First we need to figure out what values X, Y can attain. Note that X can be any of $1,2,3,4,5,6$, but Y is the sum and can only be as low as 2 and as high as 12. First we make a table of all possibilities for (X,Y) given the values of the dice. Recall X is the largest of the two, and Y is the sum of them. The possible outcomes are given by the following.

1st die\2nd die	1	2	3	4	5	6
1	$(1,2)$	$(2,3)$	$(3,4)$	$(4,5)$	$(5,6)$	$(6,7)$
2	$(2,3)$	$(2,4)$	$(3,5)$	$(4,6)$	$(5,7)$	$(6,8)$
3	$(3,4)$	$(3,5)$	$(3,6)$	$(4,7)$	$(5,8)$	$(6,9)$
4	$(4,5)$	$(4,6)$	$(4,7)$	$(4,8)$	$(5,9)$	$(6,10)$
5	$(5,6)$	$(5,7)$	$(5,8)$	$(5,9)$	$(5,10)$	$(6,11)$
6	$(6,7)$	$(6,8)$	$(6,9)$	$(6,10)$	$(6,11)$	$(6,12)$

Then we make a table of the PMF $p(x,y)$.

$X\backslash Y$	2	3	4	5	6	7	8	9	10	11	12
1	$\frac{1}{36}$	0	0	0	0	0	0	0	0	0	0
2	0	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0	0	0	0	0
3	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0	0	0
4	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0
5	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0
6	0	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Solution to Exercise 11.3(A): We integrate the PDF over $0 < x < 2$ and $0 < y < 2$ and get

$$\int_{\frac{1}{2}}^{1} \int_{\frac{2}{3}}^{\frac{4}{3}} \frac{1}{4} dy dx = \frac{1}{4}\left(1 - \frac{1}{2}\right)\left(\frac{4}{3} - \frac{2}{3}\right) = \frac{1}{12}.$$

Solution to Exercise 11.3(B): We need to find the region that is within $0 < x, y < 2$ and $y < \frac{2}{x}$. (Try to draw the region) We get two regions from this. One with bounds $0 < x < 1, 0 < y < 2$ and the region $1 < x < 2, 0 < y < \frac{2}{x}$. Then

$$\begin{aligned}
\mathbb{P}(XY < 2) &= \int_0^1 \int_0^2 \frac{1}{4} dy dx + \int_1^2 \int_0^{\frac{2}{x}} \frac{1}{4} dy dx \\
&= \frac{1}{2} + \int_1^2 \frac{1}{2x} dx \\
&= \frac{1}{2} + \frac{\ln 2}{2}.
\end{aligned}$$

Solution to Exercise 11.3(C): Recall that

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy = \int_0^2 \frac{1}{4} dy = \frac{1}{2}$$

for $0 < x < 2$ and 0 otherwise. By symmetry, f_Y is the same.

Solution to Exercise 11.4: Draw a picture of the region and note the integral needs to be set up in the following way:

$$\begin{aligned}
\mathbb{P}(X < Y) &= \int_0^{\infty} \int_0^{y} e^{-(x+y)} dx dy = \int_0^{\infty} \left[-e^{-2y} + e^{-y}\right] dy \\
&= \left(\frac{1}{2} e^{-2y} - e^{-y}\right)\Big|_0^{\infty} = 0 - \left(\frac{1}{2} - 1\right) = \frac{1}{2}.
\end{aligned}$$

Solution to Exercise 11.5: We know that

$$f_X(x) = \begin{cases} \frac{1}{4}e^{-\frac{x}{4}}, & \text{when } x \geqslant 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_Y(y) = \begin{cases} \frac{1}{3}, & \text{when } 2 < y < 5 \\ 0, & \text{otherwise} \end{cases}$$

Since X, Y are independent then $f_{X,Y} = f_X f_Y$, thus

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{12}e^{-\frac{x}{4}} & \text{when } x \geqslant 0, 2 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

Draw the region $(2X + Y < 8)$, which correspond to $0 \leqslant x$, $0 < y < 5$ and $y < 8 - 2x$. Drawing a picture of the region, we get the corresponding bounds of $2 < y < 5$ and $0 < x < 4 - \frac{y}{2}$, so that

$$\begin{aligned} \mathbb{P}(2X + Y < 8) &= \int_2^5 \int_0^{4-\frac{y}{2}} \frac{1}{12} e^{-\frac{x}{4}} dx dy \\ &= \int_2^5 \frac{1}{3}\left(1 - e^{y/8-1}\right) dy \\ &= 1 - \frac{8}{3}\left(e^{-\frac{3}{8}} - e^{-\frac{3}{4}}\right) \end{aligned}$$

Solution to Exercise 11.6(A): We have

$$f_X(x) = \begin{cases} 5x^4, & 0 \leqslant x \leqslant 1 \\ 0, & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{10}{3}y(1-y^3), & 0 \leqslant y \leqslant 1 \\ 0, & \text{otherwise} \end{cases}.$$

Solution to Exercise 11.6(B): No, since $f_{X,Y} \neq f_X f_Y$.

Solution to Exercise 11.6(C): $\mathbb{P}\left(Y \leqslant \frac{X}{2}\right) = \frac{1}{4}$.

Solution to Exercise 11.6(D): Also $\frac{1}{4}$.

Solution to Exercise 11.6(E): Use f_X and the definition of expected value, which is $5/6$.

Solution to Exercise 11.7(A): $f_X = 2x$ and $f_Y = 2y$.

Solution to Exercise 11.7(B): Yes! Since $f(x,y) = f_X f_Y$.

Solution to Exercise 11.7(C): We have $\mathbb{E}Y = \int_0^1 y \cdot 2y dy = \frac{2}{3}$.

Solution to Exercise 11.8: We get $f_X = \left(\frac{2}{3}x + \frac{2}{3}\right)$ while $f_Y = \frac{1}{3} + \frac{4}{3}y$ and $f \neq f_X f_Y$.

Solution to Exercise 11.9: If $W = X_1 + X_2$ is the sales over the next two weeks, then W is normal with mean $2,200,000 + 2,200,000 = 4,400,00$ and variance $230,000^2 + 230,000^2$.

Thus the variance is $\sqrt{230,000^2 + 230,000^2} = 325,269.1193$. Hence

$$\begin{aligned}\mathbb{P}(W > 5,000,000) &= \mathbb{P}\left(Z > \frac{4,600,000 - 4,400,000}{325,269.1193}\right)\\ &= \mathbb{P}(Z > 0.6149)\\ &\approx 1 - \Phi(0.61) = 0.27.\end{aligned}$$

Solution to Exercise 11.10: First find the Jacobian:

$$\begin{aligned}y_1 &= g_1(x_1, x_2) = 2x_1 + x_2,\\ y_2 &= g_2(x_1, x_2) = x_1 - 3x_2.\end{aligned}$$

Thus

$$J(x_1, x_2) = \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} = -7.$$

Solve for x_1, x_2 and get

$$\begin{aligned}x_1 &= \frac{3}{7}y_1 + \frac{1}{7}y_2\\ x_2 &= \frac{1}{7}y_1 - \frac{2}{7}y_2\end{aligned}$$

The joint PDF for Y_1, Y_2 is given by

$$\begin{aligned}f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(x_1, x_2)\,|J(x_1, x_2)|^{-1}\\ &= f_{X_1,X_2}\left(\frac{3}{7}y_1 + \frac{1}{7}y_2, \frac{1}{7}y_1 - \frac{2}{7}y_2\right)\frac{1}{7}.\end{aligned}$$

Since we are given the joint PDF of X_1 and X_2, then plugging them into f_{X_1,X_2}, we have

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{4}{7^3}(3y_1 + y_2)(y_1 - 2y_2) & 0 < 3y_1 + y_2 < 7, 0 < y_1 - 2y_2 < 2 \\ 0 & \text{otherwise.} \end{cases}$$

Solution to Exercise 11.11:

First find the Jacobian

$$\begin{aligned}y_1 &= g_1(x_1, x_2) = x_1 - 2x_2,\\ y_2 &= g_2(x_1, x_2) = 2x_1 + 3x_2.\end{aligned}$$

This

$$J(x_1, x_2) = \begin{vmatrix} 1 & -2 \\ 2 & 3 \end{vmatrix} = 7.$$

Now solve for x_1, x_2 and get

$$x_1 = \frac{1}{7}(3y_1 + 2y_2)$$
$$x_2 = \frac{1}{7}(-2y_1 + y_2)$$

The joint PDF for Y_1, Y_2 is given by

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(x_1, x_2)\left|J(x_1, x_2)\right|^{-1} \\ &= f_{X_1,X_2}\left(\frac{1}{7}(3y_1 + 2y_2), \frac{1}{7}(-2y_1 + y_2)\right)\frac{1}{7}. \end{aligned}$$

Since we are given the joint PDFs of X_1 and X_2, then plugging them into f_{X_1,X_2}, we have

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{3}{2} \cdot \frac{1}{7^3}\left((3y_1 + 2y_2)^2 + (-2y_1 + y_2)^2\right), & 0 < 3y_1 + 2y_2 < 7, 0 < -2y_1 + y_2 < 7 \\ 0, & \text{otherwise.} \end{cases}$$

Solution to Exercise 11.12: The joint probability mass function $\mathbb{P}(X = i, Y = j)$ of (X, Y) is

$i \backslash j$	1	2	3	4	5	6
1	1/36	1/18	1/18	1/18	1/18	1/18
2	0	1/36	1/18	1/18	1/18	1/18
3	0	0	1/36	1/18	1/18	1/18
4	0	0	0	1/36	1/18	1/18
5	0	0	0	0	1/36	1/18
6	0	0	0	0	0	1/36

The marginal probability mass functions $\mathbb{P}(X = i)$ of X is given by

i	1	2	3	4	5	6
$P(X = i)$	1/36	1/12	5/36	7/36	1/4	11/36

The marginal probability mass functions of Y is given by

i	1	2	3	4	5	6
$P(Y = i)$	11/36	1/4	7/36	5/36	1/12	1/36

The conditional probability mass function of X given that Y is 5, $\mathbb{P}(X = i|Y = 5)$, for $i = 1, \dots, 6$

i	1	2	3	4	5	6
$P(X = i\|Y = 5)$	2/9	2/9	2/9	2/9	1/9	0

Solution to Exercise 11.1*: If $u < 0$ then $F_U(u) = 0$ and $f_U(u) = 0$.

If $u > 0$ then we have that

$$F_U(u) = \mathbb{P}(X < uY) = 1 - \lambda_X\lambda_Y \int_0^\infty \int_{uy}^\infty e^{-\lambda_X x - \lambda_Y y} dx dy = 1 - \frac{\lambda_Y}{\lambda_X u + \lambda_Y} = \frac{\lambda_X u}{\lambda_X u + \lambda_Y}.$$

By differentiating with respect to u we obtain $f_U(u) = \frac{\lambda_Y \lambda_X}{(\lambda_X u + \lambda_Y)^2}$.

The second solutions to Exercise 11.1:* Introduce a variable $V = Y$ so that $X = UV$, and compute the Jacobian

$$J_h(u, v) = \det \begin{pmatrix} v & u \\ 0 & 1 \end{pmatrix} = v.$$

This implies that

$$f_U(u) = \lambda_X\lambda_Y \int_0^\infty v\, e^{-\lambda_X uv - \lambda_Y v}\, dv = \frac{\lambda_X\lambda_Y}{(\lambda_X u + \lambda_Y)^2}$$

when $u > 0$. By computing the antiderivative we obtain $F_U(u) = 1 - \frac{\lambda_Y}{\lambda_X u + \lambda_Y} = \frac{\lambda_X u}{\lambda_X u + \lambda_Y}$.

The third solution to Exercise 11.1:* Introduce a variable $V = X$ so that $Y = V/U$, and compute the Jacobian

$$J_h(u, v) = \det \begin{pmatrix} 0 & 1 \\ -v/u^2 & 1/u \end{pmatrix} = v/u^2.$$

This implies that

$$f_U(u) = \frac{\lambda_X\lambda_Y}{u^2} \int_0^\infty v\, e^{-\lambda_X v - \lambda_Y v/u}\, dv = \frac{\lambda_X\lambda_Y}{(\lambda_X + \lambda_Y/u)^2 u^2} = \frac{\lambda_X\lambda_Y}{(\lambda_X u + \lambda_Y)^2}$$

when $u > 0$. By computing the antiderivative we obtain $F_U(u) = 1 - \frac{\lambda_Y}{\lambda_X u + \lambda_Y} = \frac{\lambda_X u}{\lambda_X u + \lambda_Y}$.

Answers to Exercise 11.2*:

If $u < 0$ then $F_U(u) = 0$ and $f_U(u) = 0$. If $u > 0$ then $F_U(u) = \dfrac{u}{1 + u}$ and $f_U(u) = \dfrac{1}{(1 + u)^2}$.

CHAPTER 12

Expectations

12.1. Expectation and covariance for multivariate distributions

As we discussed earlier for two random variables X and Y and a function $g(x, y)$ we can consider $g(X, Y)$ as a random variable, and therefore

$$\mathbb{E}g(X,Y) = \sum\sum_{x,y} g(x,y)p(x,y) \tag{12.1.1}$$

in the discrete case and

$$\mathbb{E}g(X,Y) = \iint g(x,y)f(x,y)dx\,dy \tag{12.1.2}$$

in the continuous case.

In particular for $g(x, y) = x + y$ we have

$$\begin{aligned}\mathbb{E}(X+Y) &= \iint (x+y)f(x,y)dx\,dy \\ &= \iint xf(x,y)dx\,dy + \iint yf(x,y)dx\,dy.\end{aligned}$$

If we now set $g(x, y) = x$, we see the first integral on the right is $\mathbb{E}X$, and similarly the second is $\mathbb{E}Y$. Therefore

$$\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y.$$

Proposition 12.1

If X and Y are two *independent* random variables, then for any functions h and k we have

$$\mathbb{E}[h(X)k(Y)] = \mathbb{E}h(X) \cdot \mathbb{E}k(Y).$$

In particular, $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$.

Proof. We only prove the statement in the case X and Y are jointly continuous, the case of discrete random variables is left as Exercise* 12.4. By Equation (12.1.2) with $g(x, y) = h(x)k(y)$, and recalling that the joint density function factors by independence of X and Y

as we saw in (11.1), we have

$$\begin{aligned}
\mathbb{E}[h(X)k(Y)] &= \iint h(x)k(y)f_{XY}(x,y)dx\,dy \\
&= \iint h(x)k(y)f_X(x)f_Y(y)dx\,dy \\
&= \int h(x)f_X(x)\int k(y)f_Y(y)dy\,dx \\
&= \int h(x)f_X(x)(\mathbb{E}k(Y))dx \\
&= \mathbb{E}h(X)\cdot\mathbb{E}k(Y).
\end{aligned}$$

□

Note that we can easily extend Proposition 12.1 to any number of *independent* random variables.

Definition 12.1

The *covariance* of two random variables X and Y is defined by

$$\mathrm{Cov}\,(X,Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)].$$

As with the variance, $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$. It follows that if X and Y are independent, then $\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y)$, and then $\mathrm{Cov}(X,Y) = 0$.

Proposition 12.2

Suppose X, Y and Z are random variables and a and c are constants. Then
(1) $\mathrm{Cov}\,(X,X) = \mathrm{Var}\,(X)$.
(2) if X and Y are independent, then $\mathrm{Cov}\,(X,Y) = 0$.
(3) $\mathrm{Cov}\,(X,Y) = \mathrm{Cov}\,(Y,X)$.
(4) $\mathrm{Cov}\,(aX,Y) = a\,\mathrm{Cov}\,(X,Y)$.
(5) $\mathrm{Cov}\,(X+c,Y) = \mathrm{Cov}\,(X,Y)$.
(6) $\mathrm{Cov}\,(X+Y,Z) = \mathrm{Cov}\,(X,Z) + \mathrm{Cov}\,(Y,Z)$.
More generally,

$$\mathrm{Cov}\left(\sum_{i=1}^m a_iX_i, \sum_{j=1}^n b_jY_j\right) = \sum_{i=1}^m\sum_{j=1}^n a_ib_j\,\mathrm{Cov}\,(X_i,Y_j)\,.$$

Note

$$\begin{aligned}
&\mathrm{Var}(aX+bY) \\
&\quad= \mathbb{E}[((aX+bY)-\mathbb{E}(aX+bY))^2] \\
&\quad= \mathbb{E}[(a(X-\mathbb{E}X)+b(Y-\mathbb{E}Y))^2] \\
&\quad= \mathbb{E}[a^2(X-\mathbb{E}X)^2 + 2ab(X-\mathbb{E}X)(Y-\mathbb{E}Y) + b^2(Y-\mathbb{E}Y)^2] \\
&\quad= a^2\,\mathrm{Var}\,X + 2ab\,\mathrm{Cov}(X,Y) + b^2\,\mathrm{Var}\,Y.
\end{aligned}$$

We have the following corollary.

Proposition 12.3

If X and Y are independent, then

$$\operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y.$$

Proof. We have

$$\operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y + 2\operatorname{Cov}(X,Y) = \operatorname{Var} X + \operatorname{Var} Y.$$

□

Example 12.1. Recall that a binomial random variable is the sum of n independent Bernoulli random variables with parameter p. Consider the *sample mean*

$$\overline{X} := \sum_{i=1}^{n} \frac{X_i}{n},$$

where all $\{X_i\}_{i=1}^{\infty}$ are independent and have the same distribution, then $\mathbb{E}\overline{X} = \mathbb{E}X_1 = p$ and $\operatorname{Var}\overline{X} = \operatorname{Var} X_1/n = p\,(1-p)$.

12.2. Conditional expectation

Recall also that in Section 11.3 we considered conditional random variables $X \mid Y = y$. We can define its expectation as follows.

Definition (Conditional expectation)

The *conditional expectation* of X given Y is defined by

$$\mathbb{E}[X \mid Y = y] = \sum_{x} x f_{X|Y=y}(x)$$

for discrete random variables X and Y, and by

$$\mathbb{E}[X \mid Y = y] = \int x f_{X|Y=y}(x)dx$$

for continuous random variables X and Y

Here the conditional density is defined by Equation (11.3) in Section 11.3. We can think of $\mathbb{E}[X \mid Y = y]$ is the mean value of X, when Y is fixed at y. Note that unlike the expectation of a random variable which is a number, the conditional expectation, $\mathbb{E}[X \mid Y = y]$, is a *random variable* with randomness inherited from Y, not X.

12.3. Further examples and applications

12.3.1. Expectation and variance.

Example 12.2. Suppose the joint PMF of X and Y is given by

$X \diagdown Y$	0	1
0	0.2	0.7
1	0	0.1

.

To find $\mathbb{E}[XY]$ we can use the formula

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{i,j} x_i y_j p(x_i, y_j) \\ &= 0 \cdot 0 \cdot p(0,0) + 1 \cdot 0 \cdot p(1,0) + 0 \cdot 1 \cdot p(0,1) + 1 \cdot 1 \cdot p(1,1) \\ &= 0.1\end{aligned}$$

Example 12.3. Suppose X, Y are independent exponential random variables with parameter $\lambda = 1$. Set up a double integral that represents

$$\mathbb{E}\left[X^2 Y\right].$$

Since X, Y are independent then $f_{X,Y}$ factorizes

$$f_{X,Y}(x,y) = e^{-1\cdot x} e^{-1\cdot y} = e^{-(x+y)}, 0 < x, y < \infty.$$

Thus

$$\mathbb{E}\left[X^2 Y\right] = \int_0^\infty \int_0^\infty x^2 y e^{-(x+y)} dy dx.$$

Example 12.4. Suppose the joint PDF of X, Y is

$$f(x,y) = \begin{cases} 10xy^2 & 0 < x < y, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Find $\mathbb{E}XY$ and $\operatorname{Var}(Y)$.

First we draw the region of integration and then set up the integral

$$\begin{aligned}\mathbb{E}XY &= \int_0^1 \int_0^y xy\left(10xy^2\right) dx dy = 10 \int_0^1 \int_0^y x^2 y^3 dx dy \\ &= \frac{10}{3} \int_0^1 y^3 y^3 dy = \frac{10}{3}\frac{1}{7} = \frac{10}{21}.\end{aligned}$$

First note that $\operatorname{Var}(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2$. Then

$$\begin{aligned}\mathbb{E}Y^2 &= \int_0^1 \int_0^y y^2 \left(10xy^2\right) dxdy = 10 \int_0^1 \int_0^y y^4 x dxdy \\ &= 5 \int_0^1 y^4 y^2 dy = \frac{5}{7}.\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}Y &= \int_0^1 \int_0^y y \left(10xy^2\right) dxdy = 10 \int_0^1 \int_0^y y^3 x dxdy \\ &= 5 \int_0^1 y^3 y^2 dy = \frac{5}{6}.\end{aligned}$$

Therefore $\operatorname{Var}(Y) = \frac{5}{7} - \left(\frac{5}{6}\right)^2 = \frac{5}{252}$.

12.3.2. Correlation. We start with the following definition.

Definition (Correlation)

The *correlation coefficient* of X and Y is defined by

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

In addition, we say that X and Y are

positively correlated, if $\rho(X,Y) > 0$,
negatively correlated, if $\rho(X,Y) < 0$,
uncorrelated, if $\rho(X,Y) = 0$.

Proposition (Properties of the correlation coefficient)

Suppose X and Y are random variables. Then
(1) $\rho(X,Y) = \rho(Y,X)$.
(2) If X and Y are independent, then $\rho(X,Y) = 0$. The converse is not true in general.
(3) If $\rho(aX+b, cY+d) = \rho(X,Y)$ if $a, c > 0$.
(4) $-1 \leqslant \rho(X,Y) \leqslant 1$.
(5) If $\rho(X,Y) = 1$, then $Y = aX + b$ for some $a > 0$.
(6) If $\rho(X,Y) = -1$, then $Y = aX + b$ for some $a < 0$.

Proof. We only prove (4). First observe that for any random variables X and Y we can define the normalized random variables as follows.

$$\begin{aligned}U &:= \frac{X - \mathbb{E}X}{\sqrt{\operatorname{Var} X}}, \\ V &:= \frac{Y - \mathbb{E}Y}{\sqrt{\operatorname{Var} Y}}.\end{aligned}$$

Then $\mathbb{E}U = \mathbb{E}V = 0, \mathbb{E}U^2 = \mathbb{E}V^2 = 1$.

Note that we can use (3) since $U = aX + b$, where $a = \frac{1}{\sqrt{\operatorname{Var} X}} > 0$, $b = -\frac{\mathbb{E}X}{\sqrt{\operatorname{Var} X}}$, and similarly $V = cY + d$, where $c = \frac{1}{\sqrt{\operatorname{Var} Y}} > 0$, $d = -\frac{\mathbb{E}Y}{\sqrt{\operatorname{Var} Y}}$, so

$$\rho(U, V) \overset{(3)}{=} \rho(X, Y).$$

Therefore

$$\rho(X, Y) = \rho(U, V) = \frac{\operatorname{Cov}(U, V)}{\sqrt{\operatorname{Var}(U)\operatorname{Var}(V)}} = \operatorname{Cov}(U, V).$$

Note that

$$\operatorname{Cov}(U, V) = \mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V) = \mathbb{E}(UV).$$

Therefore

$$\rho(X, Y) = \mathbb{E}(UV) \tag{12.3.1}$$

To finish the proof we need an inequality

$$ab \leqslant \frac{a^2 + b^2}{2}$$

for any real numbers a and b. This follows from the fact that

$$(a - b)^2 \geqslant 0.$$

Moreover, $ab = \frac{a^2+b^2}{2}$ **if and only if** $a = b$.

Applying this to $a = U$ and $b = V$ to Equation (12.3.1) we see that

$$\rho(X, Y) = \mathbb{E}(UV) \leqslant \mathbb{E}\left(\frac{U^2}{2} + \frac{V^2}{2}\right) = 1.$$

Note that is true also for $-X$ and Y, that is,

$$\rho(-X, Y) = \frac{\operatorname{Cov}(-X, Y)}{\sqrt{\operatorname{Var}(-X)\operatorname{Var}(Y)}} = \frac{-\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = -\rho(X, Y) \leqslant 1.$$

Thus

$$\rho(X, Y) \geqslant -1.$$

□

Example 12.5. Suppose X, Y are random variables whose joint PDF is given by

$$f(x, y) = \begin{cases} \frac{1}{y} & 0 < y < 1, 0 < x < y \\ 0 & \text{otherwise} \end{cases}.$$

(a) Find the covariance of X and Y.
(b) Find $\operatorname{Var}(X)$ and $\operatorname{Var}(Y)$.
(c) Find $\rho(X, Y)$.

We start by finding expectations.

(a) Recall that $\operatorname{Cov}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$. So

$$\begin{aligned}
\mathbb{E}XY &= \int_0^1 \int_0^y xy\frac{1}{y}dxdy = \int_0^1 \frac{y^2}{2}dy = \frac{1}{6}, \\
\mathbb{E}X &= \int_0^1 \int_0^y x\frac{1}{y}dxdy = \int_0^1 \frac{y}{2}dy = \frac{1}{4}, \\
\mathbb{E}Y &= \int_0^1 \int_0^y y\frac{1}{y}dxdy = \int_0^1 ydy = \frac{1}{2}.
\end{aligned}$$

Thus

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y \\
&= \frac{1}{6} - \frac{1}{4}\frac{1}{2} \\
&= \frac{1}{24}.
\end{aligned}$$

(b) We have that

$$\begin{aligned}
\mathbb{E}X^2 &= \int_0^1 \int_0^y x^2\frac{1}{y}dxdy = \int_0^1 \frac{y^2}{3}dy = \frac{1}{9}, \\
\mathbb{E}Y^2 &= \int_0^1 \int_0^y y^2\frac{1}{y}dxdy = \int_0^1 y^2dy = \frac{1}{3}.
\end{aligned}$$

Thus recall that

$$\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \\
&= \frac{1}{9} - \left(\frac{1}{4}\right)^2 = \frac{7}{144}.
\end{aligned}$$

Also

$$\begin{aligned}
\operatorname{Var}(Y) &= \mathbb{E}Y^2 - (\mathbb{E}Y)^2 \\
&= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.
\end{aligned}$$

(c)

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\frac{1}{24}}{\sqrt{\left(\frac{7}{144}\right)\left(\frac{1}{12}\right)}} \approx 0.6547.$$

12.3.3. Conditional expectation: examples. We start with properties of conditional expectations which we introduced in Definition 12.2 for discrete and continuous random variables. We skip the proof of this as it is very similar to the case of (unconditional) expectation.

Proposition 12.4

For the conditional expectation of X given $Y = y$ it holds that

(i) for any $a, b \in \mathbb{R}$, $\mathbb{E}[aX + b \mid Y = y] = a\mathbb{E}[X \mid Y = y] + b$.

(ii) $\mathrm{Var}(X \mid Y = y) = \mathbb{E}[X^2 \mid Y = y] - (\mathbb{E}[X \mid Y = y])^2$.

Example 12.6. Let X and Y be random variables with the joint PDF

$$f_{XY}(x, y) = \begin{cases} \frac{1}{18} e^{-\frac{x+y}{6}} & \text{if } 0 < y < x, \\ 0 & \text{otherwise.} \end{cases}$$

In order to find $\mathrm{Var}(X \mid Y = 2)$, we need to compute the conditional PDF of X given $Y = 2$, i.e.

$$f_{X|Y=2}(x \mid 2) = \frac{f_{XY}(x, 2)}{f_Y(2)}.$$

To this purpose, we compute first the marginal of Y.

$$f_Y(y) = \int_y^\infty \frac{1}{18} e^{-\frac{x+y}{6}} dx = \frac{1}{3} e^{-\frac{y}{6}} \left. -e^{-\frac{y}{6}} \right|_y^\infty = \frac{1}{3} e^{-\frac{y}{3}} \quad \text{for } y \geqslant 0.$$

Then we have

$$f_{X|Y=2}(x \mid 2) = \begin{cases} \frac{1}{6} e^{\frac{2-x}{6}} & \text{if } x > 2, \\ 0 & \text{otherwise.} \end{cases}$$

Now it only remains to find $\mathbb{E}[X^2 \mid Y = 2]$ and $\mathbb{E}[X \mid Y = 2]$. Applying integration by parts twice we have

$$\mathbb{E}[X^2 \mid Y = 2] = \int_2^\infty \frac{x^2}{6} e^{\frac{2-x}{6}} dx = - \left. x^2 e^{\frac{2-x}{6}} \right|_2^\infty - \left. 12 x e^{\frac{2-x}{6}} \right|_2^\infty - \left. 12\, 6 e^{\frac{2-x}{6}} \right|_2^\infty = 4 + 24 + 72 = 100.$$

On the other hand, again applying integration by parts we get

$$\mathbb{E}[X \mid Y = 2] = \int_2^\infty \frac{x}{6} e^{\frac{2-x}{6}} dx = - \left. x e^{-\frac{x-2}{6}} \right|_2^\infty - \left. 6 e^{-\frac{x-2}{6}} \right|_2^\infty = 2 + 6 = 8.$$

Finally, we obtain $\mathrm{Var}(X \mid Y = 2) = 100 - 8^2 = 36$.

12.4. Exercises

Exercise 12.1. Suppose the joint distribution for X and Y is given by the joint probability mass function shown below:

$Y\backslash X$	0	1
0	0	0.3
1	0.5	0.2

(a) Find the covariance of X and Y.
(b) Find $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$.
(c) Find $\rho(X, Y)$.

Exercise 12.2. Let X and Y be random variables whose joint probability density function is given by

$$f(x,y) = \begin{cases} x+y & 0<x<1, 0<y<1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the covariance of X and Y.
(b) Find $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$.
(c) Find $\rho(X, Y)$.

Exercise 12.3. Let X be normally distributed with mean 1 and variance 9. Let Y be exponentially distributed with $\lambda = 2$. Suppose X and Y are independent. Find $\mathbb{E}\left[(X-1)^2 Y\right]$. (Hint: Use properties of expectations.)

Exercise* 12.1. Prove Proposition 12.2.

Exercise* 12.2. Show that if random variables X and Y are uncorrelated, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. Note that this is a more general statement than Proposition 12.3 since independent variables are uncorrelated.

Exercise* 12.3. Suppose U and V are independent random variables taking values 1 and -1 with probability $1/2$. Show that $X := U + V$ and $Y := U - V$ are dependent but uncorrelated random variables.

Exercise* 12.4. Use Equation (12.1.1) to prove Proposition 12.1 in the case when X and Y are independent discrete random variables.

12.5. Selected solutions

Solution to Exercise 12.1(A): First let us find the marginal distributions.

$Y\backslash X$	0	1	
0	0	0.3	0.3
1	0.5	0.2	0.7
	0.5	0.5	

Then

$$\begin{aligned}\mathbb{E}XY &= (0\cdot 0)\cdot 0 + (0\cdot 1)\cdot 0.5 + (1\cdot 0)\cdot 0.3 + (1\cdot 1)\cdot 0.2 = 0.2\\ \mathbb{E}X &= 0\cdot 0.5 + 1\cdot 0.5 = 0.5\\ \mathbb{E}Y &= 0\cdot 0.3 + 1\cdot 0.7 = 0.7.\end{aligned}$$

Solution to Exercise 12.1(B): First we need

$$\begin{aligned}\mathbb{E}X^2 &= 0^2 0.5 + 1^2 0.5 = 0.5\\ \mathbb{E}Y^2 &= 0^2 0.3 + 1^2 0.7 = 0.7\end{aligned}$$

Therefore

$$\begin{aligned}\operatorname{Var}(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 = 0.5 - (0.5)^2 = 0.25,\\ \operatorname{Var}(Y) &= \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = 0.7 - (0.7)^2 = 0.21,\end{aligned}$$

Solution to Exercise 12.1(C):

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\cdot\operatorname{Var}(Y)}} \approx -0.6547.$$

Solution to Exercise 12.2(A): We need to find $\mathbb{E}[XY], \mathbb{E}X$, and $\mathbb{E}Y$.

$$\begin{aligned}\mathbb{E}[XY] &= \int_0^1\int_0^1 xy\,(x+y)\,dydx = \int_0^1\left(\frac{x^2}{2}+\frac{x}{3}\right)dx = \frac{1}{3},\\ \mathbb{E}X &= \int_0^1\int_0^1 x(x+y)dydx = \int_0^1 (x^2+\frac{x}{2})dx = \frac{7}{12}\\ \mathbb{E}Y &= \frac{7}{12},\ \text{by symmetry with the } \mathbb{E}X \text{ case.}\end{aligned}$$

Therefore

$$\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \frac{1}{3} - \left(\frac{7}{12}\right)^2 = -\frac{1}{144}.$$

Solution to Exercise 12.2(B): We need $\mathbb{E}X^2$ and $\mathbb{E}Y^2$,

$$\mathbb{E}X^2 = \int_0^1 \int_0^1 x^2(x+y)dydx = \int_0^1 (x^3 + \frac{x^2}{2})dx = \frac{5}{12}$$

so we know that $\mathbb{E}Y^2 = \frac{5}{12}$ by symmetry. Therefore

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

Solution to Exercise 12.2(C):

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}} = -\frac{1}{11}.$$

Solution to Exercise 12.3: Since X, Y are independent then

$$\begin{aligned}\mathbb{E}\left[(X-1)^2 Y\right] &= \mathbb{E}\left[(X-1)^2\right]\mathbb{E}\left[Y\right] \\ &= \mathrm{Var}(X)\frac{1}{\lambda} = 9/2 = 4.5\end{aligned}$$

CHAPTER 13

Moment generating functions

13.1. Definition and examples

Definition (Moment generating function)

The *moment generating function* (MGF) of a random variable X is a function $m_X(t)$ defined by

$$m_X(t) = \mathbb{E}e^{tX},$$

provided the expectation is finite.

In the discrete case m_X is equal to $\sum_x e^{tx} p_X(x)$ and in the continuous case $\int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$.

Let us compute the moment generating function for some of the distributions we have been working with.

Example 13.1 (Bernoulli).

$$m_X(t) = e^{0\cdot t}(1-p) + e^{1\cdot t}p = e^t p + 1 - p.$$

Example 13.2 (Binomial). Using independence,

$$\mathbb{E}e^{t\sum X_i} = \mathbb{E}\prod e^{tX_i} = \prod \mathbb{E}e^{tX_i} = (pe^t + (1-p))^n,$$

where X_i are independent Bernoulli random variables. Equivalently

$$\sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}\left(pe^t\right)^k(1-p)^{n-k} = \left(pe^t + (1-p)\right)^n$$

by the Binomial formula.

Example 13.3 (Poisson).

$$\mathbb{E}e^{tX} = \sum_{k=0}^{\infty}\frac{e^{tk}e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}.$$

Example 13.4 (Exponential).

$$\mathbb{E}e^{tX} = \int_0^{\infty} e^{tx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda - t}$$

if $t < \lambda$, and ∞ if $t \geqslant \lambda$.

Example 13.5 (Standard normal). Suppose $Z \sim \mathcal{N}(0,1)$, then

$$m_Z(t) = \frac{1}{\sqrt{2\pi}} \int e^{tx} e^{-x^2/2} dx = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int e^{-(x-t)^2/2} dx = e^{t^2/2}.$$

Example 13.6 (General normal). Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$, then we can write $X = \mu + \sigma Z$, and therefore

$$m_X(t) = \mathbb{E}e^{tX} = \mathbb{E}e^{t\mu}e^{t\sigma Z} = e^{t\mu} m_Z(t\sigma) = e^{t\mu}e^{(t\sigma)^2/2} = e^{t\mu + t^2\sigma^2/2}.$$

Proposition 13.1

Suppose $X_1, ..., X_n$ are mutually independent random variables, and the random variable Y is defined by

$$Y = X_1 + ... + X_n.$$

Then

$$m_Y(t) = m_{X_1}(t) \cdot ... \cdot m_{X_n}(t).$$

Proof. By independence of $X_1, ..., X_n$ and Proposition 12.1 extended to n independent random variables we have

$$m_Y(t) = \mathbb{E}\left(e^{tX_1} \cdot ... \cdot e^{tX_n}\right) = \mathbb{E}e^{tX_1} \cdot ... \cdot \mathbb{E}e^{tX_n} = m_{X_1}(t) \cdot ... \cdot m_{X_n}(t).$$

□

Proposition 13.2

Suppose for two random variables X and Y we have $m_X(t) = m_Y(t) < \infty$ for all t in an interval, then X and Y have the same distribution.

We will not prove this, but this statement is essentially the uniqueness of the Laplace transform $\mathcal{L}$. Recall that the *Laplace transform* of a function $f(x)$ defined for all positive real numbers $s \geqslant 0$

$$(\mathcal{L}f)(s) := \int_0^\infty f(x) e^{-sx} dx$$

Thus if X is a continuous random variable with the PDF such that $f_X(x) = 0$ for $x < 0$, then

$$\int_0^\infty e^{tx} f_X(x) dx = \mathcal{L}f_X(-t).$$

Proposition 13.1 allows to show some of the properties of sums of independent random variables we proved or stated before.

Example 13.7 (Sums of independent normal random variables). If $X \sim \mathcal{N}(a, b^2)$ and $Y \sim \mathcal{N}(c, d^2)$, and X and Y are independent, then by Proposition 13.1

$$m_{X+Y}(t) = e^{at + b^2t^2/2} e^{ct + d^2t^2/2} = e^{(a+c)t + (b^2+d^2)t^2/2},$$

which is the moment generating function for $\mathcal{N}(a+c, b^2+d^2)$. Therefore Proposition 13.2 implies that $X+Y \sim \mathcal{N}(a+c, b^2+d^2)$.

Example 13.8 (Sums of independent Poisson random variables)**.** Similarly, if X and Y are independent Poisson random variables with parameters a and b, respectively, then

$$m_{X+Y}(t) = m_X(t)m_Y(t) = e^{a(e^t-1)}e^{b(e^t-1)} = e^{(a+b)(e^t-1)},$$

which is the moment generating function of a Poisson with parameter $a+b$, therefore $X+Y$ is a Poisson random variable with parameter $a+b$.

One problem with the moment generating function is that it might be infinite. One way to get around this, at the cost of considerable work, is to use the *characteristic function* $\varphi_X(t) = \mathbb{E}e^{itX}$, where $i = \sqrt{-1}$. This is always finite, and is the analogue of the *Fourier transform*.

Definition (Joint MGF)

The *joint moment generating function* of X and Y is

$$m_{X,Y}(s,t) = \mathbb{E}e^{sX+tY}.$$

If X and Y are independent, then

$$m_{X,Y}(s,t) = m_X(s)m_Y(t)$$

by Proposition 13.2. We will not prove this, but the converse is also true: if $m_{X,Y}(s,t) = m_X(s)m_Y(t)$ for *all* s and t, then X and Y are independent.

13.2. Further examples and applications

Example 13.9. Suppose that the MGF of X is given by $m_X(t) = e^{3(e^t-1)}$. Find $\mathbb{P}(X=0)$.

We can match this MGF to a known MGF of one of the distributions we considered and then apply Proposition 13.2. Observe that $m(t) = e^{3(e^t-1)} = e^{\lambda(e^t-1)}$, where $\lambda = 3$. Thus $X \sim Poisson(3)$, and therefore

$$\mathbb{P}(X=0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-3}.$$

This example is an illustration to why $m_X(t)$ is called the *moment generating function*. Namely we can use it to find all the moments of X by differentiating $m(t)$ and then evaluating at $t=0$. Note that

$$m_X'(t) = \frac{d}{dt}\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\frac{d}{dt}e^{tX}\right] = \mathbb{E}\left[Xe^{tX}\right].$$

Now evaluate at $t=0$ to get

$$m_X'(0) = \mathbb{E}\left[Xe^{0\cdot X}\right] = \mathbb{E}[X].$$

Similarly

$$m_X''(t) = \frac{d}{dt}\mathbb{E}\left[Xe^{tX}\right] = \mathbb{E}\left[X^2e^{tX}\right],$$

so that

$$m_X''(0) = \mathbb{E}\left[X^2e^{0}\right] = \mathbb{E}\left[X^2\right].$$

Continuing to differentiate the MGF we have the following proposition.

Proposition 13.3 (Moments from MGF)

For all $n \geqslant 0$ we have

$$\mathbb{E}[X^n] = m_X^{(n)}(0).$$

Example 13.10. Suppose X is a discrete random variable and has the MGF

$$m_X(t) = \frac{1}{7}e^{2t} + \frac{3}{7}e^{3t} + \frac{2}{7}e^{5t} + \frac{1}{7}e^{8t}.$$

What is the PMF of X? Find $\mathbb{E}X$.

Note that this MGF does not match any of the known MGFs directly. Reading off from the MGF we guess

$$\frac{1}{7}e^{2t} + \frac{3}{7}e^{3t} + \frac{2}{7}e^{5t} + \frac{1}{7}e^{8t} = \sum_{i=1}^{4} e^{tx_i}p(x_i)$$

then we can take $p(2) = \frac{1}{7}, p(3) = \frac{3}{7}, p(5) = \frac{2}{7}$ and $p(8) = \frac{1}{7}$. Note that these add up to 1, so this is indeed a PMF.

To find $\mathbb{E}[X]$ we can use Proposition 13.3 by taking the derivative of the moment generating function as follows.

$$m'(t) = \frac{2}{7}e^{2t} + \frac{9}{7}e^{3t} + \frac{10}{7}e^{5t} + \frac{8}{7}e^{8t},$$

so that

$$\mathbb{E}\left[X\right] = m'(0) = \frac{2}{7} + \frac{9}{7} + \frac{10}{7} + \frac{8}{7} = \frac{29}{7}.$$

Example 13.11. Suppose X has the MGF

$$m_X(t) = (1-2t)^{-\frac{1}{2}} \text{ for } t < \frac{1}{2}.$$

Find the first and second moments of X.

We first find

$$m'_X(t) = -\frac{1}{2}\left(1-2t\right)^{-\frac{3}{2}}(-2) = (1-2t)^{-\frac{3}{2}},$$
$$m''_X(t) = -\frac{3}{2}\left(1-2t\right)^{-\frac{5}{2}}(-2) = 3\left(1-2t\right)^{-\frac{5}{2}}.$$

Therefore

$$\mathbb{E}X = m'_X(0) = (1-2\cdot 0)^{-\frac{3}{2}} = 1,$$
$$\mathbb{E}X^2 = m''_X(0) = 3\left(1-2\cdot 0\right)^{-\frac{5}{2}} = 3.$$

13.3. Exercises

Exercise 13.1. Suppose that you have a fair 4-sided die, and let X be the random variable representing the value of the number rolled.

(a) Write down the moment generating function for X.
(b) Use this moment generating function to compute the first and second moments of X.

Exercise 13.2. Let X be a random variable whose probability density function is given by

$$f_X(x) = \begin{cases} e^{-2x} + \frac{1}{2}e^{-x} & x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

(a) Write down the moment generating function for X.
(b) Use this moment generating function to compute the first and second moments of X.

Exercise 13.3. Suppose that a mathematician determines that the revenue the UConn Dairy Bar makes in a week is a random variable, X, with moment generating function

$$m_X(t) = \frac{1}{(1-2500t)^4}$$

Find the standard deviation of the revenue the UConn Dairy bar makes in a week.

Exercise 13.4. Let X and Y be two independent random variables with respective moment generating functions

$$m_X(t) = \frac{1}{1-5t}, \quad \text{if } t < \frac{1}{5}, \quad m_Y(t) = \frac{1}{(1-5t)^2}, \quad \text{if } t < \frac{1}{5}.$$

Find $\mathbb{E}\,(X+Y)^2$.

Exercise 13.5. Suppose X and Y are independent Poisson random variables with parameters λ_x, λ_y, respectively. Find the distribution of $X+Y$.

Exercise 13.6. True or false? If $X \sim \operatorname{Exp}(\lambda_x)$ and $Y \sim \operatorname{Exp}(\lambda_y)$ then $X+Y \sim \operatorname{Exp}(\lambda_x+\lambda_y)$. Justify your answer.

Exercise* 13.1. Suppose the moment generating function $m_X(t)$ is defined on an interval $(-a,a)$. Show that

$$\left(m_X\left(\frac{t}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{t\mathbb{E}X}.$$

Exercise* 13.2. Suppose the moment generating function $m_X(t)$ is defined on an interval $(-a,a)$. Assume also that $\mathbb{E}X = 0$ and $\operatorname{Var}(X) = 1$. Show that

$$\left(m_X\left(\frac{t}{\sqrt{n}}\right)\right)^n \xrightarrow[n\to\infty]{} e^{\frac{t^2}{2}}.$$

13.4. Selected solutions

Solution to Exercise 13.1(A):

$$\begin{aligned} m_X(t) = \mathbb{E}\left[e^{tX}\right] &= e^{1\cdot t}\frac{1}{4} + e^{2\cdot t}\frac{1}{4} + e^{3\cdot t}\frac{1}{4} + e^{4\cdot t}\frac{1}{4} \\ &= \frac{1}{4}\left(e^{1\cdot t} + e^{2\cdot t} + e^{3\cdot t} + e^{4\cdot t}\right) \end{aligned}$$

Solution to Exercise 13.1(B): We have

$$\begin{aligned} m_X'(t) &= \frac{1}{4}\left(e^{1\cdot t} + 2e^{2\cdot t} + 3e^{3\cdot t} + 4e^{4\cdot t}\right), \\ m_X''(t) &= \frac{1}{4}\left(e^{1\cdot t} + 4e^{2\cdot t} + 9e^{3\cdot t} + 16e^{4\cdot t}\right), \end{aligned}$$

so

$$\mathbb{E}X = m_X'(0) = \frac{1}{4}(1+2+3+4) = \frac{5}{2}$$

and

$$\mathbb{E}X^2 = m_X''(0) = \frac{1}{4}(1+4+9+16) = \frac{15}{2}.$$

Solution to Exercise 13.2(A): for $t < 1$ we have

$$\begin{aligned} m_X(t) = \mathbb{E}\left[e^{tX}\right] &= \int_0^\infty e^{tx}\left(e^{-2x} + \frac{1}{2}e^{-x}\right) dx \\ &= \frac{1}{t-2}e^{tx-2x} + \frac{1}{2(t-1)}e^{tx-x}\Big|_{x=0}^{x=\infty} = \\ &= 0 - \frac{1}{2-t} + 0 - \frac{1}{2(t-1)} \\ &= \frac{1}{t-2} + \frac{1}{2(1-t)} = \frac{t}{2(2-t)(1-t)} \end{aligned}$$

Solution to Exercise 13.2(B): We have

$$\begin{aligned} m_X'(t) &= \frac{1}{(2-t)^2} + \frac{1}{2(1-t)^2} \\ m_X''(t) &= \frac{2}{(2-t)^3} + \frac{1}{(1-t)^3} \end{aligned}$$

and so $\mathbb{E}X = m_X'(0) = \frac{3}{4}$ and $\mathbb{E}X^2 = m_X'' = \frac{5}{4}$.

Solution to Exercise 13.3: We have $\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$ and $\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$. Therefore we can compute

$$\begin{aligned}
m'(t) &= 4(2500)(1-2500t)^{-5}, \\
m''(t) &= 20(2500)^2(1-2500t)^{-6}, \\
\mathbb{E}X &= m'(0) = 10,000 \\
\mathbb{E}X^2 &= m''(0) = 125,000,000 \\
\mathrm{Var}(X) &= 125,000,000 - 10,000^2 = 25,000,000 \\
\mathrm{SD}(X) &= \sqrt{25,000,000} = \mathbf{5,000}.
\end{aligned}$$

Solution to Exercise 13.4: First recall that if we let $W = X + Y$, and using that X, Y are independent, then we see that

$$m_W(t) = m_{X+Y}(t) = m_X(t)m_Y(t) = \frac{1}{(1-5t)^3},$$

recall that $\mathbb{E}[W^2] = m''_W(0)$, which we can find from

$$\begin{aligned}
m'_W(t) &= \frac{15}{(1-5t)^4}, \\
m''_W(t) &= \frac{300}{(1-5t)^5},
\end{aligned}$$

thus

$$\mathbb{E}\left[W^2\right] = m''_W(0) = \frac{300}{(1-0)^5} = 300.$$

Solution to Exercise 13.5: Since $X \sim \mathrm{Pois}(\lambda_x)$ and $Y \sim \mathrm{Pois}(\lambda_y)$ then

$$\begin{aligned}
m_X(t) &= e^{\lambda_x\left(e^t-1\right)}, \\
m_Y(t) &= e^{\lambda_y\left(e^t-1\right)}.
\end{aligned}$$

Then

$$\begin{aligned}
m_{X+Y}(t) &= m_X(t)m_Y(t) \\
&\overset{independence}{=} e^{\lambda_x\left(e^t-1\right)}e^{\lambda_y\left(e^t-1\right)} = e^{(\lambda_x+\lambda_y)\left(e^t-1\right)}.
\end{aligned}$$

Thus $X + Y \sim \mathrm{Pois}(\lambda_x + \lambda_y)$.

Solution to Exercise 13.6: We will use Proposition 13.2. Namely, we first find the MGF of $X + Y$ and compare it to the MGF of a random variable $V \sim \mathrm{Exp}(\lambda_x + \lambda_y)$. The MGF of V is

$$m_V(t) = \frac{\lambda_x + \lambda_y}{\lambda_x + \lambda_y - t} \text{ for } t < \lambda_x + \lambda_y.$$

By independence of X and Y

$$m_{X+Y}(t) = m_X(t)m_Y(t) = \frac{\lambda_x}{\lambda_x - t} \cdot \frac{\lambda_y}{\lambda_y - t},$$

but

$$\frac{\lambda_x + \lambda_y}{\lambda_x + \lambda_y - t} \neq \frac{\lambda_x}{\lambda_x - t} \cdot \frac{\lambda_y}{\lambda_y - t}$$

and hence the statement is false.

Solution to Exercise* 13.1: It is enough to show that

$$n \ln \left(m_X \left(\frac{t}{n} \right) \right) \xrightarrow[n\to\infty]{} t\mathbb{E}X.$$

Then

$$\begin{aligned}
\lim_{n\to\infty} n \ln \left(m_X \left(\frac{t}{n} \right) \right) &\overset{1/n=s}{=} \lim_{s\to 0} \frac{\ln (m_X (st))}{s} \\
&\overset{\text{L'Hôpital's rule}}{=} \lim_{s\to 0} \frac{\frac{tm'_X(st)}{m_X(st)}}{1} = \frac{tm'_X(0)}{m_X(0)} = \frac{t\mathbb{E}X}{\mathbb{E}1} = t\mathbb{E}X.
\end{aligned}$$

Solution to Exercise* 13.2: It is enough to show that

$$n \ln \left(m_X \left(\frac{t}{\sqrt{n}} \right) \right) \xrightarrow[n\to\infty]{} \frac{t^2}{2}.$$

Then

$$\begin{aligned}
\lim_{n\to\infty} n \ln \left(m_X \left(\frac{t}{\sqrt{n}} \right) \right) &= \lim_{n\to\infty} \frac{\ln \left(m_X \left(\frac{t}{\sqrt{n}} \right) \right)}{\frac{1}{n}} \overset{s=\frac{1}{\sqrt{n}}}{=} \lim_{s\to 0} \frac{\ln (m_X (st))}{s^2} \\
&\overset{\text{L'Hôpital's rule}}{=} \lim_{s\to 0} \frac{\frac{tm'_X(st)}{m_X(st)}}{2s} \overset{m_X(0)=1}{=} \frac{t}{2} \lim_{s\to 0} \frac{m'_X(st)}{s} \\
&\overset{\text{L'Hôpital's rule}}{=} \frac{t}{2} \lim_{s\to 0} \frac{tm''_X(st)}{1} = \frac{t^2 m''_X(0)}{2} = \frac{t^2}{2},
\end{aligned}$$

since by Proposition 13.3 we have $m''_X(0) = \mathbb{E}X^2 = 1$ and we are given that $\operatorname{Var} X = 1$ and $\mathbb{E}X = 0$.

CHAPTER 14

Limit laws and modes of convergence

14.1. SLLN, WLLN, CLT

Suppose we have a probability space of a sample space Ω, σ-field $\mathcal{F}$ and probability $\mathbb{P}$ defined on $\mathcal{F}$. We consider a collection of random variables X_i defined on the same probability space.

Definition 14.1 (Sums of i.i.d. random variables)

We say the sequence of random variables $\{X_i\}_{i=1}^{\infty}$ are *i.i.d.* (independent and identically distributed) if they are (mutually) independent and all have the same distribution. We call $S_n = \sum_{i=1}^{n} X_i$ the *partial sum process.*

In the case of continuous or discrete random variables, having the same distribution means that they all have the same probability density.

Theorem 14.1 (Strong law of large numbers (SLLN))

Suppose $\{X_i\}_{i=1}^{\infty}$ is a sequence of i.i.d. random variables with $\mathbb{E}|X_i| < \infty$, and let $\mu = \mathbb{E}X_i$. Then

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{} \mu.$$

The convergence here means that the sample mean $S_n(\omega)/n \to \mu$ for every $\omega \in \Omega$ except possibly for a set of ωs of probability 0.

The proof of Theorem 14.1 is quite hard, and we prove a weaker version, the weak law of large numbers (WLLN).

Theorem 14.2 (Weak law of large numbers (WLLN))

Suppose $\{X_i\}_{i=1}^{\infty}$ is a sequence of i.i.d. random variables with finite first and second moments, that is, $\mathbb{E}|X_1|$ and $\operatorname{Var} X_1$ are finite.
For every $a > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}X_1\right| > a\right) \xrightarrow[n\to\infty]{} 0.$$

It is not even that easy to give an example of random variables that satisfy the WLLN but not the SLLN. Before proving the WLLN, we need an inequality called *Markov's inequality*. We cover this in more detail later, see Proposition 15.3.

Proposition 14.1 (Markov's inequality)

If a random variable $Y \geqslant 0$, then for any $A > 0$

$$\mathbb{P}(Y > A) \leqslant \frac{\mathbb{E}Y}{A}.$$

Proof. We only consider the case of continuous random variables, the case of discrete random variables being similar. We have

$$\begin{aligned}\mathbb{P}(Y > A) = \int_A^\infty f_Y(y)\,dy \leqslant \int_A^\infty \frac{y}{A} f_Y(y)\,dy \\ \leqslant \frac{1}{A}\int_{-\infty}^\infty y f_Y(y)\,dy = \frac{1}{A}\mathbb{E}Y.\end{aligned}$$

□

We now prove the WLLN.

Proof of Theorem 14.2. Recall $\mathbb{E}S_n = n\mathbb{E}X_1$, and by the independence $\operatorname{Var} S_n = n \operatorname{Var} X_1$, so $\operatorname{Var}(S_n/n) = \operatorname{Var} X_1/n$. We have

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}X_1\right| > a\right) = \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| > a\right) \\ &= \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right|^2 > a^2\right) \overset{Markov}{\leqslant} \frac{\mathbb{E}\left|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})\right|^2}{a^2} \\ &= \frac{\operatorname{Var}(\frac{S_n}{n})}{a^2} = \frac{\frac{\operatorname{Var} X_1}{n}}{a^2} \to 0.\end{aligned}$$

The inequality step follows from Markov's inequality (Proposition 14.1) with $A = a^2$ and $Y = \left|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})\right|^2$. □

We now turn to the central limit theorem (CLT).

Theorem 14.3 (CLT)

Suppose $\{X_i\}_{i=1}^\infty$ is a sequence of i.i.d. random variables such that $\mathbb{E}X_i^2 < \infty$. Let $\mu = \mathbb{E}X_i$ and $\sigma^2 = \operatorname{Var} X_i$. Then

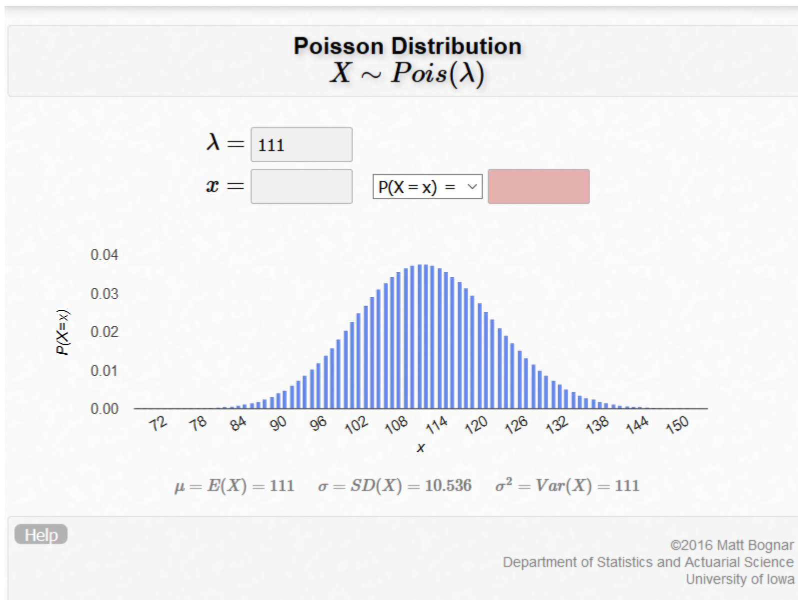
$$\mathbb{P}\left(a \leqslant \frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant b\right) \xrightarrow[n\to\infty]{} \mathbb{P}(a \leqslant Z \leqslant b)$$

for every a and b, where $Z \sim \mathcal{N}(0,1)$.

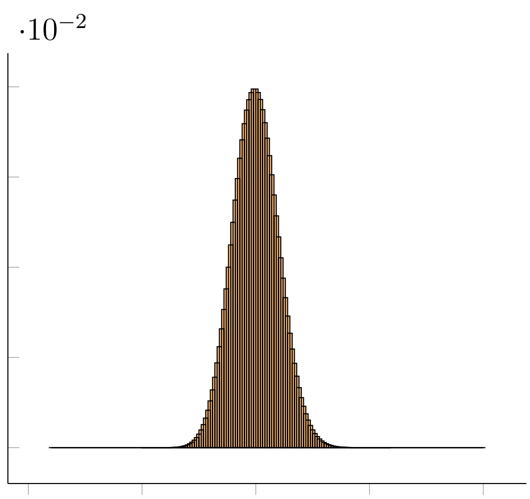
The ratio on the left is $(S_n - \mathbb{E}S_n)/\sqrt{\operatorname{Var} S_n}$. We do not claim that this ratio converges for any $\omega \in S$ (in fact, it does not), but that the probabilities converge.

Example 14.1 (Theorem 9.1). If X_i are i.i.d. Bernoulli random variables, so that S_n is a binomial, this is just the normal approximation to the binomial as in Theorem 9.1.

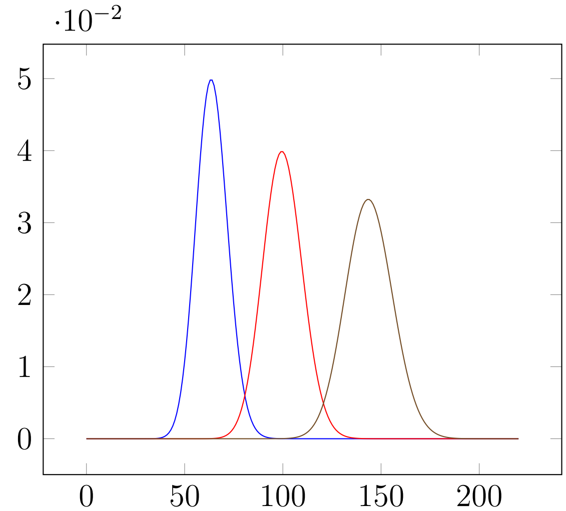
Example 14.2. Suppose in the CLT we have $X_i \sim \text{Pois}(\lambda)$, then $S_n \sim \text{Pois}(\lambda n)$ so we see that $(\text{Pois}(\lambda n) - n\lambda)/\lambda\sqrt{n}$ is close to the distribution of $\mathcal{N}(0,1)$. This is what is behind a normal distribution approximation to the Poisson distribution which can be illustrated by looking at the Poisson distribution below.



Poisson distribution for $\lambda = 111$



Histogram of the Poisson distribution for $\lambda = 100$.



Comparison of shapes of the Poisson distributions for $\lambda = 64$, $\lambda = 100$ and $\lambda = 144$.

Example 14.3. Suppose we roll a die 3600 times. Let X_i be the number showing on the i^{th} roll. We know S_n/n will be close to 3.5. What is the (approximate) probability it differs from 3.5 by more than 0.05?

We want to estimate

$$\mathbb{P}\left(\left|\frac{S_n}{n} - 3.5\right| > 0.05\right).$$

We rewrite this as

$$\mathbb{P}(|S_n - n\mathbb{E}X_1| > (0.05)(3600)) = \mathbb{P}\left(\left|\frac{S_n - n\mathbb{E}X_1}{\sqrt{n}\sqrt{\operatorname{Var} X_1}}\right| > \frac{180}{(60)\sqrt{\frac{35}{12}}}\right)$$

Note that $\frac{180}{(60)\sqrt{\frac{35}{12}}} \approx 1.7566$, so this probability can be approximately as

$$\mathbb{P}(|Z| > 1.757) \approx 0.08.$$

Example 14.4. Suppose the lifetime of a human has expectation 72 and variance 36. What is the (approximate) probability that the average of the lifetimes of 100 people exceeds 73?

We want to estimate

$$\begin{aligned}\mathbb{P}\left(\frac{S_n}{n} > 73\right) &= \mathbb{P}(S_n > 7300) \\ &= \mathbb{P}\left(\frac{S_n - n\mathbb{E}X_1}{\sqrt{n}\sqrt{\operatorname{Var} X_1}} > \frac{7300 - (100)(72)}{\sqrt{100}\sqrt{36}}\right) \\ &\approx \mathbb{P}(Z > 1.667) \approx 0.047.\end{aligned}$$

Sketch of the proof of Theorem 14.3. A typical proof of this theorem uses characteristic functions $\varphi_X(t) = \mathbb{E}e^{itX}$ (we mentioned them before Definition 13.1) which are defined for all t, but if the moment generating functions are defined on an interval $(-a, a)$, then the following idea can be used. It turns out that if

$$m_{Y_n}(t) \to m_Z(t) \text{ for every } t,$$

then $\mathbb{P}(a \leqslant Y_n \leqslant b) \to \mathbb{P}(a \leqslant Z \leqslant b)$, though *we will not prove this.*

We denote

$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$
$$W_i = \frac{X_i - \mu}{\sigma}.$$

Then $\mathbb{E}W_i = 0$, $\operatorname{Var} W_i = \frac{\operatorname{Var} X_i}{\sigma^2} = 1$, the random variables W_i are (mutually) independent, and

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n W_i}{\sqrt{n}}.$$

So there is no loss of generality in assuming that $\mu = 0$ and $\sigma = 1$. Then

$$m_{Y_n}(t) = \mathbb{E}e^{tY_n} = \mathbb{E}e^{(t/\sqrt{n})(S_n)} = m_{S_n}(t/\sqrt{n}).$$

Since the random variables X_i are i.i.d., all the X_i have the same moment generating function. Thus for $S_n = X_1 + ... + X_n$, we have

$$m_{S_n}(t) = m_{X_1}(t) \cdots m_{X_n}(t) = [m_{X_1}(t)]^n.$$

At this point we can refer to Exercise 13.2 or give a slightly different explanation as follows. If we expand e^{tX_1} as a power series in t, we have

$$m_{X_1}(t) = \mathbb{E}e^{tX_1} = 1 + t\mathbb{E}X_1 + \frac{t^2}{2!}\mathbb{E}(X_1)^2 + \frac{t^3}{3!}\mathbb{E}(X_1)^3 + \cdots .$$

We put the above together and obtain

$$m_{Y_n}(t) = m_{S_n}\left(\frac{t}{\sqrt{n}}\right) = [m_{X_1}(t/\sqrt{n})]^n =$$
$$= \left(1 + t \cdot 0 + \frac{(t/\sqrt{n})^2}{2!} + R_n\right)^n = \left(1 + \frac{t^2}{2n} + R_n\right)^n,$$

where $n^{3/2}R_n \xrightarrow[n\to\infty]{} t^3\mathbb{E}X^3$. Using the fact that

$$\left(1 + \frac{a}{n}\right)^n \xrightarrow[n\to\infty]{} e^a,$$

we see that $\left(1 + \frac{t^2}{2n} + R_n\right)^n \xrightarrow[n\to\infty]{} e^{t^2/2} = m_Z(t)$, and so we can use Proposition 13.2 to conclude the proof. □

14.2.* Convergence of random variables

In the limits laws we discussed earlier we used different modes of convergence of sequences of random variables which we now discuss separately.

Definition (Convergence in probability)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and X be a random variable on the same probability space. We say that the sequence $\{X_n\}_{n\in\mathbb{N}}$ *converges in probability* to X if for any $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0.$$

In such a case, we write $X_n \xrightarrow[n\to\infty]{P} X$.

Example 14.5 (Weak law of large numbers)**.** The convergence used in the Weak law of large numbers (Theorem 14.2) is convergence in probability. Namely, it says that the sample mean S_n/n converges to $\mathbb{E}X_1$ in probability.

Definition (Almost sure convergence)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and X be a random variable on the same probability space. We say that the sequence $\{X_n\}_{n\in\mathbb{N}}$ *converges almost surely* or *almost everywhere* or *with probability 1* to X if

$$\mathbb{P}\left(X_n \xrightarrow[n\to\infty]{} X\right) = 1,$$

and we denote it by $X_n \xrightarrow[n\to\infty]{} X$ a.s.

This mode of convergence is a slightly modified version of the concept of pointwise convergence of functions. Recall that a random variable is a function from the sample space to $\mathbb{R}$. Pointwise convergence would require that

$$X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$$

for *all* $\omega \in \Omega$. This is usually too much to assume, this is why we have almost sure convergence. Namely, define the event

$$E := \{\omega \in \Omega : X_n(\omega) \text{ does not converge }\}.$$

If E is a subset of an event of zero probability, we say that the sequence $\{X_n\}_{n\in\mathbb{N}}$ converges to X almost surely. This mode of convergence is a widely used concept in probability and statistics, but proving almost sure convergence usually requires tools from measure theory, which is beyond the scope of this course.

Example 14.6 (Strong law of large numbers)**.** This is the type of convergence appearing the Strong law of large numbers (Theorem 14.1). Note that we mentioned this right after the theorem, and now we can say that the conclusion of the SLLN is that the sample mean S_n/n converges to $\mathbb{E}X_1$ almost surely.

Definition (Convergence in distribution)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we denote by $F_n(x)$ their distribution functions. Let X is a random variable with distribution function $F_X(x)$. We say that the sequence $\{X_n\}_{n\in\mathbb{N}}$ *converges in distribution* or *converges weakly* to X if

$$F_n(x) \xrightarrow[n\to\infty]{} F_X(x)$$

for all $x \in \mathbb{R}$ at which $F_X(x)$ is continuous, and we denote it by $X_n \xrightarrow[n\to\infty]{d} X$.

Note that convergence in distribution only involves the distributions of random variables. In particular, the random variables need not even be defined on the same probability space, and moreover we do not even need the random variables to define convergence in distribution. This is in contrast to other modes of convergence we have introduced.

Example 14.7 (Central limit theorem)**.** The Central limit theorem (Theorem 14.3) says that $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges to the standard normal in distribution.

Example 14.8. Suppose $X_n \sim \operatorname{Binom}\left(n, \frac{\lambda}{n}\right)$, $n > \lambda > 0$. Then $X_n \xrightarrow[n\to\infty]{d} \operatorname{Pois}(\lambda)$.

Definition (Convergence in L^p-mean)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let X be a random variable on the same probability space. For $p \geqslant 1$ we say that the sequence $\{X_n\}_{n\in\mathbb{N}}$ *converges in the pth mean* or in the $L^p - norm$ to X if

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^p = 0$$

provided that $\mathbb{E}|X_n|^p$ and $\mathbb{E}|X|^p$ exist for all n, and we denote it by $X_n \xrightarrow[n\to\infty]{L^p} X$.

In particular, for $p = 1$ we say that X_n converges in mean, and for $p = 2$ we say that X_n converges in mean square. While we have not used this type of convergence in this course, this is used widely in probability and statistics. One of the tools used for this type of convergence is Jensen's inequality.

Relations between different modes of convergence

1. For $p > q \geqslant 1$, convergence in the pth mean implies convergence in qth mean.
2. Convergence in mean implies convergence in probability.
3. Almost sure convergence implies convergence in probability.
4. Convergence in probability implies convergence in distribution.

14.3. Further examples and applications

Example 14.9. If 10 fair dice are rolled, find the approximate probability that the sum obtained is between 30 and 40, inclusive.

We will use the ± 0.5 continuity correction because these are discrete random variables. Let X_i denote the value of the ith die. Recall that

$$\mathbb{E}(X_i) = \frac{7}{2}\,\mathrm{Var}(X_i) = \frac{35}{12}.$$

Take

$$X = X_1 + \cdots + X_n$$

to be their sum. To apply the CLT we have

$$\begin{aligned} n\mu &= 10 \cdot \frac{7}{2} = 35 \\ \sigma\sqrt{n} &= \sqrt{\frac{350}{12}}, \end{aligned}$$

thus using the continuity correction we have

$$\begin{aligned} \mathbb{P}(29.5 \leqslant X \leqslant 40.5) &= \mathbb{P}\left(\frac{29.5-35}{\sqrt{\frac{350}{12}}} \leqslant \frac{X-35}{\sqrt{\frac{350}{12}}} \leqslant \frac{40.5-35}{\sqrt{\frac{350}{12}}}\right) \\ &\approx \mathbb{P}(-1.0184 \leqslant Z \leqslant 1.0184) \\ &= \Phi(1.0184) - \Phi(-1.0184) \\ &= 2\Phi(1.0184) - 1 = 0.692. \end{aligned}$$

Example 14.10. Your instructor has 1000 Probability final exams that needs to be graded. The time required to grade an exam are all i.i.d. random variables with mean of 20 minutes and standard deviation of 4 minutes. Approximate the probability that your instructor will be able to grade at least 25 exams in the first 450 minutes of work.

Denote by X_i the time it takes to grade exam i. Then

$$X = X_1 + \cdots + X_{25}$$

is the time it takes to grade the first 25 exams. We want $\mathbb{P}(X \leq 450)$. To apply the CLT we have

$$\begin{aligned} n\mu &= 25 \cdot 20 = 500 \\ \sigma\sqrt{n} &= 4\sqrt{25} = 20. \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{P}(X \leqslant 450) &= \mathbb{P}\left(\frac{X-500}{20} \leqslant \frac{450-500}{20}\right) \\ &\approx \mathbb{P}(Z \leqslant -2.5) \\ &= 1 - \Phi(2.5) = 0.006. \end{aligned}$$

14.4. Exercises

Exercise 14.1. In a 162-game season, find the approximate probability that a team with a 0.5 chance of winning will win at least 87 games.

Exercise 14.2. An individual students MATH 3160 Final exam score at UConn is a random variable with mean 75 and variance 25, How many students would have to take the examination to ensure with probability at least 0.9 that the class average would be within 5 of 75?

Exercise 14.3. Let $X_1, X_2, ..., X_{100}$ be independent exponential random variables with parameter $\lambda = 1$. Use the central limit theorem to approximate

$$\mathbb{P}\left(\sum_{i=1}^{100} X_i > 90\right).$$

Exercise 14.4. Suppose an insurance company has 10,000 automobile policy holders. The expected yearly claim per policy holder is \$240, with a standard deviation of \$800. Approximate the probability that the total yearly claim is greater than \$2,500,000.

Exercise 14.5. Suppose that the checkout time at the UConn dairy bar has a mean of 5 minutes and a standard deviation of 2 minutes. Estimate the probability to serve at least 36 customers during a 3-hour and a half shift.

Exercise 14.6. Shabazz Napier is a basketball player in the NBA. His expected number of points per game is 15 with a standard deviation of 5 points per game. The NBA season is 82 games long. Shabazz is guaranteed a ten million dollar raise next year if he can score a total of 1300 points this season. Approximate the probability that Shabazz will get a raise next season.

Exercise* 14.1. Assuming that the moment generating functions are defined on an interval $(-a, a)$, use Exercise* 13.1 to give another proof of the WLLN.

14.5. Selected solutions

Solution to Exercise 14.1: Let X_i be 1 if the team wins the ith game and 0 if the team loses. This is a Bernoulli random variable with $p = 0.5$. Thus $\mu = p = 0.5$ and $\sigma^2 = p(1-p) = (0.5)^2$. Then

$$X = \sum_{i=1}^{162} X_i$$

is the number of games won in the season. Using the CLT with

$$n\mu = 162 \cdot 0.5 = 81$$
$$\sigma\sqrt{n} = 0.5\sqrt{162} \approx 6.36,$$

then

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{162} X_i \geqslant 87\right) &= \mathbb{P}(X \geqslant 86.5) \\ &= \mathbb{P}\left(\frac{X-81}{6.36} > \frac{86.5-81}{6.36}\right) \\ &\approx \mathbb{P}(Z > 0.86) \approx 0.1949,\end{aligned}$$

where we used a continuity correction since X is a discrete random variable.

Solution to Exercise 14.2: Now $\mu = 75$, $\sigma^2 = 25, \sigma = 5$.

$$\begin{aligned}\mathbb{P}\left(70 < \frac{\sum_{i=1}^n X_i}{n} < 80\right) \geqslant 0.9 &\Longleftrightarrow \mathbb{P}\left(70 \cdot n < \sum_{i=1}^n X_i < 80 \cdot n\right) \geqslant 0.9 \\ &\Longleftrightarrow \mathbb{P}\left(\frac{70 \cdot n - 75 \cdot n}{5\sqrt{n}} < Z < \frac{80 \cdot n - 75 \cdot n}{5\sqrt{n}}\right) \geqslant 0.9 \\ &\Longleftrightarrow \mathbb{P}\left(-5\frac{\sqrt{n}}{5} < Z < 5\frac{\sqrt{n}}{5}\right) \geqslant 0.9 \\ &\Longleftrightarrow \mathbb{P}\left(-\sqrt{n} < Z < \sqrt{n}\right) \geqslant 0.9 \\ &\Longleftrightarrow \Phi\left(\sqrt{n}\right) - \Phi\left(-\sqrt{n}\right) \geqslant 0.9 \\ &\Longleftrightarrow \Phi\left(\sqrt{n}\right) - \left(1 - \Phi\left(\sqrt{n}\right)\right) \geqslant 0.9 \\ &\Longleftrightarrow 2\Phi\left(\sqrt{n}\right) - 1 \geqslant 0.9 \\ &\Longleftrightarrow \Phi\left(\sqrt{n}\right) \geqslant 0.95.\end{aligned}$$

Using the table inversely we have that

$$\sqrt{n} \geqslant 1.65 \Longrightarrow n \geqslant 2.722$$

hence the first integer that insurers that $n \geqslant 2.722$ is

$$n = 3.$$

Solution to Exercise 14.3: Since $\lambda = 1$ then $\mathbb{E}X_i = 1$ and $\operatorname{Var}(X_i) = 1$. Use the CLT with

$$n\mu = 100 \cdot 1 = 100$$
$$\sigma\sqrt{n} = 1 \cdot \sqrt{100} = 10.$$

$$\mathbb{P}\left(\sum_{i=1}^{100} X_i > 90\right) = \mathbb{P}\left(\frac{\sum_{i=1}^{100} X_i - 100 \cdot 1}{1 \cdot \sqrt{100}} > \frac{90 - 100 \cdot 1}{1 \cdot \sqrt{100}}\right)$$
$$\approx \mathbb{P}(Z > -1) \approx 0.8413.$$

Solution to Exercise 14.4:

$$\mathbb{P}(X \geqslant 1300) = \mathbb{P}\left(\frac{X - 2400000}{80000} \geqslant \frac{2500000 - 2400000}{80000}\right)$$
$$\approx \mathbb{P}(Z \geqslant 1.25)$$
$$= 1 - \Phi(1.25) \approx 1 - 0.8944 = 0.1056.$$

Solution to Exercise 14.5: Let X_i be the time it takes to check out customer i. Then

$$X = X_1 + \cdots + X_{36}$$

is the time it takes to check out 36 customer. We want to estimate $\mathbb{P}(X \leqslant 210)$. Use the CLT with

$$n\mu = 36 \cdot 5 = 180,$$
$$\sigma\sqrt{n} = 2\sqrt{36} = 12.$$

Thus

$$\mathbb{P}(X \leqslant 210) = \mathbb{P}\left(\frac{X - 180}{12} \leqslant \frac{210 - 180}{12}\right)$$
$$\approx \mathbb{P}(Z \leqslant 2.5)$$
$$= \Phi(2.5) \approx 0.9938.$$

Solution to Exercise 14.6:

Let X_i be the number of points scored by Shabazz in game i . Then

$$X = X_1 + \cdots + X_{82}$$

is the total number of points in a whole season. We want to estimate $\mathbb{P}(X \geqslant 1800)$. Use the CLT with

$$n\mu = 82 \cdot 15 = 1230,$$
$$\sigma\sqrt{n} = 5\sqrt{82} \approx 45.28.$$

Thus

$$\begin{aligned}\mathbb{P}(X \geqslant 1300) &= \mathbb{P}\left(\frac{X-1230}{45.28} \geqslant \frac{1300-1230}{45.28}\right)\\ &\approx \mathbb{P}(Z \geqslant 1.55)\\ &= 1-\Phi(1.55) \approx 1-0.9394 = 0.0606.\end{aligned}$$

CHAPTER 15

Probability inequalities

We already used several types of inequalities, and in this Chapter we give a more systematic description of the inequalities and bounds used in probability and statistics.

15.1.* Boole's inequality, Bonferroni inequalities

Boole's inequality(or the *union bound*) states that for any at most countable collection of events, the probability that at least one of the events happens is no greater than the sum of the probabilities of the events in the collection.

Proposition 15.1 (Boole's inequality)

Suppose $(S, \mathcal{F}, \mathbb{P})$ is a probability space, and $E_1, E_2, ... \in \mathcal{F}$ are events. Then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \leqslant \sum_{i=1}^{\infty} \mathbb{P}\left(E_i\right).$$

Proof. We only give a proof for a finite collection of events, and we mathematical induction on the number of events.

For the $n=1$ we see that

$$\mathbb{P}\left(E_1\right) \leqslant \mathbb{P}\left(E_1\right).$$

Suppose that for some n and any collection of events $E_1, ..., E_n$ we have

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leqslant \sum_{i=1}^{n} \mathbb{P}\left(E_i\right).$$

Recall that by (2.1.1) for any events A and B we have

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

We apply it to $A = \bigcup_{i=1}^{n} E_i$ and $B = E_{n+1}$ and using the associativity of the union $\bigcup_{i=1}^{n+1} E_i = A \cup B$, we get that

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} E_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{n} E_i) + \mathbb{P}(E_{n+1}\right) - \mathbb{P}\left(\left(\bigcup_{i=1}^{n} E_i\right) \bigcap E_{n+1}\right).$$

By the first axiom of probability

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i \cap A_{n+1}\right) \geqslant 0,$$

and therefore we have

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} E_i\right) \leqslant \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) + \mathbb{P}\left(E_{n+1}\right).$$

Thus using the *induction hypothesis* we see that

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} E_i\right) \leqslant \sum_{i=1}^{n} \mathbb{P}\left(E_i\right) + \mathbb{P}\left(E_{n+1}\right) = \sum_{i=1}^{n+1} \mathbb{P}\left(E_i\right).$$

□

One of the interpretations of Boole's inequality is what is known as σ*-sub-additivity* in measure theory applied here to the probability measure $\mathbb{P}$.

Boole's inequality can be extended to get lower and upper bounds on probability of unions of events known as *Bonferroni inequalities*. As before suppose $(S, \mathcal{F}, \mathbb{P})$ is a probability space, and $E_1, E_2, ...E_n \in \mathcal{F}$ are events. Define

$$\begin{aligned}
S_1 &:= \sum_{i=1}^{n} \mathbb{P}\left(E_i\right), \\
S_2 &:= \sum_{1 \leqslant i < j \leqslant n} \mathbb{P}\left(E_i \cap E_j\right) \\
S_k &:= \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \mathbb{P}\left(E_{i_1} \cap \cdots \cap E_{i_k}\right), k = 3, ..., n.
\end{aligned}$$

Proposition 15.2 (Bonferroni inequalities)

For odd k in $1, ..., n$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leqslant \sum_{j=1}^{k} (-1)^{j-1} S_j,$$

for even k in $2, ..., n$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \geqslant \sum_{j=1}^{k} (-1)^{j-1} S_j.$$

We omit the proof which starts with considering the case $k = 1$ for which we need to show

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leqslant \sum_{j=1}^{1}(-1)^{j-1} S_j = S_1 = \sum_{i=1}^{n} \mathbb{P}\left(E_i\right),$$

which is Boole's inequality. When $k = 2$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \geqslant \sum_{j=1}^{2}(-1)^{j-1} S_j = S_1 - S_1 = \sum_{i=1}^{n} \mathbb{P}\left(E_i\right) - \sum_{1 \leqslant i < j \leqslant n} \mathbb{P}\left(E_i \cap E_j\right).$$

which for $n = 2$ is the inclusion-exclusion identity (Proposition 2.2).

Example 15.1. Suppose we place n distinguishable balls into m distinguishable boxes at random ($n > m$). Let E be the event that a box is empty. The sample space can be described as

$$\Omega = \{\omega = (\omega_1, ..., \omega_n) : 1 \leqslant \omega_i \leqslant m\}$$

with $\mathbb{P}(\omega) = \frac{1}{m^n}$. Denote $E_l := \{\omega : \omega_i \neq l \text{ for all } i = 1, ..., n\}$ for $l = 1, 2, ..., m$. Then, $E = E_1 \cup ... \cup E_{m-1}$ since E_m is empty, and we can include it or not, this does not change the result.

We can see that for any m we have

$$\mathbb{P}\left(E_{i_1} \cup ... \cup E_{i_k}\right) = \frac{(m-k)^n}{k^n} = \left(1 - \frac{k}{m}\right)^n.$$

Then we can use the inclusion-exclusion principle to get

$$\mathbb{P}(E) = m\left(1 - \frac{1}{m}\right)^n - \binom{m}{2}\left(1 - \frac{2}{m}\right)^n + ... + (-1)^{m-2}\binom{m}{m-1}\left(1 - \frac{m-1}{m}\right)^n$$

The last term is zero, since all boxes can not be empty. The expression is quite complicated. But if we use Bonferroni inequalities we see that

$$m\left(1 - \frac{1}{m}\right)^n - \binom{m}{2}\left(1 - \frac{2}{m}\right)^n \leqslant \mathbb{P}(E) \leqslant m\left(1 - \frac{1}{m}\right)^n$$

This gives a good estimate when n is large compared to m. For example, if $m = 10$ then

$$10 \cdot (0.9)^n - 45 \cdot (0.8)^n \leqslant \mathbb{P}(E) \leqslant 10 \cdot (0.9)^n.$$

In particular, for $n = 50$, then $45 \cdot (0.8)^{50} = 0.00064226146$, which is the difference between the left and right sides of the estimates. This gives a rather good estimate.

15.2. Markov's inequality

Proposition 15.3 (Markov's inequality)

Suppose X is a nonnegative random variable, then for any $a > 0$ we have

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}X}{a}$$

Proof. We only give the proof for a continuous random variable, the case of a discrete random variable is similar. Suppose X is a positive continuous random variable, we can write

$$\begin{aligned}\mathbb{E}X &= \int_{-\infty}^{\infty} x f_X(x)\,dx \overset{X\geqslant 0}{=} \int_0^{\infty} x f_X(x)\,dx \\ &\overset{a>0}{\geqslant} \int_a^{\infty} x f_X(x)\,dx \overset{x>a}{\geqslant} \int_a^{\infty} a f_X(x)\,dx = a\int_a^{\infty} f_X(x)\,dx = a\mathbb{P}(X\geqslant a).\end{aligned}$$

Therefore

$$a\mathbb{P}(X\geqslant a)\leqslant \mathbb{E}X$$

which is what we wanted to prove. □

Example 15.2. First we observe that Boole's inequality can be interpreted as expectations of the number of occurred events. Suppose $(S, \mathcal{F}, \mathbb{P})$ is a probability space, and $E_1, E_2, ... \in \mathcal{F}$ are events. Define

$$X_i := \begin{cases} 1, & \text{if } E_i \text{ occurs} \\ 0, & \text{otherwise.} \end{cases}$$

Then $X := X_1 + ... + X_n$ is the number of events that occur. Then

$$\mathbb{E}X = \mathbb{P}(E_1) + ... + \mathbb{P}(E_n).$$

Now we would like to prove Boole's inequality using Markov's inequality. Note that X is a nonnegative random variable, so we can apply Markov's inequality. For $a = 1$ we get

$$\mathbb{P}(X\geqslant 1)\leqslant \mathbb{E}X = \mathbb{P}(E_1) + ... + \mathbb{P}(E_n).$$

Finally we see that the event $X \geqslant 1$ means that at least one of the events $E_1, E_2, ...E_n$ occur, so

$$\mathbb{P}(X\geqslant 1) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right),$$

therefore

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \mathbb{P}(X\geqslant 1)\leqslant \mathbb{E}X = \mathbb{P}(E_1) + ... + \mathbb{P}(E_n)$$

which completes the proof.

Example 15.3. Suppose $X \sim \text{Binom}(n, p)$. We would like to use Markov's inequality to find an upper bound on $\mathbb{P}(X \geqslant qn)$ for $p < q < 1$.

Note that X is a nonnegative random variable and $\mathbb{E}X = np$. By Markov's inequality, we have

$$\mathbb{P}(X\geqslant qn)\leqslant \frac{\mathbb{E}X}{qn} = \frac{p}{q}.$$

15.3. Chebyshev's inequality

Here we revisit Chebyshev's inequality Proposition 14.1 we used previously. This results shows that the difference between a random variable and its expectation is controlled by its variance. Informally we can say that it shows how far the random variable is from its mean on average.

Proposition 15.4 (Chebyshev's inequality)

Suppose X is a random variable, then for any $b > 0$ we have

$$\mathbb{P}\left(|X - \mathbb{E}X| \geqslant b\right) \leqslant \frac{\operatorname{Var}(X)}{b^2}.$$

Proof. Define $Y := (X - \mathbb{E}X)^2$, then Y is a nonnegative random variable and we can apply Markov's inequality (Proposition 15.3) to Y. Then for $b > 0$ we have

$$\mathbb{P}\left(Y \geqslant b^2\right) \leqslant \frac{\mathbb{E}Y}{b^2}.$$

Note that

$$\begin{aligned}
&\mathbb{E}Y = \mathbb{E}\left(X - \mathbb{E}X\right)^2 = \operatorname{Var}(X), \\
&\mathbb{P}\left(Y \geqslant b^2\right) = \mathbb{P}\left((X - \mathbb{E}X)^2 \geqslant b^2\right) = \mathbb{P}\left(|X - \mathbb{E}X| \geqslant b\right)
\end{aligned}$$

which completes the proof. □

Example 15.4. Consider again $X \sim \operatorname{Binom}(n, p)$. We now will use Chebyshev's inequality to find an upper bound on $\mathbb{P}(X \geqslant qn)$ for $p < q < 1$.

Recall that $\mathbb{E}X = np$. By Chebyshev's inequality with $b = (q - p)\,n > 0$ we have

$$\begin{aligned}
&\mathbb{P}\left(X \geqslant qn\right) = \mathbb{P}\left(X - np \geqslant (q - p)\,n\right) \leqslant \mathbb{P}\left(|X - np| \geqslant (q - p)\,n\right) \\
&\leqslant \frac{\operatorname{Var}(X)}{\left((q - p)\,n\right)^2} = \frac{p\,(1 - p)\,n}{\left((q - p)\,n\right)^2} = \frac{p\,(1 - p)}{(q - p)^2\,n}.
\end{aligned}$$

15.4. Chernoff bounds

Proposition 15.5 (Chernoff bounds)

Suppose X is a random variable and we denote by $m_X(t)$ its moment generating function, then for any $a \in \mathbb{R}$

$$\mathbb{P}(X \geqslant a) \leqslant \min_{t>0} e^{-ta} m_X(t),$$
$$\mathbb{P}(X \leqslant a) \leqslant \min_{t<0} e^{-ta} m_X(t).$$

Proof.

$$\mathbb{P}(X \geqslant a) = \mathbb{P}\left(e^{tX} \geqslant e^{ta}\right), t > 0,$$
$$\mathbb{P}(X \leqslant a) = \mathbb{P}\left(e^{tX} \geqslant e^{ta}\right), t < 0.$$

Note that note that e^{tX} is a positive random variable for any $t \in \mathbb{R}$. Therefore we can apply Markov's inequality (Proposition 15.3) to see that

$$\mathbb{P}(X \geqslant a) = \mathbb{P}\left(e^{tX} \geqslant e^{ta}\right) \leqslant \frac{\mathbb{E}e^{tX}}{e^{ta}}, t > 0,$$
$$\mathbb{P}(X \leqslant a) = \mathbb{P}\left(e^{tX} \geqslant e^{ta}\right) \leqslant \frac{\mathbb{E}e^{tX}}{e^{ta}}, t < 0.$$

Recall that $\mathbb{E}e^{tX}$ is the moment generating function $m_X(t)$, and so we have

$$\mathbb{P}(X \geqslant a) \leqslant \frac{m_X(t)}{e^{ta}}, t > 0,$$
$$\mathbb{P}(X \leqslant a) \leqslant \frac{m_X(t)}{e^{ta}}, t < 0.$$

Taking the minimum over appropriate t we get the result.

□

Example 15.5. Consider again $X \sim \operatorname{Binom}(n, p)$. We now will use Chernoff bounds for $\mathbb{P}(X \geqslant qn)$ for $p < q < 1$. Recall that in Example 13.2 we found the moment generating function for X as follows

$$m_X(t) = (pe^t + (1-p))^n.$$

Thus a Chernoff bound gives

$$\mathbb{P}(X \geqslant qn) \leqslant \min_{t>0} e^{-tqn}\left(pe^t + (1-p)\right)^n.$$

To find the minimum of $g(t) = e^{-tqn}\left(pe^t + (1-p)\right)^n$ we can take its derivative and using the only critical point of this function , we can see that the minimum on $(0, \infty)$ is achieved at t_* such that

$$e^{t_*} = \frac{q(1-p)}{(1-q)p},$$

and so

$$\begin{aligned} g(t_*) &= \left(\frac{q(1-p)}{(1-q)p}\right)^{-qn} \left(p\frac{q(1-p)}{(1-q)p} + (1-p)\right)^{n} \\ &= \left(\frac{q(1-p)}{(1-q)p}\right)^{-qn} \left(\frac{1-p}{1-q}\right)^{n} = \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{-qn} \left(\frac{1-p}{1-q}\right)^{n} \\ &= \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{(1-q)n}. \end{aligned}$$

Thus the Chernoff bound gives

$$\mathbb{P}(X \geqslant qn) \leqslant \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{(1-q)n}.$$

Example 15.6 (Comparison of Markov's, Chebyshev's inequalities and Chernoff bounds)**.** These three inequalities for the binomial random variable $X \sim \operatorname{Binom}(n,p)$ give

$$\begin{aligned} &\text{Markov's inequality} && \mathbb{P}(X \geqslant qn) \leqslant \frac{p}{q}, \\ &\text{Chebyshev's inequality} && \mathbb{P}(X \geqslant qn) \leqslant \frac{p(1-p)}{(q-p)^2 n}, \\ &\text{Chernoff bound} && \mathbb{P}(X \geqslant qn) \leqslant \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{(1-q)n}. \end{aligned}$$

Clearly the right-hand sides are very different: Markov's inequality gives a bound independent of n, and the Chernoff bound is the strongest with exponential convergence to 0 as $n \to \infty$. For example, for $p = 1/2$ and $q = 3/4$ we have

$$\begin{aligned} &\text{Markov's inequality} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \frac{2}{3}, \\ &\text{Chebyshev's inequality} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \frac{4}{n}, \\ &\text{Chernoff bound} && \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \left(\frac{16}{27}\right)^{n/4}. \end{aligned}$$

For example, for $p = 1/3$ and $q = 2/3$ we have

$$\text{Markov's inequality} \qquad \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \frac{1}{2},$$

$$\text{Chebyshev's inequality} \qquad \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant \frac{2}{n},$$

$$\text{Chernoff bound} \qquad \mathbb{P}\left(X \geqslant \frac{3n}{4}\right) \leqslant 2^{-n/2}.$$

15.5. Cauchy-Schwarz inequality

This inequality appears in a number of areas of mathematics including linear algebra. We will apply it to give a different proof for the bound for correlation coefficients. Note that the Cauchy-Schwarz inequality can be easily generalized to random vectors X and Y.

Proposition 15.6 (Cauchy-Schwarz inequality)

Suppose X and Y are two random variables, then

$$(\mathbb{E}XY)^2 \leqslant \mathbb{E}X^2 \cdot \mathbb{E}Y^2,$$

and the equality holds if and only if $X = aY$ for some constant $a \in \mathbb{R}$.

PROOF. Define the random variable $U := (X - sY)^2$ which is a nonnegative random variable for any $s \in \mathbb{R}$. Then

$$0 \leqslant \mathbb{E}U = \mathbb{E}(X - sY)^2 = \mathbb{E}X^2 - 2s\mathbb{E}XY + s^2\mathbb{E}Y^2.$$

Define $g(s) := \mathbb{E}X^2 - 2s\mathbb{E}XY + s^2\mathbb{E}Y^2$ which is a quadratic polynomial in s. What we know is that $g(s)$ is nonnegative for all s. Completing the square we see that

$$g(s) = \mathbb{E}Y^2 s^2 - 2\mathbb{E}XYs + \mathbb{E}X^2 = \left(\sqrt{\mathbb{E}Y^2}s - \frac{\mathbb{E}XY}{\sqrt{\mathbb{E}Y^2}}\right)^2 + \mathbb{E}X^2 - \frac{(\mathbb{E}XY)^2}{\mathbb{E}Y^2},$$

so $g(s) \geqslant 0$ for all s if and only if

$$\mathbb{E}X^2 - \frac{(\mathbb{E}XY)^2}{\mathbb{E}Y^2} \geqslant 0,$$

which is what we needed to show.

To deal with the last claim, observe that if $U > 0$ with probability one, then $g(s) = \mathbb{E}U > 0$. This happens only if

$$\mathbb{E}X^2 - \frac{(\mathbb{E}XY)^2}{\mathbb{E}Y^2} > 0.$$

And if $\mathbb{E}X^2 - \frac{(\mathbb{E}XY)^2}{\mathbb{E}Y^2} = 0$, then $g\left(\frac{\mathbb{E}XY}{\mathbb{E}Y^2}\right) = \mathbb{E}U = 0$, which only can be true if

$$X - \frac{\mathbb{E}XY}{\mathbb{E}Y^2}Y = 0,$$

that is, X is a scalar multiple of Y. $\square$

Example 15.7. We can use the Cauchy-Schwartz inequality to prove one of the properties of correlation coefficient in Proposition 12.3.2. Namely, suppose X and Y are random variables, then $|\rho(X,Y)| \leqslant 1$. Moreover, $|\rho(X,Y)| = 1$ if and only if there are constants $a, b \in \mathbb{R}$ such that $X = a + bY$.

We will use normalized random variables as before, namely,

$$U := \frac{X - \mathbb{E}X}{\sqrt{\operatorname{Var} X}},$$
$$V := \frac{Y - \mathbb{E}Y}{\sqrt{\operatorname{Var} Y}}.$$

Then $\mathbb{E}U = \mathbb{E}V = 0, \mathbb{E}U^2 = \mathbb{E}V^2 = 1$. We can use the Cauchy-Schwartz inequality for U and V to see that

$$|\mathbb{E}UV| \leqslant \sqrt{\mathbb{E}U^2 \cdot \mathbb{E}V^2} = 1$$

and the identity holds if and only if $U = aV$ for some $a \in \mathbb{R}$.

Recall Equation (12.3.1)

$$\rho(X,Y) = \mathbb{E}(UV),$$

which gives the bound we need. Note that if $U = aV$, then

$$\frac{X - \mathbb{E}X}{\sqrt{\operatorname{Var} X}} = a\left(\frac{Y - \mathbb{E}Y}{\sqrt{\operatorname{Var} Y}}\right),$$

therefore

$$X = a\sqrt{\operatorname{Var} X}\left(\frac{Y - \mathbb{E}Y}{\sqrt{\operatorname{Var} Y}}\right) + \mathbb{E}X = a\frac{\sqrt{\operatorname{Var} X}}{\sqrt{\operatorname{Var} Y}}Y - a\frac{\sqrt{\operatorname{Var} X}}{\sqrt{\operatorname{Var} Y}}\mathbb{E}Y + \mathbb{E}X,$$

which completes the proof.

15.6. Jensen's inequality

Recall that a function $g : \mathbb{R} \longrightarrow \mathbb{R}$ is *convex* on $[a,b]$ if for each $x, y \in [a,b]$ and each $\lambda \in [0,1]$ we have

$$g(\lambda x + (1-\lambda)y) \leqslant \lambda g(x) + (1-\lambda) g(y).$$

Note that for a convex function g this property holds for any *convex linear combination* of points in $[a,b]$, that is,

$$\begin{aligned}
& g(\lambda_1 x_1 + \ldots + \lambda_n x_n) \leqslant \lambda_1 g(x_1) + \ldots + g(\lambda_n x_n), \\
& \lambda_1 + \ldots + \lambda_n = 1, \quad 0 \leqslant \lambda_1, \ldots, \lambda_n \leqslant 1, \\
& x_1, \ldots, x_n \in [a,b].
\end{aligned}$$

If g is twice differentiable, then we have a simple test to see if a function is convex, namely, g is convex if $g''(x) \geqslant 0$ for all $x \in [a, b]$. Geometrically one can show that if g is convex, then if we draw a line segment between any two points on the graph of the function, the entire segment lies above the graph of g, as we show formally bellow. A function g is *concave* if $-g$ is convex. Typical examples of convex functions are $g(x) = x^2$ and $g(x) = e^x$. Examples of concave functions are $g(x) = -x^2$ and $g(x) = \log x$. Convex and concave functions are always continuous.

Convex functions lie above tangents

Suppose $a < c < b$ and $g : [a, b] \longrightarrow \mathbb{R}$ be convex. Then there exist $A, B \in \mathbb{R}$ such that $g(c) = Ac + B$ and for all $x \in [a, b]$ we have $g(x) \geqslant Ax + B$.

Proof. For $a \leqslant x < c < y \leqslant b$ we can write c as a convex combination of x and y, namely, $C = \lambda x + (1 - \lambda) y$ with $\lambda = \frac{y-c}{y-x} \in [0, 1]$. Therefore

$$g(c) \leqslant \lambda g(x)) + (1 - \lambda) g(y)$$

which implies that

$$\frac{g(c) - g(x)}{c - x} \leqslant \frac{g(y) - g(c)}{y - c}.$$

Thus we can take

$$\sup_{x<c} \frac{g(c) - g(x)}{c - x} \leqslant A \leqslant \inf_{y>c} \frac{g(y) - g(c)}{y - c},$$

so that we have for all $x < y$ in $[a, b]$ that

$$g(x) \geqslant A(x - c) + g(c) = Ax + (g(c) - Ac).$$

□

Proposition 15.7 (Jensen's inequality)

Suppose X is a random variable such that $\mathbb{P}(a \leqslant X \leqslant b) = 1$. If $g : \mathbb{R} \longrightarrow \mathbb{R}$ is convex on $[a, b]$, then

$$\mathbb{E}g(X) \geqslant g(\mathbb{E}X).$$

If g is concave, then

$$\mathbb{E}g(X) \leqslant g(\mathbb{E}X).$$

Proof. If X is constant, then there is nothing to prove, so assume X is not constant. Then we have

$$a < \mathbb{E}X < b.$$

Denote $c := \mathbb{E}X$. Then

$$g(x) \geqslant AX + B \text{ and } g(\mathbb{E}X) = A\mathbb{E}X + B$$

for some $A, B \in \mathbb{R}$. Also note that

$$|g(X)| \leqslant |A||X| + |B| \leqslant |A \max\{|a|, |b|\} |X| + |B|,$$

so $\mathbb{E}|g(X)| < \infty$ and therefore $\mathbb{E}g(X)$ is well defined. Now we can use $AX + B \leqslant g(X)$ to see that

$$g(\mathbb{E}X) = A\mathbb{E}X + B = \mathbb{E}(AX + B) \leqslant \mathbb{E}g(X)$$

□

Example 15.8 (Arithmetic-geometric mean inequality)**.** Suppose a_1, ..., a_n are positive numbers, and X is a discrete random variable with the mass density

$$f_X(a_k) = \frac{1}{n} \text{ for } k = 1, ..., n.$$

Note that the function $g(x) = -\log x$ is a convex function on $(0, \infty)$. Jensen's inequality gives that

$$-\log\left(\frac{1}{n}\sum_{k=1}^{n} a_k\right) = -\log(\mathbb{E}X) \leqslant \mathbb{E}(-\log X) = -\frac{1}{n}\sum_{k=1}^{n}\log a_k.$$

Exponentiating this we get

$$\frac{1}{n}\sum_{k=1}^{n} a_k \geqslant \left(\prod_{k=1}^{n} a_k\right)^{1/n}.$$

Example 15.9. Suppose $p \geqslant 1$, then the function $g(x) = |x|^p$ is convex. Then

$$\mathbb{E}|X|^p \geqslant |\mathbb{E}X|^p$$

for any random variable X such that $\mathbb{E}X$ is defined. In particular,

$$\mathbb{E}X^2 \geqslant (\mathbb{E}X)^2,$$

and therefore $\mathbb{E}X^2 - (\mathbb{E}X)^2 \geqslant 0$.

15.7. Exercises

Exercise 15.1. Suppose we wire up a circuit containing a total of n connections. The probability of getting any one connection wrong is p. What can we say about the probability of wiring the circuit correctly? The circuit is wired correctly if all the connections are made correctly.

Exercise 15.2. Suppose $X \sim \operatorname{Exp}(\lambda)$. Using Markov's inequality estimate $\mathbb{P}(X \geqslant a)$ for $a > 0$ and compare it with the exact value of this probability.

Exercise 15.3. Suppose $X \sim \operatorname{Exp}(\lambda)$. Using Chebyshev's inequality estimate $\mathbb{P}(|X - \mathbb{E}X| \geqslant b)$ for $b > 0$.

Exercise 15.4. Suppose $X \sim \operatorname{Exp}(\lambda)$. Using Chernoff bounds estimate $\mathbb{P}(X \geqslant a)$ for $a > \mathbb{E}X$ and compare it with the exact value of this probability.

Exercise 15.5. Suppose $X > 0$ is a random variable such that $\operatorname{Var}(X) > 0$. Decide which of the two quantities is larger.

(A) $\mathbb{E}X^3$ or $(\mathbb{E}X)^3$?
(B) $\mathbb{E}X^{3/2}$ or $(\mathbb{E}X)^{3/2}$?
(C) $\mathbb{E}X^{2/3}$ or $(\mathbb{E}X)^{2/3}$?
(D) $\mathbb{E}\log(X+1)$ or $\log(\mathbb{E}X+1)$?
(E) $\mathbb{E}e^{X}$ or $e^{\mathbb{E}X}$?
(F) $\mathbb{E}e^{-X}$ or $e^{-\mathbb{E}X}$?

15.8. Selected solutions

Solution to Exercise 15.1: Let E_i denote the event that connection i is made correctly, so $\mathbb{P}(E_i^c) = p$. We do not anything beyond this (such as whether these events are dependent), so we will use Boole's inequality to estimate this probability as follows. The event we are interested in is $\cap_{i=1}^n E_i$.

$$\mathbb{P}\left(\bigcap_{i=1}^n E_i\right) = 1 - \mathbb{P}\left(\left(\bigcap_{i=1}^n E_i\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{i=1}^n E_i^c\right)$$
$$\overset{Boole}{\geqslant} 1 - \sum_{i=1}^n \mathbb{P}(E_i^c) = 1 - np.$$

Solution to Exercise 15.2: Markov's inequality gives

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}X}{a} = \frac{1}{a\lambda},$$

while the exact value is

$$\mathbb{P}(X \geqslant a) = \int_a^\infty \lambda e^{-\lambda x} dx = e^{-\lambda a} \leqslant \frac{1}{a\lambda}.$$

Solution to Exercise 15.3: We have $\mathbb{E}X = 1/\lambda$ and $\operatorname{Var} X = 1/\lambda^2$. By Chebyshev's inequality we have

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant b) \leqslant \frac{\operatorname{Var} X}{b^2} = \frac{1}{b^2\lambda^2}.$$

Solution to Exercise 15.4: recall first that

$$m_X(t) = \frac{\lambda}{\lambda - t} \text{ for } t < \lambda.$$

Using Chernoff bounds, we see

$$\mathbb{P}(X \geqslant a) \leqslant \min t > 0 \left(e^{-ta} m_X(t)\right) = \min_{t>0}\left(e^{-ta}\frac{\lambda}{\lambda - t}\right) \text{ for } t < \lambda.$$

To find the minimum of $e^{-ta}\frac{\lambda}{\lambda-t}$ as a function of t, we can find the critical point and see that it is $\lambda - 1/a > 0$ since we assume that $a > \mathbb{E}X = 1/\lambda$. Using this value for t we get

$$e^{-a\lambda} = \mathbb{P}(X \geqslant a) \leqslant a\lambda e^{1-a\lambda} = (a\lambda e) \cdot e^{-a\lambda} = (a\lambda e)\,\mathbb{P}(X \geqslant a).$$

Note that $a\lambda e \geqslant 1$.

Solution to Exercise 15.5(A): $\mathbb{E}X^3 > (\mathbb{E}X)^3$ since $(x^3)'' = x/3 > 0$ for $x > 0$.

Solution to Exercise 15.5(B): $\mathbb{E}X^{3/2} > (\mathbb{E}X)^{3/2}$ since $(x^{3/2})'' = \frac{3}{4\sqrt{x}} > 0$ for $x > 0$.

Solution to Exercise 15.5(C): $\mathbb{E}X^{2/3} < (\mathbb{E}X)^{2/3}$ since $(x^{2/3})'' = -\frac{2}{9x^{4/3}} < 0$ for $x > 0$.

Solution to Exercise 15.5(D): $\mathbb{E}\log(X+1) < \log(\mathbb{E}X+1)$ since $(\log(x))'' = -1/x^2 < 0$ for $x > 0$.

Solution to Exercise 15.5(E): $\mathbb{E}e^X > e^{\mathbb{E}X}$ since $(e^x)'' = e^x > 0$ for any x.

Solution to Exercise 15.5(F): $\mathbb{E}e^{-X} > e^{-\mathbb{E}X}$ since $(e^{-x})'' = e^{-x} > 0$ for any x.

Part 4

Applications of probability

CHAPTER 16

* Applications in Insurance and Actuarial Science

16.1 Introduction

Suppose that for a period, you face the risk of losing something that is unpredictable, and denote this potential loss by a random variable X. This loss may be the result of damages or injuries from (a) an automobile accident, (b) fire, theft, storm or hurricane at home, (c) premature death of head of the household, or (d) hospitalization due to an illness. Insurance allows you to exchange facing this potential loss for a fixed price or premium. It is one of the responsibilities of an actuary to assess the fair price given the nature of the risk. Actuarial science is a discipline that deals with events that are uncertain and their economic consequences; the concepts of probability and statistics provide for indispensable tools in measuring and managing these uncertainties.

16.2 The Pareto distribution

The Pareto distribution is commonly used to describe and model insurance losses. One reason is its flexibility to handle positive skewness or long distribution tails. It is possible for insurance losses to become extremely large, although such may be considered rare event. While there are several versions of the Pareto family of distributions, we consider the cumulative distribution function of X that follows a Type II Pareto:

$$F_X(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^{\alpha}, \quad \text{for } x > 0, \tag{16.0.1}$$

where $\alpha > 0$ is the shape or tail parameter and $\theta > 0$ is the scale parameter. If X follows such distribution, we write $X \sim \text{Pareto}(\alpha, \theta)$.

Example 16.2.1 Consider loss X with Pareto$(3, 100)$ distribution.

(a) Calculate the probability that loss exceeds 50.
(b) Given that loss exceeds 50, calculate the probability that loss exceeds 75.

Solution: From (16.0.1), we find

$$\mathbb{P}(X > 50) = 1 - F_X(50) = \left(\frac{100}{50+100}\right)^3 = \left(\frac{2}{3}\right)^3 = \frac{8}{27} = 0.2963.$$

and

$$\mathbb{P}(X > 75 | X > 50) = \frac{\mathbb{P}(X > 75)}{\mathbb{P}(X > 50)} = \left(\frac{150}{175}\right)^3 = \left(\frac{6}{7}\right)^3 = 0.6297.$$

 Typesetting date: September 2, 2020

By taking the derivative of (16.0.1), it can be shown that the probability density function of X is given by

$$f_X(x) = \frac{\alpha}{\theta}\left(\frac{\theta}{x+\theta}\right)^{\alpha+1}, \quad \text{for } x > 0. \tag{16.0.2}$$

Figure 16.0.1 depicts shapes of the density plot with varying parameter values of α and θ.

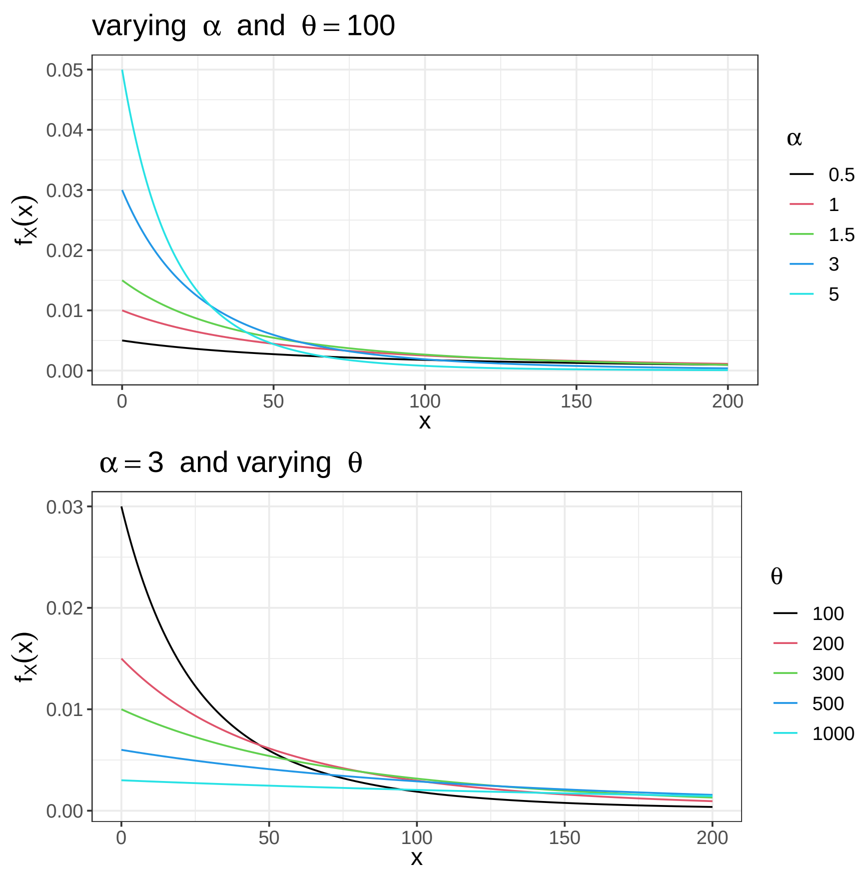


Figure 16.0.1. Density plots of the Pareto distribution.

The mean of the distribution can be shown as

$$\mathbb{E}(X) = \frac{\theta}{\alpha - 1}. \tag{16.0.3}$$

This mean exists provided $\alpha > 1$. For the variance, it can be shown that it has the expression

$$\text{Var}(X) = \frac{\alpha\theta^2}{(\alpha-1)(\alpha-2)}. \tag{16.0.4}$$

This variance exists provided $\alpha > 2$, is infinite for $1 < \alpha \leq 2$, and is otherwise undefined.

16.3 Insurance coverage modifications

In some instances, the potential loss that you face may be huge and unlimited. In this case, the cost of the insurance coverage may be burdensome. There are possible modifications to your insurance coverage so that the burden may be reduced. We introduce three possible modifications: (a) deductibles, (b) limits or caps, and (c) coinsurance.

These coverage modifications are a form of loss sharing between you, who is called the policyholder or insured, and the insurance company, which is also called the insurer. The effect is a reduced premium to the policyholder, and at the same time, because the policyholder shares in the loss, there is a perceived notion that this may alter the behavior of the policyholder. For instance, in the case of automobile insurance, the policyholder may be more careful about his or her driving behavior. Note that it is also possible to have an insurance coverage which is a combination of these three modifications.

16.3.1 Deductibles

In an *excess-of-loss* insurance contract, the insurance company agrees to reimburse the policyholder for losses beyond a pre-specified amount d. This amount d is referred to as the deductible of the contract. Given the loss is X, this amount is then shared between the policyholder, who is responsible for the first d amount, and the insurance company, which pays the excess if any. Thus, the policyholder is responsible for $\min(X, d)$ and the insurance company pays the excess, which is then equal to

$$X_I = X - \min(X, d) = \begin{cases} 0, & \text{if } X \leq d \\ X - d, & \text{if } X > d \end{cases}. \tag{16.0.5}$$

In general, we keep this notation, X_I, to denote the portion of X that the insurance company agrees to pay. The expected value of this can be expressed as

$$\mathbb{E}(X_I) = \mathbb{E}(X) - \mathbb{E}[\min(X, d)], \tag{16.0.6}$$

where $\mathbb{E}[\min(X, d)]$ is sometimes called the *limited expected value*. For any non-negative random variable X, it can be shown that

$$\mathbb{E}[\min(X, d)] = \int_0^d [1 - F_X(x)]\, dx. \tag{16.0.7}$$

This result can be proved as follows. Starting with

$$\begin{aligned} \mathbb{E}[\min(X, d)] &= \int_0^d x f_X(x) dx + \int_d^\infty d \cdot f_X(x) dx \\ &= \int_0^d x f_X(x) dx + d\,[1 - F_X(d)]\,, \end{aligned}$$

and applying integration by parts with $u = x$ and $dv = -f_X(x)dx$ so that $v = 1 - F_X(x)$, we have

$$\begin{aligned} \mathbb{E}[\min(X, d)] &= -\,x\,[1 - F_X(x)]\Big|_0^d + \int_0^d [1 - F_X(x)]\, dx + d\,[1 - F_X(d)]\,, \\ &= -d\,[1 - F_X(d)] + \int_0^d [1 - F_X(x)]\, dx + d\,[1 - F_X(d)] \\ &= \int_0^d [1 - F_X(x)]\, dx, \end{aligned}$$

which gives the desired result.

Example 16.3.1 Show that the limited expected value, with a deductible d, for a Pareto(α, θ) distribution has the following expression

$$\mathbb{E}[\min(X, d)] = \frac{\theta}{\alpha - 1}\left[1 - \left(\frac{\theta}{d+\theta}\right)^{\alpha-1}\right], \tag{16.0.8}$$

provided $\alpha \neq 1$.

Solution: From (16.0.1), we find

$$1 - F_X(x) = \left(1 + \frac{x}{\theta}\right)^{-\alpha}.$$

Applying the substitution $u = 1 + x/\theta$ so that $du = (1/\theta)dx$, we get

$$\mathbb{E}[\min(X, d)] = \int_1^{1+x/\theta} \theta u^{-\alpha} du.$$

Evaluating this integral, we get the result in (16.0.8).

Example 16.3.2 An insurance company offers an *excess-of-loss* contract against loss X. Assume that X has a Pareto distribution with mean 100 and variance 200000/3. You are given that $\alpha > 3$. The insurer's expected payment for this loss is 80. Calculate the deductible amount d.

Solution: First, we find the parameters of the distribution. From (16.0.3) and (16.0.4), we have two equations in two unknowns:

$$\frac{\theta}{\alpha - 1} = 100 \quad \text{and} \quad \frac{\alpha\theta^2}{(\alpha-1)(\alpha-2)} = 200000/3$$

This leads us to $\theta = 100(\alpha - 1)$ which results to a quadratic equation in α: $3\alpha^2 - 23\alpha + 40 = 0$. There are two possible solutions for α: either $\alpha = 5$ or $\alpha = 8/3$. Since we are given $\alpha > 3$, we use $\alpha = 5$ so that $\theta = 400$. Calculating the insurer's expected payment, we have from (16.0.6) and (16.0.8) the following:

$$\mathbb{E}(X_I) = 100\left(\frac{400}{400+d}\right)^4 = 80.$$

Solving for the deductible, we have $d = 22.95$.

16.3.2 Policy limits

An insurance contract with a policy limit, or cap, L is called a *limited* contract. For this contract, the insurance company is responsible for payment of X provided the loss does not exceed L, and if it does, the maximum payment it will pay is L. We then have

$$X_I = \min(X, L) = \begin{cases} X, & \text{if } X \leq L \\ L, & \text{if } X > L \end{cases} \tag{16.0.9}$$

Example 16.3.3 An insurance company offers a *limited* contract against loss X that follows a Pareto distribution. The parameters are assumed to be $\alpha = 3.5$ and $\theta = 250$. Calculate the company's expected payment for this loss if the limit $L = 5000$.

Solution: From (16.0.8), we have

$$\mathbb{E}(X_I) = \mathbb{E}[\min(X, 5000)] = \frac{250}{2.5}\left[1 - \left(\frac{250}{5000+250}\right)^{2.5}\right] = 99.95.$$

Example 16.3.4 Suppose loss X follows an exponential distribution with mean μ. An insurance company offers a *limited* contract against loss X. If the policy limit is $L = 500$, the company's expected payment is 216.17. If the policy limit is $L = 1000$, the company's expected payment is 245.42. Calculate the mean parameter μ.

Solution: For an exponential with mean parameter μ, its density can be expressed as $f_X(x) = (1/\mu)e^{(-x/\mu)}$, for $x > 0$. This gives a distribution function equal to $F_X(x) = 1 - e^{(-x/\mu)}$. From (16.0.7), we can show that

$$\mathbb{E}[\min(X, L)] = \mu\left[1 - e^{(-L/\mu)}\right].$$

From the given, this leads us to the following two equations: $216.17 = \mu\left[1 - e^{(-500/\mu)}\right]$ and $245.42 = \mu\left[1 - e^{(-1000/\mu)}\right]$. If we let $k = e^{(500/\mu)}$ and divide the two equations, we get the quadratic equation:

$$\frac{216.17}{245.42} = \frac{1-k}{1-k^2} \implies 1 + k = \frac{245.42}{216.17} \implies k == 0.1353102,$$

provided $k \neq 1$. When $k = 1$, μ is undefined, and therefore, $k = 0.1353102$. This gives $\mu = 249.9768 \approx 250$.

16.3.3 Coinsurance

In an insurance contract with coinsurance, the insurance company agrees to pay a fixed and pre-specified proportion of k for each loss. This proportion k must be between 0 and 100%. The company's payment for each loss is

$$X_I = kX, \tag{16.0.10}$$

where $0 < k \leq 1$. Therefore, the expected value of this payment is just a fixed proportion k of the average of the loss.

Example 16.3.5 An insurance company offers a contract with *coinsurance* of 80%. Assume loss X follows a Pareto distribution with $\theta = 400$ and $\mathbb{P}(X > 100) = 0.5120$.

(a) Calculate the company's expected payment for each loss.
(b) Suppose the company agrees to replace the contract with an *excess-of-loss* coverage. Find the deductible d so that the company has the same expected payment.

Solution: From (16.0.1), we have

$$\mathbb{P}(X > 100) = \left(\frac{400}{500}\right)^{\alpha} = 0.5120.$$

This gives $\alpha = \ln(0.5120)/\ln(0.80) = 3$. Thus, the company's expected payment for each loss is $0.80 \times (400/2) = 160$. For part (b), we use equations (16.0.6) and (16.0.8) to solve for d. We have, with deductible d, the following

$$\mathbb{E}(X_I) = \frac{400}{2}\left(\frac{400}{d+400}\right)^2 = 160.$$

Solving for the deductible, we get $d = (400/(.8)^{0.5}) - 400 = 47.2136$.

16.3.4 Combination

It is not uncommon to find insurance contracts that combine the three different coverage modifications described in the previous sections. In particular, consider the general situation where we have a combination of a deductible d, policy limit L, and coinsurance k. In this case, the insurance contract will pay for a proportion k of the loss X, in excess of the deductible d, subject to the policy limit of L. The company's payment for each loss can be written as

$$X_I = \begin{cases} 0, & \text{if } X \le d, \\ k(X-d), & \text{if } d < X \le L, \\ k(L-d), & \text{if } X > L, \end{cases} \tag{16.0.11}$$

where $0 < k \le 1$. This payment can be also be expressed as

$$X_I = k\left[\min(X, L) - \min(X, d)\right], \tag{16.0.12}$$

where its expectation can be evaluated using $\mathbb{E}(X_I) = k\left[\mathbb{E}(\min(X, L)) - \mathbb{E}(\min(X, d))\right]$.

Example 16.3.6 An insurer offers a proportional *excess-of-loss* medical insurance policy. The policy is subject to a coinsurance of 90% with a deductible of 25, but has no policy limit. Assume medical loss X follows an exponential distribution with mean 800. Calculate the expected reimbursement the policyholder will receive in the event of a loss.

Solution: Since there is no policy limit, the reimbursement can be expressed as

$$X_I = k\left[X - \min(X, d)\right]$$

For X that has an exponential distribution with mean μ, we have

$$\mathbb{E}[\min(X, d)] = \mu\left[1 - e^{(-d/\mu)}\right]$$

This leads us to the average reimbursement as $\mathbb{E}(X_I) = k\mu e^{-d/\mu} = 0.90 * 800 * e^{-25/800} = 697.85$. In effect, the policyholder can expect to be reimbursed $(697.85/800) = 87\%$ of each loss.

16.4 Loss frequency

In practice, the insurance company pays for damages or injuries to insured only if the specified insured event happens. For example, in an automobile insurance, the insurance company will pay only in the event of an accident. It is therefore important to consider also the probability that an accident occurs. This refers to the loss frequency.

For simplicity, start with the Bernoulli random variable I, which indicates an accident, or some other insured event, occurs. Assume that I follows a Bernoulli distribution with p denoting the probability the event happens, i.e., $\mathbb{P}(I = 1) = p$. If the insured event happens, the amount of loss is X, typically a continuous random variable. This amount of loss is referred to as the loss severity. Ignoring any possible coverage modifications, the insurance company will pay 0, if the event does not happen, and X, if the event happens. In effect, the insurance claim can be written as the random variable

$$Y = X \cdot I = \begin{cases} 0, & \text{if } I = 0 \ \ \text{(event does not happen)} \\ X, & \text{if } I = 1 \ \ \text{(event happens)} \end{cases} \tag{16.0.13}$$

The random variable Y is a mixed random variable and will be called the *insurance claim*. It has a probability mass at 0 and a continuous distribution for positive values. By conditioning on the Bernoulli I, it can be shown that the cumulative distribution function of Y has the expression

$$F_Y(y) = \begin{cases} 1 - p, & \text{if } y = 0 \\ (1-p) + pF_X(y), & \text{if } y > 0 \end{cases} \tag{16.0.14}$$

This result can be shown by using the law of total probability.

Denote the mean of X by μ_X and its standard deviation by $\sigma_X > 0$. It can be shown that the expected value of Y has the expression

$$\mathbb{E}(Y) = p\mu_X \tag{16.0.15}$$

and the variance is

$$\mathrm{Var}(Y) = p(1-p)\mu_X^2 + p\sigma_X^2. \tag{16.0.16}$$

Example 16.4.1 Consider an insurance contract that covers an event with probability 0.25 that it happens, and the loss severity X has a Pareto$(3, 1000)$ distribution.

(a) Calculate the probability that the insurance contract will pay an amount less than or equal to 500.
(b) Calculate the probability that the insurance contract will pay an amount larger than 750.
(c) Calculate the expected value and variance of the insurance claim.

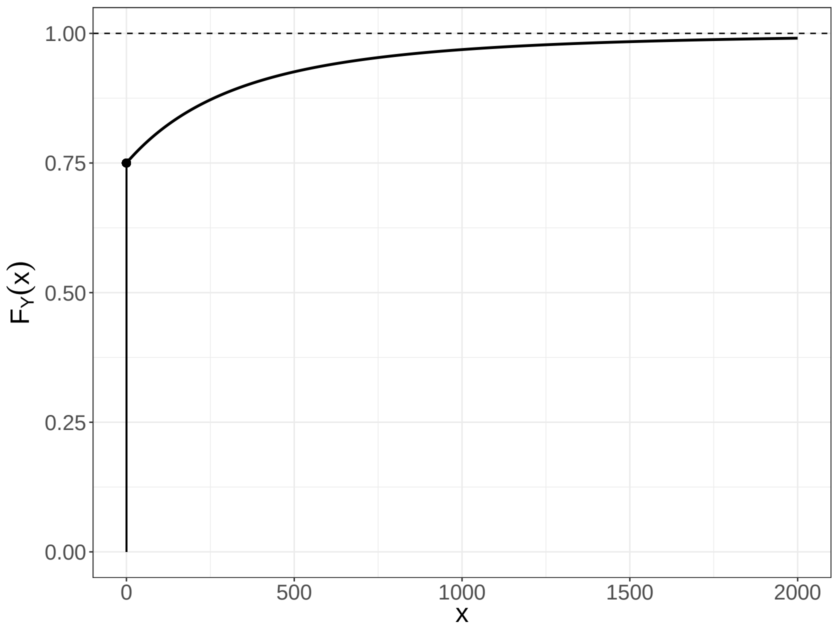


Figure 16.0.2. Cumulative distribution function of insurance claim random variable Y.

Solution: From (16.0.14), we have

$$\begin{aligned}\mathbb{P}(Y \leq 500) &= \mathbb{P}(Y = 0) + \mathbb{P}(0 < Y \leq 500) \\ &= (1 - 0.25) + 0.25 F_X(500) \\ &= 0.75 + 0.25 \left[1 - \left(\frac{1000}{500 + 1000}\right)^3\right] = 0.9259.\end{aligned}$$

For part (b), we use the complement of the cumulative distribution function:

$$\begin{aligned}\mathbb{P}(Y > 600) &= 1 - [(1 - 0.25) + 0.25 F_X(500)] \\ &= 0.25 - 0.25 \left[1 - \left(\frac{5}{8}\right)^3\right] = 0.0610.\end{aligned}$$

For the claim severity X, the mean is $\mu_X = 500$ and the variance is $\sigma_X^2 = \frac{3}{2}(1000^2)$. Therefore, from (16.0.15) and (16.0.16), we have

$$\mathbb{E}(Y) = 0.25(500) = 125$$

and

$$\text{Var}(Y) = 0.25(0.75)(500^2) + 0.25 \cdot \frac{3}{2}(1000^2) = 421,875.$$

Figure 16.0.2 shows the graph of the cumulative distribution function of Y using the parameters of this example.

Example 16.4.2 An insurance company models claims using $Y = X \cdot I$, where I is the indicator of the event happening with probability 0.20. The amount of loss X has a discrete distribution given as follows:

x	20	100	500	1000
$\mathbb{P}(X = x)$	0.30	0.50	0.15	0.05

(a) Calculate the probability that the insurance claim will be below 150.
(b) Calculate the expected value of the insurance claim.
(c) Calculate the standard deviation of the insurance claim.

Solution: We have

$$\begin{aligned}\mathbb{P}(Y \leq 150) &= \mathbb{P}(Y = 0) + 0.20 * \mathbb{P}(X \leq 150) \\ &= (1 - 0.20) + 0.20(0.30 + 0.50) = 0.96\end{aligned}$$

For part (b), first we find $\mathbb{E}(X) = 182.50$ so that $\mathbb{E}(Y) = 0.20 * 182.50 = 36.5$. For part (c), we find $\sigma_X^2 = 59381.25$ so that

$$\mathrm{Var}(Y) = 0.20(0.80) * 182.5^2 + 0.20 * 59381.25 = 17205.25.$$

Therefore, the standard deviation is $\sqrt{17205.25} = 131.17$.

16.5 The concept of risk pooling

Insurance is based on the idea of pooling several individuals willing to exchange their risks. Consider the general situation where there are n individuals in the pool. Assume that each individual faces the same loss distribution but the eventual losses from each of them will be denoted by $Y_1, Y_2, \ldots, Y_n$. In addition, assume these individual losses are independent. The total loss arising from this pool is the sum of all these individual losses, $S_n = Y_1 + Y_2 + \cdots + Y_n$. To support funding this total loss, each individual agrees to contribute an equal amount of

$$P_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} Y_k, \tag{16.0.17}$$

which is the average loss.

Notice that the contribution above is still a random loss. Indeed in the absence of risk pooling where there is a single individual in the pool, that individual is responsible for his or her own loss. However, when there is sufficiently large enough number of individuals, this average contribution becomes more predictable as shown below.

Assume that the expected value and variance of each loss are given, respectively, by

$$\mathbb{E}(Y_k) = \mu \quad \text{and} \quad \mathrm{Var}(Y_k) = \sigma^2.$$

For the pool of n individuals, the mean of the average loss is then

$$\mathbb{E}(P_n) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(Y_k) = \mu,$$

which is exactly equal to the mean of each individual loss. However, what is interesting is the variability is reduced as shown below:

$$\text{Var}(P_n) = \frac{1}{n}\sum_{k=1}^{n} \text{Var}(Y_k) = \frac{1}{n}\sigma^2.$$

The variance is further reduced as the number of individuals in the pool increases. As discussed in Chapter 14, one version of the law of large numbers is the SLLN (Strong Law of Large Numbers), that accordingly, $P_n \to \mu$ as $n \to \infty$. In words, the unpredictable loss for a single individual becomes much more predictable. This is sometimes referred to as the basic law of insurance. The origins of insurance can be traced back from the idea of pooling the contributions of several for the indemnification of losses against the misfortunes of the few.

In principle, the insurance company acts as a third party that formally makes this arrangement. The company forms a group of such homogeneous and independent risks, and is responsible for collecting the individual contributions, called the premium, as well as disbursing payments when losses occur. There are additional responsibilities of the insurance company such as ensuring enough capital to cover future claims, however, such are beyond the scope here.

Figure 16.0.3 depicts the basic law of insurance. In all situations, the mean of the average loss are the same for all cases. The variability, however, is reduced with increasing number of policyholders. As you probably suspect from these graphs, the average loss is about 1000, which is sometimes referred to as the actuarially fair premium. However, please bear in mind that there are conditions for the basic law of insurance to effectively work:

- Losses must be unpredictable.
- The individual risks must be independent.
- The individuals insured must be considered homogeneous, that is, they share common risk characteristics.
- The number of individuals in the pool must be sufficiently large.

Example 16.5.1 A company insures a group of 200 homogeneous and independent policyholders against an event that happens with probability 0.18. If the event happens and therefore a loss occurs, the amount of loss X has an exponential distribution with mean 1250.

(a) Calculate the mean of the total loss arising from the group.
(b) Calculate the variance of the total loss arising from the group.
(c) Using the Central Limit Theorem, estimate the probability that the total loss will exceed 60,000.

Solution: The total loss can be written as $S_{200} = \sum_{k=1}^{200} Y_k = \sum_{k=1}^{200} I_k X_k$, where I_k is the Bernoulli random variable for the event to happen and X_k is the loss, given an event occurs. For each k, we find that

$$\mathbb{E}(Y_k) = p\mu_X = 0.18 * 1250 = 225$$

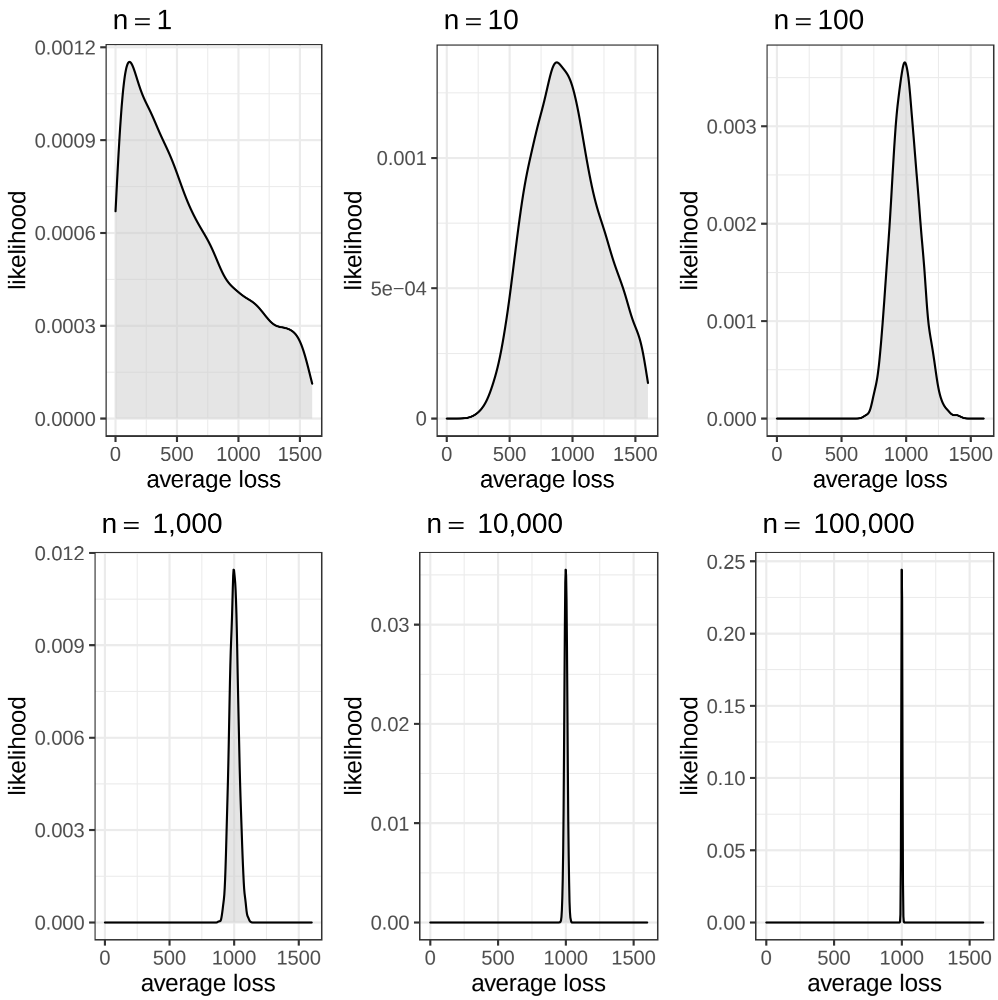


Figure 16.0.3. Distributions of the average loss as the number of policyholders increases.

and

$$\text{Var}(Y_k) = p(1-p)\mu_X^2 + p\sigma_X^2 = 0.18 * .82 * 1250^2 + 0.18 * (1250^2) = 511875$$

The mean and variance of the total variance are, respectively, given by

$$\mathbb{E}(S_{200}) = 200 * 225 = 45000 \quad \text{and} \quad \text{Var}(S_{200}) = 200 * 511875 = 102375000$$

The probability that the total loss will exceed 60,000 can be estimated as follows:

$$\mathbb{P}(S_{200} > 60000) \approx \mathbb{P}(Z > (60000 - 45000)/\sqrt{102375000}) = \mathbb{P}(Z > 1.48),$$

where Z denotes a standard normal random variable. From Table 1 on page 114, we get 0.06944.

16.6 Exercises

Exercise 16.1. Prove equations (16.0.3) and (16.0.4).

Exercise 16.2. Suppose insurance loss X has a Pareto(α, θ) distribution. You are given $\theta = 225$ and $\mathbb{P}(X \leq 125) = 0.7621$. Calculate the probability that loss will exceed 200, given it exceeds 125.

Exercise 16.3. Find an expression for the limited expected value, with a deductible d, for a Pareto(α, θ) distribution when $\alpha = 1$.

Exercise 16.4. An insurance company pays for a random loss X subject to a deductible amount of d, where $0 < d < 10$. The loss amount is modeled as a continuous random variable with density function

$$f_X(x) = \frac{1}{50}x, \quad \text{for } 0 < x \leq 10.$$

The probability that given the company will make a payment, it will pay less than or equal to 5 is 0.4092. Calculate d.

Exercise 16.5 An insurer offers a *limited* contract against loss X that follows an exponential distribution with mean 500. The limit of the contract is $L = 1000$. Calculate the probability that the loss will not reach the limit L.

Exercise 16.6 A company insures a loss X with a coinsurance of 90%. Loss X follows a distribution that is uniform on $(0, u)$, for $u > 0$. The expected payment for each loss for the company is 900. Calculate the expected amount of each loss the buyer of this policy is responsible for.

Exercise 16.7 An insurer offers a comprehensive medical insurance contract with a deductible of 25, subject to a coinsurance of 90% and a policy limit of 2000. Medical loss X follows an exponential distribution with mean 800.

(a) Calculate the expected reimbursement the policyholder will receive in the event of a loss.
(b) The insurer wants to reduce this expected reimbursement to 575 by adjusting only the level of deductible d. What amount d is needed to achieve this?

Exercise 16.8 Consider an insurance contract that covers an event with probability 0.10 that it happens, and the loss severity X has a Pareto$(2.5, 400)$ distribution.

(a) Calculate the probability that the insurance claim will exceed 150.
(b) Given the claim exceeds 150, calculate the probability that it is less than or equal to 400.

Exercise 16.9 A company insures a group of 150 homogeneous and independent policyholders against an event that happens with probability 0.10. If the event happens, the amount of loss X has a uniform distribution on $(0, 1000]$. The company collects a premium of 55 from each policyholder. Estimate the probability that the total premium collected will not be sufficient to support the total loss.

16.7 Solutions to exercises

Exercise 16.1: In writing the expressions for the first two moments, rewrite the density function as

$$f_X(x) = \frac{\alpha}{\theta}\left(1 + \frac{x}{\theta}\right)^{-\alpha-1}$$

and use the substitution $u = 1 + (x/\theta)$ in the integrals.

Exercise 16.2: From the given, we find $\alpha = 3.25$. This gives us 0.5320 for the answer.

Exercise 16.3: From (16.0.1), when $\alpha = 1$, we have $F_X(x) = 1-(\theta/(x+\theta))$. Use (16.0.7) to arrive at the following expression for the limited expected value: $\mathbb{E}[\min(X, d)] = \theta \ln(1 + \frac{d}{\theta})$.

Exercise 16.4: Observe that the insurer will pay only beyond the deductible d. So, it will pay no more than 5 if the loss does not reach $(d + 5)$. The given probability is therefore $\mathbb{P}(X \leq d + 5|X > d) = 0.4092$. It can be shown that $F_X(x) = x^2/100$. This gives $d = 1.5$.

Exercise 16.5: $\mathbb{P}(X \leq L) = 0.865$.

Exercise 16.6: The expected loss is 1000 for which insurer is responsible for 900. Therefore, the policyholder is responsible for 100, on the average.

Exercise 16.7: The expected reimbursement is $\mathbb{E}(X_I) = k\mu[e^{-d/\mu} - e^{-L/\mu}] = 0.90 * 800 * [e^{-25/800} - e^{-2000/800}] = 638.75$. To arrive at the desired d, we have $d = -\mu \cdot \log((575/(0.90 * 800) - e^{-2000/800}) = 101.63$.

Exercise 16.8: To calculate $\mathbb{P}(Y > 150)$, we can use direct application of (16.0.14) or use the law of total probability, i.e., $\mathbb{P}(Y > 150|I = 0)\mathbb{P}(I = 0) + \mathbb{P}(Y > 150|I = 1)\mathbb{P}(I = 1) = \mathbb{P}(Y > 150|I = 1)\mathbb{P}(I = 1)$. The first term vanishes because if event does not happen, then there is 0 insurance claim. Thus, using the property of Pareto, we have $\mathbb{P}(Y > 150) = 0.10 * (1 + (150/400))^{-2.5} = 0.451$. For part (b), the required probability is

$$\begin{aligned}\mathbb{P}(Y \leq 400|Y > 150) &= \frac{\mathbb{P}(150 \leq Y < 400)}{\mathbb{P}(Y > 150)} = \frac{\mathbb{P}(Y > 150) - \mathbb{P}(Y > 400)}{\mathbb{P}(Y > 150)} \\ &= 1 - \frac{\mathbb{P}(Y > 400)}{\mathbb{P}(Y > 150)} = 1 - \frac{0.10 * (1 + (400/400))^{-2.5}}{0.10 * (1 + (150/400))^{-2.5}} = 0.608.\end{aligned}$$

Exercise 16.9: The mean of the total losses is $\mathbb{E}(S_{150}) = 150 * 50 = 7500$ and the variance is $\text{Var}(S_{150}) = 150 * 30833.33 = 4625000$. Total premium collected is $55 * 150 = 8250$. The required probability then is

$$\begin{aligned}\mathbb{P}(S_{150} > 8250) &\approx \mathbb{P}(Z > (8250 - 7500)/\sqrt{4625000}) \\ &= \mathbb{P}(Z > 0.35) = 1 - \mathbb{P}(Z \leq 0.35) = 1 - 0.63307 = 0.36693.\end{aligned}$$

CHAPTER 17

* Applications of probability in finance

17.1. Coin toss games

17.1.1. The simple coin toss game. Suppose, as in Example 4.8, that we toss a fair coin repeatedly and independently. If it comes up heads, we win a dollar, and if it comes up tails, we lose a dollar. Unlike in Chapter 3 , we now can describe the solution using sums of independent random variables. We will use the partial sums process introduced in Definition 14.1 in Chapter 14

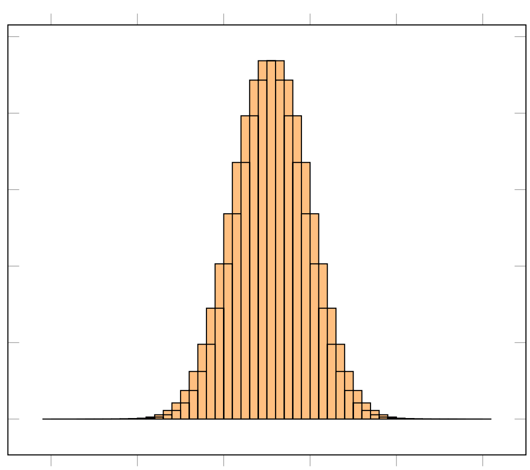
$$S_n = \sum_{i=1}^{n} X_i,$$

where $X_1, X_2, X_3, \ldots$ are i.i.d. random variables with the distribution $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. Then S_n represents the total change in the number of dollars that we have after n coin tosses: if we started with $\$M$, we will have $M + S_n$ dollars after n tosses. The name *process* is used because the amount changes over time, and *partial sums* is used because we compute S_n before we know what is the final outcome of the game. The process S_n is also commonly called *the simple random walk*.

The Central Limit Theorem tells us that S_n is approximately distributed as a normal random variable with mean 0 and variance n, that is,

$$M_n = M + S_n \sim M + \sqrt{n}Z \sim \mathcal{N}(M, n)$$

and these random variables have the distribution function $F(x) = \Phi\left(\frac{x-M}{\sqrt{n}}\right)$.



 Typesetting date: September 2, 2020

17.1.2. The coin toss game stopped at zero. Suppose the game is modified so that it is stopped when the amount of money reaches zero. Can we compute the probability distribution function of M_n, the amount of money after n coin tosses?

A useful trick, called the *Reflection Principle*, tells us that the probability to have x dollars after n coin tosses is

$$\mathbb{P}(M + S_n = x) - \mathbb{P}(M + S_n = -x) \quad \text{if } x > 0$$

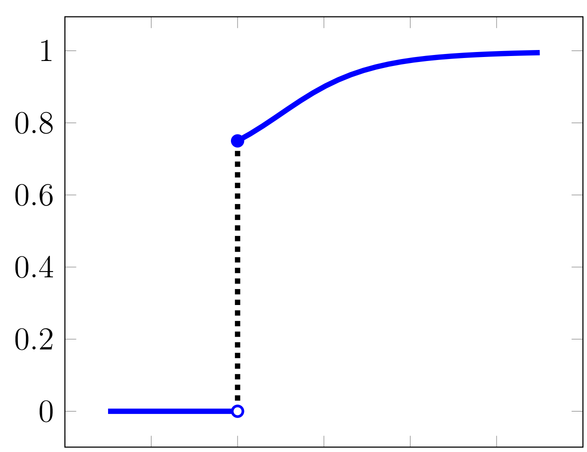
To derive this formula, we again denote by M_n the amount of money we have after n coin tosses. Then

$$\begin{aligned}\mathbb{P}(M_n = x) &= \mathbb{P}(M + S_n = x, M + S_k > 0 \text{ for all } k = 1, 2, ..., n)\\ &= \mathbb{P}(M + S_n = x) - \mathbb{P}(M + S_n = x, M + S_k = 0 \text{ for some } k = 1, 2, ..., n)\\ &= \mathbb{P}(M + S_n = x) - \mathbb{P}(M + S_n = -x, M + S_k = 0 \text{ for some } k = 1, 2, ..., n)\\ &= \mathbb{P}(M + S_n = x) - \mathbb{P}(M + S_n = -x).\end{aligned}$$

This, together with the Central Limit Theorem, implies that the cumulative probability distribution function of M_n can be approximated by

$$F(x) = \begin{cases} \Phi\left(\frac{x-M}{\sqrt{n}}\right) + \Phi\left(\frac{-x-M}{\sqrt{n}}\right) & \text{if } x \geqslant 0 \\ 0 & \text{otherwise.}\end{cases}$$

The following graph shows the approximate shape of this function.



Note that this function is discontinuous as the jump at zero represents the probability that we have lost all the money by the time n, that is,

$$\mathbb{P}(M_n = 0) \approx 2\Phi\left(-\frac{M}{\sqrt{n}}\right)$$

If we consider the limit $n \to \infty$, then $\mathbb{P}(M_n = 0) \xrightarrow[n\to\infty]{} 2\Phi(0) = 1$.

This proves that in this game all the money will be eventually lost with probability one. In fact, this conclusion is similar to the conclusion in Example 4.9.

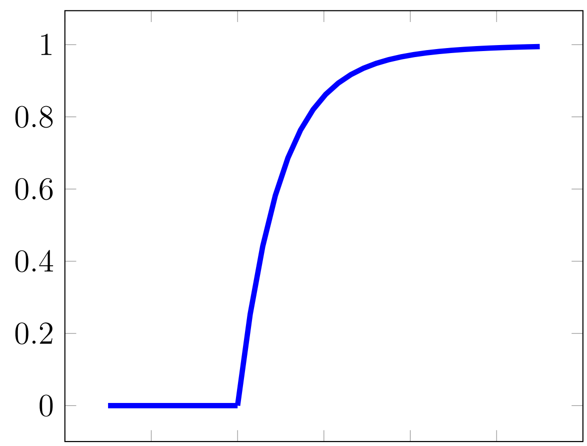
17.1.3. The coin toss game with borrowing at zero. Suppose now that the game is modified so that each time when we hit zero, instead of stopping, we borrow \$1 and continue playing. Another form of the *Reflection Principle* implies that the probability to have x dollars is

$$\mathbb{P}(M_n = x) = \mathbb{P}(M + S_n = x) + \mathbb{P}(M + S_n = -x) \quad \text{if } x > 0.$$

This formula is easy to explain because in this game the amount of money can be expressed as $M_n = |M + S_n|$. The Central Limit Theorem tells us that the cumulative probability distribution function of M_n can be approximated by

$$F(x) = \begin{cases} \Phi\left(\frac{x-M}{\sqrt{n}}\right) - \Phi\left(\frac{-x-M}{\sqrt{n}}\right) & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

The following graph shows the approximate shape of this function.



17.1.4. Probability to win N dollars before reaching as low as L dollars. Continuing the simple coin toss game, suppose now that L, M and N are integers such that $L < M < N$. If we start with \$$M$, what is the probability that we will get up to \$$N$ before we go as low as \$$L$? As in Chapter 3, we are interested in finding the function

$$y(x) = \mathbb{P}(\text{ winning \$}N \text{ before reaching \$}L \mid M = x)$$

which satisfies $N - L + 1$ linear equations

$$y(x) = \begin{cases} 0 & \text{if } x = L \\ \ldots & \\ \frac{1}{2}(y(x+1) + y(x-1)) & \text{if } L < x < N \\ \ldots & \\ 1 & \text{if } x = N \end{cases}$$

In general, in more complicated games, such a function is called a *harmonic function* because its value at a given x is the average of the neighboring values. In our game we can compute that $y(x)$ is a linear function with slope $\frac{1}{N-L}$ which gives us the formula

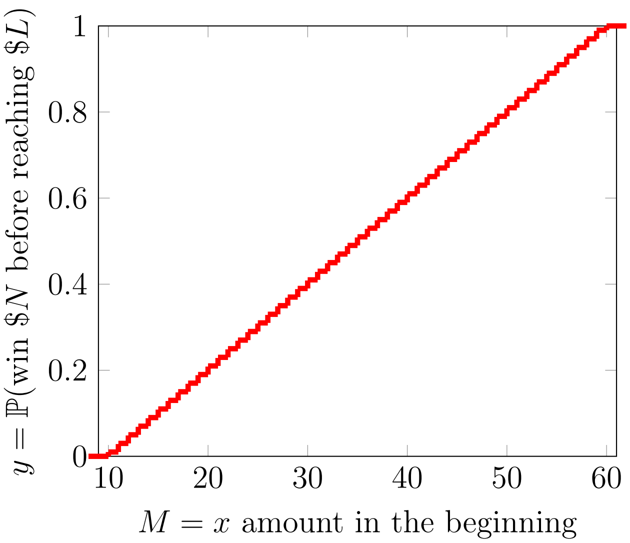
$$y(x) = \frac{x - L}{N - L}$$

and the final answer: with probability

$$\mathbb{P}(\text{ winning } \$N \text{ before reaching } \$L \mid M = x) = y(M) = \frac{\mathbf{M - L}}{\mathbf{N - L}} \tag{17.1.1}$$

we win $\$N$ before going as low as $\$L$ if we begin with $\$M$. Formula (17.1.1) applies in general to Gambler's Ruin problems, a particular case of which we consider in this section.

The following graph shows $y(x) = \frac{x-L}{N-L}$, the probability to win $\$N = \$60$ before reaching as low as $\$L = \$10$, in a game when $M_{n+1} = M_n \pm \$1$ with probability 1/2 at each step.



17.1.5. Expected playing time. Suppose we play the same simple coin toss game as in the previous section, and we would like to compute the expected number of coin tosses needed to complete the game. If we denote this expected number by $T(x)$, we will have a system of $N - L + 1$ linear equations

$$\mathbb{E}T(x) = \begin{cases} 0 & \text{if } x = L \\ \dots & \\ 1 + \frac{1}{2}\left(\mathbb{E}T(x+1) + \mathbb{E}T(x-1)\right) & \text{if } L < x < N \\ \dots & \\ 0 & \text{if } x = N \end{cases}$$

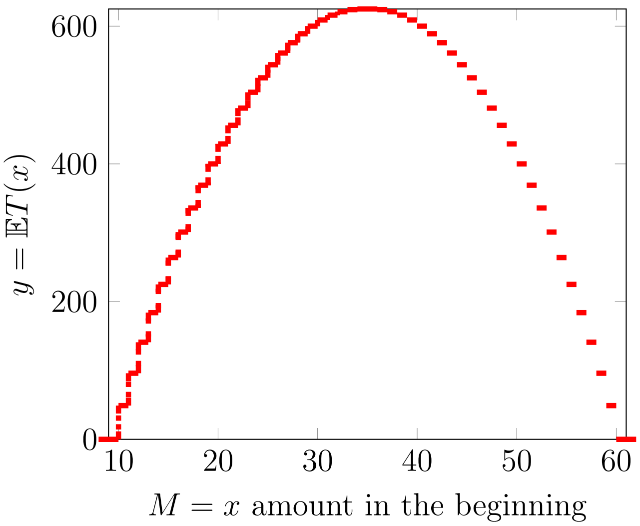
These equations have a unique solution given by the formula

$$\mathbb{E}T(x) = (x - L)(N - x)$$

and the final answer: the expected number of coin tosses is

$$\mathbb{E}\mathbf{T}(\mathbf{M}) = (\mathbf{M} - \mathbf{L})(\mathbf{N} - \mathbf{M}). \tag{17.1.2}$$

The following graph shows $\mathbb{E}T(x) = (x - L)(N - x)$, the expected number of coin tosses to win $\$N = \$60$ before reaching as low as $\$L = \$10$, in a game when $M_{n+1} = 2M_n$ or $M_{n+1} = \frac{1}{2}M_n$ with probability 1/2 at each step.



17.1.6. Doubling the money coin toss game. Let us now consider a game in which we begin with $\$M$ dollars, toss a fair coin repeatedly and independently. If it comes up heads, we *double our money*, and if it comes up tails, we *lose half of our money*. If we start with $\$M$, what is the probability that we will get up to $\$N$ before we go as low as $\$L$?

To answer this question, we first should notice that our money M_n after n coin tosses is given as a partial product process $M_n = M \cdot Y_1 \cdot Y_2 \cdot \ldots \cdot Y_n$, where $Y_1, Y_2, Y_3, \ldots$ are i.i.d. random variables with the distribution $\mathbb{P}(Y_i = 2) = \mathbb{P}(Y_i = \frac{1}{2}) = \frac{1}{2}$. If again we write $y(x) = \mathbb{P}($ winning $\$N$ before reaching $\$L)$, then

$$y(x) = \begin{cases} 0 & \text{if } x = L \\ \ldots & \\ \frac{1}{2}(y(2x) + y(\frac{1}{2}x)) & \text{if } L < x < N \\ \ldots & \\ 1 & \text{if } x = N \end{cases}$$

This function is linear if we change to the logarithmic variable $\log(x)$, which gives us the answer:

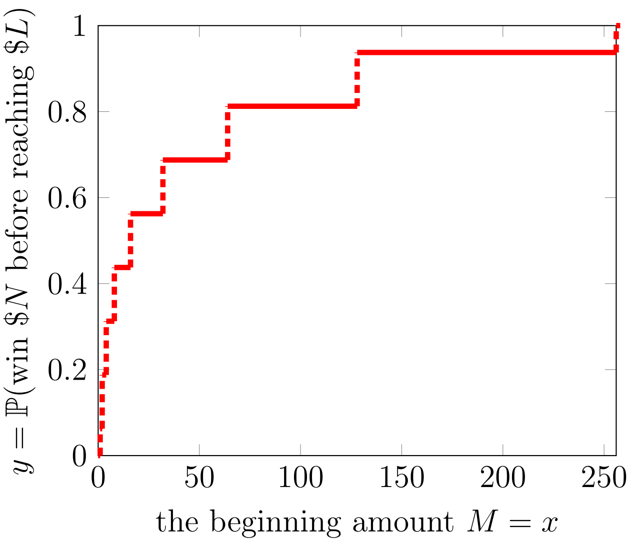
$$\mathbb{P}(\text{ winning } \$N \text{ before reaching } \$L) \approx \frac{\log(M/L)}{\log(N/L)}$$

This answer is approximate because, according to the rules, we can only have capital amounts represented by numbers $M2^k$, where k is an integer, and L, M, N maybe only approximately equal to such numbers. The exact answer is

$$\mathbb{P}(\text{ winning } \$N \text{ before reaching } \$L \mid M = x) = \frac{\mathbf{l}}{\mathbf{l} + \mathbf{w}}, \tag{17.1.3}$$

where l is the number of straight losses needed to reach $\$L$ from $\$M$ and w is the number of straight wins needed to reach $\$N$ from $\$M$. Equation (17.1.3) is again the general formula for Gambler's Ruin problems, the same as in Equation (17.1.1).

The following graph shows the probability to win $\$N = \$256$ before reaching as low as $\$L = \$1$ in a game when $M_{n+1} = 2M_n$ or $M_{n+1} = \frac{1}{2}M_n$ with probability $1/2$ at each step.


$y = \mathbb{P}$(win \$$N$ before reaching \$$L$)
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
the beginning amount $M = x$

17.2. Exercises on simple coin toss games

Exercise 17.1. In Subsection 17.1.1, what is the approximate distribution of $M_n - M_k$?

Exercise 17.2. In Subsection 17.1.1, find $\mathrm{Cov}(M_k, M_n)$.
Hint: assume $n > k$ and write $M_n = M_k + M_n - M_k = M_k + (S_n - S_k)$.

Exercise 17.3. Consider the game in which $M_n = Me^{\sigma S_n}$. Describe the rules of this game.

Exercise 17.4. In the game in Exercise 17.3, find $\mathbb{E}M_n$, $\mathbb{E}M_n^2$, $\mathrm{Var}(M_n)$.

Exercise 17.5. In the game in Exercise 17.3, how M_n and M_k are related?
Hint: assume $n > k$ and write $M_{n+1} = M_n \frac{M_{n+1}}{M_n}$. Also consider $M_n = M_k \frac{M_n}{M_k}$.

Exercise 17.6. Following Exercise 17.4, find $\mathrm{Cov}(M_n, M_k)$.

Exercise 17.7. In the game in Exercise 17.3, find the probability to win $\$N$ before reaching as low as $\$L$.

Exercise 17.8. In the game in Exercise 17.7, find the expected playing time.

Exercise 17.9. Following Exercise 17.3, use the normal approximation (as in the Central Limit Theorem) to find an approximate distribution of M_n. Then use this distribution to find approximate values of $\mathbb{E}M_n$, $\mathbb{E}M_n^2$, $\mathrm{Var}(M_n)$.

Exercise 17.10. Following Exercise 17.6, use the normal approximation (as in the Central Limit Theorem) to find the approximate value of $\mathrm{Cov}(M_n, M_k)$.

Exercise 17.11. Compare quantities in Exercises 17.4 and 17.9: which ones are larger and which ones are smaller? In which case a normal approximation gets better for larger n, and in which case it gets worse? If $n \to \infty$, how does σ need to behave in order to have an accurate normal approximation?

17.3. Problems motivated by the American options pricing

Problem 17.1. Consider the following game: a fair dice is thrown once and the player can either stop the game and receive the amount of money equals the outcome of the dice, or the player can decide to throw the dice the second time, and then receive the amount of money equals the outcome of the dice on this second throw. Compute the maximal expected value of the payoff and the corresponding optimal strategy.

Problem 17.2. Compute the maximal expected value of the payoff and the corresponding optimal strategy in the following game. A fair dice is thrown 3 times.

- After each throw except for the 3rd one, the player can either stop the game or continue.
- If the player decides to stop, then he/she receives the amount of money, which equals the current outcome of the dice (between 1 and 6).
- If the game is continued up to and including the 3rd throw, the player receives the amount of money, which equals to the outcome of the dice on the 3rd throw.

Problem 17.3.

(1) Compute the maximal expected value of the payoff and the corresponding optimal strategy in the same game as in Problem 17.2, but when up to 4, or 5, or 6 throws are allowed.
(2) Compute the maximal expected value of the payoff and the corresponding optimal strategy in the same game as in Problem 17.2, when an unlimited number of throws are allowed.

Problem 17.4. Let us consider a game where at each round, if you bet $\$x$, you get $\$2x$, if you win and $0, if you lose. Let us also suppose that at each round, the probability of winning equals to the probability of losing and is equal to 1/2. Additionally, let us assume that the outcomes of every round are independent.

In such settings, let us consider the following *doubling strategy*. Starting from a bet of $1 in the first round, you stop if you win or you bet twice as much if you lose. In such settings, if you win for the first (and only) time in the nth round, your cumulative winning is $\$2^n$. Show that

$$\mathbb{E}\,[\text{ cumulative winning }] = \infty.$$

This is called the *St. Petersburg paradox*. The paradox is in an observation that one would not pay an infinite amount to play such a game.

Notice that if the game is stopped at the nth round, the dollar amount you spent in the previous rounds is

$$2^0 + \cdots + 2^{n-2} = \left(2^0 + \cdots + 2^{n-2}\right) \frac{1-\frac{1}{2}}{1-\frac{1}{2}} = 2^{n-1} - 1.$$

Therefore, the dollar difference between the total amount won and the total amount spent is

$$2^{n-1} - (2^{n-1} - 1) = 1,$$

and *does not depend on n*. This seems to specify a riskless strategy of winning \$1. However, if one introduces a *credit constraint*, i.e., if a player can only spent \$$M$, for some fixed positive number M, then even if M is large, the expected winning becomes finite, and one cannot safely win \$1 anymore.

Problem 17.5. In the context of Problem 17.4, let G denotes the *cumulative winning*. Instead of computing the expectation of G, Daniel Bernoulli has proposed to compute the expectation of the *logarithm* of G. Show that

$$\mathbb{E}\left[\log_2(G)\right] = \log_2(g) < \infty$$

and find g.

Problem 17.6. Let us suppose that a random variable X, which corresponds to the dollar amount of winning in some lottery, has the following distribution

$$\mathbb{P}[X = n] = \frac{1}{Cn^2}, \quad n \in \mathbb{N},$$

where $C = \sum\limits_{n=1}^{\infty} \frac{1}{n^2}$, which in particular is finite. Clearly, X is finite-valued (with probability one). Show that nevertheless $\mathbb{E}[X] = \infty$.

As a historical remark, note that here $C = \zeta(2)$, where $\zeta(s) = \sum\limits_{k=1}^{\infty} \frac{1}{n^s}$ is the Riemann zeta function (or Euler-Riemann zeta function) of a complex variables s. It was first proven by Euler in 1735 that $\zeta(2) = \frac{\pi^2}{6}$.

Problem 17.7. Let us suppose that a one-year interest rate is determined at the beginning of each month. In this case $r_0, r_1, \dots, r_{N-1}$ are such interest rates, where only r_0 is non-random. Thus \$1 of investment at time zero is worth $(1 + r_0)$ at the end of the year 1, $(1+r_0)(1+r_1)$ at the end of year 2, $(1+r_0)\dots(1+r_{k-1})$ at the end of year k, and so forth. Let us suppose that $r_0 = 0.1$ and $(r_i)_{i=1,\dots,N-1}$ are independent random variables with the following Bernoulli distribution (under so-called risk-neutral measure): $r_i = 0.15$ or .05 with probability 1/2 each.

Compute the price at time 0 of the security that pays \$1 at time N. Note that such a security is called *zero-coupon bond*.

Hint: let D_N denotes the *discount factor*, i.e.,

$$D_N = \frac{1}{(1 + r_0)\dots(1 + r_{N-1})}.$$

We need to evaluate

$$\widetilde{\mathbb{E}}\left[D_N\right].$$

Problem 17.8. In the settings of Problem 17.7, let *simply compounded yield* for the zero-coupon bond with maturity N is the number y, such that

$$\widetilde{\mathbb{E}}\,[D_N] = \frac{1}{(1+y)^m}.$$

Calculate y.

Problem 17.9. In the settings of Problem 17.7, let *continuously compounded yield* for the zero-coupon bond with maturity N is the number y, such that

$$\widetilde{\mathbb{E}}\,[D_N] = e^{-yN}.$$

Find y.

17.4. Problems on the Black-Scholes-Merton pricing formula

The following problems are related to the Black-Scholes-Merton pricing formula. Let us suppose that X is a standard normal random variable and

$$S(T) = S(0)\exp\left((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}X\right), \tag{17.4.1}$$

is the price of the stock at time T, where r is the interest rate, σ is the volatility, and $S(0)$ is the initial value. Here T, r, σ, and $S(0)$ are constants.

Problem 17.10. Show that

$$\mathbb{E}\left[e^{-rT}(S(T) - K)^+\right] = S(0)\Phi(d_1) - Ke^{-rT}\Phi(d_2), \tag{17.4.2}$$

where K is a positive constant,

$$d_1 = \frac{1}{\sigma\sqrt{T}}\left(\log\left(\frac{S(0)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T\right), \quad d_2 = \frac{1}{\sigma\sqrt{T}}\left(\log\left(\frac{S(0)}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)T\right),$$

and Φ is the cumulative distribution function of a standard normal random variable, i.e.

$$\Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}z^2}dz, \quad y \in \mathbb{R}.$$

Note that (17.4.2) the Black-Scholes-Merton formula, which gives the price of a European call option in at time 0 with strike K and maturity T.

Problem 17.11. In the framework of the Black-Scholes-Merton model, i.e., with the stock price process given by (17.4.1) with $r = 0$, let us consider

$$\mathbb{E}\left[S(t)^{1/3}\right]. \tag{17.4.3}$$

Find $\hat{t} \in [0,1]$ and evaluate $\mathbb{E}\left[S(\hat{t})^{1/3}\right]$ such that

$$\mathbb{E}\left[S(\hat{t})^{1/3}\right] = \max_{t\in[0,1]} \mathbb{E}\left[S(t)^{1/3}\right].$$

Note that $\max\limits_{t\in[0,1]} \mathbb{E}\left[S(t)^{1/3}\right]$ is closely related to the payoff of the *American cube root option* with maturity 1 and $\hat{t}$ to the optimal policy.

Problem 17.12. In the framework of the Black-Scholes-Merton model, i.e. with the stock price process given by (17.4.1), let us consider

$$\max_{t\in[0,1]} \mathbb{E}\left[e^{-rt}\left(S(t) - K\right)^+\right]. \tag{17.4.4}$$

Find $\hat{t} \in [0,1]$, such that

$$\mathbb{E}\left[e^{-r\hat{t}}\left(S(\hat{t}) - K\right)^+\right] = \max_{t\in[0,1]} \mathbb{E}\left[e^{-rt}\left(S(t) - K\right)^+\right].$$

Similarly to Problem 17.11, $\max\limits_{t\in[0,1]} \mathbb{E}\left[e^{-rt}\left(S(t) - K\right)^+\right]$ is closely related to the payoff of the *American call option* with maturity 1 and $\hat{t}$ to the optimal policy.

17.5. Selected solutions

Answer to Exercise 17.4: The normal approximations imply

$$\mathbb{E}M_n = M\left(\frac{e^{\sigma}+e^{-\sigma}}{2}\right)^n,$$
$$\mathbb{E}M_n^2 = M^2\left(\frac{e^{2\sigma}+e^{-2\sigma}}{2}\right)^n,$$
$$\operatorname{Var}(M_n) = M^2\left(\left(\frac{e^{2\sigma}+e^{-2\sigma}}{2}\right)^n - \left(\frac{e^{\sigma}+e^{-\sigma}}{2}\right)^{2n}\right).$$

Answer to Exercise 17.9:

$$\mathbb{E}M_n \approx Me^{n\sigma^2/2},$$
$$\mathbb{E}M_n^2 \approx M^2e^{2n\sigma^2},$$
$$\operatorname{Var}(M_n) \approx M^2\left(e^{2n\sigma^2} - e^{n\sigma^2}\right).$$

Answer to Exercise 17.11:

$$\mathbb{E}M_n = M\left(\frac{e^{\sigma}+e^{-\sigma}}{2}\right)^n < Me^{n\sigma^2/2},$$
$$\mathbb{E}M_n^2 = M^2\left(\frac{e^{2\sigma}+e^{-2\sigma}}{2}\right)^n < M^2e^{2n\sigma^2},$$
$$\operatorname{Var}(M_n) = M^2\left(\left(\frac{e^{2\sigma}+e^{-2\sigma}}{2}\right)^n - \left(\frac{e^{\sigma}+e^{-\sigma}}{2}\right)^{2n}\right) < M^2\left(e^{2n\sigma^2} - e^{n\sigma^2}\right).$$

The normal approximations get better if σ is small, but get worse if n is large. The standard optimal regime is $n \to \infty$ and $n\sigma^2 \to 1$, which means $\sigma \sim \frac{1}{\sqrt{n}}$.

Sketch of the solution to Problem 17.1: The strategy is to select a value x and say that the player stops if this value is exceeded after the first throw, and goes to the second through if this value is not exceeded. We know that the average value of one through is $(6+1)/2$ without any strategies. The probability to exceed x is $(6-x)/6$, and the conditional expectation of the payoff is $\frac{7+x}{2}$ if x is exceeded. So the expected payoff is $\frac{7+x}{2}\cdot\frac{6-x}{6}+\frac{7}{2}\cdot\frac{x}{6}$. This gives the optimal strategies for $x=3$ and the maximal expected payoff is $EP_2 = 4.25$.

Sketch of the solution to Problem 17.2: The expected payoff is $\frac{7+x}{2}\cdot\frac{6-x}{6}+\frac{17}{4}\cdot\frac{x}{6}$. Here $\frac{17}{4}=4.25$ replaces $\frac{7}{2}=3.5$ because after the first throw the player can decide either to stop, or play the game with two throws, which was solved and the maximal expected payoff was 4.25. So in the case of three throws, we have one optimal strategy with cut off $x_1 = 4$

after the first throw, and cut off $x_2 = 3$ after the second throw, following Problem 17.1. The expected payoff of the game which allows up to three throws is

$$\mathbb{E}P_3 = \frac{7+4}{2} \cdot \frac{6-4}{6} + \frac{17}{4} \cdot \frac{4}{6} = \frac{14}{3} \approx 4.6666$$

Answer to Problem 17.3:

(1) $\mathbb{E}P_4 = \frac{89}{18} \approx 4.9444$ $\qquad$ $\mathbb{E}P_5 = \frac{277}{54} \approx 5.1296$ $\qquad$ $\mathbb{E}P_6 = \frac{1709}{324} \approx 5.2747$

(2) $\mathbb{E}P_\infty = 6$.

Sketch of the solution to Problem 17.4: By direct computation, $\mathbb{E}\left[\textit{cumulative winning}\right]$ is given by the following divergent series

$$\sum_{k=1}^{\infty} 2^k \frac{1}{2^k} = \infty.$$

Sketch of the solution to Problem 17.5:

$$\mathbb{E}\left[\log_2(G)\right] = \sum_{k=1}^{\infty} \log_2(2^k)\frac{1}{2^k} = \sum_{k=1}^{\infty} \frac{k}{2^k} = 2 = \log_2(4) < \infty.$$

In particular, $g = 4$.

Sketch of the solution to Problem 17.6: $\mathbb{E}[X] = \frac{1}{C}\sum\limits_{n=1}^{\infty}\frac{1}{n} = \infty$, as the harmonic series is divergent.

Sketch of the solution to Problem 17.7: Using independence of r_ks, we get

$$\tilde{\mathbb{E}}\left[D_N\right] = \frac{1}{1+r_0}\prod_{k=1}^{N-1}\tilde{\mathbb{E}}\left[\frac{1}{1+r_k}\right] = \frac{1}{1+r_0}\left(\tilde{\mathbb{E}}\left[\frac{1}{1+r_1}\right]\right)^{N-1} = 0.909 \times 0.911^{N-1}.$$

Sketch of the solution to Problem 17.8: Direct computations give $y = \left(\frac{1}{\tilde{\mathbb{E}}[D_N]}\right)^{\frac{1}{N}} - 1$.

Sketch of the solution to Problem 17.9: Similarly to Problem 17.8, we get $y = -\frac{\log\left(\tilde{\mathbb{E}}[D_N]\right)}{N}$.

Sketch of the solution to Problem 17.10: From formula (17.4.1), we get

$$\mathbb{E}\left[e^{-rT}(S(T)-K)^+\right] = \int_{\mathbb{R}} e^{-rT}\max\left(S(0)e^{(r-\frac{1}{2}\sigma^2)T+\sigma\sqrt{T}x} - K, 0\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx.$$

Now, integration of the right-hand side yields the result.

Sketch of the solution to Problem 17.11: Let us fix $t \in [0,1]$. From Jensen's inequality (Proposition 15.7), we get $\mathbb{E}\left[S(t)^{1/3}\right] \leq \left(\mathbb{E}\left[S(t)\right]\right)^{1/3}$. The inequality is strict for $t > 0$, by strict concavity of $x \to x^{1/3}$. The equality is achieved at $t = 0$. Therefore $\hat{t} = 0$, and $\mathbb{E}\left[S(\hat{t})^{1/3}\right] = S(0)^{1/3}$.

Note that in the settings of this problem $\hat{t}$ is actually the optimal policy of the *American cube root option* and $\mathbb{E}\left[S(\hat{t})^{1/3}\right]$ is the corresponding price. However, in general one needs to consider $\max_{\tau\in[0,1]} \mathbb{E}\left[S(\tau)^{1/3}\right]$, where τ is what is known as *stopping times*, i.e. random times which an additional structural property.

Sketch of the solution to Problem 17.12: for every $t \in [0,1]$, we have

$$e^{-rt}\left(S(t)-K\right)^{+} = \left(S(t)e^{-rt} - Ke^{-rt}\right)^{+} \leqslant \left(S(t)e^{-rt} - Ke^{-r}\right)^{+}$$

and $\mathbb{E}\left[S(t)e^{-rt}\right] = S(0)$. Now, using convexity of $x \to (x-K)^{+}$ and applying Jensen's inequality (Proposition 15.7) for conditional expectation, we deduce that

$$\mathbb{E}\left[\left(S(t)e^{-rt} - Ke^{-r}\right)^{+}\right] < \mathbb{E}\left[\left(S(1)e^{-r} - Ke^{-r}\right)^{+}\right]$$

for every $t \in [0,1)$. We conclude that

$$\hat{t} = 1 \quad \text{and} \quad \mathbb{E}\left[e^{-r\hat{t}}\left(S(\hat{t}) - K\right)^{+}\right] = \mathbb{E}\left[e^{-r}\left(S(1)-K\right)^{+}\right].$$

Table of probability distributions

Discrete random variables

Name	Abbrev.	Parameters	PMF: $\mathbb{P}[X=k],\ k \in \mathbb{N}_0$	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	MGF: $\mathbb{E}[e^{tX}],\ t \in \mathbb{R}$
Bernoulli	$\mathrm{Bern}(p)$	$p \in [0,1]$	$\binom{1}{k} p^k (1-p)^{1-k}$	p	$p(1-p)$	$(1-p)+pe^t$
Binomial	$\mathrm{Bin}(n,p)$	$n \in \mathbb{N}, p \in [0,1]$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$[(1-p)+pe^t]^n$
Poisson	$\mathrm{Pois}(\lambda)$	$\lambda > 0$	$e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ	$\exp(\lambda(e^t-1))$
Geometric	$\mathrm{Geo}(p)$	$p \in (0,1)$	$\begin{cases} (1-p)^{k-1}p, & \text{for } k \geq 1, \\ 0, & \text{else.} \end{cases}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$, for $t < -\log(1-p)$
Negative binomial	$\mathrm{NB}(r,p)$	$r \in \mathbb{N}, p \in (0,1)$	$\begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r}, & \text{if } k \geq r, \\ 0, & \text{else.} \end{cases}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$, for $t < -\log(1-p)$
Hypergeometric		$n, m, N \in \mathbb{N}_0$	$\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$	$\frac{nm}{N}$	(not tested)	(not tested)

Continuous random variables

Name	Abbrev.	Parameters	PDF.: $f(x),\ x \in \mathbb{R}$	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	MGF: $\mathbb{E}[e^{tX}],\ t \in \mathbb{R}$
Uniform	$\mathcal{U}(a,b)$	$a, b \in \mathbb{R}, a < b$	$\begin{cases} \frac{1}{b-a}, & \text{if } x \in [a,b], \\ 0, & \text{if } x \notin [a,b]. \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Normal	$\mathcal{N}(\mu,\sigma^2)$	$\mu, \sigma \in \mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2	$e^{\mu t} e^{\sigma^2 t^2/2}$
Exponential	$\mathrm{Exp}(\lambda)$	$\lambda > 0$	$\begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$, for $t < \lambda$