INTRODUCTION TO PROBABILITY THEORY AND STOCHASTIC PROCESSES FOR UNDERGRADUATE STUDENTS

ANDREY SARANTSEV

Contents

Introduction

These are lecture notes on Probability Theory and Stochastic Processes. These include both discrete- and continuous-time processes, as well as elements of Statistics. These lecture notes are intended for junior- and senior-level undergraduate courses. They contain enough material for two semesters or three quarters. These lecture notes are suitable for mathematics, applied mathematics, economics and statistics majors, but are likely too hard for business and humanities majors. Prerequisites include matrix algebra and calculus: a standard two-semester or three-quarter sequence, including multivariable calculus. We consider applications to insurance, finance and social sciences, as well as some elementary statistics. We tried to keep notes as condensed and concise as possible, very close to the actual notes written on the board during lectures.

Acknowledgements

I am thankful to Tomoyuki Ichiba and Raya Feldman, who mentored me when I started teaching at the Department of Statistics and Applied Probability, University of California, Santa Barbara. I am grateful to Judith Arms and Andrew Loveless, my teaching mentors at the Department of Mathematics, University of Washington. In addition, I would like to thank Janko Gravner from Department of Mathematics, University of California, Davis, who put his lecture notes on Probability and Markov Chains on his web page. These lecture notes use his materials a lot. Last but not least, I am thankful to Soumik Pal, my Ph.D. adviser, Ioannis Karatzas, his Ph.D. adviser, Jean-Pierre Fouque, my postdoctoral mentor, and all my graduate and undergraduate probability teachers, for helping understand and appreciate this exciting area of mathematics.

Date: April 5, 2018. Department of Statistics and Applied Probability. University of California, Santa Barbara. Email: `sarantsev@pstat.ucsb.edu` .

1. Combinatorics

1.1. **Permutations.** A *permutation* of 123 is, say, 321 or 231: numbers cannot repeat. There are $3 \cdot 2 \cdot 1 = 3! = 6$ permutations: there are 3 chioces for the first slot, 2 choices for the second slot (because one of the numbers is already in the first slot and cannot be repeated), and only one choice for the last, third slot.

In general, for $1, 2, 3, \ldots, n-1, n$ there are

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

permutations. This number is called n *factorial*. Examples:

$$1! = 1,\quad 2! = 2,\quad 3! = 6,\quad 4! = 24,\quad 5! = 120,\quad 6! = 720.$$

There is a convention that $0! = 1$. Indeed, we have the property $(n-1)!n = n!$ for $n = 2, 3, 4, \ldots$. It follows from the definition of the factorial, and this is the main property. We would like it to be true also for $n = 1$: $0! \cdot 1 = 1!$, so $0! = 1$. Factorial grows very quickly. Indeed, 100! is extremely large; no modern computer can go through permutations of $1, 2, \ldots, 100$. When a computer programming problem encounters search among permutations of n numbers, then this problem is deemed unsolvable. *Stirling's formula*:

$$\boxed{n! \backsim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \quad \text{as} \quad n \to \infty}$$

where $f(n) \backsim g(n)$ means $\lim_{n\to\infty} f(n)/g(n) = 1$.

If we have three slots for numbers $1, 2, 3, 4, 5, 6, 7$, and repetitions are not allowed, this is called an *arrangement*. Say, 364, 137, 634. There are $7 \cdot 6 \cdot 5 = 210$ such arrangements: 7 choices for the first slot, 6 for the second and 5 for the third. We can write this as

$$A_7^3 = 7 \cdot 6 \cdot 5 = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1} = \frac{7!}{(7-3)!}.$$

In general, if there are k slots for $1, 2, \ldots, n$, then the number of arrangements is

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

A permutation can be viewed as a particular case of an arrangement, when $k = n$: the number of slots is the same as the total number of elements.

1.2. **Subsets.** How many subsets of three elements are there in the set $\{1, 2, \ldots, 7\}$? The difference between an arrangement and a subset is that for a subset, order does not matter. (But in both of them, there are no repetitions.) For example, $\{3, 4, 6\}$ and $\{6, 3, 4\}$ is the same subset, but 346 and 634 are different arrangements. From any subset, we can create $3! = 6$ arrangements. So the quantity of subsets is equal to the quantity of arrangements divided by 6:

$$\frac{A_7^3}{3!} = \frac{7!}{4!3!} = \frac{210}{6} = 35.$$

In general, the quantity of subsets of k elements in $\{1, \ldots, n\}$ is equal to

$$\boxed{\binom{n}{k} = \frac{A_n^k}{k!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{(n-k)!k!}}$$

It is pronounced as "n choose k".

(i) $\binom{1}{0} = 1$, because there is only one subset of zero elements in $\{1\}$, and this is an empty set $\varnothing$.

(ii) $\binom{1}{1} = 1$, because there is only one subset of one element in $\{1\}$: the set $\{1\}$ itself.

(iii) $\binom{n}{0} = 1$, for the same reason as in (i);

(iv) $\binom{n}{n} = 1$, for the same reason as in (ii);

(v) $\binom{2}{1} = 2$, because there are two subsets of one element of $\{1, 2\}$: these are $\{1\}$ and $\{2\}$;

(vi) $\binom{n}{1} = n$, because there are n subsets of one element of $\{1, 2, \ldots, n\}$: $\{1\},\ \{2\}, \ldots, \{n\}$;

(vii) $\binom{n}{n-1} = n$, because to choose a subset of $n-1$ elements out of $\{1, 2, \ldots, n\}$, we need to throw away one element, and it can be chosen in n ways;

(viii) $\binom{4}{2} = 4!/(2!2!) = 24/4 = 6$, and these subsets of $\{1, 2, 3, 4\}$ are

$$\{1, 2\},\ \{1, 3\},\ \{1, 4\},\ \{2, 3\},\ \{2, 4\},\ \{3, 4\}.$$

1.3. **Symmetry.** We can say without calculations that $\binom{8}{2} = \binom{8}{6}$. Indeed, for every subset of $\{1, 2, \ldots, 8\}$ of two elements there is a subset of six elements: its complement. For example, $\{3, 5\}$ corresponds to $\{1, 2, 4, 6, 7, 8\}$. This is a one-to-one correspondence. So there are equally many subsets of two elements and subsets of six elements. Similarly, $\binom{8}{3} = \binom{8}{5}$. More generally,

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

1.4. **Power set.** How many subsets does the set $\{1, 2, \ldots, n\}$ contain? Answer: 2^n. Indeed, to construct an arbitrary subset E, you should answer n questions:

- Is $1 \in E$? Yes/No
- Is $2 \in E$? Yes/No
- ...
- Is $n \in E$? Yes/No

For each question, there are two possible answers. The total number of choices is $\underbrace{2 \cdot 2 \cdot \ldots \cdot 2}_{\text{n times}} = 2^n$. The set of all subsets of $\{1, \ldots, n\}$ is called a *power set*, and it contains 2^n elements. But we can also write the quantity of all subsets as the sum of binomial coefficients:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n}.$$

So we get the following identity:

$$\boxed{\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n} = 2^n}$$

Example 1.1. Let $n = 2$. Yes-Yes: $E = \{1, 2\}$, Yes-No: $E = \{2\}$, No-Yes: $E = \{1\}$, No-No: $E = \varnothing$. Total number of subsets: $2^2 = 4$. Two of them have one element: $\binom{2}{1} = 2$, one has two elements, $\binom{2}{2} = 1$, and one has zero elements, $\binom{2}{0} = 1$. Total: $1 + 2 + 1 = 4$.

1.5. **Reduction property.** We can claim that

$$\binom{5}{2} = \binom{4}{2} + \binom{4}{1}.$$

Indeed, the total number of subsets $E \subseteq \{1, 2, 3, 4, 5\}$ which contain two elements is $\binom{5}{2}$. But there are two possibilities:

Case 1. $5 \in E$. Then $E \setminus \{5\}$ is a one-element subset of $\{1, 2, 3, 4\}$; there are $\binom{4}{1}$ such subsets.

Case 2. $5 \notin E$. Then E is a two-element subset of $\{1, 2, 3, 4\}$. There are $\binom{4}{2}$ such subsets.

So $\binom{4}{1} + \binom{4}{2} = \binom{5}{2}$. In general,

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

1.6. **Pascal's triangle.**

$$\begin{array}{lccccccccc}
n = 0: & & & & & 1 & & & & \\
n = 1: & & & & 1 & & 1 & & & \\
n = 2: & & & 1 & & 2 & & 1 & & \\
n = 3: & & 1 & & 3 & & 3 & & 1 & \\
n = 4: & 1 & & 4 & & 6 & & 4 & & 1
\end{array}$$

$$\begin{array}{lccccccccc}
n = 0: & & & & & \binom{0}{0} & & & & \\
n = 1: & & & & \binom{1}{0} & & \binom{1}{1} & & & \\
n = 2: & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\
n = 3: & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\
n = 4: & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4}
\end{array}$$

Each element is the sum of two elements immediately above it: this is the reduction formula. We start from the edges, fill them with ones: $\binom{n}{0} = \binom{n}{n} = 1$, see the previous lecture. Then we fill the inside from top to bottom using this rule, which is the reduction formula.

1.7. **Newton's binomial formula.** We can expand $(x+y)^2 = x^2+2xy+y^2$, and $(x+y)^3 = x^3+3x^2y+3xy^2+y^3$. The coefficients are taken from corresponding lines in Pascal's triangle. Why is this? Let us show this for $n = 3$.

$$(x+y)^3 = (x+y)(x+y)(x+y) = xxx + xxy + xyx + yxx + xyy + yxy + yyx + yyy.$$

Each term has slots occupied by y: $xxy \leftrightarrow \{3\}$, $yxy \leftrightarrow \{1, 3\}$. If there is one slot occupied by y, this corresponds to x^2y, and there are $\binom{3}{1}$ such combinations. So we have: $\binom{3}{1}x^2y$. Other terms give us:

$$\binom{3}{0}x^3 + \binom{3}{1}x^2y + \binom{3}{2}xy^2 + \binom{3}{3}y^3.$$

The general formula looks like this:

$$\boxed{(x+y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \ldots + \binom{n}{n}y^n}$$

Let $x = y = 1$. Then we get:

$$2^n = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n}.$$

This formula was already proven above. Let $x = 1, y = -1$. Then

$$0 = \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \ldots,$$

$$\binom{n}{0} + \binom{n}{2} + \ldots = \binom{n}{1} + \binom{n}{3} + \ldots$$

The quantity of subsets with even number of elements is equal to the quantity of subsets with odd number of elements.

1.8. **Combinatorics problems.** Here, we study a few combinatorics (counting) problems, which can be reduced to counting permutations and combinations.

Example 1.2. Five women and four men take an exam. We rank them from top to bottom, according to their performance. There are no ties.

(a) How many possible rankings?
(b) What if we rank men and women separately?
(c) As in (b), but Julie has the third place in women's rankings.

(a) A ranking is just another name for permutation of nine people. The answer is $\boxed{9!}$

(b) There are 5! permutations for women and 4! permutations for men. The total number is $\boxed{5!4!}$ We should multiply them, rather than add, because men's and women's rankings are independent: we are interested in pairs: the first item is a ranking for women, the second item is a ranking for men. If we needed to choose: either rank women or rank men, then the solution would be $5! + 4!$

(c) We exclude Julie from consideration, because her place is already reserved. There are four women remaining, so the number of permutations is 4! For men, it is also 4! The answer is $\boxed{4!^2}$

Example 1.3. A licence plate consists of seven symbols: digits or letters. How many licence plates are there if the following is true:

(a) there must be three letters and four digits, and symbols may repeat?
(b) there must be three letters and four digits, and symbols may not repeat?
(c) no restrictions on the quantity of letters and numbers, and symbols may repeat?
(d) no restrictions on the quantity of letters and numbers, and symbols may not repeat?

(a) Choose three slots among seven for letters; this may be done in $\binom{7}{3}$ ways. Fill each of these three slots with letters; there are 26^3 ways to do this, since letters can repeat. Fill each of the remaining four slots with digits; there are 10^4 ways to do this, since numbers can repeat. Answer: $\boxed{\binom{7}{3} \cdot 26^3 \cdot 10^4}$

(b) This case is different from (i) because there are $26 \cdot 25 \cdot 24 = 26!/23!$ ways to fill three chosen slots for letters, and $10 \cdot 9 \cdot 8 \cdot 7 = 10!/6!$ ways to fill four chosen slots for numbers. Answer: $\boxed{\binom{7}{3} 26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 \cdot 7}$

(c) This case is easier than the previous ones, since there are 36 symbols, and each of the seven slots can be filled with any of these symbols. Answer: $\boxed{36^7}$

(d) Similarly to (c), $\boxed{36 \cdot 35 \cdot 34 \cdot 33 \cdot 32 \cdot 31 \cdot 30 = 36!/29!}$

Example 1.4. We have five women and four men. We need to choose a committee of three women and two men. How many ways are there to do this if:

(a) there are no additional restrictions?

(b) Mike and Thomas refuse to serve together?

(c) Britney and Lindsey refuse to serve together?

(d) Andrew and Anna refuse to serve together?

(a) There are $\binom{4}{2} = 6$ ways to choose two men out of four, and there are $\binom{5}{3} = 10$ ways to choose two women out of five. So the answer is $\boxed{60}$

(b) How many committees are there for which this restriction is violated, so Mike and Thomas do serve together? If they are already chosen, then we do not need to choose any other man, and there are $\binom{5}{3} = 10$ ways to choose three women out of five. So the quantity of committees where Mike and Thomas do serve together is 10. The answer is $60 - 10 = \boxed{50}$

(c) Similarly to (b), the number of committees where Britney and Lindsay serve together is $\binom{3}{1}\binom{4}{2} = 18$, because you can choose one more woman out of the remaining three in $\binom{3}{1} = 3$ ways, and the number of choices for men is $\binom{4}{2}$. So the answer is $60 - 18 = \boxed{42}$

(d) Similarly to (c), the number of committees where Andrew and Anna serve together is $\binom{3}{1}\binom{4}{2} = 18$, because you can choose one more man out of the remaining three in $\binom{3}{1} = 3$ ways, and two more women out of the remaining four in $\binom{4}{2}$ ways. So the answer is $60 - 18 = \boxed{42}$

Problems

Problem 1.1. There are 8 apartments for 6 people. Each person chooses one apartment, and each apartment can host no more than one person. How many choices are there?

For the next three problems, consider the following setup. There are 10 Swedes, 7 Finns and 6 Danes; we should choose a committee that consists of 9 people, three from each nation. Find the number of choices if:

Problem 1.2. There are no additional constraints.

Problem 1.3. Two of 10 Swedes refuse to serve together.

Problem 1.4. There is a Swede and a Dane who refuse to serve together?

For the next three problems, find the number of arrangements of letters in the words:

Problem 1.5. Seattle.

Problem 1.6. Alaska.

Problem 1.7. Spokane.

Problem 1.8. We need to choose a committee of six people: three French and three Germans, out of six French and seven Germans. How many ways are there to do this? Your answer should be in the form of a number, say 10 or 23.

Problem 1.9. Old license plates consist of six symbols: three digits and three numbers (at any place, not necessarily digits first). Symbols cannot repeat. How many such plates are there?

For the next three problems, consider the following setup. Anna, Brendan, Christine, Daniel and Emily form a band with 5 instruments: a piano, a flute, a trombone, a violin and a guitar. Find the number of arrangements if:

Problem 1.10. Each of them can play all the instruments, and there is a rule that no two people can play the same instrument.

Problem 1.11. Each of them can play all the instruments, and there is *not* a rule that no two people can play the same instrument: for example, all of them can choose piano, so there will be five pianos in the band.

Problem 1.12. Each of them can play all the instruments, and there is a rule that no two people can play the same instrument, but they need Anna or Brendan for the trombone.

For the next three problems,consider following setup. Ten married couples (husband-wife) come to the dancing lesson. They are split into ten dancing couples, each of which contains a male and a female. Partners are assigned randomly. Find the number of possible outcomes in the following cases:

Problem 1.13. There are no additional restrictions.

Problem 1.14. Mr Smith does not want to dance with his wife.

Problem 1.15. One of the women is tired, so now there are 9 dancing pairs (and one idle husband).

Problem 1.16. You need to select a governing committee board for an insurance company. This board should contain 4 actuaries, 4 experts in finance, and one CEO. There are 20 candidates: 12 actuaries and 8 experts in finance, each of whom is qualified to be a CEO. How many ways are there to select this board?

Problem 1.17. There are 10 Swedes and 7 Finns. We should choose a committee that consists of 9 people. The committee has a chairman and a vice-chairman, which should be from different nations. Moreover, apart from these two people, both nations should be represented in the remaining 15 people. How many possibilities are there?

For the next three problems, consider the following setup. Suppose you have 7 Danes and 8 Swedes. You need to choose a committee of 3 Danes and 3 Swedes. What is the number of choices in the following cases:

Problem 1.18. No additional constraints.

Problem 1.19. The committee must contain a chairperson.

Problem 1.20. The chairperson must be a Swede.

Problem 1.21. Suppose ten men and twelve women are dancing. How many ways are there to find ten pairs (man-woman), if the ordering of the pairs does not matter?

For the next two problems, consider the following setup. There are 12 weight lifters: 4 from the US, 4 from Canada, 3 from Russia and 1 from UK. We would like to rank them. How many possible rankings are there in the following cases:

Problem 1.22. We give their names in the ranking.

Problem 1.23. We give only their country in the ranking.

For the next three problems, calculate the following numbers:

Problem 1.24. $\binom{10}{4}$.

Problem 1.25. $\binom{12}{9}$.

Problem 1.26. $\binom{8}{2}$.

For the next three problems, find the sum:

Problem 1.27. $\sum_{k=0}^{2000}\binom{2000}{k}4^k$.

Problem 1.28. $\sum_{k=1}^{100}\binom{100}{k}2^k$.

Problem 1.29. $\sum_{k=2}^{2016}\binom{2016}{k}2^k3^{2015-k}$.

For the next three problems, using the binomial theorem, expan the brackets:

Problem 1.30. $(x+2)^5$.

Problem 1.31. $(2x-3)^4$.

Problem 1.32. $(1-x)^3$.

For the next four problems, find the sum:

Problem 1.33. $\binom{100}{0}+\binom{100}{2}+\binom{100}{4}+\ldots+\binom{100}{100}$.

Problem 1.34. $\binom{2017}{0}+\binom{2017}{2}+\ldots+\binom{2017}{2016}$.

Problem 1.35. $\binom{50}{1} + \binom{50}{2} + \ldots + \binom{50}{48}$.

Problem 1.36. $\binom{75}{2} + \binom{75}{4} + \ldots + \binom{75}{37}$.

For the next three problems, use the Stirling formula to approximate:

Problem 1.37. $\binom{100}{30}$.

Problem 1.38. $\binom{100}{50}$.

Problem 1.39. $\binom{2017}{1008}$.

2. Basic Probability

2.1. **Set-theoretic notation.** A *set* is a collection of its *elements*: $A = \{1, 3, 4, 5\}$, $B = \{red, blue\}$. We say: $1 \in A$ (1 belongs to A), but $2 \notin A$ (2 does not belong to A). For two sets A and B, we can define their *union*: $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$, *intersection*: $A \cap B := \{x \mid x \in A \text{ and } x \in B\}$, and *difference*: $A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$. The *empty set*, which does not contain any elements, is denoted by $\varnothing$.

Example 2.1. Let $A = \{0, 1, 2, 3\}$, $B = \{0 \le x \le 7 \mid \ x \text{ is odd}\} = \{1, 3, 5, 7\}$. Then

$$A \cup B = \{0, 1, 2, 3, 5, 7\},\ A \cap B = \{1, 3\},\ A \setminus B = \{0, 2\},\ B \setminus A = \{5, 7\}.$$

2.2. **Axioms of probability.** The following set of axioms was formulated by a Russian mathematician Andrey Kolmogorov. We have the set Ω of *elementary outcomes* $\omega \in \Omega$, and subsets $A \subseteq \Omega$ are called *events*. Each event A has a number $\mathbf{P}(A)$, which is called the *probability* of A and satisfies the following axioms:

(i) $\boxed{0 \le \mathbf{P}(A) \le 1}$ $\boxed{\mathbf{P}(\varnothing) = 0}$ $\boxed{\mathbf{P}(\Omega) = 1}$

(ii) If two events A and B are *disjoint*: $A \cap B = \varnothing$, then $\boxed{\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)}$ The same is true for three or more disjoint events, or even for infinitely many.

Because of the axiom (ii), if we taken the *complement* of the event A: $A^c = \Omega \setminus A$, then $\mathbf{P}(A) + \mathbf{P}(A^c) = \mathbf{P}(A \cap A^c) = \mathbf{P}(\Omega) = 1$, and therefore $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.

Example 2.2. Toss a coin twice. Then $\Omega = \{HH, TT, HT, TH\}$, and

$$\mathbf{P}\{HH\} = \mathbf{P}\{TT\} = \mathbf{P}\{TH\} = \mathbf{P}\{HT\} = \frac{1}{4}.$$

For example, the event

$$A = \{\text{the same result}\} = \{\text{HH, TT}\}$$

has probability $\mathbf{P}(A) = 1/2$.

Example 2.3. (Chevalier de Mere's Problem.) What is the probability that
(a) at least one six in four rolls of a fair die;
(b) at least one double-six in 24 rolls of two fair dice?

(a) Denote this event by A. The probability space is

$$\Omega = \{(i_1, i_2, i_3, i_4) \mid i_1, i_2, i_3, i_4 \in \{1, 2, 3, 4, 5, 6\}\}.$$

The probability of each elementary outcome is $1/6^4$. The complement of the event A is $A^c = \Omega \setminus A$, which means that there are no six in any of the four rolls. We have:

$$A^c = \{(i_1, i_2, i_3, i_4) \mid i_1, i_2, i_3, i_4 \in \{1, 2, 3, 4, 5\}\}.$$

The event A^c contains 5^4 outcomes. Each has probability $1/6^4$, so $\mathbf{P}(A^c) = 5^4/6^4$. But

$$\mathbf{P}(A) = 1 - \mathbf{P}(A^c) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.5177.$$

This event has probability greater than 50%, so it is good to bet on it.

(b) The probability space is

$$\Omega = \{(i_1, j_1, \ldots, i_{24}, j_{24}) \mid i_1, j_1, \ldots, i_{24}, j_{24} \in \{1, 2, 3, 4, 5, 6\}\}.$$

Each elementary outcome has probability $1/36^{24}$. The event A has complement

$$A^c = \{(i_1, j_1, \ldots, i_{24}, j_{24}) \mid (i_k, j_k) \neq (6, 6),\ k = 1, \ldots, 24\}.$$

Each toss of two dice has 35 results which are not double six. We toss them 24 times, so total number of elementary outcomes is 35^{24}. Therefore,

$$\mathbf{P}(A^c) = \frac{35^{24}}{36^{24}}, \quad \mathbf{P}(A) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.4914,$$

which is slightly less than 50%, so it is not a good bet.

Example 2.4. Suppose we have $n = 365$ days, each day is equally likely to be a birthday. There are k people. Then

$$\begin{aligned}
&\mathbf{P}(\text{there are two people with same birthdays}) \\
&\quad = 1 - \mathbf{P}(\text{there are no people with same birthdays}) \\
&\quad = 1 - \frac{n(n-1)\dots(n-k+1)}{n^k},
\end{aligned}$$

because the number of birthday arrangements (assuming they are all different) is $A_n^k = n(n-1)\dots(n-k+1)$. This probability for $k = 23$ equals 0.5073, greater than 50%!

2.3. **Inclusion-exclusion formula for two events.** Consider two events, A and B, which can intersect. Then $\mathbf{P}(A \cup B) \ne \mathbf{P}(A) + \mathbf{P}(B)$. Indeed, toss two coins, let A be the event that the first toss is H, let B be the event that the second toss is H. Then $\mathbf{P}(A \cup B) = 3/4$, but $\mathbf{P}(A) = \mathbf{P}(B) = 1/2$. We have:

$$\boxed{\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)}$$

Indeed, let $C_1 = A \setminus B$, $C_2 = A \cap B$, $C_3 = B \setminus A$. Then

$$\begin{aligned}
\mathbf{P}(A \cup B) = \mathbf{P}(C_1 \cup C_2 \cup C_3) &= \mathbf{P}(C_1) + \mathbf{P}(C_2) + \mathbf{P}(C_3), \\
\mathbf{P}(A) = \mathbf{P}(C_1 \cup C_2) &= \mathbf{P}(C_1) + \mathbf{P}(C_2), \\
\mathbf{P}(B) = \mathbf{P}(C_2 \cup C_3) &= \mathbf{P}(C_2) + \mathbf{P}(C_3), \\
\mathbf{P}(A \cap B) &= \mathbf{P}(C_2),
\end{aligned}$$

and the formula follows from here.

Example 2.5. Choose a random number from 1 to 1000. What is the probability that it is divisible either by 2 or by 3? Let

$$A = \{\text{divisible by } 2\}, \quad B = \{\text{divisible by } 3\}.$$

There are 500 numbers in A and 333 numbers in B, because $1000/3 = 333 + 1/3$ has integer part 333. More exactly,

$$A = \{2, 4, 6, \dots, 1000\}, \quad B = \{3, 6, 9, \dots, 996, 999\}.$$

$$\mathbf{P}(A) = \frac{500}{1000} = \frac{1}{2}, \quad \mathbf{P}(B) = \frac{333}{1000}.$$

In addition, $A \cap B = \{\text{divisible by } 6\}$ contains 166 numbers: $1000/6$ has integer part 166. Therefore,

$$\mathbf{P}(A \cap B) = 166/1000, \quad \mathbf{P}(A \cup B) = \frac{500}{1000} + \frac{333}{1000} - \frac{166}{1000} = \boxed{\frac{667}{1000}}$$

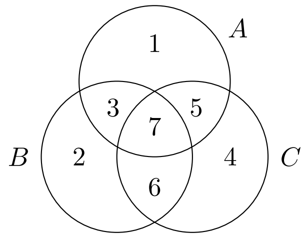
2.4. **Inclusion-exclusion formula for three events.** For any events A, B, C, we have:

$$\boxed{\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(A \cap B) - \mathbf{P}(A \cap C) - \mathbf{P}(B \cap C) + \mathbf{P}(A \cap B \cap C)}$$

Indeed, let p_1 be the probability of the event 1 on the diagram, which is $A \setminus (B \cup C)$. Let p_2 be the probability of the event 2, etc. Then

$$\begin{aligned}
\mathbf{P}(A) &= p_1 + p_3 + p_5 + p_7, \\
\mathbf{P}(B) &= p_2 + p_3 + p_6 + p_7, \\
\mathbf{P}(C) &= p_4 + p_6 + p_7 + p_5, \\
\mathbf{P}(A \cap B) &= p_3 + p_7, \\
\mathbf{P}(A \cap C) &= p_5 + p_7, \\
\mathbf{P}(B \cap C) &= p_6 + p_7, \\
\mathbf{P}(A \cap B \cap C) &= p_7.
\end{aligned}$$

Finally, $\mathbf{P}(A \cup B \cup C) = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7$. Plugging this into the formula in the box above, we can check it is indeed true.



Example 2.6. There are 20 people: 10 Norwegians, 7 Swedes and 3 Finns. We randomly choose a committee of 5 people. Assuming that the choice is uniform (that is, we choose each committee with the same probability), what is the probability that at least one nation is not represented in the committee? Here,

$$\Omega = \{\text{all subsets of 5 elements of the set of 20 people}\}.$$

It contains $\binom{20}{5}$ elementary outcomes. Since the choice is uniform, each elementary outcome (=committee) has probability $1/\binom{20}{5}$. Consider the following events:

$$F = \{\text{Finns are not represented}\},\ S = \{\text{Swedes are not represented}\},$$

$$N = \{\text{Norwegians are not represented}\}.$$

Then we need to find the probability $\mathbf{P}(F \cup N \cup S)$. What is the probability of the event F? If Finns are not represented, then we choose from $10 + 7 = 17$ people. There are $\binom{17}{5}$ elementary outcomes. Each has the same probability $1/\binom{20}{5}$. So

$$\mathbf{P}(F) = \frac{\text{number of outcomes in } F}{\text{number of all outcomes in } \Omega} = \frac{\binom{17}{5}}{\binom{20}{5}}.$$

This rule (dividing number of favorable outcomes by number of all outcomes) is valid only when the probability is uniform. If, for example, we favor Finns in the committee, then there is some other probability. Similarly,

$$\mathbf{P}(S) = \frac{\binom{13}{5}}{\binom{20}{5}},\quad \mathbf{P}(N) = \frac{\binom{10}{5}}{\binom{20}{5}}.$$

$$\mathbf{P}(S \cap F) = \frac{\binom{10}{5}}{\binom{20}{5}},$$

because $S \cap F$ means that we choose only from Norwegians, and there are 10 of them. Similarly,

$$\mathbf{P}(F \cap N) = \frac{\binom{7}{5}}{\binom{20}{5}},$$

and $\mathbf{P}(S \cap N) = 0$, because it is impossible to choose a committee of five people from three Finns. Finally, $\mathbf{P}(S \cap N \cap F) = 0$. Apply the inclusion-exclusion formula:

$$\mathbf{P}(F \cup S \cup N) = \frac{\binom{13}{5}}{\binom{20}{5}} + \frac{\binom{17}{5}}{\binom{20}{5}} + \frac{\binom{10}{5}}{\binom{20}{5}} - 0 - \frac{\binom{10}{5}}{\binom{20}{5}} - \frac{\binom{7}{5}}{\binom{20}{5}} + 0.$$

2.5. **Conditional probability.** A bag with 11 cubes:

- 3 red and fuzzy
- 2 red and smooth
- 4 blue and fuzzy
- 2 blue and smooth

Put your hand in the bag and randomly pick a cube. Let

$$R = \{\text{the cube is red}\},\ F = \{\text{the cube is fuzzy}\}.$$

Then the probability that it is red is $\mathbf{P}(R) = 5/11$. But if you feel that it is fuzzy, then the probability that it is red is $3/7$. This is called *conditional probability* of R given F:

$$\mathbf{P}(R \mid F) = \frac{\mathbf{P}(R \cap F)}{\mathbf{P}(F)} = \frac{3}{7}.$$

In general, conditional probability of A given B is defined as

$$\boxed{\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}}$$

2.6. **Bayes' formula.** Flip a fair coin. If H, roll one die. If T, roll two dice. What is the probability p that there is at least one six? Let X_1, X_2 be the first and the second dice. This is equal to

$$p = \frac{1}{2}\cdot \mathbf{P}(X_1 = 6) + \frac{1}{2}\cdot \mathbf{P}(X_1 = 6 \text{ or } X_2 = 6).$$

But $\mathbf{P}(X_1 = 6) = 1/6$, and by inclusion-exclusion formula we have:

$$\mathbf{P}(X_1 = 6\text{or } X_2 = 6) = \mathbf{P}(X_1 = 6) + \mathbf{P}(X_2 = 6) - \mathbf{P}(X_1 = 6,\ X_2 = 6) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}.$$

Thus,

$$p = \frac{1}{2}\cdot\frac{1}{6} + \frac{1}{2}\cdot\frac{11}{36} = \boxed{\frac{17}{72}}$$

In general, suppose we have events $F_1, \ldots, F_n$ such that

$$F_1 \cup F_2 \cup \ldots \cup F_n = \Omega, \quad F_i \cap F_j = \varnothing,\ i \ne j.$$

This means one and only one of the events $F_1, \ldots, F_n$ happens. Take an event A; then

$$\boxed{\mathbf{P}(A) = \sum_{k=1}^{n} \mathbf{P}(A \cap F_k) = \sum_{k=1}^{n} \mathbf{P}(A \mid F_k)\mathbf{P}(F_k)}$$

In the example above,

$$F_1 = \{\text{heads}\},\ F_2 = \{\text{tails}\},\ A = \{\text{at least one six}\}.$$

Therefore, we have:

$$\boxed{\mathbf{P}(F_1 \mid A) = \frac{\mathbf{P}(F_1 \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A \mid F_1)\mathbf{P}(F_1)}{\mathbf{P}(A \mid F_1)\mathbf{P}(F_1) + \ldots + \mathbf{P}(A \mid F_n)\mathbf{P}(F_n)}}$$

Example 2.7. 10% of people have a disease. A test gives the correct result with probability 80%. That is, if a person is sick, the test is positive with probability 80%, and if a person is healthy, the test is positive only with probability 20%. A random person is selected from the population and is tested positive. What is the probability that he is sick? Let

$$F_1 = \{\text{sick}\},\ F_2 = \{\text{healthy}\},\ A = \{\text{tested positive}\}.$$

$$\mathbf{P}(F_1) = 0.1,\ \mathbf{P}(F_2) = 0.9,\ \mathbf{P}(A \mid F_1) = 0.8,\ \mathbf{P}(A \mid F_2) = 0.2.$$

$$\mathbf{P}(F_1 \mid A) = \frac{0.1\cdot 0.8}{0.1\cdot 0.8 + 0.2\cdot 0.9} = \frac{4}{13} \approx 31\%.$$

We updated the probability of our *hypothesis* (that he is sick) from 10% to 31%, using new information that the test is positive. The probability $\mathbf{P}(F_1)$ is called a *prior probability*, and $\mathbf{P}(F_1 \mid A)$ is called a *posterior probability*.

Example 2.8. We have a fair coin and a magic coin, which always comes out H. Choose one at random (each can be chosen with probability 1/2), flip it twice. It comes out H both times. What is the probability it is fair? Let

$$F_1 = \{\text{fair}\},\ F_2 = \{\text{magic}\},\ A = \{\text{both heads}\}.$$

Then

$$\mathbf{P}(F_1) = \mathbf{P}(F_2) = \frac{1}{2},\quad \mathbf{P}(A \mid F_1) = \frac{1}{4},\ \mathbf{P}(A \mid F_2) = 1.$$

Therefore,

$$\mathbf{P}(F_1 \mid A) = \frac{\frac{1}{2}\cdot\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{4} + \frac{1}{2}\cdot 1} = \frac{1}{5} = 20\%.$$

2.7. **Independence for two events.** Events A and B are called *independent* if knowledge of whether A happened or not does not influence the probability of B:

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

Since

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

we can rewrite this as

$$\boxed{\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)}$$

Example 2.9. Toss the coin twice. Let A = first toss H, B = second toss H, C = both tosses the same. Then A and B are independent. Indeed, the probability space (the space of all outcomes) is

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

$$A = \{HH, HT\},\quad B = \{TH, HH\},\quad C = \{TT, HH\}.$$

$$\mathbf{P}(A \cap B) = \mathbf{P}\{HH\} = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = \mathbf{P}(A)\cdot\mathbf{P}(B).$$

It is obvious that these events are independent, because they result in different tosses of the coin. In some other cases, it is not obvious. For example, A and C are also independent. Indeed,

$$\mathbf{P}(A \cap C) = \mathbf{P}\{HH\} = \frac{1}{4} = \frac{1}{2}\cdot\frac{1}{2} = \mathbf{P}(A)\cdot\mathbf{P}(C).$$

Similarly, B and C are independent.

2.8. **Independence for three or more events.** Events A, B and C are called *independent* if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B),\quad \mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C),\quad \mathbf{P}(B \cap C) = \mathbf{P}(B)\mathbf{P}(C),$$
$$\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C).$$

This last condition is important, because it does not automatically follow from the first three conditions. For example, if A, B and C are the events from the example, then A and B are independent, B and C are independent, A and C are independent, so these events are *pairwise independent*. But $A \cap B \cap C$ = HH, so

$$\mathbf{P}(A \cap B \cap C) = \frac{1}{4} \ne \frac{1}{8} = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C).$$

Example 2.10. A person likes tea with probability 50%. He likes coffee with probability 60%. He likes both tea and coffee with probability 30%. What is the probability that he likes neither tea nor coffee? Let A = likes tea, B = likes coffee. Then $A \cap B$ = likes both tea and coffee, $(A \cup B)^c$ = likes neither tea nor coffee. So

$$\mathbf{P}(A) = 50\%,\ \mathbf{P}(B) = 60\%,\ \mathbf{P}(A \cap B) = 30\%.$$

Therefore, by the inclusion-exclusion formula

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) = 80\%,$$

$$\mathbf{P}((A \cup B)^c) = 1 - \mathbf{P}(A \cup B) = \boxed{20\%}$$

Note that $\mathbf{P}(A^c) = 50\%$ and $\mathbf{P}(B^c) = 40\%$. So

$$\mathbf{P}(A^c \cap B^c) = \mathbf{P}((A \cup B)^c) = 20\% = 40\% \cdot 50\% = \mathbf{P}(A^c)\mathbf{P}(B^c).$$

These events (that a person does not like tea and that he does not like coffee) are independent. Some students used this to solve the problem. But it is just a coincidence. This independence does not follow automatically. On the contrary, you need to use inclusion-exclusion formula to establish it. For example, if we set 70% for the probability that he likes coffee, then the events would not be independent.

Problems

For the next two problems, consider the following setup. Pick a random number from 1 to 1000. Find the probabilities of the following events.

Problem 2.1. This number is not divisible by any of the numbers 2, 3, 5.

Problem 2.2. This number is not divisible by 3 but is divisible either by 2 or by 5.

Problem 2.3. You have three coins, two of them fair and the third a magic one, which rolls out Heads with probability 75%. Suppose you picked randomly one coin and tossed it once. You got Heads. What is the probability that this is the magic coin?

Problem 2.4. A person likes tea with probability 50%. He likes coffee with probability 60%. He likes both tea and coffee with probability 30%. What is the probability that he likes neither tea nor coffee?

For the next five problems, consider the following setup. Take three events: A, B, C. Draw the diagram for the event:

Problem 2.5. $(A \setminus B) \setminus C$.

Problem 2.6. $A \setminus (B \setminus C)$.

Problem 2.7. $(B \setminus A) \cup (A \setminus C)$.

Problem 2.8. $A \cup (B \cap C)$

Problem 2.9. $A^c \setminus B$.

Problem 2.10. Consider the event $A \triangle B = (A \setminus B) \cup (B \setminus A)$. It is called the *symmetric difference* of A and B and means: either A or B, but not both. Express its probability as a combination of $\mathbf{P}(A)$, $\mathbf{P}(B)$ and $\mathbf{P}(A \cap B)$.

Problem 2.11. There are two computers and a printer. Consider the following events: A = first computer works, B = second computer works, C = the printer works. The system is functioning if at least one of the computers is working and the printer is working. Express this event in terms of A, B and C.

For the next two problems, consider the following variation of the birthday problem. Suppose you take n people and for each of them, the probability that he was born at any given month is $1/12$.

Problem 2.12. Calculate $p(n)$, the probability that some of them have the same month.

Problem 2.13. What is the minimal n such that $p(n) > 50\%$?

Problem 2.14. (SOA) Upon arrival at a hospital emergency room, patients are categorized according to their condition as critical, serious, or stable. In the past year, 10% of the emergency room patients were critical, 30% of them were serious, and the rest were stable. Moreover, 40% of the critical patients died, 10% of the serious patients died, and 1% of the stable patients died. Given that a patient survived, what is the probability that the patient was categorized as serious upon arrival?

Problem 2.15. (SOA) For two events A and B, determine $\mathbf{P}(A)$, if

$$\mathbf{P}(A \cup B) = 0.7, \quad \mathbf{P}(A \cup B^c) = 0.9.$$

Problem 2.16. (SOA) Among a large group of patients recovering from shoulder injuries, it is found that 22% visit both a physical therapist and a chiropractor, whereas 12% visit neither of these. The probability that a patient visits a chiropractor exceeds by 0.14 the probability that a patient visits a physical therapist. Determine the probability that a randomly chosen member of this group visits a physical therapist.

Problem 2.17. (SOA) 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. The probability that both balls are the same color is 0.44. Calculate the number of blue balls in the second urn.

Problem 2.18. (SOA) The probability that a visit to a primary care physicians (PCP) office results in neither lab work nor referral to a specialist is 35%. Of those coming to a PCPs office, 30% are referred to specialists and 40% require lab work. Determine the probability that a visit to a PCPs office results in both lab work and referral to a specialist.

Problem 2.19. An item is defective (independently of other items) with probability 0.3. You have a method of testing whether the item is defective, but it does not always give you correct answer. If the tested item is defective, the method detects the defect with probability 0.9 (and says the item is OK with probability 0.1). If the tested item is good, then the method says it is defective with probability 0.2 (and gives the right answer with probability 0.8). A box contains 3 items. You have tested all of them and the tests detect no defects. What is the probability that none of the 3 items is defective?

For the next five problems, consider the following setup. Let $\Omega = \{1, 2, \ldots, 8\}$, $A = \{2, 4, 6, 8\}$, $B = \{1, 2, 3, 4\}$, $C = \{1, 3, 5, 7\}$. Find the following events:

Problem 2.20. $A \cup B$.

Problem 2.21. $(A \cup B \cup C)^c$.

Problem 2.22. $(A \setminus B) \cap C$.

Problem 2.23. $A \setminus C$.

Problem 2.24. $C \setminus A^c$.

Problem 2.25. (SOA) A random person has a heart disease with probability 25%. A person with a heart disease is a smoker with probability twice as a person without a heart disease. What is the conditional probability that a smoker has a heart disease?

For the next three problems, toss three fair coins. Are the following events A and B independent?

Problem 2.26. A = first and second tosses resulted in exactly one H, B = second and third tosses resulted in exactly one H.

Problem 2.27. A = first and second tosses resulted in exactly one H, B = second and third tosses resulted in at least one H.

Problem 2.28. A = first toss in H, B = second and third tosses resulted in at least one H.

Problem 2.29. Pick a random number from 1 to 600. Find the probability that this number is divisible by 3 but not by 5.

Problem 2.30. Assume that 0.1% of people from a certain population have a germ. A test gives false positive (that is, shows a person has this germ if this person does not actually have a germ) in 20% of cases when the person does not have this germ. This test gives false negative (shows a person does not have this germ if this person actually has it) in 30% of cases when this person has this germ. Suppose you pick a random person from the population and apply this test twice. Both times it gives you positive result (that is, the test says that this person has this germ). What is the probability that this person actually has this germ?

For the next three problems, consider the following setup. Genes related to albinism are denoted by A (non-albino) and a (albino). Each person has two genes. People with AA, Aa or aA are non-albino, while people with aa are albino. The non-albino gene A is called *dominant*. Any parent passes one of two genes (selected randomly with equal probability) to his offspring. If a person has either Aa or aA, then he is called a *carrier*: he is a non-albino, but he can carry the albino gene to his offsprings.

Problem 2.31. If a person is non-albine, what is the conditional probability that he is a carrier?

Problem 2.32. If a couple consists of a non-albino and an albino, what is the probability that their offsping will be albino?

Problem 2.33. Suppose that a couple consists of two non-albinos. They have two offsprings. What is the probability that both offsprings are albinos?

Problem 2.34. Consider two coins: a fair one and a magic one, which always falls H. Suppose you pick a coin at random (each coin with equal probability 1/2) and toss it n times. All of the tosses are heads. You would like to convince me that the coin is magic, but I will believe you only if the probability that it is magic (given the result that all n tosses are H) exceeds 99%. How large the number n of tosses should be?

Problem 2.35. Suppose there is a rare disease which affects 1% of people. There is a test which can tell whether a person has this disease or not. If the person is healthy, the test shows he is healthy with probability 90%. If the person has this disease, the test shows this with probability 80%. A random person is selected from the population, and the test is applied twice. Both times, it shows that the person has this disease. What is the probablity that this person actually has this disease?

For the next two problems, consider the following setup. Consider a *collateralized debt obligation* (CDO) backed by ten subprime mortgages. Five of them are from California, each of which defaults with probability 50%. The other five mortgages are from Florida, each of which defaults with probability 60%. A *senior tranch* in this CDO defaults only if *all* of these mortgages default. Find the probability that the senior tranch does not default in the following two cases.

Problem 2.36. These mortgages all default independently of each other.

Problem 2.37. All five California mortgages default or do not default simultaneously, and the same is true for the five Florida mortgages, but California defaults are independent of Florida defaults.

For the next four problems, consider the following setup. Toss a coin three times. Consider the events: $A =$ first H, $B =$ second H, $C =$ third H, $D =$ even number of H. Find whether the following events are independent.

Problem 2.38. A, B, C

Problem 2.39. A and D.

Problem 2.40. A, B, D.

Problem 2.41. A, B, C, D.

Problem 2.42. (SOA) An auto owner can purchase a collision coverage and a disability coverage. These puchases are independent of each other. He is twice as likely to purchase a collision coverage than a disability coverage. The probability that he purchases both is 15%. What is the probability that he purchases neither?

Problem 2.43. Consider two independent events A and B. Find $\mathbf{P}(B)$, if you know that

$$\mathbf{P}(A) = 2\mathbf{P}(B), \quad \text{and} \quad \mathbf{P}(A \setminus B) = 0.1.$$

Problem 2.44. Suppose $\mathbf{P}(A) = 0.9$ and $\mathbf{P}(B) = 0.8$. What are the possible values of $\mathbf{P}(A \cap B)$?

Problem 2.45. Toss two dice. Let $A =$ the sum on the dice is even, $B =$ the first die is even. What is $\mathbf{P}(A \mid B)$ and $\mathbf{P}(B \mid A)$?

3. Discrete Random Variables

3.1. **Definition of random variables.** Recall the concept of a probability space: Ω consists of all possible elementary outcomes. For example, toss a coin twice. Then $\Omega = \{HH, HT, TH, TT\}$. A *random variable* is a function $X : \Omega \to \mathbb{R}$.

Example 3.1. Let Y be the number of tails. Then

$$Y(HH) = 0, \quad Y(TH) = Y(HT) = 1, \quad Y(TT) = 2.$$

The *distribution* of a random variable is all the values which it can assume together with probabilities of these values. For this variable Y, this is

$$P(Y = 0) = \frac{1}{4}, \quad \mathbf{P}(Y = 1) = \frac{1}{2}, \quad \mathbf{P}(Y = 2) = \frac{1}{4}.$$

3.2. **Expectation.** The *mean value*, or *expected value*, $\mathbf{E}X$, of X is calculated as follows: suppose that the probability of each elementary outcome $\omega \in \Omega$ is $p(\omega)$. Then

$$\mathbf{E}X = \sum_{\omega \in \Omega} P(\{\omega\}) X(\omega).$$

For example, for the variable Y in the Example 3.1,

$$\begin{aligned}\mathbf{E}Y &= \mathbf{P}\{HH\} \cdot Y(HH) + \mathbf{P}\{HT\} \cdot Y(TH) + \mathbf{P}\{TH\} \cdot Y(HT) + \mathbf{P}\{TT\} \cdot Y(TT) \\ &= \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 2 = 1.\end{aligned}$$

We can also calculate $\mathbf{E}X$ using distribution of X: if X assumes values $x_1, \ldots, x_n$ with probabilities $\mathbf{P}(X = x_1), \ldots, \mathbf{P}(X = x_n)$, then

$$\boxed{\mathbf{E}X = \sum_{k=1}^{n} x_k \cdot \mathbf{P}(X = x_k)}$$

In Example 3.1

$$\mathbf{E}Y = \mathbf{P}(Y = 0) \cdot 0 + \mathbf{P}(Y = 1) \cdot 1 + \mathbf{P}(Y = 2) \cdot 2 = \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 = 1.$$

This gives us two ways to calculate the expected value. Use whichever is more convenient. Also,

$$\mathbf{E}Y^2 = \mathbf{P}(Y = 0) \cdot 0^2 + \mathbf{P}(Y = 1) \cdot 1^2 + \mathbf{P}(Y = 2) \cdot 2^2 = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 4 = \frac{3}{2}.$$

It can happen that X takes infinitely many values, then $\mathbf{E}X$ might be infinite. For example, suppose

$$\mathbf{P}(X = 2^k) = 2^{-k}, \quad k = 1, 2, 3, \ldots$$

In other words,

$$\mathbf{P}(X=2)=\frac{1}{2},\ \ \mathbf{P}(X=4)=\frac{1}{4},\ \ \mathbf{P}(X=8)=\frac{1}{8},\dots$$

$$\mathbf{E}X=\sum_{k=1}^{\infty}2^k\mathbf{P}(X=2^k)=\sum_{k=1}^{\infty}2^k2^{-k}=\sum_{k=1}^{\infty}1=\infty.$$

In general, if we have a function $f(X)$ of X, then $f(X)$ is also a random variable. We can calculate the expectation of $f(X)$:

$$\boxed{\mathbf{E}f(X)=\sum_{k=1}^{n}f(x_k)\mathbf{P}(X=x_k)}$$

3.3. **Variance.** How to measure the deviation of a random variable X from its mean $m=\mathbf{E}X$? We take the difference $X-m$ and square it. Then we get a variable $(X-m)^2$ (which is also random) and take the expectation $\mathbf{E}(X-m)^2$. This is called the *variance* of X:

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

We can also rewrite it as

$$\begin{aligned}\operatorname{Var}X&=\mathbf{E}(X^2-2Xm+m^2)=\mathbf{E}X^2-\mathbf{E}(2Xm)+m^2=\mathbf{E}X^2-2m\mathbf{E}X+m^2\\&=\mathbf{E}X^2-2m\cdot m+m^2=\mathbf{E}X^2-m^2.\end{aligned}$$

So we have:

$$\boxed{\operatorname{Var}X=\mathbf{E}(X-\mathbf{E}X)^2=\mathbf{E}X^2-(\mathbf{E}X)^2}$$

The square root $\sigma_X=\sqrt{\operatorname{Var}X}$ is called the *standard deviation*. It has the same order as the random variable itself (for example, if X is in inches, then $\operatorname{Var}X$ is in square inches, and σ_X is in inches).

Example 3.2. For the random variable Y in Example 3.1,

$$\operatorname{Var}Y=\mathbf{E}Y^2-(\mathbf{E}Y)^2=\frac{3}{2}-1^2=\frac{1}{2}.$$

3.4. **Joint distribution.** Two variables X,Y have a *joint distribution*, which governs the probabilities $\mathbf{P}(X=x,Y=y)$. A joint distribution is a collection of all such probabilities for each x and y.

Example 3.3. Toss a coin twice. Let X be the quantity of heads, let Y be 1 if there is a head for the second toss, and 0 otherwise. Then

$$\begin{cases}X(HH)=2,\ Y(HH)=1\\X(HT)=1,\ Y(HT)=0\\X(TH)=1,\ Y(TH)=1\\X(TT)=0,\ Y(TT)=0\end{cases}$$

The joint distribution of X,Y is

$$\mathbf{P}(X=2,Y=1)=\frac{1}{4},\ \mathbf{P}(X=1,Y=0)=\frac{1}{4},\ \mathbf{P}(X=Y=1)=\frac{1}{4},\ \mathbf{P}(X=Y=0)=\frac{1}{4}.$$

We have:

$$\mathbf{E}(XY)=\frac{1}{4}\cdot2\cdot1+\frac{1}{4}\cdot1\cdot0+\frac{1}{4}\cdot1\cdot1+\frac{1}{4}\cdot0\cdot0=\frac{3}{4}.$$

If we know the joint distribution of X,Y, we can find the distribution of X and the distribution of Y, which in this case are called *marginal distributions*.

Example 3.4. Suppose the joint distribution of X and Y is given by

$$\begin{cases}\mathbf{P}(X=0,Y=0)=0.2,\\\mathbf{P}(X=1,Y=0)=0.1,\\\mathbf{P}(X=0,Y=1)=0.3,\\\mathbf{P}(X=1,Y=1)=0.4\end{cases}$$

Then the marginal distribution of X is given by

$$\mathbf{P}(X=0)=0.2+0.3=0.5,\ \mathbf{P}(X=1)=0.1+0.4=0.5,$$

and the marginal distribution of Y is given by

$$\mathbf{P}(Y=0)=0.2+0.1=0.3,\quad \mathbf{P}(Y=1)=0.3+0.4=0.7.$$

3.5. **Independent random variables.** Two discrete random variables X, Y are called *independent* if

$$\mathbf{P}(X=x, Y=y)=\mathbf{P}(X=x)\cdot\mathbf{P}(Y=y)$$

for all values x and y assumed by X and Y.

Example 3.5. Let $X=1$ if the first toss is H, otherwise 0; $Y=1$ if the second toss is H, otherwise 0. Then

$$\mathbf{P}(X=1, Y=0)=\frac{1}{4}=\frac{1}{2}\frac{1}{2}=\mathbf{P}(X=1)\mathbf{P}(Y=0).$$

Similarly for other values of x and y. So X and Y are independent. But let $Z=X+Y$. Then X and Z are not independent. Indeed,

$$\mathbf{P}(X=1, Z=2)=\mathbf{P}(X=1, Y=1)=\frac{1}{4},$$

but

$$\mathbf{P}(X=1)=\frac{1}{2},\quad \mathbf{P}(Z=2)=\mathbf{P}(X=1, Y=1)=\frac{1}{4}$$

(and $1/4\neq 1/4\cdot 1/2$). Also, X and X are not independent (because it is the same variable!). Indeed,

$$\mathbf{P}(X=1, X=1)=\frac{1}{2}\neq\frac{1}{2}\cdot\frac{1}{2}=\mathbf{P}(X=1)\mathbf{P}(X=1).$$

3.6. **Properties of expectation.** We always have: $\mathbf{E}(X+Y)=\mathbf{E}X+\mathbf{E}Y$. For product instead of sum, this is no longer true: $\mathbf{E}(XY)\neq\mathbf{E}X\cdot\mathbf{E}Y$. Indeed, let $X=Y$ be the random variable from the example above (result of first toss). Then

$$\mathbf{E}(XY)=\mathbf{E}X^2=\frac{1}{2}\cdot 1^2+\frac{1}{2}\cdot 0^2=\frac{1}{2},$$

but

$$\mathbf{E}X\cdot\mathbf{E}Y=(\mathbf{E}X)^2=(1/2)^2=\frac{1}{4}\neq\frac{1}{2}.$$

However, for independent random variables, we actually have:

$$\boxed{\mathbf{E}(XY)=\mathbf{E}X\cdot\mathbf{E}Y,\quad \text{if}\quad X, Y\quad \text{are independent}}$$

Indeed,

$$\mathbf{E}(XY)=\sum_{x,y}xy\mathbf{P}(X=x, Y=y)=\sum_{x,y}xy\mathbf{P}(X=x)\mathbf{P}(Y=y)$$
$$=\sum_x x\mathbf{P}(X=x)\sum_y y\mathbf{P}(Y=y)=\mathbf{E}X\mathbf{E}Y.$$

3.7. **Covariance.** Random variables X and Y have covariance:

$$\operatorname{Cov}(X,Y)=\mathbf{E}(X-\mathbf{E}X)(Y-\mathbf{E}Y).$$

This is a measure of the following: when one variable increases, does the other increase or decrease (on average)? We can also calculate it as

$$\operatorname{Cov}(X,Y)=\mathbf{E}(XY)-\mathbf{E}X\cdot\mathbf{E}Y.$$

If two variables X and Y are independent, then $\mathbf{E}(XY)=\mathbf{E}X\cdot\mathbf{E}Y$, and so their covariance is equal to zero: $\operatorname{Cov}(X,Y)=0$. For two random variables, if their covariance is positive, they are called *positively correlated*; if it is negative, they are called *negatively correlated*. *Correlation* is defined as

$$\operatorname{Corr}(X,Y)=\frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}X}\sqrt{\operatorname{Var}Y}}.$$

It is always between -1 and $+1$. Such extreme values ± 1 show that X and Y are linearly dependent.

Example 3.6. Consider two tosses of a coin. Let $X = 1$ if the first toss is H, $X = 0$ otherwise. Z is the quantity of heads in two tosses. Then

$$\begin{cases} X(HH) = 1, Z(HH) = 2 \\ X(HT) = 1, Z(HT) = 1 \\ X(TH) = 0, Z(TH) = 1 \\ X(TT) = 0, Z(TT) = 0 \end{cases}$$

Therefore,

$$\mathbf{E}(XZ) = \frac{1}{4}\cdot 1\cdot 2 + \frac{1}{4}\cdot 1\cdot 1 + \frac{1}{4}\cdot 0\cdot 1 + \frac{1}{4}\cdot 0\cdot 0 = \frac{3}{4},\quad \mathbf{E}X = \frac{1}{2},\quad \mathbf{E}Z = 1.$$

Thus,

$$\operatorname{Cov}(X, Z) = \frac{3}{4} - \frac{1}{2}\cdot 1 = \frac{1}{4}.$$

3.8. **Properties of covariance.** The following is easy to check:

$$\operatorname{Cov}(X + Y, Z) = \operatorname{Cov}(X, Z) + \operatorname{Cov}(Y, Z),\quad \operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X),\quad \operatorname{Cov}(X, X) = \operatorname{Var} X.$$

Example 3.7. $\mathbf{P}(X = 0, Y = 0) = 0.5$, $\mathbf{P}(X = 1, Y = 0) = 0.3$, $\mathbf{P}(X = 0, Y = 1) = 0.2$. So the distribution of X is

$$\mathbf{P}(X = 0) = 0.7,\ \mathbf{P}(X = 1) = 0.3,$$

and $\mathbf{E}X = 0.3$. The distribution of Y is

$$\mathbf{P}(Y = 0) = 0.8,\ \mathbf{P}(Y = 1) = 0.2,$$

and $\mathbf{E}Y = 0.2$. Also, always $XY = 0$, so

$$\operatorname{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y = 0 - 0.3\cdot 0.2 = -0.06.$$

3.9. **Properties of variance.** Suppose X and Y are random variables. Then

$$\boxed{\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y + 2\operatorname{Cov}(X, Y)}$$

Indeed,

$$\operatorname{Var}(X + Y) = \mathbf{E}(X + Y - \mathbf{E}X - \mathbf{E}Y)^2.$$

Applying $(a + b)^2 = a^2 + 2ab + b^2$ to $a = X - \mathbf{E}X$ and $b = Y - \mathbf{E}Y$, we have:

$$\mathbf{E}(X - \mathbf{E}X)^2 + \mathbf{E}(Y - \mathbf{E}Y)^2 + 2\mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y) = \operatorname{Var} X + \operatorname{Var} Y + 2\operatorname{Cov}(X, Y).$$

If X and Y are independent, $\operatorname{Cov}(X, Y) = 0$ and

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

Example 3.8. Toss a coin twice. Let $X = 1$, if the first toss is heads, and $X = 0$ otherwise. Let $Y = 1$ if the second toss is heads, $Y = 1$ otherwise. Then X and Y are independent. So $\operatorname{Var}(X+Y) = \operatorname{Var} X+\operatorname{Var} Y = 1/4+1/4 = 1/2$. (We calculated $\operatorname{Var} X = \operatorname{Var} Y = 1/4$ before.)

3.10. **Diversification.** Suppose you have an asset X which costs now 25; it will bring you 100 or nothing with equal probability. Then

$$\mathbf{E}X = \frac{1}{2}\cdot 100 + \frac{1}{2}\cdot 0 = 50 > 25.$$

So it seems to be a good asset. But it has high variance: it is unreliable, and you are likely to lose a lot. A way out: take two such assets X and Y, which are independent (stocks of different companies) and invest equal proportion of your wealth in each asset. You will spend 25, and your return is $S = (X + Y)/2$, with mean and variance

$$\mathbf{E}S = \frac{1}{2}\left(\mathbf{E}X + \mathbf{E}Y\right) = \frac{1}{2}\left(50 + 50\right) = 50,$$

$$\operatorname{Var} S = \frac{1}{2^2}\operatorname{Var}(X + Y) = \frac{1}{4}\left(\operatorname{Var} X + \operatorname{Var} Y\right) = \frac{1}{2}\operatorname{Var} X.$$

The variance is twice as small as before. You get the same average return but with low risk. This is called *diversification*; this is essential for finance and insurance. This is why insurance companies need large pools of clients.

3.11. **Modern portfolio theory.** This was created by Harry Markowitz in 1952, for which he won a Nobel Prize. Assume we have three assets to invest in, with returns (future price divided by current price) X, Y, Z. Their mean and variance in a year are estimated to be

$$\mathbf{E}X = 3,\ \mathbf{E}Y = 2,\ \mathbf{E}Z = 4, \quad \text{and} \quad \operatorname{Var} X = 2,\ \operatorname{Var} Y = 1,\ \operatorname{Var} Z = 4,$$

and they are independent. How to allocate your funds between these three assets? We need to choose proportions x, y, z, which satisfy

$$x, y, z \ge 0 \quad \text{and} \quad x + y + z = 1. \tag{1}$$

The return of our portfolio is $R = xX + yY + zZ$, with

$$\mathbf{E}R = 3x+2y+4z, \quad \operatorname{Var} R = \operatorname{Var}(xX)+\operatorname{Var}(yY)+\operatorname{Var}(zZ) = x^2 \operatorname{Var} X+y^2 \operatorname{Var} Y+z^2 \operatorname{Var} Z = 2x^2+y^2+4z^2. \tag{2}$$

Now we need to maximize the return $3x + 2y + 4z$ given acceptable level of variance: $2x^2 + y^2 + 4z^2 \le v$, where you can choose your v, say $v = 1$, and given constraints (1). This is done via *Lagrange multipliers*, which can be found in any optimization textbook. Let us give some examples:

Example 3.9. *Equally weighted* portfolio: $x = y = z = 1/3$. Then $m = 3$ and $\sigma^2 = 7/9 = 0.777$. Note that the variance is smaller than for each individual stock. This is the result of our diversification.

Example 3.10. Suppose the investor is risk-averse and invests mostly in the stock Y: $x = 0.25$, $y = 0.5$, $z = 0.25$. Then

$$m = 3 \cdot 0.25 + 2 \cdot 0.5 + 4 \cdot 0.25 = 2.75,$$
$$\sigma^2 = 2 \cdot 0.25^2 + 0.5^2 + 4 \cdot 0.25^2 = 0.625.$$

In this case, the mean is smaller than for the equally weighted portfolio: $2.75 < 3$, but the variance is also smaller: $0.625 < 0.777$.

But stocks can be also correlated: Assume

$$\operatorname{corr}(X, Y) = 0.4, \quad \text{and} \quad \operatorname{corr}(X, Z) = \operatorname{corr}(Y, Z) = 0.$$

Then $\operatorname{Cov}(X, Y) = \operatorname{corr}(X, Y)\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)} = 0.4 \cdot \sqrt{2 \cdot 1} = \sqrt{2}0.4$. The mean of return is also given by (2), but variance is given by

$$\begin{aligned}\operatorname{Var} R &= \operatorname{Var}(xX) + \operatorname{Var}(yY) + \operatorname{Var}(zZ) + 2\operatorname{Cov}(xX, yY) \\ &= x^2 \operatorname{Var} X + y^2 \operatorname{Var} Y + 2xy \operatorname{Cov}(X, Y) + z^2 \operatorname{Var} Z = 2x^2 + y^2 + 2 \cdot \sqrt{2}0.4xy + 4z^2.\end{aligned}$$

Example 3.11. *Equally weighted portfolio:* $x = y = z = 1/3$, and

$$\sigma^2 = 2 \cdot (1/3)^2 + (1/3)^2 + 4 \cdot (1/3)^2 + 2\sqrt{2}0.4(1/3)^2 = 7/9 + 0.8\sqrt{2}/9.$$

Variance here is larger than for independent stocks, because of positive correlation.

Problems

For the next eight problems, consider the following setup. Let X be the random number on a die: from 1 to 6.

Problem 3.1. What is the distribution of X?

Problem 3.2. Calculate $\mathbf{E}X$.

Problem 3.3. Calculate $\mathbf{E}X^2$.

Problem 3.4. Calculate $\operatorname{Var} X$.

Problem 3.5. Calculate $\mathbf{E}X^3$.

Problem 3.6. Calculate $\mathbf{E}\sin(\pi X)$.

Problem 3.7. Calculate $\mathbf{E}\cos(\pi X)$.

Problem 3.8. Calculate $\mathbf{E}(2X - 4)$.

For the next eleven problems, consider the following setup. Throw a die twice and let X, Y be the results.

Problem 3.9. What is the distribution of $X + Y$?

Problem 3.10. Calculate $\mathbf{E}(X + Y)$.

Problem 3.11. Calculate $\mathbf{E}(2X - 3Y)$.

Problem 3.12. Calculate $\mathbf{E}(XY)$.

Problem 3.13. Calculate $\mathbf{E}(X^2Y^2)$.

Problem 3.14. Calculate $\mathrm{Var}(X+Y)$.

Problem 3.15. Calculate $\mathrm{Var}(XY)$.

Problem 3.16. Find $\mathrm{Cov}(X, X+Y)$.

Problem 3.17. Find $\mathrm{Cov}(X-Y, X+Y)$.

Problem 3.18. Are X and Y independent?

Problem 3.19. Are X and $X+Y$ independent?

Problem 3.20. Toss a fair coin 3 times. Let X be the total number of heads. What is the probability space Ω, and what is the distribution of X?

Problem 3.21. (A game from the Wild West) You bet one dollar and throw three dice. If at least one of them is six, then I return you your dollar and give you as many dollars as the quantity of dice which resulted in six. For example, if there are exactly two dice which gave us six, then you win two dollars. If you do not get any dice with six, then you lose your dollar. So you either lose 1\$ or win 1\$, 2\$ or 3\$. What is the expected value of the amount you win?

Problem 3.22. An urn contains 11 balls, 3 white, 3 red, and 5 blue balls. Take out two balls at random, without replacement. You win \$1 for each red ball you select and lose a \$1 for each white ball you select. Determine the distribution of X, the amount you win.

For the next three problems, consider the following setup. An urn contains 10 balls numbered $1, \ldots, 10$. Select 2 balls at random, without replacement. Let X be the smallest number among the two selected balls.

Problem 3.23. Determine the distribution of X.

Problem 3.24. Find $\mathbf{E}X$ and $\mathrm{Var}\,X$.

Problem 3.25. Find $\mathbf{P}(X \le 3)$.

Problem 3.26. If $\mathrm{Var}\,X = 0$ and $\mathbf{E}X = 3$, then what can you say about the random variable X?

Problem 3.27. An insurance company determines that N, the number of claims received in a week, is a random variable with

$$\mathbf{P}(N = n) = \frac{1}{2^{n+1}}, \quad n = 0, 1, 2, \ldots.$$

The company also determines that the number of claims received in a given week is independent of the number of claims received in any other week. Determine the probability that exactly 7 claims will be received during a given two-week period.

Problem 3.28. (SOA) The number of injury claims per month is modeled by a random variable N with

$$\mathbf{P}(N = n) = \frac{1}{(n+1)(n+2)}, \quad n = 0, 1, 2, \ldots$$

Determine the probability of at least one claim during a particular month, given that there have been at most four claims during that month.

For the next ten problems, consider the following setup. Take two variables: X, Y, with distribution

$$\mathbf{P}(X = -1, Y = 1) = 0.2, \quad \mathbf{P}(X = 2, Y = 1) = 0.3,$$
$$\mathbf{P}(X = -1, Y = 3) = 0.1, \quad \mathbf{P}(X = 2, Y = 3) = 0.4.$$

Problem 3.29. Find the distribution of X.

Problem 3.30. Find the distribution of Y.

Problem 3.31. Calculate $\mathbf{E}X$.

Problem 3.32. Calculate $\mathbf{E}Y$.

Problem 3.33. Calculate $\mathbf{E}(XY)$.

Problem 3.34. Calculate $\mathbf{E}X^2$.

Problem 3.35. Calculate $\operatorname{Var} Y$.

Problem 3.36. Calculate $\mathbf{E}(X^2Y)$.

Problem 3.37. Calculate $\operatorname{Cov}(X, Y)$.

Problem 3.38. Are X and Y independent?

Problem 3.39. (SOA) Let X denote the size of a surgical claim and let Y denote the size of the associated hospital claim. An actuary is using a model in which

$$\mathbf{E}(X) = 5, \quad \mathbf{E}(X^2) = 27.4, \quad \mathbf{E}(Y) = 7, \quad \mathbf{E}(Y^2) = 51.4, \quad \operatorname{Var}(X+Y) = 8.$$

Let $C_1 = X + Y$ denote the size of the combined claims before the application of a 20% surcharge on the hospital portion of the claim, and let C_2 denote the size of the combined claims after the application of that surcharge. Calculate $\operatorname{Cov}(C_1, C_2)$.

For the next three problems, consider the following setup. Suppose that ρ is a distribution on $\{-1, 0, 1\}$, with probabilities $1/3$ attached to each of these three numbers. Let $X, Y, Z \backsim \rho$ be i.i.d. random variables.

Problem 3.40. Find $\mathbf{E}(X + Y + Z)$.

Problem 3.41. Find $\mathbf{E}(XYZ)$.

Problem 3.42. Find $\mathbf{P}(X = 1 \mid X + Y + Z = 0)$.

For the next five problems, consider the following setup. Let X, Y, Z be random variables with the following joint distribution:

X	Y	Z	Prob.
0	0	0	1/8
1	0	1	1/8
0	1	2	1/8
-1	1	1	1/8
3	1	0	1/2

Problem 3.43. Find the distribution and the expectation of each of these three random variables.

Problem 3.44. Find $\operatorname{Cov}(X, Y)$.

Problem 3.45. Find the distribution of $XY + Z$.

Problem 3.46. Calculate $\mathbf{E}(X + 1)^4$.

Problem 3.47. Calculate $\mathbf{E}(XYZ)$.

Problem 3.48. Toss a fair coin 3 times. Let X, Y be numbers of Heads and Tails. Find $\operatorname{Cov}(X, Y)$.

Problem 3.49. Find the distribution of $XY - Z$ for random variables X, Y, Z with joint distribution

X	Y	Z	Prob.
2	1	0	0.5
0	1	-1	0.3
1	0	-2	0.2

Problem 3.50. (SOA) A car dealership sells 0, 1, or 2 luxury cars on any day. When selling a car, the dealer also tries to persuade the customer to buy an extended warranty for the car. Let X denote the number of luxury cars sold in a given day, and let Y denote the number of extended warranties sold. We have:

$$\mathbf{P}(X = 0,\ Y = 0) = \frac{1}{6}, \quad \mathbf{P}(X = 1,\ Y = 0) = \frac{1}{12}, \quad \mathbf{P}(X = 2,\ Y = 0) = \frac{1}{12},$$
$$\mathbf{P}(X = 1,\ Y = 1) = \frac{1}{6}, \quad \mathbf{P}(X = 2,\ Y = 1) = \frac{1}{3}, \quad \mathbf{P}(X = 2,\ Y = 2) = \frac{1}{6}.$$

Calculate the variance of X.

Problem 3.51. Consider two random variables X and Y with joint distribution

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

Are they independent? Find $\operatorname{Cov}(X, Y)$.

Problem 3.52. Consider a random variable X with distribution

$$\mathbf{P}(X = -1) = 0.4, \quad \mathbf{P}(X = 0) = 0.4, \quad \mathbf{P}(X = 1) = 0.2$$

Find its expectation and the standard deviation.

Problem 3.53. Two independent random variables X and Y have joint distribution

$$\mathbf{P}(X = 2,\ Y = 0) = \frac{1}{2},\ \mathbf{P}(X = 2,\ Y = 1) = \frac{1}{4},\ \mathbf{P}(X = 1,\ Y = 0) = p,\ \mathbf{P}(X = 1,\ Y = 1) = q.$$

Find p and q.

Problem 3.54. Toss a fair coin three times. Let X be the number of Heads during the first two tosses. Let Y be the number of Tails during the last two tosses. For example, if the sequence of tosses is TTH, then $X = 0$ and $Y = 1$. Are X and Y independent? Why or why not?

For the next six problems, consider the following setup. Toss a fair coin twice and let X be the number of Heads, and let $Y = 1$ if the first toss is Tails, $Y = 0$ otherwise. Find:

Problem 3.55. The distribution of X.

Problem 3.56. The distribution of Y.

Problem 3.57. $\mathbf{E}X$.

Problem 3.58. $\operatorname{Var} X$.

Problem 3.59. $\operatorname{Cov}(X, Y)$.

Problem 3.60. $\mathbf{P}(X = 2 \mid Y = 0)$.

For the next four problems, we have a portfolio of 4 stocks, the first and the second of which have returns with mean 1 and variance 2, and the third and the fourth have returns with mean 1.2 and variance 3. For each setup, find the mean and variance of the return.

Problem 3.61. All returns are independent, equally-weighted portfolio.

Problem 3.62. All returns are independent, portfolio which invests equally in the first three stocks.

Problem 3.63. Correlation between the first two stocks is -0.8, all other correlations are zero, equally-weighted portfolio.

Problem 3.64. Correlation between the second and the third stock is 0.5, all other correlations are zero, portfolio which invests 30% in each of the first two stocks, and 20% in the last two stocks.

Problem 3.65. We invest in 10 independent assets, each has return with mean 1 and variance 7. Find the mean and variance of the equally-weighted portfolio.

Problem 3.66. We invest in 50 independent assets, 20 of which have returns with mean 0.4 and variance 1, 30 of which have returns with mean 0.5 and variance 1.4. Find the mean and variance for the equally-weighted portfolio.

Problem 3.67. We invest in N independent assets, each has return with mean 1 and variance 2. For which N do we have the variance of the equally-weighted portfolio less than or equal to 0.1?

4. Common Discrete Distributions

4.1. **Bernoulli distribution.** Suppose you have a biased coin, which falls H with probability p, $0 < p < 1$. You toss it once. Let $X = 1$ if the first toss resulted in H, $X = 0$ if the first toss resulted in T. Therefore, the distribution of X is

$$\mathbf{P}(X = 1) = p,\ \mathbf{P}(X = 0) = q = 1 - p.$$

This is called a *Bernoulli distribution* with *probability of success* p. We can calculate its expectation and variance:

$$\mathbf{E}X = 1 \cdot p + 0 \cdot q = p, \quad \mathbf{E}X^2 = 1^2 \cdot p + 0^2 \cdot q = p,$$

$$\operatorname{Var} X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = p - p^2 = p(1 - p) = pq.$$

4.2. **Binomial distribution.** Suppose you have a biased coin, which falls H with probability p, $0 < p < 1$. You toss it N times. Let X be the number of H. What is the distribution of the random variable X? We call this the sequence of *Bernoulli trials*: each trial can independently result in success (H) or failure (T), with probabilities p and $q = 1 - p$ respectively. Let $X_1 = 1$ if the first toss resulted in H, $X_1 = 0$ if the first toss resulted in T. Same for X_2, X_3, $\ldots$. Then

$$\boxed{X = X_1 + X_2 + \ldots + X_N}$$

The random variables $X_1, \ldots, X_N$ are Bernoulli, independent and identically distributed (i.i.d.): $\mathbf{P}(X_i = 1) = p$, $\mathbf{P}(X_i = 0) = q$. What is the probability that $X = k$? Let us consider the extreme cases: all H or all T.

$$\mathbf{P}(X = 0) = q^N, \quad \mathbf{P}(X = N) = p^N.$$

Now, consider the case of general k. We can choose which k tosses out of N result in H, there are $\binom{N}{k}$ choices. Each choice has probability $p^k q^{N-k}$. So

$$\boxed{\mathbf{P}(X = k) = \binom{N}{k} p^k q^{N-k}, \quad k = 0, \ldots, N}$$

We denote the distribution of X by $\mathrm{Bin}(N, p)$: *binomial distribution with parameters N and p*, or *sequence of N Bernoulli trials with probability p of success*. We write: $X \backsim \mathrm{Bin}(N, p)$. Let us calculate the expectation and variance of the binomial distribution:

$$\mathbf{E}X = \mathbf{E}X_1 + \ldots + \mathbf{E}X_N = p + p + \ldots + p = N.$$

Since $X_1, \ldots, X_N$ are independent, we have:

$$\operatorname{Var} X = \operatorname{Var} X_1 + \ldots + \operatorname{Var} X_N = pq + \ldots + pq = Npq.$$

Example 4.1. Toss twice a biased coin, when the probability of Heads is $p = 2/3$. The number of Heads is $X \backsim \mathrm{Bin}(2, 2/3)$. Therefore,

$$\mathbf{P}(X = 0) = \mathbf{P}(TT) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9},$$

$$\mathbf{P}(X = 1) = \mathbf{P}(TH, HT) = \mathbf{P}(TH) + \mathbf{P}(HT) = \frac{1}{3} \cdot \frac{2}{3} + \frac{2}{3} \cdot \frac{1}{3} = \frac{4}{9},$$

or, alternatively,

$$\mathbf{P}(X = 1) = \binom{2}{1} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^1 = 2 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9},$$

and, finally,

$$\mathbf{P}(X = 2) = \mathbf{P}(HH) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}.$$

4.3. **Poisson distribution.** Fix a parameter $\lambda > 0$. Consider a distribution on $0, 1, 2, 3, \ldots$, given by the formula

$$\boxed{\mathbf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots}$$

This is called the *Poisson distribution* with parameter λ. We write: $X \backsim \mathrm{Poi}(\lambda)$. For example,

$$\mathbf{P}(X = 0) = e^{-\lambda}, \quad \mathbf{P}(X = 1) = \lambda e^{-\lambda}, \quad \mathbf{P}(X = 2) = \frac{\lambda^2}{2} e^{-\lambda}, \quad \mathbf{P}(X = 3) = \frac{\lambda^3}{6} e^{-\lambda}.$$

Recall the Taylor series for exponential function:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \ldots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

This means that

$$\sum_{k=0}^{\infty} \mathbf{P}(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

And it should be so, since for a probability distribution, all individual probabilities should sum up to 1. What is $\mathbf{E}X$ for $X \backsim \mathrm{Poi}(\lambda)$?

$$\mathbf{E}X = \sum_{k=0}^{\infty} k \cdot \mathbf{P}(X = k) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} =$$

(because the term corresponding to $k = 0$ is zero, so we can omit it from the sum)

$$= \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} =$$

(because $k/k! = 1/(k-1)!$)

$$= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} e^{-\lambda} = \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} = \lambda.$$

This proves that

$$\boxed{\mathbf{E} X = \lambda}$$

After calculation, we can also show that

$$\boxed{\operatorname{Var} X = \lambda}$$

4.4. **Approximation theorem.** Consider a binomial distribution $\mathrm{Bin}(N, p)$. If $N \to \infty$, and $Np = \lambda$, then $\mathrm{Bin}(N, p) \to \mathrm{Poi}(\lambda)$. More precisely, for every fixed $k = 0, 1, 2, \ldots$, we have as $N \to \infty$:

$$\binom{N}{k} p^k q^{N-k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

Proof. Note that $p = \lambda/N$, $q = 1 - \lambda/N$, and

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} = \frac{N(N-1)\ldots(N-k+1)}{N^k}.$$

Therefore,

$$\binom{N}{k} p^k q^{N-k} = \frac{N(N-1)\ldots(N-k+1)}{N^k} \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^{N-k}$$
$$= \frac{\lambda^k}{k!} \cdot \frac{N(N-1)\ldots(N-k+1)}{N^k} \cdot \left(1 - \frac{\lambda}{N}\right)^{N-k}.$$

The first factor $\lambda^k/k!$ is constant (does not depend on N); the second factor is

$$\frac{N(N-1)\ldots(N-k+1)}{N^k} = \frac{N}{N} \cdot \frac{N-1}{N} \cdot \ldots \cdot \frac{N-k+1}{N} \to 1,$$

because each of these fractions tends to one: indeed, $N \to \infty$, and k is fixed. Finally, the third factor tends to $e^{-\lambda}$, because

$$\log\left(1 - \frac{\lambda}{N}\right)^{N-k} = (N-k)\log\left(1 - \frac{\lambda}{N}\right) \backsim (N-k)\left(-\frac{\lambda}{N}\right) = -\lambda\frac{N-k}{N} \to -\lambda,$$

and then we exponentiate both sides. We use the tangent line approximation:

$$\log(1+x) \backsim x,\ x \to 0,$$

which is true: let $f(x) = \log(1+x)$; then $f'(x) = 1/(1+x)$, and so $f(0) = 0$, $f'(0) = 1$, and

$$f(x) \backsim f(0) + f'(0)x = x, \quad x \to 0.$$

□

This is the meaning of a Poisson distribution: this is approximately the quantity of many events, each of which is very rare.

Example 4.2. A company which sells flood insurance has $N = 20000$ clients, and each client has probability of flood $p = 0.01\%$, independently of others. What is the distribution of the number X of floods? Approximately it is $X \backsim \mathrm{Bin}(N, p)$, but it is hard for calculations: for example,

$$\mathbf{P}(X = 10) = \binom{20000}{10} \left(\frac{1}{10000}\right)^{10} \left(\frac{9999}{10000}\right)^{19990}.$$

Use Poisson approximation: X is approximately distributed as $\mathrm{Poi}(2)$, because $\lambda = Np = 2$. So

$$\mathbf{P}(X = 10) = \frac{2^{10}}{10!} e^{-2}.$$

Also,

$$\mathbf{P}(X \geq 3) = 1 - \mathbf{P}(X = 0) - \mathbf{P}(X = 1) - \mathbf{P}(X = 2)$$

$$= 1 - e^{-2} - \frac{2^1}{1!}e^{-2} - \frac{2^2}{2!}e^{-2} = 1 - e^{-2} - 2e^{-2} - 2e^{-2} = 1 - 5e^{-2}.$$

4.5. **Poisson paradigm.** A slightly more general statement is true. If we have N (a large number of) independent events, occuring with probabilities $p_1, \ldots, p_N$, then the quantity of events which happened is approximately $\mathrm{Poi}(\lambda)$ with $\lambda = p_1 + \ldots + p_N$.

So the Poisson distribution is the *law of rare events*: if we have many events, each of which is very unlikely to occur, then the number of events is Poisson.

Example 4.3. A company which sells flood insurance has three groups of clients. The first group, $N_1 = 10000$, is low-risk: each client has probability $p_1 = 0.01\%$ of a flood, independently of others. The second group, $N_2 = 1000$ clients, is medium-risk: $p_2 = 0.05\%$. The third group, $N_3 = 100$ clients, is high-risk, $p_3 = 0.5\%$. For every flood, a company pays $100,000$\$. How much should it charge its clients so that is does not go bankrupt with probability at least 95%?

The quantity X of floods is Poisson with parameter

$$\lambda = N_1p_1 + N_2p_2 + N_3p_3 = 2.$$

Let us find k such that $P(X \le k) \ge 95\%$: after calculation (possibly using the histogram), you can find that $k = 5$. So we will have at most 5 floods with high probability, and the company will pay no more than $500,000$\$ with high probability. This sum should be taken from clients in the form of premuims. But we would like to charge high-risk clients more. Also, the company needs to cover its costs of operation and get some profit, so the real change will be greater than $500,000$\$.

4.6. **Geometric distribution.** Toss a biased coin, with probability p of Heads, and probability $q = 1-p$ of Tails. How many X tosses do you need to get your first Heads? This is called *geometric distribution with parameter* p, and denotes $X \backsim \mathrm{Geo}(p)$, meaning that X has geometric distribution with parameter p. For example, if we have the sequence of tosses TTTHTHH, then $X = 1$. We have:

$$\mathbf{P}(X = n) = \mathbf{P}(\underbrace{TT\ldots T}_{n-1}H) = pq^{n-1}, \quad n = 1, 2, 3, \ldots.$$

We can alternatively describe it as the number of Bernoulli trials you need to get your first success. [1] Let us calculate:

$$\mathbf{E}X = \sum_{n=1}^{\infty} npq^{n-1}.$$

Take a function

$$f(x) = \sum_{n=1}^{\infty} px^n = px(1 + x + x^2 + \ldots) = \frac{px}{1-x}; \quad f'(q) = \sum_{n=1}^{\infty} pnq^{n-1} = \mathbf{E}X.$$

But we can calculate $f'(q)$:

$$f'(x) = \frac{px'(1-x) - p(1-x)'x}{(1-x)^2} = \frac{p(1-x) + px}{(1-x)^2} = \frac{p}{(1-x)^2},$$

and so $f'(q) = p/(1-q)^2 = p/p^2 = 1/p$. This proves that

$$\boxed{\mathbf{E}X = \frac{1}{p}}$$

The meaning of this is commonsensical: if you have success with probability, say, $1/4$, then you need, on average, four trials to get success. Similarly, we can find that

$$\boxed{\mathrm{Var}\,X = \frac{q}{p^2}}$$

[1] Sometimes the geometric distribution is defined as the number of T, rather than tosses, to get your first H. This shifts it by 1. Then it takes values $0, 1, 2, \ldots$ In these lectures, we assume that Geometric distribution is the number of tosses until your first H.

4.7. **Negative binomial distribution.** How many times X do you need to toss a coin to get r Heads, where r is a fixed parameter? This is called *negative binomial distribution* $\mathrm{NB}(r, p)$. For example, if $r = 3$ and the sequence is TTTHTHH, then $X = 7$. What is $\mathbf{P}(X = n)$? If the nth toss resulted in rth Heads, another way to say this is the first $n-1$ tosses contain $r-1$ Heads and $n-r$ Tails, and the last, nth toss resulted in Heads. The probability that the first $n-1$ tosses resulted in $r-1$ Heads (and $n-r$ Tails) is

$$\binom{n-1}{r-1}p^{r-1}q^{n-r}.$$

Indeed, there are

$$\binom{n-1}{r-1}$$

choices for slots in the first $n-1$ tosses occupied by Heads. Each of these choices has probability $p^{r-1}q^{n-r}$. Finally, the probability of the last toss being Heads is p. So

$$\boxed{\mathbf{P}(N = n) = \binom{n-1}{r-1}p^r q^{n-r}}$$

This is true for $n = r, r+1, r+2, \ldots$. The random variable $X \backsim \mathrm{NB}(r, p)$ can take only values greater than or equal to r, because to get r Heads, you need at least r tosses. One can alternatively describe this distribution as the number of Bernoulli trials one needs to get to the rth success. For $r = 1$, this becomes the Geometric distribution with parameter p. [2]

Example 4.4. Let $p = 1/2$, $r = 2$. Then

$$\mathbf{P}(X = 3) = \binom{3-1}{2-1}\left(\frac{1}{2}\right)^2\left(1 - \frac{1}{2}\right) = 2 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{4}.$$

We can equivalently get this as follows: to get two Heads in exactly three tosses, we need either THH or HTH. Each of these sequences has probability $1/8$. So the resulting probability is $(1/8) + (1/8) = 1/4$.

For $X \backsim \mathrm{NB}(r, p)$, we have:

$$X = X_1 + \ldots + X_r, \quad X_1, \ldots, X_r \backsim \mathrm{Geo}(p) \text{ i.i.d.}$$

Recall that i.i.d. means independent identically distributed. Indeed, to get to the rth Heads, we need to get the first Heads, which required geometric number X_1 of tosses; then we need to get from first to second Heads, which required also geometric number X_2 of tosses. This second number is independent of the first one, because the coin does not have a memory, etc. Therefore,

$$\mathbf{E}X = \mathbf{E}X_1 + \ldots + \mathbf{E}X_r = \frac{r}{p},$$

and using independence, we get:

$$\operatorname{Var} X = \operatorname{Var} X_1 + \ldots + \operatorname{Var} X_r = \frac{qr}{p^2}.$$

The negative binomial distribution and the binomial distribution answer somewhat opposite questions. For the binomial distribution, you have fixed number of tries, and you ask how many successful trials you have. For the negative binomial distribution, you have fixed number of successful trials, and you ask how many trials you need.

Problems

Problem 4.1. Find $\mathbf{P}(X \le 4)$ for $X \backsim \mathrm{Bin}(3, 0.7)$.

Problem 4.2. Find $\mathbf{P}(X \le 4)$ for $X \backsim \mathrm{Bin}(5, 0.7)$.

Problem 4.3. Find $\mathbf{P}(X \le 4)$ for $X \backsim \mathrm{Poi}(1)$.

Problem 4.4. In a family with four children, how likely is the even split (two boys, two girls)? Solve this problem if there are 1.05 boys for every girl (real data for USA).

Problem 4.5. For $X \backsim \mathrm{Bin}(10, 0.1)$, calculate $\mathbf{P}(X = 1)$ in a direct way and using Poisson approximation. Compare the results. Calculate the *relative error*, which is $(x - a)/a$, where a is the exact value and x is the approximation.

---

[2] Similarly to the geometric distribution, there is ambiguity in the definition of the negative binomial distribution. Sometimes it is defined as the quantity of T, as opposed to tosses, until your first H. In this case, it is shifted by r down. Thus defined negative binomial random variable takes values $0, 1, 2, \ldots$ We define this distribution as the number of tosses until your rth H.

Problem 4.6. The monthly worldwide number of airplane crashes averages $\lambda = 3$. Assuming that it has a Poisson distribution, find the probability that there will be at least one crash. If it already occurred this month, what is the probability that there will be at least one more?

Problem 4.7. (SOA) A baseball team has scheduled its opening game for April 1. If it rains on April 1, the game is postponed and will be played on the next day that it does not rain. The team purchases insurance against rain. The policy will pay 1000 for each day, up to 2 days, that the opening game is postponed. The insurance company determines that the number of consecutive days of rain beginning on April 1 is a Poisson random variable with mean 0.6. What is the standard deviation of the amount the insurance company will have to pay?

Problem 4.8. (SOA) A study is being conducted in which the health of two independent groups of ten policyholders is being monitored over a one-year period of time. Individual participants in the study drop out before the end of the study with probability 0.2 (independently of the other participants). What is the probability that at least 9 participants complete the study in one of the two groups, but not in both groups?

Problem 4.9. (SOA) A tour operator has a bus that can accommodate 20 tourists. The operator knows that tourists may not show up, so he sells 21 tickets. The probability that an individual tourist will not show up is 0.02, independent of all other tourists. Each ticket costs 50, and is non-refundable if a tourist fails to show up. If a tourist shows up and a seat is not available, the tour operator has to pay 100 (ticket cost + 50 penalty) to the tourist. What is the expected revenue of the tour operator?

Problem 4.10. Find $\mathbf{P}(X \le 2)$ for $X \backsim \mathrm{Geo}(1/2)$.

Problem 4.11. Find $\mathbf{P}(X \le 5)$ for $X \backsim \mathrm{NB}(3, 1/3)$.

Problem 4.12. Find $\mathbf{P}(X \ge 3)$ for $X \backsim \mathrm{NB}(3, 1/3)$.

Problem 4.13. Find $\mathbf{P}(X = 0)$ for $X \backsim \mathrm{Geo}(1/4)$.

Problem 4.14. For a random variable $X \sim \mathrm{NB}(2016, 0.25)$, find $\mathbf{P}(X \le 2017)$.

Problem 4.15. There are $N = 1000$ houses, each is on fire independently of other houses with probability $p = 0.4\%$. Find the probability that there are exactly six fires.

Problem 4.16. The amount of loss in a car accident has geometric distribution with parameter 1/3. An insurance company applies a *deductible* of 2, which means the following. If the amount of loss is less than or equal to 2, then the customer pays out of his pocket and the company pays nothing. If the amount of loss exceeds 2, then the customer pays the first 2 and the company pays the rest. Given that the company has to pay something greater than zero, what is the probability that it pays 1?

Problem 4.17. Throw a die three times and let X be the quantity of even-numbered results: 2, 4 or 6. Throw it three more times and let Y be the quantity of the dice (among the last three dice, not the total of six dice) which are equal to 6. For example, if we have the sequence 364656 of six dice, then $X = 2$ and $Y = 2$. Calculate $\mathrm{Var}(X - Y)$.

Problem 4.18. You take five preliminary exams, and you will pass each exam with probability 1/3, independently of other exams. What is the probability that you will pass two or more exams?

Problem 4.19. Take a sequence of independent Bernoulli trials, each has probability $p = 0.4$ of success. Let X be the number of trials you need to get to your first success. Find $\mathbf{E}(2X + 3)^2$.

Problem 4.20. There are $N = 100$ clients, each will experience flood with probability $p = 0.5\%$, independently of others. What is the probability that exactly one client will experience flood? (Hint: use Poisson approximation.)

Problem 4.21. Suppose you throw a die and say that you have success if you get 5 or 6. You want to get two such successes. What is the probability that you need exactly 10 tries for this? (Which means that your 10th try will bring you the second success.)

Problem 4.22. A company sells car insurance. There are two types of drivers: $N_1 = 5000$ good drivers, and $N_2 = 3000$ bad drivers. This year, each good driver gets into an accident with probability $p_1 = 0.01\%$, and each bad driver gets into an accident with probability $p_2 = 0.03\%$. All accidents can happen independently of one another. Find approximately the probability that there will be at least two accidents this year.

Problem 4.23. Suppose you take a sequence a Bernoulli trials with probability $p = 0.4$ of success. Let X be the number of trials you need to make to get 4 successful trials. Find $\mathbf{E}X$ and $\mathrm{Var}\, X$.

Problem 4.24. Toss six fair coins. Let X be the number of Heads in the first two tosses. Let Y be the number of Tails in the last four tosses. Find $\mathbf{E}(X-Y)$ and $\mathrm{Var}(X-Y)$.

Problem 4.25. An insurance company sells car insurance, with an obligation to pay $100,000$\$ in case of an accident. There are four caterogies of people:

- teenagers: $N_1 = 1500$, $p = 0.05\%$;
- young adults with college education: $N_2 = 2000$, $p = 0.01\%$;
- young adults without college education: $N_3 = 1500$, $p = 0.03\%$;
- middle-aged people: $N_4 = 3000$, $p = 0.02\%$.

The first number refers to the quantity of people in each category. The second number refers to the probability of an accident for a given person in this category. The company wants to pay all the claims using money collected from premiums. It wants to be able to do this with probability 90% or more. What is the amount of money it shoud collect?

Problem 4.26. (SOA) A company buys a policy to insure its revenue in the event of major snowstorms that shut down business. The policy pays nothing for the first such snowstorm of the year and 10000 for each one thereafter, until the end of the year. The number of major snowstorms per year that shut down business is assumed to have a Poisson distribution with mean 1.5. Calculate the expected amount paid to the company under this policy during a one-year period.

Problem 4.27. There are $N_1 = 1000$ high-risk clients, each of whom can have an accident with probability $p_1 = 0.06\%$, and $N_2 = 1000$ low-risk clients, each of whom can have an accident with probability $p_2 = 0.03\%$. All accidents occur independently. Given that the number of accidents is less than 3, find the probability that there are no accidents.

Problem 4.28. A client has losses $X \sim \mathrm{Geo}(0.6)$. The insurance policy has a deductible of 3: if the losses are 3 or less, the company does not pay anything, while if the losses are greater than 3, the company pays 80% of the difference. Find the expected value of the payment.

5. Continuous Random Variables

5.1. **Continuous distributions.** A random variable X has *continuous distribution* with *density* $p(x)$ if for every a and b

$$\mathbf{P}(a \le X \le b) = \int_a^b p(x)\,\mathrm{d}x.$$

Unlike discrete random variables, this one takes any fixed value with probability zero: $\mathbf{P}(X = a) = 0$. The density must satisfy $p(x) \ge 0$ and $\int_{-\infty}^{+\infty} p(x)\,\mathrm{d}x = 1$. We compute the expectation as

$$\mathbf{E}X = \int_{-\infty}^{+\infty} xp(x)\,\mathrm{d}x.$$

Indeed, split the curve $p(x)$ into small sections $[x, x+\mathrm{d}x]$. The probability of X falling into each section is $p(x)\,dx$. This section corresponds to the value $X = x$. Sum these up; since these quantities are infinitesimal, you will obtain the integral above. You can also view it as a continuous analogy of the formula

$$\mathbf{E}X = \sum_x x \cdot \mathbf{P}(X = x)$$

for discrete random variables. Similarly,

$$\boxed{\mathbf{E}f(X) = \int_{-\infty}^{+\infty} f(x)p(x)\,\mathrm{d}x}$$

Example 5.1. Consider the density

$$p(x) = \begin{cases} cx,\ 0 \le x \le 1; \\ 0,\ \text{else.} \end{cases}$$

Let us find the constant $c > 0$: since the density must integrate up to 1, we have:

$$\int_0^1 cx\,\mathrm{d}x = 1 \ \Rightarrow\ c\int_0^1 x\,\mathrm{d}x = \frac{1}{2}c = 1 \ \Rightarrow\ c = 2.$$

Next, we can find expectation and variance:

$$\mathbf{E}X = \int_{-\infty}^{+\infty} xp(x)\,\mathrm{d}x = 2\int_0^1 x^2\,\mathrm{d}x = \frac{2}{3},$$

$$\mathbf{E}X^2 = \int_{-\infty}^{+\infty} x^2p(x)\,\mathrm{d}x = 2\int_0^1 x^3\,\mathrm{d}x = \frac{1}{2},$$

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

5.2. **Examples of continuous distributions.** There are three main classes of continuous distributions which we shall study: Normal (Gaussian), Gamma, and Uniform. Here, we shall provide densities for the Standard Normal (a particular case of Normal), Exponential (a particular case of Gamma), and Uniform.

1. Standard Normal, or Gaussian. $X \backsim \mathcal{N}(0,1)$. It has the "bell curve" density

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

The coefficient $1/\sqrt{2\pi}$ isnecessary to make this density integrate up to 1, because one can show that

$$\int_{-\infty}^{+\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}.$$

2. Exponential. $X \backsim \operatorname{Exp}(\lambda)$ (parameter $\lambda > 0$) if

$$\boxed{p(x) = \begin{cases}\lambda e^{-\lambda x}, & x \ge 0;\\ 0, & x < 0.\end{cases}}$$

We sometimes say: exponential with mean λ^{-1}, because $\mathbf{E}X = 1/\lambda$. The parameter λ is called the *rate.* [3]

3. Uniform on $[c,d]$. $X \backsim \operatorname{Uni}[c,d]$ if

$$\boxed{p(x) = \begin{cases}\frac{1}{d-c}, & c \le x \le d;\\ 0, & \text{for other } x.\end{cases}}$$

5.3. **Expectation and variance of the standard normal distribution.** Assume $X \backsim \mathcal{N}(0,1)$. Then

$$\mathbf{E}X = \int_{-\infty}^{+\infty} xe^{-x^2/2}\,\mathrm{d}x = 0,$$

because the function inside the integral is odd. We can also say that X is symmetric with respect to zero, so $\mathbf{E}X = 0$. Now,

$$\mathbf{E}X^2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^2e^{-x^2/2}\,\mathrm{d}x = 1.$$

Why is this? We know that

$$\int_{-\infty}^{+\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}.$$

Let $u = e^{-x^2/2}$, $v = x$. Integrate by parts: note that $uv = xe^{-x^2/2} = 0$ for $x = \pm\infty$. So

$$\int_{-\infty}^{+\infty} e^{-x^2/2}\,\mathrm{d}x = \int_{-\infty}^{+\infty} u\,\mathrm{d}v = uv|_{x=-\infty}^{x=+\infty} - \int_{-\infty}^{+\infty} v\,\mathrm{d}u$$
$$= -\int_{-\infty}^{+\infty} x\,\mathrm{d}e^{-x^2/2} = -\int_{-\infty}^{+\infty} x\,(-x)e^{-x^2/2}\,\mathrm{d}x = \int_{-\infty}^{+\infty} x^2e^{-x^2/2}\,\mathrm{d}x.$$

This is equal to $\sqrt{2\pi}$, which proves $\mathbf{E}X^2 = 1$. So $\operatorname{Var} X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = 1$. This proves that

$$\boxed{X \sim \mathcal{N}(0,1) \;\Rightarrow\; \mathbf{E}X = 0,\ \operatorname{Var} X = 1}$$

[3] Sometimes the exponential distribution is parametrized using the mean: $\operatorname{Exp}(\beta)$ means the distribution with density on the positive half-line: $p(x) = \beta^{-1}e^{-x/\beta}$, for $x > 0$. Then $\beta = \lambda^{-1} = \mathbf{E}X$. In our lectures, we use the parametrization with rate λ.

5.4. **Normal distribution with general parameters.** If $X \backsim \mathcal{N}(0,1)$, we say that $Y = \mu + \sigma X \backsim \mathcal{N}(\mu, \sigma^2)$. This random variable has $\mathbf{E}Y = \mu + \sigma \mathbf{E}Y = \mu$ and $\operatorname{Var} Y = \sigma^2 \operatorname{Var} X = \sigma^2$. Therefore, it is called *normal random variable with mean μ and variance σ^2*. Its density is given by

$$\boxed{p(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)}$$

Indeed,

$$\mathbf{P}(a \le Y \le b) = \mathbf{P}(a \le \mu + \sigma X \le b) = \mathbf{P}\left(\frac{a-\mu}{\sigma} \le X \le \frac{b-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} e^{-x^2/2}\,\mathrm{d}x.$$

Now let us change variables:

$$x = \frac{y-\mu}{\sigma};\quad \mathrm{d}x = \frac{\mathrm{d}y}{\sigma};\quad a \le y \le b.$$

Therefore,

$$\frac{1}{\sqrt{2\pi}} \int\limits_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} e^{-x^2/2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_a^b e^{-(y-\mu)^2/(2\sigma^2)}\,\mathrm{d}y = \int_a^b p(y)\,\mathrm{d}y.$$

This proves that Y has density $p(y)$, as shown above.

Example 5.2. For $X \sim \mathcal{N}(1,4)$, find $\mathbf{E}\left[(X-2)^3\right]$: $X = 1 + 2Z$ with $Z \sim \mathcal{N}(0,1)$, and $(X-2)^3 = (2Z-1)^3 = 8Z^3 - 12Z^2 + 6Z - 1$. Taking the expected value and using $\mathbf{E}Z = \mathbf{E}Z^3 = 0$ (by symmetry of the standard normal distribution), $\mathbf{E}Z^2 = 1$, we get: $\mathbf{E}\left[(X-2)^3\right] = -12 - 1 = \boxed{-13}$

5.5. **Cumulative distribution function and quantiles.** The *cumulative distribution function* (CDF) of a random variable X is

$$F(x) = \mathbf{P}(X \le x) = \int_{-\infty}^x p(y)\,\mathrm{d}y.$$

For some random variables, such as exponential and uniform, it can be explicitly calculated. For other variables, such as normal, it cannot be calculated in an exact form. But it is tabulated, see the *Z-table* for $X \sim \mathcal{N}(0,1)$ at the end.

Example 5.3. Calculate the CDF of $X \sim \operatorname{Exp}(2)$. For $x \le 0$, we have: $F(x) = 0$, because $X \ge 0$ always. For $x > 0$, we have: $F(x) = \int_0^x 2e^{-2x}\,\mathrm{d}x = \left(-e^{-2x}\right)\big|_{x=0}^{x=\infty} = 1 - e^{-2x}$. Therefore,

$$F(x) = \begin{cases} 0,\ x \le 0; \\ 1 - e^{-2x},\ x \ge 0. \end{cases}$$

Example 5.4. For $X \sim \mathcal{N}(1,4)$, find $\mathbf{P}(X \le 3)$: $X = 1 + 2Z$ with $Z \sim \mathcal{N}(0,1)$, and $\mathbf{P}(X \le 3) = \mathbf{P}(1 + 2Z \le 3) = \mathbf{P}(Z \le 1) = 0.8643$

An α-quantile of a distribution of X is a number x such that $\mathbf{P}(X \le x) = \alpha$. Quantiles of the standard normal distribution are given in the *Z-table*, see the end of these lecture notes. They are denoted as x_α.

Example 5.5. Find x, the 95% quantile of $X \sim \operatorname{Exp}(2)$:

$$\mathbf{P}(X \le x) = 0.95 \ \Leftrightarrow\ 1 - e^{-2x} = 0.95 \ \Leftrightarrow\ e^{-2x} = 0.05 \ \Leftrightarrow\ x = -\frac{1}{2}\ln(0.05) = \boxed{1.50}$$

Example 5.6. Find a normal distribution with mean 1 and 99% quantile 4.5. We can represent this random variable as $X = 1 + \sigma Z$, where σ needs to be found. Because

$$0.99 = \mathbf{P}(X \le 4.5) = \mathbf{P}(1 + \sigma Z \le 4.5) = \mathbf{P}(Z \le \sigma^{-1}3.5),$$

from the Z-table at the end of these lecture notes we get: $\sigma^{-1}3.5 = x_{99\%} = 2.326$, this $\sigma = 0.665$, and $X \sim \boxed{\mathcal{N}(1, 0.665)}$

5.6. **Lognormal distribution.** If $X \backsim \mathcal{N}(\mu, \sigma^2)$ is a normal random variable, then we say that the random variable $Y = e^X$ has *lognormal distribution*. The meaning of this is as follows: a normal variable X models the sum $X_1 + \ldots + X_N$ of many independent random variables (the Central Limit Theorem). And a lognormal random variable $Y = e^X$ models their product $X_1 \ldots X_N$. This is indeed true, because

$$\log(X_1 \ldots X_N) = \log X_1 + \log X_N \approx X,$$

so

$$X_1 \ldots X_N \approx e^X = Y.$$

Sometimes in real life an observed variable is the sum of many small factors (for example, the total vote for each candidate is the sum of individual votes). But in other situations, this observed variable is the *product* of small factors, such as height of a person. In this case, it is a good idea to model this variable by a lognormal distribution. Let us find the density of this lognormal distribution. The density of the normal distribution $\mathcal{N}(\mu, \sigma^2)$ is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Therefore,

$$\mathbf{P}(a \le Y = e^X \le b) = \mathbf{P}(\log a \le X \le \log b) = \int_{\log a}^{\log b} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x$$

Now, let us make a change of variables:

$$x = \log y,\ \mathrm{d}x = \frac{\mathrm{d}y}{y},\quad \log a \le x \le \log b \quad \Rightarrow \quad a \le y \le b.$$

We have:

$$\mathbf{P}(a \le Y \le b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log y-\mu)^2}{2\sigma^2}\right) \mathrm{d}y.$$

Therefore, the density of the lognormal distribution is

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log y-\mu)^2}{2\sigma^2}\right).$$

5.7. **Transformations of continuous random variables.** We can apply similar technique to find density of $Y = f(X)$, where X is a given continuous random variable, and f is a one-to-one function (that is, to every y there is no more than one x such that $f(x) = y$). Best to illustrate this with the following example: let $X \sim \mathrm{Exp}(2)$, and $f(x) = x^2$. Then for the random variable $Y = f(X) = X^2$, for any positive a and b, we have:

$$\mathbf{P}(a \le Y \le b) = \mathbf{P}\left(a^{1/2} \le X \le b^{1/2}\right) = \int_{a^{1/2}}^{b^{1/2}} 2e^{-2x}\, \mathrm{d}x.$$

Now make change of variable $x = y^{1/2}$, then

$$a^{1/2} \le x = y^{1/2} \le b^{1/2} \ \Rightarrow\ a \le y \le b,\ \mathrm{d}x = \frac{1}{2}y^{-1/2}\, \mathrm{d}y.$$

Then the integral takes the form

$$\int_a^b 2e^{-2y^{1/2}} \frac{1}{2} y^{-1/2}\, \mathrm{d}y = \int_a^b p(y)\, \mathrm{d}y,\ \text{where } p(y) := y^{-1/2} e^{-2y}.$$

Therefore, we represented the probability

$$\mathbf{P}(a \le Y \le b) = \int_a^b p(y)\, \mathrm{d}y.$$

Thus, $p(y)$ is the density of Y.

5.8. **Properties of exponential distribution.** Consider $X \backsim \mathrm{Exp}(\lambda)$, exponential distribution with parameter $\lambda > 0$. It has density

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, \ x \ge 0; \\ 0, \ x < 0. \end{cases}$$

We sometimes say: exponential with mean λ^{-1}, because after calculations:

$$\mathbf{E}X = \int_0^\infty \lambda x e^{-\lambda x}\,\mathrm{d}x = \int_0^\infty x\mathrm{d}\left(-e^{-\lambda x}\right) =$$

$$= x \cdot \left(-e^{-\lambda x}\right)\Big|_{x=0}^{x=\infty} - \int_0^\infty \left(-e^{-\lambda x}\right)\mathrm{d}x = 0 - 0 + \int_0^\infty e^{-\lambda x}\mathrm{d}x = -\frac{1}{\lambda}\int_0^\infty \mathrm{d}\left(-e^{-\lambda x}\right) =$$

$$-\frac{1}{\lambda}\left(-e^{-0} - (-e^{-\infty})\right) = \frac{1}{\lambda}.$$

We also have: for $a > 0$,

$$\mathbf{P}(X \ge a) = \int_a^\infty \lambda e^{-\lambda x}\,dx = \left(-e^{-\lambda x}\right)\Big|_{x=a}^{x=\infty} = e^{-\lambda a}.$$

The exponential distribution has *memoryless property*: for $s, t > 0$,

$$\mathbf{P}(X > t + s \mid X > s) = \mathbf{P}(X > t).$$

Indeed,

$$\mathbf{P}(X > t + s \mid X > s) = \frac{\mathbf{P}(X > t + s, X > s)}{\mathbf{P}(X > s)} = \frac{\mathbf{P}(X > t + s)}{\mathbf{P}(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbf{P}(X > t).$$

Sometimes exponential distribution is used to model the lifespan of a device, when the remaining lifespan is independent of how long it has worked so far.

5.9. **Power laws.** This is density $p(x) = cx^{-\alpha-1}$ for $x \ge x_0$. It integrates to one, therefore

$$\int_{x_0}^\infty p(x)\,\mathrm{d}x = 1 \ \Rightarrow \ \frac{cx_0^{-\alpha}}{\alpha} = 1.$$

This is sometimes called *Pareto law*, or *heavy tails*: because probabilities $\mathbf{P}(X \ge x)$ for large x are much larger than for exponential or normal distributions with same mean and variance.

Example 5.7. $p(x) = 0.5x^{-1.5}$, $x \ge 1$. Then the mean is infinite:

$$\mathbf{E}X = \int_1^\infty xp(x)\,\mathrm{d}x = 0.5\int_1^\infty x^{-0.5}\,\mathrm{d}x = c\left.\frac{x^{0.5}}{0.5}\right|_{x=1}^{x=\infty} = \infty.$$

Example 5.8. $p(x) = cx^{-2.5}$, $x \ge 2$. Here, c can be found from the total integral:

$$1 = \int_2^\infty cx^{-2.5}\,\mathrm{d}x = c\left.\frac{x^{-1.5}}{-1.5}\right|_{x=2}^{\infty} = c\frac{2^{-1.5}}{1.5} \ \Rightarrow \ c = 1.5 \cdot 2^{1.5} = 3\sqrt{2}.$$

Then the mean is finite:

$$\mathbf{E}X = \int_2^\infty xp(x)\,\mathrm{d}x = c\int_2^\infty x^{-1.5}\,\mathrm{d}x = c\left.\frac{x^{-0.5}}{-0.5}\right|_{x=2}^{x=\infty} = c\frac{2^{-0.5}}{0.5} = 6,$$

but the second moment (and therefore the variance) is infinite:

$$\mathbf{E}X^2 = \int_2^\infty x^2p(x)\,\mathrm{d}x = c\int_2^\infty x^{-0.5}\,\mathrm{d}x = c\left.\frac{x^{0.5}}{0.5}\right|_{x=2}^{\infty} = \infty.$$

5.10. **Independence for continuous distributions.** Two random variables X and Y are called *independent* if for every subsets A and B we have:

$$\mathbf{P}(X \in A,\ Y \in B) = \mathbf{P}(X \in A) \cdot \mathbf{P}(Y \in B).$$

We cannot formulate the definition as

$$\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x) \cdot \mathbf{P}(Y = y),$$

because the probability that X takes a given value is zero! For continuous random variables, expectation and variance have the same properties as for discrete ones. For all random variables X and Y,

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y,\ \operatorname{Cov}(X, Y) = \mathbf{E}(XY) - (\mathbf{E}X)(\mathbf{E}Y),$$

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

For independent random variables X and Y,

$$\mathbf{E}(XY) = (\mathbf{E}X)(\mathbf{E}Y),\ \operatorname{Cov}(X, Y) = 0,\ \operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y.$$

5.11. **Mixtures.** It is not true that all distributions are either discrete or continuous. Assume a client has an accident with probability 10%, and the losses are distributed as $X \sim \operatorname{Exp}(2)$. We have:

$$\mathbf{E}X = 0.5,\ \operatorname{Var} X = 0.25 \ \Rightarrow\ \mathbf{E}X^2 = (\mathbf{E}X)^2 + \operatorname{Var} X = 0.5.$$

Then the amount Z of losses is a random variable which has distribution

$$\mathbf{P}(Z = 0) = .9,\ \mathbf{P}(a < Z < b) = .1 \cdot \int_a^b 2e^{-2x}\,\mathrm{d}x.$$

This is a *mixture* of two distributions, one concentrated at 0 and the other exponential, with *weights* 0.9 and 0.1 respectively. Then

$$\mathbf{E}Z = .9 \cdot 0 + .1 \cdot \mathbf{E}X = .05, \mathbf{E}Z^2 = .9 \cdot 0^2 + .1 \cdot \mathbf{E}X^2 = 0.05.$$

Therefore, $\operatorname{Var} Z = \mathbf{E}Z^2 - (\mathbf{E}Z)^2 = 0.05 - 0.05^2$. We can also have mixtures of continuous distributions (which are themselves continuous).

Example 5.9. Take a random variable Z, which is a mixture of $X \sim \operatorname{Exp}(0.5)$ and $Y \sim \mathcal{N}(2, 1)$ with weights 75% and 25% respectively.[4] Recall that

$$\mathbf{E}X = \frac{1}{0.5} = 2,\ \mathbf{E}Y = 2,\ \mathbf{E}X^2 = \operatorname{Var} X + (\mathbf{E}X)^2 = \frac{1}{0.5^2} + \left(\frac{1}{0.5}\right)^2 = 8,\ \mathbf{E}Y^2 = \operatorname{Var} Y + (\mathbf{E}Y)^2 = 5.$$

The random variable Z has density

$$p_Z(x) = 0.75p_X(x) + 0.25p_Y(x) = \begin{cases} 0.75 \cdot 0.5e^{-0.5x} + 0.25 \cdot \frac{1}{\sqrt{2\pi}}e^{-(x-1)^2/2},\ x \ge 0; \\ 0.25 \cdot \frac{1}{\sqrt{2\pi}}e^{-(x-1)^2/2},\ x < 0. \end{cases}$$

Its expectation and variance are

$$\mathbf{E}Z = 0.75\mathbf{E}X + 0.25\mathbf{E}Y = 0.75 \cdot 2 + 0.25 \cdot 2 = \boxed{2}$$

$$\mathbf{E}Z^2 = 0.75\mathbf{E}X^2 + 0.25\mathbf{E}Y^2 = 0.75 \cdot 8 + 0.25 \cdot 5 = \boxed{7.25}$$

and therefore $\operatorname{Var} Z = \mathbf{E}Z^2 - (\mathbf{E}Z)^2 = 7.25 - 2^2 = \boxed{3.25}$

[4] Do not confuse Z with $W = 0.75 \cdot X + 0.25 \cdot Y$, which is *not* a mixture of distributions.

5.12. **Joint density.** Random variables X and Y have joint density $p(x, y)$ if for every subset $C \subseteq \mathbb{R}^2$, we have:

$$\mathbf{P}((X,Y) \in C) = \iint_C p(x,y)\,\mathrm{d}A.$$

For example,

$$\mathbf{P}(X \in A, Y \in B) = \int_A \int_B p(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

If X and Y are independent, then the joint density

$$\boxed{p(x,y) = p_X(x)p_Y(y)}$$

is the product of densities p_X of X and p_Y of Y. In the general case (if they are not independent), the marginal density of X can be reconstructed as follows:

$$p_X(x) = \int_{-\infty}^{+\infty} p(x,y)\,\mathrm{d}y.$$

Indeed, because for all a and b we have:

$$\mathbf{P}(a \le X \le b) = \mathbf{P}(a \le X \le b, -\infty < Y < +\infty) = \int_a^b \int_{-\infty}^{+\infty} p(x,y)\,\mathrm{d}y\,\mathrm{d}x,$$

then the expression $\int_{-\infty}^{+\infty} p(x,y)\,\mathrm{d}y$ plays the role of density for X. Similarly,

$$p_Y(y) = \int_{-\infty}^{+\infty} p(x,y)\,\mathrm{d}x$$

is the marginal density of Y. The expectation of X is given by

$$\mathbf{E}X = \int x p_X(x)\mathrm{d}x = \int\int x p(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

and the expectation of $f(X,Y)$ is given by

$$\boxed{\mathbf{E}f(X,Y) = \iint f(x,y)p(x,y)\,\mathrm{d}x\,\mathrm{d}y}$$

Example 5.10. Suppose $p(x,y) = x + y$, if $0 \le x, y \le 1$, and 0 otherwise. Then

$$p_X(x) = \int_0^1 (x+y)\mathrm{d}y = \left(xy + \frac{y^2}{2}\right)\Big|_{y=0}^{y=1} = x + \frac{1}{2}.$$

$$\mathbf{E}Y = \int_0^1 \int_0^1 (x+y)y\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \int_0^1 (xy + y^2)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \left(\frac{xy^2}{2} + \frac{y^3}{3}\right)\Big|_{y=0}^{y=1} \mathrm{d}x$$
$$= \int_0^1 \left(\frac{1}{2}x + \frac{1}{3}\right)\mathrm{d}x = \frac{1}{2}\,\frac{x^2}{2}\Big|_{x=0}^{x=1} + \frac{1}{3}\,x|_{x=0}^{x=1} = \frac{1}{4} + \frac{1}{3} = \boxed{\frac{7}{12}}$$

$$\mathbf{E}(XY) = \int_0^1 \int_0^1 (xy)(x+y)\mathrm{d}x\mathrm{d}y = \ldots = \text{calculate}$$

Example 5.11. Suppose $p(x,y) = 2e^{-x-2y}$, $x, y > 0$. Then this joint density can be represented in the product form:

$$p(x,y) = p_X(x)p_Y(y),\ p_X(x) = e^{-x},\ p_Y(y) = 2e^{-2y}.$$

Therefore, $X \sim \mathrm{Exp}(1)$ and $Y \sim \mathrm{Exp}(2)$ are independent.

Example 5.12. Consider two continuous random variables (X,Y) distributed uniformly on the triangle $\{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ 0 \le y \le x\}$. That is, their joint density is equal to

$$p(x,y) = \begin{cases} 2,\ 0 \le x \le 1,\ 0 \le y \le x; \\ 0,\ \text{else}. \end{cases}$$

Let us find $\mathbf{E}Y^2$. We have:

$$\mathbf{E}Y^2 = \int_0^1 \int_0^x 2y^2\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 2\,\frac{y^3}{3}\Big|_{y=0}^{y=x}\mathrm{d}x = \int_0^1 \frac{2}{3}x^3\,\mathrm{d}x = \frac{2}{3}\cdot\frac{x^4}{4}\Big|_{x=0}^{x=1} = \boxed{\frac{1}{6}}$$

5.13. **Convolution.** Suppose X and Y are independent discrete random variables, taking values $0, 1, 2, \ldots$. Then the distribution of $X+Y$ can be calculated as follows:

$$\mathbf{P}(X+Y=n) = \sum_{k=0}^{n} \mathbf{P}(X=k)\mathbf{P}(Y=n-k),\ n = 0, 1, 2, \ldots$$

If X and Y take values $0, \pm 1, \pm 2, \ldots$ (negative as well as positive), then

$$\mathbf{P}(X+Y=n) = \sum_{k=-\infty}^{\infty} \mathbf{P}(X=k)\mathbf{P}(Y=n-k),\ n = 0, \pm 1, \ldots$$

Example 5.13. Take $X \sim \mathrm{Poi}(\lambda)$ and $Y \sim \mathrm{Poi}(\mu)$. Then

$$\mathbf{P}(X+Y=n) = \sum_{k=0}^{n} \frac{\lambda^k}{k!}e^{-\lambda}\frac{\mu^{n-k}}{(n-k)!}e^{-\mu} = \frac{e^{-\lambda-\mu}}{n!}\sum_{k=0}^{n}\lambda^k\mu^{n-k}\frac{n!}{k!(n-k)!} = \frac{e^{-\lambda-\mu}}{n!}(\lambda+\mu)^n.$$

Therefore, $X+Y \sim \mathrm{Poi}(\lambda+\mu)$.

For continuous random variables X and Y, if they are nonnegative, with densities p_X and p_Y, we have: the density of $X+Y$ is given by

$$p_{X+Y}(x) = \int_0^x p_X(y)p_Y(x-y)\,\mathrm{d}y,\ x \ge 0.$$

If X and Y are not necessarily nonnegative, then

$$p_{X+Y}(x) = \int_{-\infty}^{\infty} p_X(y)p_Y(x-y)\,\mathrm{d}y.$$

This is called *convolution* of p_X and p_Y.

5.14. **Gamma distribution.** Take $X, Y \sim \mathrm{Exp}(\lambda)$. Then $p_X(x) = p_Y(x) = \lambda e^{-\lambda x}$, and

$$p_{X+Y}(x) = \int_0^x \lambda e^{-\lambda y}\cdot\lambda e^{-\lambda(x-y)}\mathrm{d}y = \lambda^2\int_0^x e^{-\lambda x}\mathrm{d}y = \lambda^2 x e^{-\lambda x},\ x \ge 0.$$

Next, $X, Y, Z \sim \mathrm{Exp}(\lambda)$, then

$$p_{X+Y+Z}(x) = \int_0^x p_{X+Y}(y)p_Z(x-y)\mathrm{d}y = \int_0^x \lambda^2 y e^{-\lambda y}\cdot\lambda e^{-\lambda(x-y)}\mathrm{d}y = \lambda^3 e^{-\lambda x}\int_0^x y\mathrm{d}y = \lambda^3\frac{x^2}{2}e^{-\lambda x}.$$

More generally, the sum of n i.i.d. $\mathrm{Exp}(\lambda)$ random variables has density

$$p(x) = \frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}.$$

This distribution is called *Gamma distribution* $\Gamma(n, \lambda)$. For $n = 1$, we are back to $\mathrm{Exp}(\lambda)$. [5] We changed variables $y = \lambda x$, so that $\lambda^\alpha x^{\alpha-1}\mathrm{d}x = y^{\alpha-1}\mathrm{d}y$. Let us find expectation and variance of $X \sim \Gamma(n, \lambda)$: we can represent X as

$$X = X_1 + \ldots + X_n,\ X_i \sim \mathrm{Exp}(\lambda) \text{ i.i.d.}$$

Therefore, because $\mathbf{E}X_i = \lambda^{-1}$ and $\mathrm{Var}\,X_i = \lambda^{-2}$, we have:

$$\mathbf{E}X = \mathbf{E}X_1 + \ldots + \mathbf{E}X_n = \frac{n}{\lambda},\ \mathrm{Var}\,X = \mathrm{Var}\,X_1 + \ldots + \mathrm{Var}\,X_n = \frac{n}{\lambda^2}. \tag{3}$$

Example 5.14. Find the density of $X \sim \Gamma(4, 3)$, as well as $\mathbf{E}X$ and $\mathbf{E}X^2$. The density is given by

$$p(x) = \frac{3^4}{(4-1)!}x^{4-1}e^{-3x} = \frac{3^4}{6}x^3e^{-3x} = \frac{27}{2}x^3e^{-3x}.$$

$$\mathbf{E}X = \frac{3}{4},\ \mathrm{Var}\,X = \frac{3}{4^2} \ \Rightarrow\ \mathbf{E}X^2 = (\mathbf{E}X)^2 + \mathrm{Var}\,X = \frac{9+3}{16} = \frac{3}{4}.$$

[5] Similarly to the exponential distribution, sometimes the Gamma distribution is parametrized via $\beta := \lambda^{-1}$. We use λ.

5.15. **Gamma distribution for non-integer parameter.** We can generalize it for non-integer n as follows. Define the notion of generalized factorial, or Gamma function. Note that

$$n! = I_n := \int_0^{\infty} x^n e^{-x}\mathrm{d}x.$$

Why? Integrate I_n by parts:

$$I_n = -\int_0^{\infty} x^n \mathrm{d}e^{-x} = -\left. x^n e^{-x}\right|_{x=0}^{x=\infty} + \int_0^{\infty} e^{-x}\mathrm{d}\left(x^n\right) = 0 + n\int_0^{\infty} x^{n-1}e^{-x}\mathrm{d}x = nI_{n-1}. \tag{4}$$

Combining that with $I_0 = \int_0^{\infty} e^{-x}\mathrm{d}x = 1$, we have: $I_n = n!$. Thus, define for $\alpha > 0$:

$$\alpha! := \int_0^{\infty} x^{\alpha} e^{-x}\mathrm{d}x.$$

Usually, we talk about the *Gamma function*, defined for all $\alpha > 0$:

$$\Gamma(\alpha) = (\alpha - 1)! = \int_0^{\infty} x^{\alpha-1}e^{-x}\mathrm{d}x.$$

Similarly to (4), we get

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha),\quad \alpha > 0. \tag{5}$$

Example 5.15. Let us find $\Gamma(1/2) = (-1/2)!$: we can change variables $x = u^2/2$ in the following integral

$$\Gamma(1/2) = \int_0^{\infty} x^{-1/2}e^{-x}\,\mathrm{d}x. \tag{6}$$

Then we have: $\mathrm{d}x = u\,\mathrm{d}u$, and $x^{-1/2}\,\mathrm{d}x = \sqrt{2}u^{-1}\cdot u\,\mathrm{d}u = \sqrt{2}\,\mathrm{d}u$. Thus, the integral in (6) can be written as

$$\sqrt{2}\int_0^{\infty} e^{-u^2/2}\,\mathrm{d}u = \sqrt{2}\sqrt{2\pi}\int_0^{\infty}\frac{e^{-u^2/2}}{\sqrt{2\pi}}\,\mathrm{d}u = 2\sqrt{\pi}\cdot\frac{1}{2} = \boxed{\sqrt{\pi}}$$

Example 5.16. Find $\Gamma(5/2)$: From (5) we get that

$$\Gamma(5/2) = (3/2)\Gamma(3/2) = (3/2)(1/2)\Gamma(1/2) = \boxed{\frac{3}{4}\sqrt{\pi}}$$

Define the Gamma distribution $\Gamma(\alpha, \lambda)$:

$$p(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x},\quad x > 0.$$

For $\alpha = 1$, we are back to $\mathrm{Exp}(\lambda)$. This is indeed a probability distribution, because it integrates to 1:

$$\int_0^{\infty} p(x)\mathrm{d}x = \frac{1}{\Gamma(\alpha)}\int_0^{\infty} y^{\alpha-1}e^{-y}\mathrm{d}y = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1.$$

We changed variables $y = \lambda x$, so that $\lambda^{\alpha}x^{\alpha-1}\mathrm{d}x = y^{\alpha-1}\mathrm{d}y$. Note that the expectation and variance are given by the same formulas as in (3), with n changed to α. We shall show this in Section 8.

Example 5.17. Find the density of the Gamma distribution with mean 2 and variance 4. We have:

$$\frac{\alpha}{\lambda} = 5,\quad \frac{\alpha}{\lambda^2} = 10.$$

Dividing the first equation over the second, we get: $\lambda = 0.5$, and $\alpha = 5\lambda = 2.5$. Therefore, the density is given by

$$\frac{0.5^{2.5}}{\Gamma(2.5)}x^{1.5}e^{-0.5x} = \frac{0.5^{2.5}}{3\sqrt{\pi}/4}x^{1.5}e^{-0.5x} = \boxed{3\sqrt{\frac{2}{\pi}}x^{1.5}e^{-x/2}}$$

5.16. **Service life.** Let us model the distribution of the remaining time of service of a certain device. Let T be its (random) lifetime. Assume this is a continuous random variable, with density $p(t)$ and cumulative distribution function $F(t) = \mathbf{P}(T \le t) = \int_0^t p(s)\,\mathrm{d}s$. Then $S(t) = \mathbf{P}(T \ge t) = 1 - F(t)$ is called *survival function*, and

$$f(t) = \frac{p(t)}{S(t)} = \frac{F'(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -(\ln S(t))'$$

is called the *failure rate*. If a device still works by time t, that is $T \ge t$, then for small $\mathrm{d}t$ we have:

$$\mathbf{P}(T \le t + \mathrm{d}t \mid T \ge t) = \frac{\mathbf{P}(t \le T \le t + \mathrm{d}t)}{\mathbf{P}(T \ge t)} = \frac{p(t)\,\mathrm{d}t}{S(t)} = f(t)\,\mathrm{d}t.$$

That is, among N such devices by time t, approximately $Nf(t)\,\mathrm{d}t$ will fail during time $[t, t + \mathrm{d}t]$. The simplest model is with constant failure rate $f(t) = \lambda$. This gives us exponential distribution:

$$-(\ln S(t))' = \lambda \;\Rightarrow\; \ln S(t) = \ln S(0) - \lambda t = -\lambda t \;\Rightarrow\; S(t) = e^{-\lambda t}.$$

We used $S(0) = 1$: all devices still work at time $t = 0$. Other models take into account rising failure rate: say, $f(t) = A + Be^{-Ct}$ or $f(t) = A + Bt^C$ for some constants $A, B, C > 0$.

Problems

Problem 5.1. (SOA) The loss due to a fire in a commercial building is modeled by a random variable X with density function

$$f(x) = \begin{cases} 0.005(20 - x), & 0 < x < 20; \\ 0, & \text{otherwise.} \end{cases}$$

Given that a fire loss exceeds 8, what is the probability that it exceeds 16?

Problem 5.2. (SOA) In a small metropolitan area, annual losses due to storm, fire, and theft are assumed to be independent, exponentially distributed random variables with respective means 1.0, 1.5, 2.4. Determine the probability that the maximum of these losses exceeds 3.

Problem 5.3. For $X \sim \mathcal{N}(-2, 2)$, find $\mathbf{E}X^3$.

Problem 5.4. For $X \sim \mathcal{N}(3, 4)$, find $\mathbf{E}\left[(X^2 - 1)^2\right]$.

Problem 5.5. For $X \sim \mathcal{N}(-1, 3)$, find $\mathbf{E}\left[(X + 1)^3\right]$.

Problem 5.6. For $X \sim \mathcal{N}(-2, 2)$, find $\mathbf{P}(-3 \le X \le 0)$.

Problem 5.7. For $X \sim \mathcal{N}(3, 4)$, find $\mathbf{P}(X \ge 4 \mid X \ge 3)$.

Problem 5.8. Find a normal distribution with mean -1 and 95% quantile equal to 2.3

Problem 5.9. Find a normal distribution with 1% quantile at -3.4 and 99% quantile at 2.4.

Problem 5.10. Find a normal distribution with standard deviation 2.5 and 5% quantile at -15.2.

Problem 5.11. Find an exponential distribution such that $\mathbf{P}(Z \ge 3) = 0.4$.

Problem 5.12. Find the cumulative distribution function of $Z - 3$, where $Z \sim \mathrm{Exp}(0.5)$.

Problem 5.13. Find an exponential distribution with $\mathbf{P}(X \ge 4 \mid X \ge 3) = 0.5$.

Problem 5.14. Find the parameters of the gamma distribution with mean 4 and variance 4/3.

Problem 5.15. (SOA) An actuary models the lifetime of a device using the random variable $Y = 10X^{0.8}$, where X is an exponential random variable with mean 1 year. Determine the probability density function $f(y)$, for $y > 0$, of the random variable Y.

Problem 5.16. (SOA) A device runs until either of two components fails, at which point the device stops running. The joint density function of the lifetimes of the two components, both measured in hours, is

$$p(x, y) = \frac{1}{8}\left(x + y\right), \quad 0 \le x \le 2 \;\text{ and }\; 0 \le y \le 2.$$

Calculate the probability that the device fails during its first hour of operation.

Problem 5.17. For a random variable X with density ce^x, $0 \le x \le 2$, find the constant c and find $\mathbf{E}e^X$.

Problem 5.18. An insurance company insures a large number of homes. The insured value X of a randomly selected home follows a distribution with density function

$$f(x) = \begin{cases} 2x^{-3}, & x \ge 1; \\ 0, & \text{otherwise.} \end{cases}$$

Given that a home is insured for at least 2, find the probability it is insured for less than 3.

Problem 5.19. Calculate $\mathbf{E}X^3$ and $\mathbf{E}X^4$ for $X \backsim \mathcal{N}(0,1)$. For $\mathbf{E}X^4$, use the same trick as for $\mathbf{E}X^2$ in the lecture: integration of $x^2e^{-x^2/2}$ by parts, $u = x^2e^{-x^2/2}$, $v = x$. (See lecture notes, Section 5.)

Problem 5.20. For a random variable $X \sim \mathrm{Uni}[0,1]$, find the density of $\log X$.

Problem 5.21. For a random variable $X \sim \mathrm{Exp}(3)$, find the density of $2X^2$.

Problem 5.22. (SOA) Two random variables X and Y have joint density

$$\begin{cases} \frac{1}{21}(x+y), & 1 \le x \le 3, \quad 0 \le y \le 3; \\ 0, \text{otherwise.} \end{cases}$$

Find the probability $\mathbf{P}(X \ge 2,\ Y \le 2)$.

Problem 5.23. Calculate $\mathbf{E}X$ and $\operatorname{Var} X$ for $X \backsim \mathrm{Uni}[c,d]$.

Problem 5.24. Let $X \sim \mathcal{N}(-1,4)$, $Y \sim \Gamma(5,2)$ be independent random variables. Find $\operatorname{Var}(Y - 2X - 3)$.

Problem 5.25. For X with density $p(x) = 3x^2,\ 0 \le x \le 1$, find $\mathbf{E}X^3$.

Problem 5.26. Find coefficients a and b such that for independent $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(1,2)$ we have: $aX + bY \sim \mathcal{N}(3,21)$.

Problem 5.27. Find $\mathbf{E}Y^2$ for two continuous random variables (X,Y) distributed uniformly on the triangle $\{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ 0 \le y \le x\}$. That is, their joint density is equal to

$$p(x,y) = \begin{cases} 2, & 0 \le x \le 1,\ 0 \le y \le x; \\ 0, & \text{else.} \end{cases}$$

Problem 5.28. (SOA) Calculate the expected value for a continuous random variable with density function:

$$f(x) = \begin{cases} |x|/10, & -2 \le x \le 4; \\ 0, & \text{otherwise.} \end{cases}$$

Problem 5.29. (SOA) A manufacturer's annual losses follow a distribution with density function

$$f(x) = 2.5(0.6)^{2.5}x^{-3.5}, \quad x > 0.6,$$

and $f(x) = 0$ otherwise. To cover its losses, the manufacturer purchases an insurance policy with an annual deductible of 2. Calculate the mean of the manufacturer's annual losses not paid by the insurance policy.

Problem 5.30. (SOA) An insurance policy covers a loss $X \sim \mathrm{Uni}[0,1000]$. Calculate the deductible d such that the expected payment under the policy with deductible is 25% of what it would be with no deductible.

Problem 5.31. (SOA) Let X and Y be continuous random variables with joint density function

$$f(x,y) = \frac{8}{3}xy, \quad 0 \le x \le 1, \quad x \le y \le 2x,$$

and $f(x,y) = 0$ otherwise. Calculate $\operatorname{Cov}(X,Y)$.

Problem 5.32. Two-thirds of auto insurance clients are teenagers, who have losses distributed exponentially with mean 5, and the other third are middle-aged people, who have losses distributed exponentially with mean 1. Find the density, mean, and variance of the distribution of losses for a random client.

For the next four problems, consider a client has an accident with probability 20%, and the losses are distributed exponentially with mean 10. Let X be the total amount of losses.

Problem 5.33. Find $\operatorname{Var} X$.

Problem 5.34. Find $\mathbf{E}X^3$.

Problem 5.35. Find $\mathbf{P}(X \le 2)$.

Problem 5.36. Find $\mathbf{P}(1 < X < 3)$.

For the next three problems, let X have a mixture of two distributions: $\mathrm{Bin}(4, .4)$ and $\mathcal{N}(0, 2)$, with weights 0.3 and 0.7.

Problem 5.37. Find $\mathbf{E}X$ and $\mathrm{Var}\, X$.

Problem 5.38. Find $\mathbf{P}(2 \le X \le 3)$.

Problem 5.39. Find $\mathbf{P}(X = 0 \mid X \le 2)$.

Problem 5.40. For independent $X \sim \mathcal{N}(3, 2)$, $Y \sim \mathcal{N}(-1, 4)$, $Z \sim \mathcal{N}(0, 1)$, find $\mathbf{P}(-X + 2Y - Z \le 3)$.

Problem 5.41. The failure rate for a device is $f(t) = 1 + 2t$. If it is time 2 and it is still serving, what is the probability it will serve at least 3 more units of tme?

Problem 5.42. A device has two parts, and it works as long as both parts work. Each part has constant failure rate, 2 for the first part and 3 for the second part. Find the service time distribution and failure rate for the whole device.

Problem 5.43. The failure rate for a device is $f(t) = 3 + 2t^{5/2}$. Write down, but do not evaluate, the integral for the mean service time.

Problem 5.44. We know that $\Gamma(\alpha, \beta)$ has mean 2 and standard deviation $\sqrt{5}$. Find α, β.

Problem 5.45. Find the density of $3X^4$ if X is exponential with mean 2.

Problem 5.46. Find the density of $X + Y$, where $X \sim \Gamma(3, 2)$ and $Y \sim \Gamma(5, 2)$ are independent.

For the next five problems, consider (X, Y) with density $p(x, y) = 24xy$, $0 \le x, y$; $x + y \le 1$.

Problem 5.47. Find $\mathbf{E}X$.

Problem 5.48. Find $\mathrm{Var}\, Y$.

Problem 5.49. Find $\mathrm{corr}(X, Y)$.

Problem 5.50. Find $\mathbf{E}(X^2 Y)$.

Problem 5.51. Find $\mathbf{P}(X \le 0.5 \mid Y \ge 0.5)$.

Problem 5.52. For X with $p(x) = x^{-2}$, $x \ge 1$, find $\mathbf{E}X$ and $\mathbf{P}(X \le 2)$.

Problem 5.53. For X with $p(x) = 2x^{-3}$, $x \ge 1$, find $\mathbf{E}X$ and $\mathrm{Var}\, X$.

Problem 5.54. For x with $p(x) = cx^{-3.5}$, $x \ge 3$, find c, $\mathbf{E}X$, $\mathrm{Var}\, X$.

6. Central Limit Theorem

6.1. **Statement for IID variables.** Suppose $X_1, X_2, X_3, \ldots$ are iid (independent identically distributed) random variables with mean $\mathbf{E}X_1 = \mathbf{E}X_2 = \ldots = \mu$ and variance $\mathrm{Var}\, X_1 = \mathrm{Var}\, X_2 = \ldots = \sigma^2$. Consider the sum

$$S_N := X_1 + \ldots + X_N.$$

Note that

$$\mathbf{E}S_N = \mathbf{E}X_1 + \ldots + \mathbf{E}X_N = \mu N, \quad \mathrm{Var}\, S_N = \mathrm{Var}\, X_1 + \ldots + \mathrm{Var}\, X_N = \sigma^2 N.$$

The last equation used that $X_1, X_2, \ldots$ are independent. Therefore, for every N the random variable

$$\frac{S_N - \mathbf{E}S_N}{\sqrt{\mathrm{Var}\, S_N}} = \frac{S_N - \mu N}{\sigma\sqrt{N}}$$

has mean zero and variance 1. As $N \to \infty$, we have:

$$\boxed{\frac{S_N - \mu N}{\sigma\sqrt{N}} \to \mathcal{N}(0, 1)}$$

Here $\mathcal{N}(0, 1)$ is the *standard normal distribution.* A random variable X is distributed according to $\mathcal{N}(0, 1)$: $X \backsim \mathcal{N}(0, 1)$, if

$$\mathbf{P}(a \le X \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx.$$

This convergence means that, as $N \to \infty$,

$$\mathbf{P}\left(a \le \frac{S_N - \mu N}{\sigma\sqrt{N}} \le b\right) \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,dx,$$

for every real numbers a, b. We cannot compute the antiderivative $\int e^{-x^2/2}\,dx$ as an elementary function. So there are tables of the normal distribution, in print and online (just Google them). If X_k has Bernoulli distribution:

$$\mathbf{P}(X_k = 1) = p, \quad \mathbf{P}(X_k = 0) = q, \quad p + q = 1,$$

then $\mu = \mathbf{E}X_k = 1 \cdot p + 0 \cdot q = p$, and (after calculation) $\operatorname{Var} X_k = \sigma^2 = pq$.

Example 6.1. Suppose we have a plane of $N = 200$ people. Each customer, independently of others, chooses chicken or pasta, with probability 50%. How many chicken and pasta is needed to satisfy chicken-lovers with probability 95% and pasta-lovers with probability 95% (so that everybody will be satisfied with probability 90%)?

If we prepare only 100 chicken and 100 pasta, we will likely run out ofeither chicken or pasta: it is very unlikely that *exactly* 100 people will choose chicken. The other extreme is to prepare 200 chicken and 200 pasta. This guarantees that every client will be pleased, but a lot of food (200 meals) will be thrown away. Which is not good, because airlines have tight budgets. Let us find a compromise. Let $X_k = 1$ if the kth person chooses chicken, $X_k = 0$ if he chooses pasta. Then $S_N = X_1 + \ldots + X_N$ is the total number of chicken required. The random variable X_k is Bernoulli, with $p = 1/2$, so

$$\mu = \frac{1}{2}, \quad \sigma^2 = \frac{1}{4}, \quad \sigma = \frac{1}{2};$$

$$\frac{S_N - N\mu}{\sigma\sqrt{N}} = \frac{S_N - 100}{\sqrt{200}/2} \approx \mathcal{N}(0,1).$$

We can find $x_{95\%} = 1.645$ such that

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\,du = 95\%.$$

Therefore, approximately

$$\mathbf{P}\left(\frac{S_N - 100}{\sqrt{200}/2} \le 1.645\right) \approx 95\%.$$

With this high probability 95%,

$$\frac{S_N - 100}{\sqrt{200}/2} \le 1.645 \quad \Leftrightarrow \quad S_N \le 100 + \frac{1}{2}\sqrt{200}\cdot 1.645 = 112.$$

So we need only 112 chicken: 100 chicken according to the mean, and 12 additional, as a buffer. The same applies to pasta. We will throw away $12 + 12 = 24$ meals after the trip, which is not a lot. Or maybe just give them to pilots or flight attendants.

6.2. **Normal and Poisson approximations.** Suppose you have a large quantity N of independent events, each of which happens with probability p. The number of these events is $\operatorname{Bin}(N, p)$. We have studied two approximations:

- Poisson, when $\operatorname{Bin}(N,p) \to \operatorname{Poi}(\lambda)$ if $p = \lambda/N$, as $N \to \infty$.
- Normal, when for $X \backsim \operatorname{Bin}(N,p)$, as $N \to \infty$, for constant p (independent of N):

$$\frac{X - pN}{\sqrt{pqN}} \to \mathcal{N}(0,1).$$

The first approximation is used when the events are rare, the second - when they are "usual", not rare. In practice, if you have something like $N = 1000$ and $p = 0.1\%$, you should use Poisson, and if $N = 1000$ but $p = 10\%$, you should use Normal.

6.3. **Equity buffer, or value at risk.** An insurance company has N clients. The kth client has insurance claim X_k. Suppose $X_1, X_2, X_3, \ldots$ are iid (independent identically distributed) random variables with mean $\mathbf{E}X_1 = \mathbf{E}X_2 = \ldots = \mu$ and variance $\operatorname{Var} X_1 = \operatorname{Var} X_2 = \ldots = \sigma^2$. The sum

$$S_N := X_1 + \ldots + X_N$$

is the total amount the company has to cover. How much should it charge clients so that it is not bankrupt with probability α (close to 1, say 95%)? By Central Limit Theorem,

$$\frac{S_N - \mu N}{\sigma\sqrt{N}} \approx \mathcal{N}(0,1).$$

Therefore, if x_α is such that

$$\Phi(x_\alpha) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_\alpha} e^{-y^2/2}\,dy = \alpha,$$

then with probability (approximately) α, we have:

$$\frac{S_N - \mu N}{\sigma\sqrt{N}} \le x_\alpha \ \Leftrightarrow\ S_N \le \mu N + \sigma\sqrt{N}\cdot x_\alpha = \mathrm{VaR}_\alpha\,.$$

The value VaR_α is called *equity buffer*, or *value at risk*, corresponding to the level of confidence α. The company should accumulate this amount in premiums (=money which clients pay for coverage), or using money from other sources, to be able to fulfil its obligations with probability α. Suppose that the company splits this sum between clients evenly: this makes sense, because clients pose the same risk (which is another way to say that $X_1, X_2, \ldots$ are identically distributed). Then each client pays the premium (cost of the policy)

$$\frac{1}{N}\mathrm{VaR}_\alpha = \mu + \frac{1}{\sqrt{N}}\sigma x_\alpha.$$

This premium consists of two parts: μ, which is *net premium*, covering the *average risk*, and the overhead $\frac{1}{\sqrt{N}}\sigma x_\alpha$, which covers risk. Indeed, it makes sense for an insurance company to demand at least the average of expected payoff as a premuim. But this is not enough, because the company bears clients' risk. The company has the right to demand additional payment, which is the overhead.

Note also that the larger N is, the smaller the overhead is. This is called *economies of scale*: companies need to *diversify* risks among a large pool of clients; the larger this pool is, the better.

6.4. **Extensions of the Central Limit Theorem.** The CLT (central limit theorem) is valid not only for iid (independent identically distributed) random variables. It is (sometimes) true if variables are independent but not identically distributed, or if they have slight dependence. The search for the conditions under which CLT is valid is still an active research area. For the purposes of this class, you can assume it is true if $\mathrm{Var}\,S_N$ is large. This is how it is formulated in the general form.

$$\boxed{\frac{S_N - \mathbf{E}S_N}{\sqrt{\mathrm{Var}\,S_N}} \to \mathcal{N}(0,1),\quad N \to \infty}$$

Example 6.2. An insurance company has $N_1 = 10000$ clients, each with claim Geo(1/2), and $N_2 = 40000$ clients, each with claim Geo(2/3). What is the value at risk for confidence level 95%?

Let X_k be the claim of the kth client from the first group, and let Y_k be the claim of the kth client from the second group. Then the total amount of claims is

$$S = S_{50000} = X_1 + \ldots + X_{N_1} + Y_1 + \ldots + Y_{N_2}.$$

Let us calculate $\mathbf{E}S$ and $\mathrm{Var}\,S$. Recall the expectation and variance for geometric distribution:

$$\mathbf{E}X_k = \frac{1}{1/2} = 2,\ \ \mathrm{Var}\,X_k = \frac{1-1/2}{(1/2)^2} = 2,\ \ \mathbf{E}Y_k = \frac{1}{2/3} = \frac{3}{2},\ \ \mathrm{Var}\,Y_k = \frac{1-2/3}{(2/3)^2} = \frac{3}{4}.$$

Therefore,

$$\mathbf{E}S = 2\cdot N_1 + \frac{3}{2}\cdot N_2 = 80000,\quad \mathrm{Var}\,S = 2\cdot N_1 + \frac{3}{4}\cdot N_2 = 50000.$$

By the CLT, we have:

$$\frac{S - 80000}{\sqrt{50000}} \approx \mathcal{N}(0,1).$$

Therefore, the following inequality is true with probability approximately 95%:

$$\frac{S-80000}{\sqrt{50000}} \le 1.645 \quad\Leftrightarrow\quad S \le 80000 + \sqrt{50000}\cdot 1.645.$$

The value at risk is $\boxed{80000 + \sqrt{50000}\cdot 1.645}$

Example 6.3. There are $N = 1000000$ (one million) eligible voters; each votes with probability $p = 49\%$ for Smith and with probability $q = 51\%$ for Johnson. Votes of different voters are independent. What is the probability that Smith wins? We assume all voters do vote. Let $X_k = 1$, if kth voter votes for Smith, $X_k = 0$ otherwise. Then

$$S_N = X_1 + \ldots + X_N$$

is the number of votes for Smith. Smith wins if $S_N > N/2$. But $\mathbf{P}(X_k = 1) = p$, $\mathbf{P}(X_k = 0) = q$, so $\mathbf{E}X_k = p = \mu$ and $\operatorname{Var} X_k = pq = \sigma^2$. By the CLT, we have:

$$\frac{S_N - pN}{\sqrt{pqN}} \approx \mathcal{N}(0,1).$$

Therefore,

$$\mathbf{P}(S_N > N/2) = \mathbf{P}\left(\frac{S_N - pN}{\sqrt{pqN}} > \frac{N/2 - pN}{\sqrt{pqN}}\right) = \mathbf{P}\left(\mathcal{N}(0,1) > \frac{N/2 - pN}{\sqrt{pqN}}\right).$$

But

$$\frac{N/2 - pN}{\sqrt{pqN}} = \frac{(1/2 - 0.49) \cdot 1000000}{\sqrt{0.49 \cdot 0.51 \cdot 1000000}} \approx 20.$$

Therefore,

$$\mathbf{P}(S_N > N/2) = \mathbf{P}(\mathcal{N}(0,1) > 20) = \frac{1}{\sqrt{2\pi}} \int_{20}^{\infty} e^{-y^2/2}\, dy < 10^{-50}.$$

Almost certainly, Johnson will win. Smith does not have any chance.

Problems

Problem 6.1. An insurance company has $N = 10000$ clients. Each of them can get into a car accident with probability 10%, independently of each other. If a client gets into a car accident, the amount of the claim is 1000 dollars. The company charges each of them a premium x. Calculate the minimal amount of x so that the company gets bankrupt only with probability 5%.

Problem 6.2. An insurance company has $N = 10000$ clients. Half of them file a claim which is distributed as $\operatorname{Exp}(1)$. The other half file a claim $\operatorname{Exp}(2)$; all claims are independent. Find the probability that the total amount will be between 7400 and 7560.

Problem 6.3. A health insurance company has three types of clients: low-risk, medium-risk, and high-risk. There are $N_1 = 10000$ low-risk clients, each will have a claim which is exponential with mean 100; $N_2 = 3000$ medium-risk clients, each will have a claim which is exponential with mean 200; and $N_3 = 1000$ high-risk clients, each will have a claim which is exponential with mean 1000. All claims are independent of each other. Find the value at risk for confidence level 99%.

For the next three problems, suppose you have N independent events, each of which has probability p. Will you rather use Normal or Poisson appoximation for the number of these events, if:

Problem 6.4. $N = 1000$ and $p = 0.2\%$.

Problem 6.5. $N = 10000$ and $p = 30\%$.

Problem 6.6. $N = 4000$ and $p = 0.01\%$.

Problem 6.7. There are two candidates for the local government office: Smith and Johnson. Each voter chooses to vote for Smith independently with probability 49%, and for Johnson with probability 51%. There are $N = 10000$ voters. What is the chance that Smith pulls out a victory?

Problem 6.8. (SOA) An insurance company issues 125 vision care insurance policies. The number of claims filed by a policyholder under a vision care insurance policy during one year is a Poisson random variable with mean 24. Assume the numbers of claims filed by different policyholders are mutually independent. Calculate the approximate probability that there is a total of between 3000 and 3100 claims during a one-year period.

Problem 6.9. There are two candidates for the local government office: Richard Hunter and Mary Williams. Each voter chooses Richard and Mary independently of other voters, with probability 49% and 51% respectively. There are $N = 100$ voters. What is the chance that Richard pulls out an upset victory?

Problem 6.10. An insurance company has $N = 10000$ clients. Each of them files a claim with size $\operatorname{Poi}(3)$, independently of others. Find the value at risk for the confidence level 99%: how much money the company needs to collect to be able to pay all claims with probability 99%.

Problem 6.11. Each of $N = 50000$ car owners will get into an accident independently with probability either $p_1 = 1\%$ (high-risk drivers, 10000 of them) or $p_2 = .5\%$ (low-risk drivers, 40000 of them). What is the probability that the number of accidents will differ from the mean by at least 10%?

Problem 6.12. A financial analyst models 10-year returns of 500 stocks from the Standard & Poor 500 index as independent random variables with mean 3.5 and variance 3.2. This analyst then invests into an equally weighted portfolio, so the return of this portfolio is equal to the average of these returns. Which normal distribution models the return of this portfolio? What is the probability that the actual return will be below 3?

Problem 6.13. An insurance company wants to charge its N clients a premium which corresponds to overhead 20%, that is, 20% higher than the average loss of this client. Each client sustains a loss of 250 or 100 with probabilities 10% and 20%, respectively. The company wants to be able to pay with probability 98%. What is the minimal number N of clients?

For the next two problems, there are $N = 10000$ car drivers. Each gets into an accident with probability 15%. In case of an accident, the losses are distributed uniformly on $[0, 10]$. Assume independence.

Problem 6.14. Find the probability that the total losses exceed 8000.

Problem 6.15. Find the value at risk at the confidence level 95%.

7. Introduction to Statistics

7.1. **Estimation.** Consider iid (independent identically distributed) random variables $X_1, \ldots, X_N$ with (unknown) mean and variance

$$\mu = \mathbf{E}X_1 = \mathbf{E}X_2 = \ldots = \mathbf{E}X_N, \quad \sigma^2 = \operatorname{Var} X_1 = \operatorname{Var} X_2 = \ldots = \operatorname{Var} X_N.$$

We want to estimate μ and σ^2. The estimate for μ: *empirical mean*

$$\overline{x} = \frac{X_1 + \ldots + X_N}{N}. \tag{7}$$

The estimate for σ^2:

$$s^2 = \frac{1}{N-1}\sum_{k=1}^{N}(X_k - \overline{x})^2. \tag{8}$$

Why do we divide by $N-1$, as opposed to N? Because it makes the estimate *unbiased*: $\mathbf{E}s^2 = \sigma^2$.

Good estimates should satisfy two properties:

- *Unbiased*: the expected value of the estimate should be equal to the true value of parameter. Here, it is indeed true:
$$\mathbf{E}\overline{x} = \frac{1}{N}\left(\mathbf{E}X_1 + \ldots + \mathbf{E}X_N\right) = \frac{1}{N}\left(\mu + \ldots + \mu\right) = \mu.$$
- *Consistent*: as the number of observations tends to infinity, an estimate should converge to the true value of the parameter it is trying to estimate. This is indeed true by the Law of Large Numbers:
$$\overline{x} \to \mu, \quad N \to \infty.$$

The estimate s^2 of σ^2 also satisfies these properties: it is unbiased, and consistent (harder to prove).

If we have two jointly distributed random variables (X, Y), then their covariance $\operatorname{Cov}(X, Y)$ can be estimated from a sample of i.i.d. $(x_1, y_1), \ldots, (x_N, y_N)$ as follows:

$$Q = \frac{1}{N-1}\sum_{k=1}^{N}(x_k - \overline{x})(y_k - \overline{y}). \tag{9}$$

This estimate is also unbiased and consistent.

7.2. **Confidence intervals.** Suppose you are trying to find not a precise estimate, but an interval of values in which the parameter likely lies. Let α be the level of confidence, say 95%. What is the confidence interval for μ? By the Central Limit Theorem,

$$\frac{X_1 + \ldots + X_N - \mu N}{\sigma\sqrt{N}} \approx \mathcal{N}(0, 1).$$

Therefore, let us find $x_{97.5\%}$ such that

$$\mathbf{P}\left(\mathcal{N}(0, 1) \le x_{97.5\%}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_{97.5\%}} e^{-t^2/2}\,dt = 97.5\%.$$

We split the bell curve into three parts: the right tail, which is 2.5%, the left tail, which is also 2.5%, and the middle, which is 95%. Therefore, with probability 95%,

$$-x_{97.5\%} \le \frac{X_1 + \ldots + X_N - \mu N}{\sigma\sqrt{N}} \le x_{97.5\%}.$$

$$-\frac{x_{97.5\%}\sigma}{\sqrt{N}} \le \frac{X_1 + \ldots + X_N - \mu N}{N} = \overline{x} - \mu \le \frac{x_{97.5\%}\sigma}{\sqrt{N}}.$$

Rewrite this as

$$-\frac{x_{97.5\%}\sigma}{\sqrt{N}} + \overline{x} \le \mu \le \frac{x_{97.5\%}\sigma}{\sqrt{N}} + \overline{x}.$$

This is the confidence interval for μ. If you do not know σ, then substitute s from s^2. Note that the confidence interval gets smaller as n grows. But it is larger if you set a higher confidence level.

Example 7.1. Assume we take observations $X_1, \ldots, X_N \sim \mathcal{N}(\theta, 9)$ for $N = 10000$ and are trying to estimate the unknown parameter θ. Let us find a confidence interval for the level of confidence 10%. Assume that $\overline{x} = 1.4353$. Then the confidence interval is given by

$$\left(-\frac{x_{95\%}\sqrt{9}}{\sqrt{10000}} + 1.4353, \frac{x_{95\%}\sqrt{9}}{\sqrt{10000}} + 1.4353\right) = (-0.04935 + 1.4353, 0.04935 + 1.4353).$$

7.3. **Hypotheses testing.** Suppose you want to test the hypothesis $\mu = 0$ at level of confidence 95% (or, which is the same, for $p = .05$).

Then we construct the confidence interval for $\alpha = 95\%$ and find whether $\mu = 0$ lies in this interval. If it does not, then we reject this hypothesis with confidence level 95%, or with $p = .05$. Because if we accept this hypothesis, then an unlikely event (with probability 5%) happened. If $\mu = 0$ does lie in this interval, then we do not reject this hypothesis, because it is still consistent with the experiment.

Example 7.2. A hotly debated Proposition X is on the polls for the local town referendum during this election season. A local newspaper polls $N_1 = 100$ people and 53 of them say they support Proposition X. Next week, another poll of $N_2 = 100$ people finds that 57 of them support Proposition X. The paper has a headline: "Support for Proposition X grows". You are unconvinced and would like to test the hypothesis that the level of support remained unchanged: 53%, since the last week. Let us test this for confidence level $p = 90\%$.

The hypothesis says that each vote is distributed as a Bernoulli random variable with $p = 0.53$. This distribution has mean $\mu = p = 0.53$ and variance $\sigma^2 = p(1-p) = 0.53 \cdot 0.47 = 0.2491$. Therefore, we should check whether

$$-\frac{x_{0.95\%}\sigma}{\sqrt{N_2}} + \mu \le \overline{x} \le \frac{x_{0.95\%}\sigma}{\sqrt{N_2}} + \mu.$$

If this is true, then we do not reject the hypothesis that each vote is distributed as a Bernoulli random variable with $p = 0.53$ (that is, support for Proposition X has not changed). If this is false, then we do reject this hypothesis. It turns out that

$$\frac{1.645 \cdot \sqrt{0.2491}}{\sqrt{100}} + 0.53 = 0.08 + 0.53 = 0.61 > 0.57 = \overline{x},$$

and, of course,

$$\frac{1.645 \cdot \sqrt{0.2491}}{\sqrt{100}} + 0.53 = 0.53 - 0.08 = 0.45 < 0.57.$$

Therefore, we do not reject the hypothesis $p = 0.53$.

7.4. **Bayesian methods.** These methods for confidence intervals and hypotheses testing were developed by Fisher. However, these *Fisherian methods* are sometimes inappropriate, because p is arbitrary, and because they test hypotheses without taking into account our prior beliefs in them. *Bayesian methods* aim to correct this disadvantage by using Bayes' theorem to switch from prior probabilities to posterior probabilities.

Suppose there is a poll of $N = 100$ people, with 53 expressing support for Proposition X. What is the actual level of support? Assume we have three hypotheses:

$$H_1 : p = 40\%;\ H_2 : p = 50\%;\ H_3 : p = 60\%.$$

Let us test these hypotheses for confidence level $p = 90\%$. As in the previous subsection, we reject the hypothesis H_1, but do not reject the hypotheses H_2 and H_3. Now, suppose you beileved prior to this poll that

$$\mathbf{P}(H_1) = 50\%,\ \mathbf{P}(H_2) = 30\%,\ \mathbf{P}(H_3) = 20\%.$$

How shall you update your beliefs after the poll? Let A be the event that this poll happened. By Bayes' formula,

$$\mathbf{P}(H_1 \mid A) = \frac{\mathbf{P}(A \mid H_1)\mathbf{P}(H_1)}{\mathbf{P}(A \mid H_1)\mathbf{P}(H_1) + \mathbf{P}(A \mid H_2)\mathbf{P}(H_2) + \mathbf{P}(A \mid H_3)\mathbf{P}(H_3)}.$$

However,

$$\mathbf{P}(A \mid H_1) = \binom{100}{53}0.4^{53}0.6^{47},\quad \mathbf{P}(A \mid H_2) = \binom{100}{53}0.5^{53}0.5^{47},\quad \mathbf{P}(A \mid H_3) = \binom{100}{53}0.6^{53}0.4^{47}.$$

Thus, after calculations, you get the following posterior probabilities:

$$\mathbf{P}(H_1 \mid A) = 4.7\%,\ \mathbf{P}(H_2 \mid A) = 73.7\%,\ \mathbf{P}(H_3 \mid A) = 21.5\%.$$

If another poll is conducted, then you again update your probabilities. This is the method Nate Silver uses in his blog FiveThirtyEight to predict presidential and other elections.

Problems

For the next five problems, consider the following setup. You have $N = 100$ observations, with $\bar{x} = 3.56$ and $s^2 = 18.24$.

Problem 7.1. Find the confidence interval for the confidence level $\alpha = 90\%$.

Problem 7.2. Find the confidence interval for $\alpha = 95\%$.

Problem 7.3. For what confidence level can we reject the hypothesis that $\mu = 3.5$?

Problem 7.4. Same question for $\mu = 3.6$.

Problem 7.5. Same question for $\mu = 4$.

Problem 7.6. Assume you estimate the rate parameter λ for $\mathrm{Exp}(\lambda)$ by estimating the mean λ^{-1}. Your $N = 200$ observations give you $\bar{x} = 0.545$ and $s^2 = 0.2533$. Find an estimate for λ and the confidence interval for λ with confidence level $\alpha = 98\%$.

For the next five problems, consider the following setup. There are two candidates: Richard Hunter and Mary Williams. A poll of $N = 1500$ voters shows 733 supporting Richard, and 867 supporting Mary.

Problem 7.7. How likely is it that Richard wins?

Problem 7.8. Can you reject this hypothesis at the confidence level 95%?

Problem 7.9. What is the confidence interval for the share of vote for Mary at level 98%?

Problem 7.10. Assume we have initially three hypotheses: the share of Mary's support is 25%, 50%, and 75%, with prior probabilities 20%, 60%, and 20%, respectively. What are posterior probabilities for each hypothesis?

Problem 7.11. Assume we have initially two hypotheses: the share of Mary's support is 40% and 60%, with prior probabilities 65% and 35%, respectively. What are posterior probabilities for each hypothesis?

For the next three problems, consider the following setup. Proposition A is on the ballot. A poll of $N = 200$ people finds that 112 of them support A. A week later, another poll of $M = 250$ people finds that 134 of them support A.

Problem 7.12. Can we reject the hypothesis that the support remains unchanged, with confidence level 95%?

Problem 7.13. Judging by the second poll, can we reject the hypothesis that A will fail, at confidence level 95%?

Problem 7.14. Assume you have the following hypotheses for the share of support: $p = 50\%$, $p = 60\%$, $p = 40\%$, with prior probabilities 50%, 25%, 25%. What are the posterior probabilities after the first poll? After the first and second poll?

Problem 7.15. Find the estimates $\overline{x}$ and s^2 for the selection $0, 1, 2, \ldots, N$.

8. Generating and Moment Generating Functions

8.1. **Generating functions.** Take a random variable X which takes values $0, 1, 2, \ldots$ It has a *generating function*:

$$\varphi_X(s) := \mathbf{E} s^X = \sum_{k=0}^{\infty} \mathbf{P}(X = k) s^k.$$

We can do this only for random variables taking values in the nonnegative integers, not with negative or fractional values, or continuous distributions. If you know this generating function, you know all values $p_k := \mathbf{P}(X = k)$ for each $k = 0, 1, 2, \ldots$, and so you know the distribution of X. Note that always $\varphi_X(1) = \mathbf{E} 1^X = \mathbf{E} 1 = 1$.

For a Poisson random variable: $X \sim \mathrm{Poi}(\lambda)$, we have:

$$\mathbf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

and therefore

$$\varphi_X(s) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-\lambda} e^{s\lambda} = \boxed{e^{(s-1)\lambda}}$$

For a Bernoulli random variable with probability p of success, when

$$\mathbf{P}(X = 1) = p, \quad \mathbf{P}(X = 0) = q, \quad p + q = 1,$$

we have:

$$\varphi_X(s) = \mathbf{E} s^X = p s^1 + q s^0 = \boxed{ps + q}$$

For independent X, Y, we have:

$$\varphi_{X+Y}(s) = \mathbf{E} s^{X+Y} = \mathbf{E}\left(s^X s^Y\right) = \mathbf{E} s^X \mathbf{E} s^Y = \varphi_X(s) \varphi_Y(s).$$

For a binomial random variable: $X \sim \mathrm{Bin}(N, p)$, we can represent it as a sum of N i.i.d. Bernoulli random variables: $X = X_1 + \ldots + X_N$. Therefore,

$$\varphi_X(s) = \varphi_{X_1}(s) \cdot \ldots \cdot \varphi_{X_N}(s) = (ps + q) \cdot \ldots \cdot (ps + q) = \boxed{(ps + q)^N}$$

For a geometric random variable: $X \sim \mathrm{Geo}(p)$, we have: $\mathbf{P}(X = k) = pq^{k-1}$, $k = 1, 2, \ldots$, and therefore

$$\varphi_X(s) = \sum_{k=1}^{\infty} p q^{k-1} s^k = ps \sum_{k=1}^{\infty} (qs)^{k-1} = ps \sum_{k=0}^{\infty} (qs)^k = \boxed{\frac{ps}{1 - qs}}$$

8.2. **Expectation and variance from generating functions.** Let us calculate $\varphi_X'(1)$. We have:

$$\varphi_X'(s) = \mathbf{E}\left(s^X\right)' = \mathbf{E}\left[X s^{X-1}\right],$$

because for a power function $s \mapsto s^x$ its derivative is $x s^{x-1}$. Plug in $s = 1$: $\boxed{\mathbf{E} X = \varphi_X'(1)}$ Next,

$$\varphi_X''(1) = \mathbf{E}\left(s^X\right)'' = \mathbf{E}\left[X(X-1) s^{X-2}\right].$$

Plug in $s = 1$: $\varphi_X''(1) = \mathbf{E}\left[X(X-1)\right] = \mathbf{E} X^2 - \mathbf{E} X$. Therefore, $\mathbf{E} X^2 = \mathbf{E} X + \varphi_X''(1) = \varphi_X'(1) + \varphi_X''(1)$, and $\operatorname{Var} X = \mathbf{E} X^2 - (\mathbf{E} X)^2 = \varphi_X''(1) + \varphi_X'(1) - \varphi'^2(1)$.

8.3. **Moment generating functions.** For general random variables (not necessarily taking values $0, 1, 2, \ldots$), we use the following instead of the generating function:

$$F_X(t) = \mathbf{E} e^{tX}, \; t \in \mathbb{R},$$

which is called the *moment generating function* (MGF). If X actually takes only values $0, 1, 2, \ldots$, then the generating function $\varphi_X(s) = \mathbf{E} s^X$ and the moment generating function $F_X(t)$ are related as follows: $\varphi_X(e^t) = F_X(t)$. Note that always $F_X(0) = 1$.

For an exponential random variable $X \sim \mathrm{Exp}(\lambda)$, we have:

$$F_X(t) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} \mathrm{d}x = \lambda \int_0^{\infty} e^{(t-\lambda)x} \mathrm{d}x = \boxed{\frac{\lambda}{\lambda - t}}, \quad t < \lambda.$$

Why? Because for $a > 0$ we have:

$$\int_0^{\infty} e^{-ax} \mathrm{d}x = \left(-\frac{e^{-ax}}{a}\right)\Bigg|_{x=0}^{x=\infty} = -\frac{e^{-\infty}}{a} - \left(-\frac{e^0}{a}\right) = \frac{1}{a}.$$

Apply this to $a = \lambda - t$. For the standard normal random variable $X \sim \mathcal{N}(0,1)$, we have:

$$F_X(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-x^2/2}\mathrm{d}x = \frac{1}{\sqrt{2\pi}}e^{t^2/2}\int_{-\infty}^{\infty} e^{-t^2/2+tx-x^2/2}\mathrm{d}x =$$

$$= \frac{e^{t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(x-t)^2/2}\mathrm{d}x = \frac{e^{t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-u^2}\mathrm{d}u =$$

(we changed variables $u = x - t$)

$$= \frac{e^{t^2/2}}{\sqrt{2\pi}}\sqrt{2\pi} = \boxed{e^{t^2/2}}$$

Therefore, for $Y \sim \mathcal{N}(\mu, \sigma^2)$ we have: $Y = \mu + \sigma X$, where $X \sim \mathcal{N}(0,1)$, and so

$$F_Y(t) = \mathbf{E}e^{tX} = \mathbf{E}e^{t\mu + t\sigma X} = e^{t\mu}\mathbf{E}e^{t\sigma X} = e^{t\mu}e^{t^2\sigma^2/2} = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right).$$

For independent random variables X and Y, we have:

$$F_{X+Y}(t) = \mathbf{E}e^{t(X+Y)} = \mathbf{E}e^{tX}e^{tY} = \mathbf{E}e^{tX}\mathbf{E}e^{tY} = F_X(t)F_Y(t).$$

For example, the sum of n i.i.d. $\mathrm{Exp}(\lambda)$ random variables (which has Gamma distribution $\Gamma(n, \lambda)$ has MGF

$$\left(\frac{\lambda}{\lambda - t}\right)^n.$$

More generally, the Gamma distribution $\Gamma(\alpha, \lambda)$ has MGF

$$\left(\frac{\lambda}{\lambda - t}\right)^\alpha.$$

Example 8.1. Find the distribution of a random variable X with moment generating function $F_X(t) = \exp(-4t + t^2)$. Answer: this corresponds to the normal distribution. Try $X \sim \mathcal{N}(\mu, \sigma^2)$, then we have: $F_X(t) = \exp(\mu t + \sigma^2 t^2/2)$. Comparing this with our MFG, we get: $\mu = -4$, and $\sigma^2 = 2$, so $X \sim \mathcal{N}(-4, 2)$.

Example 8.2. Find the distribution of a random variable X with

$$F_X(t) = e^{-3t}\frac{4}{4-t}.$$

For $Y = -3$, we have $F_Y(t) = e^{-3t}$. For $Z \sim \mathrm{Exp}(4)$, we have $F_Z(t) = 4/(4-t)$. Therefore, $X = Y + Z$.

Example 8.3. Same question, for

$$F_X(t) = \frac{1}{27}\left(2e^t + 3\right)^3. \tag{10}$$

Recall that for $X \sim \mathrm{Bin}(N, p)$, we have: $\varphi_X(s) = (ps+q)^N$, $q = 1 - p$. Therefore,

$$F_X(t) = \varphi_X(e^t) = (pe^t + q)^N. \tag{11}$$

Comparing (10) with (11), we see that we need $N = 3$ and $p = 2/3$.

Example 8.4. Same question, for

$$F_X(t) = \frac{2}{5}\left(e^t + e^{-t}\right) + \frac{1}{5}.$$

Answer:

$$\mathbf{P}(X = 1) = \mathbf{P}(X = -1) = \frac{2}{5},\ \mathbf{P}(X = 0) = \frac{1}{5}.$$

Example 8.5. A linear combination of independent normal random variables is itself normal. Let us illustrate this: take independent $X \sim \mathcal{N}(-1, 2)$ and $Y \sim \mathcal{N}(2, 8)$. What is the distribution of $Z = 2X - 3Y - 2$? The moment generating functions for X and Y are equal to

$$F_X(t) = \mathbf{E}e^{tX} = \exp\left(-t + t^2\right),\ F_Y(t) = \mathbf{E}e^{tY} = \exp\left(2t + 4t^2\right).$$

Therefore, the moment generating function for Z is

$$\begin{aligned}F_Z(t) &= \mathbf{E}e^{t(2X-3Y-2)} = \mathbf{E}\left[e^{(2t)X}e^{(-3t)Y}e^{-2t}\right]\\ &= e^{-2t}\mathbf{E}e^{(2t)X}\mathbf{E}e^{(-3t)Y} = e^{-2t}F_X(2t)F_Y(-3t)\\ &= \exp(-2t)\exp\left(-(2t) + (2t)^2\right)\exp\left(2(-3t) + 4(-3t)^2\right) = \exp\left(-10t + 40t^2\right)\end{aligned}$$

which corresponds to $\boxed{Z \sim \mathcal{N}(-10, 80)}$

8.4. **Expectation and moments with moment generating functions.** We have the following Taylor expansion:

$$F_X(t) = \mathbf{E}e^{tX} = \mathbf{E}\left(1 + tX + \frac{t^2X^2}{2} + \frac{t^3X^3}{6} + \ldots\right) = \mathbf{E}\sum_{k=0}^{\infty}\frac{(tX)^k}{k!} = \sum_{k=0}^{\infty}\frac{t^k}{k!}\mathbf{E}X^k.$$

Take the derivatives of this MGF:

$$F'_X(t) = \mathbf{E}\left(e^{tX}\right)' = \mathbf{E}\left[Xe^{tX}\right] \ \Rightarrow \ F'_X(0) = \mathbf{E}X,$$

$$F^{(k)}_X(t) = \mathbf{E}\left(e^{tX}\right)^{(k)} = \mathbf{E}\left[X^ke^{tX}\right] \ \Rightarrow \ F^{(k)}_X(0) = \mathbf{E}X^k.$$

So we can recreate moments $\mathbf{E}X^k$ from the MGF $F_X(t)$. In addition, $\operatorname{Var} X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = F''_X(0) - [F'_X(0)]^2$.

Example 8.6. For $X \sim \mathcal{N}(0,1)$, we have: $F_X(t) = e^{t^2/2}$, and $F'_X(t) = te^{t^2/2}$, $F''_X(t) = t'e^{t^2/2} + t(e^{t^2/2})' = (t^2+1)e^{t^2/2}$. Therefore, $\mathbf{E}X = F'_X(0) = 0$ and $\mathbf{E}X^2 = F''_X(0) = 1$, as we already know.

8.5. **Random sum of random variables.** Consider a random variable N, taking values

$$N = 0, 1, 2, \ldots$$

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, independent of N. Take a random sum

$$S := \sum_{k=1}^{N} X_k.$$

For example, if $N = 0$, we let $S := 0$. If $N = 1$, we let $S = X_1$. What is the distribution of S? Let us find its MGF $F_S(t)$, assuming we know MGF $F_{X_1}(t) = \mathbf{E}e^{tX_1}$ and generating function $\varphi_N(s) = \mathbf{E}s^N$. We have:

$$F_S(t) = \mathbf{E}e^{tS} = \sum_{k=0}^{\infty}\mathbf{E}e^{t(X_1+\ldots+X_k)}\mathbf{P}(N=k) = \sum_{k=0}^{\infty}\mathbf{E}\left[e^{tX_1}\ldots e^{tX_k}\right]\mathbf{P}(N=k) =$$

$$= \sum_{k=0}^{\infty}\mathbf{E}e^{tX_1}\ldots\mathbf{E}e^{tX_k}\mathbf{P}(N=k) = \sum_{k=0}^{\infty}\left[F_X(t)\right]^k\mathbf{P}(N=k) = \varphi_N(F_X(t)).$$

If we know MGF $F_S(t)$ of S, then we know the distribution of S. For example, if $N \sim \operatorname{Poi}(\lambda)$, then S is called a *compound Poisson random variable*. It can be used to model the total amount of claims by clients of an insurance company, because the number of these clients can be modeled by a Poisson random variable.

In particular, $F'_S(0) = \varphi'_N(F_{X_1}(0))F'_{X_1}(0) = \varphi'_N(1)F'_{X_1}(0) = \mathbf{E}N\cdot\mathbf{E}X_1$. (We used that $F_{X_1}(0) = 1$, as is true for every MGF.) That is,

$$\boxed{\mathbf{E}S = \mathbf{E}N\cdot\mathbf{E}X_1}$$

This is common sense: the mean of the random sum equals the mean of number of summands times the mean of each summand. Also, calculating $F''_S(0)$, we have:

$$\boxed{\operatorname{Var} S = \mathbf{E}N\cdot\operatorname{Var} X_1 + (\mathbf{E}X_1)^2\cdot\operatorname{Var} N}$$

Example 8.7. If $N \sim \operatorname{Poi}(\lambda)$ and $X_1 \sim \mathcal{N}(0,1)$, we have: $\varphi_N(s) = e^{\lambda(s-1)}$ and $F_{X_1}(t) = e^{t^2/2}$, so

$$F_S(t) = \varphi_N(F_{X_1}(t)) = \exp\left(\lambda\left(e^{t^2/2} - 1\right)\right).$$

We have: $\mathbf{E}S = \lambda\cdot 0 = 0$, and $\operatorname{Var} S = \lambda\cdot 1 + 0^2\cdot\lambda = \lambda$.

8.6. **Normal approximation for a random sum.** There is a version of the Central Limit Theorem that says: for large random N, $S = X_1 + \ldots + X_N$ is approximately normal. In other words,

$$\frac{S - \mathbf{E}}{\sqrt{\operatorname{Var} S}} \approx \mathcal{N}(0,1).$$

Example 8.8. Assume an insurance company receives $N \sim \operatorname{Poi}(1000)$ claims which are i.i.d. with $\mathbf{E}X = 1$ and $\operatorname{Var} X = 0.5$. Find the *value-at-risk* on the confidence level 95%, that is, a number $x_{95\%}$ such that $\mathbf{P}(S \le x_{95\%}) = 95\%$. This means that, having collected $x_{95\%}$ amount of money, the company will be able to pay all its claims with probability at least 95%. We have:

$$95\% = \mathbf{P}\left(S \le x_{95\%}\right) = \mathbf{P}\left(\frac{S - \mathbf{E}S}{\sqrt{\operatorname{Var} S}} \le \frac{x_{95\%} - \mathbf{E}S}{\sqrt{\operatorname{Var} S}}\right).$$

From the table of the normal distribution, we need to take
$$\frac{x_{95\%} - \mathbf{E}S}{\sqrt{\operatorname{Var} S}} = 1.645.$$
But we have:
$$\mathbf{E}S = \mathbf{E}N \cdot \mathbf{E}X = 1000 \cdot 1 = 1000,$$
$$\operatorname{Var} S = \mathbf{E}N \cdot \operatorname{Var} X + (\mathbf{E}X)^2 \cdot \operatorname{Var} N = 1000 \cdot 0.5 + 1^2 \cdot 1000 = 1500.$$
Thus, the value-at-risk is equal to $\boxed{1.645 \cdot \sqrt{1500} + 1000}$

8.7. **Proof of the central limit theorem.** Recall that if $X_1, X_2, \ldots$ are i.i.d. with mean μ and variance σ^2, then for $S_N := X_1 + \ldots + X_N$, we have:
$$\frac{S_N - N\mu}{\sqrt{N}\sigma} \to \mathcal{N}(0,1).$$
Let us show this for the case $\mu = 0$:
$$\frac{S_N}{\sigma\sqrt{N}} \to \mathcal{N}(0,1).$$
It suffices to prove convergence of mgf:
$$\mathbf{E}\exp\left(t\frac{S_N}{\sigma\sqrt{N}}\right) \to e^{t^2/2}.$$
Because of independence of $X_1, \ldots, X_N$ and because they are identically distributed, we have:
$$F_N(t) := \mathbf{E}\exp\left(t\frac{S_N}{\sigma\sqrt{N}}\right) = \mathbf{E}\exp\left(\frac{t}{\sigma\sqrt{N}}(X_1 + \ldots + X_N)\right) = \left[\mathbf{E}\exp\left(\frac{t}{\sigma\sqrt{N}}X_1\right)\right]^N.$$
But the moment generating function of X_1 is equal to
$$F(t) = \mathbf{E}e^{tX_1} = 1 + t\mathbf{E}X_1 + \frac{t^2}{2}\mathbf{E}X_1^2 + \ldots \approx 1 + \frac{\sigma^2 t^2}{2}.$$
This is because $\mathbf{E}X_1^2 = \operatorname{Var} X_1 = \sigma^2$. Therefore,
$$\log F_N(t) = N \log F\left(\frac{t}{\sigma\sqrt{N}}\right) \approx N \log\left(1 + \frac{t^2}{2N}\right) \approx N \cdot \frac{t^2}{2N} = \frac{t^2}{2}.$$
Thus, $F_N(t) \to e^{t^2/2}$, which completes the proof. For general μ, we can reduce it to the case $\mu = 0$ by noting that $S_N - \mu N = Y_1 + \ldots + Y_N$, where $Y_k = X_k - \mu$ has $\mathbf{E}Y_k = 0$.

Problems

Problem 8.1. An insurance company has a random number $N = 2M$ of i.i.d. claims, where $M \sim \operatorname{Poi}(1000)$. Each claim has size which is distributed as $\operatorname{Exp}(4)$. Using the normal approximation, find the value-at-risk for the level 99% of confidence.

Problem 8.2. An insurance company has a random number $N = 4M$ of i.i.d. claims, where $M \sim \operatorname{Poi}(500)$. Each claim has size which is distributed exponentially with mean 4. Using the normal approximation, find the probability that the total amount of claims is at least 600.

Problem 8.3. (SOA) An actuary determines that the claim size for a certain class of accidents is a random variable with moment generating function
$$F(t) = \frac{1}{(1 - 2500t)^4}.$$
Calculate the standard deviation for the claim size.

Problem 8.4. Consider a fair coin. Find the generating function for the third moment of Heads.

Problem 8.5. Assume a driver has accident each day with probability .1%. Using the Poisson approximation, find the generating function of the number of accidents during two years (each year is 365 days).

For the next two problems, consider the following setting. There are $N \sim \operatorname{Poi}(200)$ i.i.d. claims, each distributed as a Gamma distribution $X \sim \Gamma(2,3)$. The total amount of losses S is the sum of these claims.

Problem 8.6. Using the normal approximation, find the value at risk for the confidence level 95%, that is, the amount $x_{95\%}$ such that $\mathbf{P}(S \le x_{95\%}) = 95\%$.

Problem 8.7. Find the moment generating function of S in the problem above.

For the next two problems, take a Poisson random variable $X \sim \mathrm{Poi}(4)$.

Problem 8.8. Find the generating function of $Y = 2X$.

Problem 8.9. Find the generating function of $Z = X + 3$.

Problem 8.10. Taking the third derivative of the generating function of $X \sim \mathrm{Poi}(\lambda)$, find $\mathbf{E}X^3$.

Problem 8.11. Take two independent random variables, $X \sim \mathrm{Bin}(3, 0.4)$ and $Y \sim \mathrm{Geo}(0.3)$. Find the generating function of $X + 2Y + 1$.

Problem 8.12. A bank has $N \sim \mathrm{Poi}(1000)$ independent investments, each gives net profit distributed as $Z - 1$, where $Z \sim \mathrm{Exp}(1)$. Find the value-at-risk at the confidence level 95%, that is, the amount of capital the bank needs to have to buffer its possible losses with probability 95%.

Problem 8.13. For independent random variables

$$X \sim \mathcal{N}(-1, 3),\ Y \sim \mathcal{N}(0, 2),\ Z \sim \mathcal{N}(4, 1),$$

consider the random variable $U := 2X - 4Y - Z + 5$. Find the expectation, variance, and the moment generating function for U.

For the next three problems, consider the following setting. There are $N \sim \mathrm{Geo}(1/3)$ i.i.d. claims, each distributed as an $X \sim \mathrm{Exp}(2)$. The total claim S is the sum of these claims.

Problem 8.14. Find the moment generating function of S.

Problem 8.15. Find $\mathbf{E}S$.

Problem 8.16. Find $\mathrm{Var}\, S$.

Problem 8.17. There are $N \sim \mathrm{Poi}(100)$ i.i.d. claims, each distributed as an $X \sim \mathrm{Exp}(1)$. The total amount of losses S is the sum of these claims. Using the normal approximation, find the value at risk for the confidence level 99%, that is, the amount $x_{99\%}$ such that $\mathbf{P}(S \le x_{99\%}) = 99\%$.

Problem 8.18. Find the moment generating function of the uniform random variable on $[0, 1]$.

Problem 8.19. (SOA) Let X_1, X_2, X_3 be i.i.d. random variables with distribution

$$\mathbf{P}(X_1 = 0) = \frac{1}{3},\ \mathbf{P}(X_1 = 1) = \frac{2}{3}.$$

Calculate the moment generating function of $Y = X_1X_2X_3$.

For the following five problems, find the distributions which correspond to the given moment generating functions.

Problem 8.20. $\frac{e^{5t}}{1-t/2}$.

Problem 8.21. $\frac{4e^{-t}}{(2-t)^2}$.

Problem 8.22. $e^{4t+t^2/5}$.

Problem 8.23. $\frac{2e^t}{3-e^t}$.

Problem 8.24. $\frac{2}{5}e^t + \frac{1}{5} + \frac{2}{5}e^{-t}$.

For the next three problems, find the distribution with this generating function.

Problem 8.25. $\varphi(s) = \frac{1}{4}(s^3 + s^2 + s + 1)$.

Problem 8.26. $\varphi(s) = e^{3s^2}e^{-3}$.

Problem 8.27. $\varphi(s) = \frac{s}{1-s}$.

9. Probabilistic Inequalities and the Law of Large Numbers

9.1. **Markov's inequality.** For every nonnegative random variable $X \ge 0$, and for every $a > 0$, we have:

$$\boxed{\mathbf{P}(X \ge a) \le \frac{\mathbf{E}X}{a}}$$

Indeed, consider another random variable:

$$Y = \begin{cases} a, & X \ge a;\\ 0, & 0 \le X < a.\end{cases}$$

Then always $Y \le X$, so $\mathbf{E}Y \le \mathbf{E}X$. But $\mathbf{E}Y = a\mathbf{P}(X \ge a) + 0\mathbf{P}(0 \le X < a) = a\mathbf{P}(X \ge a)$. This gives us

$$a\mathbf{P}(X \ge a) \le \mathbf{E}X \quad \Rightarrow \quad \mathbf{P}(X \ge a) \le \frac{\mathbf{E}X}{a}.$$

Example 9.1. Consider the exponential random variable $X \backsim \mathrm{Exp}(1)$. Then

$$\mathbf{P}(X \ge 10) \le \frac{\mathbf{E}X}{10} = \frac{1}{10},$$

because $\mathbf{E}X = 1$. In reality, $\mathbf{P}(X \ge 10) = e^{-10}$, which is much smaller than $1/10$. Markov's inequality gives us very rough and crude estimates, which are one-size-fits-all solutions. But they are universal (can be applied to any variable).

9.2. **Chebyshev's inequality.** If $\mathbf{E}X = \mu$, then for every $b > 0$ we have:

$$\boxed{\mathbf{P}(|X-\mu| \ge b) \le \frac{\operatorname{Var} X}{b^2}}$$

Indeed, apply Markov's inequality to $(X-\mu)^2$ instead of X and b^2 instead of a. We get:

$$\mathbf{P}(|X-\mu| \ge b) = \mathbf{P}((X-\mu)^2 \ge b^2) \le \frac{\mathbf{E}(X-\mu)^2}{b^2} = \frac{\operatorname{Var} X}{b^2}.$$

Example 9.2. (Rule of three sigmas.) Can a random variable be away from its mean for more than three standard deviations? Let $\sigma^2 = \operatorname{Var} X$, then σ is the standard deviation. Let $b = 3\sigma$. Then

$$\mathbf{P}(|X-\mu| \ge b) \le \frac{\sigma^2}{b^2} = \frac{1}{9}.$$

But this estimate is again very crude. For example, take $X \backsim \mathcal{N}(0,1)$. Then $\mu = 0,\ \sigma = 1$, and $b = 3$, so

$$\mathbf{P}(|X-\mu| \ge b) = \mathbf{P}(X \ge 3 \text{ or } X \le -3) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-3} e^{-x^2/2}\,dx + \frac{1}{\sqrt{2\pi}}\int_3^{+\infty} e^{-x^2/2}\,dx < 1\%.$$

If you model something by a normal random variable, then it is very unlikely that it will be away for more than three sigmas from the mean.

9.3. **Law of large numbers.** Let $X_1, X_2, \ldots$ be iid (independent identically distributed) random variables with

$$\mathbf{E}X_1 = \mathbf{E}X_2 = \ldots = \mu, \quad \operatorname{Var} X_1 = \operatorname{Var} X_2 = \ldots = \sigma^2.$$

Let $S_N = X_1 + \ldots + X_N$. Then

$$\frac{S_N}{N} \to \mu,\ N \to \infty,$$

in the sense that for every $\varepsilon > 0$ we have:

$$\boxed{\mathbf{P}\left(\left|\frac{S_N}{N} - \mu\right| \ge \varepsilon\right) \to 0}$$

Indeed, apply Chebyshev's inequality to S_N/N. We have:

$$\mathbf{E}\frac{S_N}{N} = \frac{1}{N}(\mathbf{E}X_1 + \ldots + \mathbf{E}X_N) = \frac{1}{N}(\mu + \ldots + \mu) = \frac{\mu N}{N} = \mu.$$

$$\operatorname{Var}\frac{S_N}{N} = \frac{1}{N^2}\operatorname{Var} S_N = \frac{1}{N^2}\left(\operatorname{Var} X_1 + \ldots + \operatorname{Var} X_N\right) = \frac{1}{N^2}\left(\sigma^2 + \ldots + \sigma^2\right) = \frac{\sigma^2}{N}.$$

Therefore, as $N \to \infty$, we have:

$$\mathbf{P}\left(\left|\frac{S_N}{N} - \mu\right| \ge \varepsilon\right) \le \frac{\sigma^2/N}{\varepsilon^2} \to 0.$$

Central Limit Theorem tells us *how fast* S_N/N converges to μ. It converges with order $1/\sqrt{N}$. Indeed, as $N \to \infty$, by Central Limit Theorem we have:

$$\frac{\sqrt{N}}{\sigma}\left(\frac{S_N}{N} - \mu\right) = \frac{S_N - \mu N}{\sigma\sqrt{N}} \to \mathcal{N}(0,1).$$

9.4. **Large deviations.** Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbf{E}X_1 = \mu$. Then by the Law of Large Numbers for $S_N = X_1 + \ldots + X_N$ we have: $S_N/N \to \mu$. Large Deviations refer to estimating the following probabilities for $\varepsilon > 0$:

$$\mathbf{P}\left(\frac{S_N}{N} - \mu \geq \varepsilon\right) \text{ and } \mathbf{P}\left(\frac{S_N}{N} - \mu \leq -\varepsilon\right).$$

This can be done by combining Markov's inequality and moment generating functions. Instead of formulating general theory, we do a few examples.

Example 9.3. Take $X \sim \mathcal{N}(2,1)$. Then $\mathbf{E}X = 2$. Let us estimate $\mathbf{P}(X \geq 3)$: For $t > 0$, a parameter to be determined later. Then $\mathbf{P}(X \geq 3) = \mathbf{P}\left(e^{tX} \geq e^{3t}\right)$. By Markov's inequality,

$$\mathbf{P}\left(e^{tX} \geq e^{3t}\right) \leq \frac{\mathbf{E}e^{tX}}{e^{3t}}.$$

But the moment generating function for $X \sim \mathcal{N}(2,1)$ is given by $\mathbf{E}e^{tX} = e^{2t+t^2/2}$. Choose t to minimize

$$\frac{e^{2t+t^2/2}}{e^{3t}} = \exp\left(\frac{1}{2}t^2 - t\right).$$

To minimize $t^2/2 - t$, take the derivative with respect to t: $(t^2/2 - t)' = t - 1 = 0 \ \Rightarrow\ t = 1$. This gives us

$$\exp\left(\frac{1}{2}t^2 - t\right) = e^{-0.5}.$$

Therefore, $\mathbf{P}(X \geq 3) \leq \boxed{e^{-0.5}}$ Of course, in this simple case we might as well directly calculate this probability. But is more complicated settings, this direct calculation is impossible.

Example 9.4. Take $X \sim \mathcal{N}(2,1)$, and estimate $\mathbf{P}(X \leq 1)$. For $t > 0$, we have:

$$\mathbf{P}(X \leq 1) = \mathbf{P}\left(e^{-tX} \geq e^{-t}\right) = \frac{\mathbf{E}e^{-tX}}{e^{-t}} = \frac{e^{-2t+t^2/2}}{e^{-t}} = \exp\left(\frac{t^2}{2} - t\right),$$

and this problem is solved simlarly: $\mathbf{P}(X \leq 1) \leq \boxed{e^{-0.5}}$

This method works particularly well for *sums of independent random variables.*

Example 9.5. Let $X_1, \ldots, X_{10} \sim \mathrm{Exp}(3)$ be i.i.d. random variables, and let $S := X_1 + \ldots + X_{10}$. Then $\mathbf{E}S = \frac{10}{3}$. Let us estmate $\mathbf{P}(S \leq 2)$. For $t > 0$,

$$\mathbf{P}(S \leq 2) = \mathbf{P}\left(e^{-tS} \leq e^{-2t}\right) \leq \frac{\mathbf{E}e^{-tS}}{e^{-2t}}.$$

Since $X_1, \ldots, X_{10}$ are independent,

$$\begin{aligned}\mathbf{E}e^{-tS} &= \mathbf{E}\left[e^{-tX_1 - \ldots - tX_{10}}\right] = \mathbf{E}\left[e^{-tX_1}\ldots e^{-tX_{10}}\right] = \mathbf{E}\left[e^{-tX_1}\right]\cdot \ldots \cdot \mathbf{E}\left[e^{-tX_{10}}\right] \\ &= \left(\mathbf{E}e^{-tX_1}\right)^{10} = \left(\frac{3}{3+t}\right)^{10}.\end{aligned}$$

We used expression for moment generating function of the exponential random variable: $\mathbf{E}e^{-tX_1} = 3/(3+t)$. Therefore, we need to minimize

$$F(t) = \frac{\left(\frac{3}{3+t}\right)^{10}}{e^{-2t}} = e^{2t}\left(\frac{3}{3+t}\right)^{10}$$

Take the logarithm:

$$\ln F(t) = 2t - 10\ln(3+t) \ \Rightarrow\ (\ln F(t))' = 2 - \frac{10}{3+t} = 0 \ \Rightarrow\ t = 2.$$

Therefore, the minimal value is $\mathbf{P}(S \leq 2) \leq F(2) = e^{2\cdot 2}(3/5)^{10} = \boxed{0.33}$

9.5. **Chernov's inequality.** This is a special case of large deviations above. Let $X_1, \ldots, X_N$ be independent Bernoulli random variables with $\mathbf{P}(X_i = 1) = p_i$, $\mathbf{P}(X_i = 0) = 1 - p_i$. Then $S = X_1 + \ldots + X_N$ has mean $\mu = \mathbf{E}S = \mathbf{E}X_1 + \ldots + \mathbf{E}X_N = p_1 + \ldots + p_N$. Fix a $\delta \in (0, 1)$ and estimate

$$\mathbf{P}(S \ge \mu + \delta\mu).$$

Take a parameter $t > 0$:

$$\mathbf{P}(S \ge \mu + \delta\mu) \le \mathbf{P}\left(e^{tS} \ge e^{t\mu(1+\delta)}\right) \le \frac{\mathbf{E}e^{tS}}{e^{t\mu(1+\delta)}}. \tag{12}$$

Since $X_1, \ldots, X_N$ are independent,

$$\mathbf{E}e^{tS} = \mathbf{E}\left[e^{tX_1 + \ldots + tX_N}\right] = \mathbf{E}e^{tX_1} \cdot \ldots \cdot \mathbf{E}e^{tX_N}. \tag{13}$$

For each i, we have: $\mathbf{E}e^{tX_i} = p_i e^{t\cdot 1} + (1 - p_i)e^{t\cdot 0} = p_i e^t + (1 - p_i) = 1 + p_i(e^t - 1)$. Now comes the main step. Recall that

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} \ge 1 + x \text{ for } x \ge 0. \tag{14}$$

Apply inequality (14) to $x = p_i(e^t - 1)$. Then

$$\mathbf{E}e^{tX_i} \le \exp\left(p_i(e^t - 1)\right). \tag{15}$$

Substituting (15) into (13), we get:

$$\mathbf{E}e^{tS} \le \exp\left(p_1(e^t - 1)\right) \cdot \ldots \cdot \exp\left(p_N(e^t - 1)\right) = \exp\left((p_1 + \ldots + p_N)(e^t - 1)\right) = \exp(\mu(e^t - 1)). \tag{16}$$

The right-hand side of (16) depends only on μ, regardless of individual p_i. Plugging (16) into (12), we get:

$$\mathbf{P}(S \ge \mu + \delta\mu) \le \exp\left(\mu(e^t - 1 - t(1 + \delta))\right) = \exp(\mu F(t)),\ F(t) := e^t - 1 - (1 + \delta)t.$$

Minimize $F(t)$ with respect to t: $F'(t) = e^t - (1 + \delta) = 0 \;\Rightarrow\; t = \ln(1 + \delta)$. The minimal value is $F(\delta) = e^{\ln(1+\delta)} - 1 - (1 + \delta)\ln(1 + \delta) = \delta - (1 + \delta)\ln(1 + \delta)$. It is possible to show that this is less than $-\delta^2/3$ for $\delta \in (0, 1)$. Thus,

$$\boxed{\mathbf{P}(S \ge \mu + \delta\mu) \le \exp\left(-\frac{\delta^2\mu}{3}\right)}$$

Similarly, it is possible to show that

$$\boxed{\mathbf{P}(S \le \mu - \delta\mu) \le \exp\left(-\frac{\delta^2\mu}{2}\right)}$$

Example 9.6. $S \sim \mathrm{Bin}(20, 0.4)$ can be represented as $S = X_1 + \ldots + X_{20}$ with i.i.d. Bernoulli X_i with $p_i = 0.4$. Then $\mu = p_1 + \ldots + p_{20} = 0.4 \cdot 20 = 8$. Let us estimate $\mathbf{P}(S \ge 12)$: this corresponds to $8(1 + \delta) = 12 \;\Rightarrow\; \delta = 0.5$, and $\mathbf{P}(S \ge 12) \le \exp\left(-\frac{0.5^2\cdot 8}{2}\right) = \boxed{e^{-1}}$

Example 9.7. Take $S_1 \sim \mathrm{Bin}(10, 0.4)$ and $S_2 \sim \mathrm{Bin}(20, 0.6)$. Then $S_1 = X_1 + \ldots + X_{10}$ with i.i.d. Bernoulli $X_1, \ldots, X_{10}$ with $p_1 = \ldots = p_{10} = 0.4$, and $S_2 = X_{11} + \ldots + X_{30}$ with i.i.d. Bernoulli $X_{11}, \ldots, X_{30}$ with $p_{11} = \ldots = p_{30} = 0.6$. Then $\mathbf{E}S = \mu = p_1 + \ldots + p_{30} = 10\cdot 0.4 + 20\cdot 0.6 = 16$. Therefore, $\mathbf{P}(S \le 10) \le \mathbf{P}(S \le \mu(1 - \delta))$ for $\delta = 3/8$, and is estimated as $\exp(-0.5 \cdot (3/8)^2 \cdot 16) = \boxed{0.32}$

Problems

For the next three problems, estimate $\mathbf{P}(X \ge 10)$ using Markov's inequality.

Problem 9.1. $X \backsim \mathrm{Poi}(1)$.

Problem 9.2. $X \backsim \mathrm{NB}(2, 1/2)$.

Problem 9.3. $X \backsim \mathrm{Uni}(0, 10)$.

For the next seven problems, use Chebyshev's inequality.

Problem 9.4. The average height of a person from a certain group is 1.7 meters, the standard deviation is 0.25 meters. Estimate the probability that the given person has height X such that $|X - 1.7| \ge 0.3$.

Problem 9.5. Estimate the probability that a given person's height is away more than four standard deviations from its mean.

Problem 9.6. Let $X_1, \ldots, X_{100} \sim \mathrm{Poi}(2)$. Estimate the probability
$$\mathbf{P}\left(\frac{X_1 + \ldots + X_{100}}{100} \ge 2.3\right).$$

Problem 9.7. Let $X_1, \ldots, X_{1000} \sim \mathcal{N}(-5, 12)$. Estimate the probability
$$\mathbf{P}\left(\left|\frac{X_1 + \ldots + X_{1000}}{1000} + 5\right| \ge 0.1\right).$$

Problem 9.8. Let $X_1, \ldots, X_{50} \sim \mathcal{N}(0, 1)$, $Y_1, \ldots, Y_{100} \sim \mathrm{Uni}(0, 2)$ be independent. Estimate
$$\mathbf{P}\left(\left|\frac{S - \mathbf{E}S}{300}\right| \ge 0.4\right).$$

Problem 9.9. Let i.i.d. $X_1, \ldots, X_{100}$ have density $p(x) = 2(1-x)$, $0 \le x \le 1$. Estimate $\mathbf{P}(S \ge 50)$ using Chebyshev's inequality.

Problem 9.10. Let $Y_1, \ldots, Y_{150}$ each be the number of tosses for a fair coin you need to get your first Heads. Assume all these random variables are independent. Estimate $\mathbf{P}(|Y_1 + \ldots + Y_{150} - 300| \ge 50)$ using Chebyshev's inequality.

For the next five problems, estimate probabilities using the moment generating functions.

Problem 9.11. $\mathbf{P}(X_1 + \ldots + X_{20} \ge 6)$ for i.i.d. $X_i \sim \mathrm{Exp}(10)$.

Problem 9.12. $\mathbf{P}(X_1 + \ldots + X_{100} \le 5)$ for i.i.d. $X_i \sim \mathrm{Exp}(10)$.

Problem 9.13. $\mathbf{P}(X_1 + \ldots + X_{100} + Y_1 + \ldots + Y_{30} \ge 120)$ for independent $X_i \sim \mathrm{Exp}(1)$ and $Y_i \sim \mathcal{N}(0, 1)$.

Problem 9.14. $\mathbf{P}(X \ge 3)$ for $X \sim \Gamma(4, 6)$.

Problem 9.15. $\mathbf{P}(X \le -2)$ for $X \sim \mathcal{N}(-0.5, 0.3)$.

For the next four problems, apply Chernov's inequality.

Problem 9.16. $\mathbf{P}(X_1 + \ldots + X_{200} \ge 160)$, with X_i independent Bernoulli with
$$\mathbf{P}(X_i = 1) = \begin{cases} 0.7, \ i = 1, \ldots, 120; \\ 0.5, \ i = 121, \ldots, 200. \end{cases}$$

Problem 9.17. Toss a fair coin 100 times. Let X be the number of Heads. Estimate $\mathbf{P}(X \le 20)$.

Problem 9.18. Among $N = 50000$ drivers, $N_1 = 10000$ are safe: they get into an accident with probability $p = 1\%$, and $N_2 = 40000$ are dangerous, with $q = 3\%$. All accidents are independent. Let X be the total number of accidents. Find $\mathbf{E}X$ and estimate $\mathbf{P}(X \ge 1.5 \cdot \mathbf{E}X)$.

Problem 9.19. Let $X_1, \ldots, X_{50}$ be i.i.d. Bernoulli with $p_X = 0.8$, and $Y_1, \ldots, Y_{40}$ be i.i.d. Bernoulli with $p_Y = 0.6$. Estimate $\mathbf{P}(X_1 + \ldots + X_{50} + Y_1 + \ldots + Y_{40} \le 100)$.

10. Conditional Expectation and Conditional Distribution

10.1. **Conditional probability, discrete case.** Recall how to find conditional probability from Section 2.

Example 10.1. Toss a fair coin twice. Let A be the event that there are two Heads. Let $Y = 1$ if the first toss is H, 0 otherwise. Then
$$\mathbf{P}(A \mid Y = 1) = \frac{1}{2},$$
because this is the probability that the second toss is H. Next, if the first toss is not H, then we cannot have both H; thus,
$$\mathbf{P}(A \mid Y = 0) = 0.$$
We can write this as
$$\mathbf{P}(A \mid Y) = \begin{cases} \mathbf{P}(A \mid Y = 0), \ Y = 0; \\ \mathbf{P}(A \mid Y = 1), \ Y = 1; \end{cases} = \begin{cases} 0, \ Y = 0; \\ \frac{1}{2}, \ Y = 1; \end{cases} = \frac{Y}{2}.$$

The random variable $\mathbf{P}(A \mid Y)$ is called the *conditional probability of A given Y*. We can also have conditional probability depending on many random variables.

Example 10.2. Let X, Y, Z be random variables with the following joint distribution:

X	Y	Z	Prob.
0	0	0	1/8
1	0	0	1/8
0	1	2	1/8
-1	1	1	1/8
0	1	1	1/2

Find $\mathbf{P}(X = 0 \mid Y, Z)$. We have:

$$\mathbf{P}(X = 0 \mid Y = Z = 0) = \frac{1/8}{1/8 + 1/8} = \frac{1}{2}.$$

$$\mathbf{P}(X = 0 \mid Y = 1,\ Z = 2) = 1.$$

$$\mathbf{P}(X = 0 \mid Y = Z = 1) = \frac{1/2}{1/2 + 1/8} = \frac{4}{5}.$$

In other words,

$$\mathbf{P}(X = 0 \mid Y, Z) = \begin{cases} \frac{1}{2},\ Y = Z = 0; \\ 1,\ Y = 1,\ Z = 2; \\ \frac{4}{5},\ Y = Z = 1; \end{cases}$$

10.2. **Conditional expectation, discrete case.** This is defined similarly to the conditional probability.

Example 10.3. In the previous example, let us find $\mathbf{E}(X \mid Y, Z)$. We have:

$$\mathbf{E}(X \mid Y = Z = 0) = \frac{(1/8)\cdot 0 + (1/8)\cdot 1}{(1/8) + (1/8)} = \frac{1}{2},$$

$$\mathbf{E}(X \mid Y = 1,\ Z = 2) = 1,$$

$$\mathbf{E}(X \mid Y = Z = 1) = \frac{(-1)\cdot(1/8) + 0\cdot(1/2)}{(1/8) + (1/2)} = -\frac{1}{5}.$$

Example 10.4. Toss a fair coin twice, and let X be the number of Heads, let $Y = 1$ if the first toss is H, $Y = 0$ otherwise. If $Y = 0$, then $X = 0$ or $X = 1$ with probability $1/2$. If $Y = 1$, then $X = 1$ or $X = 2$ with probabilty $1/2$. Then

$$\mathbf{P}(X \mid Y = 0) = 0\cdot\frac{1}{2} + 1\cdot\frac{1}{2} = \frac{1}{2},$$

$$\mathbf{P}(X \mid Y = 1) = 1\cdot\frac{1}{2} + 2\cdot\frac{1}{2} = \frac{3}{2}.$$

We can represent this as

$$\mathbf{P}(X \mid Y) = \begin{cases} \frac{1}{2},\ Y = 0; \\ \frac{3}{2},\ Y = 1; \end{cases} = \frac{1}{2} + Y.$$

10.3. **Properties of a conditional expectation.** Take random variables $X, Y, Z_1, \ldots, Z_n$.

(a) $\mathbf{E}(X + Y \mid Z_1, \ldots, Z_n) = \mathbf{E}(X \mid Z_1, \ldots, Z_n) + \mathbf{E}(Y \mid Z_1, \ldots, Z_n)$;

(b) If X is a function of $Z_1, \ldots, Z_n$, then $\mathbf{E}(X \mid Z_1, \ldots, Z_n) = X$, because if you know $Z_1, \ldots, Z_n$, you already know the exact value of X.

(c) If X is independent of $Z_1, \ldots, Z_n$, then $\mathbf{E}(X \mid Z_1, \ldots, Z_n) = \mathbf{E}X$, because knowledge of $Z_1, \ldots, Z_n$ does not give us any additional information about the distribution of X.

(d) If X is a function of $Z_1, \ldots, Z_n$, then $\mathbf{E}(XY \mid Z_1, \ldots, Z_n) = X\mathbf{E}(Y \mid Z_1, \ldots, Z_n)$. Indeed, when you try to predict the value of XY based on the information from $Z_1, \ldots, Z_n$, then you already know the value of X, and you can assume X is just a constant, and put it outside of the expectation sign.

(e) $\mathbf{E}(\mathbf{E}(X \mid Z_1, \ldots, Z_n)) = \mathbf{E}X$. The prediction of X given $Z_1, \ldots, Z_n$ is itself a random variable, which is a function of $Z_1, \ldots, Z_n$. If you average over all possible values of $Z_1, \ldots, Z_n$, you give up your knowledge of $Z_1, \ldots, Z_n$, and arrive and the original situation, when you did not know anything. There, the best prediction for X is $\mathbf{E}X$.

Example 10.5. Take i.i.d. $X, Y, Z \sim \mathrm{Poi}(1)$. Then $\mathbf{E}X = \operatorname{Var} X = 1 \ \Rightarrow\ \mathbf{E}X^2 = (\mathbf{E}X)^2 + \operatorname{Var} X = 2$, same for Y, Z. Therefore,

$$\mathbf{E}(X + 2Y + Z \mid Z) = \mathbf{E}(X \mid Z) + 2\mathbf{E}(Y \mid Z) = \mathbf{E}(Z \mid Z) = \mathbf{E}X + 2\mathbf{E}Y + Z = \boxed{3 + Z}$$

$$\mathbf{E}(2X + Z - 1)^2 \mid Z) = \mathbf{E}(4X^2 + 4XZ - 4X + Z^2 - 2Z + 1 \mid Z)$$

$$
\begin{aligned}
&= 4\mathbf{E}(X^2 \mid Z) + 4\mathbf{E}(XZ \mid Z) - 4\mathbf{E}(X \mid Z) + \mathbf{E}(Z^2 - 2Z + 1 \mid Z)\\
&= 4\mathbf{E}X^2 + 4Z\mathbf{E}X - 4\mathbf{E}X + Z^2 - 2Z + 1\\
&= 8 + 4Z - 4 + Z^2 - 2Z + 1 = \boxed{Z^2 + 2Z + 5}
\end{aligned}
$$

10.4. **Conditional distribution.** This is the distribution of a random variable X given $Z_1, \ldots, Z_n$, which consists of all probabilities $\mathbf{P}(X = x \mid Z_1, \ldots, Z_n)$, where x is a possible value of the random variable X.

Example 10.6. Toss two fair coins, and let

$$
X = \begin{cases} 1, & \text{first H;}\\ 0, & \text{first T,} \end{cases} \qquad Y = \text{number of Heads} = \begin{cases} 2, & \text{HH;}\\ 1, & \text{HT, TH;}\\ 0, & \text{TT.} \end{cases}
$$

- if $Y = 2$, then $\mathbf{P}(X = 1 \mid Y) = \mathbf{P}(X = 1 \mid Y) = 1$, $\mathbf{E}(X \mid Y) = 1$, and $\mathrm{Var}(X \mid Y) = 0$, because a constant random variable X has zero variance;
- if $Y = 1$, then $\mathbf{P}(X = 1 \mid Y) = \mathbf{P}(X = 0 \mid Y) = 0.5$, $\mathbf{E}(X \mid Y) = 0.5$, $\mathbf{E}(X^2 \mid Y) = 1^2 \cdot 0.5 + 0^2 \cdot 0.5 = 0.5$, and $\mathrm{Var}(X \mid Y) = \mathbf{E}(X^2 \mid Y) - (\mathbf{E}(X \mid Y))^2 = 0.5 - 0.5^2 = 0.25$;
- if $Y = 0$, then $\mathbf{P}(X = 0 \mid Y) = 1$, $\mathbf{E}(X \mid Y) = 0$, and $\mathrm{Var}(X \mid Y) = 0$.

Example 10.7. Let $\Omega = \{1, 2, 3, 4, 5\}$, and $p(1) = 0.5,\ p(2) = 0.2,\ p(3) = p(4) = p(5) = 0.1$. Let

$$
X(\omega) = \begin{cases} 1, & \omega = 1;\\ 0, & \omega = 2, 3, 4, 5; \end{cases} \qquad Y(\omega) = \begin{cases} 1, & \omega = 1, 2, 3;\\ 0, & \omega = 4, 5; \end{cases} \qquad Z(\omega) = \begin{cases} 1, & \omega = 1, 2, 3, 4;\\ 0, & \omega = 5. \end{cases}
$$

Let us find the law (the joint distribution) of (X, Y) given Z. If $Z = 1$, then this event has probability

$$
\mathbf{P}(Z = 1) = p(1) + p(2) + p(3) + p(4) = 0.9,
$$

and therefore we can calculate

$$
\begin{aligned}
\mathbf{P}(X = 1, Y = 1 \mid Z = 1) &= \frac{p(1)}{\mathbf{P}(Z = 1)} = \frac{0.5}{0.9} = \frac{5}{9},\\
\mathbf{P}(X = 0, Y = 1 \mid Z = 1) &= \frac{p(2) + p(3)}{\mathbf{P}(Z = 1)} = \frac{0.3}{0.9} = \frac{1}{3},\\
\mathbf{P}(X = 0, Y = 0 \mid Z = 1) &= \frac{p(4)}{\mathbf{P}(Z = 1)} = \frac{0.1}{0.9} = \frac{1}{9}.
\end{aligned}
$$

Also, if $Z = 0$, then $X = Y = 0$ with probability 1, but

$$
\mathbf{P}(X = 0, Y = 0 \mid Z = 0) = 1.
$$

10.5. **Continuous distributions.** Assume (X, Y) have joint density $p(x, y)$. The marginal density of X is given by $p_X(x) = \int p(x, y)\mathrm{d}y$, and the marginal density of Y is given by $p_Y(y) = \int p(x, y)\mathrm{d}x$. If we know that $X = x$, then Y has *conditional density*

$$
p_{Y \mid X}(y \mid x) = \frac{p(x, y)}{p_X(x)}.
$$

Similarly, if $Y = y$, then X has *conditional density*

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

We can also calculate the conditional expectation:

$$
\mathbf{E}(Y \mid X = x) = \int y p_{Y \mid X}(y \mid x)\mathrm{d}y.
$$

Example 10.8. Take the density $p(x, y) = x + y$ for $0 \le x \le 1$ and $0 \le y \le 1$. Then

$$
p_X(x) = \int_0^1 (x + y)\mathrm{d}y = \left(xy + \frac{y^2}{2}\right)\Big|_{y=0}^{y=1} = x + \frac{1}{2},
$$

$$
p_{Y \mid X}(y \mid x) = \frac{x + y}{x + 1/2},\ 0 \le y \le 1.
$$

For example, if $x = 1/4$, then

$$
p_{Y \mid X}(y \mid x) = \frac{1/4 + y}{3/4} = \frac{1}{3} + \frac{4}{3}y,\ 0 \le y \le 1,
$$

$$\mathbf{E}\left(Y \mid X = \frac{1}{4}\right) = \int_0^1 \left(\frac{1}{3} + \frac{4}{3}y\right) y\mathrm{d}y = \left(\frac{1}{6}y^2 + \frac{4}{9}y^3\right)\Big|_{y=0}^{y=1} = \frac{1}{6} + \frac{4}{9} = \frac{11}{18}.$$

Problems

Problem 10.1. (SOA) Two random variables X and Y have joint density

$$p(x, y) = 2x,\ 0 \le x \le 1,\ x \le y \le x+1;\quad p(x, y) = 0\ \text{ otherwise.}$$

Determine the conditional variance of Y given $X = x$.

Problem 10.2. (SOA) The joint probability density of X and Y is given by

$$2e^{-x-2y},\quad x \ge 0,\ y \ge 0.$$

Calculate the variance of Y given that $X > 3$ and $Y > 3$.

Problem 10.3. Calculate $\mathbf{E}(X \mid Y)$ for two random variables X, Y, distributed as

$$\mathbf{P}((X, Y) = (1, 1)) = 0.4,\ \mathbf{P}((X, Y) = (1, 0)) = 0.5,\ \mathbf{P}((X, Y) = (0, 0)) = 0.1.$$

Problem 10.4. Find $\mathbf{E}(X^2 \mid Y = 1)$ for X and Y with joint density

$$p(x, y) = \begin{cases} x + \frac{3}{2}y^2,\ 0 \le x \le 1,\ 0 \le y \le 1; \\ 0,\ \text{otherwise.} \end{cases}$$

Problem 10.5. Suppose random variables X and Y have joint density

$$p(x, y) = c(x + 2y + 2),\quad 0 \le x \le 3,\ 0 \le y \le 2,$$

and $p(x, y) = 0$ otherwise. Find the conditional density $p_{Y|X}(y \mid x = 1)$.

Problem 10.6. (SOA) A diagnostic test for the presence of a disease has two possible outcomes: 1 for disease present and 0 for disease not present. Let X denote the disease state (0 or 1) of a patient, and let Y denote the outcome of the diagnostic test. The joint probability function of X and Y is given by

$$\mathbf{P}(X = 0,\ Y = 0) = 0.800,\quad \mathbf{P}(X = 1,\ Y = 0) = 0.050,$$
$$\mathbf{P}(X = 0,\ Y = 1) = 0.025,\quad \mathbf{P}(X = 1,\ Y = 1) = 0.125.$$

Calculate $\mathrm{Var}(Y \mid X = 1)$.

For the next three problems, consider two random variables X and Y with joint distribution

X	Y	Prob.
0	0	0.1
1	0	0.2
2	0	0.1
0	1	0.4
1	1	0.2

Problem 10.7. Find the conditional distribution and expectation of X given Y.

Problem 10.8. Find the conditional distribution and expectation of Y given X.

Problem 10.9. Find $\mathbf{E}((X + Y)^2 \mid Y)$.

Problem 10.10. For two random variables X and Y with joint density

$$p(x, y) = c(x + y),\ 0 \le x \le 1,\ 0 \le y \le 1 - x,$$

find the conditional expectation $\mathbf{E}(Y \mid X = x)$.

For the next two problems, consider independent random variables $X \sim \mathrm{Bin}(4, 0.3)$ and $Y \sim \mathrm{Poi}(5)$.

Problem 10.11. Find $\mathbf{E}(X - 2Y \mid Y)$,

Problem 10.12. Find $\mathbf{E}((X - Y)^2 \mid X)$.

Problem 10.13. Consider two continuous random variables (X, Y) distributed uniformly on the triangle $\{(x, y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ 0 \le y \le x\}$. That is, their joint density is equal to

$$p(x, y) = \begin{cases} 2,\ 0 \le x \le 1,\ 0 \le y \le x; \\ 0,\ \text{else.} \end{cases}$$

Find $\mathbf{E}(X^2 \mid Y = y)$.

11. Discrete-Time Markov Chains

11.1. **Definitions.** Assume there was only one fast-food restaurant in a small town: McDonalds. Every day customers went to this restaurant. But then another fast-food place, a Subway, opened there. People started to switch between these two places. Each next day, 20% of customers who went to McDonalds the day before, went to Subway. Conversely, 10% of customers who went to Subway the day before switched to McDonalds.

On the 0th day, before Subway opened, 100% of customers went to McDonalds and 0% to Subway. On the 1st day, 80% of customers went to McDonalds and 20% went to Subway. On the 2nd day, $0.8\cdot 0.8+0.2\cdot 0.1 = 0.66 = 66\%$ of customers went to McDonalds and $0.8 \cdot 0.2 + 0.2 \cdot 0.9 = 0.34 = 34\%$ of customers went to Subway. If at day n the share mc_n of customers went to McDonalds and the share sub_n went to Subway, then

$$\begin{cases} \mathrm{mc}_{n+1} = 0.8\,\mathrm{mc}_n + 0.1\,\mathrm{sub}_n, \\ \mathrm{sub}_{n+1} = 0.2\,\mathrm{mc}_n + 0.9\,\mathrm{sub}_n\,. \end{cases}$$

We can write this as

$$\begin{bmatrix} \mathrm{mc}_{n+1} & \mathrm{sub}_{n+1} \end{bmatrix} = \begin{bmatrix} \mathrm{mc}_n & \mathrm{sub}_n \end{bmatrix} A, \quad \text{where} \quad A = \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}$$

is called the *transition matrix*. Consider a sequence of random variables

$$X = (X_n)_{n\ge 0} = (X_0, X_1, X_2, \ldots)$$

which evolves according to the following laws. Each random variable takes values M and S. The *initial distribution* is given by

$$\mathbf{P}(X_0 = M) = 1,\ \mathbf{P}(X_0 = S) = 0.$$

The distribution of X_{n+1} given X_n is given by

$$\begin{cases} \mathbf{P}(X_{n+1} = M \mid X_n = M) = 0.8, \\ \mathbf{P}(X_{n+1} = S \mid X_n = M) = 0.2, \\ \mathbf{P}(X_{n+1} = M \mid X_n = S) = 0.1, \\ \mathbf{P}(X_{n+1} = S \mid X_n = S) = 0.9. \end{cases}$$

Then we have:

$$\mathbf{P}(X_n = M) = \mathrm{mc}_n,\ \mathbf{P}(X_n = S) = \mathrm{sub}_n\,.$$

In other words, the vector $x(n) = \begin{bmatrix} \mathrm{mc}_n & \mathrm{sub}_n \end{bmatrix}$ is the *distribution* of X_n at time n. The random process $X = (X_0, X_1, \ldots)$ is called a *Markov chain*.

11.2. **Stationary and limiting distributions.** Does $x(n)$ converge to something as $n \to \infty$? Actually, it does. To find this limit, let us find nonzero vectors $v = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ such that $vA = \lambda v$ for some real number λ. These λ and v are called *eigenvalues* and *eigenvectors*. Usually, it is formulated in terms of column vectors rather than row vectors, but here it is convenient for us to study row vectors, and it does not really make any substantial difference.

We can rewrite this as $v(A - \lambda I_2) = 0$, where I_2 is the identity 2×2 matrix. Since we multiply a matrix by a nonzero vector and get zero vector, the matrix $A - \lambda I_2$ must be *nonsingular*: $\det(A - \lambda I_2) = 0$. But

$$\det(A - \lambda I_2) = \begin{vmatrix} 0.8-\lambda & 0.2 \\ 0.1 & 0.9-\lambda \end{vmatrix} = 0 \quad\Rightarrow\quad (0.8-\lambda)(0.9-\lambda) - 0.1\cdot 0.2 = 0 \ \Rightarrow\ \lambda = 1, 0.7$$

These are called *eigenvalues*. Let us find a vector v corresponding to each of these eigenvalues. For $\lambda = 1$, we have:

$$v = vA \ \Rightarrow\ \begin{cases} v_1 = 0.8v_1 + 0.1v_2 \\ v_2 = 0.2v_1 + 0.9v_2 \end{cases} \ \Rightarrow\ 2v_1 = v_2 \ \Rightarrow\ v = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

For $\lambda = 0.7$, we have:

$$0.7w = wA \ \Rightarrow\ \begin{bmatrix} 0.7w_1 = 0.8w_1 + 0.1w_2 \\ 0.7w_2 = 0.2w_1 + 0.9w_2 \end{bmatrix} \ \Rightarrow\ w_1 + w_2 = 0 \ \Rightarrow\ w = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

Every vector $x(0)$ can be decomposed as a linear combination of these two vectors. For example, if $x(0) = \begin{bmatrix} 1 & 0 \end{bmatrix}$ (initially all customers went to McDonalds), then

$$x(0) = c_1 v + c_2 w.$$

How do we find c_1 and c_2? Solve the system of equations

$$\begin{cases} 1 = c_1 + c_2 \\ 0 = 2c_1 - c_2 \end{cases} \ \Rightarrow\ c_1 = \frac{1}{3},\ \ c_2 = \frac{2}{3}.$$

Therefore, $x(0) = \frac{1}{3}v + \frac{2}{3}w$, and

$$x(n) = x(0)A^n = \frac{2}{3}0.7^n w + \frac{1}{3}v \to \frac{1}{3}v = \begin{bmatrix}\frac{1}{3} & \frac{2}{3}\end{bmatrix} =: p.$$

This is the *limiting distribution*. Also, it is a *stationary distribution*: if $x(0) = p$, then $x(1) = p$, because $p = pA$. If we start from this distribution, then we forever remain in this distribution. Any stationary distribution must be a solution to this equation $p = pA$, and the sum $p_1 + p_2 = 1$ (because total probability is 1). Here, p is a unique solution to this problem, that is, a unique probability distribution.

Actually, p is a limit regardless of the initial distribution. Indeed, suppose that $x(0)$ were different. Then this would change only c_1 and c_2: the matrix A would remain the same, with its eigenvalues and eigenvectors $\lambda_1, \lambda_2, v, w$. So $x(n) \to c_1 v$. But $x(n)$ is a probability distribution for each n, so its components sum up to 1. The same must be true for $c_1 v$, because it is the limit of $x(n)$. However, $v_1 = \begin{bmatrix}1 & 2\end{bmatrix}$, so we must have $c_1 = 1/(1+2) = 1/3$ and $c_1 v = \begin{bmatrix}1/3 & 2/3\end{bmatrix} = p$.

You can see that in the long run (actually, in a few weeks, because 0.7^{10} is already quite small) Subway will have 2/3 of customers, and McDonalds will have only 1/3 of them. In addition, each customer will spend approximately 2/3 of days in the long-term in Subway, and 1/3 in McDonalds. Say, among the first 1500 days approximately 1000 will be spent in Subway, and 500 in McDonalds. This type of statement is true if a Markov chain has a unique stationary distribution.

Assume a person always buys coffee with the meal. The coffee at McDonalds costs 1$, and at Subway it costs 2$. Then the *long-term average* cost of coffee is given by

$$1 \cdot p_1 + 2 \cdot p_2 = \frac{5}{3}.$$

This is the approximate average cost of coffee during the first N days, when N is large.

Not all Markov chains have a unique stationary distribution. For example,

$$A = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

then every distribution is stationary: $x(1) = x(0)A = x(0)$, and so $x(n) = x(0)$. The limit coincides with the stationary distribution! This corresponds to the case where all customers are completely loyal: if they went to McDonalds yesterday, they are going to make it there today, and same for Subway.

11.3. **General finite-state Markov chains.** Such a chain has *state space* $\{1, \ldots, N\}$ of N elements; if p_{ij} is the probability of moving from state i to state j, then

$$p_{i1} + \ldots + p_{iN} = 1,\ i = 1, \ldots, N.$$

We can write these *transition probabilities* in the form of an $N \times N$ transition matrix

$$P = (p_{ij})_{i,j=1,\ldots,N}.$$

The Markov chain is a collection $X = (X_n)_{n\ge 0} = (X_0, X_1, X_2, \ldots)$ of random variables X_n, each of which takes values in the state space $\{1, \ldots, N\}$. If $X_n = i$, then $X_{n+1} = j$ with probability p_{ij}. We can write this as

$$\mathbf{P}(X_{n+1} = j \mid X_n = i) = p_{ij}.$$

We denote by $x(n)$ the distribution of X_n:

$$x(n) = \begin{bmatrix}x_1(n) & \ldots & x_N(n)\end{bmatrix},\ x_i(n) = \mathbf{P}(X_n = i).$$

These components $x_1(n), \ldots, x_N(n)$ must satisfy

$$x_1(n) + \ldots + x_N(n) = 1.$$

Also, we have the following matrix multiplication formula:

$$x(n+1) = x(n)P.$$

The difference between X_n and $x(n)$ is that X_n are random variables, and $x(n)$ are vectors.

Every Markov chain (with a finite state space) has a stationary distribution (at least one, possibly more than one). But it does not necessarily converge to this stationary distribution. However, if it converges, then each probability p_i in this stationary distribution corresponds to the long-term proportion of time spent in this state i.

11.4. **Rate of convergence.** Consider the transition matrix

$$A = \begin{bmatrix} 0 & 0.3 & 0.7 \\ 0.2 & 0.8 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

It has a unique stationary distribution $p = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}$, which can be found from

$$p = pA, \quad p_1 + p_2 + p_3 = 1.$$

Let us solve this system of equations:

$$p_1 = 0.2p_2 + p_3, \quad p_2 = 0.3p_1 + 0.8p_2, \quad p_3 = 0.7p_1.$$

Therefore, from the second equation we get:

$$0.2p_2 = 0.3p_1 \ \Rightarrow \ p_2 = 1.5p_1.$$

And $p_3 = 0.7p_1$, so $1 = p_1 + p_2 + p_3 = (1.5 + 1 + 0.7)p_1 = 3.2p_1$, and $p_1 = 1/3.2 = 5/16$, $p_2 = 15/32$, $p_3 = 7/32$. What is the rate of convergence of $x(n) = x(0)A^n$, the distribution at the nth step, to this distribution? Let us find eigenvalues and eigenvectors of A. Eigenvalues:

$$\lambda_1 = 1, \quad \lambda_2 = \frac{1}{10}\left(\sqrt{57} - 1\right), \quad \lambda_3 = \frac{1}{10}\left(-\sqrt{57} - 1\right).$$

We have: $|\lambda_2| < |\lambda_3| < 1$. Let v_1, v_2, v_3 be eigenvectors corresponding to $\lambda_1, \lambda_2, \lambda_3$:

$$v_1\lambda_1 = v_1A, \quad v_2\lambda_2 = v_2A, \quad v_3\lambda_3 = v_3A.$$

We can take $v_1 = p$ because $\lambda_1 = 1$. Then for any initial distribution $x(0)$ we can decompose

$$x(0) = c_1v_1 + c_2v_2 + c_3v_3,$$

for some numbers c_1, c_2, c_3, and

$$x(n) = x(0)A^n = c_1v_1A^n + c_2v_2A^n + c_3v_3A^n = c_1p + c_2\lambda_2^n v_2 + c_3\lambda_3^n v_3.$$

As $n \to \infty$, $x(n) \to c_1p$. But we know that $x(n) \to p$, so $c_1 = 1$. Therefore,

$$x(n) - p = c_2\lambda_2^n v_2 + c_3\lambda_3^n v_3,$$

$$|x(n) - p| \le c_2|\lambda_2|^n|v_2| + |c_3||\lambda_3|^n|v_3|.$$

As $n \to \infty$, $|\lambda_2|^n$ converges to zero faster than $|\lambda_3|^n$, because $|\lambda_2| < |\lambda_3|$. Therefore, the whole expression

$$c_2|\lambda_2|^n|v_2| + |c_3||\lambda_3|^n|v_3|$$

converges to zero with the same rate as $|c_3||\lambda_3|^n|v_3|$. We say that the rate of convergence is $|\lambda_3|^n$ (because the rest are just constants, independent of n).

11.5. **Recurrence and transience.** Let $f_i := \mathbf{P}(S \text{ ever returns to the state } i \mid S_0 = i)$ be the probability that the Markov chain returns to the position i if it started from i. The state i is called *recurrent* if $f_i = 1$, and *transient* $f_i < 1$. If the state i is recurrent, then the process returns to i. But then, starting again from there, it again returns to i, and again, etc. So the process returns to the recurrent state infinitely often. If N_i is the number of times when the process returns to i, then $N_i = \infty$, and $\mathbf{E}N_i = \infty$.

For a transient state i, with probability $1 - f_i$ the process never returns there. With probability f_i, it does return there. After this, with probability $f_i(1 - f_i)$, it never returns there again. With probability f_i^2, it returns there for the second time. With probability $f_i^2(1 - f_i)$, it never returns there again, etc. So

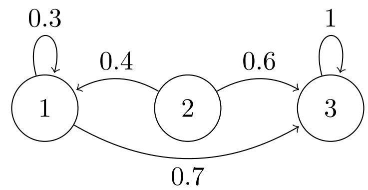
$$\mathbf{P}(N_i = k) = f_i^k(1 - f_i), \quad k = 0, 1, 2, \ldots$$

Therefore, N_i has geometric distribution with parameter $1 - f_i$, and $\mathbf{E}N_i = (1 - f_i)^{-1} < \infty$. If we start from a state and then return there infinitely many times, this state is recurrent; otherwise, it is transient.

Example 11.1. Consider the following Markov chain:

$$A = \begin{bmatrix} 0.3 & 0 & 0.7 \\ 0.4 & 0 & 0.6 \\ 0 & 0 & 1 \end{bmatrix}$$

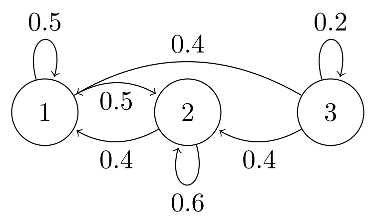
We can draw the diagram



Then you will eventually get to state 3; states 1 and 2 are transient, state 3 is recurrent. This chain is irreducible and aperiodic; therefore, it is ergodic. Let $p = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}$ be the stationary distribution. We can solve for the stationary distribution: $p = pA$, and find $p = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$.

Example 11.2. Consider the following Markov chain:

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.4 & 0.6 & 0 \\ 0.4 & 0.4 & 0.2 \end{bmatrix}$$



Stationary distribution $p = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}$ is determined from the equation $p = pA$ and the additional condition $p_1 + p_2 + p_3 = 1$. But the state 3 is transient, therefore $p_3 = 0$ (see the previous example), and so we have the system of equations

$$\begin{cases} p_1 = 0.5p_1 + 0.4p_2 + 0.4p_3 \\ p_2 = 0.5p_1 + 0.6p_2 + 0.4p_3 \\ p_3 = 0.2p_3 \end{cases} \Rightarrow\ 0.5p_1 = 0.4p_2.$$

Because $p_1 + p_2 + p_3 = 1$, we have: $p_1 + p_2 = 1$. Therefore, $p_1 = 4/9$, and $p_2 = 5/9$.

11.6. **Irreducibility.** Take a Markov chain and ignore its transient states, because the Markov chain will eventually leave them forever. In other words, suppose all states are recurrent. Take two states i, j. Suppose we can get from i to j with positive probability (in a certain number of steps): we write this as $i \to j$. Then we can also get from j to i with positive probability: $j \to i$. Otherwise, j would serve as a "sink": you leave i for j and you do not return to i with positive probability. Then the probability of return to i, starting from i, is less than 1, and this means i is transient. So if $i \to j$, then $j \to i$; we write this as $i \leftrightarrow j$ and call these states *communicating*.

We can split the state space into a few classes of communicating states. If there is more than one class, this chain is called *reducible*, and if there is only one class, it is called *irreducible*. Each such class can be considered as a Markov chain of its own. [6]

If the chain is reducible, there is more than one stationary distribution. Indeed, suppose the Markov chain has a state space $\{1, 2, 3, 4, 5\}$, where $\{1, 3, 5\}$ form an irreducible class with stationary distribution

$$\pi_1 = \frac{1}{2},\ \pi_3 = \frac{3}{8},\ \pi_5 = \frac{1}{8},$$

and $\{2, 4\}$ form an irreducible class with stationary distribution

$$\pi_2 = \frac{1}{3},\ \pi_4 = \frac{2}{3}.$$

Then there are many stationary distributions for the whole chain: add the first distribution multiplied by p_1 and the second distribution multiplied by p_2, where $p_1 + p_2 = 1$, $p_1, p_2 \ge 0$. Therefore,

$$\pi = \begin{bmatrix} \frac{1}{2}p_1 & \frac{1}{3}p_2 & \frac{3}{8}p_1 & \frac{2}{3}p_2 & \frac{1}{8}p_1 \end{bmatrix}$$

You can view it as follows: we toss a coin which has Heads with probability p_1 and Tails with probability p_2. If Heads, we move to the class $\{1, 3, 5\}$ and start from distribution $\begin{bmatrix} \pi_1 & \pi_3 & \pi_5 \end{bmatrix}$, and forever remain in this distribution

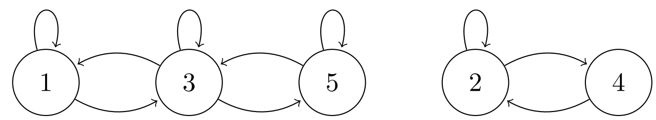
[6]Some textbooks use a different definition of irreducibility, without removal of transient states. However, we shall utilize the one given above, because it is related to stationary distributions, as shown below.

(because it is stationary). If Tails, we move to the class $\{2, 4\}$ and start from the distribution $\begin{bmatrix}\pi_2 & \pi_4\end{bmatrix}$, and forever remain in this distribution. For example, take $p_1 = 3/4$ and $p_2 = 1/4$; then

$$\pi = \begin{bmatrix}\frac{3}{8} & \frac{1}{12} & \frac{9}{32} & \frac{1}{6} & \frac{3}{32}\end{bmatrix}$$

We might also take $p_1 = 0$ and $p_2 = 1$; then we only move in the class $\{2, 4\}$, and

$$\pi = \begin{bmatrix}0 & \frac{1}{3} & 0 & \frac{2}{3} & 0\end{bmatrix}$$



We showed that if a Markov chain is reducible, then there are multiple stationary distributions. (Because there exists a stationary distribution for each communicating class, and we can combine them as above in many ways.) We can also show the following theorem.

Theorem 11.1. *If a Markov chain is irreducible, then there is only one stationary distribution.*

11.7. **Aperiodicity.** Again, ignore transient states, and suppose the Markov chain is irreducible (or, in case it is reducible, move to any class of communicating states). Start with some element i and find the number of steps needed to get there; let $\varphi_1, \varphi_2, \ldots$ be numbers of steps for different paths. Take d, the greatest common divisor of $\varphi_1, \varphi_2, \ldots$. This is called the *period* of this Markov chain. It does not depend on i. This means we can split the state space into d subclasses such that if you start in the 1st subclass, at the next step you get to the 2nd one, etc. to the dth one, and then to the first one. If $d = 1$, the Markov chain is called *aperiodic*, and if $d \ge 2$, it is called *periodic*. If there is a loop (there exists a state from which you can return to itself at the next step), then $\varphi_1 = 1$, and $d = 1$.

As mentioned before, there exists a unique stationary distribution p. However, if $d \ge 2$ (the chain is periodic), then it might not converge to this stationary distribution p. That is, not necessarily $x(n) \to p$ as $n \to \infty$. [7]

Example 11.3. Consider a Markov chain with the state space $\{1, 2\}$ and the transition matrix

$$A = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$$

This chain is 2-periodic, and its stationary distribution p satisfies

$$p = pA \;\Rightarrow\; \begin{cases} p_1 = p_2 \\ p_2 = p_1 \end{cases} \;\Rightarrow\; p_1 = p_2 = \frac{1}{2}.$$

But if $x(0) = \begin{bmatrix}1/3 & 2/3\end{bmatrix}$, then $x(1) = x(0)A = \begin{bmatrix}2/3 & 1/3\end{bmatrix}$, $x(2) = x(1)A = \begin{bmatrix}1/3 & 2/3\end{bmatrix}$, etc. So

$$x(n) = x(0) \;\text{ for even }\; n, \quad x(n) = x(1) \;\text{ for odd }\; n.$$

Therefore, $x(n)$ does not have a limit. The distribution switches between $x(0)$ and $x(1)$.

Theorem 11.2. *If the Markov chain is irreducible and aperiodic, then $x(n) \to p$ for $n \to \infty$, regardless of the initial distribution $x(0)$. Moreover, $p_i > 0$ for each state i.*

In this case, the Markov chain is called *ergodic*.

11.8. **Summary.** When we analyze a Markov chain X, the following steps are required:

Identify all transient and all recurrent states: For every state, you should find whether it is transient or recurrent. Then remove all transient states. You get a certain Markov chain Y, with all states being recurrent.

Is the resulting Markov chain Y irreducible? That is, can you get from any state to any state? If yes, the Markov chain is irreducible. If not, it splits into more than one communicating classes $C^1, \ldots, C^r$. Each such class is by itself an irreducible Markov chain, with all its states being recurrent.

Each such communicating class has a unique stationary distribution. In fact, an irreducible Markov chain has a unique stationary distribution. If the class C^i has a stationary distribution π^i, then we can weigh them (as explained earlier in this section) with certain weight coefficients $p_1, \ldots, p_r$ which satisfy

$$p_1, \ldots, p_r \ge 0, \; p_1 + \ldots + p_r = 1,$$

[7] Some textbooks use another definition of aperiodicity, without removal of transient states. However, we shall remove transient states first, because the connection to stationary distributions becomes clearer.

and get a stationary distribution π for the Markov chain Y. If a Markov chain is reducible (that is, if it consists of more than one communicating class), then it has more than one (in fact, infinitely many) stationary distributions, because $p_1, \ldots, p_r$ can be chosen in multiple ways.

Every stationary distribution of Y corresponds to a stationary distribution of X: add zeroes in the places corresponding to the removed transient states. Every stationary distribution has zeroes in components corresponding to every transient state.

The stationary distribution for an irreducible Markov chain $Z = (Z_n)_{n\ge 0}$ has the meaning of long-term time proportions spent in this state. Take the state i, and let π_i be the corresponding component of the stationary distribution π. Let $N_i(n)$ be the number of steps $k = 1, \ldots, n$ spent in i: $Z_k = i$. Then we have:

$$\frac{N_i(n)}{n} \to \pi_i \text{ a.s. as } n \to \infty.$$

In particular, if i is a transient state, then the Markov chain spends only finitely many steps in i. Say it spends M_i steps there. Then $N_i(n) = M_i$ for large enough n, so $N_i(n)$ does not grow as $n \to \infty$. Thus,

$$\frac{N_i(n)}{n} \to 0 \text{ a.s. as } n \to \infty.$$

However, if i is a recurrent state, then $\pi_i > 0$. In other words, the long-term proportion of steps spent at i is positive. During the first n steps for a large n, the Markov chain spends approximately $\pi_i n$ steps in the state i.

Finally, fix a communicating class; or, alternatively, assume the Markov chain is irreducible and does not contain any transient states. We can find its period d by fixing the state i, and taking the gcd of lengths of all paths returning from i to i, as explained above. This number is the same for all states i. Fix a state i; it is in a subclass D_0. All states j where you can go from i in one step belong to the next subclass D_1. All states where you can get in one step from some states in D_1 form the subclass D_2, etc. The whole Markov chain now is split into d subclasses $D_0, \ldots, D_{d-1}$, and the dynamics is as follows:

$$D_0 \Rightarrow D_1 \Rightarrow D_2 \Rightarrow \ldots \Rightarrow D_{d-1} \Rightarrow D_0.$$

If an irreducible Markov chain is aperiodic ($d = 1$), then for every initial distribution $x(0)$, we have convergence to the stationary distribution:

$$x(n) \to \pi,\ n \to \infty.$$

However, if this Markov chain is periodic ($d \ge 2$), then for some initial distribution $x(0)$ there is no such convergence. For example, consider the initial distribution concentrated on D_0. That is, $X_0 \in D_0$: the initial state is among the states from the subclass D_0. Then at the next step the process goes to the subclass D_1: $X_1 \in D_1$. Next,

$$X_2 \in D_2,\ X_3 \in D_3,\ \ldots, X_{d-1} \in D_{d-1},\ X_d \in D_0,\ X_{d+1} \in D_1, \ldots$$

For each step n, only some of the components of the distribution $x(n)$ are positive: those who correspond to the respective class. Others are zero. But for the stationary distribution π, all components are positive. So we cannot have convergence $x(n) \to \pi$ as $n \to \infty$.

Every comunicating class has its own period; they might be different.

The overall scheme is as follows:

(1) For each state, find whether it is recurrent or transient. Remove all transient states.
(2) Split the resulting Markov chain (with only recurrent states left) into communicating classes. If there is only one, the chain is irreducible; otherwise, it is reducible.
(3) For each communicating class, find its period and split into subclasses. If the period is 1, this class is aperiodic; otherwise, it is periodic.

11.9. **Time spent in transient states.** Assume we have a Markov chain

$$\begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}$$

States 1 and 2 are transient. If we start from state 1, then every attempt to escape succeeds with probability 0.4 and fails with probability 0.6. And once it leaves state 1, it never returns there. Therefore, time T_1 spent in 1 is distributed geometrically with parameter 0.4. In particular,

$$\mathbf{P}(T_1 = n) = 0.4 \cdot 0.6^{n-1},\ n = 1, 2, \ldots; \quad \mathbf{E}T_1 = \frac{1}{0.4} = 1.25.$$

Similarly, if we start from state 2, then time T_2 spent in 2 is $\mathrm{Geo}(0.5)$, and $\mathbf{E}T_2 = 1/0.5 = 2$.

11.10. **Comparison of hitting times.** Consider the Markov chain with four states $1, 2, 3, 4$ and transition matrix

$$\begin{bmatrix} 0.5 & 0.3 & 0.2 & 0 \\ 0.3 & 0.4 & 0.2 & 0.1 \\ * & * & * & * \\ * & * & * & * \end{bmatrix}$$

Let p_x be the probability that, starting from x, the process will reach 3 before it reaches 4. By definition,

$$p_3 = 1, \quad p_4 = 0.$$

$$\begin{cases} p_1 = 0.5p_1 + 0.3p_2 + 0.2p_3; \\ p_2 = 0.3p_1 + 0.4p_2 + 0.2p_3 + 0.1p_4 \end{cases}$$

Derive the first equation: When we start from 1, we can move back to 1 with probability 0.5, and then we start from 1 again; or we can move to 2 with probability 0.3, then we start from 2, etc. We can rewrite these equations as

$$0.5p_1 = 0.3p_2 + 0.2, \quad 0.6p_2 = 0.3p_1 + 0.2.$$

It suffices to solve them and find p_1 and p_2.

11.11. **Random walk on graphs.** Consider a graph with finitely many vertices $1, 2, \ldots, N$. A Markov chain starts from some vertex and chooses its neighbors with equal probability. We write $i \leftrightarrow j$ if the vertices i and j are connected. Let $\deg i$ be the degree of the vertex i (the number of neighbors for i). Therefore, the transition matrix is given by

$$A_{ij} = \begin{cases} (\deg i)^{-1}, & i \leftrightarrow j; \\ 0, & \text{otherwise}. \end{cases}$$

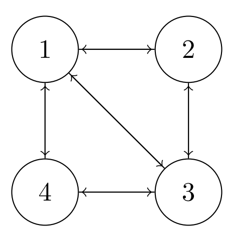
Let us take the stationary distribution

$$\pi_i := \frac{\deg i}{2M}, \; i = 1, \ldots, N.$$

where M is the total number of edges. Let us check that this is indeed a stationary distribution. For each vertex j, we have:

$$(\pi A)_j = \sum_i \pi_i A_{ij} = \sum_{i \leftrightarrow j} \pi_i A_{ij} = \sum_{i \leftrightarrow j} \frac{\deg i}{2M} \frac{1}{\deg i} = \frac{1}{2M} \sum_{i \leftrightarrow j} 1 = \frac{\deg i}{2M}.$$

For example, the following graph:



has $M = 5$, and stationary distribution

$$p_1 = p_3 = \frac{3}{10}, \; p_2 = p_4 = \frac{2}{10} = \frac{1}{5}.$$

This Markov chain (or, rather, its analogue for directed graph) is important for Google PageRank search algorithm: see section on simulation.

If the graph is connected, then the random walk on this graph is irreducible and has a unique stationary distribution. However, if the graph is disconnected, then the random walk is reducible and has more than one stationary distribution.

11.12. **Application: a Bonus-Malus system.** An insurance company classifies car drivers into three tiers: Tier 0 consists of the best drivers, tier 1 consists of not-so-good drivers, and tier 2 consists of bad drivers. This is the table which shows the change between current year and next year, depending on the number of accidents this year:

Tier	0	1	2	≥ 3
0	0	1	1	2
1	0	1	2	2
2	1	2	2	2

Say, if a driver from the first tier had no accidents this year, this driver moves to the zeroth tier next year. Assume now that the number of accidents is always $X \sim \mathrm{Poi}(2)$ for the first two tiers and $Y \sim \mathrm{Poi}(3)$ for the worst tier. Then

$$\mathbf{P}(X = 0) = e^{-2},\ \mathbf{P}(X = 1) = \mathbf{P}(X = 2) = 2e^{-2},\ \mathbf{P}(X \ge 3) = 1 - 5e^{-2},$$
$$\mathbf{P}(Y = 0) = e^{-3},\ \mathbf{P}(Y = 1) = 3e^{-3},\ \mathbf{P}(Y = 2) = \frac{9}{2}e^{-3},\ \mathbf{P}(Y \ge 3) = 1 - \frac{17}{2}e^{-3}.$$

The transition matrix is

$$P = \begin{bmatrix} e^{-2} & 4e^{-2} & 1 - 5e^{-2} \\ e^{-2} & 2e^{-2} & 1 - 3e^{-2} \\ 0 & e^{-2} & 1 - e^{-2} \end{bmatrix}$$

We can find the stationary distribution for this Markov chain: $\begin{bmatrix} p_0 & p_1 & p_2 \end{bmatrix}$ by solving the system $pP = 0$. If r_k is the premium for tier k, then the long-term average premium is $p_0 r_0 + p_1 r_1 + p_2 r_2$.

Problems

For the next two problems, consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.2 & 0.4 & 0 & 0.4 \\ 0.3 & 0 & 0.7 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.1 & 0.9 & 0 \end{bmatrix}$$

Problem 11.1. Which states are transient and which are recurrent? Is this Markov chain irreducible or reducible? What is its period?

Problem 11.2. Find the distribution of X_2 given that $X_0 = 1$.

For the next three problems, consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.1 & 0.9 \\ 1 & 0 \end{bmatrix}$$

Problem 11.3. Which states are transient and which are recurrent? Is this Markov chain irreducible or reducible? What is its period?

Problem 11.4. Find its stationary distribution.

Problem 11.5. Find its rate of convergence.

For the next two problems, consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.1 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0 & 0 & 0.1 & 0.9 \end{bmatrix}$$

Its eigenvalues and eigenvectors are given by

$$\lambda_1 = 0.659,\ \lambda_2 = 0.8,\ \lambda_3 = 1,\ \lambda_4 = 0.941.$$
$$v_1 = \begin{bmatrix} -0.271 & 0.653 & -0.653 & 0.271 \end{bmatrix}$$
$$v_2 = \begin{bmatrix} -1 & 1 & 1 & -1 \end{bmatrix}$$
$$v_3 = \begin{bmatrix} -1 & -1 & -1 & -1 \end{bmatrix}$$
$$v_4 = \begin{bmatrix} -0.653 & -0.271 & 0.271 & 0.653 \end{bmatrix}$$

Problem 11.6. Find its stationary distribution. If the payoff for the states 1 and 4 is 2.5, and for the states 2 and 3 is 0, find the long-term average payoff.

Problem 11.7. Find its rate of convergence.

For the next three problems, consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \end{bmatrix}$$

Problem 11.8. Which states are transient and which are recurrent? Is this Markov chain irreducible or reducible? What is its period?

Problem 11.9. Find its stationary distributions.

Problem 11.10. Assuming that the Markov chain starts from $X_0 = 2$, find $\mathbf{P}(X_2 = 3)$.

For the next four problems, consider a Markov chain $X = (X_n)_{n\ge 0}$ with state space $\{1, 2, 3\}$ and transition matrix

$$A = \begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.3 & 0 & 0.7 \\ 0.1 & 0.9 & 0 \end{bmatrix}$$

You can take as given the eigenvalues of this matrix:

$$\lambda_1 = 1,\ \lambda_2 \approx -0.13,\ \lambda_3 \approx -0.787.$$

Assume the initial distribution is given by

$$\mathbf{P}(X_0 = 1) = 0.2,\ \mathbf{P}(X_0 = 2) = 0.5,\ \mathbf{P}(X_0 = 3) = 0.3,$$

Problem 11.11. Find $\mathbf{P}(X_1 = 2)$.

Problem 11.12. Find $\mathbf{P}(X_2 = 3, X_3 = 2)$.

Problem 11.13. Find the stationary distribution π for this Markov chain. Assign to each state $1, 2, 3$ the benefit 0.4, 0.5, 0.1. Find the long-term average benefit.

Problem 11.14. Find the rate of convergence to this stationary distribution.

For the next three problems, consider a Markov chain $X = (X_n)_{n\ge 0}$ with state space $\{1, 2, 3, 4\}$ and transition matrix

$$A = \begin{bmatrix} 0.3 & 0 & 0.7 & 0 \\ 0.4 & 0.1 & 0.4 & 0.1 \\ 0.3 & 0 & 0.7 & 0 \\ 0 & 0.2 & 0.4 & 0.4 \end{bmatrix}$$

Problem 11.15. Find the distribution $x(2)$ of X_2, if the initial distribution is

$$x(0) = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$$

Problem 11.16. Find the stationary distribution for this Markov chain.

Problem 11.17. Find transient and recurrent states.

For the next four problems, consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Problem 11.18. Which states are transient and which are recurrent? Is this Markov chain irreducible or reducible? What is its period?

Problem 11.19. Find its stationary distribution.

Problem 11.20. Find the distribution of X_1, X_2, X_3, and X_4 if $X_0 = 1$.

Problem 11.21. Same question if $X_0 = 2$.

For the next two problems, consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.5 & 0 & 0.5 \\ 0.1 & 0 & 0.9 \end{bmatrix}$$

Problem 11.22. Find the probability that, starting from 1, it hits 2 before 3.

Problem 11.23. Find the probability that, starting from 2, it hits 1 before 3.

For the next three problems, consider a Markov chain $(X_n)_{n\ge 0}$ with three states $1, 2, 3$:

$$P = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.1 & 0.4 & 0.5 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$

Problem 11.24. Calculate the probability $\mathbf{P}(X_2 = 1 \mid X_0 = 1)$.

Problem 11.25. For the initial distribution $x(0) = [0.6, 0.4, 0]$, find the distribution of X_1.

Problem 11.26. Find the stationary distribution.

Problem 11.27. Consider a Markov chain with transition matrix

$$A = \begin{bmatrix} 0.2 & 0.8 \\ 0.7 & 0.3 \end{bmatrix}$$

Find its stationary distribution and rate of convergence.

Problem 11.28. Take a Markov chain on the state space $\{1, 2, 3\}$ with transition matrix

$$A = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Which states are recurrent and which are transient? How many communicating classes does it have, and what is the period for each of them?

For the next three problems, consider the Markov chain $(X_n)_{n\ge 0}$ on the state space $\{1, 2\}$ with

$$A = \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix}$$

Problem 11.29. Find $\mathbf{P}(X_2 = 1)$, assuming that the initial distribution is given by

$$\mathbf{P}(X_0 = 1) = 0.4, \quad \mathbf{P}(X_0 = 2) = 0.6.$$

Problem 11.30. Find its stationary distribution.

Problem 11.31. Find its rate of convergence.

Problem 11.32. Consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.5 & 0 & 0.5 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

Find the probability that, starting from 1, it hits 2 before 3.

Problem 11.33. Consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0 & 0.4 & 0.6 \\ 0 & 0 & 1 \end{bmatrix}$$

Find the mean time spent at 2, starting from 1.

Problem 11.34. Invent a Markov chain with 7 states, 3 of which are transient, 4 are recurrent, such that these recurrent states are split into two communicating classes, one of which has period 1, the other has period 2.

Problem 11.35. Consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.7 & 0.3 \end{bmatrix}$$

Let T be the time (number of steps) spent at 1 if you start this Markov chain from 1. Which distribution does T have? Find $\mathbf{E}T$ and $\operatorname{Var} T$.

Problem 11.36. Consider the Markov chain with transition matrix

$$A = \begin{bmatrix} 0.4 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Find transient and recurrent states.

In the next three problems, we take a graph. Consider the random walk on this graph. Write its transition matrix A and find all stationary distributions. Which states are transient and which are recurrent? Is this Markov chain irreducible or reducible? What is its period?

Problem 11.37. Take the graph with three vertices, 1, 2, and 3, such that 1 is connected with 2 and 3, but 2 and 3 are not connected.

Problem 11.38. Take the graph with three vertices, 1, 2, and 3, but now each two vertices are connected (that is, 1 is connected to 2 and to 3, and 2 is connected to 3).

Problem 11.39. Take the graph with four vertices, 1, 2, 3, 4, such that there are two edges: between 1 and 2, and between 3 and 4.

For the next two problems, assume we have a Bonus-Malus system of three tiers with the table below. Assume the premium paid in the kth tier is $0.5k + 1$ (measured in thousands of dollars).

Tier	0	1	≥ 2
0	0	1	1
1	0	1	2
2	1	2	2

Problem 11.40. For every tier, the distribution of accidents is given by $Z - 1$, where $Z \sim \mathrm{Geo}(0.5)$. Find the transition matrix, and its stationary distribution. Find the long-term average premium.

Problem 11.41. Each person has an accident each month with probability $(k + 1)/10$, where k is the tier. All months are independent. Find the transition matrix, and its stationary distribution. Find the long-term average premium.

12. Random Walk and Derivatives Pricing

12.1. **Construction.** Consider a sequence $X_1, X_2, \ldots$ of i.i.d. (independent identically distributed) random variables with distribution $\mathbf{P}(X_i = 1) = p$ and $\mathbf{P}(X_i = -1) = q$, $p + q = 1$. Take

$$S_0 := x, \text{ and } S_n = S_{n-1} + X_n, \ n = 1, 2, \ldots$$

$$S_1 = x + X_1, \ S_2 = x + X_1 + X_2, \ S_3 = x + X_1 + X_2 + X_3, \ldots$$

Each time, the particle moves either one step to the right (with probability p) or one step to the left (with probability q). Each next step is independent of the past. An example of the trajectory starting from $x = 0$ is:

$$(S_0 = 0, \ S_1 = 1, \ S_2 = 2, \ S_3 = 1, \ S_4 = 2, \ S_5 = 3, \ldots)$$

This process $(S_0, S_1, \ldots) = (S_n)_{n \ge 0}$ is called a *random walk*. When $p = q = 1/2$, this is called *simple* or *symmetric random walk*.

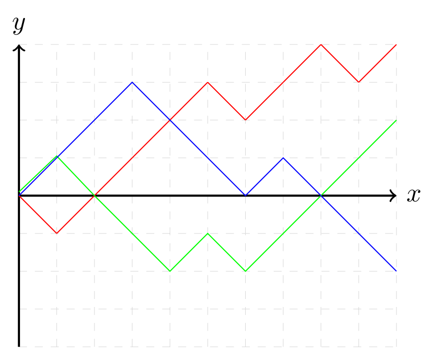
Example 12.1. Consider a random walk starting from $S_0 = -1$ with $p = 2/3$ and $q = 1/3$. Find $\mathbf{E}S_{12}$ and $\mathrm{Var}\, S_{12}$.

$$\mathbf{E}X_k = 1 \cdot p + (-1) \cdot q = \frac{1}{3},$$

$$\mathbf{E}X_k^2 = 1^2 \cdot p + (-1)^2 \cdot q = 1, \ \mathrm{Var}\, X_k = \mathbf{E}X_k^2 - (\mathbf{E}X_k)^2 = 1 - \left(\frac{1}{3}\right)^2 = \frac{8}{9}.$$

$$\mathbf{E}S_{12} = \mathbf{E}\left[-1 + X_1 + \ldots + X_{12}\right] = -1 + 12 \cdot \frac{1}{3} = \boxed{3}$$

$$\mathrm{Var}\, S_{12} = \mathrm{Var}\, X_1 + \ldots + \mathrm{Var}\, X_{12} = 12 \cdot \frac{8}{9} = \boxed{\frac{32}{3}}$$



12.2. **Transition probabilities.** The *transition probabilities* are $\mathbf{P}(S_n = z \mid S_m = y)$. Let us give an example on how to calculate them: take $\mathbf{P}(S_{10} = 4 \mid S_5 = 3)$. The quantity of paths from $(5,3)$ to $(10,4)$: suppose the path has a steps up and b steps down. The total number of steps is $a+b=5$, and $a-b=1$ (because the total ascent from 3 to 4 is 1). So $a=(5+1)/2=3$, and $b=(5-1)/2=2$. The path made 3 steps up and 2 steps down. There are five possible slots for upward steps; we must choose three upward steps, the rest will be occupied by downward steps. The number of ways to choose this is $\binom{5}{3} = (5\cdot 4)/2 = 10$. Each such path has probability p^3q^2, and so the total probability is

$$\mathbf{P}(S_{10} = 4 \mid S_5 = 3) = 10p^3q^2.$$

12.3. **Recurrence and transience.** Recall the definition of recurrent and transient states for random walk. Let N_i be the number of times when the process returns to this state. If $N_i = \infty$, and therefore $\mathbf{E}N_i = \infty$, this state is recurrent. If N_i is finite (Geometric distribution) and therefore $\mathbf{E}N_i < \infty$, it is transient. Now, let us find which of two cases holds for the random walk. Since all states are alike, let us take the state $i=0$. Let 1_A be the indicator of the event A:

$$1_A = \begin{cases} 1, & \text{if } A \text{ happened;} \\ 0, & \text{if } A \text{ did not happend} \end{cases} \qquad \text{then} \quad \mathbf{E}1_A = \mathbf{P}(A).$$

Because the random walk can return to zero only in an even number $n = 2k$ of steps.

$$N_0 = \sum_{n=1}^{\infty} 1_{\{S_n = 0\}}, \quad \mathbf{E}N_0 = \sum_{n=1}^{\infty} \mathbf{P}(S_n = 0) = \sum_{k=1}^{\infty} \mathbf{P}(S_{2k} = 0).$$

Let us calculate $\mathbf{P}(S_{2k} = 0)$. To get in $2k$ steps from 0 to 0, the process has to make k steps up and k steps down. The number of ways to choose k steps upward is $\binom{2k}{k}$. Each such path has probability p^kq^k. Therefore,

$$\mathbf{P}(S_{2k} = 0) = \binom{2k}{k} p^k q^k.$$

Now, we need to find when the series

$$\sum_{k=1}^{\infty} \binom{2k}{k} (pq)^k$$

converges or diverges. We write $a_k \sim b_k$ if $a_k/b_k \to 1$ as $k \to \infty$. By the Stirling formula, we have:

$$k! \sim \sqrt{2\pi k}(k/e)^k.$$

Therefore,

$$\binom{2k}{k} = \frac{(2k)!}{k!^2} \sim \frac{\sqrt{2\pi \cdot 2k}(2k/e)^{2k}}{\left(\sqrt{2\pi k}(k/e)^k\right)^2} = \frac{2^{2k}}{\sqrt{\pi k}}.$$

The series

$$\sum_{k=1}^{\infty} \frac{2^{2k}}{\sqrt{\pi k}} (pq)^k = \sum_{k=1}^{\infty} \frac{(4pq)^k}{\sqrt{\pi k}}$$

converges if $4pq < 1$, because of the geometric series:

$$\sum_{k=1}^{\infty} \frac{(4pq)^k}{\sqrt{\pi k}} \le \sum_{k=1}^{\infty} (4pq)^k < \infty.$$

And it diverges if $4pq = 1$, because

$$\sum_{k=1}^{\infty} \frac{1}{\sqrt{\pi k}} = \infty.$$

Now, if $p = q = 1/2$, then $4pq = 1$, and otherwise $4pq = 4p(1-p) < 1$. Indeed,

$$1 - 4p(1-p) = 1 - 4p + 4p^2 = (1-2p)^2 \ge 0,$$

and this equals to zero only if $p = 1/2$. Result: **for the symmetric random walk, every state is recurrent; otherwise, every state is transient**.

12.4. **Gambler's ruin.** Consider a random walk $X_0, X_1, X_2, \ldots$, with probability p of going up and probability q of going down. Suppose $X_0 = n$. What is the probability q_n that this random walk hits N before hitting zero? We can restate it: if a gambler wants to get N dollars, starting from n dollars, but does not want to go bankrupt before that, what is the probability that he can pull this out?

Let us deduce this for $N = 3$, $p = .25$, $q = .75$. We will simulateously deduce this for $n = 0, 1, 2, 3$. At the first step, starting from n, the random walk can either go up (with probability p) and reach $n+1$ or go down (with probability q) and reach $n-1$. Therefore,

$$q_n = pq_{n+1} + qq_{n-1}, \quad n = 1, \ldots, N-1. \tag{17}$$

Also, $q_0 = 0$, because if the gambler already had nothing in the beginning, he is bankrupt. And $q_N = 1$, because if the gambler started with N dollars, he already achieved this goal. Plugging $n = 1$ and $n = 2$ in (17), we get:

$$q_2 = .25 \cdot q_3 + .75 \cdot q_1 = .25 + .75 \cdot q_1; \quad q_1 = .25 \cdot q_2 + .75 \cdot q_0 = .25 \cdot q_2.$$

Solving this system, we get: $q_2 = .25 + .75 \cdot .25 \cdot q_2 = 1/4 + (3/16)q_2$, therefore $q_2 = 4/13$, and $q_1 = .25q_2 = 1/13$. In general case, we get with $\alpha = q/p$:

$$q_k = \begin{cases} \frac{1-\alpha^k}{1-\alpha^N}, \ \alpha \neq 1; \\ \frac{k}{N}, \ \alpha = 1. \end{cases}$$

If $p > q$, then $\alpha < 1$, and as $N \to \infty$, we have:

$$x(k) \to 1 - \alpha^k > 0.$$

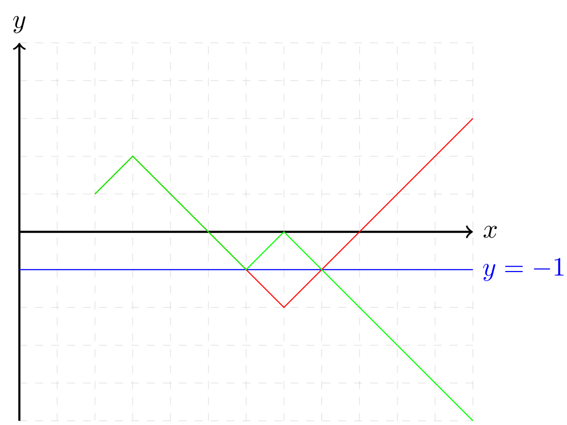
If a good gambler (who is more likely to win than lose at each step) plays against a bank with an infinite amoint of money, then the gambler can still never get bankrupt with positive probability. The above calculation works only for $\alpha \neq 1$.

12.5. **Reflection principle.** Consider $(S_n)_{n\ge 0}$, a random walk. Let us find the quantity of paths from $S_2 = 1$ to $S_{12} = 3$ such that they do not hit or cross the line $y = -1$.

First, let us find the quantity of all paths from $S_2 = 1$ to $S_{12} = 3$, that is, from $(2, 1)$ to $(12, 3)$, regardless of whether they cross or hit this line. Let n and m be the quantity of up and down steps. Then the total number of steps is $n + m = 12 - 2 = 10$, and the change in altitude is $n - m = 3 - 1 = 2$. Therefore, $n = 6$ and $m = 4$. There are $\binom{10}{6}$ ways to choose 6 upward steps out of 10, so $\binom{10}{6}$ such random walk paths.

Then, let us find the quantity of these paths that do hit or cross this line $y = -1$. Starting from this hitting point, reflect them across this line. Then we get a new random walk path, which goes from $(2, 1)$ to $(12, -5)$ (the point symmetric to $(12, 3)$ with respect to the line $y = -1$). Every path from $(2, 1)$ to $(12, 3)$ which hits $y = -1$ corresponds to a path from $(2, 1)$ to $(12, -5)$, and vice versa. Let us find the quantity of paths from $(2, 1)$ to $(12, -5)$. If it has n steps up and m steps down, then $n + m = 12 - 2 = 10$ and $n - m = 1 - (-5) = 6$. Therefore, $n = 8$ and $m = 2$. The number of such paths is $\binom{10}{8}$

Thus, the quantity of paths from $(2, 1)$ to $(12, 3)$ which do not cross the line $y = -1$ is $\binom{10}{6} - \binom{10}{8} = \frac{10\cdot 9\cdot 8\cdot 7}{4!} - \binom{10\cdot 9}{2} = 210 - 45 = 165$.



Another way to say that the path $(S_n)_{n\ge 0}$ of random walk from $(2, 1)$ to $(12, 3)$ does not cross or hit $y = -1$ is

$$S_{12} = 3, \ S_n \geq 0, \ n = 2, \ldots, 12.$$

Assume we have a random walk with $p = 0.6$ and $q = 0.4$. Every such path has 6 steps up and 4 steps down (see above), and the probability of each such path is p^6q^4. Therefore, the probability of this event is

$$\mathbf{P}\left(S_{12} = 3,\ S_n \ge 0,\ n = 2, \ldots, 12 \mid S_2 = 1\right) = \boxed{165 \cdot p^6q^4}$$

12.6. **The ballot problem.** Assume there were 1000 voters and two candidates, A and B, who got 600 and 400 votes, respectively. Therefore A won. We start counting ballots.

How many different paths of counting ballots, that is, random walks $(S_n)_{n\ge0}$ from $(0, 0)$ to $(1000, 200)$? Here, S_n is the ballots cast for A minus ballots cast for B when nth ballot has been counted. Therefore, $S_0 = 0$ and $S_{1000} = 600 - 400 = 200$. There has to be 600 ballots for A, and there are 1000 total ballots. The number of ways to choose them is $\binom{1000}{600}$.

How many such paths when A is always ahead of B? The first ballot must be for A, otherwise B immediately comes ahead. Next, how many paths from $(1, 1)$ to $(1000, 200)$ which do not hit the line $y = 0$? Similarly to the previous subsection, we can find: Each such path has n steps up and m steps down, with $n + m = 1000 - 1 = 999$, and $n - m = 200 - 1 = 199$. Therefore, $n = 599$, $m = 400$, and there are $\binom{999}{400}$ such paths.

How many such paths do cross or hit this line? Reflect it, starting from the first hitting point, across this line $y = 0$. Then we get a random walk path from $(1, 1)$ to $(1000, -200)$. It has n steps up and m steps down, with $n + m = 1000 - 1 = 999$, and $n - m = -200 - 1 = -201$. Therefore, $n = 399$ and $m = 600$. There are $\binom{999}{399}$ such paths. Therefore, the number of ways to count ballots such that A is always ahead of B is:

$$\binom{999}{400} - \binom{999}{399}$$

We can rewrite this as (because $1000! = 1000 \cdot 999!$ and $600! = 600 \cdot 599!$)

$$\frac{999!}{599!400!} - \frac{999!}{399!600!} = \frac{600}{1000}\frac{1000!}{600!400!} - \frac{400}{1000}\frac{1000!}{600!400!} = \frac{600-400}{1000}\binom{1000}{600}.$$

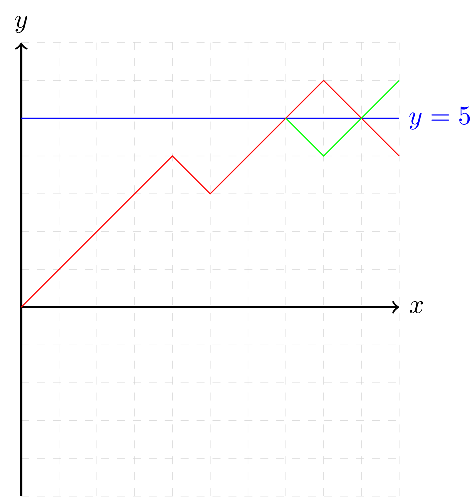
The probability that A was always ahead of B is

$$\frac{600-400}{1000} = \boxed{\frac{1}{5}}$$

12.7. **The maximum of a random walk.** Using the reflection principle, we can find the distribution of the maximum

$$M_n := \max(S_0, S_1, \ldots, S_n)$$

of the symmetric random walk, with $p = q = 0.5$. For example, let us find $\mathbf{P}(M_{10} \ge 5)$. This event $\{M_{10} \ge 5\}$ can happen in two cases: (a) $S_{10} \ge 5$; (b) $M_{10} \ge 5$ but $S_{10} < 5$. By reflection principle, the probability of the second event is equal to $\mathbf{P}(S_{10} > 5)$.



Indeed, starting from the first hitting point of the line $y = 5$, we reflect this trajectory of a random walk across this line. Thus,

$$\mathbf{P}(M_{10} \ge 5) = \mathbf{P}(S_{10} \ge 5) + \mathbf{P}(S_{10} > 5) = \mathbf{P}(S_{10} \ge 5) + \mathbf{P}(S_{10} \ge 6). \tag{18}$$

Similarly, substituting 6 instead of 5 and 7 instead of 6, we get:

$$\mathbf{P}(M_{10} \ge 6) = \mathbf{P}(S_{10} \ge 6) + \mathbf{P}(S_{10} \ge 7). \tag{19}$$

Subtract (19) from (18) and get:

$$\mathbf{P}(M_{10} = 5) = \mathbf{P}(S_{10} = 5) + \mathbf{P}(S_{10} = 6). \tag{20}$$

But S_{10} can take only even values, therefore the first term in the right-hand side of (20) is zero, and (20) takes the form

$$\mathbf{P}(M_{10} = 5) = \mathbf{P}(S_{10} = 6).$$

This we can find, as explained earlier: If the path with $S_{10} = 6$ takes a steps up and b steps down, then $a + b = 10$ and $a - b = 6$, so $a = 8$ and $b = 2$. Each given path has probability 2^{-10}. Thus

$$\mathbf{P}(S_{10} = 6) = 2^{-10}\binom{10}{2} = \frac{10 \cdot 9}{2 \cdot 1024} = 0.0439.$$

12.8. **Financial modeling.** Random walk is used for elementary modeling of movements of the stock. Assume P_n is the price of a stock at day n. Take a random walk $S = (S_n)_{n \ge 0}$ starting from $S_0 = x$, with probabilities p and q of moving up and down. Then we let

$$P_n = \exp\left(\sigma S_n + bn\right),\ n = 0, 1, 2, \dots$$

for some coefficients $\sigma > 0$ and b. The coefficient σ is called *volatility*, since it is responsible for random fluctuations. The coefficient b is called *drift*. We can alternatively write this as follows:

$$P_n = \exp\left(\sigma \sum_{k=1}^{n} X_k + bn\right) = \exp\left(\sum_{k=1}^{n} Y_k\right),$$

where $Y_k = \sigma X_k + b$. All $Y_1, Y_2, \dots$ are i.i.d. and have distribution

$$\mathbf{P}(Y_k = b + \sigma) = p,\ \mathbf{P}(Y_k = b - \sigma) = q.$$

Example 12.2. For $p = q = 0.5$ (symmetric random walk), drift $b = -2$ and volatility $\sigma = 3$, find $\mathbf{E}P_{10}$ if $P_0 = 2$. We have:

$$\mathbf{E}e^{Y_k} = 0.5e^{-2+3} + 0.5e^{-2-3} = 0.5\left(e + e^{-5}\right) = 1.363.$$

Therefore,

$$\mathbf{E}P_{10} = \mathbf{E}\left[P_0 e^{Y_1} e^{Y_2} \dots e^{Y_{10}}\right] = 2\mathbf{E}\left[e^{Y_1}\right] \cdot \ldots \cdot \mathbf{E}\left[e^{Y_{10}}\right] = 2 \cdot 1.363^{10} = 44.1.$$

Example 12.3. For $p = 0.75$, $q = 0.25$, $b = 1$ and $\sigma = 2$, given that $P_0 = 0.5$, what are the possible values of P_3, and what is the distribution of P_3? Each random variable Y_k takes values 3 and -1 with probabilities p and q, respectively. Therefore, the possible values of P_3 are

$$P_3 = e^{3+3+3}P_0 = \frac{e^9}{2},\ \text{with probability } p^3;$$

$$P_3 = e^{3+3-1}P_0 = \frac{e^5}{2},\ \text{with probability } 3p^2q;$$

$$P_3 = e^{3-1-1}P_0 = \frac{e}{2},\ \text{with probability } 3pq^2;$$

$$P_3 = e^{-1-1-1}P_0 = \frac{e^{-3}}{2},\ \text{with probability } q^3.$$

12.9. **Options and other derivatives.** Assume the price of a stock is modeled by $(P_n)_{n \ge 0}$, as above. A *European option call* is the right to buy this stock at a given price K, called the *strike*, at some future time T, called the *maturity*. A *European option put* is the right to sell this stock at the price K at time T. Let us find the values of these options.

If the stock price P_T at maturity is less than K, then it makes no sense to exercise the European option call: We can just buy the stock on the free market. Therefore, the price of the option is zero. However, the European option put is valuable, and its value is $K - P_T$. For the case $P_T > K$, the situation is reversed: The value of the European option call is $P_T - K$, and the European option put has value 0. Thus, if we denote $a_+ := \max(a, 0)$ for every real number a, then the value of the European option call is $(P_T - K)_+$, and the value of the European option put is $(K - P_T)_+$.

An *American option call/put* can be exercised at *any* time until maturity. Therefore, such option is generally more valuable than the corresponding European option. We shall discuss valuation of the American options later.

European options are particular cases of *European derivatives*, which are worth $f(P_T)$, where $f(x)$ is a certain function, and T is the maturity. For example, $f(x) = 1(x \ge a)$ is the *binary option:* pay 1 if the price at maturity P_T exceeded a, and pay nothing otherwise. A *forward contract* corresponds to $f(x) = x - K$, which is the *obligation* to buy the stock at time T for the price K. Such contracts are used by farmers and ranchers, who want to lock in good prices (to protect themselves from downside risk), and international companies (airlines, oil, steel), who want to lock in a certain exchange rate (say, be able to buy a euro next year using the current exchange rate, rather than the future exchange rate).

12.10. **Hedging a derivative.** This means buying a certain amount of this stock to be able to *replicate this derivative:* exactly match the value of this derivative at maturity. Let us illustrate this using a *one-step model:* $T = 1$, and

$$P_0 = 1, \quad P_1 = \begin{cases} 1.2 \text{ with probability } p = 40\%; \\ 0.6 \text{ with probability } q = 60\%. \end{cases}$$

We sell a European option call with maturity $T = 1$ and strike $K = 0.9$. Assume v is its current *fair value.* Then we need to use this money v to buy s stocks and c cash, and hedge the option. At time $t = 0$, the stock price is 1; therefore,

$$v = s + c. \tag{21}$$

Next, at time $t = 1$, if the stock price is 1.2, then we need:

$$1.2s + c = (1.2 - 0.9)_+ = 0.3. \tag{22}$$

Similarly, if the stock price is 0.6, then we need:

$$0.6s + c = (0.6 - 0.9)_+ = 0. \tag{23}$$

Solve these two equations: $s = 0.5$, $c = -0.3$. Thus, we need to borrow 0.3 in cash and buy 0.6 shares of stock. The fair value is $v = -0.3 + 0.5 = 0.2$. Note that this is *not* an expected value $\mathbf{E}(P_1 - K)_+$, which is equal to $0.3 \cdot 40\% + 0 \cdot 60\% = 0.12$. Rather, this is an expected value $\mathbf{E}_0(P_1 - K)_+$ with respect to the *risk-neutral probability* p_0, q_0, such that $\mathbf{E}_0 P_1 = P_0$. Let us find them:

$$\begin{cases} p_0 \cdot 1.2 + q_0 \cdot 0.6 = 1 \\ p_0 + q_0 = 1 \end{cases} \Rightarrow p_0 = \frac{2}{3},\ q_0 = \frac{1}{3} \ \Rightarrow\ \mathbf{E}_0(P_1 - K)_+ = \frac{2}{3} \cdot 0.3 + \frac{1}{3} \cdot 0 = 0.2.$$

Indeed, if we wish to get (21) as follows: multiply (22) by p_0 and (23) by q_0, and sum them; then to make the coefficient at s be equal to 1, we get: $p_0 \cdot 1.2 + q_0 \cdot 0.6 = 1$, which is the same as $\mathbf{E}_0 P_1 = P_0$. Then $v = p_0 \cdot 0.3 + q_0 \cdot 0 = \mathbf{E}_0(P_1 - K)_+$. The same can be done with any other derivative with maturity $T = 1$. Actually, the same result is true for any maturity.

This works if P_1 has only two values. If it has three or more values, or has continuous distribution (such as normal), then the system of equations in place of (22) and (23) has three or more equations, but two variables, and it does not have a solution. Therefore, not all derivatives can be replicated. Such market is called *incomplete.* There is more than one risk-neutral probabilities, because there are two equations for them, but more than two variables.

For more than one step: $T \ge 2$, we hedge for every step, starting from the end: Assume $P_n = P_0 Z_1 \ldots Z_n$, where Z_i are i.i.d. with values 2 and 0.5 with some (irrelevant) probabilities. Let $P_0 = 1$. Consider the European option call with maturity $T = 2$ and strike $K = .5$. We have three cases:

$$\begin{cases} Z_1 = Z_2 = 2: \text{ then } \quad P_2 = 4,\ (P_2 - K)_+ = 3.5; \\ Z_1 = 2,\ Z_2 = 0.5 \text{ or } Z_1 = 0.5,\ Z_2 = 2: \text{ then } \quad P_2 = 1,\ (P_2 - K)_+ = 0.5; \\ Z_1 = Z_2 = 0.5: P_2 = 0.25, \text{ then } \quad (P_2 - K)_+ = 0. \end{cases}$$

Assume we are at $P_1 = 0.5$, then hedge:

$$c + 0.5s = v,\ c + s = 0.5,\ c + 0.25s = 0 \ \Rightarrow\ v = \frac{0.5}{3}.$$

The risk-neutral probability in this case is 1/3 and 2/3 corresponding to the jumps up and down, because

$$0.5 = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 0.25.$$

Similarly, if $P_1 = 2$, then

$$c + 2s = v,\ c + s = 0.5,\ c + 4s = 3.5 \ \Rightarrow\ v = \frac{1}{3} \cdot 3.5 + \frac{2}{3} \cdot 0.5 = 1.5.$$

We have fair values corresponding to $P_1 = 0.5$ ($V_1 = 1/6$) and $P_2 = 2$ ($V_2 = 1.5$), then we can hedge on the first step, with $P_0 = 1$:

$$c + s = v,\ c + 0.5s = \frac{1}{6},\ c + 2s = 1.5.$$

At every step, we take the expectation with respect to risk-neutral probability, when $\mathbf{E}Z_1 = \mathbf{E}Z_2 = 1$, and therefore $\mathbf{E}(P_2 \mid P_1) = \mathbf{E}(P_1 Z_2 \mid P_1) = P_1 \mathbf{E}Z_2 = P_1$, and $\mathbf{E}(P_1 \mid P_0) = \mathbf{E}P_1 = \mathbf{E}Z_1 = 1$. We say that under this measure, the process $(P_n)_{n\ge 0}$ is a *martingale* (see Section 13). There is a general result (see more details in courses on stochastic finance).

Theorem 12.1. *If there exists a unique risk-neutral probability, then the market is complete: Every European derivative is replicable, and its fair price is the expectation with respect to this risk-neutral probability.*

Problems

For the next two problems, assume Peter and Alys are tossing a coin. Heads means Peter gets 1$ from Alys, Tails means the opposite. Initially, Peter has 7$, and Alys has 9$.

Problem 12.1. What is the probability that Peter becomes bankrupt if the coin is fair?

Problem 12.2. Same question for a biased coin, with probabilities $p = 0.3$ and $q = 0.7$ of Heads and Tails respectively.

Problem 12.3. Take a random walk $S = (S_n)_{n\ge 0}$ with probability $p = 1/3$ of going up, starting from $S_0 = 1$. Find the probability that this random walk hits 10 before -10.

For the next three problems, consider a random walk $(S_n)_{n\ge 0}$ with probability $p = 0.6$ going upward and probability $q = 0.4$ going downward.

Problem 12.4. Find the probability that $\mathbf{P}(S_{14} = 4 \mid S_2 = 2)$.

Problem 12.5. Assume the random walk starts at $S_0 = 0$. Find $\mathbf{E}S_3$ and $\operatorname{Var} S_3$.

Problem 12.6. Find $\mathbf{E}(S_6 - S_2)$ and $\operatorname{Var}(S_6 - S_2)$.

For the next two problems, consider a symmetric random walk $(S_n)_{n\ge 0}$, starting from $S_0 = 1$.

Problem 12.7. Find the probability that $S_{10} = 1$.

Problem 12.8. Find $\mathbf{E}S_n$ and $\operatorname{Var} S_n$.

For the next three problems, consider a simple random walk $(S_n)_{n\ge 0}$ starting from $S_0 = 0$, with $M_n = \max(S_0, S_1, \dots, S_n)$.

Problem 12.9. Find $\mathbf{P}(M_9 = 7)$.

Problem 12.10. Find $\mathbf{P}(M_{20} = 5)$.

Problem 12.11. Find $\mathbf{P}(M_{10} = 6,\ S_{10} \ge 4)$. Hint: Decompose into $\mathbf{P}(M_{10} = 6\, 4 \le S_{10} \le 5) - \mathbf{P}(S_{10} \ge 6)$ and apply reflection principle to the first term.

For the next two problems, consider a random walk $(S_n)_{n\ge 0}$ with probability $p = 0.7$ going upward and probability $q = 0.3$ going downward.

Problem 12.12. Find the probability that, starting from $S_2 = 2$, the trajectory ends at $S_{13} = 3$.

Problem 12.13. Same question, under additional condition that the trajectory does not touch $y = 1$.

Problem 12.14. Consider a symmetric random walk $(S_n)_{n\ge 0}$ starting from $S_0 = 0$. What is the probability

$$\mathbf{P}(S_{10} = 2,\ S_n > -2,\ n = 0, \dots, 10).$$

For the next two problems, assume the following setting. There are two candidates, A and B, and 500 votes. A gets 350 and B gets 150. Assume the ballots are counted one by one.

Problem 12.15. Find the total quantity of ways to count the ballots.

Problem 12.16. Find the probability that A was always ahead of B during the vote counting.

For the next two problems, consider a random walk $(S_n)_{n\ge 0}$, starting from $S_0 = 2$, with probabilties $p = 0.7$ and $q = 0.3$ of up and down jumps.

Problem 12.17. Find $\mathbf{P}(S_{15} = 3)$.

Problem 12.18. Find $\mathbf{E}(S_{14} - S_1)$ and $\mathrm{Var}(S_{14} - S_1)$.

Problem 12.19. Take a random walk with $p = 0.3$ and $q = 0.7$. Find the probability that the random walk, starting from $S_0 = 3$, comes to $S_{11} = 4$ without crossing or hitting the line $y = 5$.

For the next three problems, consider the stock market model with

$$p = 0.6,\ q = 0.4,\ b = 1.5,\ \sigma = 0.5,\ P_0 = 2.$$

Problem 12.20. Find the distribution of P_1 and of P_2.

Problem 12.21. Find $\mathbf{E}P_{10}$ and $\mathrm{Var}\,P_{10}$.

Problem 12.22. Find $\mathbf{P}(P_{10} = 2e^{15})$.

For the next five problems, consider a European option put with maturity T and strike $K = 1.6$, on the market modeled as geometric random walk, with dynamics $P_0 = 1.5$, $Z_k = 2$ or 1 with equal probabilities.

Problem 12.23. Find its fair price and the replicating strategy for $T = 1$.

Problem 12.24. Find its fair price and the replicating strategy for $T = 1$.

Problem 12.25. Find its fair price for $T = 3$ via the risk-neutral probability.

Problem 12.26. Hedge a binary option with strike $K = 1.6$ and maturity $T = 2$ from the same market as above.

Problem 12.27. Find the fair price for a European option call with $T = 3$, $P_0 = K = 1$, $P_1 = 2$ or 0.3.

13. Discrete-Time Martingales

13.1. **Definitions.** A process $X = (X_n)_{n\ge 0}$ is called a *martingale* if for each $n = 0, 1, 2, \ldots$, we have:

$$\mathbf{E}(X_{n+1} \mid X_0, \ldots, X_n) = X_n. \tag{24}$$

That is, if the best prediction of the next value, given all history, is the current value. If we have

$$\mathbf{E}(X_{n+1} \mid X_0, \ldots, X_n) \ge X_n, \tag{25}$$

then the process is called a *submartingale*. If we have:

$$\mathbf{E}(X_{n+1} \mid X_0, \ldots, X_n) \ge X_n,$$

then the process is called a *supermartingale*.

Example 13.1. Take independent random variables $Z_1, Z_2, \ldots$. Let

$$X_0 := 0, \quad \text{and} \quad X_n := Z_1 + \ldots + Z_n \quad \text{for} \quad n = 1, 2, \ldots$$

Then $X_{n+1} = X_n + Z_{n+1}$. Therefore,

$$\mathbf{E}(X_{n+1} \mid X_0, \ldots, X_n) = \mathbf{E}(X_n \mid X_0, \ldots, X_n) + \mathbf{E}(Z_{n+1} \mid X_0, \ldots, X_n) = X_n + \mathbf{E}Z_{n+1}.$$

because Z_{n+1} is independent of $X_0, \ldots, X_n$. Thus, $(X_n)_{n\ge 0}$ is a martingale if and only if all $\mathbf{E}Z_1 = \mathbf{E}Z_2 = \ldots = 0$. It is a submartingale if all $\mathbf{E}Z_k \ge 0$, and a supermartingale if all $\mathbf{E}Z_k \le 0$.

Example 13.2. Again, take independent random variables $Z_1, Z_2, \ldots$. Let

$$Y_n := e^{X_n}$$

Then $Y_{n+1} = Y_n e^{Z_{n+1}}$. Therefore,

$$\mathbf{E}(Y_{n+1} \mid Y_0, \ldots, Y_n) = \mathbf{E}(Y_n e^{Z_{n+1}} \mid Y_0, \ldots, Y_n) = Y_n \mathbf{E}(e^{Z_{n+1}} \mid Y_0, \ldots, Y_n) = Y_n \mathbf{E}\left[e^{Z_{n+1}}\right],$$

because Z_{n+1} is independent of $Y_0, \ldots, Y_n$. Thus, $(X_n)_{n\ge 0}$ is a martingale if and only if all $\mathbf{E}e^{Z_1} = \mathbf{E}e^{Z_2} = \ldots = 1$. It is a submartingale if all $\mathbf{E}e^{Z_k} \ge 1$, and a supermartingale if all $\mathbf{E}Z_k \le 1$.

For a submartingale $X = (X_n)_{n\ge 0}$, taking expectations in (25), we get:

$$\mathbf{E}(X_{n+1}) \ge \mathbf{E}X_n.$$

Therefore,

$$\mathbf{E}X_0 \le \mathbf{E}X_1 \le \mathbf{E}X_2 \le \ldots$$

For a martingale, we have the equality. Our goal now is to show this is true for random times τ instead of fixed times n.

13.2. **Stopping times.** A random variable τ which takes values $0, 1, 2, \ldots$ is called a *stopping time* if the event $\{\tau = n\}$ depends only on $X_0, \ldots, X_n$. If X_n is the price of your stock on day n, then τ can be the moment when you decide to sell your stock. But you can decide whether to do this right now (on day n) or not based only on the current information, that is, on $X_0, \ldots, X_n$; not on X_{n+1}, for example.

Example 13.3. The moment $\tau := \min\{n : X_n > 1\}$ is a stopping time, because $\{\tau = n\}$ means that $X_0 \le 1, X_1 \le 1, \ldots, X_{n-1} \le 1, X_n > 1$. So $\{\tau = n\}$ depends only on $X_0, \ldots, X_n$.

Example 13.4. Similarly, $\tau = \min\{n : a \le X_n \le b\}$ is a stopping time, for fixed a and b. More generally, for every subset $D \subseteq \mathbb{R}$, $\tau = \min\{n : X_n \in D\}$ is a stopping time.

Example 13.5. However, $\tau = \max\{n : X_n > 1\}$ is not a stopping time. Indeed, consider the event $\{\tau = 1\}$. This means that $X_2 \le 1$, so it depends on the future values of X.

13.3. **Optional stopping theorem.** Consider a stopping time τ whch is bounded from above by some constant N: $\tau \le N$. If $X = (X_n)_{n \ge 0}$ is a martingale, then

$$\mathbf{E}X_\tau = \mathbf{E}X_N = \mathbf{E}X_0. \tag{26}$$

Indeed, τ can take values $0, 1, \ldots, N-1, N$. Therefore,

$$\mathbf{E}X_N = \sum_{n=0}^N \mathbf{E}\left[X_N 1_{\{\tau = n\}}\right], \text{ and } \mathbf{E}X_\tau = \sum_{n=0}^N \mathbf{E}\left[X_\tau 1_{\{\tau = n\}}\right]. \tag{27}$$

But $\{\tau = n\}$ depends only on $X_0, \ldots, X_n$, and $\mathbf{E}\left[X_N \mid X_0, \ldots, X_n\right] = X_n$. Therefore

$$\mathbf{E}\left[X_N 1_{\{\tau=n\}}\right] = \mathbf{E}\left[\mathbf{E}\left[X_N 1_{\{\tau=n\}} \mid X_0, \ldots, X_n\right]\right] = \mathbf{E}\left[1_{\{\tau=n\}}\mathbf{E}\left[X_N \mid X_0, \ldots, X_n\right]\right] = \mathbf{E}\left[1_{\{\tau=n\}} X_n\right]. \tag{28}$$

Summing (28) over $n = 0, \ldots, N$, and using (27), we complete the proof of (26).

For a submartingale, we have $\mathbf{E}X_0 \le \mathbf{E}X_\tau \le \mathbf{E}X_N$, and for a supermartingale, these inequalities are reversed. Assume X is the price of a stock. If it is a martingale, then the optional stopping theorem means the following: Suppose you need to sell the stock by day N. Then you cannot devise a strategy (a rule) which will make you more (expected) profit than if you simply sell the stock at the initial day. No matter how you observe and analyze the behavior of this stock, you will not gain extra profit. However, if the price is a submartingale, then you should hold it and sell it at the last day. If the price is a supermartingale, you should sell it immediately.

The condition that τ is bounded is important. Consider the symmetric random walk from Section 12:

$$X_n = Z_1 + \ldots + Z_n,\ \mathbf{P}(Z_n = \pm 1) = \frac{1}{2}.$$

Since $\mathbf{E}Z_n = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$, the process $X = (X_n)_{n \ge 0}$ is a martingale. Let $\tau = \min\{n : X_n = 1\}$. As noted in Section 12, this stopping time is well defined, since this random walk will eventually hit level 1. But this hitting can happen very late; this stopping time τ is unbounded. By construction, $X_\tau = 1$, so $\mathbf{E}X_\tau = 1 \ne 0 = \mathbf{E}X_0$. If you are an investor trying to sell the stock with price X_n (forgetting for a minute about the fact that stock prices cannot go negative), then your strategy τ is to sell the stock when it hits level 1. But until then, your investment might go far into the negative terrirory. You will need to borrow potentially unlimited amount of money, and no lender will agree to do this.

13.4. **The Jensen inequality.** Recall that $g : \mathbb{R} \to \mathbb{R}$ is a *convex function* if

$$g(\lambda x + (1-\lambda)y) \le \lambda g(x) + (1-\lambda)g(y).$$

For example, $g(x) = x$ and $g(x) = x^2$ are convex functions, while $g(x) = \sin x$ is not. Equivalently, if you connect any two points on the graph of g by a segment, then it lies above the graph. For a twice differentiable function g, it is convex if and only if its second derivative is nonnegative: $g''(x) \ge 0$ for all x. Jensen's inequality says that if g is a convex function, then

$$\mathbf{E}g(Z) \ge g(\mathbf{E}Z). \tag{29}$$

Indeed, let $m = \mathbf{E}Z$. Since g is convex, there exists a real-valued a such that for all real x we have: $g(x) - g(m) \ge a(x - m)$. (The graph of g lies above the tangent line at point $x = m$.) Plugging in $x = Z$, we have: $g(Z) - g(m) \ge a(Z - m)$. Take the expectations:

$$\mathbf{E}g(Z) - g(m) \ge a\mathbf{E}(Z - m) = 0.$$

Therefore,

$$\mathbf{E}g(Z) \ge g(m) = g(\mathbf{E}Z).$$

One example of this is a well-known fact that $\mathbf{E}Z^2 \ge (\mathbf{E}Z)^2$. This is true, and $\mathbf{E}Z^2 - (\mathbf{E}Z)^2 = \operatorname{Var} Z \ge 0$. This immediately follows from (29): jsut apply $g(x) = x^2$.

Similarly, we can show (29) for conditional expectation instead of the unconditional:

$$\mathbf{E}\left[g(Z) \mid Y_1, \ldots, Y_n\right] \ge g\left(\mathbf{E}\left[Z \mid Y_1, \ldots, Y_n\right]\right). \tag{30}$$

13.5. **Preservation of the martingale property.** Take a martingale $X = (X_n)_{n\ge 0}$. Apply a convex function g to (24). By Jensen's inequality (30), we have:

$$\mathbf{E}\left[g(X_{n+1}) \mid X_0, \ldots, X_n\right] \ge g\left(\mathbf{E}\left[X_{n+1} \mid X_0, \ldots, X_n\right]\right) = g(X_n).$$

Therefore, $g(X) = (g(X_n))_{n\ge 0}$ is a submartingale.

Let us apply this to option pricing. We already discussed European options and other European derivatives in Section 12. Recall that a *call option* is the right to buy a stock at a certain *strike price* K. A European call option has *maturity time* N, when you can *exercise* this option: demand to buy this stock at price K. If the market price S_N of this stock at time N is less than K, then you can just buy the stock at the market price and forget about your option. Then your option does not have value. However, if the market price $S_N \ge K$, then you should exercise this option, and its value is $S_N - K$. In general, the value of this option is $\max(S_N - K, 0) = g(S_N)$, where $g(x) = \max(x - K, 0)$.

An American call option is different from a European one in the following way: the former can be exercised at any time until maturity N, while the latter must be exercised at maturity. Therefore, let τ be the time you decide to exercise your American call option to get the best expected value $\mathbf{E}g(S_\tau)$. When is the best exercise time τ? This is a stopping time, since your decision to exercise at time n or not is based only on your observations until time n, but not on future observations. But the function g is convex (draw a graph and check). Frequently, the stock price X is modeled by a martingale. Then $g(S) = (g(S_n))_{n\ge 0}$ is a submartingale. By the optional stopping theorem,

$$\mathbf{E}g\left(S_\tau\right) \le \mathbf{E}g\left(S_N\right).$$

Therefore, the best time to exercise your American call option is at maturity $n = N$. In fact, American and European call options are of the same value in this case. Additional freedom to choose exercise time does not give you anything.

13.6. **Doob's inequalities.** These are generalizations of Markov's and Chebyshev's inequalities. Take a nonnegative submartingale $X = (X_n)_{n\ge 0}$ and a number $\lambda > 0$. Then

$$\mathbf{P}\left(\max_{k=0,\ldots,n} X_k \ge \lambda\right) \le \frac{\mathbf{E}X_n}{\lambda}.$$

Indeed, consider the stopping time $\tau := \min\{k = 0, \ldots, n : X_k \ge \lambda\}$, with $\tau = n$ if $X_0 < \lambda, \ldots, X_n < \lambda$. If the event

$$A = \left\{\max_{k=0,\ldots,n} X_k \ge \lambda\right\}$$

has happened, then $X_\tau \ge \lambda$, and

$$\lambda\mathbf{P}(A) \le \mathbf{E}\left(X_\tau 1_A\right).$$

But X is nonnegative, so

$$\mathbf{E}\left(X_\tau 1_A\right) \le \mathbf{E}X_\tau 1_A + \mathbf{E}X_\tau 1_{A^c} = \mathbf{E}X_\tau.$$

Because of the optional stopping theorem, $\mathbf{E}X_\tau \le \mathbf{E}X_n$. Comparing this, we get:

$$\lambda\mathbf{P}(A) \le \mathbf{E}\left(X_\tau 1_A\right) \le \mathbf{E}X_\tau \le \mathbf{E}X_n.$$

It suffices to divide by λ. A corollary: for a nonnegative submartingale or a martingale $X = (X_n)_{n\ge 0}$, if $p \ge 1$, then

$$\mathbf{P}\left(\max_{k=0,\ldots,n} X_k \ge \lambda\right) \le \frac{\mathbf{E}|X_n|^p}{\lambda^p}.$$

Indeed, the function $x \mapsto |x|^p$ is convex for $p \ge 1$, and for $x \ge 0$ it is equal to x^p, so it is nondecreasing.

Letting $p = 2$ and recalling the very first example of a martingale: $X_n = Z_1 + \ldots + Z_n$ for independent $Z_1, \ldots, Z_n$, we have the following *Kolmogorov's inequality*:

$$\mathbf{P}\left(\max\left(X_0, \ldots, X_n\right) \ge \lambda\right) \le \frac{1}{\lambda^2}\mathbf{E}X_n^2 = \frac{1}{\lambda^2}\sum_{k=1}^{N}\mathbf{E}Z_k^2.$$

Example 13.6. Simple random walk S_n starting from $S_0 = 0$ is a martingale. Estimate by Kolmogorov's inequality:

$$\mathbf{P}\left[\max_{0\le n\le 10}|S_n| \ge 5\right] \le \frac{\mathbf{E}S_{10}^2}{5^2}. \tag{31}$$

Since $\mathbf{E}S_{10} = 0$, $\mathbf{E}S_{10}^2 = \operatorname{Var} S_{10} = \operatorname{Var}(X_1 + \ldots + X_{10}) = \operatorname{Var} X_1 + \ldots + \operatorname{Var} X_{10} = 1 + \ldots + 1 = 10$, where $X_i = \pm 1$ with equal probabilities and are independent. Thus the right-hand side in (31) is $10/25 = \boxed{0.4}$

Example 13.7. Now, consider $M_n = e^{S_n - cn}$. Need to find n such that this is a martingale:

$$\mathbf{E}e^{X_1 - c} = 1 \ \Rightarrow\ \frac12\left(e^1 + e^{-1}\right) = e^c \ \Rightarrow\ c = \ln\left(e^1 + e^{-1}\right) - \ln 2.$$

Apply Doob's martingale inequality with $f(x) = x^3$. For $x > 0$, this function is convex, since $f''(x) = 6x > 0$; and only this matters, because $M_n > 0$ always. Thus for $\lambda > 0$ we have:

$$\mathbf{P}\left[\max_{0\le n\le 100} M_n \ge \lambda\right] \le \frac{\mathbf{E}\left[M_{100}^3\right]}{\lambda^3}. \tag{32}$$

Now we calculate the right-hand side of (32):

$$M_{100}^3 = e^{3S_{100}}e^{-300c} = e^{3X_1}e^{3X_2}\ldots e^{3X_{100}}e^{-300c}. \tag{33}$$

All terms in the right-hand side of (33) are independent, so we can calculate the expectation:

$$\mathbf{E}\left[M_3^{100}\right] = e^{-300c}\left(\mathbf{E}\left[e^{3X_1}\right]\right)^{100} = e^{-300c}\left(\frac12\left[e^3 + e^{-3}\right]\right)^{100}.$$

Problems

For the next three problems, find the values of parameters for which the process $S = (S_n)_{n\ge 0}$, defined as

$$S_n := X_1 + \ldots + X_n \quad \text{for} \quad n = 1, 2, \ldots, \quad \text{and} \quad S_0 := 0,$$

is a martingale, submartingale, supermartingale.

Problem 13.1. Let $X_1, X_2, \ldots$, the i.i.d. Bernoulli with $\mathbf{P}(X_i = 1) = p$ and $\mathbf{P}(X_i = 0) = q$.

Problem 13.2. Let $X_1, X_2, \ldots \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d. normal.

Problem 13.3. Let $X_n = Z_n - 2$, where $Z_n \sim \operatorname{Exp}(\lambda)$ are i.i.d.

For the next three problems, find the values of parameters for which the process $M = (M_n)_{n\ge 0}$, defined as

$$M_n := e^{2S_n - 3n},\ n = 0, 1, 2, \ldots$$

is a martingale, submartingale, supermartingale.

Problem 13.4. Same as Problem 13.1.

Problem 13.5. Same as Problem 13.2.

Problem 13.6. Same as Problem 13.3.

For the next seven problems, take a process $X = (X_n)_{n\ge 0}$ and find whether τ is a stopping time.

Problem 13.7. $\tau := \inf\{n \mid X_n \ge 2\}$.

Problem 13.8. $\tau := \inf\{n \mid X_n^2 - 3X_n + 1 = 0\}$.

Problem 13.9. $\tau : \inf\{n \mid X_{n+1} \ge 0\}$.

Problem 13.10. $\tau := \inf\{n \mid X_{n-1} \ge 0\}$.

Problem 13.11. $\tau := \sup\{n \mid X_{n-1} = 1\}$.

Problem 13.12. $\tau := \inf\{n \mid X_n + X_{n+1} = 0\}$.

Problem 13.13. $\tau := \inf\{n \mid X_n = X_{n-1}\}$.

Problem 13.14. $\tau := \min(\inf\{n \mid X_n = 3\}, 4)$.

Problem 13.15. $\tau := 4 - \min(\inf\{n \mid X_n = 3\}, 4)$.

For the next four problems, find whether the following functions are convex.

Problem 13.16. $g(x) = x^4 + x^2$.

Problem 13.17. $g(x) = x^4 - x^2$.

Problem 13.18. $g(x) = \cos x$.

Problem 13.19. $g(x) = e^{2x+1}$.

For the next two problems, consider the following setting. Toss a fair coin repeatedly. Let X_n be the number of Heads during the first n tosses for $n = 1, 2, \ldots$, and let $X_0 := 0$.

Problem 13.20. Find a constant c such that the process $(Y_n)_{n\ge 0}$ is a martingale: $Y_n := 3X_n - cn$.

Problem 13.21. Find c such that $M = (M_n)_{n\ge 0}$ is a martingale: $M_n := e^{2X_n + cn}$.

For the next five problems, consider a derivative $f(P_t)$ of the stock with price P_t at time t, which forms a martingale. Is it optimal to exercise the American version at maturity time?

Problem 13.22. $f(x) = x^2 + x$.

Problem 13.23. A put option.

Problem 13.24. A binary option.

Problem 13.25. $f(x) = 4x^3$.

Problem 13.26. $f(x) = (x-1)^3$.

For the next three problems, apply Kolmogorov's inequality to the martingale $(M_n)_{n\ge 0}$ to estimate from above:

$$\mathbf{P}\left(\max(|M_0|, \ldots, |M_{100}|) \ge \lambda\right). \tag{34}$$

Problem 13.27. $M_n = 2 + X_1 + \ldots + X_n$, $X_i \sim \mathcal{N}(0, 2)$ i.i.d.

Problem 13.28. $M_n = e^{S_n - cn}$, $S_n = -1 + X_1 + \ldots + X_n$, $X_i \sim \mathcal{N}(0, 4)$ i.i.d. and find the c such that this is a martingale.

Problem 13.29. $M_n = c^{-n} Z_1 \cdot \ldots \cdot Z_n$, $Z_i \sim \mathrm{Exp}(2)$ i.i.d. and find the c such that this is a martingale.

For the next two problems, apply Doob's martingale inequality with a given convex function $f(x)$ to estimate (34):

Problem 13.30. $M_n = e^{S_n - cn}$, $S_n = -1 + X_1 + \ldots + X_n$, $X_i \sim \mathcal{N}(0, 4)$ i.i.d., $f(x) = x^3$.

Problem 13.31. $M_n = c^{-n} Z_1 \cdot \ldots \cdot Z_n$, $Z_i \sim \mathrm{Exp}(2)$ i.i.d., $f(x) = x^4$.

14. Poisson Process and Compound Poisson Process

14.1. **Poisson process.** Fix a positive real number λ. A collection $N = (N(t),\, t \ge 0)$, of random variables with values in $\mathbb{Z}_+ := \{0, 1, 2, \ldots\}$ is called a *Poisson process* if: $N(0) = 0$, and for every $0 \le s < t$, $N(t) - N(s)$ is distributed as a Poisson random variable $\mathrm{Poi}(\lambda(t-s))$ and is independent of $N(u)$, $u \le s$. In particular,

$$N(t) \sim \mathrm{Poi}(\lambda t);\ \mathbf{P}(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \ldots$$

This process is used, among other applications, to model arrival of insurance claims, and a stock price. Since a Poisson random variable with parameter μ has mean and variance μ, we have: $\mathbf{E} N(t) = \mathrm{Var}\, N(t) = \lambda t$.

Example 14.1. Take $\lambda = 3$. What is the probability that $N(5) = 3$, given $N(2) = 1$, for the same Poisson process?

$$\mathbf{P}(N(5) = 3 \mid N(2) = 1) = \mathbf{P}(N(5) - N(2) = 2 \mid N(2) = 1) = \mathbf{P}(N(5) - N(2) = 2)$$
$$= \mathbf{P}(\mathrm{Poi}(3 \cdot 3) = 2) = \frac{9^2}{2!} e^{-9} = \frac{81}{2} e^{-9}.$$

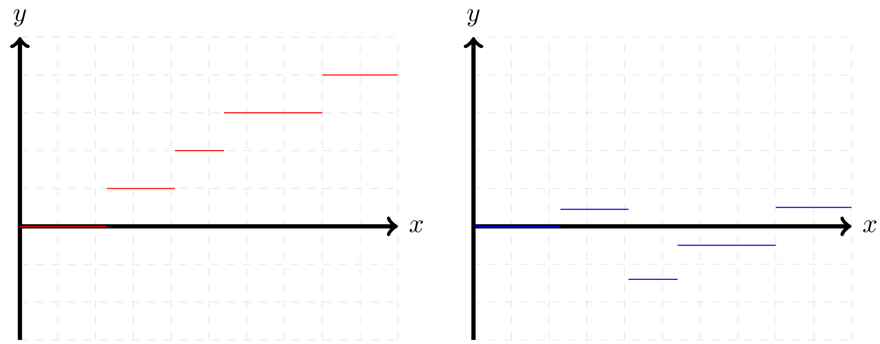
$$\mathbf{P}(N(5) = 3,\, N(2) = 1) = \mathbf{P}(N(5) = 3 \mid N(2) = 1) \cdot \mathbf{P}(N(2) = 1) = \frac{81}{2} e^{-9} \cdot \frac{6^1}{1!} e^{-6} = \boxed{243 e^{-15}}$$

Example 14.2. Take $\lambda = 3$. Then $N(2) \sim \mathrm{Poi}(6)$, therefore $\mathbf{E} N(2) = 6$ and $\mathrm{Var}\, N(2) = 6$. Thus

$$\mathbf{E} N^2(2) = \mathrm{Var}\, N(2) + (\mathbf{E} N(2))^2 = 6 + 6^2 = \boxed{42}$$

Also, by independence of $N(2)$ and $N(5) - N(2)$, we have:

$$\mathbf{E}\left[N(2) N(5)\right] = \mathbf{E} N^2(2) + \mathbf{E}\left[N(2)(N(5) - N(2))\right] = 42 + \mathbf{E} N(2) \cdot \mathbf{E}\left[N(5) - N(2)\right] = 42 + 2\lambda \cdot (5-2)\lambda = \boxed{96}$$



Left picture: Poisson process with intensity $\lambda = .5$. Right picture: Compound Poisson process with the same Poisson process (and the same trajectory) as on the left picture, and with increments distributed as $\mathcal{N}(0, 1)$

14.2. **Jump times of Poisson process.** Note that Poisson random variable is nonnegative, and so for $s \le t$, we have: $N(t) - N(s) \ge 0$. Therefore, the process N is nondecreasing: $N(t) \ge N(s)$. It starts from 0, then jumps to 1 after waiting an exponential time $\tau_1 \sim \mathrm{Exp}(\lambda)$. Why? Because the probability that this process has not yet jumped by time t is equal to the probability that $N(t) = 0$. Letting $k = 0$ into the formula above, we get: $\mathbf{P}(N(t) = 0) = e^{-\lambda t}$. Therefore,

$$\mathbf{P}(\tau_1 > t) = e^{-\lambda t}.$$

This means that $\tau_1 \sim \mathrm{Exp}(\lambda)$. This random variable has cumulative distribution function and density

$$F(t) = \mathbf{P}(\tau_1 \le t) = 1 - e^{-\lambda t},\ p(t) = F'(t) = \lambda e^{-\lambda t},\ t > 0.$$

This random variable has $\mathbf{E}\tau_1 = \lambda^{-1}$ and $\mathrm{Var}\,\tau_1 = \lambda^{-2}$.

After jumping to τ_1, the process waits some random time and jumps from 1 to 2. If τ_2 is the time when it jumps from 1 to 2, then $\tau_2 - \tau_1$ is the time which it waits between its first and second jumps. It turns out that $\tau_2 - \tau_1$ is independent from τ_1 and is also distributed as $\mathrm{Exp}(\lambda)$. The reason for this is when the process N jumps to 1, it "forgets the past" and behaves as if instead of 1 it was at 0 and nothing yet has happened. This follows from the property above: that $N(t) - N(s)$ for $t > s$ is independent of $N(u)$ for $u \le s$. This important property is called the *Markov property*, after a Russian mathematician Andrey Markov.

Similarly, if τ_k is the time of jump from $k-1$ to k, then $\tau_k - \tau_{k-1}, \tau_{k-1} - \tau_{k-2}, \ldots, \tau_2 - \tau_1, \tau_1$ are i.i.d. $\mathrm{Exp}(\lambda)$ random variables. It is possible to show that this process does not jump, say, from 0 to 2: all its jumps have magnitude 1.

Let us now find the distribution of τ_2. This is a sum of two independent random variables τ_1 and $\tau_2 - \tau_1$, both with density $p_1(t) = \lambda e^{-\lambda t}$ for $t > 0$. As mentioned in subsection 5.8, the density $p_2(x)$ of τ_2 is then the *convolution* of these exponential densities:

$$p_2(x) = \int_0^x p_1(y) p_1(x - y)\,\mathrm{d}y = \lambda^2 x e^{-\lambda x}.$$

It was calculated in subsection 5.9. More generally, the density p_k of τ_k is given by

$$p_k(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x},\ x > 0.$$

This is called the *Gamma distribution* $\Gamma(k, \lambda)$ with parameters k and λ. It has expectation and variance (because of independence):

$$\mathbf{E}\tau_k = \mathbf{E}\tau_1 + \mathbf{E}(\tau_2 - \tau_1) + \ldots + \mathbf{E}(\tau_k - \tau_{k-1}) = k\lambda^{-1},$$

$$\mathrm{Var}\,\tau_k = \mathrm{Var}\,\tau_1 + \mathrm{Var}(\tau_2 - \tau_1) + \ldots + \mathrm{Var}(\tau_k - \tau_{k-1}) = k\lambda^{-2}.$$

Example 14.3. What is the probability that the Poisson process with $\lambda = 3$, by time $t = 4$, has jumped two or more times?

$$\mathbf{P}(\tau_2 \le 4) = \mathbf{P}(N(4) \ge 2) = 1 - \mathbf{P}(N(4) = 0) - \mathbf{P}(N(4) = 1).$$

Since $N(4) \sim \mathrm{Poi}(12)$, this becomes

$$1 - \frac{12^0}{0!} e^{-12} - \frac{12^1}{1!} e^{-12} = \boxed{1 - 13e^{-12}}$$

14.3. **Compound Poisson process.** Now, we modify the Poisson process to let it jump not only by 1 upward, but in a more general way. Define a sequence of i.i.d. (independent identically distributed) random variables $Z_1, Z_2, \ldots$, independent of $N(t)$, $t \ge 0$). Then a *compound Poisson process* is defined as

$$X(t) = \sum_{k=1}^{N(t)} Z_k.$$

It starts from $X(0) = 0$, then waits time τ_1 and jumps to $X(\tau_1) = Z_1$. Next, it waits time $\tau_2 - \tau_1$ and jumps to $X(\tau_2) = Z_1 + Z_2$, then waits time $\tau_3 - \tau_2$ and jumps to $X(\tau_3) = Z_1 + Z_2 + Z_3$, etc.

This process also has the property that $X(t) - X(s)$ is independent of $X(u)$, $u \le s$ (Markov property), but it is distributed usually not as a Poisson random variable. The distribution of $X(t) - X(s)$ is the same as the distribution of $X(t-s)$, because they are sums of $N(t) - N(s)$ and, respectively, $N(t-s)$i.i.d. random variables $Z_1, Z_2, \ldots$ but $N(t) - N(s)$ and $N(t-s)$ have the same distribution.

To find the distribution of $X(t)$, let us recall the theory of Section 8: random sum of random numbers. The generating function $\varphi_t(s)$ of $N(t) \sim \mathrm{Poi}(\lambda t)$ is given by

$$\varphi_t(s) = \mathbf{E} s^{N(t)} = e^{\lambda t(s-1)}.$$

Assume $F_Z(y) = \mathbf{E} e^{yZ_i}$ is the moment generating function of each Z_i; it is independent of i, since $Z_1, Z_2, \ldots$ have the same distribution. Then by results of Section 8 (random sum of random variables), we have: the moment generating function of $X(t)$ is equal to

$$G_t(y) := \mathbf{E} e^{yX(t)} = \pi_t(F_Z(y)) = \exp\left(\lambda t(F_Z(y) - 1)\right).$$

We can also find expectation and variance, using the formulas from Section 8. Assume $\mathbf{E} Z_k = \mu$, and $\mathrm{Var}\, Z_k = \sigma^2$. Because $\mathbf{E} N(t) = \mathrm{Var}\, N(t) = \lambda t$,

$$\mathbf{E} X(t) = \mathbf{E} N(t) \cdot \mathbf{E} Z_k = \lambda \mu t, \tag{35}$$

$$\mathrm{Var}\, X(t) = \mathbf{E} N(t) \cdot \mathrm{Var}\, Z_k + \mathrm{Var}\, N(t) \cdot (\mathbf{E} Z_k)^2 = \lambda(\sigma^2 + m^2)t. \tag{36}$$

14.4. **Sampling from a Poisson process.** One can consider the Poisson process $N = (N(t),\, t \ge 0)$ with intensity λ, as the process of arrival of customers with interarrival times independent $\mathrm{Exp}(\lambda)$. Now assume that "good" customers arrive with probability p; that is, each customer is "good" with probability p, independently of other customers and of the Poisson process. Then the number of "good" customers arrived by time t is

$$M(t) = \sum_{k=1}^{N(t)} Z_k, \quad Z_k = \begin{cases} 1, & \text{if } k\text{th customer good;} \\ 0, & \text{else.} \end{cases}$$

This is a compound Poisson process with $M(t)$ having moment generating function

$$G_t(y) = \mathbf{E} e^{yM(t)} = \exp\left(\lambda t(F_Z(y) - 1)\right), \quad F_Z(y) := \mathbf{E} e^{Z_k y} = pe^y + (1-p).$$

Then we get:

$$G_t(y) = \exp\left(\lambda p t(F_Z(y) - 1)\right).$$

Therefore, $M = (M(t),\, t \ge 0)$ is itself a Poisson process, but with intensity λp.

14.5. **Sum of two Poisson processes.** Take two independent Poisson processes N_1 and N_2 with intensities λ_1 and λ_2. Then $N = N_1 + N_2$ is also a Poisson process with intensity $\lambda = \lambda_1 + \lambda_2$. Why? First, $N(0) = N_1(0) + N_2(0) = 0$. Next, for $t > s$ we have:

$$N(t) - N(s) = (N_1(t) - N_1(s)) + (N_2(t) - N_2(s)). \tag{37}$$

The expression in (37) is independent of $N(u) = N_1(u) + N_2(u)$ for $u \le s$. Indeed, $N_1(t) - N_1(s)$ is independent of $N_1(u)$, because N_1 is a Poisson process; and $N_1(t) - N_1(s)$ is independent of N_2, because these two Poisson processes are independent. Therefore, $N_1(t) - N_1(s)$ is independent from $N(u)$. Same for $N_2(t) - N_2(s)$. Finally, from (37) we get: This is the sum of two independent Poisson random variables:

$$N_1(t) - N_1(s) \sim \mathrm{Poi}(\mu_1), \quad N_2(t) - N_2(s) \sim \mathrm{Poi}(\mu_2),$$

$$\mu_1 = \lambda_1(t-s), \ \mu_2 = \lambda_2(t-s).$$

Therefore, this sum is also a Poisson random variable with parameter $\mu_1 + \mu_2 = (t-s)(\lambda_1 + \lambda_2)$. Which completes the proof that N is a Poisson process with intensity $\lambda = \lambda_1 + \lambda_2$.

14.6. **Central Limit Theorem.** We can apply Central Limit Theorem to a sum of large quantity of i.i.d. random variables, even if the sum itself is random (that is, random quantity of summands). Then we get: For a compound Poisson process, if $\mathbf{E}Z_k = \mu$ and $\operatorname{Var} Z_k = \sigma^2$,

$$\frac{X(t) - \mathbf{E}X(t)}{\sqrt{\operatorname{Var} X(t)}} \Rightarrow \mathcal{N}(0,1).$$

We can use this normal approximation to find distribution of $X(t)$ for large t, when it becomes inconvenient to use moment generating function.

Example 14.4. An insurance company receives claims as a Poisson process with $\lambda = 2$. Each claim has mean $\mu = 3$ and variance $\sigma^2 = 4$. What is the *value at risk* $\mathrm{VaR}_{95\%}$ for two years ($t = 2$) at confidence level 95%? That is, which capital does the company need to accumulate so that it can pay its obligations? From (35), (36), we have:

$$\mathbf{E}X(t) = \lambda\mu t = 12,\ \operatorname{Var} X(t) = \lambda(\mu^2 + \sigma^2)t = 52.$$

Therefore, from the Central Limit Theorem we get:

$$\frac{X(2) - 12}{\sqrt{52}} \approx \mathcal{N}(0,1).$$

Looking at the table of normal distribution, the quantile corresponding to 95% is $x_{95\%} = 1.645$. Therefore, the following event happens with probability approximately 95%:

$$\frac{X(2) - 12}{\sqrt{52}} \le x_{95\%} \ \Leftrightarrow\ X(2) \le 12 + \sqrt{52}\cdot 1.645 = \boxed{23.86}$$

14.7. **Cramer-Lundberg model.** Assume the initial capital of an insurance company is u, the constant flow rate of premiums is c, and the claims arrive according to the Poisson process $N = (N(t),\, t \ge 0)$ and have i.i.d. sizes $Z_1, Z_2, \ldots$ Then the capital of the company at time t is

$$X(t) = u + ct - \sum_{k=1}^{N(t)} Z_k.$$

The *ruin time* is the first jump when $X(t) < 0$. This can be infinite, when ruin actually never happens.

Example 14.5. Assume $u = 1$, $c = 2$, $\lambda = 2$, $Z_k \sim \mathrm{Exp}(1.5)$. Then

$$X(3) = 7 - \sum_{k=1}^{N(3)} Z_k,$$

Since $\mathbf{E}N(3) = \operatorname{Var} N(3) = 6$ and $\mathbf{E}Z_k = 1/1.5$, $\operatorname{Var} Z_k = (1/1.5)^2$, we can calculate the mean and variance:

$$\mathbf{E}X(3) = 7 - \mathbf{E}Z_k \cdot \mathbf{E}N(3) = 7 - \frac{1}{1.5}\cdot 2\cdot 3 = 3,$$

$$\operatorname{Var} X(3) = \mathbf{E}N(3)\cdot \operatorname{Var} Z_k + \operatorname{Var} N(3)\cdot(\mathbf{E}Z_k)^2 = 6\cdot\frac{1}{1.5^2} + 6\cdot\frac{1}{1.5^2} = 16/3.$$

Problems

For the next four problems, take a Poisson process $N = (N(t),\, t \ge 0)$ with intensity $\lambda = 4$.

Problem 14.1. Find the distribution of $N(5) - N(2)$, its expectation and variance.

Problem 14.2. Find the distribution of $\tau_5 - \tau_2$, its expectation and variance.

Problem 14.3. Find $\mathbf{P}(N(5) = 6 \mid N(3) = 1)$.

Problem 14.4. Show that $(N(3t),\, t \ge 0)$ is also a Poisson process, and find its intensity.

For the next three problems, take a Poisson process $N = (N(t),\, t \ge 0)$ with intensity $\lambda = 2$.

Problem 14.5. Find $\mathbf{E}\,[N(1)N(2)]$.

Problem 14.6. Find $\mathbf{P}(N(3) = 1,\, N(4) = 3)$.

Problem 14.7. Find $\mathbf{E}N^2(0.5)$.

Problem 14.8. Take two independent Poisson processes $N_1 = (N_1(t),\, t \ge 0)$ and $N_2 = (N_2(t),\, t \ge 0)$, with intensities $\lambda_1 = 3$ and $\lambda_2 = 4$. Show that $N(t) = N_1(2t) + N_2(3t)$ is a Poisson process, and find its intensity.

For the next two problems, take a compound Poisson process $S = (S(t),\, t \ge 0)$ with intensity $\lambda = 4$ and $Z \sim \mathrm{Exp}(2)$.

Problem 14.9. Find $\mathbf{E} S(t)$ and $\operatorname{Var} S(t)$.

Problem 14.10. Find the moment generating function of $S(t)$, and $\mathbf{E} S^3(t)$.

Problem 14.11. Find the value-at-risk for $S(t)$ for $t = 10$, for confidence level 99%. That is, we approximate $S(t)$ with a normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu = \mathbf{E} S(t)$, $\sigma^2 = \operatorname{Var} S(t)$. Find a number $x_{99\%}$ such that $\mathbf{P}(X \le x_{99\%}) = 99\%$.

Problem 14.12. Consider a Poisson process $N = (N(t),\, t \ge 0)$, with intensity $\lambda = 3$. Find the distribution of the random variable $\tau_4 - \tau_2$, its expectation and variance.

Problem 14.13. Consider a Poisson process $N = (N(t),\, t \ge 0)$, with intensity $\lambda = 3$. Find $\mathbf{P}(N(5/2) = 3 \mid N(1) = 1)$.

Problem 14.14. Take the compound Poisson process $S = (S(t),\, t \ge 0)$ with intensity $\lambda = 2$ and $Z \sim \mathrm{Exp}(3)$. Find $\mathbf{E} S(t)$ and $\operatorname{Var} S(t)$.

Problem 14.15. Consider a Poisson process $N = (N(t),\, t \ge 0)$ with intensity $\lambda = 5$. Let τ_3 be the first moment when this process jumps to the level 3. Find $\mathbf{E}\tau_3^3$.

Problem 14.16. Let N be the Poisson process from Problem 1, and let $Z_k \sim \mathcal{N}(-1, 3)$ be i.i.d. random variables, independent of N. Take a compound Poisson process

$$X(t) = \sum_{k=1}^{N(t)} Z_k.$$

Find the moment generating function of $X(3)$, $\mathbf{E} X(3)$, and $\operatorname{Var} X(3)$.

Problem 14.17. For a Poisson process $N = (N(t),\, t \ge 0)$ with intensity $\lambda = 2$, find the first three moments:

$$\mathbf{E} N(3),\ \mathbf{E} N^2(3),\ \mathbf{E} N^3(3).$$

Problem 14.18. For a compound Poisson process

$$X(t) = \sum_{k=1}^{N(t)} Z_k,$$

where N has intensity $\lambda = 1$ and $Z_k \sim \mathcal{N}(2, 3)$ i.i.d., find $\mathbf{E} X(t)$ and $\operatorname{Var} X(t)$.

15. Continuous-Time Markov Chains

15.1. **Properties of exponential random variables.** Recall that $\tau \sim \mathrm{Exp}(\lambda)$ is an exponential random variable with intensity λ if it has density $\lambda e^{-\lambda x}$ for $x > 0$. Then it has the property $\mathbf{P}(\tau > t) = e^{-\lambda t}$. It is commonly used to model waiting time until something happens.

1. *Memoryless property*: for $s, t \ge 0$,

$$\mathbf{P}(\tau > t + s \mid \tau > s) = \mathbf{P}(\tau > t).$$

The meaning is that if we already waited time s, the remaining wait time is distributed in the same way as if we had not waited at all. In other words, how much more we need to wait is independent of how long we already waited. (This is the only distribution which has such property.)

This property follows from

$$\mathbf{P}(\tau > t + s \mid \tau > s) = \frac{\mathbf{P}(\tau > t + s,\, \tau > s)}{\mathbf{P}(\tau > s)} = \frac{\mathbf{P}(\tau > t + s)}{\mathbf{P}(\tau > s)}$$
$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbf{P}(\tau > t).$$

2. If we have two independent exponential random variables

$$X_1 \sim \mathrm{Exp}(\lambda_1),\ \text{and}\ X_2 \sim \mathrm{Exp}(\lambda_2),$$

then $\min(X_1, X_2) \sim \mathrm{Exp}(\lambda_1 + \lambda_2)$:

$$\mathbf{P}(\min(X_1, X_2) > t) = \mathbf{P}(X_1 > t,\, X_2 > t) = \mathbf{P}(X_1 > t)\mathbf{P}(X_2 > t)$$

$$= e^{-\lambda_1 t}e^{-\lambda_2 t} = e^{-(\lambda_1+\lambda_2)t}.$$

3. If we have two independent exponential random variables $X_1 \sim \mathrm{Exp}(\lambda_1)$ and $X_2 \sim \mathrm{Exp}(\lambda_2)$, then

$$\mathbf{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}.$$

Indeed, we condition on the value of X_1:

$$\mathbf{P}(X_1 < X_2) = \int_0^\infty \mathbf{P}(X_2 > x)\lambda_1 e^{-\lambda_1 x}\,\mathrm{d}x = \int_0^\infty e^{-\lambda_2 x}\lambda_1 e^{-\lambda_1 x}\,\mathrm{d}x$$

$$= \lambda_1 e^{-(\lambda_1+\lambda_2)x}\,\mathrm{d}x = \frac{\lambda_1}{\lambda_1+\lambda_2}.$$

15.2. **Definition and construction of continuous-time Markov chains.** A continuous-time Markov chain on the state space $\{1, 2, \ldots, N\}$ is a process $X = (X(t),\, t \ge 0)$, where for every $t \ge 0$ $X(t)$ is a random variable which takes values $1, \ldots, N$, such that the behavior of $X(t)$ for $t \ge s$ is determined only by the knowledge of $X(s)$; we do not need additional knowledge of $X(u)$ for $u < s$. This is called the *Markov property*.

If the process stays at, for example, 1, then it can stay there an exponential amount of time τ_1 (with some intensity λ_1) before jumping to some other state. This is because of the Markov property: how much more we need to wait at state 1 does not depend on how long we waited to far. Therefore, every state i has an associated waiting time $\mathrm{Exp}(\lambda_i)$: an exponential "clock". When this clock rings, we have to leave i. As we leave i, we have a choice of going to other states $j \ne i$, with the probability of going from i to j being p_{ij}. The parameter $\lambda_{ij} = p_{ij}\lambda_i$ is called the *intensity of move* from i to j. We usually write them in the form of the *generating matrix,* or simply a *generator:*

$$A = (a_{ij}),\; a_{ij} = \lambda_{ij},\; i \ne j; a_{ii} = -\lambda_i.$$

This matrix has the property that its off-diagonal elements are nonnegative, and the sum of each row is equal to zero.

We can construct this Markov chain alternatively: for each state i we can generate $N-1$ independent exponential clocks $\tau_{ij} \sim \mathrm{Exp}(\lambda_{ij})$, $j \ne i$. Whichever rings first (whichever is the minimal of them) we move to the corresponding state j. This construction gives us the same result as above, because by the property 2 above we have:

$$\sum_{j\ne i}\tau_{ij} \sim \mathrm{Exp}\left(\sum_{j\ne i}\lambda_{ij}\right) = \mathrm{Exp}(\lambda_i).$$

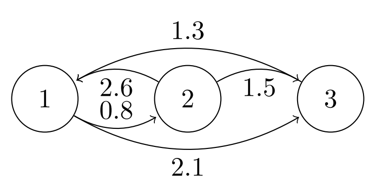
From Property 3 above, the probability that clock τ_{ij} will ring first among these clocks is

$$\mathbf{P}\left(\tau_{ij} = \min_{k\ne i}\tau_{ik}\right) = \frac{\lambda_{ij}}{\sum_{k\ne i}\lambda_{ik}} = \frac{\lambda_{ij}}{\lambda_i} = p_{ij}.$$

The parameter λ_i is called the *intensity of exit* from state i. The parameter λ_{ij} is called the *intensity of moving* from i to j.

Example 15.1. State space $\{1, 2, 3\}$, with generating matrix

$$A = \begin{bmatrix} -2.9 & 0.8 & 2.1 \\ 2.6 & -4.1 & 1.5 \\ 1.3 & 0 & -1.3 \end{bmatrix}$$



Generate two independent exponential clocks $\tau_{12} \sim \mathrm{Exp}(0.8)$, $\tau_{13} \sim \mathrm{Exp}(2.1)$. Then $\tau_1 = \min(\tau_{12}, \tau_{13}) \sim \mathrm{Exp}(2.9)$, so we will have to wait τ_1 time until leaving 1. Upon leaving 1, we will go to 2 with probability

$$\mathbf{P}(\tau_{12} < \tau_{13}) = \frac{0.8}{0.8+2.1} = \frac{0.8}{2.9}$$

and to 3 with probability

$$\mathbf{P}(\tau_{13} < \tau_{12}) = \frac{2.1}{0.8+2.1} = \frac{2.1}{2.9}.$$

Similarly with exit from 2. Exiting from 3, we will always go to 1, because the intensity of moving from 3 to 2 is 0. We can include this in our framework by letting $\tau_{32} \sim \mathrm{Exp}(0) = \infty$. (An exponential distribution $\mathrm{Exp}(\lambda)$ with intensity $\lambda = 0$ is concentrated at infinity.)

15.3. **Transition matrix.** For states i and j, let $p_{ij}(t) = \mathbf{P}(X(t) = j \mid X(0) = i)$ be the probability that the Markov chain, starting from state i, will be at time t at state j. These probabilities form the *transition matrix:* $P(t) = (p_{ij}(t))_{i,j=1,\ldots,N}$.

Let the row-vector in $\mathbb{R}^N$:

$$x(t) = \begin{bmatrix} \mathbf{P}(X(t) = 1) & \mathbf{P}(X(t) = 2) & \ldots & \mathbf{P}(X(t) = N) \end{bmatrix}$$

be the distribution of $X(t)$. If we start from the initial distribution $x(0) \in \mathbb{R}^N$, then the distribution $x(t)$ can be found as

$$x(t) = x(0)P(t). \tag{38}$$

Indeed, assume you want to be at state j at time t. The probability $\mathbf{P}(X(t) = j)$ of this is the jth element of the vector $x(t)$. Then you can achieve this by being at state i initially (which happens with probability $x_i(0) = \mathbf{P}(X(0) = i)$), and moving from state i to state j in time t (which happens with probability $p_{ij}(t)$). Summing over all $i = 1, \ldots, N$, we have:

$$x_j(t) = \sum_{i=1}^N p_{ij}(t)x_i(0),$$

which is the same as (38). We can similarly show *Chapman-Kolmorogov equations:*

$$P(t+s) = P(t)P(s) \quad \text{for} \quad t, s \ge 0.$$

Note that $p_{ii}(0) = 1$ and $p_{ij}(0) = 0$ for $i \ne j$. Therefore,

$$P(0) = I_N, \tag{39}$$

which is the $N \times N$-identity matrix.

15.4. **Forward and backward Kolmorogov equations.** Taking the derivatives entrywise, we get: $P'(t) = (p'_{ij}(t))$. The following system of *Kolmogorov equations* hold true:

$$P'(t) = P(t)A \quad \text{(forward Kolmorogov equations)} \tag{40}$$

$$P'(t) = AP(t) \quad \text{(backward Kolmorogov equations)} \tag{41}$$

Indeed, how does the matrix $P(t)$ change from t to $t + \mathrm{d}t$? In other words, how can a process get from state i to state j in time $t + \mathrm{d}t$?

Case 1. It can already be at state j by time t. This happens with probability $p_{ij}(t)$. And the probability that it will stay at state j from time t to time $t + \mathrm{d}t$ is $\mathbf{P}(\tau_j > \mathrm{d}t) = e^{-\lambda_j \mathrm{d}t} \approx 1 - \lambda_j \mathrm{d}t$. This gives us the term $p_{ij}(t)(1 - \lambda_j \mathrm{d}t)$.

Case 2. It can be at a different state $k \ne j$ by time t. This happens with probability $p_{ik}(t)$. and the probability that the process will jump from k to j during time $[t, t + \mathrm{d}t]$ is $\mathbf{P}(\tau_{kj} \le \mathrm{d}t) = 1 - e^{-\lambda_{kj}\,\mathrm{d}t} = \lambda_{kj}\,\mathrm{d}t$. This gives us the term $p_{ik}(t)\lambda_{kj}\,\mathrm{d}t$.

Summing these terms, we have:

$$p_{ij}(t + \mathrm{d}t) = p_{ij}(t)(1 - \lambda_j \mathrm{d}t) + \sum_{k \ne j} p_{ik}(t)\lambda_{kj}\,\mathrm{d}t.$$

Subtracting p_{ij}, and dividing by $\mathrm{d}t$, we get:

$$p'_{ij}(t) = \frac{p_{ij}(t + \mathrm{d}t) - p_{ij}(t)}{\mathrm{d}t} = -\lambda_j p_{ij}(t) + \sum_{k \ne j} p_{ik}(t)\lambda_{kj}(t) = \sum_{k=1}^N p_{ik}(t)a_{kj}.$$

This shows, by conditioning on $X(t)$, that $P'(t) = P(t)A$; this proves (40). One can also show (41) by conditioning on $X(\mathrm{d}t)$.

Finally, in light of (39) and (40), we have:

$$P'(0) = A. \tag{42}$$

15.5. **Stationary distribution.** Take a distribution π. Assume we start from there: $x(0) = \pi$. If the Markov chain remains there for each time $t \ge 0$: $x(t) = \pi$, then we have a *stationary distribution* p. In light of (38), we can rewrite this as

$$\pi = \pi P(t). \tag{43}$$

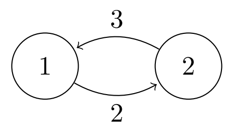
Take a derivative in (43) at $t = 0$. The left-hand side is constant, so its derivative is zero (row vector). The right-hand side by (42) has derivative πA. Therefore,

$$0 = \pi A. \tag{44}$$

Theorem 15.1. *A continuous-time Markov chain on the state space $\{1, \dots, N\}$ always has a stationary distribution.*

Example 15.2. Take the generator

$$A = \begin{bmatrix} -2 & 2 \\ 3 & -3 \end{bmatrix}$$



Solve the system of equations for $\pi = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix}$:

$$0 = \pi A \ \Rightarrow\ -2\pi_1 + 3\pi_2 = 0 \ \Rightarrow\ \pi_1 = \frac{3}{2}\pi_2.$$

But since this is a probability distribution, it should also satisfy

$$\pi_1 + \pi_2 = 1.$$

Therefore,

$$\boxed{\pi_1 = \frac{3}{5},\ \pi_2 = \frac{2}{5}}$$

15.6. **Convergence.** The vector $x(t)$, which is the distribution of $X(t)$, satisfies the ODE itself: Differentiating (38), and using (40), we have:

$$x'(t) = x(0)P'(t) = x(0)P(t)A = x(t)A. \tag{45}$$

Let us solve it, if we know the initial distribution $\pi(0)$. Assume that this Markov chain has a unique stationary distribution π; in other words, it has only one eigenvector corresponding to the eigenvalue $\lambda_1 = 0$. Other eigenvalues $\lambda_2, \dots, \lambda_N$ are negative (or are complex but have negative real parts - the analysis is the same) and have eigenvectors $u_2, \dots, u_N$:

$$u_i A = \lambda_i u_i, \quad i = 2, \dots, N.$$

Then the vectors $\pi, u_2, \dots, u_N$ form a basis of $\mathbb{R}^N$. We can find coefficients $c_1(t), \dots, c_N(t)$ such that

$$x(t) = c_1(t)\pi + c_2(t)u_2 + \ldots + c_N(t)u_N. \tag{46}$$

Let us derive the ODE for each $c_i(t)$:

$$\begin{aligned} x'(t) =& c_1'(t)\pi + c_2'(t)u_2 + \ldots + c_N'(t)u_N = c_1(t)\pi A + c_2(t)u_2 A + \ldots + c_N(t)u_N A \\ &= 0 + \lambda_2 c_2(t)u_2 + \ldots + \lambda_N c_N(t)u_N. \end{aligned} \tag{47}$$

Comparing left and right-hand sides of (47), we have:

$$c_1'(t) = 0 \ \Rightarrow\ c_1(t) = c_1(0),$$

$$c_i'(t) = \lambda_i c_i(t) \ \Rightarrow\ c_i(t) = c_i(0)e^{\lambda_i t},\ i = 2, \dots, N. \tag{48}$$

We can find $c_1(0), \dots, c_N(0)$ by decomposing the initial distribution $x(0) = c_1(0)\pi + c_2(0)u_2 + \ldots + c_N(0)u_N$ as a linear combination of the basis vectors. Then plug these initial conditions into (48). We have:

$$x(t) = c_1\pi + c_2(0)e^{\lambda_2 t}u_2 + \ldots + c_N(0)e^{\lambda_N t}u_N. \tag{49}$$

Because $\lambda_2, \dots, \lambda_N < 0$, we have: $x(t) \to c_1\pi$ as $t \to \infty$. Actually, $c_1 = 1$, because all $x(t)$ and π have the same property: sum of components equals to one. Moreover, from (49) we get:

$$|x(n) - \pi| \le c_2(0)|u_2|e^{\lambda_2 t} + \ldots + c_N(0)|u_N|e^{\lambda_N t}.$$

Each summand converges to zero with rate $e^{\lambda_i t}$. The slowest convergence is for $\lambda = \lambda_i$ with smallest absolute value (closest to zero). This gives general convergence rate $e^{\lambda t}$. To find convergence rate, take among nonzero eigenvalues the closest λ to zero, and then take $e^{\lambda t}$.

The matrix $P(t)$ can be found in the same way, because its ith row is, in fact, $x(t)$ for

$$x(0) = \begin{bmatrix} 0 & 0 & \ldots 0 & 1 & 0 & \ldots & 0 \end{bmatrix}, \tag{50}$$

where the unity is on the ith place. Indeed, $p_{ij}(t)$ is the probability of $X(t) = j$, given that $X(0) = i$; and if $X(0) = i$ with probabiltiy one, then $X(0)$ has distribution (50).

Example 15.3. A Markov chain on two states $\{1, 2\}$, with

$$A = \begin{bmatrix} -2 & 2 \\ 3 & -3 \end{bmatrix}$$

Let the initial distribution be $x(0) = \begin{bmatrix} 1 & 0 \end{bmatrix}$ Then $\pi = \begin{bmatrix} \frac35 & \frac25 \end{bmatrix}$ (found above). The eigenvalues of A: $\det(A - \lambda I_2) = 0 \Rightarrow \lambda^2 + 5\lambda = 0 \Rightarrow \lambda_{1,2} = 0, -5$. An eigenvector corresponding to $\lambda_1 = 0$ is π. An eigenvector corresponding to $\lambda_2 = -5$ can be found from

$$-5u = uA \Rightarrow \begin{cases} -5u_1 = -2u_1 + 3u_2 \\ -5u_2 = 2u_1 - 3u_2 \end{cases} \Rightarrow 3u_1 + 3u_2 = 0 \Rightarrow u = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

We can decompose $x(0) = c_1\pi + c_2 u$ by comparing componentwise:

$$\begin{cases} 1 = \frac35 c_1 + c_2 \\ 0 = \frac25 c_1 - c_2 \end{cases} \Rightarrow c_1 = 1,\ c_2 = \frac{2}{5}.$$

Then

$$x(t) = \pi + \frac{2}{5}e^{-5t}u_2 \Leftrightarrow \begin{cases} x_1(t) = \frac35 + \frac25 e^{-5t} \\ x_2(t) = \frac25 - \frac25 e^{-5t} \end{cases}$$

The rate of convergence is e^{-5t}.

15.7. **Relation between discrete- and continuous-time Markov chains.** Take a continuous-time Markov chain on three states $1, 2, 3$, with generator

$$A = \begin{bmatrix} -2.9 & 0.8 & 2.1 \\ 1.3 & -3.9 & 2.6 \\ 3 & 0 & -3 \end{bmatrix}$$

Let $\tau_0 = 0 < \tau_1 < \tau_2 < \ldots$ be the jump times. For example, if the Markov chain starts from 1: $X(0) = 1$, then $\tau_1 \sim \mathrm{Exp}(2.9)$. Next, if $X(\tau_1) = 2$ (that is, if X makes its first jump from 1 to 2), then $\tau_2 - \tau_1 \sim \mathrm{Exp}(3.9)$, etc. Let

$$Y_n := X(\tau_n),\ n = 0, 1, 2, \ldots$$

Then $Y = (Y_n)_{n\ge0} = (Y_0, Y_1, Y_2, \ldots)$ is a discrete-time Markov chain, with transition matrix

$$P = \begin{bmatrix} 0 & \frac{0.8}{2.1} & \frac{2.1}{2.9} \\ \frac{1.3}{3.9} & 0 & \frac{2.6}{3.9} \\ 1 & 0 & 0 \end{bmatrix}$$

Indeed, if the continuous-time Markov chain X jumps from 1, then it goes to 2 with probability $0.8/2.9$, and to 3 with probability $2.1/2.9$.

Remark 15.1. This discrete-time Markov chain Y, by construction, cannot go in one step from a state i to the same state i. It has to jump somewhere else (although it can return to the same state in more than one step).

When we switch from a continuous-time to this discrete-time Markov chain, we lose information about *when* jumps occurred. We only know *from where to where* the process has jumped.

We can also move backward, from a discrete-time to a continuous-time Markov chain. Assume we have a discrete-time Markov chain

$$P = \begin{bmatrix} 0 & 0.7 & 0.3 \\ 0.4 & 0 & 0.6 \\ 1 & 0 & 0 \end{bmatrix}$$

It has to have zeros on the main diagonal, which corresponds to the property from Remark 15.1. Choose any intensities $\lambda_1, \lambda_2, \lambda_3 > 0$ of exiting from states $1, 2, 3$; for example,

$$\lambda_1 = 2,\ \lambda_2 = 3,\ \lambda_3 = 4.$$

We have complete freedom in choosing these intensities; by doing this, we restore the information lost when switching from continuous time to discrete time (see above). Then the corresponding continuous-time Markov chain will have generator

$$A = \begin{bmatrix} -2 & 2\cdot 0.7 & 2\cdot 0.3 \\ 3\cdot 0.4 & -3 & 3\cdot 0.6 \\ 4 & 0 & -4 \end{bmatrix}$$

There is also a relation between stationary distributions of the discrete-time and the continuous-time Markov chains: Assume

$$\pi = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix}$$

is a stationary distribution for the discrete-time Markov chain. Then we have:

$$\pi = \pi P \ \Rightarrow \ \begin{cases} \pi_1 = 0.4\pi_2 + \pi_3 \\ \pi_2 = 0.7\pi_1 \\ \pi_3 = 0.3\pi_1 + 0.6\pi_2 \end{cases}$$

Take $\rho_i = \pi_i/\lambda_i$, $i = 1, 2, 3$, and $\rho = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 \end{bmatrix}$ Then

$$\begin{cases} -2\rho_1 + 3\cdot 0.4\rho_2 + 4\rho_3 = -\pi_1 + 0.4\pi_2 + \pi_3 = 0 \\ 2\cdot 0.7\rho_1 - 3\rho_2 = 0.7\pi_1 - \pi_2 = 0 \\ 2\cdot 0.3\rho_1 + 3\cdot 0.6\rho_2 - 4\rho_3 = 0.3\pi_1 + 0.6\pi_2 - \pi_3 = 0 \end{cases} \quad \Rightarrow \ \rho A = 0.$$

This means ρ, or, rather,

$$\rho' = \begin{bmatrix} \rho_1' & \rho_2' & \rho_3' \end{bmatrix}, \ \ \rho_i' = \frac{\rho_i}{\rho_1 + \rho_2 + \rho_3}, \ i = 1, 2, 3$$

(we divide each ρ_i to make them sum up to 1) is a stationary distribution for the continuous-time Markov chain. Conversely, if ρ is a stationary distribution for a continuous-time Markov chain, then $\pi = \begin{bmatrix} \lambda_1\rho_1 & \lambda_2\rho_2 & \lambda_3\rho_3 \end{bmatrix}$ is a stationary distribution for the corresponding discrete-time Markov chain.

There is a heuristical explanation of this relation between π and ρ. If the Markov chain is ergodic, then it converges to its stationary distribution as time goes to infinity. And the long-term share of time spent at state i is equal to π_i for the discrete-time Markov chain or ρ_i' for the continuous-time Markov chain. Assume that λ_i is very large compared to other λ_j. Then, as long as the continuous-time Markov chain jumps to i, it almost immediately jumps out of i. The share of time spent at i is small. This corresponds to a small $\rho_i = \pi_i/\lambda_i$, which comes from the large λ_i.

15.8. **Reducible and irreducible Markov chains.** There is a well-developed theory of discrete-time Markov chains, see Section 11. We can apply the same theory to continuous-time Markov chains. A state i is called *recurrent* for a continuous-time Markov chain, if it is recurrent for the corresponding discrete-time Markov chain; that is, if the probability of ever returning to i if we start from i is 1. Otherwise, it is called *transient*. After removal of all transient states, if all remaining states are connected, then the Markov chain is called *irreducible*; otherwise, it is called *reducible*, and it splits into corresponding *communicating classes*. An irreducible Markov chain (both discrete and continuous time) has a unique stationary distribution, but a reducible one has infinitely many stationary distributions.

Take a continuous-time Markov chain on $\{1, 2, 3, 4\}$ with generating matrix

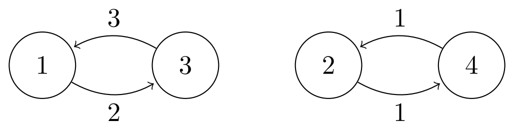
$$A = \begin{bmatrix} -2 & 0 & 2 & 0 \\ 0 & -1 & 0 & 1 \\ 3 & 0 & -3 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Then this Markov chain splits into two parts: a Markov chain with states 1 and 3, and with transition matrix

$$A_1 = \begin{bmatrix} -2 & 2 \\ 3 & -3 \end{bmatrix}$$

and a Markov chain with states 2 and 4, and with transition matrix

$$A_2 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$



These two parts are not connected to each other. We can find a stationary distribution for the first part:

$$\begin{bmatrix}\pi_1 & \pi_3\end{bmatrix} A_1 = \begin{bmatrix}0 & 0\end{bmatrix} \Rightarrow -2\pi_1 + 3\pi_3 = 0 \Rightarrow \pi_1 = \frac{3}{2}\pi_3,$$

and because $\pi_1 + \pi_3 = 1$, we have: $\pi_1 = \frac{3}{5}$, $\pi_3 = \frac{2}{5}$. Similarly, the second part has the following stationary distribution:

$$\begin{bmatrix}\pi_2 & \pi_4\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

Now, let us construct a stationary distribution for the whole Markov chain. Let p_1 be the probability that we are in part $1-3$. Denote it by p_1. Let p_2 be the probability that we are in $2-4$. Then $p_1 + p_2 = 1$ and $p_1, p_2 \geq 0$. Given that we are in $1-3$, we are at 1 with probability $3/5$ (conditional probability). Therefore, the unconditional probability of being in 1 is $(3/5)p_1$. Similarly, we can find it for other states:

$$\begin{bmatrix}\frac{3}{5}p_1 & \frac{1}{2}p_2 & \frac{2}{5}p_1 & \frac{1}{2}p_2\end{bmatrix}$$

This is an infinite family of stationary distributions. Members of this family include:

$$\begin{bmatrix}\frac{3}{5} & 0 & \frac{2}{5} & 0\end{bmatrix} \quad \text{for} \quad p_1 = 1,\, p_2 = 0;$$

$$\begin{bmatrix}0 & \frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix} \quad \text{for} \quad p_1 = 0,\, p_2 = 1;$$

$$\begin{bmatrix}\frac{3}{10} & \frac{1}{4} & \frac{1}{5} & \frac{1}{4}\end{bmatrix} \quad \text{for} \quad p_1 = p_2 = \frac{1}{2}.$$

Another way to find this would be simply to solve the system

$$\pi A = \begin{bmatrix}0 & 0 & 0 & 0\end{bmatrix} \quad \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1.$$

This would give us a free variable:

$$-2\pi_1 + 3\pi_3 = 0,\ -\pi_2 + \pi_4 = 0,\quad \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1.$$

$$\pi_1 = \frac{3}{2}\pi_3,\ \pi_4 = \pi_3,\ \frac{5}{2}\pi_3 + 2\pi_4 = 1 \Rightarrow \pi_4 = \frac{1}{2} - \frac{5}{4}\pi_3.$$

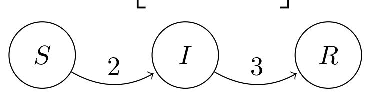
$$\begin{bmatrix}\pi_1 & \pi_2 & \pi_3 & \pi_4\end{bmatrix} = \begin{bmatrix}\frac{3}{2}\pi_3 & \frac{1}{2} - \frac{5}{4}\pi_3 & \pi_3 & \frac{1}{2} - \frac{5}{4}\pi_3\end{bmatrix}$$

$$= \pi_3\begin{bmatrix}\frac{3}{2} & -\frac{5}{4} & 1 & -\frac{5}{4}\end{bmatrix} + \begin{bmatrix}0 & \frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix}$$

This is the same answer, but in a different form.

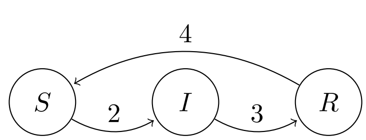
15.9. **Applications.** Consider a common infection model, with three states: S = susceptible, I = infected, R = recovered, and intensity 2 of moving from S to I, and intensity 3 from I to R; the state R is absorbing, and S, I are transient. The generator is

$$A = \begin{bmatrix}-2 & 2 & 0\\ 0 & -3 & 3\\ 0 & 0 & 0\end{bmatrix}$$



If a virus mutates and can threaten again after recovery, we can add, say, intensity 4 transition from R to S:

$$A = \begin{bmatrix}-2 & 2 & 0\\ 0 & -3 & 3\\ 4 & 0 & -4\end{bmatrix}$$



Problems

For the next three problems, take a continuous-time Markov chain on the state space $\{1, 2, 3\}$ with the generating matrix

$$A = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 3 & 2 & -5 \end{bmatrix}$$

Problem 15.1. Let T be the time this Markov chain spends at state 1 after getting there. Find the distribution of T, its expectation and variance.

Problem 15.2. As the Markov chain exits 3, what is the probability it jumps to 2?

Problem 15.3. Find its stationary distribution.

For the next five problems, consider a continuous-time Markov chain on the state space $\{1, 2\}$ with

$$A = \begin{bmatrix} -2 & 2 \\ 1 & -1 \end{bmatrix}$$

Assume it starts from the distribution $x(0) = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}$.

Problem 15.4. Let T be the time this Markov chain spends at state 2 after getting there. Find the distribution of T, its expectation and variance.

Problem 15.5. Find the distribution $x(t)$ at time t.

Problem 15.6. Find its stationary distribution and rate of convergence.

Problem 15.7. Write the transition matrix for corresponding discrete-time Markov chain.

Problem 15.8. Find the stationary distribution for that discrete-time Markov chain.

Problem 15.9. In an S-I-R model with infection rate 2 and recovery rate 3, find the stationary distribution and the distribution at time t, if we start from susceptible state.

Problem 15.10. In an S-I-R-S model with infection rate 2, recovery rate 3, and mutation rate 4, find the stationary distribution.

Problem 15.11. Consider a continuous-time Markov chain on the state space $\{1, 2, 3\}$ with the generator

$$A = \begin{bmatrix} -3 & 1 & 2 \\ 0 & -2 & 2 \\ 5 & 2 & -7 \end{bmatrix}$$

Its eigenvalues and eigenvectors are given by

$$\lambda_1 = 0, \quad \lambda_2 = -3, \quad \lambda_3 = -9$$

$$v_1 = \begin{bmatrix} 10 & 11 & 6 \end{bmatrix} \quad v_2 = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \quad v_3 = \begin{bmatrix} -5 & -1 & 6 \end{bmatrix}$$

Find its stationary distribution and rate of convergence.

Problem 15.12. Consider a continuous-time Markov chain on the state space $\{1, 2, 3, 4\}$ with the generator

$$A = \begin{bmatrix} -2 & 0 & 2 & 0 \\ 0 & -1 & 0 & 1 \\ 3 & 0 & -3 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

It has more than one stationary distributions. Find them all.

For the next seven problems, consider a continuous-time Markov chain on the state space $\{1, 2, 3, 4\}$ with the following generator:

$$A = \begin{bmatrix} -2 & 1 & 1 & 0 \\ 3 & -5 & 1 & 1 \\ 1.5 & 1.5 & -3 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix}$$

This matrix A has eigenvalues and eigenvectors

$$\lambda_1 = 0, \; v_1 = \begin{bmatrix} 8 & 4 & 3 & 1 \end{bmatrix}$$

$$\lambda_2 = -3, \; v_2 = \begin{bmatrix} -1 & 0 & 0 & -1 \end{bmatrix}$$

$$\lambda_3 = -4,\ v_3 = \begin{bmatrix} 4 & -1 & -4 & 1 \end{bmatrix}$$
$$\lambda_4 = -6,\ v_4 = \begin{bmatrix} 2 & -3 & 0 & 1 \end{bmatrix}$$

Problem 15.13. Find the stationary distribution π and the rate of convergence to π.

Problem 15.14. Let T_2 be the (random) time spent in state 2 before leaving it.

Problem 15.15. Find the distribution of T_2, its expectation and variance.

Problem 15.16. Having left state 2, what is the probability that the Markov chain jumps to 1?

Problem 15.17. Find the transition matrix for the corresponding discrete-time chain.

Problem 15.18. Find the stationary distribution for this Markov chain.

Problem 15.19. Find $x(t)$ by solving the system $x'(t) = x(t)A$, for the initial distribution

$$x(0) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

Hint: Decompose $x(0) = c_1v_1 + c_2v_2 + c_3v_3 + c_4v_4$ for some numbers c_1, c_2, c_3, c_4.

Problem 15.20. Find the stationary distribution for a continuous-time Markov chain on the state space $\{1, 2, 3\}$ with the generating matrix

$$A = \begin{bmatrix} -3 & 2 & 1 \\ 1 & -1 & 0 \\ 0 & 2 & -2 \end{bmatrix}$$

For the next three problems, consider a continuous-time Markov chain with two states 1 and 2, with generator

$$A = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

Problem 15.21. For the initial distribution $x(0) = \begin{bmatrix} 1 & 0 \end{bmatrix}$ find $x(t)$ for every $t \ge 0$.

Problem 15.22. Find the stationary distribution.

Problem 15.23. Find the rate of convergence.

For the next four problems, consider a continuous-time Markov chain on state space $\{1, 2, 3\}$ with generator

$$A = \begin{bmatrix} -2 & 1 & 1 \\ 0 & -1 & 1 \\ 1 & 2 & -3 \end{bmatrix}$$

with eigenvalues and eigenvectors

$$\lambda_1 = 0,\ \lambda_2 = -2,\ \lambda_3 = -4,$$
$$v_1 = \begin{bmatrix} 1 & 5 & 2 \end{bmatrix}\quad v_2 = \begin{bmatrix} -1 & 1 & 0 \end{bmatrix}\quad v_3 = \begin{bmatrix} -1 & -1 & 2 \end{bmatrix}$$

Problem 15.24. Find its stationary distribution.

Problem 15.25. Find the rate of convergence.

Problem 15.26. Find the transition matrix for the corresponding discrete-time Markov chain.

Problem 15.27. Find the stationary distribution for this discrete-time Markov chian.

Problem 15.28. Consider the Markov chain with generator

$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

This matrix has eigenvalues

$$\lambda_1 = 0,\ \lambda_2 = -2,\ \lambda_3 = -\sqrt{2} - 2,\ \lambda_4 = \sqrt{2} - 2,$$

and eigenvector corresponding to $\lambda_1 = 0$:

$$v_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$$

Find its stationary distribution and the rate of convergence.

Problem 15.29. Take a discrete-time Markov chain with the transition matrix

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 0.7 & 0.3 & 0 \end{bmatrix}$$

Find the generator for the corresponding continuous-time Markov chain with intensities $\lambda_1 = 3$, $\lambda_2 = \lambda_3 = 2$ of exits from the states 1, 2, and 3.

16. Queueing Theory

16.1. **M/M/1 queue.** Assume we have a server (cashier) which serves customers, one by one. Each customer is served for a random time, while others in the queue wait. This random time is independent for each customer and is distributed as $\mathrm{Exp}(\mu)$. Customers arrive (from the shop) with intensity λ; that is, the interarrival time of each next customer is distributed as $\mathrm{Exp}(\lambda)$, independently of other interarrival times, and of serving times. Let $X(t)$ be the number of customers in the queue at time t (including the customer who is currently being served). Then $X(t)$ can take values $0, 1, 2, \ldots$ Actually, this is a continuous-time Markov chain on the state space $\{0, 1, 2, \ldots\}$ with transition intensities

$$\lambda_{n,n+1} = \lambda,\ \lambda_{n,n-1} = \mu.$$

It has generating matrix

$$A = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \ldots \\ \mu & -\lambda-\mu & \lambda & 0 & \ldots \\ 0 & \mu & -\lambda-\mu & \lambda & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Let us find its stationary distribution

$$\pi = \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 & \ldots \end{bmatrix}$$

It has to satisfy $\pi A = \begin{bmatrix} 0 & 0 & 0 & \ldots \end{bmatrix}$ Therefore,

$$-\lambda\pi_0 + \mu\pi_1 = 0, \tag{51}$$

$$\lambda\pi_{n-1} - (\lambda+\mu)\pi_n + \mu\pi_{n+1} = 0,\ n \ge 1. \tag{52}$$

Let us try to find a solution to this system of equations. First, consider the case $\lambda < \mu$: the arrival intensity is less than the service intensity. Try $\pi_n = c\rho^n,\ n = 0, 1, 2, \ldots$ Then plugging into the equation (52), we get:

$$\lambda \cdot c\rho^{n-1} - (\lambda+\mu)\cdot c\rho^n + \mu \cdot c\rho^{n+1} = 0.$$

Canceling $c\rho^{n-1}$, we have: $\lambda - (\lambda+\mu)\rho + \mu\rho^2 = 0$. Solving this quadratic equation, we get: $\rho_1 = 1,\ \rho_2 = \lambda/\mu$. Therefore, we get the following solutions of (52):

$$\pi_n = c_1 \cdot 1^n = c_1,\ \text{and}\ \pi_n = c_2\left(\frac{\lambda}{\mu}\right)^n.$$

And their sum is also a solution to (52); actually, it is the most general solution:

$$\pi_n = c_1 + c_2\left(\frac{\lambda}{\mu}\right)^n. \tag{53}$$

Plug into (51) to find c_1 and c_2:

$$\pi_0 = c_1 + c_2,\ \pi_1 = c_1 + c_2\frac{\lambda}{\mu} \ \Rightarrow\ -\lambda(c_1+c_2) + \mu\left(c_1 + c_2\frac{\lambda}{\mu}\right) = (\mu-\lambda)c_1 \ \Rightarrow\ c_1 = 0.$$

Therefore, we have:

$$\pi_n = c_2\left(\frac{\lambda}{\mu}\right)^n,\quad n = 0, 1, 2, \ldots$$

Next, because of $\sum_{n=0}^{\infty}\pi_n = 1$, we have:

$$c_2\sum_{n=0}^{\infty}\left(\frac{\lambda}{\mu}\right)^n = 1 \ \Rightarrow\ c_2\left(1-\frac{\lambda}{\mu}\right)^{-1} = 1 \ \Rightarrow\ c_2 = 1 - \frac{\lambda}{\mu}.$$

$$\boxed{\pi_n = \left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n,\quad n = 0, 1, 2, \ldots}$$

This is the distribution of $N = Z - 1$, where Z has a geometric distribution $\mathrm{Geo}(\rho)$ with parameter $\rho = 1 - \lambda/\mu \in (0, 1)$. We know mean and variance of the geometric distribution. Therefore, the distribution π has mean and variance

$$\mathbf{E}N = \mathbf{E}Z - 1 = \frac{1}{\rho} - 1,\ \mathrm{Var}\,N = \mathrm{Var}\,Z = \frac{1-\rho}{\rho^2}.$$

For the case $\lambda \ge \mu$, the intensity of arrival is greater than or equal to the intensity of service. One can show that there is no stationary distribution, and the queue, on average, grows infinitely large as time goes to infinity: $\mathbf{E}X(t) \to \infty$ as $t \to \infty$.

In the stationary distribution, the outflow of customers also forms a Poisson process with intensity λ, just like the inflow (Burke's theorem). This follows from the fact that the system is in the equilibrium.

16.2. **Finite queues.** Now, assume new people do not come when N people already are in the queue. Then $X = (X(t),\, t \ge 0)$ is a continuous-time Markov chain with state space $\{0, 1, \ldots, N-1, N\}$, with generator

$$A = \begin{bmatrix} -\lambda & \lambda & 0 & \ldots & 0 & 0 \\ \mu & -\lambda-\mu & \lambda & \ldots & 0 & 0 \\ 0 & \mu & -\lambda-\mu & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ldots & -\lambda-\mu & \lambda \\ 0 & 0 & 0 & \ldots & \mu & -\mu \end{bmatrix}$$

Let us find its stationary distribution $\pi = \begin{bmatrix}\pi_0 & \pi_1 & \ldots & \pi_N\end{bmatrix}$ It has to satisfy $\pi A = \begin{bmatrix}0 & 0 & \ldots & 0\end{bmatrix}$ We can write this as (51) and (52) for $n = 1, \ldots, N-1$; the last column gives us another boundary condition

$$\lambda\pi_{N-1} - \mu\pi_N = 0. \tag{54}$$

The general solution to the difference equation (52) is given by (53). The boundary condition (51) again gives us $c_1 = 0$, so we have

$$\pi_n = c\left(\frac{\lambda}{\mu}\right)^n,\ n = 0, \ldots, N.$$

This solution also satisfies (54). And to find c, we need to use

$$\pi_0 + \ldots + \pi_N = 1\ \Rightarrow\ c\left(1 + \rho + \ldots + \rho^N\right) = 1,\quad \rho := \frac{\lambda}{\mu}. \tag{55}$$

Summing this finite geometric series, we get (for $\lambda \ne \mu\ \Leftrightarrow\ \rho \ne 1$):

$$c\frac{\rho^{N+1} - 1}{\rho - 1} = 1\ \Rightarrow\ c = \frac{\rho - 1}{\rho^{N+1} - 1}.$$

The answer is:

$$\boxed{\pi_n = \frac{\rho - 1}{\rho^{N+1} - 1}\rho^n,\ n = 0, \ldots, N}$$

Separate case: $\lambda = \mu$. Then we have:

$$\pi_{n-1} - 2\pi_n + \pi_{n+1} = 0,\ n = 1, 2, \ldots, N-1. \tag{56}$$

From (51) and (54), we have:

$$\pi_0 = \pi_1,\ \pi_{N-1} = \pi_N. \tag{57}$$

We can just let $\pi_n = 1$ for all $n = 0, \ldots, N$. Then these conditions (56) and (57) are satisfied. But we need also to normalize it, so that (55) holds. Then we get:

$$\boxed{\pi_0 = \ldots = \pi_N = \frac{1}{N+1}}$$

This stationary distribution exists even if $\lambda \ge \mu$, because the queue cannot grow indefinitely.

16.3. **Finite queues with varying intensities.** Assume we have a lemonade stand. People come there with intensity $\lambda = 2$, and are served with intensity $\mu = 2$, but with the following qualifications. If there are currently no people in the queue, then the newcomer joins the queue. If there are one or two people in the queue, the newcomer joint the queue with probability 1/2. If there are three or more people, then the newcomer does not join the queue. Let $X(t)$ be the number of people in the queue at time t. Then X is a continuous-time Markov chain with state space $\{0, 1, 2, 3\}$. When $X(t) = 1$, the intensity of jumping to 2 is $2 \cdot 0.5 = 1$. Indeed, if two people per minute on average come to the lemonade stand, but each of them chooses whether to stay or not independently with probability 0.5, then the effective rate of people joining the queue is 1 per minite. Similarly, when $X(t) = 2$, the intensity of jumping to 3 is $2 \cdot 0.5 = 1$. The intensity from 3 to 2, from 2 to 1, from 1 to 0, is 2. The generator is

$$A = \begin{bmatrix} -2 & 2 & 0 & 0 \\ 2 & -3 & 1 & 0 \\ 0 & 2 & -3 & 1 \\ 0 & 0 & 2 & -2 \end{bmatrix}$$

To find the stationary distribution $\pi = \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 \end{bmatrix}$, solve $\pi A = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$

Problems

Problem 16.1. Consdier the unbounded model of the queue: M/M/1, with $\lambda = 2$ and $\mu = 3$. Find the stationary distribution for the length of the queue, and the expectation and variance corresponding to this distribution.

Problem 16.2. Consider a queue of at most two people. Incoming intensity is $\lambda = 2$, but if there are no people, then the person joins the queue with probability 1, and if there is one person in the queue, then the newcomer joins the queue with probability 0.5. The service intensity is $\mu = 1$. Write the generator for this Markov chain, and find the stationary distribution.

Problem 16.3. Consider a queue with incoming intensity $\lambda = 2$, and service intensity μ. Let N be the number of people in this queue, under the stationary distribution. Find μ such that $\mathbf{E}N = 2$.

Problem 16.4. The incoming stream of people into the queue has intensity $\lambda = 4$. The service intensity is $\mu = 3$. However, a person does not enter the queue if there are two people or more. Find the stationary distribution of the number of people.

Problem 16.5. Consider a queue with service intensity $\mu = 4$, and incoming intensity $\lambda = 2$. Let N be the number of people in this queue in the stationary distribution. Find $\mathbf{E}N$ and $\operatorname{Var} N$.

17. Brownian Motion and the Black-Scholes Model

17.1. **Continuous-time random walk.** Take a compound Poisson process

$$S(t) = \sum_{k=1}^{N(t)} Z_k,$$

with $N = (N(t),\, t \ge 0)$ being a Poisson process with intensity $\lambda = 1$, and with $Z_1, Z_2, \ldots$ are i.i.d. variables with $\mathbf{P}(Z_i = 1) = \mathbf{P}(Z_i = -1) = 0.5$. This process S is called a *continuous-time random walk*, by analogy with a *discrete-time random walk* $X_n = Z_1 + \ldots + Z_n$. In fact, if you take $0 =: \tau_0 < \tau_1 < \ldots$ to be the jump times of this continuous-time random walk S, then the corresponding discrete-time Markov chain is nothing else but a discrete-time random walk: $X_n = S(\tau_n)$.

$$\mathbf{E}Z_k = 1 \cdot 0.5 + (-1) \cdot 0.5 = 0,\ \mathbf{E}Z_k^2 = 1^2 \cdot 0.5 + (-1)^2 \cdot 0.5 = 1,$$

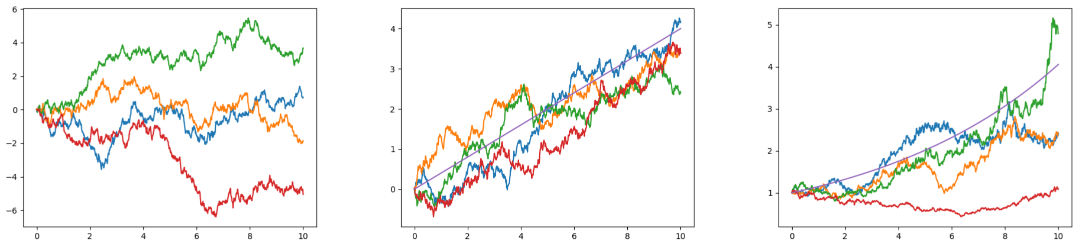
$$\text{and so}\quad \operatorname{Var} Z_k = \mathbf{E}Z_k^2 - (\mathbf{E}Z_k)^2 = 1.$$

Therefore,

$$\mathbf{E}S(t) = \lambda t \cdot \mathbf{E}Z_k = 0,\ \operatorname{Var} S(t) = t\left(\lambda(\mathbf{E}Z_k)^2 + \lambda^2 \operatorname{Var} Z_k\right) = t.$$

17.2. **Zooming out: Scaling limit is a Brownian motion.** To make jumps very small but very frequent, take a large n, and define

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{N(nt)} Z_k = \frac{1}{\sqrt{n}} S(nt).$$



Left: standard Brownian motion. Center: Brownian motion with drift $g = 0.4$ and diffusion $\sigma = .5$, starting from 0, and the drift line $x = gt$. Right: geometric Brownian motion with drift $g = 0.1$ and diffusion $\sigma = 0.2$, starting from 1, and the expectation $x = e^{(g+\sigma^2/2)t}$.

This is also a compound Poisson process: Poisson process $(N(tn),\, t \ge 0)$ having intensity $\lambda n = n$, so jumps occur every n^{-1} times, on average; and each jump is up or down with equal probability, with size $n^{-1/2}$ of jump. We have:

$$\mathbf{E} S_n(t) = \frac{1}{\sqrt{n}} \mathbf{E} S(nt) = 0,\ \operatorname{Var} S_n(t) = \left(\frac{1}{\sqrt{n}}\right)^2 \operatorname{Var} S(nt) = \frac{1}{n} \cdot nt = t.$$

The random variable $S_n(t)$, ofr a large n, is the sum of many very small summands. By Central Limit Theorem, as $n \to \infty$, it converges to the normal distribution with the same mean and variance as $S_n(t)$. In other words,

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{N(nt)} Z_k \ \Rightarrow\ \mathcal{N}(0, t).$$

Similarly, for $t > s$, as $n \to \infty$,

$$S_n(t) - S_n(s) = \frac{1}{\sqrt{n}} \sum_{N(ns)+1}^{N(nt)} Z_k \ \Rightarrow\ \mathcal{N}(0, t - s).$$

And this difference is independent of $S_n(u)$ for $u \le s$. Therefore, $S_n(t) \Rightarrow W(t)$ as $n \to \infty$. Here, we have the following properties of W:

(a) $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ for $t > s$, and is independent of $W(u)$, $u \le s$;
(b) $W(0) = 0$, because $S_n(0) = 0$.
(c) W has continuous trajectories, because the jumps of S_n are very small: only of size n^{-1}.

This process W which satisfies (a), (b), (c), is called a *Brownian motion.* This is the most important random (stochastic) process in Probability Theory.

This movement is a sum of very frequent but very small jumps. It was first used by Einstein to model the motion of a small dust particle hit by molecules (many small hits). Brownian motion is also used to model the stock price, because its movement is influenced by a lot of new bits of information, which are random (if they were not random, they would be predictable, that is, not new).

Brownian motion is a Markov process: Future movement $W(t)$, $t > s$, does not depend on the past $W(t)$, $t < s$, if we know the present $W(s)$. This is because $W(t) - W(s)$ is independent of $W(u)$, $u \le s$. The state space of the Brownian motion is the whole real line: $\mathbb{R}$.

17.3. **Expectation, variance, and transition density.** Assume $W(s) = x$. Then $W(t) \sim \mathcal{N}(x, t - s)$, because $W(t) - W(s) \sim \mathcal{N}(0, t - s)$. Therefore, the density of $W(t)$ given $W(s) = x$ is

$$p(s, t, x, y) = \frac{1}{\sqrt{2\pi(t - s)}} \exp\left(-\frac{(y - x)^2}{2(t - s)}\right).$$

In other words, for every $a < b$,

$$\mathbf{P}(a \le W(t) \le b \mid W(s) = x) = \int_a^b p(s, t, x, y)\,\mathrm{d}y.$$

We often write this as

$$p(s, t, x, y) = \varphi(t - s, y - x),\ \varphi(t, z) := \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{z^2}{2t}\right).$$

Example 17.1. We have:

$$\mathbf{E}W(2)W(3) = \mathbf{E}W^2(2) + \mathbf{E}\left[W(2)(W(3)-W(2))\right] = \operatorname{Var} W(2) + \mathbf{E}(W(3)-W(2))\mathbf{E}W(2) = 2.$$

More generally,

$$\mathbf{E}W(s)W(t) = \min(s,t).$$

Example 17.2. For every number x, we have:

$$\begin{aligned}\mathbf{E}\left[W(3) \mid W(2)=x\right] &= \mathbf{E}\left[W(3)-W(2) \mid W(2)=x\right] + \mathbf{E}\left[W(2) \mid W(2)=x\right] \\ &= \mathbf{E}(W(3)-W(2)) + x = 0 + x = x.\end{aligned}$$

Example 17.3. $\mathbf{E}\left[W^2(5) \mid W(3)=4\right] = \mathbf{E}\left(\xi+4\right)^2 = \mathbf{E}\xi^2 + 8\mathbf{E}\xi + 16 = 2+0+16 = 18$, where $\xi = W(5)-W(3) \sim \mathcal{N}(0,2)$.

Example 17.4. $\mathbf{E}(W(6)-W(4))W(5) = \mathbf{E}W(5)W(6) - \mathbf{E}W(4)W(5) = \min(5,6)-\min(4,5) = 1.$

Example 17.5. $\mathbf{P}(W(3)>0) = 0.5$, because $W(3) \sim \mathcal{N}(0,3)$ is symmetric with respect to zero.

Example 17.6. Let us find $\mathbf{P}(W(3)>1 \mid W(1)=2)$. We have: $\xi := W(3)-W(1) \sim \mathcal{N}(0,2)$. Therefore,

$$\mathbf{P}(W(3)>1 \mid W(1)=2) = \mathbf{P}(\xi+2>1) = \mathbf{P}(\xi>-1).$$

But $\xi = \sqrt{2}Z$ for $Z \sim \mathcal{N}(0,1)$, so

$$\mathbf{P}(\xi>-1) = \mathbf{P}(\sqrt{2}Z>-1) = \mathbf{P}(Z>-1/\sqrt{2}) = \mathbf{P}(Z>-0.707) = 1-0.7611 = 0.2389.$$

Example 17.7. Let us find the density of $W^3(2)$, given $W(1)=2$. We have: $W^3(2) = (\xi+2)^3$ for $\xi \sim \mathcal{N}(0,1)$. Then the density of ξ is

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

and the density of $(\xi+2)^3$ can be found as follows:

$$\begin{aligned}&\mathbf{P}\left(a \le (\xi+2)^3 \le b\right) \\ &= \mathbf{P}\left(a^{1/3}-2 \le \xi \le b^{1/3}-2\right) = \int_{a^{1/3}-2}^{b^{1/3}-2} f(x)\,\mathrm{d}x \\ &= \int_a^b f(y^{1/3}-2)\frac{1}{3}y^{-2/3}\mathrm{d}y.\end{aligned}$$

where we change variables $x = y^{1/3}-2$, so

$$\mathrm{d}x = \frac{1}{3}y^{-2/3}\mathrm{d}y,$$

$$a^{1/3}-2 \le x \le b^{1/3}-2 \ \Leftrightarrow\ a^{1/3}-2 \le y^{1/3}-2 \le b^{1/3}-2 \ \Rightarrow\ a \le y \le b$$

Therefore, the density of $(\xi+2)^3$ is given by

$$f(y^{1/3}-2)\frac{1}{3}y^{-2/3} = \frac{1}{3\sqrt{2\pi}y^{2/3}}\exp\left(-\frac{(y^{1/3}-2)^2}{2}\right).$$

17.4. **Brownian motion with drift and diffusion.** Take coefficients g and σ, a starting point x, and define

$$X(t) = x + gt + \sigma W(t).$$

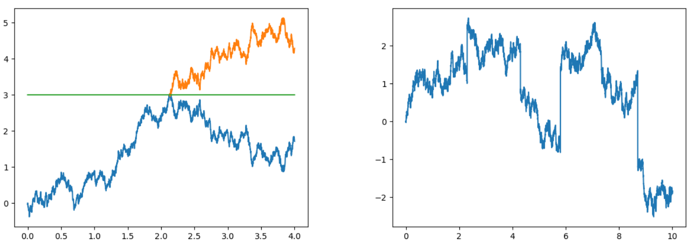
This is a *Brownian motion starting from* x, with *drift coefficient* g, and *diffusion coefficient* σ^2. The *standard Brownian motion* has $x=0$, $g=0$, $\sigma=1$. When we say "Brownian motion" without specifying the coefficients and the starting point, we mean the standard Brownian motion. This process has properties similar to the standard Brownian motion: for $t>s$,

$$X(t)-X(s) = g(t-s) + \sigma(W(t)-W(s)) \sim \mathcal{N}\left(g(t-s), \sigma^2(t-s)\right),$$

and this random variable is independent of $X(u)$, $u \le s$.

Example 17.8. Consider $x=0$, $g=-1$, $\sigma=2$. Find $\mathbf{E}(X(3) \mid X(2)=2)$. We have: $X(t) = -t+2W(t)$, and so $X(3)-X(2) = -1+2(W(3)-W(2)) \sim \mathcal{N}(-1,4)$. Therefore, $\mathbf{E}(X(3) \mid X(2)=2) = \mathbf{E}(X(3)-X(2))+2 = -1+2 = 1$. Next, find $\mathbf{E}(X^2(5) \mid X(2)=2)$: if $\xi := X(5)-X(2) \sim \mathcal{N}(-3,12)$, then we have $X^2(5) = (2+\xi)^2$, and the answer is

$$\mathbf{E}(2+\xi)^2 = 4 + 4\mathbf{E}\xi + \mathbf{E}\xi^2 = 4 + 4\cdot(-3) + (-3)^2 + 12 = 13.$$



Left: Illustration of the reflection principle for the Brownian motion. Right: A Lévy process: a standard Brownian motion plus a compound Poisson process with intensity of jumps $\lambda = 0.5$.

17.5. **Levy processes.** Combine a Brownian motion with drift and diffusion together with a compound Poisson process:

$$L(t) = x + gt + \sigma W(t) + \sum_{k=1}^{N(t)} Z_k. \tag{58}$$

Here, W is a standard Brownian motion, N is a Poisson process with intensity λ, and $Z_1, Z_2, \ldots$ are i.i.d. Also, W, N, Z_k are all independent. This process behaves as a Brownian motion between jumps, which occur every $\mathrm{Exp}(\lambda)$ times, and the displacement during the kth jump is Z_k. This process also has the Markov property: $L(t) - L(s)$ is independent of $L(u)$, $u \le s$. Also, it has *stationary increments*: the distribution of $L(t) - L(s)$ depends only on $t - s$ for $t > s$. For example, the distribution of $L(6) - L(4)$ and of $L(5) - L(3)$ is the same.

Example 17.9. $x = 1$, $g = 2$, $\sigma = 3$, $\lambda = 2$, $Z_k \sim \mathrm{Exp}(2)$. Then

$$\mathbf{E}L(t) = x + gt + \sigma\mathbf{E}W(t) + \mathbf{E}\sum_{k=1}^{N(t)} Z_k = 1 + 2t + \lambda t\mathbf{E}Z_k = 1 + 2t + 2t\cdot\frac{1}{2} = 1 + 3t,$$

$$\begin{aligned}\operatorname{Var} L(t) &= \sigma^2 \operatorname{Var} W(t) + \operatorname{Var}\sum_{k=1}^{N(t)} Z_k = \sigma^2 t + t\lambda\left((\mathbf{E}Z_k)^2 + \operatorname{Var} Z_k\right)\\ &= 9t + 2t\left(0.5^2 + \frac{1}{4}\right) = 10t.\end{aligned}$$

Let $G(u) := \mathbf{E}e^{uZ_k}$ be the moment generating function of Z_k. It is equal to

$$G(u) = \frac{2}{2-u}.$$

Then the moment generating function of $L(t)$ is given by

$$\begin{aligned}F_t(u) := \mathbf{E}\exp(uL(t)) &= \mathbf{E}\exp\left[u\left(x + gt + \sigma W(t)\right)\right]\mathbf{E}\exp\left[u\sum_{k=1}^{N(t)} Z_k\right]\\ &= \exp\left(ux + gut + \frac{\sigma^2 u^2}{2}t\right)\exp\left(\lambda t(G(u) - 1)\right)\\ &= e^{ux}\exp\left(t\left[gu + \sigma^2u^2/2 + \lambda(G(u) - 1)\right]\right) = \exp\left(u + t\left[\frac{13}{2}u + 2\left(\frac{2}{2-u} - 1\right)\right]\right).\end{aligned}$$

Any process with stationary independent increments is called a Levy process. One can show that if such process has at most finitely many jumps on a finite time interval, then it can be represented as (58).

17.6. **Reflection principle and the maximum of Brownian motion.** Try to find the distribution of $M(t) = \max_{0\le s\le t} W(s)$. For example, find

$$\mathbf{P}(M(4) \ge 3,\ 1 \le W(4) \le 2). \tag{59}$$

The trajectory of the Brownian motion which satisfies (59) crosses the line $y = 3$. Let us take the first moment of crossing and reflect the trajectory of the Brownian motion across this line, starting from this moment. Then we get another trajectory of Brownian motion, because it is symmetric (can go up or down with equal probability). But it reaches between 4 and 5 at time $t = 4$, because 4 and 5 are symmetric to the points 2 and 1 respectively, with respect to this line. Conversely, every such trajectory of a Brownian motion with $4 \le W(4) \le 5$, after being reflected after its first crossing of line $y = 3$, becomes a Brownian motion which satisfies (59). Therefore, the probability from (59) is equal to (with $W(4) = 2Z$, $Z \sim \mathcal{N}(0,1)$):

$$\mathbf{P}(4 \le W(4) \le 5) = \frac{1}{\sqrt{2\pi}2}\int_4^5 e^{-x^2/8}\,\mathrm{d}x = \mathbf{P}(2 \le Z \le 2.5) = 0.9938 - 0.9772 = \boxed{0.0166}$$

Similarly,

$$\mathbf{P}(M(4) \ge 3,\, W(4) \le 3) = \mathbf{P}(W(4) \ge 3) = \frac{1}{\sqrt{2\pi}2}\int_3^\infty e^{-x^2/8}\,\mathrm{d}x.$$

Note that if $W(4) \ge 3$, then certainly $M(4) \ge 3$. And

$$\mathbf{P}(M(4) \ge 3) = \mathbf{P}(M(4) \ge 3,\, W(4) \le 3) + \mathbf{P}(W(4) \ge 3) = 2\mathbf{P}(W(4) \ge 3).$$

Similarly, for every $t > 0$, and $y \ge 0$,

$$\mathbf{P}(M(t) \ge y) = 2\mathbf{P}(W(t) \ge y) = \mathbf{P}(W(t) \ge y) + \mathbf{P}(W(t) \le -y) = \mathbf{P}(|W(t)| \ge y).$$

That is, $M(t)$ and $|W(t)|$ have the same distribution. It has density

$$\frac{2}{\sqrt{2\pi t}}e^{-x^2/2t}\,\mathrm{d}x,\ \ x \ge 0.$$

17.7. **Reflected Brownian motion.** This is the process $|W| = (|W(t)|,\, t \ge 0)$. It is a Markov process, because the behavior of $|W(t)|$ for $t \ge s$ depends only on $|W(s)|$; if you know $W(u)$, $u < s$, this does not give you any additional information. The state space of this process is $\mathbb{R}_+ := [0, \infty)$. *Skorohod representation:*

$$|W(t)| = B(t) + \ell(t),$$

where B is another Brownian motion (not W!), and ℓ is a continuous nondecreasing process with $\ell(0) = 0$, with ℓ increasing only when $|W| = 0$. As long as $|W|$ stays inside the positive half-line, it behaves as a Brownian motion. But when it hits zero, it "wants" to go down, but is not allowed to, because it has to stay positive. Then we add a push $\mathrm{d}\ell(t)$ to make it positive, and prevent it from crossing $y = 0$. The process ℓ is called a *local time* of $|W|$ at zero.

The transition density of $|W(t)|$: for any $0 \le a < b$,

$$\begin{aligned}
&\mathbf{P}\left(a \le |W(t)| \le b \mid |W(s)| = x\right) \\
&\quad = \mathbf{P}\left(a \le W(t) \le b \mid W(s) = x\right) + \mathbf{P}\left(-b \le W(t) \le -a \mid W(s) = x\right) \\
&\quad = \frac{1}{\sqrt{2\pi(t-s)}}\int_a^b \exp\left(-\frac{(y-x)^2}{2(t-s)}\right)\mathrm{d}y + \frac{1}{\sqrt{2\pi(t-s)}}\int_{-b}^{-a} \exp\left(-\frac{(y-x)^2}{2(t-s)}\right)\mathrm{d}y \\
&\quad = \frac{1}{\sqrt{2\pi(t-s)}}\int_a^b \left[\exp\left(-\frac{(y-x)^2}{2(t-s)}\right) + \exp\left(-\frac{(-y-x)^2}{2(t-s)}\right)\right]\mathrm{d}y
\end{aligned}$$

so given $|W(s)| = x$, $|W(t)|$ has density

$$p(s,t,x,y) = \frac{1}{\sqrt{2\pi(t-s)}}\left[\exp\left(-\frac{(y-x)^2}{2(t-s)}\right) + \exp\left(-\frac{(y+x)^2}{2(t-s)}\right)\right].$$

Example 17.10. Let us find $\mathbf{P}(|W(2)| \le 1 \mid |W(1)| = 1)$. Let $\xi := W(2) - W(1) \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned}
\mathbf{P}(|W(2)| \le 1 \mid |W(1)| = 1) &= \mathbf{P}(-1 \le W(2) \le 1 \mid W(1) = 1) \\
&= \mathbf{P}(-1 \le \xi - 1 \le 1) = \mathbf{P}(0 \le \xi \le 2) = \frac{1}{\sqrt{2\pi}}\int_0^2 e^{-x^2/2}\,\mathrm{d}x = 0.48.
\end{aligned}$$

Example 17.11. Let us find $\mathbf{E}|W(3)|$. We can represent $W(3) = \sqrt{3}\xi$, $\xi \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned}
\mathbf{E}|W(3)| &= \sqrt{3}\mathbf{E}|\xi| = \sqrt{3}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty |y|e^{-y^2/2}\,\mathrm{d}y \\
&= \sqrt{3}\frac{2}{\sqrt{2\pi}}\int_0^\infty ye^{-y^2/2}\,\mathrm{d}y = \sqrt{3}\frac{2}{\sqrt{2\pi}}\left(-e^{-y^2/2}\right)\Big|_{y=0}^\infty = \frac{2\sqrt{3}}{\sqrt{2\pi}}.
\end{aligned}$$

17.8. **Transformations of Brownian motion.** Note that the process $(W(9t),\, t \ge 0)$ is also a Brownian motion with diffusion coefficient 9 and drift coefficient 0. In other words, the following processes have the same distribution:

$$(W(9t),\, t \ge 0) \text{ and } (3W(t),\, t \ge 0).$$

Indeed, $W(9t)$ and $3W(t)$ have the same distribution: $\mathcal{N}(0, 9t)$. Moreover, $W(9t) - W(9s)$ and $3W(t) - 3W(s)$ have the same distribution: $\mathcal{N}(0, 9(t-s))$, for $t > s$. And $W(9t) - W(9s)$ is independent of $W(9u)$, $u \le s$.

Consider two independent Brownian motions: $X = (X(t),\, t \ge 0)$ is a Brownian motion with drift $g_1 = -2$ and diffusion $\sigma_1^2 = 3$, starting from $x_1 = -1$, and $Y = (Y(t),\, t \ge 0)$ is a Brownian motion with drift $g_2 = 3$ and diffusion $\sigma_2^2 = 4$, starting from $x_2 = 2$. Then take the process

$$Z = (Z(t),\, t \ge 0) = 2X - 3Y = (2X(t) - 3Y(t),\, t \ge 0).$$

This is also a Brownian motion, starting from

$$Z(0) = 2X(0) - 3Y(0) = 2 \cdot (-1) - 3 \cdot 2 = -8.$$

For $t > s$, we have:

$$Z(t) - Z(s) \sim 2(X(t) - X(s)) - 3(Y(t) - Y(s)).$$

But the following increments are normally distributed and are independent:

$$X(t) - X(s) \sim \mathcal{N}(-2(t-s), 3(t-s)),\ \ Y(t) - Y(s) \sim \mathcal{N}(3(t-s), 4(t-s)).$$

Therefore, using results of Section 9, we get: $Z(t) - Z(s)$ is also normal with

$$\begin{aligned}\mathbf{E}(Z(t) - Z(s)) &= 2\mathbf{E}(X(t) - X(s)) - 3\mathbf{E}(Y(t) - Y(s)) \\ &= -2(t-s) \cdot 2 + 3(t-s) \cdot (-3) = -13(t-s),\end{aligned}$$

$$\begin{aligned}\operatorname{Var}(Z(t) - Z(s)) &= 2^2 \operatorname{Var}(X(t) - X(s)) + (-3)^2 \operatorname{Var}(Y(t) - Y(s)) \\ &= 2^2 \cdot 3(t-s) + (-3)^2 \cdot 4(t-s) = 48(t-s).\end{aligned}$$

And $Z(t) - Z(s)$ is independent of $Z(u)$, $u \le s$, because X and Y satisfy the same property. Therefore, Z is a Brownian motion with drift coefficient -13 and diffusion coefficient 48, starting from -8.

17.9. **Derivatives in the Black-Scholes model.** This is an extension of stochastic finance theory from Section 12 for the continuous case. We model the stock price as the geometric Brownian motion:

$$S(t) = S_0 \exp(gt + \sigma W(t)).$$

Here, σ is called *volatility*; this parameter shows the magnitude of fluctuations in the price. One can see the definitions of European options and derivatives in Section 12. We would like to hedge a derivative $D = f(S(T))$, where $f(x)$ is some real-valued function. In particular, we are interested in the following functions, with $a_+ := \max(a, 0)$: (a) $f(x) = (x - K)_+$: *European option-call* with *strike* K; (b) $f(x) = (K - x)_+$: *European option-put* with *strike* K; (c) $f(x) = 1(x > K)$: *Binary option* with *barrier* K.

If we had random walk instead of a Brownian motion, we would switch to a risk-neutral probability, under which $\mathbf{E}S(t) = S_0$. Here, we have a (geometric) Brownian motion, which is a limit of geometric random walks with small but frequent jumps. It is possible to show that in this case, risk-neutral probability corresponds to a change in drift. That is, we need to find new g_* such that $\mathbf{E}S(t) = S_0$. We can do this as follows: Calculating the moment generating function of $W(t) \sim \mathcal{N}(0, t)$, we get:

$$\mathbf{E}\exp(g_* t + \sigma W(t)) = e^{g_* t} e^{\sigma^2 t/2} = 1 \ \Rightarrow\ g_* = -\frac{\sigma^2}{2}.$$

Therefore, under the new risk-neutral probability, the stock price is modeled by the following geometric Brownian motion:

$$S(t) = S_0 \exp\left(\sigma W(t) - \frac{\sigma^2}{2} t\right),\ t \ge 0. \tag{60}$$

And we need to take the expected value: $v = \mathbf{E} f(S(T))$. This is the fair price. We can find the way to *hedge*, or *replicate* these derivatives, using stochastic calculus in the next section.

Example 17.12. For the binary option with barrier $K = 3$, maturity $T = 2$, traded upon the stock with volatility $\sigma = 3$ and initial price $S_0 = 1.5$, its risk-neutral dynamics is given by $S(t) = 1.5\,\exp\left(3W(t) - 9t/2\right)$, therefore the fair price of this option is (with $W(2) = \sqrt{2}Z$ for $Z \sim \mathcal{N}(0,1)$):

$$\begin{aligned}\mathbf{E}1(S(2) > 3) &= \mathbf{P}\left(1.5\,\exp\left(3W(2) - 9\cdot 2/2\right) > 3\right) = \mathbf{P}\left(\exp\left(3\sqrt{2}Z - 9\right) > 2\right)\\ &= \mathbf{P}\left(3\sqrt{2}Z - 9 > \ln 2\right) = \mathbf{P}\left(Z > \frac{3}{\sqrt{2}} + \frac{\ln 2}{3\sqrt{2}}\right) = \mathbf{P}(Z > 2.28) = 0.011\end{aligned}$$

For the European option call, we get the celebrated *Black-Scholes formula*, which got a Nobel Prize in Economics in 1997.

Example 17.13. Try strike $K = 1.5$, maturity $T = 3$, volatility $\sigma = 2$, and current price $S_0 = 1$. Then $S(t) = \exp(2W(t) - 2t)$, and $S(T) = \exp(2W(3) - 6)$. Then the fair price is

$$v = \mathbf{E}(S(T) - K)_+ = \mathbf{E}\left(\exp(2W(3) - 6) - 1.5\right)_+.$$

The option is executed if its terminal price at maturity is greater than strike, that is,

$$3\exp(2W(3) - 6) - 1.5 > 0 \ \Leftrightarrow W(3) > 3 + \log 1.5 = 3.405,$$

The probability density function of $W(3)$ is given by $(6\pi)^{-1/2}\exp(-x^2/6)$. Then

$$\begin{aligned}v &= \mathbf{E}\left(\exp(2W(3) - 6) - 1.5\right)_+ = \frac{1}{\sqrt{6\pi}}\int_{3+\log 1.5}^{\infty}\left[e^{2x-6} - 1.5\right]e^{-x^2/6}\,\mathrm{d}x\\ &= \frac{1}{\sqrt{6\pi}}\int_{3+\log 1.5}^{\infty} e^{2x-6}\,e^{-x^2/6}\,\mathrm{d}x - \frac{1}{\sqrt{6\pi}}\int_{3+\log 1.5}^{\infty} 1.5\,e^{-x^2/6}\,\mathrm{d}x.\end{aligned} \tag{61}$$

The density $(6\pi)^{-1/2}e^{-x^2/6}$ is of $\sqrt{3}Z$, $Z \sim \mathcal{N}(0,1)$. The second term in the right-hand side of (61) is equal to:

$$\frac{1}{\sqrt{6\pi}}\int_{3+\log 1.5}^{\infty} 1.5e^{-x^2/6}\,\mathrm{d}x = \mathbf{P}(\sqrt{3}Z > 3 + \log(1.5)) = \mathbf{P}(\sqrt{3}Z > 3 + \log 1.5) = \mathbf{P}\left(Z > \sqrt{3} + \frac{\log 1.5}{\sqrt{3}}\right).$$

The first term in the right-hand side of (61) is equal to:

$$\begin{aligned}\frac{1}{\sqrt{6\pi}}\int_{3+\log 1.5}^{\infty} e^{2x-6}\,e^{-x^2/6}\,\mathrm{d}x &= \frac{1}{\sqrt{6\pi}}\int_{3+\log 1.5}^{\infty} e^{-x^2/6+2x-6}\,\mathrm{d}x = \frac{1}{\sqrt{6\pi}}\int_{3+\log 1.5}^{\infty} e^{-(x-6)^2/6}\,\mathrm{d}x\\ &= \mathbf{P}(\sqrt{3}Z + 6 > 3 + \log 1.5) = \mathbf{P}\left(Z > -\sqrt{3} + \frac{\log 1.5}{\sqrt{3}}\right).\end{aligned}$$

Therefore, the answer is

$$\begin{aligned}v &= \mathbf{P}\left(Z > -\sqrt{3} + \frac{\log 1.5}{\sqrt{3}}\right) - 1.5\cdot\mathbf{P}\left(Z > \sqrt{3} + \frac{\log 1.5}{\sqrt{3}}\right)\\ &= \mathbf{P}(Z > -1.50) - 1.5\cdot\mathbf{P}(Z > 1.97) = 0.933 - 1.5\cdot 0.024 = \boxed{0.897}\end{aligned}$$

We can also find the fair price of a European derivative at any time t, if we know the price $S(t) = x$ at this time: This is a function

$$u(t,x) = \mathbf{E}(f(S(T)) \mid S(t) = x),$$

where the expectation is taken again with respect to the risk-neutral probability, that is, for the process $S(t)$ from (60). Then we simply treat time t as the initial time, and $S(t) = x$ as the initial price x. The hedging of such a derivative: that is, replicating it with a combination of cash and this stock, is slightly harder, and is deferred to the next section on stochastic calculus.

17.10. **Drawbacks of the Black-Scholes model.** *Heavy Tails.* Large fluctuations of stocks occur more frequently than prescribed by the normal distributions. A better distribution is $\mathbf{P}(\ln(S(t_2)/S(t_1)) \ge x) \approx cx^{-\alpha}$, and same for $\mathbf{P}(\ln(S(t_2)/S(t_1)) \le -x)$; here, $c, \alpha > 0$ are some constants.

Short-term dependence. Increments of logarithms are *not* independent: say $\ln S(2.1) - \ln S(2)$ and $\ln S(2.2) - \ln S(2.1)$ are dependent, although they should be independent if we lived in the Black-Scholes model.

Volatility is not constant. If we calculate σ^2 from real data, it should be constant in the Black-Scholes model, but it is not. In fact, the graph $(S(t), \sigma(t))$, of the stock price and volatility, is not a horizontal line, but a decreasing one: *"volatility skew"* or a U-shaped curve: *"volatility smile"*. To explain this, several techniques are available, including *stochastic volatility models:* when $\sigma(t)$ is modeled itself by a stochastic process, for example using second Brownian motion $B(t)$, correlated with $W(t)$.

Bid-ask spread. The stock price is actually two different prices: *bid*, the price offered for buying, and *ask*, the price offered for selling. Selling - buying occurs when a bid is equal to an ask. The dynamics of bid offers and ask offers is modeled separately.

Problems

For the next nine problems, let W be the standard Brownian motion.

Problem 17.1. Find $\mathbf{E}W^3(3)$.

Problem 17.2. Find $\mathbf{E}W(2)(W(1)+W(3))$.

Problem 17.3. Find $\mathbf{E}W(1)W^2(2)$.

Problem 17.4. Find $\mathbf{E}(W(6) \mid W(2)=1)$.

Problem 17.5. Find $\mathrm{Var}(W(6) \mid W(2)=1)$.

Problem 17.6. Find $\mathbf{P}(4 \le W(5) \le 5 \mid W(2)=1)$.

Problem 17.7. Find $\mathbf{E}\left(e^{W(5)} \mid W(3)=-2\right)$.

Problem 17.8. Find the density of $W(3)$, given $W(1)=1$.

Problem 17.9. Find the density of $e^{W(3)}$, given $W(1)=1$.

For the next three problems, take a Brownian motion $X=(X(t),\, t\ge 0)$ with drift $g=2$ and diffusion $\sigma^2=3$, starting from $x=3$.

Problem 17.10. Find the density of $X(4)$ given $X(2)=0$.

Problem 17.11. Find $\mathbf{E}(X^2(3) \mid X(1)=1)$.

Problem 17.12. Find $\mathbf{E}(X(2)X(3) \mid X(1)=1)$.

For the next three problems, consider a Levy process $L=(L(t),\, t\ge 0)$, starting from $x=1$, with $g=-1$, $\sigma=2$, $\lambda=3$, with $Z_k \sim \mathcal{N}(0,1)$.

Problem 17.13. Find $\mathbf{E}L(4)$.

Problem 17.14. Find $\mathrm{Var}\, L(4)$.

Problem 17.15. Find the moment generating function of $L(4)$.

For the next four problems, let X be the Brownian motion with drift $g=3$ and diffusion $\sigma^2=4$, starting from $x=1$.

Problem 17.16. Find the distribution, expectation, variance, and the moment generating function of $X(t)$.

Problem 17.17. Calculate $\mathbf{E}(X^2(5)X(6) \mid X(3)=-1)$.

Problem 17.18. Find the density of $X^3(4)$, given $X(1)=-1$.

Problem 17.19. Find $\mathbf{P}(-3 \le X(5) \le 2)$.

For the next eight problems, take the standard Brownian motion $W=(W(t),\, t\ge 0)$, and let $M(t) := \max\limits_{0\le s\le t} W(s)$.

Problem 17.20. Find $\mathbf{P}(M(1)\le 3)$.

Problem 17.21. Find $\mathbf{P}(M(3)\ge 5,\, 2\le W(3)\le 4)$.

Problem 17.22. Find $\mathbf{P}(M(5)\ge 5,\, W(5)\le 5)$.

Problem 17.23. Find $\mathbf{P}(1\le W(3)\le 3,\, M(3)\ge 3)$.

Problem 17.24. Find the expectation and variance of $|W(2)|$.

Problem 17.25. Find $\mathbf{P}(M(4)\ge 1,\, -1\le W(4)\le 0)$.

Problem 17.26. Find $\mathbf{P}(M(2)>1, W(2)<0.5)$.

Problem 17.27. Find $\mathbf{E}\left[W(5)(W(6)+3) \mid W(2)=-1\right]$.

For the next two problems, consider a geometric Brownian motion X with drift $g=-1$ and diffusion $\sigma^2=3$, starting from $X(0)=2$.

Problem 17.28. Find $\mathbf{P}(X(4) > 3)$.

Problem 17.29. Find $\mathbf{E}X(4)$.

For the next three problems, consider a Levy process $L = (L(t),\, t \ge 0)$, starting from $x = -2$, with $g = 1$, $\sigma^2 = 3$, $\lambda = 4$, with $Z_k \sim \mathrm{Geo}(2/3)$.

Problem 17.30. Find $\mathbf{E}L(t)$.

Problem 17.31. Find $\operatorname{Var} L(t)$.

Problem 17.32. Find the moment generating function of $L(t)$.

Problem 17.33. Take a Brownian motion $X = (X(t),\, t \ge 0)$ with drift $g = 2$, diffusion $\sigma = 4$, starting from $x = 1$. Find $\mathbf{P}(-1 \le X(5) \le 5 \mid X(2) = 0)$.

For the next two problems, consider a Brownian motion $X = (X(t),\, t \ge 0)$, with drift $g = 2$ and diffusion $\sigma = 3$, starting from $X(0) = -2$.

Problem 17.34. Find $\mathbf{E}X(4)X(6)$.

Problem 17.35. Find the probability density function of $e^{X(t)}$.

Problem 17.36. For a geometric Brownian motion $X(t)$ with drift $g = 0$ and diffusion $\sigma^2 = 2$, starting from $X(0) = 2$, find $\mathbf{P}(1 \le X(3) \le 4)$.

Problem 17.37. Find the fair value of the European option call if $\sigma^2 = 3$, $T = 1$, $S_0 = 1$, and $K = 2$.

Problem 17.38. Find the fair value of the European option put if $\sigma^2 = 3$, $T = 1$, $S_0 = 1$, and $K = 2$.

Problem 17.39. Find the value of the binary option with $\sigma^2 = 3$, $T = 1$, $S_0 = 1$, and $K = 2$.

Problem 17.40. What volatility σ does the stock have if its current price is 1, and the binary option with maturity $T = 3$ and the barrier $K = 2$ trades at 0.3? Such σ is called *implied volatility.*

18. Stochastic Calculus

In this section, our aim is to differentiate Brownian motion $W = (W(t),\, t \ge 0)$. We cannot do this in the usual sense, because this is not differentiable. Indeed,

$$W(t+\mathrm{d}t) - W(t) \sim \mathcal{N}(0, \mathrm{d}t) \ \Rightarrow\ \frac{W(t+\mathrm{d}t) - W(t)}{\mathrm{d}t} \sim \mathcal{N}\left(0, \frac{1}{\mathrm{d}t^2}\cdot \mathrm{d}t\right) = \mathcal{N}\left(0, \frac{1}{\mathrm{d}t}\right). \tag{62}$$

As $\mathrm{d}t \to 0$, the variance $1/\mathrm{d}t \to \infty$. Therefore, this ratio in (62) does not converge anywhere.

18.1. **Quadratic variation.** For a function $f : [0, \infty) \to \mathbb{R}$, define its *quadratic variation* on the interval $[0, t]$:

$$\langle f \rangle_t := \lim_{N\to\infty} \sum_{k=0}^{N-1} \left(f\left(\frac{(k+1)t}{N}\right) - f\left(\frac{kt}{N}\right)\right)^2.$$

We split the interval $[0, t]$ into N equal subintervals

$$\left[0, \frac{t}{N}\right], \left[\frac{t}{N}, \frac{2t}{N}\right], \ldots, \left[\frac{(N-1)t}{N}, t\right].$$

and calculate the increment of f on each such subinterval; then square each and sum them. Then let the size of intervals go to zero. For smooth functions f, one can show that this quantity is zero. But for the Brownian motion W, we have independent identically distributed increments:

$$W\left(\frac{(k+1)t}{N}\right) - W\left(\frac{kt}{N}\right) \sim \mathcal{N}\left(0, \frac{t}{N}\right).$$

Therefore, we can represent

$$W\left(\frac{(k+1)t}{N}\right) - W\left(\frac{kt}{N}\right) = (t/N)^{1/2} Z_k,\ k = 0, \ldots, N-1,$$

where $Z_1, \ldots, Z_N$ are i.i.d. $\mathcal{N}(0, 1)$. Therefore, the sum of squares by the Law of Large Numbers converges as $N \to \infty$:

$$\left((t/N)^{1/2} Z_1\right)^2 + \ldots + \left((t/N)^{1/2} Z_N\right)^2 = \frac{t}{N}\left(Z_1^2 + \ldots + Z_N^2\right) \to t\mathbf{E}Z_1 = t.$$

We have shown that $\langle W\rangle_t = t$.

We can also symbolically represent quadratic variation as

$$\langle f\rangle_t = \int_0^t (\mathrm{d}f)^2\,.$$

In other words, $\mathrm{d}\langle f\rangle_t = (\mathrm{d}f)^2$. For smooth functions f, we have: $\mathrm{d}f = f'(t)\,\mathrm{d}t$, and therefore

$$\mathrm{d}\langle f\rangle_t = (f'(t)\,\mathrm{d}t)^2 = f'^2(t)\,(\mathrm{d}t)^2 = 0.$$

Indeed, in such calculations we must consider $(\mathrm{d}t)^2 = 0$, and in general $(\mathrm{d}t)^a = 0$ for every $a > 1$.

18.2. **Construction of a stochastic integral.** For a deterministic function or a random process $H = (H(t),\, t \ge 0)$, define the *stochastic integral*, or *Itô's integral:*

$$\begin{aligned}\int_0^t H(s)\,\mathrm{d}W(s) &= \lim_{N\to\infty}\sum_{k=0}^{N-1} H\left(\frac{kt}{N}\right)\left(W\left(\frac{(k+1)t}{N}\right) - W\left(\frac{kt}{N}\right)\right)\\ &= H(0)\left(W\left(\frac{t}{N}\right) - W(0)\right) + H\left(\frac{t}{N}\right)\left(W\left(\frac{2t}{N}\right) - W\left(\frac{t}{N}\right)\right)\\ &+ \ldots + H\left(\frac{t(N-1)}{N}\right)\left(W(t) - W\left(\frac{t(N-1)}{N}\right)\right).\end{aligned}\tag{63}$$

This is similar to a Riemann sum for the integral $\int_0^t H(s)\,\mathrm{d}s$.

Example 18.1. Let $H(s) \equiv c$. Then the sum in (63) becomes $c(W(t) - W(0)) = cW(t)$. That is,

$$\int_0^t c\mathrm{d}W(s) = cW(t).$$

18.3. **Stochastic integral of a deterministic function.** More generally, take a deterministic (non-random) function $f : [0,t] \to \mathbb{R}$, and let $H := f$. Then the sum in (63) is a linear combination of independent normal random variables

$$Z_k := W\left(\frac{(k+1)t}{N}\right) - W\left(\frac{kt}{N}\right) \sim \mathcal{N}\left(0, \frac{t}{N}\right),\quad k = 0,\ldots,N-1.$$

Each of which has $\mathbf{E}Z_k = 0$ and $\operatorname{Var} Z_k = t/N$. We can rewrite this sum as

$$S_N := f(0)\,Z_0 + f\left(\frac{t}{N}\right)\,Z_1 + \ldots + f\left(\frac{(N-1)t}{N}\right)\,Z_{N-1}.$$

It also has normal distribution (as a linear combination of independent normal random variables), with mean and variance

$$\mathbf{E}S_N = f(0)\,\mathbf{E}Z_0 + \ldots + f\left(\frac{(N-1)t}{N}\right)\,\mathbf{E}Z_{N-1} = 0,$$

$$\operatorname{Var} S_N = f^2(0)\,\operatorname{Var} Z_0 + \ldots + f^2\left(\frac{(N-1)t}{N}\right)\,\operatorname{Var} Z_{N-1} = \frac{t}{N}\sum_{k=0}^{N-1} f^2\left(\frac{kt}{N}\right).$$

Note that $\operatorname{Var} S_N$ is a Riemann sum for the integral $I := \int_0^t f^2(s)\,\mathrm{d}s$. Therefore, $\operatorname{Var} S_N \to I$. And the stochastic integral is distributed as the normal random variable:

$$\int_0^t f(s)\,\mathrm{d}W(s) \sim \mathcal{N}(0, I).$$

Example 18.2. Consider the integral $\int_0^2 (3-t)\,\mathrm{d}W(t)$. It is distributed as $\mathcal{N}(0,\sigma^2)$ with

$$\sigma^2 = \int_0^2 (3-s)^2\,\mathrm{d}s = \int_0^2 \left(9 - 6s + s^2\right)\,\mathrm{d}s = \left(9s - 3s^2 + \frac{s^3}{3}\right)\Big|_{s=0}^{s=2} = \frac{26}{3}.$$

Example 18.3. Consider the integral $\int_0^1 t\,\mathrm{d}X(t)$, where X is a Brownian motion with drift $g = -1$ and diffusion $\sigma^2 = 4$. Then we have: $\mathrm{d}X(t) = -\mathrm{d}t + 2\mathrm{d}W(t)$, and therefore

$$\int_0^1 t\,\mathrm{d}X(t) = -\int_0^1 t\,\mathrm{d}t + 2\int_0^1 t\,\mathrm{d}W(t) \sim \mathcal{N}\left(-\int_0^1 t\,\mathrm{d}t, 4\int_0^1 t^2\,\mathrm{d}t\right) = \mathcal{N}\left(-\frac{1}{2}, \frac{4}{3}\right).$$

If the integrator $H = (H(t),\, t \ge 0)$ is random, then similarly we can show that

$$\mathbf{E}\int_0^T X(t)\,\mathrm{d}W(t) = 0 \text{ and } \mathbf{E}\left[\int_0^T X(t)\,\mathrm{d}W(t)\right]^2 = \mathbf{E}\int_0^t X^2(t)\,\mathrm{d}t.$$

However, in this case the distribution of this stochastic integral is not necessarily normal.

18.4. **Relation between stochastic integral and quadratic variation.** The stochastic integral

$$X(t) = \int_0^t H(s)\,\mathrm{d}W(s),\ t \ge 0 \tag{64}$$

can be itself viewed as a random process. We can write it in the form

$$\mathrm{d}X(t) = H(t)\,\mathrm{d}W(t).$$

Its quadratic variation is equal to

$$\langle X\rangle_t = \int_0^t (\mathrm{d}X(s))^2 = \int_0^t H^2(s)\,(\mathrm{d}W(s))^2 = \int_0^t H^2(s)\mathrm{d}\langle W\rangle_s = \int_0^t H^2(s)\,\mathrm{d}s.$$

Alternatively, we can write this as

$$\mathrm{d}\langle X\rangle_t = H^2(t)\,\mathrm{d}t.$$

When X is a sum of a stochastic integral and a usual integral:

$$X(t) = X(0) + \int_0^t G(s)\,\mathrm{d}s + \int_0^t H(s)\,\mathrm{d}W(s), \tag{65}$$

we can rewrite this as $\mathrm{d}X(t) = G(t)\,\mathrm{d}t + H(t)\,\mathrm{d}W(t)$. Therefore,

$$\mathrm{d}\langle X\rangle_t = (\mathrm{d}X(t))^2 = G^2(t)(\mathrm{d}t)^2 + 2G(t)H(t)\mathrm{d}t\mathrm{d}W(t) + H^2(t)\,\mathrm{d}t.$$

The term $(\mathrm{d}t)^2$ should be neglected in such calculation, as explained before. The term $\mathrm{d}t\mathrm{d}W(t)$ is of order $(\mathrm{d}t)^{3/2}$, because $\mathrm{d}W(t) = W(t+\mathrm{d}t) - W(t) \sim \mathcal{N}(0, \mathrm{d}t)$ is of order $(\mathrm{d}t)^{1/2}$. So it should also be neglected. At the end, we get:

$$\mathrm{d}\langle X\rangle_t = H^2(t)\,\mathrm{d}t.$$

In other words, the processes (64) and (65) have the same quadratic variation. Adding a usual (non-stochastic) integral does not influence the quadratic variation.

Take expectation in (65): $\mathbf{E}(X(t) - X(0)) = \int_0^t \mathbf{E}G(s)\,\mathrm{d}s$. In particular, if

$$\mathbf{E}X(t) = \mathbf{E}X(0) \quad \text{for allt } t \ge 0, \quad \text{then} \quad \mathbf{E}G(s) = 0 \quad \text{for all} \quad s \ge 0. \tag{66}$$

18.5. **Itô's formula.** This is the main formula in stochastic calculus. Take a stochastic process $X = (X(t),\, t \ge 0)$ with quadratic variation $\langle X\rangle_t$. Consider a function $f : \mathbb{R} \to \mathbb{R}$. Apply f to $X(t)$ and get $Y(t) = f(X(t))$. Then

$$\mathrm{d}Y(t) = f'(X(t))\,\mathrm{d}X(t) + \frac{1}{2}f''(X(t))\,\mathrm{d}\langle X\rangle_t. \tag{67}$$

Let us explain this: use Taylor decomposition

$$f(y) - f(x) \approx f'(x)(y-x) + \frac{1}{2}f''(x)(y-x)^2 \text{ for } y \approx x.$$

$$\begin{aligned}
\mathrm{d}Y(t) =& Y(t+\mathrm{d}t) - Y(t) = f(X(t+\mathrm{d}t)) - f(X(t)) \\
&= f'(X(t))(X(t+\mathrm{d}t) - X(t)) + \frac{1}{2}f''(X(t))\left(X(t+\mathrm{d}t) - X(t)\right)^2 \\
&= f'(X(t))\,\mathrm{d}X(t) + \frac{1}{2}f''(X(t))(\mathrm{d}X(t))^2 = f'(X(t))\,\mathrm{d}X(t) + \frac{1}{2}f''(X(t))\mathrm{d}\langle X\rangle_t.
\end{aligned}$$

Example 18.4. Let $f(x) = x^2$ and $X(t) = W(t)$. Then $f'(x) = 2x$ and $f''(x) = 2$. Also, $\langle W\rangle_t = t$. Therefore, applying (67), we get:

$$\mathrm{d}Y(t) = \mathrm{d}W^2(t) = 2W(t)\,\mathrm{d}W(t) + \mathrm{d}t.$$

We can write this as (because $W(0) = 0$):

$$W^2(t) = 2\int_0^t W(s)\,\mathrm{d}W(s) + t \ \Rightarrow\ \int_0^t W(s)\,\mathrm{d}W(s) = \frac{W^2(t) - t}{2}. \tag{68}$$

For smooth functions f with $f(0) = 0$, we have:

$$\int_0^t f(s)\,\mathrm{d}f(s) = \int_0^t f(s)f'(s)\,\mathrm{d}s = \frac{f^2(t)}{2}.$$

This additional term $-t/2$ is due to difference between ordinary and stochastic calculus. From this expression (68), we can immediately calculate the quadratic variation of $W^2(t)$:

$$\langle W^2\rangle_t = \int_0^t 4W^2(s)\,\mathrm{d}s.$$

For a function $f(t,x)$, Itô's formula takes form

$$\mathrm{d}f(t,X(t)) = \frac{\partial f}{\partial t}(t,X(t))\,\mathrm{d}t + \frac{\partial f}{\partial x}(t,X(t))\,\mathrm{d}X(t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,X(t))\mathrm{d}\langle X\rangle_t. \tag{69}$$

Example 18.5. Find quadratic variation of the process $tW(t)$. Apply Itô's formula to $f(t,x) = tx$ and $X(t) = W(t)$, with $\langle W\rangle_t = t$. We have:

$$\frac{\partial f}{\partial t} = x,\ \frac{\partial f}{\partial x} = t,\ \frac{\partial^2 f}{\partial x^2} = 0.$$

Therefore,

$$\mathrm{d}(tW(t)) = W(t)\,\mathrm{d}t + t\,\mathrm{d}W(t)\ \Rightarrow\ \langle (tW(t))\rangle_t = \int_0^t s^2\,\mathrm{d}s = \frac{t^3}{3}.$$

Let us introduce a *geometric Brownian motion* with drift g and diffusion σ^2, starting from x:

$$X(t) = xe^{gt+\sigma W(t)}.$$

Example 18.6. Apply Itô's formula to geometric Brownian motion $X(t)$, starting from $x = 3$, with parameters $g = 2$ and $\sigma = 3$, and to the function $f(x) = x^2 t$. We have:

$$X(t) = 3e^{2t+3W(t)} = e^{Y(t)},\ Y(t) := \ln 3 + 2t + 3W(t).$$

We need to solve this problem in three steps:

1. Find $\mathrm{d}X(t)$ and $\langle X\rangle_t$. To this end, apply Itô's formula to $g(x) := e^x$ and $Y(t)$. Then $g'(x) = g''(x) = e^x$, and $g'(Y(t)) = g''(Y(t)) = X(t)$. Also, $\mathrm{d}Y(t) = 2\mathrm{d}t + 3\mathrm{d}W(t)$, so $\langle Y\rangle_t = 9t$. Therefore,

$$\begin{aligned}\mathrm{d}X(t) &= \mathrm{d}g(Y(t)) = X(t)\,\mathrm{d}Y(t) + \frac{1}{2}X(t)\mathrm{d}\langle Y\rangle_t \\ &= X(t)\left(2\mathrm{d}t + 3\mathrm{d}W(t) + \frac{9}{2}\mathrm{d}t\right) = X(t)\left(\frac{13}{2}\mathrm{d}t + 3\mathrm{d}W(t)\right).\end{aligned}$$

From here, we get:

$$\langle X\rangle_t = \int_0^t (9X^2(s))\,\mathrm{d}s\ \Rightarrow\ \mathrm{d}\langle X\rangle_t = 9X^2(t)\,\mathrm{d}t.$$

2. Calculate

$$\frac{\partial f}{\partial t} = x^2,\ \frac{\partial f}{\partial x} = 2tx,\ \frac{\partial^2 f}{\partial x^2} = 2t.$$

3. Apply (69). We have:

$$\begin{aligned}\mathrm{d}f(t,X(t)) =& X^2(t)\,\mathrm{d}t + 2tX(t)\mathrm{d}X(t) + t\,\mathrm{d}\langle X\rangle_t \\ &= 2X^2(t)\,\mathrm{d}t + 13tX^2(t)\,\mathrm{d}t + 6tX^2(t)\,\mathrm{d}W(t) + 9tX^2(t)\,\mathrm{d}t \\ &= (2+22t)X^2(t)\,\mathrm{d}t + 6tX^2(t)\,\mathrm{d}W(t).\end{aligned}$$

18.6. **Hedging a European derivative.** Consider a European derivative from the previous section: $f(S(T))$, where $S(t) = S_0 e^{gt+\sigma W(t)}$ is the Black-Scholes model of a stock price, T is the maturity, and $f(x)$ is a real-valued function. Assume at time t we constuct a portfolio: We split our wealth $V(t)$ at time t between $H(t)$ shares of the stock and $V(t) - H(t)S(t)$ in cash. Then the change in this wealth during time $[t, t+\mathrm{d}t]$ is equal to $H(t)(S(t+\mathrm{d}t) - S(t)) = H(t)\,\mathrm{d}S(t)$. Thus $\mathrm{d}V(t) = H(t)\,\mathrm{d}S(t)$. We need $V(T) = f(S(T))$: at time T our wealth should exactly match the derivative. Then $V(0) = v$ would be the *fair price* at time $t = 0$, and $V(t)$ is the *fair price* at time t. The function $H(t)$ gives our *hedging strategy*, or *replicating portfolio*. Let us find $V(t)$ and $H(t)$. As discussed in the previous section, the fair price at time t given $S(t) = x$ is given by

$$u(t,x) = \mathbf{E}(f(S(T)) \mid S(t) = x),\ S(t) = S_0\exp(\sigma W(t) - \sigma^2 t/2);\quad u(t,S(t)) = V(t). \tag{70}$$

Note that $\mathbf{E}u(t, S(t)) = \mathbf{E}f(S(T))$, because the expectation of a conditional expectation is always equal to the *unconditional expectation*. Therefore, $\mathbf{E}u(t, S(t))$ does not depend on t: It is constant. Now, decompose $u(t, S(t))$ according to Itô's formula:

$$\mathrm{d}u(t, S(t)) = \frac{\partial u}{\partial t}(t, S(t))\,\mathrm{d}t + \frac{\partial u}{\partial x}(t, S(t))\,\mathrm{d}S(t) + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t, S(t))\,\mathrm{d}\langle S\rangle_t. \tag{71}$$

Apply Itô's formula with $f(t, x) = e^{\sigma x - \sigma^2 t/2}$ to $x = W(t)$. Then we have:

$$\frac{\partial f}{\partial t} = -\frac{\sigma^2}{2}f,\ \frac{\partial f}{\partial x} = \sigma f,\ \frac{\partial^2 f}{\partial x^2} = \sigma^2 f.$$

$$\mathrm{d}S(t) = \mathrm{d}f(t, W(t)) = \left[\frac{\partial f}{\partial t}(t, S(t)) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, S(t))\right]\,\mathrm{d}t + \frac{\partial f}{\partial x}(t, S(t))\,\mathrm{d}W(t) = \sigma f(t, S(t))\,\mathrm{d}W(t) = \sigma S(t)\,\mathrm{d}W(t).$$

Therefore, $\mathrm{d}\langle S\rangle_t = \sigma^2 S^2(t)\,\mathrm{d}t$. Thus we can rewrite (71) as

$$\mathrm{d}u(t, S(t)) = \left[\frac{\partial u}{\partial t}(t, S(t)) + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 u}{\partial x^2}(t, S(t))\right]\,\mathrm{d}t + \frac{\partial u}{\partial x}(t, S(t))\cdot\sigma S(t)\,\mathrm{d}W(t). \tag{72}$$

From (66), we get: For every $t \ge 0$,

$$\mathbf{E}\left[\frac{\partial u}{\partial t}(t, S(t)) + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 u}{\partial x^2}(t, S(t))\right] = 0.$$

For example, for $t = 0$ we have: $S(t) = S_0$, and the expression inside the expectation is a constant. Therefore, this constant is zero:

$$\frac{\partial u}{\partial t}(0, S_0) + \frac{1}{2}\sigma^2 S_0^2\frac{\partial^2 u}{\partial x^2}(0, S_0) = 0.$$

We can start from time t instead of $t = 0$, and apply this to $S(t)$ instead of S_0. Therefore,

$$\frac{\partial u}{\partial t}(t, S(t)) + \frac{1}{2}\sigma^2 S_0^2\frac{\partial^2 u}{\partial x^2}(t, S(t)) = 0.$$

Since this is true for all t and $S(t)$, the fair price function $u(t, x)$ satisfies the following partial differential equation:

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sigma^2 x^2\frac{\partial^2 u}{\partial x^2} = 0. \tag{73}$$

This is called the *Black-Scholes equation.* Therefore, we can rewrite (72) as follows:

$$\mathrm{d}u(t.S(t)) = \frac{\partial u}{\partial x}(t, S(t))\cdot\sigma S(t)\,\mathrm{d}W(t) = \frac{\partial u}{\partial x}(t, S(t))\,\mathrm{d}S(t).$$

Therefore, $\frac{\partial u}{\partial x}(t, S(t)) = H(t)$. This is called *delta hedging:* The derivatives of $u(t, x)$ with respect to volatility σ, current price x, current time t, are called *delta*, *vega*, and *theta*, respectively. They measure sensitivity of the option price with respect to the parameters of the model. Collectively, they are called *"greeks"*. The *delta derivative* gives us the quantity of shares we need to have at time t in our portfolio.

Example 18.7. $f(x) = x^2$, $\sigma = T = 1$. Then $S(t) = e^{W(t)-t/2}$, and the fair price is given by

$$V(t) = u(t, S(t)) = \mathbf{E}\left[\left(e^{W(1)-1/2}\right)^2 \mid S(t)\right] = \mathbf{E}\left[e^{2W(1)-1} \mid W(t)\right] = e^{2W(t)-1}\mathbf{E}\,e^{2(W(1)-W(t))}.$$

In the last derivation, we used that $W(1) = W(t) + (W(1) - W(t))$, and $W(1) - W(t)$ is independent of $W(t)$. Continuing our calculations, we have: Since $W(1) - W(t) \sim \mathcal{N}(0, 1-t)$, calculate the moment generating function:

$$V(t) = e^{2W(t)-1}e^{2^2(1-t)/2} = e^{2W(t)-1}e^{2(1-t)} = e^{2W(t)+1-2t}.$$

Now express this in terms of $S(t)$: $W(t) = t/2 + \ln S(t)$, and thus

$$V(t) = \exp\left(2(t/2 + \ln S(t)) + 1 - 2t\right) = \exp\left(2\ln S(t) + 1 - t\right) = S^2(t)e^{1-t}.$$

Therefore, the fair price function $u(t, x) = x^2 e^{1-t}$, and the delta hedging gives us:

$$\frac{\partial u}{\partial x} = 2xe^{1-t},\ H(t) = \frac{\partial u}{\partial x}(t, S(t)) = 2S(t)e^{1-t}.$$

As an example, if $S(0.5) = 1.4$, then at time $t = 0.5$ we need to have $2\cdot 1.4\cdot e^{1-0.5} = 4.616$ shares of this stock.

Problems

For the next nine problems, apply Itô's formula to calculate $\mathrm{d}f(X(t))$:

Problem 18.1. $f(x) = e^x$ and $X(t) = W^2(t)$.

Problem 18.2. $f(x) = (x+1)^2$ and $X(t) = W(t)$.

Problem 18.3. $f(x) = \log x$ and $X(t) = W(t) + 1$.

Problem 18.4. $f(x) = \sin x$ and $X(t) = W^2(t) + t$.

Problem 18.5. $f(x) = \cos x$ and $X(t) = \int_0^t (W(s)+1)\,\mathrm{d}s + W(t)$.

Problem 18.6. $f(x) = e^x$ and $X(t) = \int_0^t W(s)\,\mathrm{d}W(s) + \int_0^t 3s^2\,\mathrm{d}W(s) - \cos t$.

Problem 18.7. $f(x) = x^2/2$ and $X(t) = \int_0^t (W^2(s) + e^s)\,\mathrm{d}W(s) + \int_0^t s^3\,\mathrm{d}s + e^t$.

Problem 18.8. $f(x) = \log x$ and $X(t) = \int_0^t \left(e^{W(s)} + e^{-2s}\right)\,\mathrm{d}W(s) + 2t + 2$.

Problem 18.9. $f(x) = \log x$ and $X(t)$ being a Brownian motion with drift $g = 3$ and diffusion $\sigma^2 = 4$, starting from $x = 1$.

For the following four problems, apply Itô's formula to calculate $\mathrm{d}f(t, X(t))$.

Problem 18.10. $f(t, x) = e^{x+t}$ and $X(t) = W(t)$.

Problem 18.11. $f(t, x) = (x+t)^2$ and $X(t) = W(t)$.

Problem 18.12. $f(t, x) = x^2 t$ and $X(t) = e^{W(t)-3t+1}$.

Problem 18.13. $f(t, x) = te^x$ and $X(t) = 2W(t) - 3t + 1$.

For the following ten problems, find the quadratic variation $\langle X\rangle_t$ for the processes:

Problem 18.14. A Brownian motion with drift $g = 2$ and diffusion $\sigma^2 = 4$, starting from -1.

Problem 18.15. A geometric Brownian motion with drift -4 and diffusion 1, starting from 3;

Problem 18.16. $X(t) = \frac{1}{2}W^2(t) + 3t^2$.

Problem 18.17. $X(t) = tW(t)$.

Problem 18.18. $X(t) = \int_0^t \left(e^{W(s)} + e^s\right)\,\mathrm{d}W(s) + \int_0^t e^{2W(s)}\,\mathrm{d}s + 3$.

Problem 18.19. $X(t) = W^3(t)$.

Problem 18.20. A geometric Brownian motion with drift $g = 4$ and diffusion $\sigma^2 = 0.5$, starting from $x = 0.3$.

Problem 18.21. $X(t) = tW^2(t) + t^3 e^t$.

Problem 18.22. $X(t) = Y^4(t)$, where $Y = (Y(t),\, t \ge 0)$ is a Brownian motion with drift $g = 2$ and diffusion $\sigma = 3$, starting from $Y(0) = -2$.

For the next three problems, find the distribution of the following random variables:

Problem 18.23. $\int_0^2 (3-t)\,\mathrm{d}W(t)$.

Problem 18.24. $\int_0^1 e^{-t}\,\mathrm{d}W(t)$.

Problem 18.25. $\int_0^1 t\,\mathrm{d}X(t)$, where X is a Brownian motion with drift $g = 2$ and diffusion $\sigma^2 = 4$, starting from -1.

For the next two problems, apply Itô's formula to find the following stochastic ddifferentials:

Problem 18.26. $\mathrm{d}\cos(t + W(t))$.

Problem 18.27. $\mathrm{d}X(t)$, where $X(t) := \sqrt{W(t)^2 + 1}$.

For the next four problems, find the fair price function at time t, and the delta hedging for the derivative $f(S(T))$ with maturity T, initial stock price S_0 and volatility σ.

Problem 18.28. $T = 4$, $\sigma = \sqrt{3}$, $S_0 = 1$, European option call with strike $K = 1$.

Problem 18.29. $T = 1$, $\sigma = 1.5$, $S_0 = 2$, $f(x) = x^3$.

Problem 18.30. $T = 4$, $\sigma = 2$, $S_0 = 1$, barrier option with $K = 2$.

Problem 18.31. $T = 0.5$, $\sigma = 1.5$, $S_0 = 0.4$, $f(x) = x^2 + 2x$

19. Stochastic Differential Equations

19.1. **Definition.** Take two functions $g, \sigma : \mathbb{R} \to \mathbb{R}$. An equation

$$\mathrm{d}X(t) = g(X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t) \tag{74}$$

is called a *stochastic differential equation* (SDE) with *drift* g and *diffusion* σ. If we impose the *initial condition* $X(0) = x_0$, then this equation (under some conditions) has a unique solution. One can rewrite (74) together with this initial condition in the form

$$X(t) = x_0 + \int_0^t g(X(s))\,\mathrm{d}s + \int_0^t \sigma(X(s))\,\mathrm{d}W(s). \tag{75}$$

Example 19.1. If $g(x) \equiv g$ and $\sigma(x) \equiv \sigma$ are constant functions, then the equation (75) takes the form

$$X(t) = x_0 + \int_0^t g\,\mathrm{d}s + \int_0^t \sigma\,\mathrm{d}W(s) = x_0 + gt + \sigma W(t).$$

This is a Brownian motion with drift g and diffusion σ.

19.2. **Geometric Brownian motion.** Let $g(x) = g_0 x$ and $\sigma(x) = \sigma_0 x$ for some constants g_0 and σ_0. Then (74) takes the form

$$\mathrm{d}X(t) = X(t)\left[g_0\,\mathrm{d}t + \sigma_0\,\mathrm{d}W(t)\right]. \tag{76}$$

Let us show this is a geometric Brownian motion. From (76), we get:

$$\mathrm{d}\langle X\rangle_t = \sigma_0^2 X^2(t)\,\mathrm{d}t.$$

Apply Itô's formula to (76) with $f(x) = \log x$. Then

$$f'(x) = \frac{1}{x},\ f''(x) = -\frac{1}{x^2},$$

$$\mathrm{d}\log X(t) = \mathrm{d}f(X(t)) = f'(X(t))\,\mathrm{d}X(t) + \frac{1}{2}f''(X(t))\,\mathrm{d}\langle X\rangle_t \tag{77}$$

$$= \frac{1}{X(t)}\cdot X(t)\left[g_0\,\mathrm{d}t + \sigma_0\,\mathrm{d}W(t)\right] - \frac{1}{2X^2(t)}\sigma_0^2 X^2(t)\,\mathrm{d}t \tag{78}$$

$$= \mu\,\mathrm{d}t + \sigma_0 \mathrm{d}W(t), \tag{79}$$

where $\mu = g_0 - \sigma_0^2/2$ is the drift. Therefore,

$$\log X(t) = \log X(0) + \mu t + \sigma_0 W(t) \ \Rightarrow\ X(t) = X(0)e^{\mu t + \sigma_0 W(t)}.$$

Example 19.2. Let $\mathrm{d}X(t) = 4X(t)\,\mathrm{d}t + 0.5X(t)\,\mathrm{d}W(t)$, and $X(0) = 3$. Then $g_0 = 4$, $\sigma_0 = 0.5$, and $\mu = \frac{31}{8}$. Therefore,

$$X(t) = 3e^{\frac{31}{8}t + 0.5W(t)}.$$

Recall the moment generating functionfor normal random variables:

$$Z \sim \mathcal{N}(m, \rho^2) \ \Rightarrow\ \mathbf{E}e^{uZ} = \exp\left(mu + \frac{1}{2}\rho^2 u^2\right). \tag{80}$$

$$Z = \frac{31}{8}t + 0.5W(t) \ \Rightarrow\ m = \mathbf{E}Z = \frac{31}{8}t,\ \rho^2 = \operatorname{Var} Z = \operatorname{Var}(0.5W(t)) = 0.5^2 \operatorname{Var} W(t) = \frac{t}{4}. \tag{81}$$

Plugging this Z from (81) and $u = 1$ into (80), we get after calculation: $\mathbf{E}X(t) = 3\mathbf{E}e^Z = 3e^{4t}$. Similarly, plugging $u = 2$ into (81), we can find the second moment:

$$\mathbf{E}X^2(t) = 3^2\mathbf{E}e^{2Z} = \exp\left(\frac{33}{4}t\right).$$

19.3. **Ornstein-Uhlenbeck process.** Assume $g(x) = c(m-x)$ and $\sigma(x) = \sigma$ for $c, \sigma > 0$. Then the equation (74) takes the form

$$\mathrm{d}X(t) = c(m - X(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t). \tag{82}$$

This has *mean-reverting property:* if $X(t) > m$, then X tends to move, on average, down to m; if $X(t) < m$, then X tends to move, on average, up to m. Therefore, this process tends to oscillate around m. Compare this with Brownian motion, which just diffuses (as a random walk) to infinity, without any limits. Or with geometric Brownian motion, which does the same, but on the logarithmic scale.

The equation (82) is an example of a *linear* SDE, which has drift $g(x)$ and diffusion $\sigma(x)$ to be linear functions of x.

19.4. **Linear equations.** An SDE is called *linear* if the drift and diffusion coefficients are linear functions. For such SDE, we can find their mean and variance. As an example, consider the process

$$\mathrm{d}X(t) = (X(t) - 2)\,\mathrm{d}t + 2X(t)\,\mathrm{d}W(t),\ \ X(0) = -3. \tag{83}$$

Find $m(t) := \mathbf{E}X(t)$. We have: $m(0) = -3$, and

$$X(t) = -3 + \int_0^t (X(s) - 2)\,\mathrm{d}s + 2\int_0^t X(s)\,\mathrm{d}W(s).$$

Taking the expectation, note that the stochastic integral has expectation zero, and

$$m(t) = -3 + \int_0^t (m(s) - 2)\,\mathrm{d}s \ \Rightarrow\ m'(t) = m(t) - 2,\ m(0) = -3.$$

Let us now solve this linear SDE using integrating factor. Because $m' = m \ \Rightarrow\ m(t) = Ce^t$, the integrating factor is $m(t) = e^{-t}$, and so

$$\begin{aligned} m' - m = -2 \ &\Rightarrow\ e^{-t}m' - e^{-t}m = -2e^{-t} \ \Rightarrow\ (e^{-t}m)' = -2e^{-t} \\ &\Rightarrow\ e^{-t}m = \int (-2e^{-t})\,\mathrm{d}t \ \Rightarrow\ e^{-t}m = 2e^{-t} + C \ \Rightarrow\ \boxed{m = 2 + Ce^t} \end{aligned}$$

Because $m(0) = -3$, we find $2 + C = -3 \ \Rightarrow\ C = -5$. Thus,

$$m(t) = 2 - 5e^t.$$

Now, apply Itô's formula to (83) with $f(x) = x^2$, $f'(x) = 2x$, $f''(x) = 2$:

$$\mathrm{d}X^2(t) = f'(X(t))\,\mathrm{d}X(t) + \frac{1}{2}f''(X(t))\,\mathrm{d}\langle X\rangle_t = 2X(t)\,\mathrm{d}X(t) + \mathrm{d}\langle X\rangle_t.$$

But $d\langle X\rangle_t = 4X^2(t)\,\mathrm{d}t$. Therefore,

$$\begin{aligned} \mathrm{d}X^2(t) =& 2X(t)\left[(X(t) - 2)\,\mathrm{d}t + 2X(t)\,\mathrm{d}W(t)\right] + 4X^2(t)\,\mathrm{d}t \\ =& \left[6X^2(t) - 4X(t)\right]\,\mathrm{d}t + 4X^2(t)\,\mathrm{d}W(t). \end{aligned}$$

We can rewrite this as

$$X^2(t) = X^2(0) + \int_0^t \left[6X^2(s) - 4X(s)\right]\,\mathrm{d}s + \int_0^t 4X^2(s)\,\mathrm{d}W(s).$$

Take expectation and use that $X(0) = -3$:

$$\mathbf{E}X^2(t) = (-3)^2 + \int_0^t \mathbf{E}\left[6X^2(s) - 4X(s)\right]\,\mathrm{d}s.$$

Denote $a(t) := \mathbf{E}X^2(t)$. Then

$$a(t) = 9 + \int_0^t \left[6a(s) - 4m(s)\right]\,\mathrm{d}s.$$

We can rewrite this as

$$a'(t) = 6a(t) - 4m(t) = 6a(t) - 4(2 - 5e^t),\ \ a(0) = 9.$$

Now it remains to solve this linear ODE. Using the same integrating factor method, we solve this equation.

19.5. **General remarks.** Every SDE is a continuous-time Markov process with state space $\mathbb{R}$, the whole real line, unlike Markov chains, which have state space $\{1, 2, 3\}$ or another finite or countable set. This Markov process has transition density $p(t, x, y)$: the density of $X(t)$ at point y given that it started from $X(0) = x$. We already know this transition density for Brownian motion. For the general SDE, this transition density is a solution of two PDEs, called *forward* and *backward Kolmogorov equations.*

Sometimes this SDE has a stationary distribution. For example, Brownian motion and geometric Brownian motion do not have a stationary distribution, but Ornstein-Uhlenbeck process does have one, and it is normal:

$$\mathcal{N}(m, \rho^2), \text{ with } \rho^2 := \frac{\sigma^2}{2c}.$$

The rate of convergence to such stationary distribution is known for the Ornstein-Uhlenbeck process, but not for the general SDE.

Problems

Problem 19.1. Show that the solution of an SDE

$$\mathrm{d}X(t) = 3X(t)\,\mathrm{d}t + 2X(t)\,\mathrm{d}W(t)$$

is a geometric Brownian motion, and find its drift and diffusion coefficients.

Problem 19.2. Find $\mathbf{E}X(t)$ for the solution of the following SDE:

$$\mathrm{d}X(t) = (2X(t) + 3)\,\mathrm{d}t + X(t)\,\mathrm{d}W(t),\ X(0) = 1.$$

Problem 19.3. Take the solution to the following SDE:

$$\mathrm{d}X(t) = X(t)\left(\mathrm{d}t + 2\mathrm{d}W(t)\right),\ X(0) = 1.$$

Show that X is a geometric Brownian motion, and find $\operatorname{Var} X(2)$ from the explicit solution.

For the next two problems, consider the SDE

$$\mathrm{d}X(t) = (2 + 3X(t))\,\mathrm{d}t + (X(t) - 1)\,\mathrm{d}W(t),\ X(0) = 0.$$

Problem 19.4. Find $\mathbf{E}X(t)$.

Problem 19.5. Find $\operatorname{Var} X(t)$.

For the next three problems, consider the following Ornstein-Uhlenbeck process:

$$\mathrm{d}X(t) = 2(1 - X(t))\,\mathrm{d}t + 2\mathrm{d}W(t),\ X(0) = 0.$$

Problem 19.6. Find $\mathbf{E}X(t)$.

Problem 19.7. Find $\operatorname{Var} X(t)$.

Problem 19.8. Find the stationary distribution, and show that $\mathbf{E}X(t)$ and $\operatorname{Var} X(t)$ converge to the mean and variance of this stationary distribution, respectively.

Problem 19.9. Solve the equation and show that its solution is a geometric Brownian motion:

$$\mathrm{d}X(t) = X(t)\left[-3 + \mathrm{d}W(t)\right],\ X(0) = 2.$$

Find $\mathbf{E}X^3(t)$ from this explicit formula.

20. Continuous-Time Martingales

The content of this section largely duplicates Section 13, adapting it for continuous time. We did not avoid some repetitions, doing this for the convenience of a reader.

20.1. **Definition and examples.** A process $M = (M(t),\, t \ge 0)$ is called a *martingale* if for every $0 \le s \le t$, we have:

$$\mathbf{E}\left[M(t) \mid M(u),\, 0 \le u \le s\right] = M(s) \ \Leftrightarrow\ \mathbf{E}\left[M(t) - M(s) \mid M(u),\, 0 \le u \le s\right] = 0. \tag{84}$$

If in (84) we have $\ge$ or $\le$ instead of equality, then M is called a *submartingale* or *supermartingale,* respectively. The equation (84) means that the best prediction of the future value at time t is the current value at time s. In particular, for a martingale M, we have $\mathbf{E}M(t) = \mathbf{E}M(0)$.

Example 20.1. A Brownian motion $W = (W(t),\, t \ge 0)$ is a martingale: Since $W(t) - W(s)$ is independent of $W(u)$, $0 \le u \le s$, and $W(t) - W(s) \sim \mathcal{N}(0, t-s)$, we have:

$$\mathbf{E}\left[W(t) \mid W(u),\, 0 \le u \le s\right] = \mathbf{E}\left[W(t) - W(s) \mid W(u),\, 0 \le u \le s\right] + \mathbf{E}\left[W(s) \mid W(u),\, 0 \le u \le s\right]$$
$$= \mathbf{E}\left[W(t) - W(s)\right] + W(s) = W(s).$$

Example 20.2. Compensated Poisson process $M(t) = N(t) - \lambda t$, where $N(t)$ is a Poisson process with intensity λ. Then $N(t) - N(s) \sim \mathrm{Poi}(\lambda(t-s))$ is independent of $N(u)$, $0 \le u \le s$, and we can write $M(t) - M(s) = N(t) - N(s) - \lambda(t-s)$. Thus,

$$\mathbf{E}\left[M(t) - M(s) \mid M(u),\, 0 \le u \le s\right] = \mathbf{E}\left[N(t) - N(s) \mid N(u),\, 0 \le u \le s\right] - \lambda(t-s)$$
$$= \mathbf{E}\left[N(t) - N(s)\right] - \lambda(t-s) = 0.$$

Note that in both of these examples, we used the property that $M(t) - M(s)$ is independent of $M(u)$, $0 \le u \le s$. Thus for every Lévy process $L(t)$, if $\mathbf{E}\left[L(t) - L(s)\right] = 0$, then this is a martingale. If $\mathbf{E}\left[L(t) - L(s)\right] \ge 0$, then L is a submartingale.

Example 20.3. Consider a Lévy process $L(t) = 1 + 2t + 2W(t) + \sum_{k=1}^{N(t)} Z_k$, with N a Poisson process with intensity $\lambda = 3$. Then it is a martingale if and only if $\mathbf{E}\left[L(t) - L(s)\right] = 2(t-s) + \lambda \cdot \mathbf{E} Z_1 = 0$.

We can consider a geometric Brownian motion $X(t) = X(0)\exp\left(gt + \sigma W(t)\right)$. When is it a martingale? We have:

$$X(t) = X(s)\exp\left(g(t-s) + \sigma(W(t) - W(s))\right), \quad s < t.$$

And $W(t) - W(s)$ is independent of $X(u)$, $0 \le u \le s$. Therefore, we can write

$$\mathbf{E}\left[X(t) \mid X(u),\, 0 \le u \le s\right] = X(s)\mathbf{E}\exp\left(g(t-s) + \sigma(W(t) - W(s))\right) = X(s)\exp\left((t-s)(g + \sigma^2/2)\right).$$

Indeed, $W(t) - W(s) \sim \mathcal{N}(0, t-s)$ and therefore $\mathbf{E}\exp(\sigma(W(t) - W(s))) = \exp(\sigma^2(t-s)/2)$. Thus, this geometric Brownian motion is a martingale if and only if $g + \sigma^2/2 = 0$. More generally, take a Lévy process $L = (L(t),\, t \ge 0)$. Then $(e^{L(t)},\, t \ge 0)$ is a martingale if and only if

$$\mathbf{E} e^{L(t) - L(s)} = 1 \quad \text{for every } 0 \le s < t.$$

Example 20.4. Consider $L(t) = W(t) + ct + \sum_{k=1}^{N(t)} Z_k$ with $Z_k \sim \mathcal{N}(0,1)$ and N having intensity 2. Then we can find the moment generating function, as in Section 17: Because $N(t) - N(s) \sim \mathrm{Poi}(2(t-s))$, and Z_k has moment generating function $e^{u^2/2}$

$$\mathbf{E}\left[e^{L(t) - L(s)}\right] = e^{c(t-s)}\mathbf{E}\left[e^{W(t) - W(s)}\right]\mathbf{E}\left[\exp\left(\sum_{k=N(s)+1}^{N(t)} Z_k\right)\right]$$
$$= e^{c(t-s)}\exp\left(\frac{1}{2}(t-s)\right)\exp\left(2(t-s)e^{1/2}\right) = \exp\left[\left(c + \frac{1}{2} + 2e^{1/2}\right)(t-s)\right].$$

Therefore, we need $c = -0.5 - 2e^{1/2}$ for e^L to be a martingale.

For every random variable ξ and a process $X = (X(t),\, t \ge 0)$, the following conditional expectation is a martingale:

$$M(t) = \mathbf{E}\left[\xi \mid X(u),\, 0 \le u \le t\right].$$

Indeed, $M(t)$ is the best prediction of ξ at time t; and the best prediction of *this* at earlier time s is equal to the best prediction of ξ at time s. In particular, take a stock price $S = (S(t),\, t \ge 0)$. This last remark applies to the fair price process $V(t)$ of a European derivative $f(S(T))$ with maturity T. Indeed, this fair price, as shown in Section 17, is

$$V(t) = \mathbf{E}_*\left[f(S(T)) \mid S(u),\, 0 \le u \le t\right],$$

where $\mathbf{E}_*$ is the expectation taken with respect to the risk-neutral probability, under which S is a martingale.

Finally, let us mention that any stochastic integral with respect to the Brownian motion is a martingale:

$$M(t) = \int_0^t X(s)\,\mathrm{d}W(s),\ t \ge 0,$$

if only $X(t)$ is itself dependent on $W(u)$, $0 \le u \le t$. Indeed, fix $0 \le s < t$ and split the interval $[s,t]$ into many small subintervals: $s = u_0 < u_1 < \ldots < u_N = t$. We can express the difference

$$M(t) - M(s) = \int_s^t X(u)\,\mathrm{d}W(u) \approx \sum_{k=0}^{N-1} X(u_k)\left(W(u_{k+1}) - W(u_k)\right). \tag{85}$$

If X is deterministic (non-random), then each summand in the right-hand side of (85) is an increment of the Brownian motion, which is independent of $W(u)$, $0 \le u \le s$ and has expectation zero. Therefore, its conditional expectation with respect to $M(u)$, $0 \le u \le s$, is zero. For stochastic (random) X, it is also possible to show, although not so easy.

20.2. **Optional stopping theorem.** For every continuous-time martingale $M = (M(t),\, t \ge 0)$, $(M(\varepsilon n),\, n = 0, 1, 2, \ldots)$ is a discrete-time martingale. And we can approximate a continuous-time martingale by such discrete-time martingales. Therefore, we can extend an optional stopping theorem from Section 13 for continuous-time martingales.

We say that τ is a *stopping time* with respect to a process $X = (X(t),\, t \ge 0)$, if for every $t \ge 0$, the event $\{\tau \le t\}$ depends only on $X(u)$, $0 \le u \le t$. One can think about this as time to sell the stock, based only on its observed prices so far.

Example 20.5. Stopping times τ:

$$\inf\{t \ge 0 \mid X(t) \le 0\}, \quad \inf\{t \ge 0 \mid X(t) \in [0,1]\},$$
$$\min(\inf\{t \ge 0 \mid X(t) = 0\}, 3), \quad \inf\{t \ge 0 \mid X(t-1) \ge 0\}$$

Example 20.6. These are *not* stopping times:

$$\tau = \inf\{t \ge 0 \mid X(t+1) \ge 0\}, \quad \tau = \sup\{t \ge 0 \mid X(t) = 0\}.$$

Optional stopping theorem. If $M(t)$ is a martingale, and τ is bounded, or $M(t)$, $0 \le t \le \tau$, is bounded, then

$$\mathbf{E}M(\tau) = \mathbf{E}M(t) \quad \text{for all} \quad t \ge 0.$$

The boundedness condition is essential. We can weaken it, but not get rid of it entirely. Indeed, consider a Brownian motion W starting with zero. Eventually it will hit 1, and so we let $\tau = \inf\{t \ge 0 \mid W(t) = 1\}$. Then $W(\tau) = 1$ but $W(0) = 0$. This does not contradict the optional stopping theorem, however, since τ is unbounded (we can wait for a very long time until we hit 1; in fact $\mathbf{E}\tau = \infty$), and $(W(t),\, 0 \le t \le \tau)$ is also unbounded (the Brownian motion can go arbitrarily far into the minus until it hits the level 1).

This corresponds to the *doubling strategy,* already discussed in Section 13: Play the coin until you end up with cumulative win. But this is only feasible if you are prepared to wait for unilimited time (unfeasible) and borrow arbitrarily much from someone (unfeasible). If you decide to wait only until time $T = 100$, then you get $\tau' := \min(\tau, 100)$, which is bounded. Then we have $\mathbf{E}W(\tau') = \mathbf{E}W(0) = 0$. Or if you decide to borrow up to 100, then you get $\tau'' := \inf\{t \ge 0 \mid W(t) \ge 1 \text{ or } W(t) \le -100\}$. Then $W(t)$ is bounded for $t \le \tau''$, and $\mathbf{E}W(\tau'') = \mathbf{E}W(0) = 0$.

20.3. **The Jensen inequality.** Recall from Section 13 that $g : \mathbb{R} \to \mathbb{R}$ is a *convex function* if

$$g(\lambda x + (1-\lambda)y) \le \lambda g(x) + (1-\lambda)g(y).$$

For example, $g(x) = x$ and $g(x) = x^2$ are convex functions, while $g(x) = \sin x$ is not. Equivalently, if you connect any two points on the graph of g by a segment, then it lies above the graph. For a twice differentiable function g, it is convex if and only if its second derivative is nonnegative: $g''(x) \ge 0$ for all x. Jensen's inequality says that if g is a convex function, then

$$\mathbf{E}g(Z) \ge g(\mathbf{E}Z). \tag{86}$$

One example of this is a well-known fact that $\mathbf{E}Z^2 \ge (\mathbf{E}Z)^2$. This is true, and $\mathbf{E}Z^2 - (\mathbf{E}Z)^2 = \operatorname{Var} Z \ge 0$. This immediately follows from (29): jsut apply $g(x) = x^2$.

Similarly, we can show (86) for conditional expectation instead of the unconditional:

$$\mathbf{E}\left[g(Z) \mid Y_1, \ldots, Y_n\right] \ge g\left(\mathbf{E}\left[Z \mid Y_1, \ldots, Y_n\right]\right). \tag{87}$$

20.4. **Jensen's inequality.** Take a martingale $M = (M(t),\, t \ge 0)$. Apply a convex function g to (84). By Jensen's inequality (87), we have:

$$\mathbf{E}\left[g(M(t)) \mid M(u),\, 0 \le u \le s\right] \ge g\left(\mathbf{E}\left[M(t) \mid M(u),\, 0 \le u \le s\right]\right) = g(M(t)).$$

Therefore, $g(X) = (g(M(t)),\, t \ge 0)$ is a submartingale.

Let us apply this to option pricing. We already discussed European options and other European derivatives in Section 17. Recall that a *call option* is the right to buy a stock at a certain *strike price* K. A European call option has *maturity time* T, when you can *exercise* this option: demand to buy this stock at price K. If the market price $S(T)$ of this stock at time T is less than K, then you can just buy the stock at the market price and forget about your option. Then your option does not have value. However, if the market price $S(T) \ge K$, then you should exercise this option, and its value is $S(T) - K$. In general, the value of this option is $(S(T) - K)_+ = g(S(T))$, where $g(x) = (x - K)_+$, $a_+ := \max(a, 0)$.

An American call option is different from a European one in the following way: the former can be exercised at any time until maturity N, while the latter must be exercised at maturity. Therefore, let τ be the time you decide to exercise your American call option to get the best expected value $\mathbf{E} g\left(S(\tau)\right)$. When is the best exercise time τ? This is a stopping time, since your decision to exercise at time n or not is based only on your observations until time n, but not on future observations. But the function g is convex (draw a graph and check). Frequently, the stock price X is modeled by a martingale. Then $g(S) = (g(S(t)),\, t \ge 0)$, is a submartingale. By the optional stopping theorem,

$$\mathbf{E} g\left(S(\tau)\right) \le \mathbf{E} g\left(S(T)\right).$$

Therefore, the best time to exercise your American call option is at maturity $t = T$. In fact, American and European call options are of the same value in this case. Additional freedom to choose exercise time does not give you anything.

20.5. **Doob's inequalities.** These are generalizations of Markov's and Chebyshev's inequalities. Take a nonnegative submartingale $X = (X(t),\, t \ge 0)$ and a number $\lambda > 0$. Then

$$\mathbf{P}\left(\max_{0 \le t \le T} X(t) \ge \lambda\right) \le \frac{\mathbf{E}X(T)}{\lambda}.$$

This is proved by approximating the continuous-time submartingale $(X(t),\, 0 \le t \le T)$ by discrete-time submartingales $(X(\varepsilon n),\ n = 0, 1, \ldots)$ for small $\varepsilon > 0$. In particular, applying a convex function $f(x) = x^2$ to a martingale $M = (M(t),\, t \ge 0)$, we get a submartingale $M^2(t)$. Therefore, we get *Kolmogorov's inequality:*

$$\mathbf{P}\left(\max_{0 \le t \le T} |M(t)| \ge \lambda\right) = \mathbf{P}\left(\max_{0 \le t \le T} M^2(t) \ge \lambda^2\right) \le \frac{\mathbf{E}M^2(T)}{\lambda^2}.$$

Example 20.7. For a geometric Brownian motion $M(t) = e^{W(t) - t/2}$, which is a martingale, apply Kolmorogov's inequality:

$$\mathbf{P}\left(\max_{0 \le t \le 4} M(t) \ge 20\right) \le \frac{1}{20^2}\mathbf{E}\left[e^{W(4)-2}\right]^2 = \frac{1}{400}\mathbf{E}\left[e^{2W(4)-4}\right]$$
$$= \frac{1}{400}\exp\left(4 \cdot 4/2 - 4\right) = \boxed{0.136}$$

Example 20.8. In the previous example, we can also apply the function $f(x) = x^3$, which is convex on $x > 0$: $f''(x) = 6x > 0$. Thus

$$\mathbf{P}\left(\max_{0 \le t \le 4} M(t) \ge 20\right) \le \frac{1}{20^3}\mathbf{E}\left[e^{W(4)-2}\right]^3 = \frac{1}{8000}\mathbf{E}\left[e^{3W(4)-6}\right]$$
$$= \frac{1}{8000}\exp\left(9 \cdot 4/2 - 6\right) > 1,$$

so it is useless...

Problems

For the next eight problems, find the value of parameter c when $(X(t),\, t \ge 0)$ is a martingale, submartingale, or supermartingale.

Problem 20.1. X is a Brownian motion with drift c and diffusion 25, starting from -2.

Problem 20.2. $X(t) = N(t) - 4t$, where $N(t)$ is the Poisson process with intensity c.

Problem 20.3. Compound Poisson process: $X(t) = \sum_{k=1}^{N(t)} Z_k$, where N is a Poisson process with intensity $\lambda = 3$, and $Z_k \sim \mathcal{N}(c, 2.3)$.

Problem 20.4. $X(t) = ct + \sum_{k=1}^{N(t)} Z_k$, where N is a Poisson process with intensity $\lambda = 1.5$, and $Z_k \sim \Gamma(3, 12)$.

Problem 20.5. $X(t) = 2 + ct + W(t) + \sum_{k=1}^{N(t)} Z_k$, where N has intensity $\lambda = 1$, and $Z_k \sim \mathcal{N}(-1, 4)$.

Problem 20.6. $X(t) = 2e^{3W(t) - ct}$

Problem 20.7. $X(t) = \exp\left(1 + 3W(t) + ct + \sum_{k=1}^{N(t)} Z_k\right)$, where N has intensity $\lambda = .5$, and $Z_k \sim \mathrm{Exp}(2)$.

Problem 20.8. $X(t) = \exp\left(W(t) - 3t + \sum_{k=1}^{N(t)} Z_k\right)$, where N has intensity $\lambda = 2.5$, and $Z_k \sim \mathcal{N}(c, 1)$.

For the next eight problems, take a process $X = (X(t),\ t \ge 0)$ and find whether τ is a stopping time.

Problem 20.9. $\tau := \inf\{t \ge 0 \mid X(t) \ge 2.45\}$.

Problem 20.10. $\tau := \inf\{t \ge 0 \mid X^2(t) - 2X^2(t)) = 0\}$.

Problem 20.11. $\tau := \inf\{t \ge 0 \mid X(2t) \le -0.3\}$.

Problem 20.12. $\tau := \inf\{t \ge 2 \mid X(.5t - 1) \in [0, 5.67]\}$.

Problem 20.13. $\tau := \sup\{t \le 3 \mid X(t) = 1\}$.

Problem 20.14. $\tau := \inf\{t \ge 2 \mid X(t) = X(t - 0.6) = 0\}$.

Problem 20.15. $\tau := \min(\inf\{t \ge 1.4 \mid X(t) \le 3\}, 3)$.

Problem 20.16. $\tau := \max(\inf\{t \ge 1.4 \mid X(t) \le 3\}, 3)$.

For the next four problems, find whether the following functions are convex.

Problem 20.17. $g(x) = x^4 + 2x^2 - 23$.

Problem 20.18. $g(x) = x^4 - 2x^2$.

Problem 20.19. $g(x) = -\ln(x)$.

Problem 20.20. $g(x) = 3e^{-2x}$.

For the next six problems, apply Kolmogorov's inequality to estimate the following probabilities:

Problem 20.21. $\mathbf{P}(\max_{0 \le t \le 1} |W(t)| \ge 2)$.

Problem 20.22. Take a Poisson process N with intensity $\lambda = 2$. Estimate $\mathbf{P}(\max\limits_{0 \le t \le 3} |N(t) - 2t| \ge 10)$.

Problem 20.23. Take a Poisson process N with intensity 1, and $Z_k \sim \mathcal{N}(0, 2)$. Estimate

$$\mathbf{P}\big(\max_{0 \le t \le 3} \big|\sum_{k=1}^{N(t)} Z_k\big| \ge 4\big).$$

Problem 20.24. $\mathbf{P}(\max_{0 \le t \le 2} e^{2W(t) - 2t} \ge 30)$.

Problem 20.25. In the Black-Scholes model with volatility $\sigma = .4$ and initial stock price $S_0 = 2$, the probability that the fair price of an option $f(S(T))$ with maturity $T = 4.5$ and $f(x) = x^2$ never exceeds 100.

Problem 20.26.

$$\mathbf{P}\Big(\max_{0 \le t \le 1} \Big|\int_0^t W(t)\, \mathrm{d}W(t)\Big| \ge 100\Big).$$

Problem 20.27. Apply Doob's inequality with $f(x) = x^4$ to estimate $\mathbf{P}(\max_{0 \le t \le 1} |W(t)| \ge 2)$.

Problem 20.28. Apply Doob's inequality with $f(x) = x^5$ to estimate

$$\mathbf{P}\left(\max_{0 \le t \le 3} e^{4W(t) - 8t} \ge 200\right).$$

21. Appendix. Simulations

In this section, we outline simple ways to simulate random variables and random proceeces. At the beginning of each subsection, we indicate to which section of the main theoretical text it corresponds.

21.1. **Bernoulli and binomial random variables.** (Section 4) We can simulate a Bernoulli random variable Z, which takes values 1 and 0 with probabilities p and $1-p$, as follows: Simulate $U \sim \mathrm{Uni}[0, 1]$. If $U \le p$, then $X := 1$; else $X := 0$. A binomial random variable $X \sim \mathrm{Bin}(N, p)$ can be simulated as the sum of N i.i.d. Bernoulli random variables: $X = Z_1 + \ldots + Z_N$.

21.2. **Discrete random variables.** (Sections 3, 4) Consider a random variable X which takes values $0, 1, 2, \ldots$ with probabilities $p_0, p_1, p_2, \ldots$ Split $[0, 1]$ into

$$[0, p_0],\ [p_0, p_0 + p_1],\ [p_0 + p_1, p_0 + p_1 + p_2],\ \ldots$$

Simulate $U \sim \mathrm{Uni}[0, 1]$ and consider these intervals one by one: If $0 \le U \le p_0$, then let $X := 0$; if $p_0 \le U \le p_0 + p_1$, then let $X := 1$; etc.

21.3. **Joint discrete distributions.** (Sections 3, 4) Similarly we can simulate jointly distributed discrete random variables. As an example, simulate (X, Y) with joint distribution

X	Y	Prob.
0	0	0.5
0	1	0.3
1	0	0.2

Simulate $U \sim \mathrm{Uni}[0, 1]$. Next,

$$\begin{cases} 0 \le U \le 0.5 \ \Rightarrow X := 0,\, Y := 0; \\ 0.5 \le U \le 0.8 \ \Rightarrow \ X := 0,\, Y := 1; \\ 0.8 \le U \le 1 \ \Rightarrow \ X := 1,\, Y := 0. \end{cases}$$

21.4. **Continuous random variables: the inverse function method.** (Section 5) Consider a random variable X with cumulative distribution function $F(x) := \mathbf{P}(X \le x)$, with density $p(x) = F'(x)$. Let us simulate it. Construct the inverse function:

$$F^{-1}(u) := x, \quad \text{which solves} \quad F(x) = u.$$

Note that F is increasing, so $F^{-1}(u) \le x$ if and only if $u \le F(x)$. Take a uniform $[0, 1]$ random variable U. Then we can take $X = F^{-1}(U)$, because it has the correct cumulative distribution function:

$$\mathbf{P}\left(F^{-1}(U) \le x\right) = \mathbf{P}\left(U \le F(x)\right) = F(x),$$

because for this uniform random variable U we have: $\mathbf{P}(U \le a) = a, \quad 0 \le a \le 1$.

Example 21.1. Try to simulate $X \sim \mathrm{Exp}(\lambda)$. It has density $p(x) = \lambda e^{-\lambda x}$ and cumulative distribution function

$$F(x) = \int_0^x p(y)\mathrm{d}y = \int_0^x \lambda e^{-\lambda y}\mathrm{d}y = \left(-e^{-\lambda y}\right)\Big|_{y=0}^{y=x} = 1 - e^{-\lambda x}, \quad x \ge 0.$$

(And $F(x) = 0$ for $x \le 0$.) Now, let us find the inverse function. To this end, let us solve $F(x) = u$:

$$1 - e^{-\lambda x} = u \ \Rightarrow \ e^{-\lambda x} = 1 - u \ \Rightarrow \ x = -\lambda^{-1}\ln(1 - u).$$

Now, let $\boxed{X = -\lambda^{-1}\ln(1 - U)}$

21.5. **Rejection method.** (Section 5) Sometimes it is too difficult to simulate X using the inverse function method. For example, let $X \sim \mathcal{N}(0, 1)$, then

$$F(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-y^2/2}\mathrm{d}y,$$

which is difficult to calculate and even more difficult to invert. You can do this, using tables, but there is a better way. Let $f(x)$ be the density of X. In this case, it is

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

It is bounded from above by $c = (2\pi)^{-1/2}$. Moreover, $-10 \le X \le 10$ with very high probability. Then we can simulate independent $X \sim \mathrm{Uni}[-10, 10]$ and $Y \sim \mathrm{Uni}[0, c]$. The point (X, Y) is then uniformly distributed in the rectangle $[-10, 10] \times [0, c]$. We *accept* X if this point lies below the density curve: that is, if $f(X) \ge Y$. Otherwise, we *reject* this point and repeat this again, until we get an acceptable point (X, Y).

This method works equally well for multivariate distributions.

Example 21.2. Simulate (X, Y) with density $p(x, y) = x + y$ for $x, y \in [0, 1]$. Then $p \le 2$. Simulate independent $X \sim \mathrm{Uni}[0, 1]$, $Y \sim \mathrm{Uni}[0, 1]$, $Z \sim \mathrm{Uni}[0, 2]$, and accept (X, Y) if $p(X, Y) = X + Y \ge Z$.

21.6. **Value at risk.** (Section 6) Assume we have i.i.d. claims $X_1, \ldots, X_n$, and we would like to find value-at-risk corresponding to the level of confidence 95% by simulation. Simulate these claims, rank them from top to bottom, and find the $.05 \cdot n$-th ranked value. This is the value-at-risk for the current simulation. Repeat this a lot of times and average the resulting values-at-risk.

21.7. **Monte Carlo method.** (Section 8) Assume you have an integral

$$I = \int_0^1 f(x)\mathrm{d}x,$$

which you cannot calculate directly. One way to approximate this integral is to split $[0, 1]$ into small subintervals and use Simpson's rule or other similar rules. Another way is to generate N i.i.d. uniform random variables $X_1, \ldots, X_N$ on $[0, 1]$ and let

$$I_N := \frac{1}{N}\sum_{k=1}^{N} f(X_k).$$

By Law of Large Numbers, because $\mathbf{E}f(X_k) = \int_0^1 f(x)\mathrm{d}x$, and $f(X_1), \ldots, f(X_N)$ are i.i.d., we have: $I_N \to I$. More generally, if we have

$$I := \int_{-\infty}^{\infty} f(x)p(x)\mathrm{d}x,$$

where p is some probability density, let us generate N i.i.d. random variables $X_1, \ldots, X_N$ with density p, and let

$$I_N := \frac{1}{N}\sum_{k=1}^{N} f(X_k) \to I, \quad N \to \infty,$$

because $\mathbf{E}f(X_k) = \int f(x)p(x)\mathrm{d}x = I$. This convergence is with speed $N^{-1/2}$, as follows from the Central Limit Theorem.

Example 21.3. Let $I = \int_{-\infty}^{\infty} e^{-x^4}\mathrm{d}x$. Generate $X_1, \ldots, X_N \sim \mathcal{N}(0, 1)$, and represent I as

$$I = \sqrt{2\pi}\int_{-\infty}^{\infty} f(x)p(x)\mathrm{d}x, \quad f(x) = e^{-x^4+x^2/2}, \quad p(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

Then we can approximate

$$I \approx I_N := \frac{1}{N}\sum_{k=1}^{N} f(X_k).$$

This method works equally well for multiple integrals. This is a particular advantage of Monte Carlo, because the usual methods become slow when the dimension grows (*"curse of dimensionality"*).

Example 21.4. Consider the following integral on the 10-dimensional block $[0, 1]^{10}$ in $\mathbb{R}^{10}$:

$$I = \int_0^1\int_0^1 \ldots \int_0^1 e^{x_1^2+x_2^2+\ldots+x_{10}^2}\mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_{10}.$$

To split it into subblocks $[0, 0.1]^{10}$ and similar, you need 10^{10} blocks. But it is just as easy to use Monte Carlo for this case as for dimension one. Generate i.i.d. 10-dimensional vectors $U_i = (U_{i1}, \ldots, U_{i10})$, with each component i.i.d. uniform on $[0, 1]$. Then let

$$I_N := \frac{1}{N}\sum_{i=1}^{N}\exp\left(U_{i1}^2 + U_{i2}^2 + \ldots + U_{i10}^2\right) \to I.$$

21.8. **Discrete-time Markov chains.** (Section 11) Consider the example with McDonalds-Subway. Let 0 and 1 correspond to McDonalds and Subway. Assume that the initial distribution is $x(0) = \begin{bmatrix}0.7 & 0.3\end{bmatrix}$. Let $U_0 \sim \mathrm{Uni}[0, 1]$. If $U_0 \le 0.7$, we let $X_0 := 0$; otherwise $X_0 := 1$. Recall the transition matrix:

$$P = \begin{bmatrix}0.8 & 0.2\\ 0.1 & 0.9\end{bmatrix}$$

Simulate $U_1 \sim \mathrm{Uni}[0, 1]$ independently of U_0. If $X_0 = 0$, let $X_1 := 0$ if $U_1 \le 0.8$, $X_1 := 1$ otherwise. If $X_0 = 1$, let $X_1 := 0$ if $U_1 \le 0.1$, and $X_1 := 1$ otherwise. If this Markov chain has more than two (say four) states, then we should consider four cases for U_1. In the same way, we simulate $X_2, X_3, \ldots$

Assume the Markov chain is irreducible. Then the occupation time shares converge to the stationary distribution π. One can empirically calculate the stationary distribtuion: Fix a large number N of steps. For each state i, find the number M_i of times $n = 1, \ldots, N$ when $X_n = i$. Then $p_i := M_i/N$ is an approximation for the stationary distribution π.

21.9. **Google's PageRank.** (Section 11) Take an oriented graph (edges are arrows), where states are Web pages and arrows are hyperlinks. Let us rank pages. A page is ranked highly if there are a lot of links pointing to it. However, the importance of each link is different. A link from an important Web page is also important. But if this important Web page provides lots of links, each of them is not so important. (A recommendation letter from Bill Gates carries more weight than the one from elementary school teacher. However, if Bill Gates is generous and writes thousands of letters every year, maybe his letter does not carry much weight after all.)

Let $r(i)$ be the rank of the ith Web page. It is called PageRank in honor of Larry Page (who founded Google with Sergey Brin as Ph.D. students at Stanford in 1998). Let $|i|$ be the number of links from page i. Then

$$r(i) = \sum_{j\to i}\frac{r(j)}{|j|}.$$

This is a mathematical formulation of the idea above. If we normalize $r(i)$: divide it by $R = \sum r(i)$, the sum over all existing Web pages, then we have a probability distribution $r(i)$ on the set of all Web pages. However, it is very hard to solve this system of equation explicitly. Brin and Page invented a Markov chain, which they called a *random crawler*: if at a certain moment it is at Web page i, it chooses randomly (uniformly) among the links from this page and moves along this link. This is, in effect, the random walk on a (oriented) graph. The transition matrix A of this Markov chain is given by

$$A_{ij} = \begin{cases}\frac{1}{|i|},\ i\to j;\\ 0,\ \text{otherwise.}\end{cases}$$

But we can rewrite the equation for r above as

$$r(i) = \sum_{j\to i} r(j)A_{ji} = \sum_j r(j)A_{ji}\quad\Rightarrow\quad r = rA.$$

Therefore, r is a stationary distribution for A. If we run this random crawler long enough, it converges to r. (In fact, we need to make some adjustments to make sure this process indeed converges.)

21.10. **Random walk.** (Section 12) Consider a random walk with $p = 0.75$, starting from $S_0 = 1$. We simulate i.i.d. random variables $Z_1, Z_2, \ldots$ with distribution

$$\mathbf{P}(Z_k = 1) = .75,\ \mathbf{P}(Z_k = -1) = .25.$$

We do this by taking i.i.d. $U_1, U_2, \ldots \sim \mathrm{Uni}[0,1]$ and letting $Z_k = 1$ if $U_k \le .75$, $Z_k = -1$ if $U > .75$. Then let

$$S_0 := 1,\ S_1 = S_0 + Z_1,\ S_2 = S_1 + Z_2, \ldots$$

21.11. **Poisson process.** (Section 14) Simulate i.i.d $T_1, T_2, \ldots \sim \mathrm{Exp}(\lambda)$ and let

$$\tau_1 := T_1,\ \tau_2 := \tau_1 + T_2,\ \tau_3 := \tau_2 + T_3, \ldots$$

Then we find $N(t)$ by comparing t and τ_k until $t < \tau_k$. If k is the first such that $t < \tau_k$, then $N(t) = k - 1$. For example, if $\tau_3 \le t$, but $t < \tau_4$, then $N(t) = 3$. For a compound Poisson process

$$X(t) = \sum_{k=1}^{N(t)} Z_k,$$

we aslo need to simulate i.i.d. steps $Z_1, Z_2, \ldots$ and let $X(t) = Z_1 + \ldots + Z_{k-1}$ (and $X(t) = 0$ in case $k = 0$).

21.12. **Continuous-time Markov chains.** (Section 15) The difference from the discrete-time case is that we cannot simulate the value $X(t)$ for every time t. Instead, we do exactly as for the Poisson process: We simulate the jump times, and the values of this process at these jump times. Essentially, we switch from continuous to discrete time. As an example, consider a continuous-time Markov chain $X = (X(t),\, t \ge 0)$ with generator

$$A = \begin{bmatrix}-3 & 1 & 2\\ 0 & -1 & 1\\ 1 & 1 & -2\end{bmatrix}$$

The corresponding discrete-time Markov chain $Y = (Y_n)_{n\ge 0}$ has transition matrix

$$P = \begin{bmatrix}0 & 1/3 & 2/3\\ 0 & 0 & 1\\ 0.5 & 0.5 & 0\end{bmatrix}$$

Simulate $Y_0 = X(0)$ as shown above. Assume $X(0) = Y_0 = 1$. Then the Markov chain will spend time $T_1 \sim \mathrm{Exp}(3)$ in this state. After exiting this state, it will go to 2 with probability 1/3 and to 3 with probability 2/3. One can simulate Y_1 as shown in the subsection devoted to discrete-time Markov chains: Let $U_1 \sim \mathrm{Uni}[0,1]$; if $U_1 \le 2/3$ then $Y_1 := 3$; otherwise $Y_1 := 2$. Assume $Y_1 = 3$. Then the Markov chain will spend time $T_2 \sim \mathrm{Exp}(2)$ in this state. After exiting this state, it will go to 1 and 2 with equal probabilities .5. Let $\tau_1 := T_1$ and $\tau_2 := \tau_1 + T_2$. In this way we can simulate jump times τ_k and jump positions Y_k. For every fixed $t \ge 0$, find the k such that $\tau_k \le t < \tau_{k+1}$ and assign $X(t) := Y_k$.

Similarly to discrete-time Markov chains, we can estimate the stationary distribution for irreducible continuous-time Markov chains by calculating the share of time spent in each state over a long time interval. To this end, we fix a large number N of jumps. Take each state i, and find all $k = 1, \dots, N$ such that $Y_k = i$. Sum T_{k+1} for all such k, and divide the sum over N. The result will be p_i, and $\begin{bmatrix} p_1 & p_2 & \dots & p_N \end{bmatrix}$ is an approximation to the stationary distribution.

21.13. **Brownian motion.** (Section 17) It can be approximated by a discrete-time random walk: Take t_0 to be the step size. Then

$$W(kt_0) \approx \sqrt{t_0}\,(Z_1 + \ldots + Z_k)\,,$$

where $Z_1, Z_2, \ldots$ are i.i.d. taking values ± 1 with equal probability. This follows from the Central Limit Theorem: $\mathbf{E} Z_k = 0$ and $\operatorname{Var} Z_k = 1$, therefore

$$\frac{Z_1 + \ldots + Z_k}{\sqrt{k}} \approx \mathcal{N}(0,1).$$

And $Z_1 + \ldots + Z_k \approx \sqrt{k}\mathcal{N}(0,1) = \mathcal{N}(0,k)$. But $W(kt_0) \sim \mathcal{N}(0, t_0 k) = \mathcal{N}(0,k)\sqrt{t_0}$. Thus, the simulation method is as follows: take a small t_0, say $t_0 = .01$, and let $W((k+1)t_0) := W(kt_0) + \sqrt{t_0} Z_{k+1}$, with initial value $W(0) := 0$.

21.14. **Stochastic integrals.** (Section 18) We approximate them by taking a small step t_0, as in the previous subsection. For example,

$$\int_0^2 W(s)\,\mathrm{d}W(s) \approx \sum_{k=0}^{2t_0-1} W(kt_0)\left(W((k+1)t_0) - W(kt_0)\right). \tag{88}$$

Having simulated the Brownian motion as in the previous subsection, we can then calculate the sum in (88).

21.15. **Stochastic differential equations.** (Section 19) Similarly to the Brownian motion, we take a small time step t_0. As an example, take

$$\mathrm{d}X(t) = -3X(t)\,\mathrm{d}t + X(t)\,\mathrm{d}W(t),\ \ X(0) = 1. \tag{89}$$

We approximate the equation (89) as follows:

$$X((k+1)t_0) - X(kt_0) = -3X(kt_0)t_0 + X(kt_0)Z_{k+1},\ \ k = 0, 1, 2, \ldots$$

where $Z_{k+1} := W((k+1)t_0) - W(kt_0)$ can be approximated by $\sqrt{t_0} Z_{k+1}$, with $Z_{k+1} = \pm 1$ with equal probabilities. Start from $X(0) = 1$, simulate Z_1 and calculate $X(t_0)$, then simulate Z_2 and calculate $X(2t_0)$, etc.

Problems

Problem 21.1. (Section 3) Simulate the random variables (X, Y) 10000 times, if they take values

$$(1, 0),\ (0, 1),\ (-1, 0),\ (0, -1)$$

with equal probability 0.25. Calculate the empirical covariance $\operatorname{Cov}(X, Y)$ from (9), and compare with the theoretical value.

Problem 21.2. (Section 3) Simulate the die flip 10000 times, and find the empirical mean and variance, using (7) and (8). Compare with theoretical value.

Problem 21.3. (Section 4) Simulate a geometric random variable 10000 times with probability of success $p = 20\%$, by simulating a sequence of Bernoulli trials and then taking the number of the first successful trial. Compute empirical mean and variance using (7) and (8), and compare with theoretical values.

Problem 21.4. (Section 4) Simulate 1000 times a Poisson random variable $\mathrm{Poi}(1.5)$ by splitting $[0, 1]$ according to this distribution. Find its empirical mean, and compare with theoretical value. Simulate also $\mathrm{Bin}(100, 0.15)$ by 100 Bernoulli trials at a time, repeat 1000 times. Compare empirical distributions with each other: find the difference

$$|\mathbf{P}(\mathrm{Poi}(1.5) = k) - \mathbf{P}(\mathrm{Bin}(100, 0.15) = k)|, \quad , k = 0, 1, 2, 3, 4.$$

For each distribution, find its empirical mean, and compare with theoretical value.

Problem 21.5. (Section 5) Simulate 10000 times the random variable X with density
$$p(x) = \frac{2}{\pi(1+x^2)},\ x \ge 0,$$
using the inverse function method. Calculate an empirical probability $\mathbf{P}(X \ge 2)$, and compare with theoretical value.

Problem 21.6. (Section 5) Simulate 1000 times the Gamma variable $\Gamma(3, 2.6)$, using that it is the sum of three exponential random variables, which can be generated by the inverse function method. Compare the empirical and theoretical mean and variance, using (7) and (8).

Problem 21.7. (Section 5) Simulate 1000 times the Gamma variable $\Gamma(3, 2.6)$ using rejection method. Cut its tail at some large enough point. Compare the empirical and theoretical mean and variance, using (7) and (8).

Problem 21.8. (Section 5) Simulate the following jointly distributed random variables 10000 times using the rejection method:
$$p(x, y) = y,\ x \le y \le x + 1,\ 0 \le x \le 1.$$
Calculate empirical covariance as in (9), and compare with the theoretical result.

Problem 21.9. (Section 5) Using Monte Carlo method with 10000 tries, simulate the double integral
$$\int_0^1\int_0^1 (x+y)\,\mathrm{d}x\,\mathrm{d}y.$$
Compare with theoretical value.

Problem 21.10. (Section 6) Simulate 100 i.i.d. claims of size $Z \sim \mathrm{Exp}(1.5)$. Repeat this 1000 times. Find the value-at-risk for confidence level 95%. Compare with theoretical value.

Problem 21.11. (Section 11) Simulate 10000 steps of a Markov chain with transition matrix
$$P = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.4 & 0.6 & 0 \\ 0.1 & 0.9 & 0 \end{bmatrix}$$
starting from 1. Find the stationary distribution by calculating the share of time spent in each state for this simulation. Compare with theoretical result.

Problem 21.12. (Section 11) Simulate 10000 steps of a Markov chain with transition matrix
$$P = \begin{bmatrix} 0.5 & 0.3 & 0.1 & 0.1 \\ 0.3 & 0.6 & 0.1 & 0 \\ 0.2 & 0.2 & 0 & 0.6 \\ 0 & 0.2 & 0.8 & 0 \end{bmatrix}$$
starting from 1. Find empirical probability of hitting 2 before 3. Compare with theoretical result.

Problem 21.13. (Section 12) Simulate 1000 times the random walk with $p = 0.3$ and $q = 0.7$, starting from 2. Find the empirical probability that it hits -1 before 4. Compare with theoretical value.

Problem 21.14. (Section 12) Simulate 10 steps of a random walk with $p = 0.6$ and $q = 0.4$, starting from -1. Repeat this simulation 10000 times. Find the empirical probability $\mathbf{P}(S_{10} = -3)$. Compare with theoretical value.

Problem 21.15. (Section 12) Simulate a stock price 1000 times given by geometric symmetric random walk, starting from $P_0 = 1$. Find the empirical fair value of the European option call at maturity $T = 30$ with strike $K = 2$.

Problem 21.16. (Section 14) Simulate 1000 times the first 50 jumps of a Poisson process N with intensity $\lambda = 2$. Calculate the empirical expectation $\mathbf{E}\,[N(1)N(2)]$, and compare it with the true value.

Problem 21.17. (Section 14) Simulate 1000 times the first 50 jumps of a compound Poisson process X with increments $\mathcal{N}(2.5, 4)$, and intensity $\lambda = 0.5$. Use this to find empirical value $\mathrm{Var}\,X(4)$ as in (8), and compare this with the true value.

Problem 21.18. (Section 14) An insurance company receives claims during the year according to the Poisson process $N(t)$ with intensity $\lambda = 2$. Each claim is distributed as $\mathrm{Exp}(2.3)$. We measure time in days (365 days in a year). From 100 simulations, find the value-at-risk for confidence level 90%, and compare with theoretical value.

Problem 21.19. (Section 15) Until time $T = 100$, simulate a continuous-time Markov chain with generator

$$A = \begin{bmatrix} -3 & 1 & 2 \\ 2 & -5 & 3 \\ 5 & 0 & -5 \end{bmatrix}$$

starting from state 1. Find the stationary distribution by calculating the share of time spent in each state. Compare with the theoretical value.

Problem 21.20. (Section 15) Simulate 100 times, starting from state 2, a continuous-time Markov chain with generator

$$A = \begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}$$

Find the empirical distribution of this chain at time $t = 10$. Compare with the theoretical value.

Problem 21.21. (Section 16) Simulate the M/M/1 queue with arrival intensity $\lambda = 2$ and service intensity $\mu = 4$, until time horizon $T = 100$. Find the empirical stationary distribution by calculating the share of time spent in each state. Compare this with the theoretical value.

Problem 21.22. (Section 17) Simulate a Brownian motion with step size .01 for time horizon $T = 10$. Repeat this 1000 times. Calculate the empirical probability $\mathbf{P}(W(10) > 2)$, and compare with true value.

Problem 21.23. (Section 17) Simulate a geometric Brownian motion X with drift $g = -.3$ and variance $\sigma^2 = 2$, starting from $X(0) = 0.3$, until for time horizon $T = 1$, with step size .01. Repeat this 1000 times. Calculate the empirical value of $\mathbf{E}X^2(1)$, and compare with the true value.

Problem 21.24. (Section 17) Find an empirical fair price of a *barrier option* which pays 1 if the price of a stock at some time until maturity $T = 2.5$ exceeds $K = 2$, if the initial price is $S_0 = 1.4$, and the volatility is $\sigma = 0.4$.

Problem 21.25. (Section 18) Simulate the following Itô integral 1000 times with step size 0.01:

$$\int_0^2 e^{W(s)}\,\mathrm{d}W(s).$$

Calculate the empirical variance as in (8), and compare with the true result.

Problem 21.26. (Section 19) Simulate the following Ornstein-Uhlenbweck process:

$$\mathrm{d}X(t) = -2X(t)\,\mathrm{d}t + 3\,\mathrm{d}W(t),\ \ X(0) = -2,$$

with step 0.01, for time horizon $T = 1$. Repeat this 1000 times. Find empirical mean $\mathbf{E}X(1)$, and compare with the true value.

Cumulative Probabilities of the Standard Normal Distribution

The table gives the probabilities $\alpha = \Phi(z)$ to the left of given z–values for the standard normal distribution.

For example, the probability that a standard normal random variable Z is less than 1.53 is found at the intersection of the 1.5 rows and the 0.03 column, thus $\Phi(1.53) = P(Z \leq 1.53) = 0.9370$. Due to symmetry it holds $\Phi(-z) = 1 - \Phi(z)$ for all z.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Quantiles of the Standard Normal Distribution

For selected probabilities α, the table shows the values of the quantiles z_α such that $\Phi(z_\alpha) = P(Z \leq z_\alpha) = \alpha$, where Z is a standard normal random variable.

The quantiles satisfy the relation $z_{1-\alpha} = -z_\alpha$.

α	0.9	0.95	0.975	0.99	0.995	0.999
z_α	1.282	1.645	1.960	2.326	2.576	3.090