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 strongly 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 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 the 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 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.

Our 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. Introduction

The first basic principle is to multiply.

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

Example 1.2: How many license plates of 3 letters followed by 3 numbers are possible?

Answer. $(26)^3(10)^3$, because 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.

How many ways can one arrange a, b, c? One can have

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

There are 3 possibilities for the first position. Once we have chosen the first position, there are 2 possibilities for the second position, and once we have chosen the first two possibilities, 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 letters, there are $n!$ possibilities.

Example 1.3: What is the number of possible batting orders with 9 players?

Answer. $9! = 362880$

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.

Answer. $4! \cdot (4! \cdot 3! \cdot 2! \cdot 1!) = 6912$. 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 the letters M, C, P, B, and there are 4! ways of doing that.

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

If there were 3 a's, 4 b's, and 2 c's, we would have

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

What we just did was called the ***number of permutations***.

Now let us look at what are known as ***combinations***. 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 for the first position, 4 for the second, and 3 for the third, for a total of $5 \times 4 \times 3$. But suppose the letters selected were a, b, c. If order doesn't matter, we will have the letters a, b, c 6 times, because there are 3! ways of arranging 3 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 $\binom{5}{3}$, read "*5 choose 3*." Sometimes this is written $C_{5,3}$ or ${}_5C_3$. More generally,

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

Example 1.6: How many ways can one choose a committee of 3 out of 10 people?

Answer. $\binom{10}{3} = 120$.

Example 1.7: Suppose there are 8 men and 8 women. How many ways can we choose a committee that has 2 men and 2 women?

Answer. We can choose 2 men in $\binom{8}{2}$ ways and 2 women in $\binom{8}{2}$ ways. The number of committees is then the product: $\binom{8}{2} \cdot \binom{8}{2} = 784$.

Example 1.8: Suppose one has 9 people and one wants to divide them into one committee of 3, one of 4, and a last 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

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

In general, to divide n objects into one group of n_1, one group of n_2, ..., and a kth group of n_k, where $n = n_1 + \cdots + n_k$, the answer is

$$\frac{n!}{n_1!n_2!\cdots n_k!}.$$

These are known as multinomial coefficients.

Another example: 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?

Answer. (a) This is just 10! (b) Consider the Americans as a group and each Canadian as a group; this gives 7 groups, which can be arranged 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! (c) This is the answer to (a) minus the answer to (b): $10! - 4!7!$ (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 $\binom{4}{3} + \binom{6}{3}$. (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. First, suppose one has 8 o's and 2 $|$'s. 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 o's. So we have $\binom{10}{8}$.

Next, suppose one has 8 indistinguishable balls. How many ways can one put them in 3 boxes? Let us make sequences of o's and $|$'s; any such sequence that has $|$ at each side, 2 other $|$'s, and 8 o's 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: the answer is $\binom{10}{8}$.

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. Counting Principle revisited. We need a way to help us count faster rather than counting by hand one by one. We define the following counting principle.

FACT. *(Basic Counting Principle) Suppose 2 experiments are to be performed. If one experiment can result in* **m** *possibilities, and the second experiment can result in* **n** *possibilities, then together there are* **mn** *possibilities.*

One can visualize the Basic Counting Principle by using the Box Method. In the Box Method, each box represents the number of possibilities in that experiment.

$$\begin{array}{ccccc} \boxed{Experiment 1} & \boxed{Experiment 2} & = & \boxed{\text{Experiment 1 and 2 together}} \\ \boxed{\mathbf{n}} & \boxed{\mathbf{m}} & & = \boxed{nm} \end{array}$$

Example 1.9: There are 20 teachers and 100 students in a school. How many ways can we pick a teacher and student of the year?

Answer. Use the box method: $20 \times 100 = 2000$.

FACT. *The counting principle can be generalized to any number of experiments: for* $\boldsymbol{k}$ *experiment we have* $\boldsymbol{n_1 \cdots n_k}$ *possibilities.*

Example 1.10: A college planning committee consists of 3 freshmen, 4 sophomores, 5 juniors, and 2 seniors. A subcommittee of 4 consists 1 person from each class. How many choices are possible?

Answer. Box Method gives $3 \times 4 \times 5 \times 2 = 120$.

Example 1.11: 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. Question: What if no repetition is allowed?

Answer. the Box Method again: $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8$

Example 1.12: How many functions defined on k points are possible if each functional value is either 0 or 1.

Answer. 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. Recall how many different ordered arrangements of the letters a, b, c are possible:

- $abc, acb, bac, bca, cab, cba$, and each arrangement is a permutation.
- We also can use the Box Method to figure this out: $3 \cdot 2 \cdot 1 = 6$.

FACT. *With n objects. There are*

$$\boxed{n\,(n-1)\cdots 3 \cdot 2 \cdot 1 = n!}$$

different permutations of the n objects.

($\star$) *Note that* **order matters** *when it comes to* **permutations**.

Example 1.13: What is the number of possible batting order with 9 players?

Answer. 9! by the Box Method or permutations.

Example 1.14: 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.

Answer. 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,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 1.15: How many ways can one arrange the letters a, a, b, b, c, c? Let us first re-label them A, a, B, b, C, c. Then there are $6! = 720$, ways to arrange these letters. But we have repeats: we could have Aa or aA. So we have a repeat for each possibility ans (**so we have to divide!**).

Answer. $6!/(2!)^3 = 60$.

Example 1.16: How many different letter arrangements can be formed from the word PEPPER?

Answer. There 3 P's 2 E's and one R. So $\frac{6!}{3!2!1!} = 30$.

Example 1.17: 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 don't distinguish players from the same country?

Answer. $\frac{11!}{4!4!3!}$.

FACT. *There are*

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

different permutations of n objects of which n_1 are alike, n_2 are alike, n_r are alike.

1.2.3. Combinations. We are often interested in selecting r objects from a total of n objects and the order of these objects does not matter.

FACT. *If $r \leq n$, then*

$$\boxed{\binom{n}{r} = \frac{n!}{(n-r)!r!}}$$

called n choose r, represents the number of possible ***combinations*** *of objects taken r at a time from n objects.*

($\star$) The ***order DOES NOT matter*** *for combinations.*

Recall in Permutations order did matter.

Example 1.18: How many ways can one choose a committee of 3 out of 10 people?

Answer. $\binom{10}{3} = \frac{10!}{3!7!} = \frac{10\cdot9\cdot8}{3\cdot2} = 10 \cdot 3 \cdot 4 = 120.$

Example 1.19: Suppose there are 9 men and 8 women. How many ways can we choose a committee that has 2 men and 3 women?

Answer. 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 1.20: Suppose somebody has n friends, of whom k are be invited to a meeting.

Answer.

a How many choices exist if 2 of the friends will not attend together?
  - Box it: [*none*] + [one of them] [*others*]
  - $\binom{n-2}{k} + \binom{2}{1} \cdot \binom{n-2}{k-2}$ (recall that when we have OR, use +)

b How many choices exist if 2 of the friends will only attend together?
  - Box it: [*none*] + [with both]
  - $\binom{n-2}{k} + 1 \cdot 1 \cdot \binom{n-2}{k-2}$

The value of $\binom{n}{r}$ are called binomials coefficients because of their prominence in the binomial theorem.

THEOREM. *(The Binomial Theorem)*

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

Proof. To see this, the left hand side is $(x+y)(x+y)\cdots(x+y)$. This will be the sum of 2^n terms, and each term will have n factors. How many terms have k x's and $n-k$ y's? This 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's? (Box it) The answer is $\begin{pmatrix} n \\ k \end{pmatrix}$, so the coefficient of $x^k y^{n-k}$ should be $\begin{pmatrix} n \\ k \end{pmatrix}$.

Example 1.21: Using Combinatorics: Let's prove

$$\begin{pmatrix} 10 \\ 4 \end{pmatrix} = \begin{pmatrix} 9 \\ 3 \end{pmatrix} + \begin{pmatrix} 9 \\ 4 \end{pmatrix}$$

with no algebra:

Answer. The left hand side (LHS) represents the number of committees having 4 people out of the 10. Let's interpret the right hand side (RHS). Let's say Tom Brady will be in one of these committees and he's special, so we want to know when he'll be there or not. When he's there, then there are $1 \cdot \begin{pmatrix} 9 \\ 3 \end{pmatrix}$ number of ways that contain Tom Brady while $\begin{pmatrix} 9 \\ 4 \end{pmatrix}$ is the number of committees that do not contain Tom Brady and contain 4 out of the remaining people. Adding it up gives us the number of committees having 4 people out of the 10.

Example 1.22: The more general equation is

$$\boxed{\begin{pmatrix} n \\ r \end{pmatrix} = \begin{pmatrix} n-1 \\ r-1 \end{pmatrix} + \begin{pmatrix} n-1 \\ r \end{pmatrix}}$$

Example 1.23: Expand $(x+y)^3$.

Answer. $(x+y)^3 = y^3 + 3xy^2 + 3x^2y + x^3$.

1.2.4. Multinomial Coefficients.

Example 1.24: Suppose one has 9 people and wants to divide them into one committee of 3, one of 4, and a last of 2. How many different ways are there?

Answer. (Box it) 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 there is 1 one to choose that. So the answer is

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

<u>**In general:**</u> if we are to divide n objects into a group of n_1, a group of n_2, $\ldots$ and a kth group of n_k, where $n = n_1 + \cdots + n_k$, then the answer can be given in $\dfrac{n!}{n_1!n_2!\cdots n_k!}$ ways. These are known as multinomial coefficients. We write them as

$$\boxed{\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!n_2!\cdots n_k!}}$$

Example 1.25: Suppose we are to assign 10 police officers: 6 patrols, 2 in station, 2 in schools. Then there are $\dfrac{10!}{6!2!2!}$ different assignments.

Example 1.26: There are 10 flags: 5 Blue one, 3 red one, and 2 yellow. These flags are indistinguishable, except for their color. How may different ways can we order them on a flag pole?

Answer. $\frac{10!}{5!3!2!}$.

Example 1.27: Suppose one has n indistinguishable balls. How many ways can one put them in k boxes, assuming $n > k$?

> <u>**Solution 1:**</u> Let us make sequences of o's and $|$'s; any such sequence that has $|$ at each side, $k-1$ other $|$'s, and n o's represents a way of arranging balls into boxes. For example, one may have
>
> $$|\; oo \;|\; ooo \;|\; ooo \;|$$
>
> if $n = 8$ and $k = 3$. How many different ways can we arrange this, if we have to start with $|$ and end with $|$? In between, we are only arranging $n+k-1$ symbols, of which only n are o's. So the question is: how many ways out of $n+k-1$ **spaces** can one pick n **of them** into which to put an o? Answer: $\binom{n+k-1}{n}$.
>
> <u>**Solution 2:**</u> Look at spaces between that have a $|$. There are $k-1$ spaces, and so the answer is $\binom{n+k-1}{k-1} = \binom{n+k-1}{n}$.

1.3. Exercises

Exercise 1.1: Suppose a License plate must consist of 7 numbers or letter. 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 groomsman 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 party's 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 party's are possible?
(D) Suppose that one possible groomsman and one possible bridesmaid refuse to serve together. How many wedding party's are possible?

Exercise 1.9: There are 52 cards in a standard deck of playing cards. The poker hand is 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

1.4. Selected solutions

Solution to Exercise 1.1(A): 26^7

Solution to Exercise 1.1(B): $10^3 \cdot 26^4$

Solution to Exercise 1.1(C): $10 \cdot 9 \cdot 8 \cdot 26 \cdot 25 \cdot 24 \cdot 23$

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: $\begin{pmatrix} 52 \\ 5 \end{pmatrix}$

Solution to Exercise 1.10: $\begin{pmatrix} 30 \\ 2 \end{pmatrix}$

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

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

CHAPTER 2

The probability set-up

2.1. Introduction and basic theory

We will have a sample space, denoted 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 use the following usual notation: $A \cup B$ is the union of A and B and denotes the points of S that are in A or B or both. $A \cap B$ is the intersection of A and B and is the set of points that are in both A and B. $\emptyset$ denotes the empty set. $A \subset B$ means that A is contained in B and A^c is the complement of A, that is, the points in S that are not in A. We extend the definition to have $\cup_{i=1}^n A_i$ is the union of $A_1, \cdots, A_n$, and similarly $\cap_{i=1}^n A_i$.

An exercise is to show that

$$(\cup_{i=1}^n A_i)^c = \cap_{i=1}^n A_i^c \qquad \text{and} \qquad (\cap_{i=1}^n A_i)^c = \cup_{i=1}^n A_i^c.$$

These are called DeMorgan's laws.

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

(i) $\emptyset, S$ *is 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* $\cup_{i=1}^\infty A_i$ *and* $\cap_{i=1}^\infty A_i$ *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.

So now we have a space S, a σ-field $\mathcal{F}$, and we need to talk about what a probability is. There are three axioms:

(1) $0 \le \mathbb{P}(E) \le 1$ *for all events* E*.*
(2) $\mathbb{P}(S) = 1$.
(3) If the E_i *are pairwise disjoint,* $\mathbb{P}(\cup_{i=1}^\infty E_i) = \sum_{i=1}^\infty \mathbb{P}(E_i)$.

Pairwise disjoint means that $E_i \cap E_j = \emptyset$ unless $i = j$.

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 times, 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.

There are some easy consequences of the axioms.

Proposition 2.1: (1) $\mathbb{P}(\emptyset) = 0$.
(2) If $A_1, \ldots, A_n$ are pairwise disjoint, $\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$.
(3) $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
(4) If $E \subset F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$.
(5) $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.

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

The second follows if we let $A_{n+1} = A_{n+2} = \cdots = \emptyset$. We still have pairwise disjointness and $\cup_{i=1}^\infty A_i = \cup_{i=1}^n A_i$, and $\sum_{i=1}^\infty \mathbb{P}(A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$, using (1).

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 $\mathbb{P}(F) = \mathbb{P}(E) + \mathbb{P}(F \cap E^c) \geq \mathbb{P}(E)$ 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 very 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?

Answer. 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?

Answer. The probability of 4 aces and 1 king is $\binom{4}{4}\binom{4}{1} \Big/ \binom{52}{5}$. 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

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

Example 2.3: What is the probability that in a poker hand we get exactly 3 of a kind (and the other two cards are of different ranks)?

Answer. The probability of 3 aces, 1 king and 1 queen is

$$\binom{4}{3}\binom{4}{1}\binom{4}{1} \Big/ \binom{52}{5}.$$

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

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

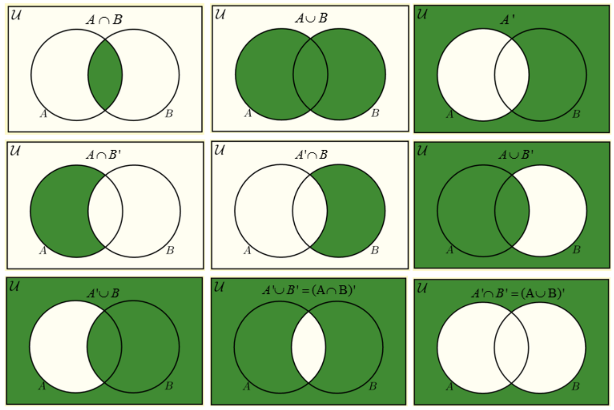
Example 2.4: In a class of 30 people, what is the probability everyone has a different birthday? (We assume each day is equally likely.)

Answer. 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

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

2.2. Further examples and applications

2.2.1. Events. An event A is a subset of S. In this case we use the notation $A \subset S$ meaning that A is a subset of S.



A picture of Venn diagrams from

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

Example 2.5: Roll two dice. Examples of events are

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

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 and their consequences.

Example 2.7: Coin tosses. Recall that if we toss a coin with each side being equally likely. Then, $S = \{H, T\}$ and

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

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

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

An important consequence of the axioms that appeared in Proposition 2.1 is the so-called *inclusion-exclusion identity* for two events $A, B \subseteq S$,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \tag{2.2.1}$$

Let us see how we can combine the different consequences listed 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:

- Home game has .5 chance of winning
- Away game has .4 chance of winning.
- There is .3 chances that UConn wins both games.

What is the probability that UConn loses both games?

Let us write $A_1 =$"win home game", and $A_2 =$"win away game". Then, from past experience we know that $\mathbb{P}(A_1) = .5$, $\mathbb{P}(A_2) = .4$ and $\mathbb{P}(A_1 \cap A_2) = .3$. Notice that the event "loses both games" can be expressed as $A_1^c \cap A_2^c$. Thus we want to find out $\mathbb{P}(A_1^c \cap A_2^c)$. Simplifying as much as possible (de Morgan's laws!) and using consequence (3) from 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 (2.2.1) tells us

$$\mathbb{P}(A_1 \cup A_2) = .5 + .4 - .3 = .6,$$

and hence $\mathbb{P}(A_1^c \cap A_2^c) = 1 - .6 = .4$.

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 $A, B, C \subseteq S$,

$$\begin{aligned} \mathbb{P}(A \cup B \cup C) = {} & \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ & - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C). \end{aligned} \tag{2.2.2}$$

Exercise 2.1: Prove Proposition 2.2 by grouping $A \cup B \cup C$ as $A \cup (B \cup C)$ and using the formula (2.2.1) for two sets.

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.

Proposition 2.3: The probability of an event $E \subseteq S$ is

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

There are different ways how to count the number of outcomes. If nothing is explicitly said we can always choose (and specify!) the way we regard our experiment.

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

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

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?

We may distinguish two cases:

a. The order in which the balls are drawn is important. Then,

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

b. The order is not important. In this case

$$P(E) = \frac{(1 \text{ white})(2 \text{ black})}{\binom{11}{3}} = \frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{4}{11}.$$

2.3. Exercises

Exercise 2.2: 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 2.3: 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 2.4: Forty percent of the students at a certain college are members neither of an academic club nor 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 2.5: 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 2.6: 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 nt 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 10's and exactly three Aces?
(C) all 5 cards are either face cards or no-face cards?

Exercise 2.7: 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 2.8: 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 2.9: 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 2.10: In a seminar attended by 8 students, what is the probability that at least two of them have birthday in the same month?

Exercise 2.11: 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 2.12: 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 2.13: 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 2.14: 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 2.15: 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)?

2.4. Selected solutions

Solution to Exercise 2.2(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 2.2(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 2.3(A): Since $A \cap B = \emptyset$, $A^c \subseteq B$ hence $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) = 0.4$.

Solution to Exercise 2.3(B): By de Morgan's laws and property (3) of Proposition 2.1,

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

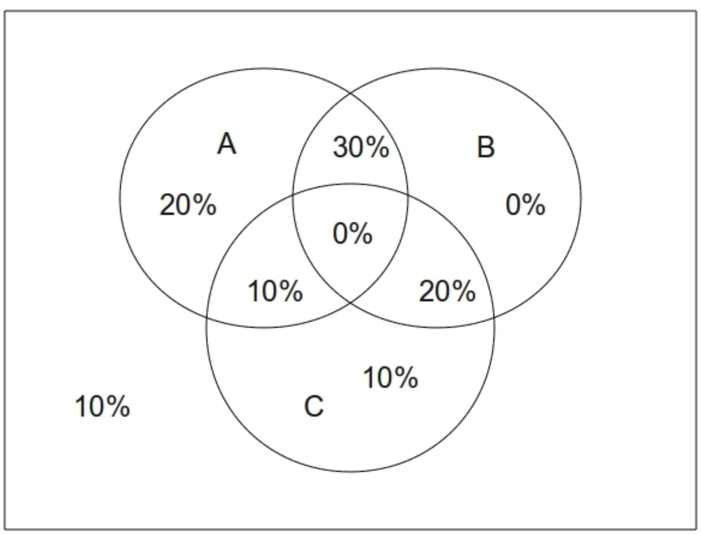
Solution to Exercise 2.4(A): $\mathbb{P}(A \cup B) = 1 - .4 = .6$

Solution to Exercise 2.4(B): Notice that

$$.6 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = .5 + .3 - \mathbb{P}(A \cap B)$$

Thus, $\mathbb{P}(A \cap B) = .2$.

Solution to Exercise 2.5(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 2.5(B): The Venn diagram tells us that $100\% - (10\% + 20\% + 30\%) = 40\%$ of the households subscribe to at most one paper.

Solution to Exercise 2.6(A): $\dfrac{\binom{26}{5}}{\binom{52}{5}}$

Solution to Exercise 2.6(B): $\dfrac{\binom{4}{2}\cdot\binom{4}{3}}{\binom{52}{5}}$

Solution to Exercise 2.6(C): $\dfrac{\binom{12}{5}}{\binom{52}{5}}+\dfrac{\binom{40}{5}}{\binom{52}{5}}$

Solution to Exercise 2.7(A): $13\binom{4}{2}\binom{12}{3}\binom{4}{1}\binom{4}{1}\binom{4}{1}\Big/\binom{52}{5}$

Solution to Exercise 2.7(B): $\binom{13}{2}\binom{4}{2}\binom{4}{2}\binom{44}{1}\Big/\binom{52}{5}$

Solution to Exercise 2.7(C): $13\binom{4}{3}\binom{12}{2}\binom{4}{1}\binom{4}{1}\Big/\binom{52}{5}$

Solution to Exercise 2.7(D): $4\binom{13}{5}\Big/\binom{52}{5}$

Solution to Exercise 2.7(E): $13\cdot 12\binom{4}{3}\binom{4}{2}\Big/\binom{52}{5}$

Solution to Exercise 2.8(A): $\frac{1}{70}$

Solution to Exercise 2.8(B): $\binom{14}{3}\binom{5}{2}\binom{5}{2}\binom{5}{2}\Big/\binom{70}{6}$

Solution to Exercise 2.8(C): $14\binom{5}{3}13\binom{5}{2}12\binom{5}{1}\Big/\binom{70}{6}$

Solution to Exercise 2.9: Simple inspection we can see that the only possibilities

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

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

Solution to Exercise 2.10: $1-\frac{12\cdot 11\cdot 10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5}{12^8}$

Solution to Exercise 2.11(A): $\dfrac{\binom{10}{2}\binom{12}{5}\binom{15}{2}}{\binom{37}{9}}$

Solution to Exercise 2.11(B): $1 - \dfrac{\binom{27}{9}}{\binom{37}{9}} - \dfrac{\binom{10}{1}\binom{27}{8}}{\binom{37}{9}}$

Solution to Exercise 2.12: $\frac{8 \cdot 7 \cdot 6 \cdot 5}{8^4}$

CHAPTER 3

Independence

3.1. Introduction

Definition 3.1: We say that E and F are **independent events** if

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

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

Proof.

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

□

We say E, F, and G 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: Suppose you roll two dice, E is that the sum is 7, F that the first is a 4, and G that the second is a 3. E and F are independent, as are E and G and F and G, but E, F and G are not.

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

Definition 3.3: Let $A_1, \ldots, A_n \subset S$ be a collection of n events. We say that they are independent if for all possible subcollections $i_1, \ldots, i_r \in \{1, \ldots, n\}$, $1 \leq r \leq n$, it holds that

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

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

Answer. 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. Suppose you have n independent trials, where the probability of a success is p. Then 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 $\binom{n}{k}p^k(1-p)^{n-k}$.

Proposition 3.4: If an experiment with probability of success p that is repeated n times independently, the probability of obtaining k successes for any $0 \leq k \leq n$ is given by

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

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

Example 3.5: 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 \$50. What's the probability you will get to \$200 before you go broke?

Answer. It is easier to solve a slightly harder problem. Let $y(x)$ be the probability you get to 200 before 0 if you start with x dollars. Clearly $y(0) = 0$ and $y(200) = 1$. If you start with x dollars, then with probability $\frac{1}{2}$ you get a heads and will then have $x+1$ dollars. With probability $\frac{1}{2}$ you get a tails and will then have $x-1$ dollars. So we have

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

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

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

This says succeeding slopes of the graph of $y(x)$ are constant (remember that x must be an integer). In other words, $y(x)$ must be a line. Since $y(0) = 0$ and $y(200) = 1$, we have

$$y(x) = x/200,$$

and therefore $y(50) = 1/4$.

Example 3.6: Suppose we are in the same situation, but you are allowed to go arbitrarily far in debt. Let $y(x)$ be the probability you ever get to \$200. What is a formula for $y(x)$?

Answer. Just as above, we have the equation $y(x) = \frac{1}{2}y(x+1) + \frac{1}{2}y(x-1)$. This implies $y(x)$ is linear, and as above $y(200) = 1$. Now the slope of y cannot be negative, or else we would have $y > 1$ for some x and that is not possible. Neither can the slope be positive, or else we would have $y < 0$, and again this is not possible, because probabilities must be between 0 and 1. Therefore the slope must be 0, or $y(x)$ is constant, or

$$y(x) = 1 \quad \text{for all } x.$$

In other words, one is certain to get to \$200 eventually (provided, of course, that one is allowed to go into debt). There is nothing special about the figure 300. Another way of seeing this is to compute as above the probability of getting to 200 before L and then letting $L \to -\infty$.

3.2. Further examples and explanations

3.2.1. Independent Events.

Example 3.7: A card is drawn from an ordinary deck of cards (52 cards). Consider the events

$$F = \text{"a face is drawn"}, \qquad R = \text{"a red color is drawn"}.$$

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 3.8: 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

$$A_T = \text{"the first coin lands tails"}, \quad B_T = \text{"the second coin lands heads"}$$

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:

$$\mathbb{P}(A_T\cap B_H) = .4999\cdot .5002 = \mathbb{P}(A_T)\,\mathbb{P}(B_h).$$

Example 3.9: Two dice are simultaneously rolled. Consider the events

$$A_1 = \text{"the sum is 9"}, \quad A_2 = \text{"the first die lands even"}, \quad A_3 = \text{"the second die lands a 3"}.$$

Are A_1, A_2 and A_3 independent?

Answer. Notice that, on the one hand, $\mathbb{P}(A_1\cap A_2\cap A_3) = \dfrac{1}{36}$. On the other hand,

$$\mathbb{P}(A_1)\,\mathbb{P}(A_2)\,\mathbb{P}(A_3) = \frac{4}{36}\frac{1}{2}\frac{1}{6} = \frac{1}{36\cdot 3} \neq \frac{1}{36},$$

so they are not independent.

Example 3.10: 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?

Answer. Define the events

$$R_i = \text{"a red ball is drawn from urn } i\text{"}, \quad B_i = \text{"a blue ball is drawn from urn } i\text{"},$$

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

Answer. 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 3.2, prove that E and F are independent if and only if E and F^c are independent.

Exercise 3.4: Prove the following statements about Example 3.3:

(a) E and F are independent,
(b) E and G are independent,
(c) F and G are independent,
(d) E, F and G are not 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: 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 3.5 for a solution).

Exercise 3.8: 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 3.9: You make successive independent flips of a coin that lands on heads with probability p. What is the probability that the 3rd head appears on the 7th flip? (Express your answers in terms of p; do not assume $p = 1/2$.)

Exercise 3.10: 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 3.11: 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

$$.64 = .4 + \mathbb{P}(B) - .4\mathbb{P}(B).$$

Solving for $\mathbb{P}(B)$ we have $\mathbb{P}(B) = .4$.

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 3.2 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 3.2 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: Rolling two dice, the sample space S has 36 elements, which are all possible pairs of numbers between 1 and 6. All possible outcomes (pairs) are equally likely. The event E has 6 possible outcomes: $(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)$. The events F and G have also 6 possible outcomes each (check this!). Thus,

$$\mathbb{P}(E) = \frac{6}{36} = \frac{1}{6} = \mathbb{P}(F) = \mathbb{P}(G).$$

Thus, on the one hand,

$$\mathbb{P}(E) \cdot \mathbb{P}(F) = \mathbb{P}(E) \cdot \mathbb{P}(G) = \mathbb{P}(F) \cdot \mathbb{P}(G) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

On the other hand, since $E \cap F = \{(4,3)\} = E \cap G = F \cap G$, we have

$$\mathbb{P}(E \cap F) = \mathbb{P}(E \cap G) = \mathbb{P}(F \cap G) = \frac{1}{36}$$

hence E, F, G are *pairwise* independent. However, $E \cap F \cap G = \{(4,3)\}$, so that

$$\mathbb{P}(E \cap F \cap G) = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(E) \cdot \mathbb{P}(F) \cdot \mathbb{P}(G),$$

and therefore E, F, G are *not* all together independent.

Solution to Exercise 3.8: These are Bernouilli trials. Each game is a trial and the probability of loosing is $p = 0.45$. Using Proposition 3.4 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 3.9: 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

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

Using Berouilli trials, $\mathbb{P}(\text{“ 2 heads in 6 trials”}) = \binom{6}{2}p^2(1-p)^4$ 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 3.11:

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

CHAPTER 4

Conditional probability

4.1. Introduction

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 make the following definition.

Definition 4.1: If $\mathbb{P}(F) > 0$, we define

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

$\mathbb{P}(E \mid F)$ is read "the probability of E given F."

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

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 4.1: 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?

Answer. 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. The $\mathbb{P}(A \cap B) = \binom{3}{2}\binom{2}{0} \Big/ \binom{5}{2}$. Then $\mathbb{P}(A \mid B) = \frac{3/10}{3/5} = 1/2$.

Example 4.2: 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?

Answer. 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 4.3: 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 mean HIV positive, and T mean tests positive.

$$\mathbb{P}(D \mid T) = \frac{\mathbb{P}(D \cap T)}{\mathbb{P}(T)} = \frac{(.99)(.003)}{(.99)(.003) + (.01)(.997)} \approx 23\%.$$

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

Example 4.4: 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?

Answer. Let D be the families that own a dog, and C the families that own a cat. We are given $\mathbb{P}(D) = .36, \mathbb{P}(C) = .30, \mathbb{P}(C \mid D) = .22$ 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 $\mathbb{P}(D \cap C) = \mathbb{P}(C \mid D)\mathbb{P}(D) = (.22)(.36) = .0792$. So $\mathbb{P}(D \mid C) = .0792/.3 = .264 = 26.4\%$.

Example 4.5: 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?

Answer. 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 $\mathbb{P}(A \mid W) = .30, \mathbb{P}(A \mid M) = .25, \mathbb{P}(W) = .60$ 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,

$$\mathbb{P}(W \cap A) = \mathbb{P}(A \mid W)\mathbb{P}(W) = (.30)(.60) = .18.$$

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,

$$\mathbb{P}(M \cap A) = \mathbb{P}(A \mid M)\mathbb{P}(M) = (.25)(.40) = .10.$$

So

$$\mathbb{P}(A) = \mathbb{P}(W \cap A) + \mathbb{P}(M \cap A) = .18 + .10 = .28.$$

Finally,

$$\mathbb{P}(W \mid A) = \frac{\mathbb{P}(W \cap A)}{\mathbb{P}(A)} = \frac{.18}{.28}.$$

Proposition 4.2: If $\mathbb{P}(E) > 0$, then

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

This formula is known as **Bayes' rule**.

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 of the 60's called *Let's Make a Deal*.

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?

Answer. Let's suppose you choose door 1, since the same analysis applies whichever door you chose. Strategy one 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.

Strategy 2 is to change. Let's 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.

4.2. Further examples and applications

4.2.1. Conditional Probabilities.

Example 4.6: Landon is 80% sure he forgot his textbook either at the Union or at Monteith. He is 40% sure that the book is at the union, and 40% sure that it is at 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?

Answer. Calling $U =$ "textbook at the Union", and $U =$ "textbook at 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 4.7: 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?

Answer. Let $A =$ "Sarah has two aces", and let $B =$ "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 4.8: A total of 500 married couples are poled about their salaries with the following results:

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

(a) Find the probability that a Husband earns less than \$25K.
Answer.

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

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

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

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

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

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

Lemma 4.3: Let $E, F \subseteq S$ 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. (i) is a rewriting of $\mathbb{P}(F \mid E) = \frac{\mathbb{P}(E\cap F)}{\mathbb{P}(E)}$. 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 4.9: 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?

Answer. Let B = "Phan takes Biology", and C = "Phan takes Chemistry", and A = "score 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 relation $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F \mid E)$ from Lemma 4.3(i) can be generalized to any number of events in what is sometimes called the *multiplication rule*.

Proposition 4.4 (Multiplication rule)**:** Let $E_1, E_2, \ldots, E_n \subseteq S$ be events. Then,

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbb{P}(E_1)\,\mathbb{P}(E_2 \mid E_1)\,\mathbb{P}(E_3 \mid E_1 \cap E_2) \cdots \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

Example 4.10: 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?
Answer. In this case, we have the sequence of events B_1, B_2, B_3, where B_i = "the i-th 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?
Answer. Writing R_i = "the i-th 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 relation from Lemma 4.3(ii) can be generalized by partitioning the sample space S into several pairwise disjoint sets $F_1, \dots, F_n$ (instead of simply F and F^c).

Proposition 4.5 (Law of total probabilities)**:** Let $F_1, \dots, F_n \subseteq S$ be mutually exclusive and exhaustive events, i.e. $S = \bigcup_{i=1}^n F_i$. Then, for any event $E \subseteq S$ it holds that

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

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

Example 4.11: An insurance company classifies insurers into "accident prone" or "not accident prone". Their current risk model works with the following probabilities.

- The probability that an "accident prone" insurer has an accident within a year is 0.4
- The probability that a "non-accident prone" insurer 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?
Answer. Write A_1 = "policy holder will have an accident within a year" and let A = "policy holder is accident prone". Applying Lemma 4.3(ii) we have

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

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

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

Using the Law of total probabilities from Proposition 4.5 one can generalize Bayes's rule, which appeared in Proposition 4.2.

Proposition 4.6 (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

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

Example 4.12: Suppose a factory has machines I, II, and III that produce "iSung" phones. The factory's record show that

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

An iSung is selected at random from the factory.

(a) What is probability that the iSung selected is defective?
Answer. By the law of total probabilities,

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

(b) Given that the iSung is defective, what is the conditional probability that it was produced by machine III?
Answer. 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 4.13: 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?

Answer. Let K = "student knows the answer" and C = "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's the probability they actually knew 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)?

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: Let E be the event "there is at least one 6" and F be the event }the sum is at least 9" We want to calculate $\mathbb{P}(F \mid E)$. 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: Let B_i denote the event "box i 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. Let E be the event "a student is exposed to the flu" and let F be the event "a student gets the flu". We know that $P(E) = 60\%$ and $P(F \mid E) = 15\%$. This means that $P(E \cap F) = (.6)(.15) = .09$, and it is clear that $P(E^c \cap F) = 0$. Since $P(F) = P(E \cap F) + P(E^c \cap F)$, we see that$P(F) = .09$.

Solution to Exercise 4.4(B): Suppose we look at students who have not gotten the flu shot. Let E be the event "a student is exposed to the flu" and let F be the event "a student gets the flu". We know that $P(E) = 60\%$ and $P(F \mid E) = 80\%$. This means that $P(E \cap F) = (.6)(.8) = .48$, and it is clear that $P(E^c \cap F) = 0$. Since $P(F) = P(E \cap F) + P(E^c \cap F)$, we see that $P(F) = .48$.

Solution to Exercise 4.5: Let M be the event “an American is a man" and let$C$ be the event “” an American is color blind.". Then

$$\begin{aligned}\mathbb{P}(M \mid C) &= \frac{\mathbb{P}(C \mid M)\,\mathbb{P}(M)}{\mathbb{P}(C \mid M)\,\mathbb{P}(M)+\mathbb{P}(C \mid M^c)\,\mathbb{P}(M^c)}\\ &= \frac{(.05)\,(.49)}{(.05)\,(.49)+(.0025)\,(.51)} \approx .9505.\end{aligned}$$

Solution to Exercise 4.6(A): Let H be the event “works over 5000 hours”. Let X be the event comes from factory X“ and Y be the event “comes fom factory Y”. Then by the Law of Total Probability

$$\begin{aligned}\mathbb{P}(H) &= \mathbb{P}(H \mid X)\,\mathbb{P}(X)+\mathbb{P}(H \mid Y)\,\mathbb{P}(Y)\\ &= (.99)\,(.6)+(.95)\,(.4)\\ &= .974.\end{aligned}$$

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

$$\begin{aligned}\mathbb{P}(Y \mid H) &= \frac{\mathbb{P}(H \mid Y)\,\mathbb{P}(Y)}{\mathbb{P}(H)}\\ &= \frac{(.95)\,(.4)}{.974} \approx .39.\end{aligned}$$

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

$$\begin{aligned}\mathbb{P}(X \mid H^c) &= \frac{\mathbb{P}(H^c \mid X)\,\mathbb{P}(X)}{\mathbb{P}(H^c)} = \frac{\mathbb{P}(H^c \mid X)\,\mathbb{P}(X)}{1-\mathbb{P}(H)}\\ &= \frac{(1-.99)\,(.6)}{1-.974} = \frac{(.01)\,(.6)}{.026}\\ &\approx .23\end{aligned}$$

Solution to Exercise 4.7: Let $D = \{\text{Bolt is defective}\}$, $A = \{\text{bolt is form machine } A\}$, $B = \{\text{bolt is from machine } C\}$. Then by Baye’s theorem

$$\begin{aligned}\mathbb{P}(A \mid D) &= \frac{\mathbb{P}(D \mid A)\,\mathbb{P}(A)}{\mathbb{P}(D \mid A)\,\mathbb{P}(A)+\mathbb{P}(D \mid B)\,\mathbb{P}(B)+\mathbb{P}(D \mid C)\,\mathbb{P}(C)}\\ &= \frac{(.05)\,(.25)}{(.05)\,(.25)+(.04)\,(.35)+(.02)\,(.4)}\\ &= .362.\end{aligned}$$

Solution to Exercise 4.8: Let C be the vent the students the problem correct and K the event the students knows the answer. Using Bayes' theorem we have

$$\begin{aligned}
&P(K \mid C) \\
&= \frac{P(C \mid K)P(K)}{P(C)} \\
&= \frac{P(C \mid K)P(K)}{P(C \mid K)P(K) + P(C \mid \text{Eliminates})P(\text{Eliminates}) + P(C \mid \text{Guess})P(\text{Guess})} \\
&= \frac{1 \cdot \frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4}} = \frac{24}{31} \approx .774 = 77.4\%.
\end{aligned}$$

Solution to Exercise 4.9: Let + signiffy a positive test result, and D means dissease is present. Then

$$\begin{aligned}
\mathbb{P}(D \mid +) &= \frac{\mathbb{P}(+ \mid D)\, P(D)}{P(+ \mid D)\, P(D) + P(+ \mid D^c)P(D^c)} \\
&= \frac{(.95)\,(.01)}{(.95)\,(.01) + (.005)\,(.99)} \\
&= .657.
\end{aligned}$$

CHAPTER 5

Random variables

5.1. Introduction

A random variable is a real-valued function on S. Random variables are usually denoted by $X, Y, Z, \ldots$

Example 5.1: If one rolls a die, let X denote the outcome (i.e., either 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.

A discrete random variable is one that can only take countably many values. For a discrete random variable, we define the probability mass function or the density by $p(x) = \mathbb{P}(X = x)$. Here $\mathbb{P}(X = x)$ is an abbreviation for $\mathbb{P}(\{\omega \in S : X(\omega) = x\})$. This type of abbreviation is standard. Note $\sum_i p(x_i) = 1$ since X must equal something.

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

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

to be the expected value or expectation or mean of X.

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?

Answer.

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

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?

Answer. 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 $\mathbb{E}\, X$?

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

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

Let's give a proof of the linearity of expectation in the case when X and Y both take only finitely many values.

Let $Z = X + Y$, let $a_1, \dots, a_n$ be the values taken by X, $b_1, \dots, b_m$ be the values taken by Y, and $c_1, \dots, c_\ell$ the values taken by Z. Since there are only finitely many values, we can interchange the order of summations freely.

We write

$$\begin{aligned}
\mathbb{E}\, Z &= \sum_{k=1}^{\ell} c_k\mathbb{P}(Z = c_k) = \sum_{k=1}^{\ell}\sum_{i=1}^{n} c_k\mathbb{P}(Z = c_k, X = a_i) \\
&= \sum_k\sum_i c_k\mathbb{P}(X = a_i, Y = c_k - a_i) \\
&= \sum_k\sum_i\sum_{j=1}^{m} c_k\mathbb{P}(X = a_i, Y = c_k - a_i, Y = b_j) \\
&= \sum_i\sum_j\sum_k c_k\mathbb{P}(X = a_i, Y = c_k - a_i, Y = b_j).
\end{aligned}$$

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

$$\begin{aligned}\sum_k c_k \mathbb{P}(X = a_i, Y = c_k - a_i, Y = b_j) \\ &= \sum_k (a_i + b_j)\mathbb{P}(X = a_i, Y = c_k - a_i, Y = b_j) \\ &= (a_i + b_j)\mathbb{P}(X = a_i, Y = b_j).\end{aligned}$$

Substituting,

$$\begin{aligned}\mathbb{E}\, Z &= \sum_i \sum_j (a_i + b_j)\mathbb{P}(X = a_i, Y = b_j) \\ &= \sum_i a_i \sum_j \mathbb{P}(X = a_i, Y = b_j) + \sum_j b_j \sum_i \mathbb{P}(X = a_i, Y = b_j) \\ &= \sum_i a_i \mathbb{P}(X = a_i) + \sum_j b_j \mathbb{P}(Y = b_j) \\ &= \mathbb{E}\, X + \mathbb{E}\, Y.\end{aligned}$$

It turns out there is a formula for the expectation of random variables like X^2 and e^X. To see how this works, let us first look at an example.

Suppose we roll a die and let X be the value that is showing. We want the expectation $\mathbb{E}\, X^2$. 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

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

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

Theorem 5.1: $\mathbb{E}\, g(X) = \sum g(x)p(x)$.

Proof. Let $Y = g(X)$. Then

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

$\square$

Example 5.8: $\mathbb{E}\, X^2 = \sum x^2 p(x)$.

$\mathbb{E}\, X^n$ is called the nth moment of X. If $M = \mathbb{E}\, X$, then

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

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

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

Example 5.9: 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 $\mathrm{Var} X = \mathbb{E}\, X^2 = (1)^2 \frac{1}{2} + (-1)^2 \frac{1}{2} = 1$.

Example 5.10: 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

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

Note that the expectation of a constant is just the constant. An alternate expression for the variance is

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

5.2. Further examples and applications

5.2.1. Random Variables. 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. A random variable is thus a function $X : S \to \mathbb{R}$ and we can think of it as a numerical value that is random. We will use capital letters to denote random variables.

Example 5.11: 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 not on the sample space $S = \{H, T\}$. This shows the need to covert outcomes to numerical values.

Example 5.12: 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 5.13: 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?

Answer.

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

Example 5.14: Toss a coin n times. Let X be the number of heads that occur. This random variable can take the values $0, 1, 2, \ldots, n$. From the binomial formula we can conclude that $\mathbb{P}(X = k) = \frac{1}{2^n}\binom{n}{k}$.

5.2.2. Discrete Random Variables.

Definition 5.2: A random variable that has countably many possible values is called a **discrete random variable**.

Example 5.15: 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

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

Oftentimes the probability mass function (p.m.f.) will already be given and we can then use it to compute probabilities.

Example 5.16: The p.m.f. of X is given by $p_X(i) = e^{-\lambda}\frac{\lambda^i}{i!}$ for $i = 0, 1, 2, \ldots$ and some positive real number λ.

(a) Find $\mathbb{P}(X = 0)$.
Answer. By definition we have $\mathbb{P}(X = 0) = p_X(0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-\lambda}$.

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

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

5.2.3. Expected Value. One of the most important concepts in probability is that of expectation. Given a random variable X, one can ask what is the average value of X, that is, what is the **expected value** of X. For a random variable X with p.m.f. $p_X(x)$, we defined its **expectation**, or **expected value** of X as

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

This formula works if the state space S is countable.

Example 5.17: 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 5.18: 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 5.19: 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.

Answer. 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,

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

5.2.4. The cumulative distribution function (c.d.f.)

Definition 5.3: Let X be a random variable. The cumulative distribution function (c.d.f.), 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 p.m.f. , then

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

Example 5.20: Suppose that X has the following p.m.f.:

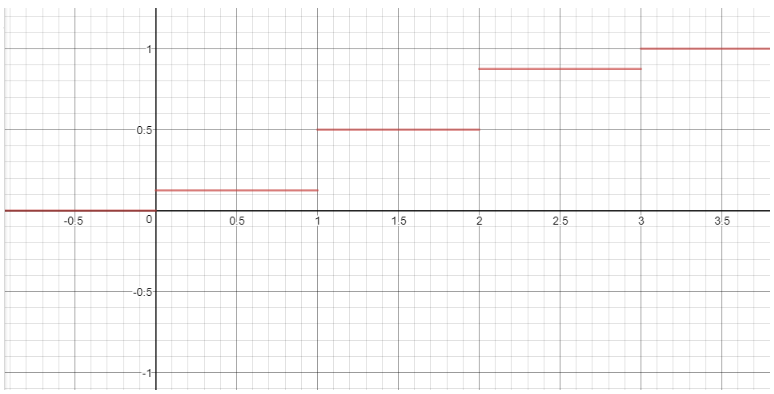
$$\begin{aligned}
p_X(0) = \mathbb{P}(X = 0) &= \frac{1}{8} \\
p_X(1) = \mathbb{P}(X = 1) &= \frac{3}{8} \\
p_X(2) = \mathbb{P}(X = 2) &= \frac{3}{8} \\
p_X(3) = \mathbb{P}(X = 3) &= \frac{1}{8}.
\end{aligned}$$

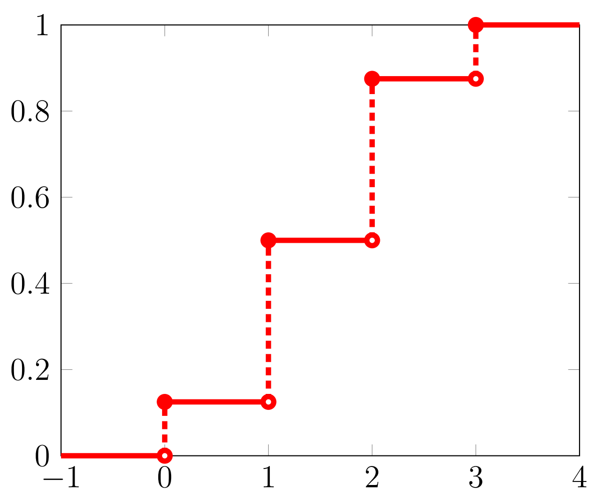
Find the c.d.f of X and plot the graph of the c.d.f.

Answer. Summing up the probabilities up to the value of x we get the following:

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

Two graphs of this function are given here:





Note that this is a step function.

Here are some properties of the c.d.f.:

1. F is nondecreasing , that is if $x < y$, then $F(x) \le F(y)$.
2. $\lim_{x\to\infty} F(x) = 1$.
3. $\lim_{x\to-\infty} F(x) = 0$.
4. F is right continuous. That is for any decreasing sequence $x_n \downarrow x$, then $\lim\limits_{n\to\infty} F_X(x_n) = F_X(x)$.

Example 5.21: Let X have distribution

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

(a) Compute $\mathbb{P}(X < 3)$.
Answer. We have that $\mathbb{P}(X < 3) = \lim_{n \to \infty} \mathbb{P}\left(X \leq 3 - \frac{1}{n}\right) = \lim_{n \to \infty} F_X\left(3 - \frac{1}{n}\right) = \frac{11}{12}$.

(b) Compute $\mathbb{P}(X = 1)$.
Answer. We have that

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

(c) Compute $\mathbb{P}(2 < X \leq 4)$.
Answer. We have that

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

5.2.5. Expectated Value of Sums of Random Variables. Recall our current definition of $\mathbb{E}X$. If, informally, we can list out $X = x_1, x_2, \ldots$, the probability mass function of X will be given by $p_X(x_i)$, $i = 1, 2, \ldots$ and $\mathbb{E}X = \sum_{i=1}^{\infty} x_i p(x_i)$. In this section we will introduce another definition of expectation. It will allow us to prove the linearity of the expectation in a different way. That is, the goal is to show (again) that if $Z = X + Y$ then $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$.

If S is a countable sample space, then

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

In a little bit we will prove that this definition is consistent with the previous definition. But first, let us do some examples.

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

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

The difference between the two definitions are the following.

- We list all the values that X can attain and use the $p_X(x)$ for only those values. In a sense we are looking at the range of X.
- We list all the possible outcomes that are in the domain of X.

Proposition 5.4: If X is a discrete random variable and S is countable, then the two definitions are equivalent.

Proof. We start with the first definition. Let $X = x_1, x_2, \ldots$

$$\begin{aligned}\mathbb{E}X &= \sum_{x_i} x_i p(x_i) = \sum_{x_i} x_i \mathbb{P}(X = x_i) = \sum_{x_i} x_i \sum_{\omega\in\{\omega: X(\omega)=x_i\}} \mathbb{P}(\omega) \\ &= \sum_{x_i} \sum_{\omega\in\{\omega: X(\omega)=x_i\}} x_i \mathbb{P}(\omega) = \sum_{x_i} \sum_{\omega\in\{\omega: X(\omega)=x_i\}} X(\omega)\mathbb{P}(\omega) \\ &= \sum_{\omega\in S} X(\omega)\mathbb{P}(\omega),\end{aligned}$$

where we used that each $S_i = \{\omega : X(\omega) = x_i\}$ are mutually exclusive events that union up to S.

Using this definition, we can easily prove linearity of the expectation.

Theorem 5.5: (Linearity) If X and Y are discrete random variables and $a \in \mathbb{R}$ then

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

Proof. (i) We have that

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

(ii) If $a \in \mathbb{R}$, then

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

Using induction, linearity holds a sequence of random variables $X_1, X_2, \ldots, X_n$.

Theorem 5.6: If $X_1, X_2, \ldots, X_n$ are random variables, then

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n].$$

5.2.6. 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 5.23: Let X be a random variable whose p.m.f. is given by

$$\begin{aligned}\mathbb{P}(X=-1) &= .2,\\ \mathbb{P}(X=0) &= .5\\ \mathbb{P}(X=1) &= .3\end{aligned}$$

Let $Y = X^2$. Find $\mathbb{E}[Y]$.

Answer. Note that Y takes the values 0^2, $(-1)^2$ and 1^2, which reduce to 0 or 1. Also notice that $p_Y(1) = .2 + .3 = .5$ and $p_Y(0) = .5$. Thus, $\mathbb{E}[Y] = 0 \cdot .5 + 1 \cdot .5 = .5$.

Note that $\mathbb{E}X^2 = .5$. While $(\mathbb{E}X)^2 = .01$ since $\mathbb{E}X = .3 - .2 = .1$. 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 5.1. 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

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

Note that

$$\mathbb{E}X^2 = \sum x_i^2 p_X(x_i)$$

will be useful.

Example 5.24: Let us revisit the previous example. Let X denote a random variable such that

$$\begin{aligned}\mathbb{P}(X=-1) &= .2\\ \mathbb{P}(X=0) &= .5\\ \mathbb{P}(X=1) &= .3\end{aligned}$$

Let $Y = X^2$. Find $\mathbb{E}Y$.

Answer. We have that $\mathbb{E}X^2 = \sum x_i^2 p_X(x_i) = (-1)^2(.2) + 0^2(.5) + 1^2(.3) = .5$.

Definition 5.7: The quantity $\mathbb{E}X^n$ for $n \geq 1$, is called the n**th moment** of X. The first moment, that is $\mathbb{E}X$, is also called the **mean** of X.

From Theorem 5.1 we know that the nth moment can be calculated by

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

5.2.7. 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.

Recall the alternative formula to compute the variance

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

Example 5.25: Calculate $\operatorname{Var}(X)$ if X represents the outcome when a fair die is rolled.

Answer. Previously we calculated that $\mathbb{E}X = \frac{7}{5}$. Thus we only need to calculate 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

$$\operatorname{Var}(X) = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2 = \frac{91}{6} - \left(\frac{7}{5}\right)^2 = \frac{35}{12}.$$

Another useful formula is the following.

Proposition 5.8: For constants $a, b \in \mathbb{R}$ we have that $\operatorname{Var}(aX + b) = a^2\operatorname{Var}(X)$.

Proof. We compute

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

Definition 5.9: The **standard deviation** of X is defined as

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

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 it's 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: $p(0)=\frac{1}{3}$, $p(1)=\frac{1}{6}$, $p(2)=\frac{1}{4}$, $p(3)=\frac{1}{4}$. Find its expected value, variance, and standard deviation. Also, plot its c.d.f.

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) $\mathbb{E}\left[X^2\right]$
(b) $\mathbb{E}[3X+2]$
(c) $\mathbb{E}\left[(X+2)^2\right]$
(d) $\operatorname{Var}[-X]$
(e) $SD(2X)$.

Exercise 5.11: Does there exist a random variable X such that $\mathbb{E}[X]=4$ and $\mathbb{E}\left[X^2\right]=10$? 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 Bernouilli 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,

$$\mathbb{P}(X = -2) = \mathbb{P}(2\spadesuit) = \frac{\binom{13}{2}}{\binom{52}{2}} \qquad \mathbb{P}(X = -1) = \mathbb{P}(1\spadesuit \text{ and } 1(\diamondsuit \text{ or } \clubsuit)) = \frac{13 \cdot 26}{\binom{52}{2}}$$

$$\mathbb{P}(X = 0) = \mathbb{P}(2(\diamondsuit \text{ or } \clubsuit)) = \frac{\binom{26}{2}}{\binom{52}{2}} \qquad \mathbb{P}(X = 1) = \mathbb{P}(1\heartsuit \text{ and } 1\spadesuit) = \frac{13 \cdot 13}{\binom{52}{2}}$$

$$\mathbb{P}(X = 2) = \mathbb{P}(1\heartsuit \text{ and } 1(\diamondsuit \text{ or } \clubsuit)) = \mathbb{P}(X = -1) \quad \mathbb{P}(X = 4) = \mathbb{P}(2\heartsuit) = \mathbb{P}(X = -2).$$

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.

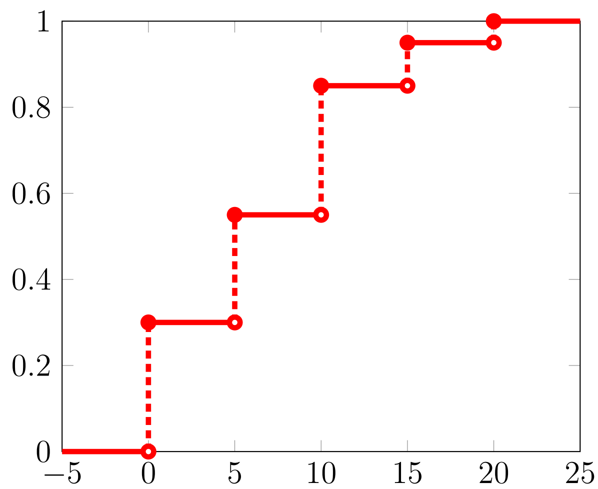
Solution to Exercise 5.3: The random variable X can take the values $0, 5, 10, 15$ and 20 depending on which test was applied to each bank, and if the bank fails the evaluation or not. Let us write B_i ="i-th bank fails" and T_i="test i applied". Then, $\mathbb{P}(T_1) = \mathbb{P}(T_2) = 0.5$, $\mathbb{P}(B_1) = \mathbb{P}(B_1 \mid T_1) = \mathbb{P}(B_1 \mid T_2) = 0.4$ and $\mathbb{P}(B_2) = \mathbb{P}(B_2 \mid T_1) = \mathbb{P}(B_2 \mid T_2) = 0.5$. 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 probability distribution function is given by

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

The graph of the probability distribution function is:



Solution to Exercise 5.4: Let X denote the number of students in the bus that carries the random 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): Then

$$\mathbb{P}(X=4)=\mathbb{P}(BB)=\frac{\binom{4}{2}}{\binom{14}{2}}=\frac{6}{91} \qquad \mathbb{P}(X=0)=\mathbb{P}(OO)=\frac{\binom{2}{2}}{\binom{14}{2}}=\frac{1}{91}$$

$$\mathbb{P}(X=2)=\mathbb{P}(BO)=\frac{\binom{4}{1}\binom{2}{1}}{\binom{14}{2}}=\frac{8}{91} \qquad \mathbb{P}(X=-1)=\mathbb{P}(WO)=\frac{\binom{8}{1}\binom{2}{1}}{\binom{14}{2}}=\frac{16}{91}$$

$$\mathbb{P}(X=1)=\mathbb{P}(BW)=\frac{\binom{4}{1}\binom{8}{1}}{\binom{14}{2}}=\frac{32}{91} \qquad \mathbb{P}(X=-2)=\mathbb{P}(WW)=\frac{\binom{8}{2}}{\binom{14}{2}}=\frac{28}{91}$$

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 $\mathbb{E}X = 35 \cdot \left(\frac{1}{38}\right) - 1 \cdot \frac{37}{38} = -\$.0526$.

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 $\mathbb{E}X = 1 \cdot \left(\frac{18}{38}\right) - 1 \cdot \frac{20}{38} = -\$.0526$.

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

$$\mathbb{E}X = (200 - 10,000) \cdot (.08) + 200 \cdot (.92) = -600.$$

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

$$100 = \mathbb{E}X = (P - 10,000) \cdot (.08) + P \cdot (.92)$$

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

Solution to Exercise 5.9: Let's apply the formulas

$$\mathbb{E}X = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{34}{24}.$$

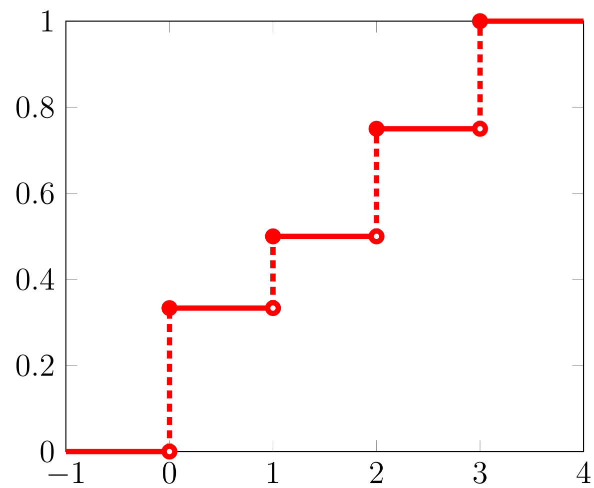
Now to calculate variance we have

$$\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 \\
&= 0^2 \cdot \frac{1}{3} - 1^2\frac{1}{6} + 2^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{1}{4} - \left(\frac{34}{24}\right)^2 \\
&= \frac{82}{24} - \frac{34^2}{24^2} \\
&= \frac{812}{24^2}.
\end{aligned}$$

Taking the square root gives us

$$\mathrm{SD}(X) = \frac{2\sqrt{203}}{24}.$$

The plot of the c.d.f. is:



Solution to Exercise 5.10(A): Since $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 = 12$ then

$$\mathbb{E}\left[X^2\right] = \mathrm{Var}(X) + (\mathbb{E}X)^2 = 12 + 50^2 = 2512.$$

Solution to Exercise 5.10(B):

$$\mathbb{E}\left[3X + 2\right] = 3\mathbb{E}\left[X\right] + \mathbb{E}\left[2\right] = 3 \cdot 50 + 2 = 152.$$

Solution to Exercise 5.10(C):

$$\mathbb{E}\left[(X+2)^2\right] = \mathbb{E}\left[X^2\right] + 4\mathbb{E}\left[X\right] + 4 = 2512 + 4 \cdot 50 + 4 = 2716.$$

Solution to Exercise 5.10(D):

$$\mathrm{Var}\left[-X\right] = (-1)^2\,\mathrm{Var}(X) = 12$$

Solution to Exercise 5.10(E):

$$SD\left(2X\right) = \sqrt{\mathrm{Var}(2X)} = \sqrt{2^2\mathrm{Var}(X)} = \sqrt{48} = 2\sqrt{12}.$$

Solution to Exercise 5.11: Using the hint let's compute the variance of this random variable which would be $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 = 10 - 4^2 = -6$. But we know a random variable cannot have a negative variance. Thus no such random variable exists.

CHAPTER 6

Some discrete distributions

6.1. Introduction

Bernoulli distribution. A r.v. X such that $\mathbb{P}(X=1)=p$ and $\mathbb{P}(X=0)=1-p$ is said to be a Bernoulli r.v. with parameter p. Note $\mathbb{E}\,X=p$ and $\mathbb{E}\,X^2=p$, so $\mathrm{Var}X=p-p^2=p(1-p)$.

Binomial distribution. A r.v. X has a binomial distribution with parameters n and p if $\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$. The number of successes in n trials is a binomial. 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's, so $\mathbb{E}\,X=\mathbb{E}\,Y_1+\cdots+\mathbb{E}\,Y_n=np$. We haven't defined what it means for r.v.'s to be independent, but here we mean that the events $(Y_k=1)$ are independent. The cumbersome way is as follows.

$$
\begin{aligned}
\mathbb{E}\,X&=\sum_{k=0}^{n}k\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=1}^{n}k\binom{n}{k}p^k(1-p)^{n-k}\\
&=\sum_{k=1}^{n}k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\
&=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!((n-1)-(k-1))!}p^{k-1}(1-p)^{(n-1)-(k-1)}\\
&=np\sum_{k=0}^{n-1}\frac{(n-1)!}{k!((n-1)-k)!}p^k(1-p)^{(n-1)-k}\\
&=np\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{(n-1)-k}=np.
\end{aligned}
$$

To get the variance of X, we have

$$
\mathbb{E}\,X^2=\sum_{k=1}^{n}\mathbb{E}\,Y_k^2+\sum_{i\neq j}\mathbb{E}\,Y_iY_j.
$$

Now

$$
\begin{aligned}
\mathbb{E}\,Y_iY_j&=1\cdot\mathbb{P}(Y_iY_j=1)+0\cdot\mathbb{P}(Y_iY_j=0)\\
&=\mathbb{P}(Y_i=1,Y_j=1)=\mathbb{P}(Y_i=1)\mathbb{P}(Y_j=1)=p^2
\end{aligned}
$$

using independence. The square of $Y_1 + \cdots + Y_n$ 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_iY_j with $i \neq j$. Hence

$$\mathrm{Var} X = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = np + (n^2 - n)p^2 - (np)^2 = np(1-p).$$

Later we will see that the variance of the sum of independent r.v.'s is the sum of the variances, so we could quickly get $\mathrm{Var} X = np(1-p)$. 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. X is Poisson with parameter λ if

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

Note $\sum_{i=0}^{\infty} \lambda^i / i! = e^{\lambda}$, so the probabilities add up to one.

To compute expectations,

$$\mathbb{E}\, X = \sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda.$$

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 $\mathrm{Var} X = \lambda$.

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?

Answer. If X is the number of homicides, 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's the probability that next year there will be exactly 2 large earthquakes?

Answer. $\lambda = \mathbb{E}\, X = 1$, so $\mathbb{P}(X = 2) = e^{-1}(\frac{1}{2})$.

We have the following proposition.

Proposition 6.1: If X_n is binomial 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.

The above 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 Poissons.

Proof. For simplicity, let us suppose $\lambda = np_n$. In the general case we use $\lambda_n = np_n$. We write

$$\begin{aligned}\mathbb{P}(X_n = i) &= \frac{n!}{i!(n-i)!} p_n^i (1-p_n)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{i!} \Big(\frac{\lambda}{n}\Big)^i \Big(1-\frac{\lambda}{n}\Big)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}.\end{aligned}$$

The first factor tends to 1 as $n \to \infty$. $(1-\lambda/n)^i \to 1$ as $n \to \infty$ and $(1-\lambda/n)^n \to e^{-\lambda}$ as $n \to \infty$. $\square$

Uniform distribution. Let $\mathbb{P}(X = k) = \frac{1}{n}$ for $k = 1, 2, \ldots, n$. This is the distribution of the number showing on a die (with $n = 6$), for example.

Geometric distribution. Here $\mathbb{P}(X = i) = (1-p)^{i-1}p$ for $i = 1, 2, \ldots$. In Bernoulli trials, if we let X be the first time we have a success, then X will be geometric. 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 $(1-p)^{k-1}p$. Since $\sum_{n=0}^\infty nr^n = 1/(1-r)^2$ (differentiate the formula $\sum r^n = 1/(1-r)$), we see that $\mathbb{E}\, X = 1/p$. Similarly we have $\mathrm{Var} X = (1-p)/p^2$.

Negative binomial distribution. Let r and p be parameters and set

$$\mathbb{P}(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \qquad n = r, r+1, \ldots.$$

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. Set

$$\mathbb{P}(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}.$$

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.

6.2. Further examples and applications

6.2.1. Bernouli 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, calculate the probability that there are fewer than 3 hurricanes in a 20-year period.

Answer. We have that $X \sim bin(20, .05)$ then

$$\begin{aligned}
\mathbb{P}(X<3) &= \mathbb{P}(X \leq 2) \\
&= \binom{20}{0}(.05)^0(.95)^{20} + \binom{20}{1}(.05)^1(.95)^{19} + \binom{20}{2}(.05)^2(.95)^{18} \\
&= .9245.
\end{aligned}$$

Example 6.4: Phan has a .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?

Answer. If $X \sim bin(10, .6)$ then

$$\begin{aligned}
\mathbb{P}(X \geq 2) &= 1 - \mathbb{P}(X=0) - \mathbb{P}(X=1) \\
&= 1 - \binom{10}{0}(.6)^0(.4)^{10} - \binom{10}{1}(.6)^1(.4)^9 \\
&= .998.
\end{aligned}$$

6.2.2. The Poisson Distribution.

Here are some examples that usually obey Poisson distribution and so can be modeled as Poisson r.v.:

(1) The number of misprints on a random page of a book.
(2) The number of 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.

All of these are Poisson for the same reason. Each event has 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. Here we are using the fact that the Poisson distribution approximates the binomial distribution.

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?

Answer. Let X be the number of texts in a 9−minute period. Then $\lambda = 3 \cdot 2 = 6$ and

$$\begin{aligned}\mathbb{P}(X \geq 5) &= 1 - \mathbb{P}(X \leq 4) \\ &= 1 - \sum_{n=0}^{4} \frac{6^n e^{-6}}{n!} \\ &= 1 - .285 = .715.\end{aligned}$$

Example 6.6: Let $X_1, ..., X_k$ be independent Poisson r.v., each with expectation λ_1. What is the distribution of the r.v. $Y = X_1 + ... + X_k$?

Answer. The distribution of Y is Poisson with expectation $\lambda = k\lambda_1$. To show this, we consider Proposition 6.1 where, in two different ways, we sum $n = mk$ Bernoulli r.v. with parameter $p_n = k\lambda_1/n = \lambda_1/m = \lambda/n$. 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 separate the same $n = mk$ Bernoulli r.v. in k groups, each group having m Bernoulli r.v. Then the limit gives us the distribution of $X_1 + ... + X_k$.

Example 6.7: Let $X_1, \ldots, X_k$ are independent Poisson r.v., each with expectation $\lambda_1, \ldots, \lambda_k$, respectively. What is the distribution of the r.v. $Y = X_1 + ... + X_k$?

Answer. The distribution of Y is Poisson with expectation $\lambda = \lambda_1 + ... + \lambda_k$. To show this, we again consider Proposition 6.1 with parameter $p_n = \lambda/n$. If n is large, we can separate these n Bernoulli r.v. in k groups, each having $n_i \approx \lambda_i n/\lambda$ Bernoulli r.v. 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 so 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 r.v.?

Answer. Yes, and this is really simple. Recall again from Proposition 6.1 that we can approximate Poisson Y with parameter λ by binomial r.v. with parameters $(n, p_n = \lambda/n)$. Each such 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	Abbrev.	Parameters	p.m.f. ($k \in \mathbb{N}_0$)	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
Bernoulli	$\mathrm{Ber}(p)$	$p \in [0,1]$	$\binom{1}{k} p^k (1-p)^{1-k}$	p	$p(1-p)$
Binomial	$\mathrm{bin}(n,p)$	$n \in \mathbb{N}$ $p \in [0,1]$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Poisson	$\mathrm{Pois}(\lambda)$	$\lambda > 0$	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ
Geometric	$\mathrm{Geo}(p)$	$p \in (0,1)$	$\begin{cases} (1-p)^{k-1} p, & \text{for } k \geq 1, \\ 0, & \text{else.} \end{cases}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
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 .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. Calculate 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: $\binom{6}{4}(.75)^4(.25)^2$

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): $\mathbb{P}(X\geq 3)=1-\mathbb{P}(X\leq 2)=1-\binom{7}{0}\left(\frac{4}{9}\right)^0\left(\frac{5}{9}\right)^7-\binom{7}{1}\left(\frac{4}{9}\right)^1\left(\frac{5}{9}\right)^6-\binom{7}{2}\left(\frac{4}{9}\right)^2\left(\frac{5}{9}\right)^5$

Solution to Exercise 6.5(A): $e^{-.2}$

Solution to Exercise 6.5(B): $1-e^{-.2}-.2e^{-.2}=1-1.2e^{-.2}$.

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 $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - e^{.5}\frac{(-.5)^0}{0!} - e^{.5}\frac{(-.5)^1}{1!}$.

Solution to Exercise 6.8(B): In two weeks the average number of accidents will be $\lambda = .5 + .5 = 1$. 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 $\lambda = 4 \cdot (.5) = 2$. 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 .8, so we take this as our λ . We are asking for $\mathbb{P}(R = 3)$, which is $e^{-.8}\frac{(.8)^3}{3!} \approx .0383$.

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 $\mathbb{P}(C \geq 1)$, which is $1 - \mathbb{P}(C = 0) = 1 - e^{-1.2}\frac{(1.2)^0}{0!} \approx .6988$.

Solution to Exercise 6.9(C): This calls for a Poisson random variable B. The average number of bits per cookie is $.8 + 1.2 = 2$, so we take this as our λ. We are asking for $\mathbb{P}(B \leq 2)$, 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 $X \sim \operatorname{bin}(5000, p)$ with $p = 1/17000 \approx 0.00029$. The binomial distribution gives us

$$\mathbb{P}(X=0) = \left(\frac{16999}{17000}\right)^{5000} \approx 0.745 \quad \mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X=0) = 1 - \left(\frac{16999}{17000}\right)^{5000} \approx 0.255$$

$$\mathbb{P}(X > 1) = \mathbb{P}(X \geqslant 1) - \mathbb{P}(X = 1) =$$

$$= 1 - \left(\frac{16999}{17000}\right)^{5000} - \begin{pmatrix} 5000 \\ 1 \end{pmatrix} \left(\frac{16999}{17000}\right)^{4999} \left(\frac{1}{17000}\right)^1 \approx 0.035633$$

Approximating the distribution of X by a Poisson with parameter $\lambda = \frac{5000}{17000} = \dfrac{5}{17}$ 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

$$\mathbb{P}(X \geq 2) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) = \frac{\binom{3}{2} \cdot 17}{\binom{20}{5}} + \frac{1}{\binom{20}{5}}.$$

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, $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-0.2251}$.

(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, $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-2500}$.

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. Introduction

A r.v. X is said to have a continuous distribution if there exists a nonnegative function f such that

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$$

for every a and b. (More precisely, such an X is said to have an absolutely continuous distribution.) f is called the density function for X. Note $\int_{-\infty}^{\infty} f(x)dx = \mathbb{P}(-\infty < X < \infty) = 1$. In particular, $\mathbb{P}(X = a) = \int_a^a f(x)dx = 0$ for every a.

Example 7.1: Suppose we are given $f(x) = c/x^3$ for $x \geq 1$. 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 $c = 2$.

Define $F(y) = \mathbb{P}(-\infty < X \leq y) = \int_{-\infty}^{y} f(x)dx$. F is called the distribution function of X. We can define F for any random variable, not just continuous ones, by setting $F(y) = \mathbb{P}(X \leq y)$. In the case of discrete random variables, this 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 by

$$\mathbb{E}\, X = \int_{-\infty}^{\infty} xf(x)dx.$$

In the example above,

$$\mathbb{E}\, X = \int_1^{\infty} x\frac{2}{x^3}dx = 2\int_1^{\infty} x^{-2}dx = 2.$$

We give another definition of the expectation in the continuous case. First suppose X is nonnegative. Define $X_n(\omega)$ to be $k/2^n$ if $k/2^n \leq X(\omega) < (k+1)/2^n$. We are approximating X from below by the largest multiple of 2^{-n}. Each X_n is discrete and the X_n increase to X. We define $\mathbb{E}\, X = \lim_{n\to\infty} \mathbb{E}\, X_n$.

Let us argue that this agrees with the first definition in this case. We have

$$\begin{aligned}\mathbb{E}\, X_n &= \sum_{k/2^n} \frac{k}{2^n}\mathbb{P}(X_n = k/2^n)\\ &= \sum_{k/2^n} \frac{k}{2^n}\mathbb{P}(k/2^n \le X < (k+1)/2^n)\\ &= \sum \frac{k}{2^n}\int_{k/2^n}^{(k+1)/2^n} f(x)dx\\ &= \sum \int_{k/2^n}^{(k+1)/2^n} \frac{k}{2^n} f(x)dx.\end{aligned}$$

If $x \in [k/2^n, (k+1)/2^n)$, then x differs from $k/2^n$ by at most $1/2^n$. So the last integral differs from

$$\sum \int_{k/2^n}^{(k+1)/2^n} x f(x)dx$$

by at most $\sum (1/2^n)\mathbb{P}(k/2^n \le X < (k+1)/2^n) \le 1/2^n$, which goes to 0 as $n \to \infty$. On the other hand,

$$\sum \int_{k/2^n}^{(k+1)/2^n} x f(x)dx = \int_0^M x f(x)dx,$$

which is how we defined the expectation of X.

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 $\mathbb{E}\, X$.

To show linearity, if X and Y are bounded positive random variables, then 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

$$\begin{aligned}\mathbb{E}\,(X+Y) &= \lim_{m\to\infty} \mathbb{E}\,(X_m + Y_m)\\ &= \lim_{m\to\infty} \mathbb{E}\, X_m + \lim_{m\to\infty} \mathbb{E}\, Y_m = \mathbb{E}\, X + \mathbb{E}\, Y.\end{aligned}$$

If X is not necessarily positive, we have a similar definition; we will not do the details. This second definition of expectation is mostly useful for theoretical purposes and much less so for calculations.

Similarly to the discrete case, we have

Proposition 7.1: $\mathbb{E}\, g(X) = \int g(x)f(x)dx$.

As in the discrete case,

$$\mathrm{Var} X = \mathbb{E}\,[X - \mathbb{E}\, X]^2.$$

As an example of these calculations, let us look at the 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 $a \le x \le b$ and 0 otherwise.

To calculate the expectation of X,

$$\begin{aligned}\mathbb{E}\,X &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx \\ &= \frac{1}{b-a}\int_a^b x\,dx \\ &= \frac{1}{b-a}\Big(\frac{b^2}{2} - \frac{a^2}{2}\Big) = \frac{a+b}{2}.\end{aligned}$$

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

$$\mathrm{Var} X = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 = \frac{(b-a)^2}{12}.$$

7.2. Further examples and applications

Example 7.2: Suppose X has the following p.d.f.

$$f(x)=\begin{cases}\frac{2}{x^3} & x\geq 1\\ 0 & x\leq 1.\end{cases}$$

Find the c.d.f of X, that is, find $F_X(x)$. Use this c.d.f to find $\mathbb{P}(3\leq X\leq 4)$.

Answer. We have $F_X(x)=0$ if $x\leqslant 1$ and will need to compute

$$F_X(x)=\mathbb{P}(X\leq x)=\int_1^x \frac{2}{y^3}dy=1-\frac{1}{x^2}$$

when $x\geqslant 1$. We can use this formula to find the following probability:

$$\mathbb{P}(3\leq X\leq 4)=\mathbb{P}(X\leq 4)-\mathbb{P}(X<3)=F_X(4)-F_X(3)=\left(1-\frac{1}{4^2}\right)-\left(1-\frac{1}{3^2}\right)=\frac{7}{144}.$$

Example 7.3: Suppose X has density

$$f(x)=\begin{cases}2x & 0\leq x\leq 1\\ 0 & \text{otherwise}\end{cases}.$$

Find $\mathbb{E}X$.

Answer. We have that

$$\mathbb{E}[X]=\int xf(x)dx=\int_0^1 x\cdot 2x\,dx=\frac{2}{3}.$$

Example 7.4: The density of X is given by

$$f(x)=\begin{cases}\frac{1}{2} & \text{if } 0\leq x\leq 2\\ 0 & \text{otherwise}\end{cases}.$$

Find $\mathbb{E}\left[e^X\right]$.

Answer. Using proposition 7.1 with $g(x)=e^x$ we have

$$\mathbb{E}e^X=\int_0^2 e^x\cdot\frac{1}{2}dx=\frac{1}{2}\left[e^2-1\right].$$

Example 7.5: Suppose X has density

$$f(x)=\begin{cases}2x & 0\leq x\leq 1\\ 0 & \text{otherwise}\end{cases}.$$

Find $\mathrm{Var}(X)$.

Answer. From Example 7.3 we found $\mathbb{E}[X]=\frac{2}{3}$. Now

$$\mathbb{E}\left[X^2\right]=\int_0^1 x^2\cdot 2xdx=2\int_0^1 x^3dx=\frac{1}{2}.$$

Thus

$$\operatorname{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

Example 7.6: Suppose X has density

$$f(x) = \begin{cases} ax+b & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

and that $\mathbb{E}\left[X^2\right] = \frac{1}{6}$. Find the values of a and b.

Answer. 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 give 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

$$f(x)=\begin{cases} cx\,(5-x) & 0\leq x\leq 5,\\ 0 & \text{otherwise.}\end{cases}$$

(a) What is the value of c?
(b) What is the cumulative distribution function of X? That is, find $F_X(x)=\mathbb{P}\,(X\leq x)$.
(c) Use your answer in part (b) to find $\mathbb{P}\,(2\leq X\leq 3)$.
(d) What is $\mathbb{E}\,[X]$?
(e) What is $\operatorname{Var}(X)$?

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

$$f(x)=\begin{cases} \frac{10}{x^2} & x\geq 10,\\ 0 & x\leq 10.\end{cases}$$

(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 $F_X(x)=\mathbb{P}\,(X\leq x)$.
(c) Use part (b) to help you find $\mathbb{P}\,(X\geq 35)$?

Exercise 7.3: Suppose the random variable X has a density function

$$f(x)=\begin{cases} \frac{2}{x^2} & x>2,\\ 0 & x\leq 2.\end{cases}$$

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

$$f(x)=\begin{cases} a+bx^2 & 0\leq x\leq 1,\\ 0 & \text{otherwise.}\end{cases}$$

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 $\mathbb{E}[X] = 6\mathrm{Var}(X)$. Find the value of a.

Exercise 7.7: Suppose you order a pizza from your favorite pizzaria 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: You arrive at a bus stop at $10:05$ am, and the bus arrives at some (random) time uniformly distributed between $10:00$ and $10:20$. Given that when you arrive today to the station the bus is not there yet (you are lucky today), what is the probability that you have to wait more than 4 minutes? **Hint:** the event 'today you are lucky' can be expressed as $(X > 5)$, where X denotes the arrival time of the bus at the station (in minutes past $10:00$ a.m.).

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

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)dy \\ &= \int_0^x \frac{6}{125} y(5-y)\,dx = \frac{6}{125}\left[\left(\frac{5y^2}{2} - \frac{y^3}{3}\right)\right]_0^x \\ &= \frac{6}{125}\left(\frac{5x^2}{2} - \frac{x^3}{3}\right). \end{aligned}$$

Solution to Exercise 7.1(C): We have

$$\begin{aligned} \mathbb{P}(2 \leq X \leq 3) &= \mathbb{P}(X \leq 3) - \mathbb{P}(X < 2) \\ &= \frac{6}{125}\left(\frac{5 \cdot 3^2}{2} - \frac{3^3}{3}\right) - \frac{6}{125}\left(\frac{5 \cdot 2^2}{2} - \frac{2^3}{3}\right) \\ &= .296. \end{aligned}$$

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

$$\operatorname{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 = 7.5 - (2.5)^2 = 1.25.$$

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

$$F(x) = \mathbb{P}(X \leq x) = \int_{10}^{x} \frac{10}{y^2} dy = 1 - \frac{10}{x}$$

for $x > 10$, and $F(x) = 0$ for $x < 10$.

Solution to Exercise 7.2(C): We have

$$\begin{aligned}\mathbb{P}(X \geq 35) &= 1-\mathbb{P}(X<35) = 1-F_X(35)\\ &= 1-\left(1-\frac{10}{35}\right)=\frac{10}{35}.\end{aligned}$$

Solution to Exercise 7.5: 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 give us

$$\frac{7}{10} = \int_0^1 x\left(a+bx^2\right)dx = \frac{a}{2}+\frac{b}{4}.$$

Solving these equations gives us

$$a = \frac{1}{5}, \text{ and } b = \frac{12}{5}.$$

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

$$\begin{aligned}Var(X) &= \left(\frac{1}{3}a^2+\frac{1}{3}a+\frac{1}{3}\right)-\left(\frac{1}{2}a+\frac{1}{2}\right)^2\\ &= \frac{1}{12}a^2-\frac{1}{6}a+\frac{1}{12}.\end{aligned}$$

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

$$f(x) = \begin{cases} \frac{cx}{5} & 0 \le x \le 10, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\int_0^{10} c\frac{x}{5} dx = 10c$, we have $c = \frac{1}{10}$. Now, applying Proposition 7.1,

$$\mathbb{E}[Y] = \int_0^{10} \frac{3}{50} x^3 dx = 150.$$

CHAPTER 8

Normal distribution

8.1. Introduction

A r.v. is a standard normal (written $\mathcal{N}(0,1)$) if it has density

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

A synonym for normal is Gaussian. The first thing to do is show that this is a density. 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,

$$I^2 = \int_0^{\pi/2} \int_0^\infty re^{-r^2/2}dr = \pi/2.$$

So $I = \sqrt{\pi/2}$, hence $\int_{-\infty}^\infty e^{-x^2/2}dx = \sqrt{2\pi}$ as it should.

Note

$$\int xe^{-x^2/2}dx = 0$$

by symmetry, so $\mathbb{E}\,Z = 0$. For the variance of Z, we use integration by parts:

$$\mathbb{E}\,Z^2 = \frac{1}{\sqrt{2\pi}}\int x^2e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int x\cdot xe^{-x^2/2}dx.$$

The integral is equal to

$$-xe^{-x^2/2}\Big]_{-\infty}^{\infty} + \int e^{-x^2/2}dx = \sqrt{2\pi}.$$

Therefore $\mathrm{Var}Z = \mathbb{E}\,Z^2 = 1$.

We say X is a $\mathcal{N}(\mu,\sigma^2)$ if $X = \sigma Z + \mu$, where Z is a $\mathcal{N}(0,1)$. We see that

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \le x) = \mathbb{P}(\mu + \sigma Z \le x) \\ &= \mathbb{P}(Z \le (x-\mu)/\sigma) = F_Z((x-\mu)/\sigma) \end{aligned}$$

if $\sigma > 0$. (A similar calculation holds if $\sigma < 0$.) Then by the chain rule X has density

$$f_X(x) = F_X'(x) = F_Z'((x-\mu)/\sigma) = \frac{1}{\sigma}f_Z((x-\mu)/\sigma).$$

This is equal to

$$\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}.$$

$\mathbb{E}\,X = \mu + \mathbb{E}\,Z$ and $\text{Var}X = \sigma^2\text{Var}Z$, so

$$\mathbb{E}\,X = \mu, \qquad \text{Var}X = \sigma^2.$$

If X is $\mathcal{N}(\mu, \sigma^2)$ 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 is $\mathcal{N}(\mu, \sigma^2)$ and $Z = (X - \mu)/\sigma$, then Z is $\mathcal{N}(0, 1)$.

The distribution function of a standard $\mathcal{N}(0, 1)$ is often denoted $\Phi(x)$, 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

$$\begin{aligned}
\Phi(-x) = \mathbb{P}(Z \le -x) &= \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
&= \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \mathbb{P}(Z \ge x) \\
&= 1 - \mathbb{P}(Z < x) = 1 - \Phi(x).
\end{aligned}$$

Example 8.1: Find $\mathbb{P}(1 \le X \le 4)$ if X is $\mathcal{N}(2, 25)$.

Answer. Write $X = 2 + 5Z$. So

$$\begin{aligned}
\mathbb{P}(1 \le X \le 4) &= \mathbb{P}(1 \le 2 + 5Z \le 4) \\
&= \mathbb{P}(-1 \le 5Z \le 2) = \mathbb{P}(-0.2 \le Z \le .4) \\
&= \mathbb{P}(Z \le .4) - \mathbb{P}(Z \le -0.2) \\
&= \Phi(0.4) - \Phi(-0.2) = .6554 - [1 - \Phi(0.2)] \\
&= .6554 - [1 - .5793].
\end{aligned}$$

Example 8.2: Find c such that $\mathbb{P}(|Z| \ge c) = .05$.

Answer. By symmetry we want c such that $\mathbb{P}(Z \ge c) = .025$ or $\Phi(c) = \mathbb{P}(Z \le c) = .975$. From the table 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 8.1: We have the following bound. For $x > 0$

$$\mathbb{P}(Z \ge x) = 1 - \Phi(x) \le \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.$$

Proof. If $y \geq x$, then $y/x \geq 1$, and then

$$\begin{aligned}\mathbb{P}(Z \geq x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\,dy \\ &\leq \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{y}{x} e^{-y^2/2}\,dy \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.\end{aligned}$$

□

This is a good estimate when $x \geq 3.5$.

In particular, for x large,

$$\mathbb{P}(Z \geq x) = 1 - \Phi(x) \leq e^{-x^2/2}.$$

8.2. Further examples and applications

Example 8.3: Suppose X is normal with mean 6. If $\mathbb{P}(X > 16) = .0228$, then what is the standard deviation of X?

Answer. We apply the fact that $\frac{X-\mu}{\sigma} = Z$ is $\mathcal{N}(0,1)$ and get

$$\begin{aligned}
\mathbb{P}(X > 16) = .0228 \quad &\Longleftrightarrow \quad \mathbb{P}\left(\frac{X-6}{\sigma} > \frac{16-6}{\sigma}\right) = .0228 \\
&\Longleftrightarrow \quad \mathbb{P}\left(Z > \frac{10}{\sigma}\right) = .0228 \\
&\Longleftrightarrow \quad 1 - \mathbb{P}\left(Z \leq \frac{10}{\sigma}\right) = .0228 \\
&\Longleftrightarrow \quad 1 - \Phi\left(\frac{10}{\sigma}\right) = .0228 \\
&\Longleftrightarrow \quad \Phi\left(\frac{10}{\sigma}\right) = .9772.
\end{aligned}$$

Using the standard normal table we see that $\Phi(2) = .9772$, thus we must 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 $\mathbb{P}(X > 0) = .8888$, approximately what is $Var(X)$?

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 $Z \in \mathcal{N}(0,1)$, 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):

$$\begin{aligned}\mathbb{P}(X>5) &= \mathbb{P}\left(Z>\frac{5-10}{6}\right)=\mathbb{P}(Z>-.8333)\\ &= 1-\mathbb{P}(Z\leq -.8333)=1-\Phi(-.8333)\\ &= 1-(1-\Phi(.8333))=\Phi(.8333)=.7977\end{aligned}$$

Solution to Exercise 8.1(B): $2\Phi(1)-1=.6827$

Solution to Exercise 8.1(C): $1-\Phi(.3333)=.3695$

Solution to Exercise 8.2: We have $\mu=200$ and $\sigma=\sqrt{64}=8$. Then

$$\begin{aligned}\mathbb{P}(X>210) &= \mathbb{P}\left(Z>\frac{210-200}{8}\right)=\mathbb{P}(Z>1.25)\\ &= 1-\Phi(1.25)=.1056.\end{aligned}$$

Solution to Exercise 8.3(A): We have $\sigma=\sqrt{225}=15$. Since $\mathbb{P}(X>10)=.5$ then we must have that $\mu=10$ since the pdf of the normal distribution is symmetric. Then

$$\begin{aligned}\mathbb{P}(T>32) &= \mathbb{P}\left(Z>\frac{32-10}{15}\right)\\ &= 1-\Phi(1.47)=.0708.\end{aligned}$$

Solution to Exercise 8.3(B): We have $\mathbb{P}(T<0)=\Phi(-.67)=1-\Phi(.67)=.2514$.

Solution to Exercise 8.4(A): First we need to figure out what μ and σ are. Note that

$$\begin{aligned}\mathbb{P}(X\leq 80,000)=.33 &\iff \mathbb{P}\left(Z<\frac{80,000-\mu}{\sigma}\right)=.33\\ &\iff \Phi\left(\frac{80,000-\mu}{\sigma}\right)=.33\end{aligned}$$

and since $\Phi(.44)=.67$ then $\Phi(-.44)=.33$. Then we must have

$$\frac{80,000-\mu}{\sigma}=-.44.$$

Similarly, since

$$\begin{aligned}\mathbb{P}(X>120,000)=.33 &\iff 1-\mathbb{P}(X\leq 120,000)=.33\\ &\iff 1-\Phi\left(\frac{120,000-\mu}{\sigma}\right)=.33\\ &\iff \Phi\left(\frac{120,000-\mu}{\sigma}\right)=.67\end{aligned}$$

Now again since $\Phi(.44)=.67$ then

$$\frac{120,000-\mu}{\sigma}=.44.$$

Solving the equations

$$\frac{80,000-\mu}{\sigma}=-.44 \text{ and } \frac{120,000-\mu}{\sigma}=.44,$$

simultaneously we have that

$$\mu=100,000 \text{ and } \sigma \approx 45,454.5.$$

Then

$$\mathbb{P}(X>100,000)=.5.$$

Solution to Exercise 8.4(B): We have

$$\mathbb{P}(70,000<X<80,000) \approx .0753.$$

Solution to Exercise 8.5: Since $\mathbb{P}(X>0)=.8888$, then

$$\begin{aligned}
\mathbb{P}(X>0)=.8888 \quad &\Longleftrightarrow \quad \mathbb{P}\left(Z>\frac{0-5}{\sigma}\right)=.8888\\
&\Longleftrightarrow \quad 1-\mathbb{P}\left(Z\leq-\frac{5}{\sigma}\right)=.8888\\
&\Longleftrightarrow \quad 1-\Phi\left(-\frac{5}{\sigma}\right)=.8888\\
&\Longleftrightarrow \quad 1-\left(1-\Phi\left(\frac{5}{\sigma}\right)\right)=.8888\\
&\Longleftrightarrow \quad \Phi\left(\frac{5}{\sigma}\right)=.8888.
\end{aligned}$$

Using the table we see that $\Phi(1.22)=.8888$, 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

$$\begin{aligned}
\mathbb{P}(X>c)=.10 \quad &\Longleftrightarrow \quad \mathbb{P}\left(Z>\frac{c-12}{2}\right)=.10\\
&\Longleftrightarrow \quad 1-\mathbb{P}\left(Z\leq\frac{c-12}{2}\right)=.10\\
&\Longleftrightarrow \quad \mathbb{P}\left(Z\leq\frac{c-12}{2}\right)=.9\\
&\Longleftrightarrow \quad \Phi\left(\frac{c-12}{2}\right)=.9
\end{aligned}$$

Using the table we see that $\Phi(1.28)=.90$, 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

$$\mathbb{P}(X > 12.5) = \mathbb{P}\Big(\frac{X-12}{\sigma} > \frac{0.5}{\sigma}\Big) = 1 - \Phi\Big(\frac{0.5}{\sigma}\Big) = 0.1$$

From the table we get $\frac{0.5}{\sigma} \approx 0.8159$ hence $\sigma \approx 0.4079$.

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, $\Phi\Big(\frac{-\mu}{10}\Big) = 0.8\Phi(-3)$ which implies

$$\Phi\Big(\frac{\mu}{10}\Big) = 0.2 + 0.8\Phi(3) \approx 0.998.$$

From the table we thus get $\mu/10 \approx 2.88$ and hence the standard deviation of C is $\mu/3 \approx 9.6$.

CHAPTER 9

Normal approximation to the binomial

9.1. Introduction

A special case of the central limit theorem is

Theorem 9.1: If S_n is a binomial with parameters n and p, then

$$\mathbb{P}\Big(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\Big) \to \mathbb{P}(a \le Z \le b),$$

as $n \to \infty$, where Z is a $\mathcal{N}(0,1)$.

This approximation is good if $np(1-p) \ge 10$ and gets better the larger this quantity gets. Note np is the same as $\mathbb{E}\, S_n$ and $np(1-p)$ is the same as $\mathrm{Var} S_n$. So the ratio is also equal to $(S_n - \mathbb{E}\, S_n)/\sqrt{\mathrm{Var} S_n}$, 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 the case of the Poisson approximation.

Example 9.1: Suppose a fair coin is tossed 100 times. What is the probability there will be more than 60 heads?

Answer. $np = 50$ and $\sqrt{np(1-p)} = 5$. We have

$$\mathbb{P}(S_n \ge 60) = \mathbb{P}((S_n - 50)/5 \ge 2) \approx \mathbb{P}(Z \ge 2) \approx .0228.$$

Example 9.2: Suppose a die is rolled 180 times. What is the probability a 3 will be showing more than 50 times?

Answer. 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?

Answer. Here S_n is the number of successes, $n = 100$, and $p = .75$. We have

$$\mathbb{P}(S_n \geq 70) = \mathbb{P}((S_n - 75)/\sqrt{300/16} \geq -1.154)$$
$$\approx \mathbb{P}(Z \geq -1.154) \approx .87.$$

(The last figure came from a table.)

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?

Answer. We write $\mathbb{P}(49 \leq S_n \leq 51) = \mathbb{P}(48.5 \leq S_n \leq 51.5)$ and then continue as above.

9.2. 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.

9.3. Selected solutions

Solution to Exercise 9.1(A): The probability of rolling two ves 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 want the number of successes to be 0 or 1, so we want $\mathbb{P}(0 \leq S_{180} \leq 1)$. Since the binomial is integer-valued, we apply the continuity correction and calculate $\mathbb{P}(-.5 \leq S_{180} \leq 1.5)$instead. We calculate that the expected value is $\mu = 180 \cdot p = 5$ and the standard deviation $\sigma = \sqrt{180p(1-p)} \approx 2.205$. Now, as always, we convert this question to a question about the standard normal random variable Z,

$$\begin{aligned} \mathbb{P}(-.5 \leq S_{180} \leq 1.5) &= \mathbb{P}\left(\frac{.5-5}{2.205} \leq Z \leq \frac{1.5-5}{2.205}\right) = \mathbb{P}(-2.49 < Z < -1.59) \\ &= (1 - \Phi(1.59)) - (1 - \Phi(2.49)) \\ &= (1 - .9441) - (1 - .9936) = .0495. \end{aligned}$$

Solution to Exercise 9.1(C): We use $\lambda = np = 5$ (note that we calculated this already in (b)!). Now we see that

$$\mathbb{P}(E) \approx e^{-5}\frac{5^0}{0!} + e^{-5}\frac{5^1}{1!} \approx .0404.$$

Solution to Exercise 9.2: Let X denote the number of left-handed students in the class. Since $X \sim \text{bin}(150, 0.1)$ using Theorem 9.1 we have

(a) $\mathbb{P}(X > 25) = \mathbb{P}\left(\frac{X-15}{\sqrt{13.5}} > \frac{10}{\sqrt{13.5}}\right) \approx 1 - \Phi(2.72) \approx 0.0032.$

(b) $\mathbb{P}(15 \leq X \leq 20) = \mathbb{P}(14.5 < X < 20.5) = \Phi\left(\frac{5.5}{\sqrt{13.5}}\right) - \Phi\left(\frac{-0.5}{\sqrt{13.5}}\right) \approx \Phi(1.5) - 1 + \Phi(0.14) \approx 0.4889.$

Solution to Exercise 9.3: Let X denote the number of black markers. Since $X \sim \text{bin}(50, 0.6)$ we have

$$\mathbb{P}(X > N) \approx 1 - \Phi\Big(\frac{N-30}{2\sqrt{3}}\Big) > 0.2 \quad \text{and} \quad \mathbb{P}(X > N+1) \approx 1 - \Phi\Big(\frac{N-29}{2\sqrt{3}}\Big) < 0.2.$$

From this we deduce that $N \leq 32.909$ and $N \geq 31.944$ so that $N = 32$.

CHAPTER 10

Some continuous distributions

10.1. Introduction

We look at some other continuous random variables besides normals.

Uniform distribution. Here $f(x) = 1/(b-a)$ if $a \le x \le b$ and 0 otherwise. To compute expectations, $\mathbb{E}\, X = \frac{1}{b-a}\int_a^b x\, dx = (a+b)/2$.

Exponential distribution. An exponential with parameter λ has density $f(x) = \lambda e^{-\lambda x}$ if $x \ge 0$ and 0 otherwise. We have

$$\mathbb{P}(X > a) = \int_a^\infty \lambda e^{-\lambda x} dx = e^{-\lambda a}$$

and we readily compute $\mathbb{E}\, X = 1/\lambda$, $\mathrm{Var} X = 1/\lambda^2$. Examples where an exponential r.v. 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 memory-less. This means that $\mathbb{P}(X > s+t \mid X > t) = \mathbb{P}(X > s)$, 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. To prove this,

$$\begin{aligned}\mathbb{P}(X > s+t \mid X > t) &= \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} \\ &= \mathbb{P}(X > s).\end{aligned}$$

Gamma distribution. A gamma distribution with parameters λ and t has density

$$f(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{t-1}}{\Gamma(t)}$$

if $x \ge 0$ and 0 otherwise. Here $\Gamma(t) = \int_0^\infty e^{-y} y^{t-1} dt$ is the Gamma function, which interpolates the factorial function.

An exponential is the time for something to occur. A gamma is the time for t events to occur. A gamma with parameters $\frac{1}{2}$ and $\frac{n}{2}$ is known as a χ_n^2, a chi-squared r.v. with n degrees of freedom. Gammas and chi-squared's come up frequently in statistics.

Another distribution that arises in statistics is the ***beta distribution***:

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \qquad 0 < x < 1,$$

where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}$.

Cauchy distribution. Here

$$f(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}.$$

What is interesting about the Cauchy is that it does not have finite mean, that is, $\mathbb{E}\,|X| = \infty$.

Often it is important to be able to compute the density of $Y = g(X)$. Let us give a couple of examples.

If X is uniform on $(0,1]$ and $Y = -\log X$, then $Y > 0$. If $x > 0$,

$$\begin{aligned} F_Y(x) &= \mathbb{P}(Y \le x) = \mathbb{P}(-\log X \le x) \\ &= \mathbb{P}(\log X \ge -x) = \mathbb{P}(X \ge e^{-x}) = 1 - \mathbb{P}(X \le e^{-x}) \\ &= 1 - F_X(e^{-x}). \end{aligned}$$

Taking the derivative,

$$f_Y(x) = \frac{d}{dx} F_Y(x) = -f_X(e^{-x})(-e^{-x}),$$

using the chain rule. Since $f_X = 1$, this gives $f_Y(x) = e^{-x}$, or Y is exponential with parameter 1.

For another example, suppose X is $\mathcal{N}(0,1)$ and $Y = X^2$. Then

$$\begin{aligned} F_Y(x) &= \mathbb{P}(Y \le x) = \mathbb{P}(X^2 \le x) = \mathbb{P}(-\sqrt{x} \le X \le \sqrt{x}) \\ &= \mathbb{P}(X \le \sqrt{x}) - \mathbb{P}(X \le -\sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x}). \end{aligned}$$

Taking the derivative and using the chain rule,

$$f_Y(x) = \frac{d}{dx} F_Y(x) = f_X(\sqrt{x})\Big(\frac{1}{2\sqrt{x}}\Big) - f_X(-\sqrt{x})\Big(-\frac{1}{2\sqrt{x}}\Big).$$

Remembering that $f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ and doing some algebra, we end up with

$$f_Y(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2},$$

which is a Gamma with parameters $\frac{1}{2}$ and $\frac{1}{2}$. (This is also a χ^2 with one degree of freedom.)

One more example. Suppose X is uniform on $(-\pi/2, \pi/2)$ and $Y = \tan X$. Then

$$F_Y(x) = \mathbb{P}(X \le \tan^{-1} x) = F_X(\tan^{-1} x),$$

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 10.1: Suppose that the length of a phone call in minutes is an exponential r.v with average length 10 minutes.

(1) What's probability of your phone call being more than 10 minutes?
Answer. Here $\lambda = \frac{1}{10}$ thus

$$\mathbb{P}(X > 10) = e^{-\left(\frac{1}{10}\right)10} = e^{-1} \approx .368.$$

(2) Between 10 and 20 minutes?
Answer. We have that

$$\mathbb{P}(10 < X < 20) = F(20) - F(10) = e^{-1} - e^{-2} \approx .233.$$

Example 10.2: 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.

Answer. Note that $\lambda = \frac{1}{4}$.

$$\begin{aligned}
\mathbb{P}\left(X > 5 + 3 \mid X > 3\right) &= \frac{\mathbb{P}\left(X > 8\right)}{\mathbb{P}\left(X > 3\right)} \\
&= \frac{e^{-\frac{1}{4}\cdot 8}}{e^{-\frac{1}{4}\cdot 3}} = e^{-\frac{1}{4}\cdot 5} \\
&= \mathbb{P}\left(X > 5\right).
\end{aligned}$$

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 u-phone functions is exponentially distributed with parameter $\lambda = \frac{1}{8}$. If Pat buys a used u-phone, 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've 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 p.d.f. of Y, $f_Y(y)$.

Exercise 10.5: Suppose that X has an exponential distribution with parameter $\lambda = 1$. Let $c > 0$. 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

$$f_X(t) = \begin{cases} \frac{8}{t^3} & t > 2, \\ 0 & \text{otherwise.} \end{cases}$$

(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 $E|X - a|$ where X is exponential with rate λ.

10.4. Selected solutions

Solution to Exercise 10.1(A): We have

$$\begin{aligned}\mathbb{P}(X>3) &= 1-\mathbb{P}(0<X<3)\\ &= 1-\int_0^3 \frac{1}{2}e^{-\frac{x}{2}}dx\\ &= e^{-\frac{3}{2}} \approx .2231.\end{aligned}$$

Solution to Exercise 10.1(B): There are two ways to do this. The longer one is to calculate $\mathbb{P}(X>5 \mid X>2)$. 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 (a).

Solution to Exercise 10.2: e^{-1}

Solution to Exercise 10.3(A): Recall a formula $\mathbb{P}(X>a)=e^{-\lambda a}$, then

$$\mathbb{P}(X>5)=e^{-\frac{10}{11}}.$$

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 $\mathbb{E}[X]=1$ then we know that λ=1. Then it's pdf and cdf is

$$\begin{aligned}f_X(x) &= e^{-x}, x\geq 0\\ F_X(x) &= 1-e^{-x},\ \ x\geq 0.\end{aligned}$$

By using the given relation,

$$F_y(y)=\mathbb{P}(Y\leq y)=\mathbb{P}\left(e^X\leq y\right)=\mathbb{P}(X\leq \ln y)=F_X(\ln y)$$

and so

$$F_Y(y)=1-e^{-\ln y}=1-\frac{1}{y},\ \text{when}\ \ln(y)\geq 0,$$

taking derivatives we get

$$f_Y(y)=\frac{dF_y(y)}{dy}=\frac{1}{y^2},\ \text{when}\ y\geq 1.$$

Solution to Exercise 10.5: Since X is exponential with parameter 1, then its pdf and cdf are

$$\begin{aligned}f_X(x) &= e^{-x}, x\geq 0\\ F_X(x) &= 1-e^{-x},\ \ x\geq 0.\end{aligned}$$

By using the given relation,

$$F_Y(y)=\mathbb{P}(Y\leq y)=\mathbb{P}\left(\frac{X}{c}\leq y\right)=\mathbb{P}(X\leq cy)=F_X(cy)$$

and so

$$F_Y(y) = 1 - e^{-cy}, \text{ when } cy \geq 0,$$

taking derivatives we get

$$f_Y(y) = \frac{dF_y(y)}{dy} = ce^{-cy}, \text{ when } y \geq 0.$$

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 it's pdf and cdf are

$$\begin{aligned} f_X(x) &= 1 \quad , \quad 0 \leq x < 1 \\ F_X(x) &= x \quad , \quad 0 \leq x < 1. \end{aligned}$$

By using the given relation,

$$F_y(y) = \mathbb{P}\left(Y \leq y\right) = \mathbb{P}\left(e^X \leq y\right) = \mathbb{P}\left(X \leq \ln y\right) = F_X\left(\ln y\right)$$

and so

$$F_Y(y) = \ln y, \text{ when } 0 \leq \ln y < 1.$$

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

$$\mathbb{P}(X < 4) = \int_2^4 \frac{8}{t^3} dt = \Big[-\frac{4}{t^2}\Big]_2^3 = \frac{3}{4} \quad \text{and} \quad \mathbb{P}(2 < X < 3) = \int_2^3 \frac{8}{t^3} dt = \Big[-\frac{4}{t^2}\Big]_2^3 = \frac{5}{9},$$

the probability we look for is $\frac{20}{27} \approx 0.74074074$.

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

$$\mathbb{P}(X > 3) = \int_3^\infty \frac{8}{t^3} dt = \Big[-\frac{4}{t^2}\Big]_3^\infty = \frac{4}{9} \quad \text{and} \quad \mathbb{P}(3 < X < 4) = \int_3^4 \frac{8}{t^3} dt = \Big[-\frac{4}{t^2}\Big]_3^4 = \frac{7}{36},$$

the probability we look for is $\frac{7}{16} = 0.4375$.

Part 3

Multivariate Discrete and Continuous Random Variables

CHAPTER 11

Multivariate distributions

11.1. Introduction

We want to discuss collections of random variables $(X_1, X_2, \ldots, X_n)$, which are known as random vectors. In the discrete case, we can define the density $p(x, y) = \mathbb{P}(X = x, Y = y)$. Remember that here the comma means “and". In the continuous case a density is a function such that

$$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dy\, dx.$$

Example 11.1: If $f_{X,Y}(x, y) = ce^{-x}e^{-2y}$ for $0 < x < \infty$ and $x < y < \infty$, what is c?

Answer. 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 is

$$\int_0^\infty ce^{-x}\tfrac{1}{2}e^{-2x} dx = \frac{c}{6},$$

so $c = 6$.

The multivariate distribution function of (X, Y) is defined by $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$. 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.

We have

$$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) dy\, dx,$$

or

$$\mathbb{P}((X, Y) \in D) = \int \int_D f_{X,Y} dy\, dx$$

when D is the set $\{(x,y) : a \le x \le b, c \le y \le d\}$. One can show this holds when D is any set. For example,

$$\mathbb{P}(X < Y) = \int\int_{\{x<y\}} f_{X,Y}(x,y) dy\, dx.$$

If one has the ***joint density*** $f_{X,Y}(x,y)$ of X and Y, one can recover the ***marginal*** densities of X and of Y:

$$\boxed{f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx.}$$

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.

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.

In the discrete case we say 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. In the continuous case, 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 reals. The left hand side is an abbreviation for

$$\mathbb{P}(\{\omega : X(\omega) \text{ is in } A \text{ and } Y(\omega) \text{ is in } B\})$$

and similarly for the right hand side.

In the discrete case, if we have independence,

$$\begin{aligned} p_{X,Y}(x,y) &= \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &= p_X(x) p_Y(y). \end{aligned}$$

In other words, the joint density $p_{X,Y}$ factors. In the continuous case,

$$\begin{aligned}\int_a^b \int_c^d f_{X,Y}(x,y)dy\,dx &= \mathbb{P}(a \le X \le b, c \le Y \le d)\\ &= \mathbb{P}(a \le X \le b)\mathbb{P}(c \le Y \le d)\\ &= \int_a^b f_X(x)dx \int_c^d f_Y(y)dy\\ &= \int_a^b \int_c^d f_X(x)f_Y(y)\,dy\,dx..\end{aligned}$$

One can conclude from this by taking partial derivatives that

$$f_{X,Y}(x,y) = f_X(x)f_Y(y),$$

or again the joint density factors. Going the other way, one can also see that if the joint density factors, then one has independence.

Example 11.2: 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.

Answer. 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 will be independent. 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 $0 \le \Theta \le \pi/2$ and $0 \le X \le D/2$. The integral is

$$\int_0^{\pi/2} \int_0^{L\cos\theta/2} \frac{4}{\pi D} dx\,d\theta = \frac{2L}{\pi D}.$$

If X and Y are independent, then

$$\begin{aligned}\mathbb{P}(X+Y \le a) &= \int\int_{\{x+y\le a\}} f_{X,Y}(x,y)dx\,dy\\ &= \int\int_{\{x+y\le a\}} f_X(x)f_Y(y)dx\,dy\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)dx\,dy\\ &= \int F_X(a-y)f_Y(y)dy.\end{aligned}$$

Differentiating with respect to a, we have the ***convolution formula*** for the density of $X+Y$:

$$f_{X+Y}(a) = \int f_X(a-y)f_Y(y)dy.$$

There are a number of cases where this is interesting.

(1) If X is a gamma with parameters s and λ and Y is a gamma with parameters t and λ, then straightforward integration shows that $X+Y$ is a gamma with parameters $s+t$ and λ. In particular, the sum of n independent exponentials with parameter λ is a gamma with parameters n and λ.

(2) If Z is a $\mathcal{N}(0,1)$, then $F_{Z^2}(y) = \mathbb{P}(Z^2 \le y) = \mathbb{P}(-\sqrt{y} \le Z \le \sqrt{y}) = F_Z(\sqrt{y}) - F_Z(-\sqrt{y})$. Differentiating shows that $f_{Z^2}(y) = ce^{-y/2}(y/2)^{(1/2)-1}$, or Z^2 is a gamma with parameters $\frac{1}{2}$ and $\frac{1}{2}$. So using (1) above, if Z_i are independent $\mathcal{N}(0,1)$'s, then $\sum_{i=1}^n Z_i^2$ is a gamma with parameters $n/2$ and $\frac{1}{2}$, i.e., a χ_n^2.

(3) 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 $\mathcal{N}(\sum \mu_i, \sum \sigma_i^2)$.

(4) The analogue for discrete random variables is easier. If X and Y takes 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 with parameter λ and Y is a Poisson with parameter μ, we see that $X+Y$ is a Poisson with parameter $\lambda+\mu$. To check this, use the above formula to get

$$\begin{aligned}
\mathbb{P}(X+Y=r) &= \sum_{k=0}^{r} \mathbb{P}(X=k)\mathbb{P}(Y=r-k) \\
&= \sum_{k=0}^{r} e^{-\lambda}\frac{\lambda^k}{k!} e^{-\mu}\frac{\mu^{r-k}}{(r-k)!} \\
&= e^{-(\lambda+\mu)}\frac{1}{r!}\sum_{k=0}^{r} \begin{pmatrix} r \\ k \end{pmatrix} \lambda^k \mu^{r-k} \\
&= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^r}{r!}
\end{aligned}$$

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.

If X and Y are independent normals, then $-Y$ is also a normal (with $\mathbb{E}(-Y) = -\mathbb{E}Y$ and $\mathrm{Var}(-Y) = (-1)^2\mathrm{Var}Y = \mathrm{Var}Y$), and so $X-Y$ is also normal.

To define a ***conditional density*** in the discrete case, we write

$$\boxed{p_{X|Y=y}(x \mid y) = \mathbb{P}(X = x \mid Y = y).}$$

This is equal to

$$\frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p(x,y)}{p_Y(y)}.$$

Analogously, we define in the continuous case

$$\boxed{f_{X|Y=y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}.}$$

Just as in the one-dimensional case, there is a ***change of variables formula***.

Let us recall how the formula goes in one dimension. If X has a density f_X and $Y = g(X)$, then

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) \\ &= \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)). \end{aligned}$$

Taking the derivative, using the chain rule, and recalling that the derivative of $g^{-1}(y)$ is

$$(g^{-1}(y))' = \frac{1}{g'(x)} = \frac{1}{g'(g^{-1}(y))}.$$

Here we use that $y = g(x)$, $x = g^{-1}(y)$, and assume that $g(x)$ is an increasing function.

Thus, we have

$$(*) \qquad \boxed{f_Y(y) = f_X(g^{-1}(y))\frac{1}{|g'(g^{-1}(y))|} = \frac{f_X(x)}{|g'(x)|}}$$

which is defined on the range of the random variable X.

The higher dimensional case is very analogous. Suppose $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$. Let h_1 and h_2 be such that $X_1 = h_1(Y_1, Y_2)$ and $X_2 = h_2(Y_1, Y_2)$. This plays the role of g^{-1}.

Let J be the Jacobian of the mapping $(x_1, x_2) \to (g_1(x_1, x_2), g_2(x_1, x_2))$, so that

$$\boxed{J = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1}.}$$

This is the analogue of $g'(g^{-1}(y)) = g'(x)$. Using the change of variables theorem from multivariable calculus, we have

$$\boxed{f_{Y_1,Y_2}(y_1, y_2) = \frac{f_{X_1,X_2}}{|J|}}$$

which is defined on the range of the random variables (Y_1, Y_2) and is analogous to (*).

Example 11.3: 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

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{\sqrt{2\pi}}e^{-(\frac{3}{7}y_1+\frac{1}{7}y_2)^2/2}\frac{1}{\sqrt{8\pi}}e^{-(\frac{1}{7}y_1-\frac{2}{7}y_2)^2/8}\frac{1}{7}.$$

11.2. Further examples and applications

Example 11.4: 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 p.m.f. of X and Y.

Answer. 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:

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 p.m.f. 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 11.5: 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 joint pdf:
Answer. The region that we integrate over in the first quadrant thus

$$\begin{aligned} 1 &= \int_0^\infty \int_0^\infty ce^{-x}e^{-2y}dxdy = c\int_0^\infty e^{-2y}\left[-e^{-x}\right]_0^\infty dy \\ &= c\int_0^\infty e^{-2y}dy = c\left[-\frac{1}{2}e^{-2y}\right]_0^\infty = c\frac{1}{2}. \end{aligned}$$

Then $c = 2$.

(b) Find $\mathbb{P}(X < Y)$.

Answer. We need to draw the region Let $D = \{(x, y) \mid 0 < x < y, 0 < y < \infty\}$ and set up the the double integral:

$$\begin{aligned}
\mathbb{P}(X < Y) &= \int\int_D f(x, y)dA \\
&= \int_0^\infty \int_0^y 2e^{-x}e^{-2y}dxdy \\
&= \text{work} \\
&\quad \frac{1}{3}.
\end{aligned}$$

(c) Set up the double integral for $\mathbb{P}(X > 1, Y < 1)$
Answer.

$$\mathbb{P}(X > 1, Y < 1) = \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}dxdy = (1 - e^{-1})e^{-2}.$$

(d) Find the marginal $f_X(x)$:
Answer. We have

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^\infty f(x, y)dy = \int_0^\infty 2e^{-x}e^{-2y}dy \\
&= 2e^{-x}\left[\frac{-1}{2}e^{-2y}\right]_0^\infty = 2e^{-x}\left[0 + \frac{1}{2}\right] \\
&= e^{-x}.
\end{aligned}$$

Example 11.6: Let X, Y be r.v. with joint pdf

$$f(x, y) = 6e^{-2x}e^{-3y} \;\; 0 < x < \infty, 0 < y < \infty.$$

Are X, Y independent?

Answer.

Find f_X and f_Y and see if $f = f_X f_Y$. First

$$\begin{aligned}
f_X(x) &= \int_0^\infty 6e^{-2x}e^{-3y}dy = 2e^{-2x}, \\
f_Y(y) &= \int_0^\infty 6e^{-2x}e^{-3y}dx = 3e^{-2y}.
\end{aligned}$$

which are both exponential. Since $f = f_X f_Y$ then yes they are independent!

Example 11.7: Let X, Y have

$$f_{X,Y}(x, y) = x + y, \;\; 0 < x < 1, 0 < y < 1$$

Are X, Y independent?

Answer. Note that there is no way to factor $x + y = f_X(x)f_Y(y)$, hence they can't be independent.

PROPOSITION 11.1. *If* $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ *are* independent *for* $1 \leq i \leq n$ *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 11.8: 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?
Answer.
By the proposition $T + H \sim \mathcal{N}(160, 61)$. Thus

$$\begin{aligned} \mathbb{P}\left(\frac{T+H}{2} \leq 90\right) &= \mathbb{P}(T + H \leq 180) \\ &= \mathbb{P}\left(Z \leq \frac{180 - 160}{\sqrt{61}}\right) = \Phi\left(\frac{180 - 160}{\sqrt{61}}\right) \\ &\approx \Phi(2.56) \approx .9961 \end{aligned}$$

(b) What is the probability that H. will have scored higher than T.?
Answer.
Using $H - T \sim \mathcal{N}(-30, 61)$ we compute

$$\begin{aligned} \mathbb{P}(H > T) &= \mathbb{P}(H - T > 0) \\ &= 1 - \mathbb{P}(H - T < 0) \\ &= 1 - \mathbb{P}\left(Z \leq \frac{0 - (-30)}{\sqrt{61}}\right) \\ &\approx 1 - \Phi(3.84) \approx 0.00006 \end{aligned}$$

–

(c) Answer question (b) if $T \sim \mathcal{N}(90, 64)$ and $H \sim \mathcal{N}(70, 225)$.
Answer.
By the proposition $T - H \sim \mathcal{N}(-20, 289)$ and so

$$\begin{aligned} \mathbb{P}(H > T) &= \mathbb{P}(H - T > 0) \\ &= 1 - \mathbb{P}(H - T < 0) \\ &= 1 - \mathbb{P}\left(Z \leq \frac{0 - (-20)}{17}\right) \\ &\approx 1 - \Phi(1.18) \approx 0.11900 \end{aligned}$$

Example 11.9: Suppose X_1, X_2 have joint distribution

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} x_1 + \frac{3}{2}(x_2)^2 & 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Find the joint pdf of $Y_1 = X_1 + X_2$ and $Y_2 = X_1^2$.

Answer.

Step 1: Find the Jacobian:

$$\begin{aligned} y_1 &= g_1(x_1, x_2) = x_1 + x_2, \\ y_2 &= g_2(x_1, x_2) = x_1^2. \end{aligned}$$

So

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ & \\ 2x_1 & 0 \end{vmatrix} = -2x_1$$

Step 2: Solve for x_1, x_2 and get

$$\begin{aligned} x_1 &= \sqrt{y_2}, \\ x_2 &= y_1 - \sqrt{y_2}. \end{aligned}$$

Step 3: The joint pdf of Y_1, Y_2 is given by the formula:

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(x_1, x_2) \left|J(x_1, x_2)\right|^{-1} \\ &= f_{X_1,X_2}(\sqrt{y_2}, y_1 - \sqrt{y_2}) \frac{1}{2x_1} \\ &= \begin{cases} \dfrac{1}{2\sqrt{y_2}} \left[\sqrt{y_2} + \frac{3}{2}\left(y_1 - \sqrt{y_2}\right)^2\right] & 0 \le y_2 \le 1,\ 0 \le y_1 - \sqrt{y_2} \le 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

11.3. 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) Calculate $\mathbb{P}\left(\frac{1}{2} < X < 1, \frac{2}{3} < Y < \frac{4}{3}\right)$.
(B) Calculate $\mathbb{P}(XY < 2)$.
(C) Calculate 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

$$f(x, y) = e^{-(x+y)}, \;\; 0 \leq x < \infty, 0 \leq y < \infty.$$

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

$$f(x, y) = \begin{cases} 10x^2 y & 0 \leq y \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(A) Find the marginal pdfs, $f_X(x)$ and $f_Y(x)$.
(B) Are X and Y independent random variables?
(C) Find $\mathbb{P}\left(Y \leq \frac{X}{2}\right)$.
(D) Find $\mathbb{P}\left(Y \leq \frac{X}{4} \mid Y \leq \frac{X}{2}\right)$.
(E) Find $\mathbb{E}[X]$.

Exercise 11.7: Consider X and Y given by the joint density

$$f(x, y) = \begin{cases} 4xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(A) Find the joint pdf's, 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

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

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 \ j	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Also find the marginal probability mass functions of X and Y. Finally, find the conditional probability mass function of X given that Y is 5, $P(X = i|Y = 5)$, for $i = 1, \ldots, 6$.

11.4. 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 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 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:

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 the bounds 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: are

$$\begin{aligned} \mathbb{P}(X < Y) &= \int_0^\infty \int_0^y e^{-(x+y)} dx dy = \int_0^\infty \left[-e^{-2y} + e^{-y}\right] dy \\ &= \left[\frac{1}{2} e^{-2y} - e^{-y}\right]_0^\infty = 0 - \left(\frac{1}{2} - 1\right) = \frac{1}{2}. \end{aligned}$$

Solution to Exercise 11.5: We know that

$$f_X(x) = \begin{cases} \frac{1}{4}e^{-\frac{x}{4}} & \text{when } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

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

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{12}e^{-\frac{x}{4}} & \text{when } x \geq 0, 2 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

Draw the region $(2X + Y < 8)$, which correspond to $0 \leq x$, $0 < y < 5$ 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

$$\begin{aligned} \mathbb{P}\left(2X + Y < 8\right) &= \int_2^5 \int_0^{4-\frac{y}{2}} \frac{1}{12} e^{-\frac{x}{4}} dxdy \\ &= \int_2^5 \frac{1}{3}\left(1 - e^{y/8-1}\right) dx \\ &= 1 - \frac{8}{3}\left(e^{-\frac{3}{8}} - e^{-\frac{3}{4}}\right) \end{aligned}$$

Solution to Exercise 11.6(A): We have

$$\begin{aligned} f_X(x) &= \begin{cases} 5x^4 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}, \\ f_Y(y) &= \begin{cases} \frac{10}{3}y(1-y^3) & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

Solution to Exercise 11.6(B): No, since $f_{X,Y} \neq f_X f_Y$.

Solution to Exercise 11.6(C): $\mathbb{P}\left(Y \leq \frac{X}{2}\right) = \frac{1}{4}$.

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

$$\begin{aligned}
\mathbb{P}(W > 5,000,000) &= \mathbb{P}\left(Z > \frac{4,600,000 - 4,400,000}{325,269.1193}\right) \\
&= \mathbb{P}(Z > .6149) \\
&\approx 1 - \Phi(.61) \\
&= .27
\end{aligned}$$

Solution to Exercise 11.10:

Step 1: Find the Jacobian:

$$\begin{aligned}
y_1 &= g_1(x_1, x_2) = 2x_1 + x_2, \\
y_2 &= g_2(x_1, x_2) = x_1 - 3x_2.
\end{aligned}$$

So

$$J(x_1, x_2) = \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} = -7.$$

Step 2: Solve for x_1, x_2 and get

$$\begin{aligned}
x_1 &= \frac{3}{7}y_1 + \frac{1}{7}y_2 \\
x_2 &= \frac{1}{7}y_1 - \frac{2}{7}y_2
\end{aligned}$$

Step 3: The joint pdf of Y_1, Y_2 is given by the formula:

$$\begin{aligned}
f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(x_1, x_2) \left|J(x_1, x_2)\right|^{-1} \\
&= f_{X_1,X_2}\left(\frac{3}{7}y_1 + \frac{1}{y}y_2, \frac{1}{7}y_1 - \frac{2}{7}y_2\right)\frac{1}{7}.
\end{aligned}$$

Since we are given the joint pdf of X_1 and X_2, then plugging it these 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:

Step 1: Find the Jacobian:

$$\begin{aligned}
y_1 &= g_1(x_1, x_2) = x_1 - 2x_2, \\
y_2 &= g_2(x_1, x_2) = 2x_1 + 3x_2.
\end{aligned}$$

So

$$J(x_1, x_2) = \begin{vmatrix} 1 & -2 \\ 2 & 3 \end{vmatrix} = 7.$$

Step 2: Solve for x_1, x_2 and get

$$\begin{aligned} x_1 &= \frac{1}{7}\left(3y_1 + 2y_2\right) \\ x_2 &= \frac{1}{7}\left(-2y_1 + y_2\right) \end{aligned}$$

Step 3: The joint pdf of Y_1, Y_2 is given by the formula:

$$\begin{aligned} f_{Y_1,Y_2}\left(y_1, y_2\right) &= f_{X_1,X_2}\left(x_1, x_2\right)\left|J\left(x_1, x_2\right)\right|^{-1} \\ &= f_{X_1,X_2}\left(\frac{1}{7}\left(3y_1 + 2y_2\right), \frac{1}{7}\left(-2y_1 + y_2\right)\right)\frac{1}{7}. \end{aligned}$$

Since we are given the joint pdf of X_1 and X_2, then plugging it these into f_{X_1,X_2}, we have

$$f_{Y_1,Y_2}\left(y_1, y_2\right) = \begin{cases} \frac{3}{2} \cdot \frac{1}{7^3}\left(\left(3y_1 + 2y_2\right)^2 + \left(-2y_1 + y_2\right)^2\right) & 0 < 3y_1 + 2y_2 < 7, 0 < -2y_1 + y_2 < 7 \\ 0 & \text{otherwise.} \end{cases}$$

CHAPTER 12

Expectations

12.1. Introduction

As in the one variable case, we have

$$\mathbb{E}\, g(X,Y) = \sum\sum g(x,y)p(x,y)$$

in the discrete case and

$$\mathbb{E}\, g(X,Y) = \int\int g(x,y)f(x,y)dx\,dy$$

in the continuous case.

If we set $g(x,y) = x+y$, then

$$\begin{aligned}\mathbb{E}\,(X+Y) &= \int\int (x+y)f(x,y)dx\,dy \\ &= \int\int xf(x,y)dx\,dy + \int\int yf(x,y)dx\,dy.\end{aligned}$$

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 independent, then

$$\mathbb{E}\,[h(X)k(Y)] = \mathbb{E}\,h(X)\mathbb{E}\,k(Y).$$

In particular, $\mathbb{E}\,(XY) = (\mathbb{E}\,X)(\mathbb{E}\,Y)$.

Proof. By the above with $g(x,y) = h(x)k(y)$,

$$\begin{aligned}\mathbb{E}\,[h(X)k(Y)] &= \int\int h(x)k(y)f(x,y)dx\,dy \\ &= \int\int h(x)k(y)f_X(x)f_Y(y)dx\,dy \\ &= \int h(x)f_X(x)\int k(y)f_Y(y)dy\,dx \\ &= \int h(x)f_X(x)(\mathbb{E}\,k(Y))dx \\ &= \mathbb{E}\,h(X)\mathbb{E}\,k(Y).\end{aligned}$$

□

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 $\mathrm{Cov}(X,Y) = 0$.

Note

$$\begin{aligned}
&\mathrm{Var}(X+Y)\\
&\quad= \mathbb{E}\,[((X+Y)-\mathbb{E}\,(X+Y))^2]\\
&\quad= \mathbb{E}\,[((X-\mathbb{E}\,X)+(Y-\mathbb{E}\,Y))^2]\\
&\quad= \mathbb{E}\,[(X-\mathbb{E}\,X)^2 + 2(X-\mathbb{E}\,X)(Y-\mathbb{E}\,Y) + (Y-\mathbb{E}\,Y)^2]\\
&\quad= \mathrm{Var}X + 2\mathrm{Cov}(X,Y) + \mathrm{Var}Y.
\end{aligned}$$

We have the following corollary.

Proposition 12.2: If X and Y are independent, then

$$\mathrm{Var}(X+Y) = \mathrm{Var}X + \mathrm{Var}Y.$$

Proof. We have

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

□

Since a binomial is the sum of n independent Bernoulli's, its variance is $np(1-p)$. If we write $\overline{X} = \sum_{i=1}^n X_i/n$ and the X_i are independent and have the same distribution ($\overline{X}$ is called the sample mean), then $\mathbb{E}\,\overline{X} = \mathbb{E}\,X_1$ and $\mathrm{Var}\overline{X} = \mathrm{Var}X_1/n$.

We define the conditional expectation of X given Y by

$$\mathbb{E}\,[X \mid Y=y] = \int x f_{X|Y=y}(x)\,dx.$$

12.2. Further examples and applications

12.2.1. Expectation and variance.

Example 12.1: Suppose the joint p.m.f of X and Y is given by

$X \diagdown Y$	0	1
0	.2	.7
1	0	.1

.

Find $\mathbb{E}[XY]$.

Answer.

Using the formula we have

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{i,j} x_i y_j p(x_i, y_j) \\ &= 0 \cdot 0p(0,0) + 1 \cdot 0p(1,0) + 0 \cdot 1p(0,1) + 1 \cdot 1p(1,1) \\ &= .1\end{aligned}$$

Example 12.2: Suppose X, Y are independent exponential r.v. with parameter $\lambda = 1$. Set up a double integral that represents

$$\mathbb{E}\left[X^2 Y\right].$$

Answer.

Since X, Y are independent then

$$f_{X,Y}(x,y) = e^{-1x} e^{-1y} = e^{-(x+y)}. \quad 0 < x, y < \infty.$$

Thus

$$\mathbb{E}\left[X^2 Y\right] = \int_0^\infty \int_0^\infty x^2 y e^{-(x+y)} dy dx.$$

Example 12.3: 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 $\operatorname{Var}(Y)$.

Answer.

We first draw the region (try it!) and then set up the integral

$$\begin{aligned}\mathbb{E}XY &= \int_0^1\int_0^y xy\left(10xy^2\right)dxdy = 10\int_0^1\int_0^y x^2y^3dxdy \\ &= \frac{10}{3}\int_0^1 y^3y^3dy = \frac{10}{3}\frac{1}{7} = \frac{10}{21}.\end{aligned}$$

First note that $\operatorname{Var}(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2$. Then

$$\begin{aligned}\mathbb{E}Y^2 &= \int_0^1\int_0^y y^2\left(10xy^2\right)dxdy = 10\int_0^1\int_0^y y^4xdxdy \\ &= 5\int_0^1 y^4y^2dy = \frac{5}{7}.\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}Y &= \int_0^1\int_0^y y\left(10xy^2\right)dxdy = 10\int_0^1\int_0^y y^3xdxdy \\ &= 5\int_0^1 y^3y^2dy = \frac{5}{6}.\end{aligned}$$

So that $\operatorname{Var}(Y) = \frac{5}{7} - \left(\frac{5}{6}\right)^2 = \frac{5}{252}$.

12.2.2. Correlation. We define the ***correlation coefficient*** of X and Y by

$$\boxed{\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}}$$

Note that we always have

$$-1 \leqslant \rho(X,Y) \leqslant 1.$$

Moreover, we say that X and Y are

positively correlated if $\rho(X,Y) > 0$,
negatively correlated if $\rho(X,Y) < 0$,
uncorrelated if $\rho(X,Y) = 0$.

Example 12.4: 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.
Answer.

Recall that $\mathrm{Cov}\,(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$. 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

$$\begin{aligned}
\mathrm{Cov}\,(X,Y) &= \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y \\
&= \frac{1}{6} - \frac{1}{4}\frac{1}{2} \\
&= \frac{1}{24}.
\end{aligned}$$

(b) Compute $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$.

Answer.

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

$$\begin{aligned}
\mathrm{Var}\,(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \\
&= \frac{1}{9} - \left(\frac{1}{4}\right)^2 = \frac{7}{144}
\end{aligned}$$

Also

$$\begin{aligned}
\mathrm{Var}\,(Y) &= \mathbb{E}Y^2 - (\mathbb{E}Y)^2 \\
&= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.
\end{aligned}$$

(c) Calculate $\rho(X,Y)$.

Answer.

$$\rho\,(X,Y) = \frac{\mathrm{Cov}\,(X,Y)}{\sqrt{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)}} = \frac{\frac{1}{24}}{\sqrt{\left(\frac{7}{144}\right)\left(\frac{1}{12}\right)}} \approx .6547.$$

12.2.3. Conditional expectation. Recall that for two random variables X and Y we can consider the conditional distribution of X given $Y = y$ for suitable values of y. It is therefore natural to define the *conditional expectation* of X given $Y = y$. This definition follows the same idea as the "usual" expectation.

If X and Y are discrete with probability mass distribution functions p_X and p_Y respectively, the *conditional expectation of* X given $Y = y$ (for $p_Y(y) \neq 0$) is defined as

$$\mathbb{E}[X \mid Y = y] = \sum_x x p_{X|Y=y}(x|y), \tag{12.2.1}$$

where $p_{X|Y=y}(x|y) = \mathbb{P}(X = x \mid Y = y)$ was defined in the previous chapter as the conditional density of X given $Y = y$.

Analogously, if X and Y are continuous with probability density functions f_X and f_Y, respectively, then the *conditional expectation of* X given $Y = y$ (for $f_Y(y) \neq 0$) is defined as

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y=y}(x|y) dx, \tag{12.2.2}$$

where $f_{X|Y=y}(x|y)$ is the conditional density of X given $Y = y$ as defined in the previous chapter.

All properties of the "usual" expectation that we know will also be satisfied in this case.

Proposition 12.3: 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 12.5: Let X and Y be random variables with joint p.d.f.

$$f_{XY}(x, y) = \begin{cases} \frac{1}{18} e^{-\frac{x+y}{6}} & \text{if } 0 < y < x, \\ 0 & \text{otherwise.} \end{cases}$$

In order to find $\mathrm{Var}(X \mid Y = 2)$, we need to compute the conditional p.d.f. of X given $Y = 2$, i.e.

$$f_{X|Y=2}(x|2) = \frac{f_{XY}(x, 2)}{f_Y(2)}.$$

To this purpose, we compute first the marginal of Y.

$$f_Y(y) = \int_y^{\infty} \frac{1}{18} e^{-\frac{x+y}{6}} dx = \frac{1}{3} e^{-\frac{y}{6}} \Big[- e^{-\frac{y}{6}} \Big]_y^{\infty} = \frac{1}{3} e^{-\frac{y}{3}} \qquad \text{for } y \geq 0.$$

Then, we have

$$f_{X|Y=2}(x|2) = \begin{cases} \frac{1}{6} e^{\frac{2-x}{6}} & \text{if } x > 2, \\ 0 & \text{otherwise.} \end{cases}$$

Now it only remains to compute $\mathbb{E}[X^2 \mid Y = 2]$ and $\mathbb{E}[X \mid Y = 2]$. Applying twice integration by parts we have

$$\mathbb{E}[X^2 \mid Y = 2] = \int_2^\infty \frac{x^2}{6} e^{\frac{2-x}{6}} dx = -\left[x^2 e^{\frac{2-x}{6}}\right]_2^\infty - \left[12xe^{\frac{2-x}{6}}\right]_2^\infty - 12\left[6e^{\frac{2-x}{6}}\right]_2^\infty = 4+24+72 = 100.$$

On the other hand, again applying integration by parts we get

$$\mathbb{E}[X \mid Y = 2] = \int_2^\infty \frac{x}{6} e^{\frac{2-x}{6}} dx = -\left[xe^{-\frac{x-2}{6}}\right]_2^\infty - \left[6e^{-\frac{x-2}{6}}\right]_2^\infty = 2 + 6 = 8.$$

Finally, we obtain $\mathrm{Var}(X \mid Y = 2) = 100 - 8^2 = 36$.

12.3. 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	.3
1	.5	.2

(a) Calculate the covariance of X and Y.
(b) Calculate $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$.
(c) Calculate $\rho(X, Y)$.

Exercise 12.2: Let X and Y be random variable 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) Calculate the covariance of X and Y.
(b) Calculate $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$.
(c) Calculate $\rho(X, Y)$.

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.)

12.4. Selected solutions

Solution to Exercise 12.1(A): First let's calculate the marginal distributions:

$Y \backslash X$	0	1	
0	0	.3	.3
1	.5	.2	.7
	.5	.5	

Then

$$\begin{aligned}\mathbb{E}XY &= (0\cdot 0)\,0+(0\cdot 1)\,.5+(1\cdot 0)\,.3+(1\cdot 1)\,.2=.2\\ \mathbb{E}X &= 0\cdot .5+1\cdot .5=.5\\ \mathbb{E}Y &= 0\cdot .3+1\cdot .7=.7.\end{aligned}$$

Solution to Exercise 12.1(B): First need

$$\begin{aligned}\mathbb{E}X^2 &= 0^2.5+1^2.5=.5\\ \mathbb{E}Y^2 &= 0^2.3+1^2.7=.7\end{aligned}$$

Therefore

$$\begin{aligned}\operatorname{Var}(X) &= \mathbb{E}X^2-(\mathbb{E}X)^2=.5-(.5)^2=.25,\\ \operatorname{Var}(Y) &= \mathbb{E}Y^2-(\mathbb{E}Y)^2=.7-(.7)^2=.21,\end{aligned}$$

Solution to Exercise 12.1(C):

$$\rho(X,Y)=\frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\cdot\operatorname{Var}(Y)}}=\approx -.6547.$$

Solution to Exercise 12.2(A): We'll need $\mathbb{E}[XY], \mathbb{E}X$, 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

$$\operatorname{Var}(X) = \operatorname{Var}(Y) = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

Solution to Exercise 12.2(C):

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)}} = -\frac{1}{11}.$$

Solution to Exercise 12.3: Since X, Y are independent then

$$\begin{aligned} \mathbb{E}\left[(X-1)^2 Y\right] &= \mathbb{E}\left[(X-1)^2\right]\mathbb{E}[Y] \\ &= \operatorname{Var}(X)\mu_Y \\ &= 9/2 = 4.5 \end{aligned}$$

CHAPTER 13

Moment generating functions

13.1. Introduction

We define the moment generating function m_X by

$$m_X(t) = \mathbb{E}\, e^{tX},$$

provided this is finite. In the discrete case this is equal to $\sum e^{tx} p(x)$ and in the continuous case $\int e^{tx} f(x) dx$.

Let us compute the moment generating function for some of the distributions we have been working with.

1. Bernoulli: $pe^t + (1-p)$.

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 the X_i are independent Bernoulli's.

3. Poisson:

$$\mathbb{E}\, e^{tX} = \sum \frac{e^{tk} e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

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 $t \geq \lambda$.

5. $\mathcal{N}(0,1)$:

$$\frac{1}{\sqrt{2\pi}} \int e^{tx} e^{-x^2/2} dx = e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int e^{-(x-t)^2/2} dx = e^{t^2/2}.$$

6. $\mathcal{N}(\mu, \sigma^2)$: Write $X = \mu + \sigma Z$. Then

$$\mathbb{E}\, e^{tX} = \mathbb{E}\, e^{t\mu} e^{t\sigma Z} = e^{t\mu} e^{(t\sigma)^2/2} = e^{t\mu + t^2\sigma^2/2}.$$

Proposition 13.1: If X and Y are independent, then

$$m_{X+Y}(t) = m_X(t) m_Y(t).$$

Proof. By independence and Proposition 12.1,

$$m_{X+Y}(t) = \mathbb{E}\, e^{tX} e^{tY} = \mathbb{E}\, e^{tX} \mathbb{E}\, e^{tY} = m_X(t) m_Y(t).$$

$\square$

Proposition 13.2: If $m_X(t) = m_Y(t) < \infty$ for all t in an interval, then X and Y have the same distribution.

We will not prove this, but this is essentially the uniqueness of the Laplace transform. Note $\mathbb{E}\, e^{tX} = \int e^{tx} f_X(x) dx$. If $f_X(x) = 0$ for $x < 0$, this is $\int_0^\infty e^{tx} f_X(x) dx = \mathcal{L} f_X(-t)$, where $\mathcal{L} f_X$ is the Laplace transform of f_X.

We can use this to verify some of the properties of sums we proved before. For example, if X is a $\mathcal{N}(a, b^2)$ and Y is a $\mathcal{N}(c, d^2)$ and X and Y are independent, then

$$m_{X+Y}(t) = e^{at+b^2t^2/2} e^{ct+d^2t^2/2} = e^{(a+c)t+(b^2+d^2)t^2/2}.$$

Proposition 13.2 then implies that $X + Y$ is a $\mathcal{N}(a + c, b^2 + d^2)$.

Similarly, if X and Y are independent Poisson random variables with parameters a and b, resp., 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 $a + b$.

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.

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 13.1: Suppose that m.g.f of X is given by $m(t) = e^{3(e^t-1)}$. Find $\mathbb{P}(X=0)$.

Answer. We can match this m.g.f to a known m.g.f in our table. Looks like

$$m(t) = e^{3(e^t-1)} = e^{\lambda(e^t-1)} \quad \text{where } \lambda = 3.$$

Thus $X \sim Poisson(3)$. Thus

$$\mathbb{P}(X=0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-3}.$$

□

We remark that $m_X(t)$ is called the moment generating function because we can find all the moments of X by differentiating $m(t)$ and then evaluating at $t=0$. Note that

$$\begin{aligned} m'(t) &= \frac{d}{dt}\mathbb{E}\left[e^{tX}\right] \\ &= \mathbb{E}\left[\frac{d}{dt}e^{tX}\right] \\ &= \mathbb{E}\left[Xe^{tX}\right]. \end{aligned}$$

Now evaluate at $t=0$ and get

$$m'(0) = \mathbb{E}\left[Xe^{0\cdot X}\right] = \mathbb{E}[X].$$

Similarly,

$$\begin{aligned} m''(t) &= \frac{d}{dt}\mathbb{E}\left[Xe^{tX}\right] \\ &= \mathbb{E}\left[X^2e^{tX}\right] \end{aligned}$$

so that

$$m''(0) = \mathbb{E}\left[X^2e^0\right] = \mathbb{E}\left[X^2\right].$$

Continuing to differentiate the m.g.f. we have the following proposition.

Proposition 13.1. *For all $n \geq 0$ we have*

$$\mathbb{E}[X^n] = m^{(n)}(0).$$

Example 13.2: Suppose X is a discrete random variable and has the m.g.f.

$$m_X(t) = \frac{1}{7}e^{2t} + \frac{3}{7}e^{3t} + \frac{2}{7}e^{5t} + \frac{1}{7}e^{8t}.$$

What is the p.m.f of X? Find $\mathbb{E}X$.

Answer. This doesn't match any of the known mg.f.s. Reading off from the mgf we have

$$\frac{1}{7}e^{2t} + \frac{3}{7}e^{3t} + \frac{2}{7}e^{5t} + \frac{1}{7}e^{8t} = \sum_{i=1}^{4} e^{tx_i}p(x_i)$$

then $p(2) = \frac{1}{7}, p(3) = \frac{3}{7}, p(5) = \frac{2}{7}$ and $p(8) = \frac{1}{7}$.

To find $\mathbb{E}[X]$ we use Proposition 13.1 by taking the derivative of the moment generating function:

$$m'(t) = \frac{2}{7}e^{2t} + \frac{9}{7}e^{3t} + \frac{10}{7}e^{5t} + \frac{8}{7}e^{8t},$$

so that

$$\mathbb{E}[X] = m'(0) = \frac{2}{7} + \frac{9}{7} + \frac{10}{7} + \frac{8}{7} = \frac{29}{7}.$$

Example 13.3: Suppose X has m.g.f

$$m_X(t) = (1-2t)^{-\frac{1}{2}} \text{ for } t < \frac{1}{2}.$$

Find the first and second moments of X.

Answer. We have

$$\begin{aligned} m_X'(t) &= -\frac{1}{2}(1-2t)^{-\frac{3}{2}}(-2) = (1-2t)^{-\frac{3}{2}}, \\ m_X''(t) &= -\frac{3}{2}(1-2t)^{-\frac{5}{2}}(-2) = 3(1-2t)^{-\frac{5}{2}}. \end{aligned}$$

So that

$$\begin{aligned} \mathbb{E}X &= m_X'(0) = (1-2\cdot 0)^{-\frac{3}{2}} = 1, \\ \mathbb{E}X^2 &= m_X''(0) = 3(1-2\cdot 0)^{-\frac{5}{2}} = 3. \end{aligned}$$

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

$$m_X(t) = \frac{1}{(1 - 2500t)^4}$$

Calculate 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 random variables with parameters λ_x, λ_y, respectively. Find the distribution of $X + Y$.

Exercise 13.6: True or False? If $X \sim \text{Exp}(\lambda_x)$ and $Y \sim \text{Exp}(\lambda_y)$ then $X + Y \sim \text{Exp}(\lambda_x + \lambda_y)$. Explain your answer.

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

$$\begin{aligned} m_X'(t) &= \frac{1}{(2-t)^2} + \frac{1}{2(1-t)^2} \\ m_X''(t) &= \frac{2}{(2-t)^3} + \frac{1}{(1-t)^3} \end{aligned}$$

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 want $SD(X) = \sqrt{\mathrm{Var}(X)}$. But $\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$. We compute

$$\begin{aligned}
m'(t) &= 4\,(2500)\,(1-2500t)^{-5}\,, \\
m''(t) &= 20\,(2500)^2\,(1-2500t)^{-6}\,, \\
\mathbb{E}X &= m'(0) = 10,000 \\
\mathbb{E}X^2 &= m''(0) = 125,000,000 \\
\mathrm{Var}(X) &= 125,000,000 - 10,000^2 = 25,000,000 \\
SD(X) &= \sqrt{25,000,000} = \mathbf{5,000}.
\end{aligned}$$

Solution to Exercise 13.4: First recall that if we let $W = X + Y$ t, and X, Y independent then

$$m_W(t) = m_{X+Y}(t) = m_X(t)m_Y(t) = \frac{1}{(1-5t)^3},$$

recall that $\mathbb{E}\left[W^2\right] = m_W''(0)$. Thus we need to compute some derivatives

$$\begin{aligned}
m_W'(t) &= \frac{15}{(1-5x)^4}, \\
m_W''(t) &= \frac{300}{(1-5x)^5},
\end{aligned}$$

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

$$m_X(t) = e^{\lambda_x\left(e^t-1\right)} \text{ and } m_Y(t) = e^{\lambda_y\left(e^t-1\right)}.$$

Then

$$\begin{aligned}
m_{X+Y}(t) &= m_X(t)m_Y(t), \quad \text{by independence} \\
&= e^{\lambda_x\left(e^t-1\right)}e^{\lambda_y\left(e^t-1\right)} \\
&= e^{(\lambda_x+\lambda_y)\left(e^t-1\right)}.
\end{aligned}$$

Thus $X + Y \sim \mathrm{Pois}\,(\lambda_x + \lambda_y)$.

Solution to Exercise 13.6: We compute the MGF of $X+Y$ and compare it to the MGF of a random variable $Z \sim \mathrm{Exp}\,(\lambda_x + \lambda_y)$. The MGF of Z is $m_Z(t) = \dfrac{\lambda_x + \lambda_y}{\lambda_x + \lambda_y - t}$ for $t < \lambda_x + \lambda_y$. By independence $m_{X+Y}(t) = m_X(t)m_Y(t) = \dfrac{\lambda_x}{\lambda_x - t} \cdot \dfrac{\lambda_y}{\lambda_y - t}$ but $\dfrac{\lambda_x + \lambda_y}{\lambda_x + \lambda_y - t} \neq \dfrac{\lambda_x}{\lambda_x - t} \cdot \frac{\lambda_y}{\lambda_y - t}$ and hence the statement is false.

CHAPTER 14

Limit laws

14.1. Introduction

Suppose X_i are independent and have the same distribution. In the case of continuous or discrete random variables, this means they all have the same density. We say the X_i are i.i.d., which stands for "independent and identically distributed." Let $S_n = \sum_{i=1}^n X_i$. S_n is called the partial sum process.

Theorem 14.1: Suppose $\mathbb{E}\,|X_i| < \infty$ and let $\mu = \mathbb{E}\,X_i$. Then

$$\frac{S_n}{n} \to \mu.$$

This is known as the strong law of large numbers (SLLN). The convergence here means that $S_n(\omega)/n \to \mu$ for every $\omega \in S$, where S is the probability space, except possibly for a set of ω 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). The WLLN states that for every $a > 0$,

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mathbb{E}\,X_1\Big| > a\Big) \to 0$$

as $n \to \infty$. 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 Chebyshev's inequality.

Proposition 14.2: If $Y \geq 0$, then for any A,

$$\mathbb{P}(Y > A) \leq \frac{\mathbb{E}\,Y}{A}.$$

Proof. We do the case for continuous densities, the case for discrete densities being similar. We have

$$\begin{aligned}\mathbb{P}(Y > A) &= \int_A^\infty f_Y(y)\,dy \leq \int_A^\infty \frac{y}{A} f_Y(y)\,dy \\ &\leq \frac{1}{A}\int_{-\infty}^\infty y f_Y(y)\,dy = \frac{1}{A}\mathbb{E}\,Y.\end{aligned}$$

□

We now prove the WLLN.

Theorem 14.3: Suppose the X_i are i.i.d. and $\mathbb{E}\,|X_1|$ and $\mathrm{Var}X_1$ are finite. Then for every $a > 0$,

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mathbb{E}\,X_1\Big| > a\Big) \to 0$$

as $n \to \infty$.

Proof. Recall $\mathbb{E}\,S_n = n\mathbb{E}\,X_1$ and by the independence, $\mathrm{Var}S_n = n\mathrm{Var}X_1$, so $\mathrm{Var}(S_n/n) = \mathrm{Var}X_1/n$. We have

$$\begin{aligned}\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mathbb{E}\,X_1\Big| > a\Big) &= \mathbb{P}\Big(\Big|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\Big| > a\Big)\\ &= \mathbb{P}\Big(\Big|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\Big|^2 > a^2\Big)\\ &\le \frac{\mathbb{E}\,|\frac{S_n}{n} - \mathbb{E}\,(\frac{S_n}{n})|^2}{a^2}\\ &= \frac{\mathrm{Var}(\frac{S_n}{n})}{a^2}\\ &= \frac{\frac{\mathrm{Var}X_1}{n}}{a^2} \to 0.\end{aligned}$$

The inequality step follows from Proposition 14.2 with $A = a^2$ and $Y = |\frac{S_n}{n} - \mathbb{E}\,(\frac{S_n}{n})|^2$. $\square$

We now turn to the central limit theorem (CLT).

Theorem 14.4: Suppose the X_i are i.i.d. Suppose $\mathbb{E}\,X_i^2 < \infty$. Let $\mu = \mathbb{E}\,X_i$ and $\sigma^2 = \mathrm{Var}X_i$. Then

$$\mathbb{P}\Big(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\Big) \to \mathbb{P}(a \le Z \le b)$$

for every a and b, where Z is a $\mathcal{N}(0,1)$.

The ratio on the left is $(S_n - \mathbb{E}\,S_n)/\sqrt{\mathrm{Var}S_n}$. We do not claim that this ratio converges for any ω (in fact, it doesn't), but that the probabilities converge.

Example 14.1: If the 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.

Example 14.2: 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's the probability it differs from 3.5 by more than 0.05?

Answer. We want

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - 3.5\Big| > .05\Big).$$

We rewrite this as

$$\mathbb{P}(|S_n - n\mathbb{E}\, X_1| > (.05)(3600)) = \mathbb{P}\Big(\Big|\frac{S_n - n\mathbb{E}\, X_1}{\sqrt{n}\sqrt{\mathrm{Var} X_1}}\Big| > \frac{180}{(60)\sqrt{\frac{35}{12}}}\Big)$$

$$r \approx \mathbb{P}(|Z| > 1.756) \approx .08.$$

Example 14.3: Suppose the lifetime of a human has expectation 72 and variance 36. What is the probability that the average of the lifetimes of 100 people exceeds 73?

Answer. We want

$$\begin{aligned}\mathbb{P}\Big(\frac{S_n}{n} > 73\Big) &= \mathbb{P}(S_n > 7300)\\ &= \mathbb{P}\Big(\frac{S_n - n\mathbb{E}\, X_1}{\sqrt{n}\sqrt{\mathrm{Var} X_1}} > \frac{7300 - (100)(72)}{\sqrt{100}\sqrt{36}}\Big)\\ &\approx \mathbb{P}(Z > 1.667) \approx .047.\end{aligned}$$

The idea behind proving the central limit theorem is the following. It turns out that if $m_{Y_n}(t) \to m_Z(t)$ for every t, then $\mathbb{P}(a \le Y_n \le b) \to \mathbb{P}(a \le Z \le b)$. (We won't prove this.) We are going to let $Y_n = (S_n - n\mu)/\sigma\sqrt{n}$. Let $W_i = (X_i - \mu)/\sigma$. Then $\mathbb{E}\, W_i = 0$, $\mathrm{Var} W_i = \frac{\mathrm{Var} X_i}{\sigma^2} = 1$, the W_i are 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 X_i are i.i.d., all the X_i have the same moment generating function. Since $S_n = X_1 + \cdot + X_n$, then

$$m_{S_n}(t) = m_{X_1}(t) \cdots m_{X_n}(t) = [m_{X_1}(t)]^n.$$

If we expand e^{tX_1} as a power series,

$$m_{X_1}(t) = \mathbb{E}\, e^{tX_1} = 1 + t\mathbb{E}\, X_1 + \frac{t^2}{2!}\mathbb{E}\,(X_1)^2 + \frac{t^3}{3!}\mathbb{E}\,(X_1)^3 + \cdots.$$

We put the above together and obtain

$$\begin{aligned}m_{Y_n}(t) &= m_{S_n}(t/\sqrt{n})\\ &= [m_{X_1}(t/\sqrt{n})]^n\\ &= \Big[1 + t\cdot 0 + \frac{(t/\sqrt{n})^2}{2!} + R_n\Big]^n\\ &= \Big[1 + \frac{t^2}{2n} + R_n]^n,\end{aligned}$$

where $|R_n|/n \to 0$ as $n \to \infty$. This converges to $e^{t^2/2} = m_Z(t)$ as $n \to \infty$.

14.2. Further examples and applications

Example 14.4: If 10 fair dice are rolled, find the approximate probability that the sum obtained is between 30 and 40, inclusive.

Answer. We will need to use the $\pm.5$ continuity correction because these are discrete random variables. Let X_i denote the value of the ith die. Recall that

$$\mathbb{E}(X_i) = \frac{7}{2} \quad \text{Var}(X_i) = \frac{35}{12}.$$

Take

$$X = X_1 + \cdots + X_n$$

to be their sum. Using the CLT we need

$$\begin{aligned} n\mu &= 10 \cdot \frac{7}{2} = 35 \\ \sigma\sqrt{n} &= \sqrt{\frac{350}{12}} \end{aligned}$$

thus using the continuity correction, then

$$\begin{aligned} \mathbb{P}(29.5 \leq X \leq 40.5) &= \mathbb{P}\left(\frac{29.5-35}{\sqrt{\frac{350}{12}}} \leq \frac{X-35}{\sqrt{\frac{350}{12}}} \leq \frac{40.5-35}{\sqrt{\frac{350}{12}}}\right) \\ &\approx \mathbb{P}(-1.0184 \leq Z \leq 1.0184) \\ &= \Phi(1.0184) - \Phi(-1.0184) \\ &= 2\Phi(1.0184) - 1 = .692. \end{aligned}$$

Example 14.5: Your instructor has 1000 Probability final exams that needs to be graded. The time required to grade an exam are all i.i.s. 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 minutes in the first 450 minutes of work.

Answer. Let X_i be 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 $\mathbb{P}(X \leq 450)$. Using the CLT, we need

$$\begin{aligned} n\mu &= 25 \cdot 20 = 500 \\ \sigma\sqrt{n} &= 4\sqrt{25} = 20. \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{P}(X \leq 450) &= \mathbb{P}\left(\frac{X-500}{20} \leq \frac{450-500}{20}\right) \\ &\approx \mathbb{P}(Z \leq -2.5) \\ &= 1 - \Phi(2.5) \\ &= .006. \end{aligned}$$

14.3. 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 .9 that the class average would be within 5 of 75?

Exercise 14.3: Let $X_1, X_2, \ldots, X_{100}$ 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.

14.4. Selected solutions

Solution to Exercise 14.1: Let X_i be 1 if the team win's the ith game and 0 if the team loses. This is a Bernoulli R.V. with $p = .5$. Thus $\mu = p = .5$ and $\sigma^2 = p(1-p) = (.5)^2$. Then

$$X = \sum_{i=1}^{162} X_i$$

is the number of games won in the season. Using CLT

$$\begin{aligned} n\mu &= 162 \cdot .5 = 81 \\ \sigma\sqrt{n} &= .5\sqrt{162} = 6.36 \end{aligned}$$

then

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{162} X_i \geq 87\right) &= \mathbb{P}(X \geq 86.5) \\ &= \mathbb{P}\left(\frac{X-81}{6.36} > \frac{86.5-81}{6.36}\right) \\ &\approx \mathbb{P}(Z > .86) = .1949 \end{aligned}$$

where we used a correction since X is a discrete r.v.

Solution to Exercise 14.2: Now $\mu = 75$, $\sigma^2 = 25, \sigma = 5$.

$$\begin{aligned} \mathbb{P}\left(70 < \frac{\sum_{i=1}^n X_i}{n} < 80\right) \geq .9 \quad &\iff \quad \mathbb{P}\left(70 \cdot n < \sum_{i=1}^n X_i < 80 \cdot n\right) \geq .9 \\ &\iff \quad \mathbb{P}\left(\frac{70 \cdot n - 75 \cdot n}{5\sqrt{n}} < Z < \frac{80 \cdot n - 75 \cdot n}{5\sqrt{n}}\right) \geq .9 \\ &\iff \quad \mathbb{P}\left(-5\frac{\sqrt{n}}{5} < Z < 5\frac{\sqrt{n}}{5}\right) \geq .9 \\ &\iff \quad \mathbb{P}\left(-\sqrt{n} < Z < \sqrt{n}\right) \geq .9 \\ &\iff \quad \Phi\left(\sqrt{n}\right) - \Phi\left(-\sqrt{n}\right) \geq .9 \\ &\iff \quad \Phi\left(\sqrt{n}\right) - \left(1 - \Phi\left(\sqrt{n}\right)\right) \geq .9 \\ &\iff \quad 2\Phi\left(\sqrt{n}\right) - 1 \geq .9 \\ &\iff \quad \Phi\left(\sqrt{n}\right) \geq .95 \end{aligned}$$

Using the table inversely we have that

$$\sqrt{n} \geq 1.65 \implies n \geq 2.722$$

hence the first integer that insurers that $n \geq 2.722$ is

$$n = 3.$$

Solution to Exercise 14.3: Since $\lambda = 1$ then $\mathbb{E}X_i = 1$ and $\mathrm{Var}(X_i) = 1$. Use CLT

$$\begin{aligned} n\mu &= 100 \cdot 1 = 100 \\ \sigma\sqrt{n} &= 1 \cdot \sqrt{100} = 10. \end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{100} X_i > 90\right) &= \mathbb{P}\left(\frac{\sum_{i=1}^{100} X_i - 100\cdot 1}{1\cdot\sqrt{100}} > \frac{90-100\cdot 1}{1\cdot\sqrt{100}}\right) \\
&\approx \mathbb{P}(Z > -1) = .8413.
\end{aligned}$$

Solution to Exercise 14.4:

$$\begin{aligned}
\mathbb{P}(X \geq 1300) &= \mathbb{P}\left(\frac{X-2400000}{80000} \geq \frac{2500000-2400000}{80000}\right) \\
&\approx \mathbb{P}(Z \geq 1.25) \\
&= 1-\Phi(1.25) = 1-.8944 \\
&= .1056.
\end{aligned}$$

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 $\mathbb{P}(X \leq 210)$. Use CLT,

$$\begin{aligned}
n\mu &= 36\cdot 5 = 180 \\
\sigma\sqrt{n} &= 2\sqrt{36} = 12.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{P}(X \leq 210) &= \mathbb{P}\left(\frac{X-180}{12} \leq \frac{210-180}{12}\right) \\
&\approx \mathbb{P}(Z \leq 2.5) \\
&= \Phi(2.5) \\
&= .9938.
\end{aligned}$$

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 $\mathbb{P}(X \geq 1800)$. Use CLT,

$$\begin{aligned}
n\mu &= 82\cdot 15 = 1230 \\
\sigma\sqrt{n} &= 5\sqrt{82} = 45.28.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{P}(X \geq 1300) &= \mathbb{P}\left(\frac{X-1230}{45.28} \geq \frac{1300-1230}{45.28}\right) \\
&\approx \mathbb{P}(Z \geq 1.55) \\
&= 1-\Phi(1.55) = 1-.9394 \\
&= .0606.
\end{aligned}$$

Part 4

Applications of Probability Distributions and Random Variables

CHAPTER 15

Application of Probability in Finance

15.1. Introduction

15.1.1. The simple coin toss game. Suppose, as in Example 3.5, 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 (Chapter 14)

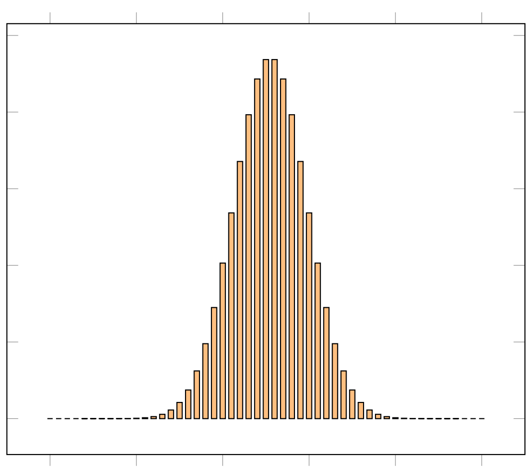
$$S_n = \sum_{i=1}^{n} X_i$$

where $X_1, X_2, X_3, \ldots$ are independent random variables with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. 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

$$M_n = M + S_n \quad \sim \quad M + \sqrt{n}Z \quad \sim \quad \mathcal{N}(M, n)$$

and these random variables have the distribution function computed as $F(x) = \Phi\left(\dfrac{x - M}{\sqrt{n}}\right)$.



15.1.2. 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

$$\mathbb{P}(M + S_n = x) - \mathbb{P}(M + S_n = -x) \qquad \text{if } x > 0$$

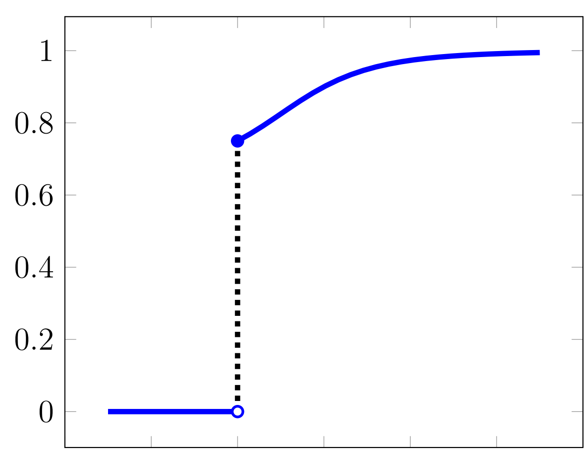
To obtain this formula, we again denote by M_n the amount of money we have after n coin tosses. Then

$$\begin{aligned}
\mathbb{P}(M_n = x) &= \mathbb{P}(M + S_n = x, M + S_k > 0 \text{ for all } k = 1, 2, ..., n) \\
&= \mathbb{P}(M + S_n = x) - \mathbb{P}(M + S_n = x, M + S_k = 0 \text{ for some } k = 1, 2, ..., n) \\
&= \mathbb{P}(M + S_n = x) - \mathbb{P}(M + S_n = -x, M + S_k = 0 \text{ for some } k = 1, 2, ..., n) \\
&= \mathbb{P}(M + S_n = x) - \mathbb{P}(M + S_n = -x).
\end{aligned}$$

This, together with the Central Limit Theorem, implies that the cumulative probability distribution function of M_n can be approximated by

$$F(x) = \begin{cases} \Phi\left(\dfrac{x - M}{\sqrt{n}}\right) + \Phi\left(\dfrac{-x - M}{\sqrt{n}}\right) & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

The following graph shows the approximate shape of this function.



Note that this function is discontinuous as the jump at zero represents the probability thatwe have lost all the money by the time n:

$$\mathbb{P}(M_n = 0) \approx 2\Phi\left(-\frac{M}{\sqrt{n}}\right)$$

If we consider the limit $n \to \infty$, then $\mathbb{P}(M_n = 0) \to 2\Phi(0) = 1$.

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 3.6.

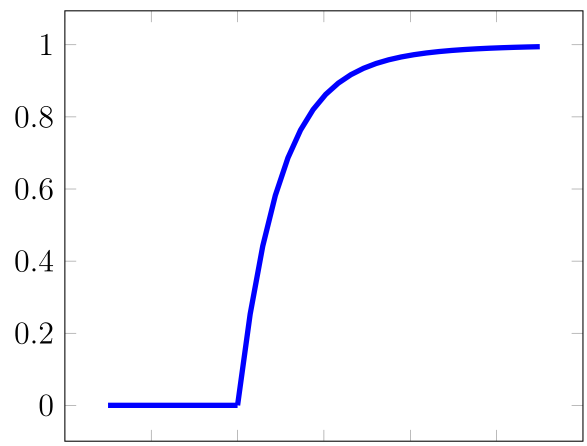
15.1.3. The coin toss game with borrowing at zero. Suppose now that the game is modified so that each time when we hit zero, instead of stopping, we borrow \$1 and continue playing. Another form of the "Reflection Principle" implies that the probability to have x dollars is

$$\mathbb{P}(M_n = x) = \mathbb{P}(M + S_n = x) + \mathbb{P}(M + S_n = -x) \qquad \text{if } x > 0$$

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

$$F(x) = \begin{cases} \Phi\left(\dfrac{x-M}{\sqrt{n}}\right) - \Phi\left(\dfrac{-x-M}{\sqrt{n}}\right) & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

The following graph shows the approximate shape of this function.



15.1.4. Probability to win $\$N$ before reaching as low as $\$L$. Continuing the simple coin toss game, suppose L, M and N are integer numbers 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 computing the function

$$y(x) = \mathbb{P}(\text{win } \$N \text{ before reaching } \$L \mid \text{given } M = x)$$

which satisfies $N - L + 1$ linear equations

$$y(x) = \begin{cases} 0 & \text{if } x = L \\ \dots & \\ \frac{1}{2}(y(x+1) + y(x-1)) & \text{if } L < x < N \\ \dots & \\ 1 & \text{if } x = N \end{cases}$$

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

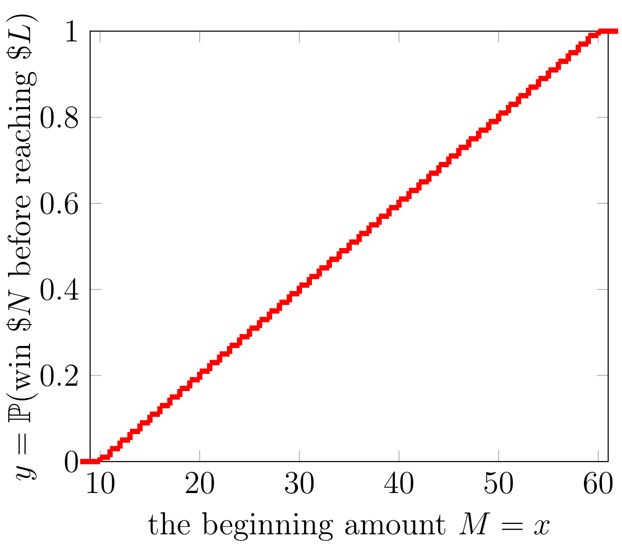
$$y(x) = \frac{x - L}{N - L}$$

and the final answer: with probability

$$\mathbb{P}(\text{win } \$N \text{ before reaching } \$L \mid \text{given } M = x) = y(M) = \frac{\boldsymbol{M - L}}{\boldsymbol{N - L}} \tag{15.1.1}$$

we win $\$N$ before going as low as $\$L$ if we begin with $\$M$. Formula (15.1.1) applies in general to the Gambler's Ruin problems, a particular case of which we consider in this subsection.

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.



15.1.5. Expected playing time. Suppose we play the same simple coin toss game as in the previous subsection, 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

$$\mathbb{E}\,T(x) = \begin{cases} 0 & \text{if } x = L \\ \dots & \\ 1 + \frac{1}{2}\big(\mathbb{E}\,T(x+1) + \mathbb{E}\,T(x-1)\big) & \text{if } L < x < N \\ \dots & \\ 0 & \text{if } x = N \end{cases}$$

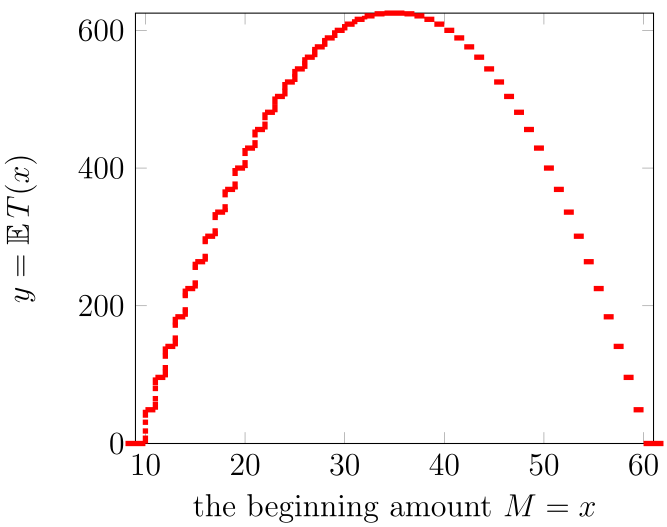
These equations have a unique solution given by the formula

$$\mathbb{E}\,T(x) = (x-L)(N-x)$$

and the final answer: the expected number of coin tosses is

$$\mathbb{E}\,\boldsymbol{T(M) = (M-L)(N-M)} \tag{15.1.2}$$

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.



15.1.6. 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 independent random variables with where $\mathbb{P}(Y_i = 2) = \mathbb{P}(Y_i = \frac{1}{2}) = \frac{1}{2}$. If again write $y(x) = \mathbb{P}(\text{win } \$N \text{ before reaching } \$L)$ 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 will be linear if we change to the logarithmic variable $\log(x)$, which gives us the answer:

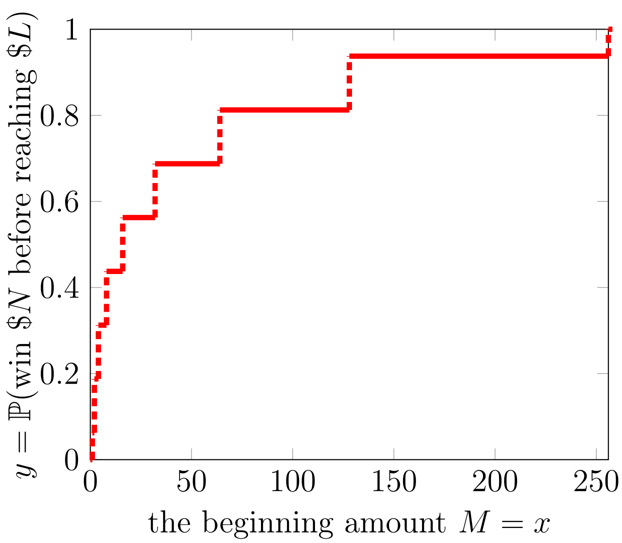
$$\mathbb{P}(\text{win } \$N \text{ before reaching } \$L \mid \text{given } M = x) \approx \frac{\log(M/L)}{\log(N/L)}$$

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

$$\mathbb{P}(\text{win } \$N \text{ before reaching } \$L \mid \text{given } M = x) = \frac{\boldsymbol{\ell}}{\boldsymbol{\ell} + \boldsymbol{w}} \tag{15.1.3}$$

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 $\$M$. Formula (15.1.3) is again the general formula for the Gambler's Ruin problems, the same as in formula (15.1.1).

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.



15.2. Exercises on simple coin toss games

Exercise 15.1: In Subsection 15.1.1, what is the approximate distribution of $M_n - M_k$?

Exercise 15.2: In Subsection 15.1.1, compute $\mathrm{Cov}(M_k, M_n)$.
Hint: assume $n > k$ and write $M_n = M_k + M_n - M_k = M_k + (S_n - S_k)$.

Exercise 15.3: Consider the game in which $M_n = Me^{\sigma S_n}$. Describe the rules of this game.

Exercise 15.4: In the game in Exercise 15.3, find $\mathbb{E}\, M_n$, $\mathbb{E}\, M_n^2$, $\mathrm{Var}(M_n)$.

Exercise 15.5: In the game in Exercise 15.3, how M_n and M_k are related?
Hint: assume $n > k$ and write $M_{n+1} = M_n \dfrac{M_{n+1}}{M_n}$. Also consider $M_n = M_k \dfrac{M_n}{M_k}$.

Exercise 15.6: Following Exercise 15.4, find $\mathrm{Cov}(M_n, M_k)$.

Exercise 15.7: In the game in Exercise 15.3, find the probability to win $\$N$ before reaching as low as $\$L$.

Exercise 15.8: In the game in Exercise 15.7, find the expected playing time.

Exercise 15.9: Following Exercise 15.3, use the Normal Approximation (the Central Limit Theorem) to find the approximate distribution of M_n. Then use this distribution to find the approximate values of $\mathbb{E}\, M_n$, $\mathbb{E}\, M_n^2$, $\mathrm{Var}(M_n)$.

Exercise 15.10: Following Exercise 15.6, use the Normal Approximation (the Central Limit Theorem) to find the approximate value of $\mathrm{Cov}(M_n, M_k)$.

Exercise 15.11: Comparing Exercises 15.4 and 15.9, which quantities are larger and which are smaller? In which case the Normal Approximation gets better, 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?

15.3. Problems motivated by the American options pricing

Problem 15.1: 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 15.2: 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 15.3:

(1) Compute the maximal expected value of the payoff and the corresponding optimal strategy in the same game as in Problem 15.2, 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 15.2, when an unlimited number of throws are allowed.

Problem 15.4: 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

$$\mathbb{E}\left[\textit{cumulative winning}\right] = \infty.$$

This is called the St. Petersburg paradox. The paradox is in an observation that one wouldn't pay an infinite amount to play such a game.

Notice that if the game is stopped at the nth round, you spent in the previous rounds the dollar amount

$$2^0 + \cdots + 2^{n-2} = \left(2^0 + \cdots + 2^{n-2}\right)\frac{1-\frac{1}{2}}{1-\frac{1}{2}} = 2^{n-1} - 1.$$

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 15.5: In the context of Problem 15.4, 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 compute g.

Problem 15.6: 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 15.7: Let us suppose that a one-year interest rate is determined at the beginning of each month. In this case $r_0, r_1, \ldots, r_{N-1}$ 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, $(1+r_0)\ldots(1+r_{k-1})$ at the end of year k, and so forth. Let us suppose that $r_0 = .1$ and $(r_i)_{i=1,\ldots,N-1}$ are independent random variables with the following Bernoulli distribution (under so-called risk-neutral measure): $r_i = .15$ 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.,

$$D_N = \frac{1}{(1+r_0)\ldots(1+r_{N-1})}.$$

You need to evaluate

$$\tilde{\mathbb{E}}\left[D_N\right].$$

Problem 15.8: In the settings of Problem 15.7, let *simply compounded yield* for the zero-coupon bond with maturity N is the number y, such that

$$\tilde{\mathbb{E}}\left[D_N\right] = \frac{1}{(1+y)^m}.$$

Calculate y.

Problem 15.9: In the settings of Problem 15.7, let *continuously compounded yield* for the zero-coupon bond with maturity N is the number y, such that

$$\tilde{\mathbb{E}}\left[D_N\right] = e^{-yN}.$$

Calculate y.

15.4. Problems about 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

$$S(T) = S(0)\exp\left((r - \tfrac{1}{2}\sigma^2)T + \sigma\sqrt{T}X\right), \tag{15.4.1}$$

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 15.10: Show that

$$\mathbb{E}\left[e^{-rT}(S(T) - K)^+\right] = S(0)\Phi(d_1) - Ke^{-rT}\Phi(d_2), \tag{15.4.2}$$

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 (15.4.2) 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 15.11: In the framework of the Black-Scholes-Merton model, i.e., with the stock price process given by (15.4.1) with $r = 0$, let us consider

$$\mathbb{E}\left[S(t)^{1/3}\right]. \tag{15.4.3}$$

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 15.12: In the framework of the Black-Scholes-Merton model, i.e., with the stock price process given by (15.4.1), let us consider

$$\max_{t\in[0,1]} \mathbb{E}\left[e^{-rt}\left(S(t) - K\right)^+\right]. \tag{15.4.4}$$

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 15.11, $\max\limits_{t\in[0,1]} \mathbb{E}\left[e^{-rt}\left(S(t) - K\right)^+\right]$ is closely related to the payoff of the *American call option* with maturity 1 and $\hat{t}$ to the optimal policy.

15.5. Selected solutions

Answer to Exercise 15.4: The Normal Approximations imply

$$\mathbb{E}\, M_n = M\left(\frac{e^{\sigma}+e^{-\sigma}}{2}\right)^n,$$

$$\mathbb{E}\, M_n^2 = M^2\left(\frac{e^{2\sigma}+e^{-2\sigma}}{2}\right)^n,$$

$$\mathrm{Var}(M_n) = M^2\left(\left(\frac{e^{2\sigma}+e^{-2\sigma}}{2}\right)^n - \left(\frac{e^{\sigma}+e^{-\sigma}}{2}\right)^{2n}\right).$$

Answer to Exercise 15.9:

$$\mathbb{E}\, M_n \approx M e^{n\sigma^2/2},$$

$$\mathbb{E}\, M_n^2 \approx M^2 e^{2n\sigma^2},$$

$$\mathrm{Var}(M_n) \approx M^2\left(e^{2n\sigma^2} - e^{n\sigma^2}\right).$$

Answer to Exercise 15.11:

$$\mathbb{E}\, M_n = M\left(\frac{e^{\sigma}+e^{-\sigma}}{2}\right)^n < M e^{n\sigma^2/2},$$

$$\mathbb{E}\, M_n^2 = M^2\left(\frac{e^{2\sigma}+e^{-2\sigma}}{2}\right)^n < M^2 e^{2n\sigma^2},$$

$$\mathrm{Var}(M_n) = M^2\left(\left(\frac{e^{2\sigma}+e^{-2\sigma}}{2}\right)^n - \left(\frac{e^{\sigma}+e^{-\sigma}}{2}\right)^{2n}\right) < M^2\left(e^{2n\sigma^2} - e^{n\sigma^2}\right).$$

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 15.1: 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 15.2: 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 15.1. The expected payoff of the game which allows up to three throws is

$$EP_3 = \frac{7+4}{2}\cdot\frac{6-4}{6}+\frac{17}{4}\cdot\frac{4}{6} = \frac{14}{3} \approx 4.6666$$

Answer to Problem 15.3:

(1) $EP_4 = \dfrac{89}{18} \approx 4.9444 \qquad EP_5 = \dfrac{277}{54} \approx 5.1296 \qquad EP_6 = \dfrac{1709}{324} \approx 5.2747$

(2) $EP_\infty = 6$.

Sketch of the solution to Problem 15.4: 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 15.5:

$$\mathbb{E}\left[\log_2(G)\right] = \sum_{k=1}^{\infty} \log_2(2^k)\frac{1}{2^k} = \sum_{k=1}^{\infty} \frac{k}{2^k} = 2 = \log_2(4) < \infty.$$

In particular, $g = 4$.

Sketch of the solution to Problem 15.6: $\mathbb{E}[X] = \frac{1}{C}\sum\limits_{n=1}^{\infty} \frac{1}{n} = \infty$, as the harmonic series is divergent.

Sketch of the solution to Problem 15.7: Using independence of r_k's, 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 15.8: Direct computations give $y = \left(\frac{1}{\tilde{\mathbb{E}}[D_N]}\right)^{\frac{1}{N}} - 1$.

Sketch of the solution to Problem 15.9: Similarly to Problem 15.8, we get $y = -\frac{\log\left(\tilde{\mathbb{E}}[D_N]\right)}{N}$.

Sketch of the solution to Problem 15.10: From formula (15.4.1), 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 15.11: Let us fix $t \in [0,1]$. From Jensen's inequality, 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 reached 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 $\max\limits_{\tau\in[0,1]} \mathbb{E}\left[S(\tau)^{1/3}\right]$, where τ are so-called *stopping times*, i.e., random times which an additional structural property.

Sketch of the solution to Problem 15.12: for every $t \in [0,1]$, we have

$$e^{-rt}\left(S(t)-K\right)^{+} = \left(S(t)e^{-rt} - Ke^{-rt}\right)^{+} \leq \left(S(t)e^{-rt} - Ke^{-r}\right)^{+}$$

and $\mathbb{E}\left[S(t)e^{-rt}\right] = S(0)$. Now, using convexity of $x \to (x-K)^{+}$ and applying Jensen's inequality 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

$$\hat{t} = 1 \quad \text{and} \quad \mathbb{E}\left[e^{-r\hat{t}}\left(S(\hat{t})-K\right)^{+}\right] = \mathbb{E}\left[e^{-r}\left(S(1)-K\right)^{+}\right].$$

Appendix: a table of common distributions and a table of the standard normal cumulative distribution function

The following tables contain information from `https://www.wikipedia.org/`

Table of probability distributions

Discrete random variables

Name	Abbrev.	Parameters	p.m.f.: $\mathbb{P}[X=k],\ k \in \mathbb{N}_0$	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	m.g.f.: $\mathbb{E}[e^{tX}],\ t \in \mathbb{R}$
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	p.d.f.: $f(x),\ x \in \mathbb{R}$	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	m.g.f.: $\mathbb{E}[e^{tX}],\ t \in \mathbb{R}$
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 $t < \lambda$

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55966	0.56360	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	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	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	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	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	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	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	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	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	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	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	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	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	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	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	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	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	0.99990	0.99990	0.99990	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	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998