Lecture Notes for Introductory Probability

Janko Gravner
Mathematics Department
University of California
Davis, CA 95616
gravner@math.ucdavis.edu

June 9, 2011

These notes were started in January 2009 with help from Christopher Ng, a student in Math 135A and 135B classes at UC Davis, who typeset the notes he took during my lectures. This text is not a treatise in elementary probability and has no lofty goals; instead, its aim is to help a student achieve the proficiency in the subject required for a typical exam and basic real-life applications. Therefore, its emphasis is on examples, which are chosen without much redundancy. A reader should strive to *understand* every example given and be able to *design and solve* a similar one. Problems at the end of chapters and on sample exams (the solutions to all of which are provided) have been selected from actual exams, hence should be used as a test for preparedness.

I have only one tip for studying probability: *you cannot do it half-heartedly*. You have to devote to this class several hours per week of concentrated attention to understand the subject enough so that standard problems become routine. If you think that coming to class and reading the examples while also doing something else is enough, you're in for an unpleasant surprise on the exams.

This text will always be available free of charge to UC Davis students. Please contact me if you spot any mistake. I am thankful to Marisano James for numerous corrections and helpful suggestions.

1 Introduction

The theory of probability has always been associated with gambling and many most accessible examples still come from that activity. You should be familiar with the basic tools of the gambling trade: a coin, a (six-sided) die, and a full deck of 52 cards. A fair coin gives you Heads (H) or Tails (T) with equal probability, a fair die will give you 1, 2, 3, 4, 5, or 6 with equal probability, and a shuffled deck of cards means that any ordering of cards is equally likely.

Example 1.1. Here are typical questions that we will be asking and that you will learn how to answer. This example serves as an illustration and you should not expect to understand how to get the answer yet.

Start with a shuffled deck of cards and distribute all 52 cards to 4 players, 13 cards to each. What is the probability that each player gets an Ace? Next, assume that you are a player and you get a single Ace. What is the probability now that each player gets an Ace?

Answers. If any ordering of cards is equally likely, then any position of the four Aces in the deck is also equally likely. There are $\binom{52}{4}$ possibilities for the positions (slots) for the 4 aces. Out of these, the number of positions that give each player an Ace is 13^4: pick the first slot among the cards that the first player gets, then the second slot among the second player's cards, then the third and the fourth slot. Therefore, the answer is $\frac{13^4}{\binom{52}{4}} \approx 0.1055$.

After you see that you have a single Ace, the probability goes up: the previous answer needs to be divided by the probability that you get a single Ace, which is $\frac{13 \cdot \binom{39}{3}}{\binom{52}{4}} \approx 0.4388$. The answer then becomes $\frac{13^4}{13 \cdot \binom{39}{3}} \approx 0.2404$.

Here is how you can quickly *estimate* the second probability during a card game: give the second ace to a player, the third to a different player (probability about 2/3) and then the last to the third player (probability about 1/3) for the approximate answer $2/9 \approx 0.22$.

History of probability

Although gambling dates back thousands of years, the birth of modern probability is considered to be a 1654 letter from the Flemish aristocrat and notorious gambler Chevalier de Méré to the mathematician and philosopher Blaise Pascal. In essence the letter said:

> I used to bet even money that I would get at least one 6 in four rolls of a fair die. The probability of this is 4 times the probability of getting a 6 in a single die, i.e., $4/6 = 2/3$; clearly I had an advantage and indeed I was making money. Now I bet even money that within 24 rolls of two dice I get at least one double 6. This has the same advantage ($24/6^2 = 2/3$), but now I am losing money. Why?

As Pascal discussed in his correspondence with Pierre de Fermat, de Méré's reasoning was faulty; after all, if the number of rolls were 7 in the first game, the logic would give the nonsensical probability 7/6. We'll come back to this later.

Example 1.2. In a family with 4 children, what is the probability of a 2:2 boy-girl split?

One common wrong answer: $\frac{1}{5}$, as the 5 possibilities for the number of boys are not equally likely.

Another common guess: close to 1, as this is the most "balanced" possibility. This represents the mistaken belief that symmetry in probabilities should very likely result in symmetry in the outcome. A related confusion supposes that events that are probable (say, have probability around 0.75) occur nearly certainly.

Equally likely outcomes

Suppose an experiment is performed, with n possible outcomes comprising a set S. *Assume* also that all outcomes are equally likely. (Whether this assumption is realistic depends on the context. The above Example 1.2 gives an instance where this is not a reasonable assumption.) An *event* E is a set of outcomes, i.e., $E \subset S$. If an event E consists of m different outcomes (often called "good" outcomes for E), then the probability of E is given by:

$$P(E) = \frac{m}{n}. \tag{1.1}$$

Example 1.3. A fair die has 6 outcomes; take $E = \{2, 4, 6\}$. Then $P(E) = \frac{1}{2}$.

What does the answer in Example 1.3 mean? Every student of probability should spend some time thinking about this. The fact is that it is very difficult to attach a meaning to $P(E)$ if we roll a die a single time or a few times. The most straightforward interpretation is that for *a very large number* of rolls about half of the outcomes will be even. Note that this requires at least the concept of a limit! This *relative frequency* interpretation of probability will be explained in detail much later. For now, take formula (1.1) as the definition of probability.

2 Combinatorics

Example 2.1. Toss three fair coins. What is the probability of exactly one Heads (H)?

There are 8 equally likely outcomes: HHH, HHT, HTH, HTT, THH, THT, TTH, TTT. Out of these, 3 have exactly one H. That is, $E = \{\text{HTT, THT, TTH}\}$, and $P(E) = 3/8$.

Example 2.2. Let us now compute the probability of a 2:2 boy-girl split in a four-children family. We have 16 outcomes, which we will *assume* are equally likely, although this is not quite true in reality. We list the outcomes below, although we will soon see that there is a better way.

BBBB	BBBG	BBGB	BBGG
BGBB	BGBG	BGGB	BGGG
GBBB	GBBG	GBGB	GBGG
GGBB	GGBG	GGGB	GGGG

We conclude that

$$\begin{aligned} P(\text{2:2 split}) &= \frac{6}{16} = \frac{3}{8}, \\ P(\text{1:3 split or 3:1 split}) &= \frac{8}{16} = \frac{1}{2}, \\ P(\text{4:0 split or 0:4 split}) &= \frac{2}{16} = \frac{1}{8}. \end{aligned}$$

Example 2.3. Roll two dice. What is the most likely sum?

Outcomes are ordered pairs (i, j), $1 \le i \le 6$, $1 \le j \le 6$.

sum	no. of outcomes
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1

Our answer is 7, and $P(\text{sum} = 7) = \frac{6}{36} = \frac{1}{6}$.

How to count?

Listing all outcomes is very inefficient, especially if their number is large. We will, therefore, learn a few counting techniques, starting with a trivial, but conceptually important fact.

Basic principle of counting. If an experiment consists of two stages and the first stage has m outcomes, while the second stage has n outcomes regardless of the outcome at the first stage, then the experiment as a whole has mn outcomes.

Example 2.4. Roll a die 4 times. What is the probability that you get different numbers?

At least at the beginning, you should divide every solution into the following three steps:

Step 1: Identify the set of equally likely outcomes. In this case, this is the set of all ordered 4-tuples of numbers $1, \dots, 6$. That is, $\{(a, b, c, d) : a, b, c, d \in \{1, \dots, 6\}\}$.

Step 2: Compute the number of outcomes. In this case, it is therefore 6^4.

Step 3: Compute the number of good outcomes. In this case it is $6 \cdot 5 \cdot 4 \cdot 3$. Why? We have 6 options for the first roll, 5 options for the second roll since its number must differ from the number on the first roll; 4 options for the third roll since its number must not appear on the first two rolls, etc. Note that the *set* of possible outcomes changes from stage to stage (roll to roll in this case), but their *number* does not!

The answer then is $\frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4} = \frac{5}{18} \approx 0.2778$.

Example 2.5. Let us now compute probabilities for de Méré's games.

In Game 1, there are 4 rolls and he wins with at least one 6. The number of good events is $6^4 - 5^4$, as the number of *bad* events is 5^4. Therefore

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

In Game 2, there are 24 rolls of two dice and he wins by at least one pair of 6's rolled. The number of outcomes is 36^{24}, the number of bad ones is 35^{24}, thus the number of good outcomes equals $36^{24} - 35^{24}$. Therefore

$$P(\text{win}) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.4914.$$

Chevalier de Méré overcounted the good outcomes in both cases. His count $4 \cdot 6^3$ in Game 1 selects a die with a 6 and arbitrary numbers for other dice. However, many outcomes have more than one six and are hence counted more than once.

One should also note that both probabilities are barely different from $1/2$, so de Méré was gambling *a lot* to be able to notice the difference.

Permutations

Assume you have n objects. The number of ways to fill n ordered slots with them is

$$n \cdot (n-1) \dots 2 \cdot 1 = n!,$$

while the number of ways to fill $k \leq n$ ordered slots is

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

Example 2.6. Shuffle a deck of cards.

- $P(\text{top card is an Ace}) = \frac{1}{13} = \frac{4 \cdot 51!}{52!}$.
- $P(\text{all cards of the same suit end up next to each other}) = \frac{4! \cdot (13!)^4}{52!} \approx 4.5 \cdot 10^{-28}$. This event never happens in practice.
- $P(\text{hearts are together}) = \frac{40!13!}{52!} = 6 \cdot 10^{-11}$.

To compute the last probability, for example, collect all hearts into a block; a good event is specified by ordering 40 items (the block of hearts plus 39 other cards) and ordering the hearts within their block.

Before we go on to further examples, let us agree that when the text says *without further elaboration*, that a random choice is made, this means that *all available choices are equally likely*. Also, in the next problem (and in statistics in general) *sampling with replacement* refers to choosing, at random, an object from a population, noting its properties, putting the object back into the population, and then repeating. *Sampling without replacement* omits the putting back part.

Example 2.7. A bag has 6 pieces of paper, each with one of the letters, E, E, P, P, P, and R, on it. Pull 6 pieces at random out of the bag (1) without, and (2) with replacement. What is the probability that these pieces, in order, spell $PEPPER$?

There are two problems to solve. For sampling without replacement:

1. An outcome is an ordering of the pieces of paper $E_1E_2P_1P_2P_3R$.
2. The number of outcomes thus is $6!$.
3. The number of good outcomes is $3!2!$.

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

For sampling with replacement, the answer is $\frac{3^3 \cdot 2^2}{6^6} = \frac{1}{2 \cdot 6^3}$, quite a lot smaller.

Example 2.8. Sit 3 men and 3 women at random (1) in a row of chairs and (2) around a table. Compute P(all women sit together). In case (2), also compute P(men and women alternate).

In case (1), the answer is $\frac{4!3!}{6!} = \frac{1}{5}$.

For case (2), pick a man, say John Smith, and sit him first. Then, we reduce to a row problem with 5! outcomes; the number of good outcomes is $3! \cdot 3!$. The answer is $\frac{3}{10}$. For the last question, the seats for the men and women are fixed after John Smith takes his seat and so the answer is $\frac{3!2!}{5!} = \frac{1}{10}$.

Example 2.9. A group consists of 3 Norwegians, 4 Swedes, and 5 Finns, and they sit at random around a table. What is the probability that all groups end up sitting together?

The answer is $\frac{3!\cdot 4!\cdot 5!\cdot 2!}{11!}$. Pick, say, a Norwegian (Arne) and sit him down. Here is how you count the good events. There are 3! choices for ordering the group of Norwegians (and then sit them down to one of both sides of Arne, depending on the ordering). Then, there are 4! choices for arranging the Swedes and 5! choices for arranging the Finns. Finally, there are 2! choices to order the two blocks of Swedes and Finns.

Combinations

Let $\binom{n}{k}$ be the number of different subsets with k elements of a set with n elements. Then,

$$\begin{aligned}\binom{n}{k} &= \frac{n(n-1)\dots(n-k+1)}{k!}\\ &= \frac{n!}{k!(n-k)!}\end{aligned}$$

To understand why the above is true, first choose a subset, then order its elements in a row to fill k ordered slots with elements from the set with n objects. Then, distinct choices of a subset and its ordering will end up as distinct orderings. Therefore,

$$\binom{n}{k}k! = n(n-1)\dots(n-k+1).$$

We call $\binom{n}{k} = n$ *choose* k or a *binomial coefficient* (as it appears in the binomial theorem: $(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$). Also, note that

$$\binom{n}{0} = \binom{n}{n} = 1 \quad \text{and} \quad \binom{n}{k} = \binom{n}{n-k}.$$

The *multinomial coefficients* are more general and are defined next.

The number of ways to divide a set of n elements into r (distinguishable) subsets of $n_1, n_2, \ldots, n_r$ elements, where $n_1 + \ldots + n_r = n$, is denoted by $\binom{n}{n_1 \ldots n_r}$ and

$$\begin{aligned}\binom{n}{n_1 \ldots n_r} &= \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\ldots-n_{r-1}}{n_r} \\ &= \frac{n!}{n_1!n_2!\ldots n_r!}\end{aligned}$$

To understand the slightly confusing word *distinguishable*, just think of painting n_1 elements red, then n_2 different elements blue, etc. These colors distinguish among the different subsets.

Example 2.10. A fair coin is tossed 10 times. What is the probability that we get exactly 5 Heads?

$$P(\text{exactly 5 Heads}) = \frac{\binom{10}{5}}{2^{10}} \approx 0.2461,$$

as one needs to choose the position of the five heads among 10 slots to fix a good outcome.

Example 2.11. We have a bag that contains 100 balls, 50 of them red and 50 blue. Select 5 balls at random. What is the probability that 3 are blue and 2 are red?

The number of outcomes is $\binom{100}{5}$ and all of them are equally likely, which is a reasonable interpretation of "select 5 balls at random." The answer is

$$P(\text{3 are blue and 2 are red}) = \frac{\binom{50}{3}\binom{50}{2}}{\binom{100}{5}} \approx 0.3189$$

Why should this probability be less than a half? The probability that 3 are blue and 2 are red is equal to the probability of 3 are red and 2 are blue and they cannot both exceed $\frac{1}{2}$, as their sum cannot be more than 1. It cannot be exactly $\frac{1}{2}$ either, because other possibilities (such as all 5 chosen balls red) have probability greater than 0.

Example 2.12. Here we return to Example 1.1 and solve it more slowly. Shuffle a standard deck of 52 cards and deal 13 cards to each of the 4 players.

What is the probability that each player gets an Ace? We will solve this problem in two ways to emphasize that you often have a choice in your set of equally likely outcomes.

The first way uses an outcome to be an ordering of 52 cards:

1. There are 52! equally likely outcomes.
2. Let the first 13 cards go to the first player, the second 13 cards to the second player, etc. Pick a slot within each of the four segments of 13 slots for an Ace. There are 13^4 possibilities to choose these four slots for the Aces.
3. The number of choices to fill these four positions with (four different) Aces is 4!.
4. Order the rest of the cards in 48! ways.

The probability, then, is $\frac{13^4 4! 48!}{52!}$.

The second way, via a small leap of faith, assumes that each set of the four *positions of the four Aces* among the 52 shuffled cards is equally likely. You may choose to believe this intuitive fact or try to write down a formal proof: the number of permutations that result in a given set F of four positions is independent of F. Then:

1. The outcomes are the positions of the 4 Aces among the 52 slots for the shuffled cards of the deck.
2. The number of outcomes is $\binom{52}{4}$.
3. The number of good outcomes is 13^4, as we need to choose one slot among 13 cards that go to the first player, etc.

The probability, then, is $\frac{13^4}{\binom{52}{4}}$, which agrees with the number we obtained the first way.

Let us also compute P(one person has all four Aces). Doing the problem the second way, we get

1. The number of outcomes is $\binom{52}{4}$.
2. To fix a good outcome, pick one player ($\binom{4}{1}$ choices) and pick four slots for the Aces for that player ($\binom{13}{4}$ choices).

The answer, then, is $\frac{\binom{4}{1}\binom{13}{4}}{\binom{52}{4}} = 0.0106$, a lot smaller than the probability of each player getting an Ace.

Example 2.13. Roll a die 12 times. P(each number appears exactly twice)?

1. An outcome consists of filling each of the 12 slots (for the 12 rolls) with an integer between 1 and 6 (the outcome of the roll).
2. The number of outcomes, therefore, is 6^{12}.
3. To fix a good outcome, pick two slots for 1, then pick two slots for 2, etc., with $\binom{12}{2}\binom{10}{2}\dots\binom{2}{2}$ choices.

The probability, then, is $\frac{\binom{12}{2}\binom{10}{2}\dots\binom{2}{2}}{6^{12}}$.

What is P(1 appears exactly 3 times, 2 appears exactly 2 times)?

To fix a good outcome now, pick three slots for 1, two slots for 2, and fill the remaining 7 slots by numbers $3, \dots, 6$. The number of choices is $\binom{12}{3}\binom{9}{2}4^7$ and the answer is $\frac{\binom{12}{3}\binom{9}{2}4^7}{6^{12}}$.

Example 2.14. We have 14 rooms and 4 colors, white, blue, green, and yellow. Each room is painted at random with one of the four colors. There are 4^{14} equally likely outcomes, so, for example,

$$P(\text{5 white, 4 blue, 3 green, 2 yellow rooms}) = \frac{\binom{14}{5}\binom{9}{4}\binom{5}{3}\binom{2}{2}}{4^{14}}.$$

Example 2.15. A middle row on a plane seats 7 people. Three of them order chicken (C) and the remaining four pasta (P). The flight attendant returns with the meals, but has forgotten who ordered what and discovers that they are all asleep, so she puts the meals in front of them at random. What is the probability that they all receive correct meals?

A reformulation makes the problem clearer: we are interested in P(3 people who ordered C get C). Let us label the people $1, \dots, 7$ and assume that 1, 2, and 3 ordered C. The outcome is a selection of 3 people from the 7 who receive C, the number of them is $\binom{7}{3}$, and there is a *single* good outcome. So, the answer is $\frac{1}{\binom{7}{3}} = \frac{1}{35}$. Similarly,

$$P(\text{no one who ordered C gets C}) = \frac{\binom{4}{3}}{\binom{7}{3}} = \frac{4}{35},$$

$$P(\text{a single person who ordered C gets C}) = \frac{3 \cdot \binom{4}{2}}{\binom{7}{3}} = \frac{18}{35},$$

$$P(\text{two persons who ordered C get C}) = \frac{\binom{3}{2} \cdot 4}{\binom{7}{3}} = \frac{12}{35}.$$

Problems

1. A California licence plate consists of a sequence of seven symbols: number, letter, letter, letter, number, number, number, where a letter is any one of 26 letters and a number is one among $0, 1, \dots, 9$. Assume that all licence plates are equally likely. (a) What is the probability that all symbols are different? (b) What is the probability that all symbols are different and the first number is the largest among the numbers?

2. A tennis tournament has $2n$ participants, n Swedes and n Norwegians. First, n people are chosen at random from the $2n$ (with no regard to nationality) and then paired randomly with the other n people. Each pair proceeds to play one match. An outcome is a *set* of n (ordered) pairs, giving the winner and the loser in each of the n matches. (a) Determine the number of outcomes. (b) What do you need to assume to conclude that all outcomes are equally likely? (c) Under this assumption, compute the probability that all Swedes are the winners.

3. A group of 18 Scandinavians consists of 5 Norwegians, 6 Swedes, and 7 Finns. They are seated at random around a table. Compute the following probabilities: (a) that all the Norwegians sit together, (b) that all the Norwegians and all the Swedes sit together, and (c) that all the Norwegians, all the Swedes, and all the Finns sit together.

4. A bag contains 80 balls numbered $1, \ldots, 80$. Before the game starts, you choose 10 different numbers from amongst $1, \ldots, 80$ and write them on a piece of paper. Then 20 balls are selected (without replacement) out of the bag at random. (a) What is the probability that all your numbers are selected? (b) What is the probability that none of your numbers is selected? (c) What is the probability that exactly 4 of your numbers are selected?

5. A full deck of 52 cards contains 13 hearts. Pick 8 cards from the deck at random (a) without replacement and (b) with replacement. In each case compute the probability that you get no hearts.

Solutions to problems

1. (a) $\frac{10\cdot 9\cdot 8\cdot 7\cdot 26\cdot 25\cdot 24}{10^4\cdot 26^3}$, (b) the answer in (a) times $\frac{1}{4}$.

2. (a) Divide into two groups (winners and losers), then pair them: $\binom{2n}{n}\cdot n!$. Alternatively, pair the first player, then the next available player, etc., and, then, at the end choose the winners and the losers: $(2n-1)(2n-3)\cdots 3\cdot 1\cdot 2^n$. (Of course, these two expressions are the same.) (b) All players are of equal strength, equally likely to win or lose any match against any other player. (c) The number of good events is $n!$, the choice of a Norwegian paired with each Swede.

3. (a) $\frac{13!\cdot 5!}{17!}$, (b) $\frac{8!\cdot 5!\cdot 6!}{17!}$, (c) $\frac{2!\cdot 7!\cdot 6!\cdot 5!}{17!}$.

4. (a) $\frac{\binom{70}{10}}{\binom{80}{20}}$, (b) $\frac{\binom{70}{20}}{\binom{80}{20}}$, (c) $\frac{\binom{10}{4}\binom{70}{16}}{\binom{80}{20}}$.

5. (a) $\frac{\binom{39}{8}}{\binom{52}{8}}$, (b) $\left(\frac{3}{4}\right)^8$.

3 Axioms of Probability

The question here is: how can we mathematically define a random experiment? What we have are *outcomes* (which tell you exactly what happens), *events* (sets containing certain outcomes), and *probability* (which attaches to every event the likelihood that it happens). We need to agree on which properties these objects must have in order to compute with them and develop a theory.

When we have finitely many equally likely outcomes all is clear and we have already seen many examples. However, as is common in mathematics, infinite sets are much harder to deal with. For example, we will soon see what it means to choose a random point within a unit circle. On the other hand, we will also see that there is no way to choose at random a positive integer — remember that "at random" means all choices are equally likely, unless otherwise specified.

Finally, a *probability space* is a triple $(\Omega, \mathcal{F}, P)$. The first object Ω is an arbitrary set, repesenting the set of outcomes, sometimes called the *sample space*.

The second object $\mathcal{F}$ is a collection of events, that is, a set of subsets of Ω. Therefore, an event $A \in \mathcal{F}$ is necessarily a subset of Ω. Can we just say that each $A \subset \Omega$ is an event? In this course *you can assume so without worry, although there are good reasons for not assuming so in general*! I will give the definition of what properties $\mathcal{F}$ needs to satisfy, but this is only for illustration and you should take a course in measure theory to understand what is really going on. Namely, $\mathcal{F}$ needs to be a *σ-algebra*, which means (1) $\emptyset \in \mathcal{F}$, (2) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$, and (3) $A_1, A_2, \dots \in \mathcal{F} \implies \cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

What is important is that you can take the complement A^c of an event A (i.e., A^c happens when A does not happen), unions of two or more events (i.e., $A_1 \cup A_2$ happens when either A_1 or A_2 happens), and intersections of two or more events (i.e., $A_1 \cap A_2$ happens when both A_1 and A_2 happen). We call events $A_1, A_2, \dots$ *pairwise disjoint* if $A_i \cap A_j = \emptyset$ if $i \neq j$ — that is, at most one of these events can occur.

Finally, the probability P is a number attached to every event A and satisfies the following three axioms:

<u>Axiom 1</u>. For every event A, $P(A) \geq 0$.

<u>Axiom 2</u>. $P(\Omega) = 1$.

<u>Axiom 3</u>. If $A_1, A_2, \dots$ is a sequence of pairwise disjoint events, then

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

Whenever we have an abstract definition such as this one, the first thing to do is to look for examples. Here are some.

Example 3.1. $\Omega = \{1, 2, 3, 4, 5, 6\}$,

$$P(A) = \frac{(\text{number of elements in } A)}{6}.$$

The random experiment here is rolling a fair die. Clearly, this can be generalized to any finite set with equally likely outcomes.

Example 3.2. $\Omega = \{1, 2, \ldots\}$ and you have numbers $p_1, p_2, \ldots \geq 0$ with $p_1 + p_2 + \ldots = 1$. For any $A \subset \Omega$,

$$P(A) = \sum_{i \in A} p_i.$$

For example, toss a fair coin repeatedly until the first Heads. Your outcome is the number of tosses. Here, $p_i = \frac{1}{2^i}$.

Note that p_i cannot be chosen to be equal, as you cannot make the sum of infinitely many equal numbers to be 1.

Example 3.3. Pick a point from inside the unit circle centered at the origin. Here, $\Omega = \{(x, y) : x^2 + y^2 < 1\}$ and

$$P(A) = \frac{\text{(area of A)}}{\pi}.$$

It is important to observe that if A is a singleton (a set whose element is a single point), then $P(A) = 0$. This means that we cannot attach the probability to outcomes — you hit a single point (or even a line) with probability 0, but a "fatter" set with positive area you hit with positive probability.

Another important theoretical remark: this is a case where A cannot be an arbitrary subset of the circle — for some sets area cannot be defined!

Consequences of the axioms

(C0) $P(\emptyset) = 0$.

Proof. In Axiom 3, take all sets to be $\emptyset$.

(C1) If $A_1 \cap A_2 = \emptyset$, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Proof. In Axiom 3, take all sets other than first two to be $\emptyset$.

(C2)

$$\boxed{P(A^c) = 1 - P(A).}$$

Proof. Apply (C1) to $A_1 = A$, $A_2 = A^c$.

(C3) $0 \leq P(A) \leq 1$.

Proof. Use that $P(A^c) \geq 0$ in (C2).

(C4) If $A \subset B$, $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

Proof. Use (C1) for $A_1 = A$ and $A_2 = B \setminus A$.

(C5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. Let $P(A \setminus B) = p_1$, $P(A \cap B) = p_{12}$ and $P(B \setminus A) = p_2$, and note that $A \setminus B$, $A \cap B$, and $B \setminus A$ are pairwise disjoint. Then $P(A) = p_1 + p_{12}$, $P(B) = p_2 + p_{12}$, and $P(A \cup B) = p_1 + p_2 + p_{12}$.

(C6)

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = {} & P(A_1) + P(A_2) + P(A_3) \\ & - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_2) \\ & + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

and more generally

$$\boxed{\begin{aligned} P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) + \dots \\ + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n). \end{aligned}}$$

This is called the *inclusion-exclusion formula* and is commonly used when it is easier to compute probabilities of intersections than of unions.

Proof. We prove this only for $n = 3$. Let $p_1 = P(A_1 \cap A_2^c \cap A_3^c)$, $p_2 = P(A_1^c \cap A_2 \cap A_3^c)$, $p_3 = P(A_1^c \cap A_2^c \cap A_3)$, $p_{12} = P(A_1 \cap A_2 \cap A_3^c)$, $p_{13} = P(A_1 \cap A_2^c \cap A_3)$, $p_{23} = P(A_1^c \cap A_2 \cap A_3)$, and $p_{123} = P(A_1 \cap A_2 \cap A_3)$. Again, note that all sets are pairwise disjoint and that the right hand side of (6) is

$$\begin{aligned} & (p_1 + p_{12} + p_{13} + p_{123}) + (p_2 + p_{12} + p_{23} + p_{123}) + (p_3 + p_{13} + p_{23} + p_{123}) \\ & - (p_{12} + p_{123}) - (p_{13} + p_{123}) - (p_{23} + p_{123}) \\ & + p_{123} \\ & = p_1 + p_2 + p_3 + p_{12} + p_{13} + p_{23} + p_{123} = P(A_1 \cup A_2 \cup A_3). \end{aligned}$$

Example 3.4. Pick an integer in $[1, 1000]$ at random. Compute the probability that it is divisible neither by 12 nor by 15.

The sample space consists of the 1000 integers between 1 and 1000 and let A_r be the subset consisting of integers divisible by r. The cardinality of A_r is $\lfloor 1000/r \rfloor$. Another simple fact is that $A_r \cap A_s = A_{\mathrm{lcm}(r,s)}$, where lcm stands for the least common multiple. Our probability equals

$$\begin{aligned} 1 - P(A_{12} \cup A_{15}) &= 1 - P(A_{12}) - P(A_{15}) + P(A_{12} \cap A_{15}) \\ &= 1 - P(A_{12}) - P(A_{15}) + P(A_{60}) \\ &= 1 - \frac{83}{1000} - \frac{66}{1000} + \frac{16}{1000} = 0.867. \end{aligned}$$

Example 3.5. Sit 3 men and 4 women at random in a row. What is the probability that either all the men or all the women end up sitting together?

Here, $A_1 = \{$all women sit together$\}$, $A_2 = \{$all men sit together$\}$, $A_1 \cap A_2 = \{$both women and men sit together$\}$, and so the answer is

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = \frac{4! \cdot 4!}{7!} + \frac{5! \cdot 3!}{7!} - \frac{2! \cdot 3! \cdot 4!}{7!}.$$

Example 3.6. A group of 3 Norwegians, 4 Swedes, and 5 Finns is seated at random around a table. Compute the probability that at least one of the three groups ends up sitting together.

Define $A_N = \{$Norwegians sit together$\}$ and similarly A_S, A_F. We have

$$\begin{aligned} &P(A_N) = \frac{3! \cdot 9!}{11!},\ P(A_S) = \frac{4! \cdot 8!}{11!},\ P(A_N) = \frac{5! \cdot 7!}{11!}, \\ &P(A_N \cap A_S) = \frac{3! \cdot 4! \cdot 6!}{11!},\ P(A_N \cap A_F) = \frac{3! \cdot 5! \cdot 5!}{11!},\ P(A_S \cap A_F) = \frac{4! \cdot 5! \cdot 4!}{11!}, \\ &P(A_N \cap A_S \cap A_F) = \frac{3! \cdot 4! \cdot 5! \cdot 2!}{11!}. \end{aligned}$$

Therefore,

$$P(A_N \cup A_S \cup A_F) = \frac{3! \cdot 9! + 4! \cdot 8! + 5! \cdot 7! - 3! \cdot 4! \cdot 6! - 3! \cdot 5! \cdot 5! - 4! \cdot 5! \cdot 4! + 3! \cdot 4! \cdot 5! \cdot 2!}{11!}.$$

Example 3.7. *Matching problem.* A large company with n employees has a scheme according to which each employee buys a Christmas gift and the gifts are then distributed at random to the employees. What is the probability that someone gets his or her own gift?

Note that this is different from asking, assuming that you are one of the employees, for the probability that *you* get your own gift, which is $\frac{1}{n}$.

Let $A_i = \{i$th employee gets his or her own gift$\}$. Then, what we are looking for is

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

We have

$$\begin{aligned}
P(A_i) &= \frac{1}{n} \quad \text{(for all } i\text{)},\\
P(A_i \cap A_j) &= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \quad \text{(for all } i<j\text{)},\\
P(A_i \cap A_j \cap A_k) &= \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)} \quad \text{(for all } i<j<k\text{)},\\
&\dots\\
P(A_1 \cap \dots \cap A_n) &= \frac{1}{n!}.
\end{aligned}$$

Therefore, our answer is

$$\begin{aligned}
& n \cdot \frac{1}{n} - \binom{n}{2} \cdot \frac{1}{n(n-1)} + \binom{n}{3} \cdot \frac{1}{n(n-1)(n-2)} - \ldots + (-1)^{n-1}\frac{1}{n!}\\
= \;& 1 - \frac{1}{2!} + \frac{1}{3!} + \ldots + (-1)^{n-1}\frac{1}{n!}\\
\to \;& 1 - \frac{1}{e} \approx 0.6321 \quad \text{(as } n \to \infty\text{)}.
\end{aligned}$$

Example 3.8. *Birthday Problem.* Assume that there are k people in the room. What is the probability that there are two who share a birthday? We will ignore leap years, assume all birthdays are equally likely, and generalize the problem a little: from n possible birthdays, sample k times with replacement.

$$P(\text{a shared birthday}) = 1 - P(\text{no shared birthdays}) = 1 - \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k}.$$

When $n = 365$, the lowest k for which the above exceeds 0.5 is, famously, $k = 23$. Some values are given in the following table.

k	prob. for $n = 365$
10	0.1169
23	0.5073
41	0.9032
57	0.9901
70	0.9992

Occurences of this problem are quite common in various contexts, so we give another example. Each day, the Massachusetts lottery chooses a four digit number at random, with leading 0's allowed. Thus, this is sampling with replacement from among $n = 10,000$ choices each day. On February 6, 1978, the *Boston Evening Globe* reported that

> "During [the lottery's] 22 months' existence [...], no winning number has ever been repeated. [David] Hughes, the expert [and a lottery official] doesn't expect to see duplicate winners until about half of the 10,000 possibilities have been exhausted."

What would an informed reader make of this? Assuming $k = 660$ days, the probability of no repetition works out to be about $2.19 \cdot 10^{-10}$, making it a remarkably improbable event. What happened was that Mr. Hughes, not understanding the Birthday Problem, did not check for repetitions, confident that there would not be any. Apologetic lottery officials announced later that there indeed were repetitions.

Example 3.9. *Coupon Collector Problem.* Within the context of the previous problem, assume that $k \geq n$ and compute $P(\text{all } n \text{ birthdays are represented})$.

More often, this is described in terms of cereal boxes, each of which contains one of n different cards (coupons), chosen at random. If you buy k boxes, what is the probability that you have a complete collection?

When $k = n$, our probability is $\frac{n!}{n^n}$. More generally, let

$$A_i = \{i\text{th birthday is missing}\}.$$

Then, we need to compute

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

Now,

$$\begin{aligned}
P(A_i) &= \frac{(n-1)^k}{n^k} \quad \text{(for all } i\text{)} \\
P(A_i \cap A_j) &= \frac{(n-2)^k}{n^k} \quad \text{(for all } i < j\text{)} \\
&\dots \\
P(A_1 \cap \dots \cap A_n) &= 0
\end{aligned}$$

and our answer is

$$\begin{aligned}
& 1 - n\left(\frac{n-1}{n}\right)^k + \binom{n}{2}\left(\frac{n-2}{n}\right)^k - \ldots + (-1)^{n-1}\binom{n}{n-1}\left(\frac{1}{n}\right)^k \\
&= \sum_{i=0}^{n} \binom{n}{i} (-1)^i \left(1 - \frac{i}{n}\right)^k.
\end{aligned}$$

This must be $\frac{n!}{n^n}$ for $k = n$, and 0 when $k < n$, neither of which is obvious from the formula. Neither will you, for large n, get anything close to the correct numbers when $k \leq n$ if you try to compute the probabilities by computer, due to the very large size of summands with alternating signs and the resulting rounding errors. We will return to this problem later for a much more efficient computational method, but some numbers are in the two tables below. Another remark for those who know a lot of combinatorics: you will perhaps notice that the above probability is $\frac{n!}{n^k} S_{k,n}$, where $S_{k,n}$ is the Stirling number of the second kind.

k	prob. for $n = 6$
13	0.5139
23	0.9108
36	0.9915

k	prob. for $n = 365$
1607	0.0101
1854	0.1003
2287	0.5004
2972	0.9002
3828	0.9900
4669	0.9990

More examples with combinatorial flavor

We will now do more problems which would rather belong to the previous chapter, but are a little harder, so we do them here instead.

Example 3.10. Roll a die 12 times. Compute the probability that a number occurs 6 times and two other numbers occur three times each.

The number of outcomes is 6^{12}. To count the number of good outcomes:

1. Pick the number that occurs 6 times: $\binom{6}{1} = 6$ choices.
2. Pick the two numbers that occur 3 times each: $\binom{5}{2}$ choices.
3. Pick slots (rolls) for the number that occurs 6 times: $\binom{12}{6}$ choices.
4. Pick slots for one of the numbers that occur 3 times: $\binom{6}{3}$ choices.

Therefore, our probability is $\frac{6\binom{5}{2}\binom{12}{6}\binom{6}{3}}{6^{12}}$.

Example 3.11. You have 10 pairs of socks in the closet. Pick 8 socks at random. For every i, compute the probability that you get i complete pairs of socks.

There are $\binom{20}{8}$ outcomes. To count the number of good outcomes:

1. Pick i pairs of socks from the 10: $\binom{10}{i}$ choices.
2. Pick pairs which are represented by a single sock: $\binom{10-i}{8-2i}$ choices.
3. Pick a sock from each of the latter pairs: 2^{8-2i} choices.

Therefore, our probability is $\frac{2^{8-2i}\binom{10-i}{8-2i}\binom{10}{i}}{\binom{20}{8}}$.

Example 3.12. *Poker Hands*. In the definitions, the word *value* refers to A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, 2. This sequence orders the cards in descending consecutive values, with one exception: an Ace may be regarded as 1 for the purposes of making a straight (but note that, for example, K, A, 1, 2, 3 is *not* a valid straight sequence — A can only begin or end a straight). From the lowest to the highest, here are the hands:
(a) *one pair*: two cards of the same value plus 3 cards with different values

$$J\spadesuit\ J\clubsuit\ 9\heartsuit\ Q\clubsuit\ 4\spadesuit$$

(b) *two pairs*: two pairs plus another card of different value

$$J\spadesuit\ J\clubsuit\ 9\heartsuit\ 9\clubsuit\ 3\spadesuit$$

(c) *three of a kind*: three cards of the same value plus two with different values

$$Q\spadesuit\ Q\clubsuit\ Q\heartsuit\ 9\clubsuit\ 3\spadesuit$$

(d) *straight*: five cards with consecutive values

$$5\heartsuit\ 4\clubsuit\ 3\clubsuit\ 2\heartsuit\ A\spadesuit$$

(e) *flush*: five cards of the same suit

$$K\clubsuit\ 9\clubsuit\ 7\clubsuit\ 6\clubsuit\ 3\clubsuit$$

(f) *full house*: a three of a kind and a pair

$$J\clubsuit\ J\diamondsuit\ J\heartsuit\ 3\clubsuit\ 3\spadesuit$$

(g) *four of a kind*: four cards of the same value

$$K\clubsuit\ K\diamondsuit\ K\heartsuit\ K\clubsuit\ 10\spadesuit$$

(e) *straight flush*: five cards of the same suit with consecutive values

$$A\spadesuit\ K\spadesuit\ Q\spadesuit\ J\spadesuit\ 10\spadesuit$$

Here are the probabilities:

hand	no. combinations	approx. prob.
one pair	$13 \cdot \binom{12}{3} \cdot \binom{4}{2} \cdot 4^3$	0.422569
two pairs	$\binom{13}{2} \cdot 11 \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot 4$	0.047539
three of a kind	$13 \cdot \binom{12}{2} \cdot \binom{4}{3} \cdot 4^2$	0.021128
straight	$10 \cdot 4^5$	0.003940
flush	$4 \cdot \binom{13}{5}$	0.001981
full house	$13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2}$	0.001441
four of a kind	$13 \cdot 12 \cdot 4$	0.000240
straight flush	$10 \cdot 4$	0.000015

Note that the probabilities of a straight and a flush above include the possibility of a straight flush.

Let us see how some of these are computed. First, the number of all outcomes is $\binom{52}{5} = 2,598,960$. Then, for example, for the *three of a kind*, the number of good outcomes may be obtained by listing the number of choices:

1. Choose a value for the triple: 13.
2. Choose the values of other two cards: $\binom{12}{2}$.
3. Pick three cards from the four of the same chosen value: $\binom{4}{3}$.
4. Pick a card (i.e., the suit) from each of the two remaining values: 4^2.

One could do the same for *one pair*:

1. Pick a number for the pair: 13.
2. Pick the other three numbers: $\binom{12}{3}$
3. Pick two cards from the value of the pair: $\binom{4}{2}$.
4. Pick the suits of the other three values: 4^3

And for the *flush*:

1. Pick a suit: 4.
2. Pick five numbers: $\binom{13}{5}$

Our final worked out case is *straight flush*:

1. Pick a suit: 4.
2. Pick the beginning number: 10.

We end this example by computing the probability of not getting any hand listed above, that is,

$$\begin{aligned}
P(\text{nothing}) &= P(\text{all cards with different values}) - P(\text{straight or flush}) \\
&= \frac{\binom{13}{5} \cdot 4^5}{\binom{52}{5}} - (P(\text{straight}) + P(\text{flush}) - P(\text{straight flush})) \\
&= \frac{\binom{13}{5} \cdot 4^5 - 10 \cdot 4^5 - 4 \cdot \binom{13}{5} + 40}{\binom{52}{5}} \\
&\approx 0.5012.
\end{aligned}$$

Example 3.13. Assume that 20 Scandinavians, 10 Finns, and 10 Danes, are to be distributed at random into 10 rooms, 2 per room. What is the probability that exactly $2i$ rooms are mixed, $i = 0, \dots 5$?

This is an example when careful thinking about what the outcomes should be really pays off. Consider the following model for distributing the Scandinavians into rooms. First arrange them at random into a row of 20 slots $S1, S2, \dots, S20$. Assume that room 1 takes people in slots $S1, S2$, so let us call these two slots $R1$. Similarly, room 2 takes people in slots $S3, S4$, so let us call these two slots $R2$, etc.

Now, it is clear that we only need to keep track of the distribution of 10 D's into the 20 slots, corresponding to the positions of the 10 Danes. Any such distribution constitutes an outcome and they are equally likely. Their number is $\binom{20}{10}$.

To get $2i$ (for example, 4) mixed rooms, start by choosing $2i$ (ex., 4) out of the 10 rooms which are going to be mixed; there are $\binom{10}{2i}$ choices. You also need to decide into which slot in each of the $2i$ chosen mixed rooms the D goes, for 2^{2i} choices.

Once these two choices are made, you still have $10 - 2i$ (ex., 6) D's to distribute into $5 - i$ (ex., 3) rooms, as there are two D's in each of these rooms. For this, you need to choose $5 - i$ (ex., 3) rooms from the remaining $10 - 2i$ (ex., 6), for $\binom{10-2i}{5-i}$ choices, and this choice fixes a good outcome.

The final answer, therefore, is

$$\frac{\binom{10}{2i} 2^{2i} \binom{10-2i}{5-i}}{\binom{20}{10}}.$$

Problems

1. Roll a single die 10 times. Compute the following probabilities: (a) that you get at least one 6; (b) that you get at least one 6 *and* at least one 5; (c) that you get three 1's, two 2's, and five 3's.

2. Three married couples take seats around a table at random. Compute P(no wife sits next to her husband).

3. A group of 20 Scandinavians consists of 7 Swedes, 3 Finns, and 10 Norwegians. A committee of five people is chosen at random from this group. What is the probability that at least one of the three nations is *not* represented on the committee?

4. Choose each digit of a 5 digit number at random from digits $1, \dots 9$. Compute the probability that no digit appears more than twice.

5. Roll a fair die 10 times. (a) Compute the probability that at least one number occurs exactly 6 times. (b) Compute the probability that at least one number occurs exactly once.

6. A lottery ticket consists of two rows, each containing 3 numbers from $1, 2, \ldots, 50$. The drawing consists of choosing 5 different numbers from $1, 2, \ldots, 50$ at random. A ticket wins if its first row contains *at least two* of the numbers drawn *and* its second row contains *at least two* of the numbers drawn. The four examples below represent the four types of tickets:

Ticket 1	Ticket 2	Ticket 3	Ticket 4
1 2 3	1 2 3	1 2 3	1 2 3
4 5 6	1 2 3	2 3 4	3 4 5

For example, if the numbers 1, 3, 5, 6, 17 are drawn, then Ticket 1, Ticket 2, and Ticket 4 all win, while Ticket 3 loses. Compute the winning probabilities for each of the four tickets.

Solutions to problems

1. (a) $1 - (5/6)^{10}$. (b) $1 - 2 \cdot (5/6)^{10} + (2/3)^{10}$. (c) $\binom{10}{3}\binom{7}{2}6^{-10}$.

2. The complement is the union of the three events $A_i = \{\text{couple } i \text{ sits together}\}$, $i = 1, 2, 3$. Moreover,

$$\begin{aligned}
&P(A_1) = \frac{2}{5} = P(A_2) = P(A_3), \\
&P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{3! \cdot 2! \cdot 2!}{5!} = \frac{1}{5}, \\
&P(A_1 \cap A_2 \cap A_3) = \frac{2! \cdot 2! \cdot 2! \cdot 2!}{5!} = \frac{2}{15}.
\end{aligned}$$

For $P(A_1 \cap A_2)$, for example, pick a seat for husband$_3$. In the remaining row of 5 seats, pick the ordering for couple$_1$, couple$_2$, and wife$_3$, then the ordering of seats within each of couple$_1$ and couple$_2$. Now, by inclusion-exclusion,

$$P(A_1 \cup A_2 \cup A_3) = 3 \cdot \frac{2}{5} - 3 \cdot \frac{1}{5} + \frac{2}{15} = \frac{11}{15},$$

and our answer is $\frac{4}{15}$.

3. Let A_1 = the event that Swedes are not represented, A_2 = the event that Finns are not represented, and A_3 = the event that Norwegians are not represented.

$$\begin{aligned}
P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\
&\quad + P(A_1 \cap A_2 \cap A_3) \\
&= \frac{1}{\binom{20}{5}} \left[\binom{13}{5} + \binom{17}{5} + \binom{10}{5} - \binom{10}{5} - 0 - \binom{7}{5} + 0 \right]
\end{aligned}$$

4. The number of bad events is $9 \cdot \binom{5}{3} \cdot 8^2 + 9 \cdot \binom{5}{4} \cdot 8 + 9$. The first term is the number of numbers in which a digit appears 3 times, but no digit appears 4 times: choose a digit, choose 3

positions filled by it, then fill the remaining position. The second term is the number of numbers in which a digit appears 4 times, but no digit appears 5 times, and the last term is the number of numbers in which a digit appears 5 times. The answer then is

$$1 - \frac{9 \cdot \binom{5}{3} \cdot 8^2 + 9 \cdot \binom{5}{4} \cdot 8 + 9}{9^5}.$$

5. (a) Let A_i be the event that the number i appears exactly 6 times. As A_i are pairwise disjoint,

$$P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6) = 6 \cdot \frac{\binom{10}{6} \cdot 5^4}{6^{10}}.$$

(b) (a) Now, A_i is the event that the number i appears exactly once. By inclusion-exclusion,

$$\begin{aligned}
&P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6) \\
&= 6P(A_1) \\
&- \binom{6}{2} P(A_1 \cap A_2) \\
&+ \binom{6}{3} P(A_1 \cap A_2 \cap A_3) \\
&- \binom{6}{4} P(A_1 \cap A_2 \cap A_3 \cap A_4) \\
&+ \binom{6}{5} P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) \\
&- \binom{6}{6} P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5 \cap A_6) \\
&= 6 \cdot 10 \cdot \frac{5^9}{6^{10}} \\
&- \binom{6}{2} \cdot 10 \cdot 9 \cdot \frac{4^8}{6^{10}} \\
&+ \binom{6}{3} \cdot 10 \cdot 9 \cdot 8 \cdot \frac{3^7}{6^{10}} \\
&- \binom{6}{4} \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot \frac{2^6}{6^{10}} \\
&+ \binom{6}{5} \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot \frac{1}{6^{10}} \\
&- 0.
\end{aligned}$$

6. Below, a *hit* is shorthand for a chosen number.

$$\begin{aligned}
P(\text{ticket 1 wins}) &= P(\text{two hits on each line}) + P(\text{two hits on one line, three on the other}) \\
&= \frac{3 \cdot 3 \cdot 44 + 2 \cdot 3}{\binom{50}{5}} = \frac{402}{\binom{50}{5}},
\end{aligned}$$

and

$$\begin{aligned}P(\text{ticket 2 wins}) &= P(\text{two hits among 1, 2, 3}) + P(\text{three hits among 1, 2, 3}) \\ &= \frac{3\cdot\binom{47}{3}+\binom{47}{2}}{\binom{50}{5}} = \frac{49726}{\binom{50}{5}},\end{aligned}$$

and

$$\begin{aligned}P(\text{ticket 3 wins}) &= P(\text{2, 3 both hit}) + P(\text{1, 4 both hit and one of 2, 3 hit}) \\ &= \frac{\binom{48}{3}+2\cdot\binom{46}{2}}{\binom{50}{5}} = \frac{19366}{\binom{50}{5}},\end{aligned}$$

and, finally,

$$\begin{aligned}P(\text{ticket 4 wins}) &= P(\text{3 hit, at least one additional hit on each line}) + P(\text{1, 2, 4, 5 all hit}) \\ &= \frac{(4\binom{45}{2}+4\cdot 45+1)+45}{\binom{50}{5}} = \frac{4186}{\binom{50}{5}}\end{aligned}$$

4 Conditional Probability and Independence

Example 4.1. Assume that you have a bag with 11 cubes, 7 of which have a fuzzy surface and 4 are smooth. Out of the 7 fuzzy ones, 3 are red and 4 are blue; out of 4 smooth ones, 2 are red and 2 are blue. So, there are 5 red and 6 blue cubes. Other than color and fuzziness, the cubes have no other distinguishing characteristics.

You plan to pick a cube out of the bag at random, but forget to wear gloves. Before you start your experiment, the probability that the selected cube is red is 5/11. Now, you reach into the bag, grab a cube, and notice it is fuzzy (but you do not take it out or note its color in any other way). Clearly, the probability should now change to 3/7!

Your experiment clearly has 11 outcomes. Consider the events R, B, F, S that the selected cube is red, blue, fuzzy, and smooth, respectively. We observed that $P(R) = 5/11$. For the probability of a red cube, *conditioned on it being fuzzy*, we do not have notation, so we introduce it here: $P(R|F) = 3/7$. Note that this also equals

$$\frac{P(R \cap F)}{P(F)} = \frac{P(\text{the selected ball is red and fuzzy})}{P(\text{the selected ball is fuzzy})}.$$

This conveys the idea that *with additional information the probability must be adjusted*. This is common in real life. Say bookies estimate your basketball team's chances of winning a particular game to be 0.6, 24 hours before the game starts. Two hours before the game starts, however, it becomes known that your team's star player is out with a sprained ankle. You cannot expect that the bookies' odds will remain the same and they change, say, to 0.3. Then, the game starts and at half-time your team leads by 25 points. Again, the odds will change, say to 0.8. Finally, when complete information (that is, the outcome of your experiment, the game in this case) is known, all probabilities are trivial, 0 or 1.

For the general definition, take events A and B, and assume that $P(B) > 0$. The *conditional probability of A given B* equals

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}.}$$

Example 4.2. Here is a question asked on Wall Street job interviews. (This is the original formulation; the macabre tone is not unusual for such interviews.)

"Let's play a game of Russian roulette. You are tied to your chair. Here's a gun, a revolver. Here's the barrel of the gun, six chambers, all empty. Now watch me as I put *two* bullets into the barrel, into *two adjacent chambers*. I close the barrel and spin it. I put a gun to your head and pull the trigger. Click. Lucky you! Now I'm going to pull the trigger one more time. Which would you prefer: that I spin the barrel first or that I just pull the trigger?"

Assume that the barrel rotates clockwise after the hammer hits and is pulled back. You are given the choice between an unconditional and a conditional probability of death. The former, if the barrel is spun again, remains 1/3. The latter, if the trigger is pulled without the extra spin, equals the probability that the hammer clicked on an empty slot, which is next to a bullet in the counterclockwise direction, and equals 1/4.

For a fixed condition B, and acting on events A, the conditional probability $Q(A) = P(A|B)$ satisfies the three axioms in Chapter 3. (This is routine to check and the reader who is more theoretically inclined might view it as a good exercise.) Thus, Q is another probability assignment and all consequences of the axioms are valid for it.

Example 4.3. Toss two fair coins, blindfolded. Somebody tells you that you tossed at least one Heads. What is the probability that both tosses are Heads?

Here $A = \{\text{both H}\}$, $B = \{\text{at least one H}\}$, and

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\text{both H})}{P(\text{at least one H})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

Example 4.4. Toss a coin 10 times. If you know (a) that exactly 7 Heads are tossed, (b) that at least 7 Heads are tossed, what is the probability that your first toss is Heads?

For (a),

$$P(\text{first toss H}|\text{exactly 7 H's}) = \frac{\binom{9}{6} \cdot \frac{1}{2^{10}}}{\binom{10}{7} \cdot \frac{1}{2^{10}}} = \frac{7}{10}.$$

Why is this not surprising? Conditioned on 7 Heads, they are equally likely to occur on any given 7 tosses. If you choose 7 tosses out of 10 at random, the first toss is included in your choice with probability $\frac{7}{10}$.

For (b), the answer is, after canceling $\frac{1}{2^{10}}$,

$$\frac{\binom{9}{6} + \binom{9}{7} + \binom{9}{8} + \binom{9}{9}}{\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}} = \frac{65}{88} \approx 0.7386.$$

Clearly, the answer should be a little larger than before, because this condition is more advantageous for Heads.

Conditional probabilities are sometimes *given*, or can be easily determined, especially in sequential random experiments. Then, we can use

$$\begin{aligned}
&P(A_1 \cap A_2) = P(A_1)P(A_2|A_1),\\
&P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2),\\
&\text{etc.}
\end{aligned}$$

Example 4.5. An urn contains 10 black and 10 white balls. Draw 3 (a) without replacement, and (b) with replacement. What is the probability that all three are white?

We already know how to do part (a):

1. Number of outcomes: $\binom{20}{3}$.

2. Number of ways to select 3 balls out of 10 white ones: $\binom{10}{3}$.

Our probability is then $\frac{\binom{10}{3}}{\binom{20}{3}}$.

To do this problem another way, imagine drawing the balls sequentially. Then, we are computing the probability of the intersection of the three events: P(1st ball is white, 2nd ball is white, and 3rd ball is white). The relevant probabilities are:

1. P(1st ball is white) $= \frac{1}{2}$.

2. P(2nd ball is white|1st ball is white) $= \frac{9}{19}$.

3. P(3rd ball is white|1st two picked are white) $= \frac{8}{18}$.

Our probability is, then, the product $\frac{1}{2} \cdot \frac{9}{19} \cdot \frac{8}{18}$, which equals, as it must, what we obtained before.

This approach is particularly easy in case (b), where the previous colors of the selected balls do not affect the probabilities at subsequent stages. The answer, therefore, is $\left(\frac{1}{2}\right)^3$.

Theorem 4.1. *First Bayes' formula. Assume that $F_1, \dots, F_n$ are pairwise disjoint and that $F_1 \cup \dots \cup F_n = \Omega$, that is, exactly one of them always happens. Then, for an event A,*

$$\boxed{P(A) = P(F_1)P(A|F_1) + P(F_2)P(A|F_2) + \dots + P(F_n)P(A|F_n)\ .}$$

Proof.

$$\begin{aligned} P(F_1)P(A|F_1) + P(F_2)P(A|F_2) + \dots + P(F_n)P(A|F_n) &= P(A \cap F_1) + \dots + P(A \cap F_n) \\ &= P((A \cap F_1) \cup \dots \cup (A \cap F_n)) \\ &= P(A \cap (F_1 \cup \dots \cup F_n)) \\ &= P(A \cap \Omega) = P(A) \end{aligned}$$

□

We call an instance of using this formula "computing the probability by conditioning on which of the events F_i happens." The formula is useful in sequential experiments, when you face different experimental conditions at the second stage depending on what happens at the first stage. Quite often, there are just two events F_i, that is, an event F and its complement F^c, and we are thus conditioning on whether F happens or not.

Example 4.6. Flip a fair coin. If you toss Heads, roll 1 die. If you toss Tails, roll 2 dice. Compute the probability that you roll exactly one 6.

Here you condition on the outcome of the coin toss, which could be Heads (event F) or Tails (event F^c). If $A = \{\text{exactly one } 6\}$, then $P(A|F) = \frac{1}{6}$, $P(A|F^c) = \frac{2\cdot 5}{36}$, $P(F) = P(F^c) = \frac{1}{2}$ and so

$$P(A) = P(F)P(A|F) + P(F^c)P(A|F^c) = \frac{2}{9}.$$

Example 4.7. Roll a die, then select at random, without replacement, as many cards from the deck as the number shown on the die. What is the probability that you get at least one Ace?

Here $F_i = \{\text{number shown on the die is } i\}$, for $i = 1, \dots, 6$. Clearly, $P(F_i) = \frac{1}{6}$. If A is the event that you get at least one Ace,

1. $P(A|F_1) = \frac{1}{13}$,
2. In general, for $i \geq 1$, $P(A|F_i) = 1 - \frac{\binom{48}{i}}{\binom{52}{i}}$.

Therefore, by Bayes' formula,

$$P(A) = \frac{1}{6}\left(\frac{1}{13} + 1 - \frac{\binom{48}{2}}{\binom{52}{2}} + 1 - \frac{\binom{48}{3}}{\binom{52}{3}} + 1 - \frac{\binom{48}{4}}{\binom{52}{4}} + 1 - \frac{\binom{48}{5}}{\binom{52}{5}} + 1 - \frac{\binom{48}{6}}{\binom{52}{6}}\right).$$

Example 4.8. *Coupon collector problem*, revisited. As promised, we will develop a computationally much better formula than the one in Example 3.9. This will be another example of *conditioning*, whereby you (1) *reinterpret* the problem as a sequential experiment and (2) use Bayes' formula with "conditions" F_i being relevant events at the first stage of the experiment.

Here is how it works in this example. Let $p_{k,r}$ be the probability that exactly r (out of a total of n) birthdays are represented among k people; we call the event A. We will fix n and let k and r be variable. Note that $p_{k,n}$ is what we computed by the inclusion-exclusion formula.

At the first stage you have $k-1$ people; then the k'th person arrives on the scene. Let F_1 be the event that there are r birthdays represented among the $k-1$ people and let F_2 be the event that there are $r-1$ birthdays represented among the $k-1$ people. Let F_3 be the event that any other number of birthdays occurs with $k-1$ people. Clearly, $P(A|F_3) = 0$, as the newcomer contributes either 0 or 1 new birthdays. Moreover, $P(A|F_1) = \frac{r}{n}$, the probability that the newcomer duplicates one of the existing r birthdays, and $P(A|F_2) = \frac{n-r+1}{n}$, the probability that the newcomer does not duplicate any of the existing $r-1$ birthdays. Therefore,

$$p_{k,r} = P(A) = P(A|F_1)P(F_1) + P(A|F_2)P(F_2) = \frac{r}{n} \cdot p_{k-1,r} + \frac{n-r+1}{n} \cdot p_{k-1,r-1},$$

for $k, r \geq 1$, and this, together with the boundary conditions

$$\begin{aligned} p_{0,0} &= 1, \\ p_{k,r} &= 0, \text{ for } 0 \leq k < r, \\ p_{k,0} &= 0, \text{ for } k > 0, \end{aligned}$$

makes the computation fast and precise.

Theorem 4.2. *Second Bayes' formula. Let $F_1, \ldots, F_n$ and A be as in Theorem 4.1. Then*

$$\boxed{P(F_j|A) = \frac{P(F_j \cap A)}{P(A)} = \frac{P(A|F_j)P(F_j)}{P(A|F_1)P(F_1) + \ldots + P(A|F_n)P(F_n)}\ .}$$

An event F_j is often called a *hypothesis*, $P(F_j)$ its *prior probability*, and $P(F_j|A)$ its *posterior* probability.

Example 4.9. We have a fair coin and an unfair coin, which always comes out Heads. Choose one at random, toss it twice. It comes out Heads both times. What is the probability that the coin is fair?

The relevant events are $F = \{\text{fair coin}\}$, $U = \{\text{unfair coin}\}$, and $B = \{\text{both tosses H}\}$. Then $P(F) = P(U) = \frac{1}{2}$ (as each coin is chosen with equal probability). Moreover, $P(B|F) = \frac{1}{4}$, and $P(B|U) = 1$. Our probability then is

$$\frac{\frac{1}{2} \cdot \frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot 1} = \frac{1}{5}.$$

Example 4.10. A factory has three machines, M_1, M_2 and M_3, that produce items (say, lightbulbs). It is impossible to tell which machine produced a particular item, but some are defective. Here are the known numbers:

machine	proportion of items made	prob. any made item is defective
M_1	0.2	0.001
M_2	0.3	0.002
M_3	0.5	0.003

You pick an item, test it, and find it is defective. What is the probability that it was made by machine M_2?

The best way to think about this random experiment is as a two-stage procedure. First you choose a machine with the probabilities given by the proportion. Then, that machine produces an item, which you then proceed to test. (Indeed, this is the same as choosing the item from a large number of them and testing it.)

Let D be the event that an item is defective and let M_i also denote the event that the item was made by machine i. Then, $P(D|M_1) = 0.001$, $P(D|M_2) = 0.002$, $P(D|M_3) = 0.003$, $P(M_1) = 0.2$, $P(M_2) = 0.3$, $P(M_3) = 0.5$, and so

$$P(M_2|D) = \frac{0.002 \cdot 0.3}{0.001 \cdot 0.2 + 0.002 \cdot 0.3 + 0.003 \cdot 0.5} \approx 0.26.$$

Example 4.11. Assume 10% of people have a certain disease. A test gives the correct diagnosis with probability of 0.8; that is, if the person is sick, the test will be positive with probability 0.8, but if the person is not sick, the test will be positive with probability 0.2. A *random person* from the population has tested positive for the disease. What is the probability that he is actually sick? (No, it is not 0.8!)

Let us define the three relevant events: $S = \{\text{sick}\}$, $H = \{\text{healthy}\}$, $T = \{\text{tested positive}\}$.

Now, $P(H) = 0.9$, $P(S) = 0.1$, $P(T|H) = 0.2$ and $P(T|S) = 0.8$. We are interested in

$$P(S|T) = \frac{P(T|S)P(S)}{P(T|S)P(S) + P(T|H)P(H)} = \frac{8}{26} \approx 31\%.$$

Note that the prior probability $P(S)$ is very important! Without a very good idea about what it is, a positive test result is difficult to evaluate: a positive test for HIV would mean something very different for a random person as opposed to somebody who gets tested because of risky behavior.

Example 4.12. *O. J. Simpson's first trial*, 1995. The famous sports star and media personality O. J. Simpson was on trial in Los Angeles for murder of his wife and her boyfriend. One of the many issues was whether Simpson's history of spousal abuse could be presented by prosecution at the trial; that is, whether this history was "probative," i.e., it had some evidentiary value, or whether it was merely "prejudicial" and should be excluded. Alan Dershowitz, a famous professor of law at Harvard and a consultant for the defense, was claiming the latter, citing the statistics that $< 0.1\%$ of men who abuse their wives end up killing them. As J. F. Merz and J. C. Caulkins pointed out in the journal *Chance* (Vol. 8, 1995, pg. 14), this was the wrong probability to look at!

We need to start with the fact that a woman is murdered. These numbered $4,936$ in 1992, out of which $1,430$ were killed by partners. In other words, if we let

$$A = \{\text{the (murdered) woman was abused by the partner}\},$$
$$M = \{\text{the woman was murdered by the partner}\},$$

then we estimate the prior probabilities $P(M) = 0.29$, $P(M^c) = 0.71$, and what we are interested in is the posterior probability $P(M|A)$. It was also commonly estimated at the time that about 5% of the women had been physically abused by their husbands. Thus, we can say that $P(A|M^c) = 0.05$, as there is no reason to assume that a woman murdered by somebody else was more or less likely to be abused by her partner. The final number we need is $P(A|M)$. Dershowitz states that "a considerable number" of wife murderers had previously assaulted them, although "some" did not. So, we will (conservatively) say that $P(A|M) = 0.5$. (The two-stage experiment then is: choose a murdered woman at random; at the first stage, she is murdered by her partner, or not, with stated probabilities; at the second stage, she is among the abused women, or not, with probabilities depending on the outcome of the first stage.) By Bayes' formula,

$$P(M|A) = \frac{P(M)P(A|M)}{P(M)P(A|M) + P(M^c)P(A|M^c)} = \frac{29}{36.1} \approx 0.8.$$

The law literature studiously avoids quantifying concepts such as probative value and reasonable doubt. Nevertheless, we can probably say that 80% is considerably too high, compared to the prior probability of 29%, to use as a *sole* argument that the evidence is not probative.

Independence

Events A and B are *independent* if $P(A \cap B) = P(A)P(B)$ and *dependent* (or *correlated*) otherwise.

Assuming that $P(B) > 0$, one could rewrite the condition for independence,

$$P(A|B) = P(A),$$

so the probability of A is unaffected by knowledge that B occurred. Also, if A and B are independent,

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c),$$

so A and B^c are also independent — knowing that B has not occurred also has no influence on the probability of A. Another fact to notice immediately is that disjoint events with nonzero probability cannot be independent: given that one of them happens, the other *cannot* happen and thus its probability drops to zero.

Quite often, independence is an *assumption* and it is *the* most important concept in probability.

Example 4.13. Pick a random card from a full deck. Let $A = \{\text{card is an Ace}\}$ and $R = \{\text{card is red}\}$. Are A and R independent?

We have $P(A) = \frac{1}{13}$, $P(R) = \frac{1}{2}$, and, as there are two red Aces, $P(A \cap R) = \frac{2}{52} = \frac{1}{26}$. The two events are independent — the proportion of aces among red cards is the same as the proportion among all cards.

Now, pick two cards out of the deck sequentially without replacement. Are $F = \{\text{first card is an Ace}\}$ and $S = \{\text{second card is an Ace}\}$ independent?

Now $P(F) = P(S) = \frac{1}{13}$ and $P(S|F) = \frac{3}{51}$, so they are not independent.

Example 4.14. Toss 2 fair coins and let $F = \{\text{Heads on 1st toss}\}$, $S = \{\text{Heads on 2nd toss}\}$. These are independent. You will notice that here the independence is in fact an assumption.

How do we define independence of more than two events? We say that events $A_1, A_2, \ldots, A_n$ are *independent* if

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}),$$

where $1 \le i_1 < i_2 < \ldots < i_k \le n$ are arbitrary indices. The occurrence of any combination of events does not influence the probability of others. Again, it can be shown that, in such a collection of independent events, we can replace an A_i by A_i^c and the events remain independent.

Example 4.15. Roll a four sided fair die, that is, choose one of the numbers 1, 2, 3, 4 at random. Let $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$. Check that these are pairwise independent (each pair is independent), but not independent.

Indeed, $P(A) = P(B) = P(C) = \frac{1}{2}$ and $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$ and pairwise independence follows. However,

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

The simple reason for lack of independence is

$$A \cap B \subset C,$$

so we have complete information on the occurrence of C as soon as we know that A and B both happen.

Example 4.16. You roll a die, your friend tosses a coin.

- If you roll 6, you win outright.
- If you do not roll 6 and your friend tosses Heads, you lose outright.
- If neither, the game is repeated until decided.

What is the probability that you win?

One way to solve this problem certainly is this:

$$\begin{aligned}P(\text{win}) &= P(\text{win on 1st round}) + P(\text{win on 2nd round}) + P(\text{win on 3rd round}) + \dots \\ &= \frac{1}{6} + \left(\frac{5}{6}\cdot\frac{1}{2}\right)\frac{1}{6} + \left(\frac{5}{6}\cdot\frac{1}{2}\right)^2\frac{1}{6} + \dots,\end{aligned}$$

and then we sum the geometric series. *Important note: we have implicitly assumed independence between the coin and the die, as well as between different tosses and rolls. This is very common in problems such as this!*

You can avoid the nuisance, however, by the following trick. Let

$$\begin{aligned}D &= \{\text{game is decided on 1st round}\}, \\ W &= \{\text{you win}\}.\end{aligned}$$

The events D and W are independent, which one can certainly check by computation, but, in fact, there is a very good reason to conclude so immediately. The crucial observation is that, provided that the game is not decided in the 1st round, you are thereafter facing the same game with the same winning probability; thus

$$P(W|D^c) = P(W).$$

In other words, D^c and W are independent and then so are D and W, and therefore

$$P(W) = P(W|D).$$

This means that one can solve this problem by computing the relevant probabilities for the 1st round:

$$P(W|D) = \frac{P(W \cap D)}{P(D)} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{5}{6}\frac{1}{2}} = \frac{2}{7},$$

which is our answer.

Example 4.17. *Craps.* Many casinos allow you to bet even money on the following game. Two dice are rolled and the sum S is observed.

- If $S \in \{7, 11\}$, you win immediately.
- If $S \in \{2, 3, 12\}$, you lose immediately.
- If $S \in \{4, 5, 6, 8, 9, 10\}$, the pair of dice is rolled repeatedly until one of the following happens:
  - S repeats, in which case you win.
  - 7 appears, in which case you lose.

What is the winning probability?

Let us look at all possible ways to win:

1. You win on the first roll with probability $\frac{8}{36}$.
2. Otherwise,
   - you roll a 4 (probability $\frac{3}{36}$), then win with probability $\frac{\frac{3}{36}}{\frac{3}{36}+\frac{6}{36}} = \frac{3}{3+6} = \frac{1}{3}$;
   - you roll a 5 (probability $\frac{4}{36}$), then win with probability $\frac{4}{4+6} = \frac{2}{5}$;
   - you roll a 6 (probability $\frac{5}{36}$), then win with probability $\frac{5}{5+6} = \frac{5}{11}$;
   - you roll a 8 (probability $\frac{5}{36}$), then win with probability $\frac{5}{5+6} = \frac{5}{11}$;
   - you roll a 9 (probability $\frac{4}{36}$), then win with probability $\frac{4}{4+6} = \frac{2}{5}$;
   - you roll a 10 (probability $\frac{3}{36}$), then win with probability $\frac{3}{3+6} = \frac{1}{3}$.

Using Bayes' formula,

$$P(\text{win}) = \frac{8}{36} + 2\left(\frac{3}{36}\cdot\frac{1}{3} + \frac{4}{36}\cdot\frac{2}{5} + \frac{5}{36}\cdot\frac{5}{11}\right) \approx 0.4929,$$

a decent game by casino standards.

Bernoulli trials

Assume n independent experiments, each of which is a success with probability p and, thus, failure with probability $1-p$.

> In a sequence of n Bernoulli trials, $P(\text{exactly } k \text{ successes}) = \binom{n}{k} p^k(1-p)^{n-k}$.

This is because the successes can occur on any subset S of k trials out of n, i.e., on any $S \subset \{1, \ldots, n\}$ with cardinality k. These possibilities are disjoint, as exactly k successes cannot occur on two different such sets. There are $\binom{n}{k}$ such subsets; if we fix such an S, then successes must occur on k trials in S and failures on all $n-k$ trials not in S; the probability that this happens, by the assumed independence, is $p^k(1-p)^{n-k}$.

Example 4.18. A machine produces items which are independently defective with probability p. Let us compute a few probabilities:

1. $P(\text{exactly two items among the first 6 are defective}) = \binom{6}{2}p^2(1-p)^4$.
2. $P(\text{at least one item among the first 6 is defective}) = 1 - P(\text{no defects}) = 1-(1-p)^6$
3. $P(\text{at least 2 items among the first 6 are defective}) = 1-(1-p)^6 - 6p(1-p)^5$
4. $P(\text{exactly 100 items are made before 6 defective are found})$ equals

$$P(\text{100th item defective, exactly 5 items among 1st 99 defective}) = p \cdot \binom{99}{5} p^5(1-p)^{94}.$$

Example 4.19. *Problem of Points*. This involves finding the probability of n successes before m failures in a sequence of Bernoulli trials. Let us call this probability $p_{n,m}$.

$$\begin{aligned} p_{n,m} &= P(\text{in the first } m+n-1 \text{ trials, the number of successes is } \geq n) \\ &= \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k(1-p)^{n+m-1-k}. \end{aligned}$$

The problem is solved, but it needs to be pointed out that computationally this is not the best formula. It is much more efficient to use the recursive formula obtained by conditioning on the outcome of the first trial. Assume $m, n \geq 1$. Then,

$$\begin{aligned} p_{n,m} &= P(\text{first trial is a success}) \cdot P(n-1 \text{ successes before } m \text{ failures}) \\ &\quad + P(\text{first trial is a failure}) \cdot P(n \text{ successes before } m-1 \text{ failures}) \\ &= p \cdot p_{n-1,m} + (1-p) \cdot p_{n,m-1}. \end{aligned}$$

Together with boundary conditions, valid for $m, n \geq 1$,

$$p_{n,0} = 0,\ p_{0,m} = 1,$$

which allows for very speedy and precise computations for large m and n.

Example 4.20. *Best of 7.* Assume that two equally matched teams, A and B, play a series of games and that the first team that wins four games is the overall winner of the series. As it happens, team A lost the first game. What is the probability it will win the series? Assume that the games are Bernoulli trials with success probability $\frac{1}{2}$.

We have

$$\begin{aligned} P(A \text{ wins the series}) &= P(4 \text{ successes before 3 failures}) \\ &= \sum_{k=4}^{6} \binom{6}{k} \left(\frac{1}{2}\right)^6 = \frac{15+6+1}{2^6} \approx 0.3438. \end{aligned}$$

Example 4.21. *Banach Matchbox Problem.* A mathematician carries two matchboxes, each originally containing n matches. Each time he needs a match, he is equally likely to take it from either box. What is the probability that, upon reaching for a box and finding it empty, there are exactly k matches still in the other box? Here, $0 \le k \le n$.

Let A_1 be the event that matchbox 1 is the one discovered empty and that, at that instant, matchbox 2 contains k matches. Before this point, he has used $n + n - k$ matches, n from matchbox 1 and $n - k$ from matchbox 2. This means that he has reached for box 1 exactly n times in $(n + n - k)$ trials and for the last time at the $(n + 1 + n - k)$th trial. Therefore, our probability is

$$2 \cdot P(A_1) = 2 \cdot \frac{1}{2} \binom{2n-k}{n} \frac{1}{2^{2n-k}} = \binom{2n-k}{n} \frac{1}{2^{2n-k}}.$$

Example 4.22. Each day, you independently decide, with probability p, to flip a fair coin. Otherwise, you do nothing. (a) What is the probability of getting exactly 10 Heads in the first 20 days? (b) What is the probability of getting 10 Heads before 5 Tails?

For (a), the probability of getting Heads is $p/2$ independently each day, so the answer is

$$\binom{20}{10} \left(\frac{p}{2}\right)^{10} \left(1 - \frac{p}{2}\right)^{10}.$$

For (b), you can disregard days at which you do not flip to get

$$\sum_{k=10}^{14} \binom{14}{k} \frac{1}{2^{14}}.$$

Example 4.23. You roll a die and your score is the number shown on the die. Your friend rolls five dice and his score is the number of 6's shown. Compute (a) the probability of event A that the two scores are equal and (b) the probability of event B that your friend's score is strictly larger than yours.

In both cases we will condition on your friend's score — this works a little better in case (b) than conditioning on your score. Let F_i, $i = 0, \ldots, 5$, be the event that your friend's score is i. Then, $P(A|F_i) = \frac{1}{6}$ if $i \geq 1$ and $P(A|F_0) = 0$. Then, by the first Bayes' formula, we get

$$P(A) = \sum_{i=1}^{5} P(F_i) \cdot \frac{1}{6} = \frac{1}{6}(1 - P(F_0)) = \frac{1}{6} - \frac{5^5}{6^6} \approx 0.0997.$$

Moreover, $P(B|F_i) = \frac{i-1}{6}$ if $i \geq 2$ and 0 otherwise, and so

$$\begin{aligned}
P(B) &= \sum_{i=1}^{5} P(F_i) \cdot \frac{i-1}{6} \\
&= \frac{1}{6}\sum_{i=1}^{5} i \cdot P(F_i) - \frac{1}{6}\sum_{i=1}^{5} P(F_i) \\
&= \frac{1}{6}\sum_{i=1}^{5} i \cdot P(F_i) - \frac{1}{6} + \frac{5^5}{6^6} \\
&= \frac{1}{6}\sum_{i=1}^{5} i \cdot \binom{5}{i}\left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{5-i} - \frac{1}{6} + \frac{5^5}{6^6} \\
&= \frac{1}{6} \cdot \frac{5}{6} - \frac{1}{6} + \frac{5^5}{6^6} \approx 0.0392.
\end{aligned}$$

The last equality can be obtained by computation, but we will soon learn why the sum has to equal $\frac{5}{6}$.

Problems

1. Consider the following game. Pick one card at random from a full deck of 52 cards. If you pull an Ace, you win outright. If not, then you look at the value of the card (K, Q, and J count as 10). If the number is 7 or less, you lose outright. Otherwise, you select (at random, without replacement) that number of additional cards from the deck. (For example, if you picked a 9 the first time, you select 9 more cards.) If you get at least one Ace, you win. What are your chances of winning this game?

2. 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 it is good 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?

3. A chocolate egg either contains a toy or is empty. Assume that each egg contains a toy with probability p, independently of other eggs. You have 5 eggs; open the first one and see if it has a toy inside, then do the same for the second one, etc. Let E_1 be the event that you get at least 4 toys and let E_2 be the event that you get at least two toys in succession. Compute $P(E_1)$ and $P(E_2)$. Are E_1 and E_2 independent?

4. You have 16 balls, 3 blue, 4 green, and 9 red. You also have 3 urns. For each of the 16 balls, you select an urn at random and put the ball into it. (Urns are large enough to accommodate any number of balls.) (a) What is the probability that no urn is empty? (b) What is the probability that each urn contains 3 red balls? (c) What is the probability that each urn contains all three colors?

5. Assume that you have an n–element set U and that you select r independent random subsets $A_1, \dots, A_r \subset U$. All A_i are chosen so that all 2^n choices are equally likely. Compute (in a simple closed form) the probability that the A_i are pairwise disjoint.

Solutions to problems

1. Let

$$\begin{aligned} F_1 &= \{\text{Ace first time}\}, \\ F_8 &= \{8 \text{ first time}\}, \\ F_9 &= \{9 \text{ first time}\}, \\ F_{10} &= \{10, \text{J, Q, or K first time}\}. \end{aligned}$$

Also, let W be the event that you win. Then

$$\begin{aligned} P(W|F_1) &= 1, \\ P(W|F_8) &= 1 - \frac{\binom{47}{8}}{\binom{51}{8}}, \\ P(W|F_9) &= 1 - \frac{\binom{47}{9}}{\binom{51}{9}}, \\ P(W|F_{10}) &= 1 - \frac{\binom{47}{10}}{\binom{51}{10}}, \end{aligned}$$

and so,

$$P(W) = \frac{4}{52} + \frac{4}{52}\left(1 - \frac{\binom{47}{8}}{\binom{51}{8}}\right) + \frac{4}{52}\left(1 - \frac{\binom{47}{9}}{\binom{51}{9}}\right) + \frac{16}{52}\left(1 - \frac{\binom{47}{10}}{\binom{51}{10}}\right).$$

2. Let $F = \{\text{none is defective}\}$ and $A = \{\text{test indicates that none is defective}\}$. By the second Bayes' formula,

$$\begin{aligned} P(F|A) &= \frac{P(A \cap F)}{P(A)} \\ &= \frac{(0.7 \cdot 0.8)^3}{(0.7 \cdot 0.8 + 0.3 \cdot 0.1)^3} \\ &= \left(\frac{56}{59}\right)^3. \end{aligned}$$

3. $P(E_1) = 5p^4(1-p) + p^5 = 5p^4 - 4p^5$ and $P(E_2) = 1 - (1-p)^5 - 5p(1-p)^4 - \binom{4}{2}p^2(1-p)^3 - p^3(1-p)^2$. As $E_1 \subset E_2$, E_1 and E_2 are not independent.

4. (a) Let A_i = the event that the i-th urn is empty.

$$P(A_1) = P(A_2) = P(A_3) = \left(\frac{2}{3}\right)^{16},$$

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \left(\frac{1}{3}\right)^{16},$$

$$P(A_1 \cap A_2 \cap A_3) = 0.$$

Hence, by inclusion-exclusion,

$$P(A_1 \cup A_2 \cup A_3) = \frac{2^{16} - 1}{3^{15}},$$

and

$$\begin{aligned} P(\text{no urns are empty}) &= 1 - P(A_1 \cup A_2 \cup A_3) \\ &= 1 - \frac{2^{16} - 1}{3^{15}}. \end{aligned}$$

(b) We can ignore other balls since only the red balls matter here.
Hence, the result is:

$$\frac{\frac{9!}{3!3!3!}}{3^9} = \frac{9!}{8 \cdot 3^{12}}.$$

(c) As

$$P(\text{at least one urn lacks blue}) = 3\left(\frac{2}{3}\right)^3 - 3\left(\frac{1}{3}\right)^3,$$

$$P(\text{at least one urn lacks green}) = 3\left(\frac{2}{3}\right)^4 - 3\left(\frac{1}{3}\right)^4,$$

$$P(\text{at least one urn lacks red}) = 3\left(\frac{2}{3}\right)^9 - 3\left(\frac{1}{3}\right)^9,$$

we have, by independence,

$$\begin{aligned} P(\text{each urn contains all 3 colors}) &= \left[1-\left(3\left(\frac{2}{3}\right)^3-3\left(\frac{1}{3}\right)^3\right)\right]\times \\ &\quad \left[1-\left(3\left(\frac{2}{3}\right)^4-3\left(\frac{1}{3}\right)^4\right)\right]\times \\ &\quad \left[1-\left(3\times\left(\frac{2}{3}\right)^9-3\times\left(\frac{1}{3}\right)^9\right)\right]. \end{aligned}$$

5. This is the same as choosing an $r \times n$ matrix in which every entry is independently 0 or 1 with probability 1/2 and ending up with at most one 1 in every column. Since columns are independent, this gives $((1+r)2^{-r})^n$.

Interlude: Practice Midterm 1

This practice exam covers the material from the first four chapters. Give yourself 50 minutes to solve the four problems, which you may assume have equal point score.

1. Ten fair dice are rolled. What is the probability that:

(a) At least one 1 appears.

(b) Each of the numbers 1, 2, 3 appears exactly twice, while the number 4 appears four times.

(c) Each of the numbers 1, 2, 3 appears at least once.

2. Five married couples are seated at random around a round table.

(a) Compute the probability that all couples sit together (i.e., every husband-wife pair occupies adjacent seats).

(b) Compute the probability that at most one wife does not sit next to her husband.

3. Consider the following game. A player rolls a fair die. If he rolls 3 or less, he loses immediately. Otherwise he selects, at random, as many cards from a full deck as the number that came up on the die. The player wins if all four Aces are among the selected cards.

(a) Compute the winning probability for this game.

(b) Smith tells you that he recently played this game once and won. What is the probability that he rolled a 6 on the die?

4. A chocolate egg either contains a toy or is empty. Assume that each egg contains a toy with probability $p \in (0,1)$, independently of other eggs. Each toy is, with equal probability, red, white, or blue (again, independently of other toys). You buy 5 eggs. Let E_1 be the event that you get at most 2 toys and let E_2 be the event that you get you get at least one red and at least one white and at least one blue toy (so that you have a complete collection).

(a) Compute $P(E_1)$. Why is this probability very easy to compute when $p = 1/2$?

(b) Compute $P(E_2)$.

(c) Are E_1 and E_2 independent? Explain.

Solutions to Practice Midterm 1

1. Ten fair dice are rolled. What is the probability that:

   (a) At least one 1 appears.

   Solution:

$$1 - P(\text{no 1 appears}) = 1 - \left(\frac{5}{6}\right)^{10}.$$

   (b) Each of the numbers 1, 2, 3 appears exactly twice, while the number 4 appears four times.

   Solution:

$$\frac{\binom{10}{2}\binom{8}{2}\binom{6}{2}}{6^{10}} = \frac{10!}{2^3 \cdot 4! \cdot 6^{10}}.$$

   (c) Each of the numbers 1, 2, 3 appears at least once.

   Solution:

   Let A_i be the event that the number i does not appear. We know the following:

$$P(A_1) = P(A_2) = P(A_3) = \left(\frac{5}{6}\right)^{10},$$

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \left(\frac{4}{6}\right)^{10},$$

$$P(A_1 \cap A_2 \cap A_3) = \left(\frac{3}{6}\right)^{10}.$$

Then,

$$\begin{aligned}
& P(1,\ 2,\ \text{and } 3 \text{ each appear at least once}) \\
= \;& P((A_1 \cup A_2 \cup A_3)^c) \\
= \;& 1 - P(A_1) - P(A_2) - P(A_3) \\
& + P(A_1 \cap A_2) + P(A_2 \cap A_3) + P(A_1 \cap A_3) \\
& - P(A_1 \cap A_2 \cap A_3) \\
= \;& 1 - 3 \cdot \left(\frac{5}{6}\right)^{10} + 3 \cdot \left(\frac{4}{6}\right)^{10} - \left(\frac{3}{6}\right)^{10}.
\end{aligned}$$

2. Five married couples are seated at random around a round table.

   (a) Compute the probability that all couples sit together (i.e., every husband-wife pair occupies adjacent seats).

   Solution:

   Let i be an integer in the set $\{1, 2, 3, 4, 5\}$. Denote each husband and wife as h_i and w_i, respectively.

   i. Fix h_1 onto one of the seats.
   ii. There are 9! ways to order the remaining 9 people in the remaining 9 seats. This is our sample space.
   iii. There are 2 ways to order w_1.
   iv. Treat each couple as a block and the remaining 8 seats as 4 pairs (where each pair is two adjacent seats). There are 4! ways to seat the remaining 4 couples into 4 pairs of seats.
   v. There are 2^4 ways to order each h_i and w_i within its pair of seats.

   Therefore, our solution is
$$\frac{2 \cdot 4! \cdot 2^4}{9!}.$$

   (b) Compute the probability that at most one wife does not sit next to her husband.

   Solution:

   Let A be the event that all wives sit next to their husbands and let B be the event that exactly one wife does not sit next to her husband. We know that $P(A) = \frac{2^5 \cdot 4!}{9!}$ from part (a). Moreover, $B = B_1 \cup B_2 \cup B_3 \cup B_4 \cup B_5$, where B_i is the event that w_i

does not sit next to h_i and the remaining couples sit together. Then, B_i are disjoint and their probabilities are all the same. So, we need to determine $P(B_1)$.

i. Fix h_1 onto one of the seats.
ii. There are 9! ways to order the remaining 9 people in the remaining 9 seats.
iii. Consider each of remaining 4 couples and w_1 as 5 blocks.
iv. As w_1 cannot be next to her husband, we have 3 positions for w_1 in the ordering of the 5 blocks.
v. There are 4! ways to order the remaining 4 couples.
vi. There are 2^4 ways to order the couples within their blocks.

Therefore,

$$P(B_1) = \frac{3 \cdot 4! \cdot 2^4}{9!}.$$

Our answer, then, is

$$5 \cdot \frac{3 \cdot 4! \cdot 2^4}{9!} + \frac{2^5 \cdot 4!}{9!}.$$

3. Consider the following game. The player rolls a fair die. If he rolls 3 or less, he loses immediately. Otherwise he selects, at random, as many cards from a full deck as the number that came up on the die. The player wins if all four Aces are among the selected cards.

(a) Compute the winning probability for this game.

<u>Solution</u>:

Let W be the event that the player wins. Let F_i be the event that he rolls i, where $i = 1, \ldots, 6$; $P(F_i) = \frac{1}{6}$.
Since we lose if we roll a 1, 2, or 3, $P(W|F_1) = P(W|F_2) = P(W|F_3) = 0$. Moreover,

$$P(W|F_4) = \frac{1}{\binom{52}{4}},$$

$$P(W|F_5) = \frac{\binom{5}{4}}{\binom{52}{4}},$$

$$P(W|F_6) = \frac{\binom{6}{4}}{\binom{52}{4}}.$$

Therefore,

$$P(W) = \frac{1}{6} \cdot \frac{1}{\binom{52}{4}} \left(1 + \binom{5}{4} + \binom{6}{4}\right).$$

(b) Smith tells you that he recently played this game once and won. What is the probability that he rolled a 6 on the die?

Solution:

$$\begin{aligned} P(F_6|W) &= \frac{\frac{1}{6} \cdot \frac{1}{\binom{52}{4}} \cdot \binom{6}{4}}{P(W)} \\ &= \frac{\binom{6}{4}}{1 + \binom{5}{4} + \binom{6}{4}} \\ &= \frac{15}{21} \\ &= \frac{5}{7}. \end{aligned}$$

4. A chocolate egg either contains a toy or is empty. Assume that each egg contains a toy with probability $p \in (0, 1)$, independently of other eggs. Each toy is, with equal probability, red, white, or blue (again, independently of other toys). You buy 5 eggs. Let E_1 be the event that you get at most 2 toys and let E_2 be the event that you get you get at least one red and at least one white and at least one blue toy (so that you have a complete collection).

(a) Compute $P(E_1)$. Why is this probability very easy to compute when $p = 1/2$?

Solution:

$$\begin{aligned} P(E_1) &= P(0 \text{ toys}) + P(1 \text{ toy}) + P(2 \text{ toys}) \\ &= (1-p)^5 + 5p(1-p)^4 + \binom{5}{2} p^2 (1-p)^3. \end{aligned}$$

When $p = \frac{1}{2}$,

$$\begin{aligned} P(\text{at most 2 toys}) &= P(\text{at least 3 toys}) \\ &= P(\text{at most 2 eggs are empty}) \end{aligned}$$

Therefore, $P(E_1) = P(E_1^c)$ and so $P(E_1) = \frac{1}{2}$.

(b) Compute $P(E_2)$.

Solution:

Let A_1 be the event that red is missing, A_2 the event that white is missing, and A_3 the event that blue is missing.

$$\begin{aligned} P(E_2) &= P((A_1 \cup A_2 \cup A_3)^c) \\ &= 1 - 3 \cdot \left(1 - \frac{p}{3}\right)^5 + 3 \cdot \left(1 - \frac{2p}{3}\right)^5 - (1-p)^5. \end{aligned}$$

(c) Are E_1 and E_2 independent? Explain.

Solution:

No: $E_1 \cap E_2 = \emptyset$.

5 Discrete Random Variables

A *random variable* is a number whose value depends upon the outcome of a random experiment. Mathematically, a random variable X is a real-valued function on Ω, the space of outcomes:

$$X : \Omega \to \mathbb{R}.$$

Sometimes, when convenient, we also allow X to have the value ∞ or, more rarely, $-\infty$, but this will not occur in this chapter. The crucial theoretical property that X should have is that, for each interval B, the set of outcomes for which $X \in B$ is an event, so we are able to talk about its probability, $P(X \in B)$. Random variables are traditionally denoted by capital letters to distinguish them from deterministic quantities.

Example 5.1. Here are some examples of random variables.

1. Toss a coin 10 times and let X be the number of Heads.
2. Choose a random point in the unit square $\{(x, y) : 0 \le x, y \le 1\}$ and let X be its distance from the origin.
3. Choose a random person in a class and let X be the height of the person, in inches.
4. Let X be value of the NASDAQ stock index at the closing of the next business day.

A *discrete random variable* X has finitely or countably many values x_i, $i = 1, 2, \ldots$, and $p(x_i) = P(X = x_i)$ with $i = 1, 2, \ldots$ is called the *probability mass function* of X. Sometimes X is added as the subscript of its p. m. f., $p = p_X$.

A probability mass function p has the following properties:

1. For all i, $p(x_i) > 0$ (we do not list values of X which occur with probability 0).
2. For any interval B, $P(X \in B) = \sum_{x_i \in B} p(x_i)$.
3. As X must have some value, $\sum_i p(x_i) = 1$.

Example 5.2. Let X be the number of Heads in 2 fair coin tosses. Determine its p. m. f.

Possible values of X are 0, 1, and 2. Their probabilities are: $P(X = 0) = \frac{1}{4}$, $P(X = 1) = \frac{1}{2}$, and $P(X = 2) = \frac{1}{4}$.

You should note that the random variable Y, which counts the number of Tails in the 2 tosses, has the same p. m. f., that is, $p_X = p_Y$, but X and Y are far from being the same random variable! In general, random variables may have the same p. m. f., but may not even be defined on the same set of outcomes.

Example 5.3. An urn contains 20 balls numbered $1, \ldots, 20$. Select 5 balls at random, without replacement. Let X be the largest number among selected balls. Determine its p. m. f. and the probability that at least one of the selected numbers is 15 or more.

The possible values are $5, \dots, 20$. To determine the p. m. f., note that we have $\binom{20}{5}$ outcomes, and, then,

$$P(X = i) = \frac{\binom{i-1}{4}}{\binom{20}{5}}.$$

Finally,

$$P(\text{at least one number 15 or more}) = P(X \geq 15) = \sum_{i=15}^{20} P(X = i).$$

Example 5.4. An urn contains 11 balls, 3 white, 3 red, and 5 blue balls. Take out 3 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 p. m. f. of X, the amount you win.

The number of outcomes is $\binom{11}{3}$. X can have values -3, -2, -1, 0, 1, 2, and 3. Let us start with 0. This can occur with one ball of each color or with 3 blue balls:

$$P(X = 0) = \frac{3 \cdot 3 \cdot 5 + \binom{5}{3}}{\binom{11}{3}} = \frac{55}{165}.$$

To get $X = 1$, we can have 2 red and 1 white, or 1 red and 2 blue:

$$P(X = 1) = P(X = -1) = \frac{\binom{3}{2}\binom{3}{1} + \binom{3}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{39}{165}.$$

The probability that $X = -1$ is the same because of symmetry between the roles that the red and the white balls play. Next, to get $X = 2$ we must have 2 red balls and 1 blue:

$$P(X = -2) = P(X = 2) = \frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}} = \frac{15}{165}.$$

Finally, a single outcome (3 red balls) produces $X = 3$:

$$P(X = -3) = P(X = 3) = \frac{1}{\binom{11}{3}} = \frac{1}{165}.$$

All the seven probabilities should add to 1, which can be used either to check the computations or to compute the seventh probability (say, $P(X = 0)$) from the other six.

Assume that X is a discrete random variable with possible values x_i, $i = 1, 2 \ldots$. Then, the *expected value*, also called *expectation*, *average*, or *mean*, of X is

$$EX = \sum_i x_i P(X = x_i) = \sum_i x_i p(x_i).$$

For any function, $g : \mathbb{R} \to \mathbb{R}$,

$$Eg(X) = \sum_i g(x_i) P(X = x_i).$$

Example 5.5. Let X be a random variable with $P(X = 1) = 0.2$, $P(X = 2) = 0.3$, and $P(X = 3) = 0.5$. What is the expected value of X?

We can, of course, just use the formula, but let us instead proceed intuitively and see that the definition makes sense. What, then, should the average of X be?

Imagine a large number n of repetitions of the experiment and measure the realization of X in each. By the frequency interpretation of probability, about $0.2n$ realizations will have $X = 1$, about $0.3n$ will have $X = 2$, and about $0.5n$ will have $X = 3$. The average value of X then should be

$$\frac{1 \cdot 0.2n + 2 \cdot 0.3n + 3 \cdot 0.5n}{n} = 1 \cdot 0.2 + 2 \cdot 0.3 + 3 \cdot 0.5 = 2.3,$$

which of course is the same as the formula gives.

Take a discrete random variable X and let $\mu = EX$. How should we measure the deviation of X from μ, i.e., how "spread-out" is the p. m. f. of X?

The most natural way would certainly be $E|X - \mu|$. The only problem with this is that absolute values are annoying. Instead, we define the *variance* of X

$$\mathrm{Var}(X) = E(x - \mu)^2.$$

The quantity that has the correct units is the *standard deviation*

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

We will give another, more convenient, formula for variance that will use the following property of expectation, called *linearity*:

$$E(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 EX_1 + \alpha_2 EX_2,$$

valid for any random variables X_1 and X_2 and nonrandom constants α_1 and α_2. This property will be explained and discussed in more detail later. Then

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X - \mu)^2] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - \mu^2 = E(X^2) - (EX)^2
\end{aligned}$$

In computations, bear in mind that variance cannot be negative! Furthermore, the only way that a random variable has 0 variance is when it is equal to its expectation μ with probability 1 (so it is not really random at all): $P(X = \mu) = 1$. Here is the summary:

$$\boxed{\text{The variance of a random variable } X \text{ is } \mathrm{Var}(X) = E(X - EX)^2 = E(X^2) - (EX)^2.}$$

Example 5.6. Previous example, continued. Compute $\text{Var}(X)$.

$$E(X^2) = 1^2 \cdot 0.2 + 2^2 \cdot 0.3 + 3^2 \cdot 0.5 = 5.9,$$

$(EX)^2 = (2.3)^2 = 5.29$, and so $\text{Var}(X) = 5.9 - 5.29 = 0.61$ and $\sigma(X) = \sqrt{\text{Var}(X)} \approx 0.7810$.

We will now look at some famous probability mass functions.

5.1 Uniform discrete random variable

This is a random variable with values $x_1, \ldots, x_n$, each with equal probability $1/n$. Such a random variable is simply the random choice of one among n numbers.

Properties:

1. $EX = \frac{x_1+\ldots+x_n}{n}$.
2. $\text{Var}X = \frac{x_1^2+\ldots+x_n^2}{n} - \left(\frac{x_1+\ldots+x_n}{n}\right)^2$.

Example 5.7. Let X be the number shown on a rolled fair die. Compute EX, $E(X^2)$, and $\text{Var}(X)$.

This is a standard example of a discrete uniform random variable and

$$\begin{aligned} EX &= \frac{1+2+\ldots+6}{6} = \frac{7}{2}, \\ EX^2 &= \frac{1+2^2+\ldots+6^2}{6} = \frac{91}{6}, \\ \text{Var}(X) &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}. \end{aligned}$$

5.2 Bernoulli random variable

This is also called an *indicator* random variable. Assume that A is an event with probability p. Then, I_A, the *indicator* of A, is given by

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

Other notations for I_A include 1_A and χ_A. Although simple, such random variables are very important as building blocks for more complicated random variables.

Properties:

1. $EI_A = p$.
2. $\text{Var}(I_A) = p(1-p)$.

For the variance, note that $I_A^2 = I_A$, so that $E(I_A^2) = EI_A = p$.

5.3 Binomial random variable

A Binomial(n,p) random variable counts the number of successes in n independent trials, each of which is a success with probability p.

Properties:

1. Probability mass function: $P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$, $i = 0, \ldots, n$.
2. $EX = np$.
3. $\mathrm{Var}(X) = np(1-p)$.

The expectation and variance formulas will be proved in Chapter 8. For now, take them on faith.

Example 5.8. Let X be the number of Heads in 50 tosses of a fair coin. Determine EX, $\mathrm{Var}(X)$ and $P(X \le 10)$? As X is Binomial($50, \frac{1}{2}$), so $EX = 25$, $\mathrm{Var}(X) = 12.5$, and

$$P(X \le 10) = \sum_{i=0}^{10} \binom{50}{i} \frac{1}{2^{50}}.$$

Example 5.9. Denote by d the dominant gene and by r the recessive gene at a single locus. Then dd is called the pure dominant genotype, dr is called the hybrid, and rr the pure recessive genotype. The two genotypes with at least one dominant gene, dd and dr, result in the phenotype of the dominant gene, while rr results in a recessive phenotype.

Assuming that both parents are hybrid and have n children, what is the probability that at least two will have the recessive phenotype? Each child, independently, gets one of the genes at random from each parent.

For each child, independently, the probability of the rr genotype is $\frac{1}{4}$. If X is the number of rr children, then X is Binomial(n, $\frac{1}{4}$). Therefore,

$$P(X \ge 2) = 1 - P(X = 0) - P(X = 1) = 1 - \left(\frac{3}{4}\right)^n - n \cdot \frac{1}{4}\left(\frac{3}{4}\right)^{n-1}.$$

5.4 Poisson random variable

A random variable is Poisson(λ), with parameter $\lambda > 0$, if it has the probability mass function given below.

Properties:

1. $P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}$, for $i = 0, 1, 2, \ldots$.

2. $EX = \lambda$.
3. $\mathrm{Var}(X) = \lambda$.

Here is how we compute the expectation:

$$EX = \sum_{i=1}^{\infty} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = e^{-\lambda} \lambda \, e^{\lambda} = \lambda,$$

and the variance computation is similar (and a good exercise!).

The Poisson random variable is useful as an approximation to a Binomial random variable when the number of trials is large and the probability of success is small. In this context it is often called the *law of rare events*, first formulated by L. J. Bortkiewicz (in 1898), who studied deaths by horse kicks in the Prussian cavalry.

Theorem 5.1. *Poisson approximation to Binomial. When n is large, p is small, and $\lambda = np$ is of moderate size, Binomial(n,p) is approximately Poisson(λ):*

If X is Binomial(n, p), with $p = \frac{\lambda}{n}$, then, as $n \to \infty$, $$P(X = i) \to e^{-\lambda} \frac{\lambda^i}{i!},$$ *for each fixed $i = 0, 1, 2, \dots$.*

Proof.

$$\begin{aligned} P(X = i) &= \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\dots(n-i+1)}{i!} \cdot \frac{\lambda^i}{n^i} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-i} \\ &= \frac{\lambda^i}{i!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \frac{n(n-1)\dots(n-i+1)}{n^i} \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^i} \\ &\to \frac{\lambda^i}{i!} \cdot e^{-\lambda} \cdot 1 \cdot 1, \end{aligned}$$

as $n \to \infty$. □

The Poisson approximation is quite good: one can prove that the error made by computing a probability using the Poisson approximation instead of its exact Binomial expression (in the context of the above theorem) is no more than

$$\min(1, \lambda) \cdot p.$$

Example 5.10. Suppose that the probability that a person is killed by lighting in a year is, independently, 1/(500 million). Assume that the US population is 300 million.

1. Compute P(3 or more people will be killed by lightning next year) exactly.

   If X is the number of people killed by lightning, then X is Binomial(n, p), where $n = 300$ million and $p = 1/$ (500 million), and the answer is

   $$1 - (1-p)^n - np(1-p)^{n-1} - \binom{n}{2} p^2 (1-p)^{n-2} \approx 0.02311530.$$

2. Approximate the above probability.

   As $np = \frac{3}{5}$, X is approximately Poisson$(\frac{3}{5})$, and the answer is

   $$1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2} e^{-\lambda} \approx 0.02311529.$$

3. Approximate P(two or more people are killed by lightning within the first 6 months of next year).

   This highlights the interpretation of λ as a *rate*. If lightning deaths occur at the rate of $\frac{3}{5}$ a year, they should occur at half that rate in 6 months. Indeed, assuming that lightning deaths occur as a result of independent factors in disjoint time intervals, we can imagine that they operate on different people in disjoint time intervals. Thus, *doubling* the time interval is the same as doubling the number n of people (while keeping p the same), and then np also doubles. Consequently, halving the time interval has the same p, but half as many trials, so np changes to $\frac{3}{10}$ and so $\lambda = \frac{3}{10}$ as well. The answer is

   $$1 - e^{-\lambda} - \lambda e^{-\lambda} \approx 0.0369.$$

4. Approximate P(in exactly 3 of next 10 years exactly 3 people are killed by lightning).

   In every year, the probability of exactly 3 deaths is approximately $\frac{\lambda^3}{3!} e^{-\lambda}$, where, again, $\lambda = \frac{3}{5}$. Assuming year-to-year independence, the number of years with exactly 3 people killed is approximately Binomial$(10, \frac{\lambda^3}{3!} e^{-\lambda})$. The answer, then, is

   $$\binom{10}{3} \left(\frac{\lambda^3}{3!} e^{-\lambda}\right)^3 \left(1 - \frac{\lambda^3}{3!} e^{-\lambda}\right)^7 \approx 4.34 \cdot 10^{-6}.$$

5. Compute the expected number of years, among the next 10, in which 2 or more people are killed by lightning.

   By the same logic as above and the formula for Binomal expectation, the answer is

   $$10(1 - e^{-\lambda} - \lambda e^{-\lambda}) \approx 0.3694.$$

Example 5.11. *Poisson distribution and law.* Assume a crime has been committed. It is known that the perpetrator has certain characteristics, which occur with a small frequency p (say, 10^{-8}) in a population of size n (say, 10^8). A person who matches these characteristics has been found at random (e.g., at a routine traffic stop or by airport security) and, since p is so small, charged with the crime. *There is no other evidence*. What should the defense be?

Let us start with a mathematical model of this situation. Assume that N is the number of people with given characteristics. This is a Binomial random variable but, given the assumptions, we can easily assume that it is Poisson with $\lambda = np$. Choose a person from among these N, label that person by C, the criminal. Then, choose at random another person, A, who is arrested. The question is whether $C = A$, that is, whether the arrested person is guilty. Mathematically, we can formulate the problem as follows:

$$P(C = A \,|\, N \geq 1) = \frac{P(C = A, N \geq 1)}{P(N \geq 1)}.$$

We need to condition as the experiment cannot even be performed when $N = 0$. Now, by the first Bayes' formula,

$$\begin{aligned} P(C = A, N \geq 1) &= \sum_{k=1}^{\infty} P(C = A, N \geq 1 \,|\, N = k) \cdot P(N = k) \\ &= \sum_{k=1}^{\infty} P(C = A \,|\, N = k) \cdot P(N = k) \end{aligned}$$

and

$$P(C = A \,|\, N = k) = \frac{1}{k},$$

so

$$P(C = A, N \geq 1) = \sum_{k=1}^{\infty} \frac{1}{k} \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

The probability that the arrested person is guilty then is

$$P(C = A | N \geq 1) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \cdot \sum_{k=1}^{\infty} \frac{\lambda^k}{k \cdot k!}.$$

There is no closed-form expression for the sum, but it can be easily computed numerically. The defense may claim that the probability of innocence, $1-$(the above probability), is about 0.2330 when $\lambda = 1$, presumably enough for a reasonable doubt.

This model was in fact tested in court, in the famous *People v. Collins* case, a 1968 jury trial in Los Angeles. In this instance, it was claimed by the prosecution (on flimsy grounds) that $p = 1/12,000,000$ and n would have been the number of adult couples in the LA area, say $n = 5,000,000$. The jury convicted the couple charged for robbery on the basis of the prosecutor's claim that, due to low p, "the chances of there being another couple [with the specified characteristics, in the LA area] must be one in a billion." The Supreme Court of California reversed the conviction and gave two reasons. The first reason was insufficient foundation for the estimate of p. The second reason was that the probability that another couple with matching characteristics existed was, in fact,

$$P(N \geq 2 \,|\, N \geq 1) = \frac{1 - e^{-\lambda} - \lambda e^{-\lambda}}{1 - e^{-\lambda}},$$

much larger than the prosecutor claimed, namely, for $\lambda = \frac{5}{12}$ it is about 0.1939. This is about twice the (more relevant) probability of innocence, which, for this λ, is about 0.1015.

5.5 Geometric random variable

A Geometric(p) random variable X counts the number trials required for the first success in independent trials with success probability p.

Properties:

1. Probability mass function: $P(X = n) = p(1-p)^{n-1}$, where $n = 1, 2, \ldots$.
2. $EX = \frac{1}{p}$.
3. $\text{Var}(X) = \frac{1-p}{p^2}$.
4. $P(X > n) = \sum_{k=n+1}^{\infty} p(1-p)^{k-1} = (1-p)^n$.
5. $P(X > n + k | X > k) = \frac{(1-p)^{n+k}}{(1-p)^k} = P(X > n)$.

We omit the proofs of the second and third formulas, which reduce to manipulations with geometric series.

Example 5.12. Let X be the number of tosses of a fair coin required for the first Heads. What are EX and $\text{Var}(X)$?

As X is Geometric($\frac{1}{2}$), $EX = 2$ and $\text{Var}(X) = 2$.

Example 5.13. You roll a die, your opponent tosses a coin. If you roll 6 you win; if you do not roll 6 and your opponent tosses Heads you lose; otherwise, this round ends and the game repeats. On the average, how many rounds does the game last?

$$P(\text{game decided on round 1}) = \frac{1}{6} + \frac{5}{6} \cdot \frac{1}{2} = \frac{7}{12},$$

and so the number of rounds N is Geometric($\frac{7}{12}$), and

$$EN = \frac{12}{7}.$$

Problems

1. Roll a fair die repeatedly. Let X be the number of 6's in the first 10 rolls and let Y the number of rolls needed to obtain a 3. (a) Write down the probability mass function of X. (b) Write down the probability mass function of Y. (c) Find an expression for $P(X \geq 6)$. (d) Find an expression for $P(Y > 10)$.

2. A biologist needs at least 3 mature specimens of a certain plant. The plant needs a year to reach maturity; once a seed is planted, any plant will survive for the year with probability 1/1000 (independently of other plants). The biologist plants 3000 seeds. A year is deemed a success if three or more plants from these seeds reach maturity.

(a) Write down the *exact* expression for the probability that the biologist will indeed end up with at least 3 mature plants.

(b) Write down a relevant *approximate* expression for the probability from (a). Justify briefly the approximation.

(c) The biologist plans to do this year after year. What is the approximate probability that he has at least 2 successes in 10 years?

(d) Devise a method to determine the number of seeds the biologist should plant in order to get at least 3 mature plants in a year with probability at least 0.999. (Your method will probably require a lengthy calculation – do not try to carry it out with pen and paper.)

3. You are dealt one card at random from a full deck and your opponent is dealt 2 cards (without any replacement). If you get an Ace, he pays you \$10, if you get a King, he pays you \$5 (regardless of his cards). If you have neither an Ace nor a King, but your card is red and your opponent has no red cards, he pays you \$1. In all other cases you pay him \$1. Determine your expected earnings. Are they positive?

4. You and your opponent both roll a fair die. If you both roll the same number, the game is repeated, otherwise whoever rolls the larger number wins. Let N be the number of times the two dice have to be rolled before the game is decided. (a) Determine the probability mass function of N. (b) Compute EN. (c) Compute $P(\text{you win})$. (d) Assume that you get paid \$10 for winning in the first round, \$1 for winning in any other round, and nothing otherwise. Compute your expected winnings.

5. Each of the 50 students in class belongs to exactly one of the four groups A, B, C, or D. The membership numbers for the four groups are as follows: A: 5, B: 10, C: 15, D: 20. First, choose one of the 50 students at random and let X be the size of that student's group. Next, choose one of the four groups at random and let Y be its size. (Recall: all random choices are with equal probability, unless otherwise specified.) (a) Write down the probability mass functions for X and Y. (b) Compute EX and EY. (c) Compute $\text{Var}(X)$ and $\text{Var}(Y)$. (d) Assume you have

s students divided into n groups with membership numbers $s_1, \ldots, s_n$, and again X is the size of the group of a randomly chosen student, while Y is the size of the randomly chosen group. Let $EY = \mu$ and $\text{Var}(Y) = \sigma^2$. Express EX with s, n, μ, and σ.

Solutions

1. (a) X is Binomial$(10, \frac{1}{6})$:

$$P(X = i) = \binom{10}{i} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{10-i},$$

where $i = 0, 1, 2, \ldots, 10$.

(b) Y is Geometric$(\frac{1}{6})$:

$$P(Y = i) = \frac{1}{6} \left(\frac{5}{6}\right)^{i-1},$$

where $i = 1, 2, \ldots$.

(c)

$$P(X \geq 6) = \sum_{i=6}^{10} \binom{10}{i} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{10-i}.$$

(d)

$$P(Y > 10) = \left(\frac{5}{6}\right)^{10}.$$

2. (a) The random variable X, the number of mature plants, is Binomial$\left(3000, \frac{1}{1000}\right)$.

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - (0.999)^{3000} - 3000(0.999)^{2999}(0.001) - \binom{3000}{2}(0.999)^{2998}(0.001)^2. \end{aligned}$$

(b) By the Poisson approximation with $\lambda = 3000 \cdot \frac{1}{1000} = 3$,

$$P(X \geq 3) \approx 1 - e^{-3} - 3e^{-3} - \frac{9}{2}e^{-3}.$$

(c) Denote the probability in (b) by s. Then, the number of years the biologists succeeds is approximately Binomial$(10, s)$ and the answer is

$$1 - (1 - s)^{10} - 10s(1 - s)^9.$$

(d) Solve

$$e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2}{2}e^{-\lambda} = 0.001$$

for λ and then let $n = 1000\lambda$. The equation above can be solved by rewriting

$$\lambda = \log 1000 + \log(1 + \lambda + \frac{\lambda^2}{2})$$

and then solved by iteration. The result is that the biologist should plant 11,229 seeds.

3. Let X be your earnings.

$$P(X = 10) = \frac{4}{52},$$

$$P(X = 5) = \frac{4}{52},$$

$$P(X = 1) = \frac{22}{52} \cdot \frac{\binom{26}{2}}{\binom{51}{2}} = \frac{11}{102},$$

$$P(X = -1) = 1 - \frac{2}{13} - \frac{11}{102},$$

and so

$$EX = \frac{10}{13} + \frac{5}{13} + \frac{11}{102} - 1 + \frac{2}{13} + \frac{11}{102} = \frac{4}{13} + \frac{11}{51} > 0$$

4. (a) N is Geometric($\frac{5}{6}$):

$$P(N = n) = \left(\frac{1}{6}\right)^{n-1} \cdot \frac{5}{6},$$

where $n = 1, 2, 3, \ldots$.

(b) $EN = \frac{6}{5}$.

(c) By symmetry, $P(\text{you win}) = \frac{1}{2}$.

(d) You get paid \$10 with probability $\frac{5}{12}$, \$1 with probability $\frac{1}{12}$, and 0 otherwise, so your expected winnings are $\frac{51}{12}$.

5. (a)

x	$P(X = x)$	$P(Y = x)$
5	0.1	0.25
10	0.2	0.25
15	0.3	0.25
20	0.4	0.25

(b) $EX = 15$, $EY = 12.5$.

(c) $E(X^2) = 250$, so $\text{Var}(X) = 25$. $E(Y^2) = 187.5$, so $\text{Var}(Y) = 31.25$.

(d) Let $s = s_1 + \ldots + s_n$. Then,

$$E(X) = \sum_{i=1}^{n} s_i \cdot \frac{s_i}{s} = \frac{n}{s} \sum_{i=1}^{n} s_i^2 \cdot \frac{1}{n} = \frac{n}{s} \cdot EY^2 = \frac{n}{s}(\mathrm{Var}(Y) + (EY)^2) = \frac{n}{s}(\sigma^2 + \mu^2).$$

6 Continuous Random Variables

A random variable X is *continuous* if there exists a nonnegative function f so that, for every interval B,

$$P(X \in B) = \int_B f(x)\,dx,$$

The function $f = f_X$ is called the *density* of X.

We will assume that a density function f is continuous, apart from finitely many (possibly infinite) jumps. Clearly, it must hold that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Note also that

$$\begin{aligned}
&P(X \in [a,b]) = P(a \le X \le b) = \int_a^b f(x)\,dx, \\
&P(X = a) = 0, \\
&P(X \le b) = P(X < b) = \int_{-\infty}^b f(x)\,dx.
\end{aligned}$$

The function $F = F_X$ given by

$$F(x) = P(X \le x) = \int_{-\infty}^x f(s)\,ds$$

is called the *distribution function* of X. On an open interval where f is continuous,

$$F'(x) = f(x).$$

Density has the same role as the probability mass function for discrete random variables: it tells which values x are relatively more probable for X than others. Namely, if h is very small, then

$$P(X \in [x, x+h]) = F(x+h) - F(x) \approx F'(x) \cdot h = f(x) \cdot h.$$

By analogy with discrete random variables, we define,

$$\begin{aligned}
EX &= \int_{-\infty}^{\infty} x \cdot f(x)\,dx, \\
Eg(X) &= \int_{-\infty}^{\infty} g(x) \cdot f(x)\,dx,
\end{aligned}$$

and variance is computed by the same formula: $\mathrm{Var}(X) = E(X^2) - (EX)^2$.

Example 6.1. Let

$$f(x) = \begin{cases} cx & \text{if } 0 < x < 4, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Determine c. (b) Compute $P(1 \le X \le 2)$. (c) Determine EX and $\text{Var}(X)$.

For (a), we use the fact that density integrates to 1, so we have $\int_0^4 cx\,dx = 1$ and $c = \frac{1}{8}$. For (b), we compute

$$\int_1^2 \frac{x}{8}\,dx = \frac{3}{16}.$$

Finally, for (c) we get

$$EX = \int_0^4 \frac{x^2}{8}\,dx = \frac{8}{3}$$

and

$$E(X^2) = \int_0^4 \frac{x^3}{8}\,dx = 8.$$

So, $\text{Var}(X) = 8 - \frac{64}{9} = \frac{8}{9}$.

Example 6.2. Assume that X has density

$$f_X(x) = \begin{cases} 3x^2 & \text{if } x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

Compute the density f_Y of $Y = 1 - X^4$.

In a problem such as this, compute first the distribution function F_Y of Y. Before starting, note that the density $f_Y(y)$ will be nonzero only when $y \in [0,1]$, as the values of Y are restricted to that interval. Now, for $y \in (0,1)$,

$$F_Y(y) = P(Y \le y) = P(1 - X^4 \le y) = P(1 - y \le X^4) = P((1-y)^{\frac{1}{4}} \le X) = \int_{(1-y)^{\frac{1}{4}}}^1 3x^2\,dx\,.$$

It follows that

$$f_Y(y) = \frac{d}{dy}F_Y(y) = -3((1-y)^{\frac{1}{4}})^2\frac{1}{4}(1-y)^{-\frac{3}{4}}(-1) = \frac{3}{4}\frac{1}{(1-y)^{\frac{1}{4}}},$$

for $y \in (0,1)$, and $f_Y(y) = 0$ otherwise. Observe that it is immaterial how $f(y)$ is defined at $y = 0$ and $y = 1$, because those two values contribute nothing to any integral.

As with discrete random variables, we now look at some famous densities.

6.1 Uniform random variable

Such a random variable represents the choice of a random number in $[\alpha, \beta]$. For $[\alpha, \beta] = [0,1]$, this is ideally the output of a computer random number generator.

Properties:

1. Density: $f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } x \in [\alpha, \beta], \\ 0 & \text{otherwise.} \end{cases}$

2. $EX = \frac{\alpha+\beta}{2}$.

3. $\text{Var}(X) = \frac{(\beta-\alpha)^2}{12}$.

Example 6.3. Assume that X is uniform on $[0, 1]$. What is $P(X \in \mathbb{Q})$? What is the probability that the binary expansion of X starts with 0.010?

As $\mathbb{Q}$ is countable, it has an enumeration, say, $\mathbb{Q} = \{q_1, q_2, \dots\}$. By Axiom 3 of Chapter 3:

$$P(X \in \mathbb{Q}) = P(\cup_i \{X = q_i\}) = \sum_i P(X = q_i) = 0.$$

Note that you cannot do this for sets that are not countable or you would "prove" that $P(X \in \mathbb{R}) = 0$, while we, of course, know that $P(X \in \mathbb{R}) = P(\Omega) = 1$. As X is, with probability 1, irrational, its binary expansion is uniquely defined, so there is no ambiguity about what the second question means.

Divide $[0, 1)$ into 2^n intervals of equal length. If the binary expansion of a number $x \in [0, 1)$ is $0.x_1x_2\dots$, the first n binary digits determine which of the 2^n subintervals x belongs to: if you know that x belongs to an interval I based on the first $n-1$ digits, then nth digit 1 means that x is in the right half of I and nth digit 0 means that x is in the left half of I. For example, if the expansion starts with 0.010, the number is in $[0, \frac{1}{2}]$, then in $[\frac{1}{4}, \frac{1}{2}]$, and then finally in $[\frac{1}{4}, \frac{3}{8}]$.

Our answer is $\frac{1}{8}$, but, in fact, we can make a more general conclusion. If X is uniform on $[0, 1]$, then any of the 2^n possibilities for its first n binary digits are equally likely. In other words, the binary digits of X are the result of an infinite sequence of independent fair coin tosses. Choosing a uniform random number on $[0, 1]$ is thus equivalent to tossing a fair coin infinitely many times.

Example 6.4. A uniform random number X divides $[0, 1]$ into two segments. Let R be the ratio of the smaller versus the larger segment. Compute the density of R.

As R has values in $(0, 1)$, the density $f_R(r)$ is nonzero only for $r \in (0, 1)$ and we will deal only with such r's.

$$\begin{aligned}
F_R(r) &= P(R \le r) = P\left(X \le \frac{1}{2}, \frac{X}{1-X} \le r\right) + P\left(X > \frac{1}{2}, \frac{1-X}{X} \le r\right) \\
&= P\left(X \le \frac{1}{2}, X \le \frac{r}{r+1}\right) + P\left(X > \frac{1}{2}, X \ge \frac{1}{r+1}\right) \\
&= P\left(X \le \frac{r}{r+1}\right) + P\left(X \ge \frac{1}{r+1}\right) \quad \left(\text{since } \frac{r}{r+1} \le \frac{1}{2} \text{ and } \frac{1}{r+1} \ge \frac{1}{2}\right) \\
&= \frac{r}{r+1} + 1 - \frac{1}{r+1} = \frac{2r}{r+1}
\end{aligned}$$

For $r \in (0,1)$, the density, thus, equals

$$f_R(r) = \frac{d}{dr} F_R(r) = \frac{2}{(r+1)^2}.$$

6.2 Exponential random variable

A random variable is Exponential(λ), with parameter $\lambda > 0$, if it has the probability mass function given below. This is a distribution for the *waiting time* for some random event, for example, for a lightbulb to burn out or for the next earthquake of at least some given magnitude.

Properties:

1. Density: $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$
2. $EX = \frac{1}{\lambda}$.
3. $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.
4. $P(X \geq x) = e^{-\lambda x}$.
5. Memoryless property: $P(X \geq x + y | X \geq y) = e^{-\lambda x}$.

The last property means that, if the event has not occurred by some given time (no matter how large), the distribution of the remaining waiting time is the same as it was at the beginning. There is no "aging."

Proofs of these properties are integration exercises and are omitted.

Example 6.5. Assume that a lightbulb lasts on average 100 hours. Assuming exponential distribution, compute the probability that it lasts more than 200 hours and the probability that it lasts less than 50 hours.

Let X be the waiting time for the bulb to burn out. Then, X is Exponential with $\lambda = \frac{1}{100}$ and

$$P(X \geq 200) = e^{-2} \approx 0.1353,$$
$$P(X \leq 50) = 1 - e^{-\frac{1}{2}} \approx 0.3935.$$

6.3 Normal random variable

A random variable is Normal with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ or, in short, X is $N(\mu, \sigma^2)$, if its density is the function given below. Such a random variable is (at least approximately) very common. For example, measurement with random error, weight of a randomly caught yellow-billed magpie, SAT (or some other) test score of a randomly chosen student at UC Davis, etc.

Properties:

1. Density:
$$f(x) = f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$
where $x \in (-\infty, \infty)$.
2. $EX = \mu$.
3. $\mathrm{Var}(X) = \sigma^2$.

To show that
$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$
is a tricky exercise in integration, as is the computation of the variance. Assuming that the integral of f is 1, we can use symmetry to prove that EX must be μ:
$$\begin{aligned}
EX &= \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{\infty} (x-\mu) f(x)\, dx + \mu \int_{-\infty}^{\infty} f(x)\, dx \\
&= \int_{-\infty}^{\infty} (x-\mu)\, \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx + \mu \\
&= \int_{-\infty}^{\infty} z \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz + \mu \\
&= \mu,
\end{aligned}$$
where the last integral was obtained by the change of variable $z = x - \mu$ and is zero because the function integrated is odd.

Example 6.6. Let X be a $N(\mu, \sigma^2)$ random variable and let $Y = \alpha X + \beta$, with $\alpha > 0$. How is Y distributed?

If X is a "measurement with error" $\alpha X + \beta$ amounts to changing the units and so Y should still be normal. Let us see if this is the case. We start by computing the distribution function of Y,
$$\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(\alpha X + \beta \le y) \\
&= P\left(X \le \frac{y-\beta}{\alpha}\right) \\
&= \int_{-\infty}^{\frac{y-\beta}{\alpha}} f_X(x)\, dx
\end{aligned}$$

and, then, the density

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y-\beta}{\alpha}\right)\cdot\frac{1}{\alpha} \\ &= \frac{1}{\sqrt{2\pi}\sigma\alpha}\, e^{-\frac{(y-\beta-\alpha\mu)^2}{2\alpha^2\sigma^2}}. \end{aligned}$$

Therefore, Y is normal with $EY = \alpha\mu + \beta$ and $\text{Var}(Y) = (\alpha\sigma)^2$.

In particular,

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

has $EZ = 0$ and $\text{Var}(Z) = 1$. Such a $N(0,1)$ random variable is called *standard* Normal. Its distribution function $F_Z(z)$ is denoted by $\Phi(z)$. Note that

$$\begin{aligned} f_Z(z) &= \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2} \\ \Phi(z) = F_Z(z) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-x^2/2}\, dx. \end{aligned}$$

The integral for $\Phi(z)$ cannot be computed as an elementary function, so approximate values are given in tables. Nowadays, this is largely obsolete, as computers can easily compute $\Phi(z)$ very accurately for any given z. You should also note that it is enough to know these values for $z > 0$, as in this case, by using the fact that $f_Z(x)$ is an even function,

$$\Phi(-z) = \int_{-\infty}^{-z} f_Z(x)\, dx = \int_{z}^{\infty} f_Z(x)\, dx = 1 - \int_{-\infty}^{z} f_Z(x)\, dx = 1 - \Phi(z).$$

In particular, $\Phi(0) = \frac{1}{2}$. Another way to write this is $P(Z \geq -z) = P(Z \leq z)$, a form which is also often useful.

Example 6.7. What is the probability that a Normal random variable differs from its mean μ by more than σ? More than 2σ? More than 3σ?

In symbols, if X is $N(\mu, \sigma^2)$, we need to compute $P(|X-\mu| \geq \sigma)$, $P(|X-\mu| \geq 2\sigma)$, and $P(|X-\mu| \geq 3\sigma)$.

In this and all other examples of this type, the letter Z will stand for an $N(0,1)$ random variable.

We have

$$P(|X-\mu| \geq \sigma) = P\left(\left|\frac{X-\mu}{\sigma}\right| \geq 1\right) = P(|Z| \geq 1) = 2P(Z \geq 1) = 2(1-\Phi(1)) \approx 0.3173.$$

Similarly,

$$\begin{aligned} P(|X-\mu| \geq 2\sigma) = 2(1-\Phi(2)) \approx 0.0455, \\ P(|X-\mu| \geq 3\sigma) = 2(1-\Phi(3)) \approx 0.0027. \end{aligned}$$

Example 6.8. Assume that X is Normal with mean $\mu = 2$ and variance $\sigma^2 = 25$. Compute the probability that X is between 1 and 4.

Here is the computation:

$$\begin{aligned} P(1 \le X \le 4) &= P\left(\frac{1-2}{5} \le \frac{X-2}{5} \le \frac{4-2}{5}\right) \\ &= P(-0.2 \le Z \le 0.4) \\ &= P(Z \le 0.4) - P(Z \le -0.2) \\ &= \Phi(0.4) - (1 - \Phi(0.2)) \\ &\approx 0.2347\ . \end{aligned}$$

Let S_n be a Binomial(n, p) random variable. Recall that its mean is np and its variance $np(1-p)$. If we pretend that S_n is Normal, then $\frac{S_n - np}{\sqrt{np(1-p)}}$ is standard Normal, i.e., $N(0, 1)$. The following theorem says that this is approximately true if p is fixed (e.g., 0.5) and n is large (e.g., $n = 100$).

Theorem 6.1. *De Moivre-Laplace Central Limit Theorem.*

Let S_n be Binomial(n, p), where p is fixed and n is large. Then, $\frac{S_n - np}{\sqrt{np(1-p)}} \approx N(0, 1)$; more precisely,

$$P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right) \to \Phi(x)$$

as $n \to \infty$, for every real number x.

We should also note that the above theorem is an analytical statement; it says that

$$\sum_{k: 0 \le k \le np + x\sqrt{np(1-p)}} \binom{n}{k} p^k (1-p)^{n-k} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{s^2}{2}}\, ds$$

as $n \to \infty$, for every $x \in \mathbb{R}$. Indeed it can be, and originally was, proved this way, with a lot of computational work.

An important issue is the quality of the Normal approximation to the Binomial. One can prove that the difference between the Binomial probability (in the above theorem) and its limit is at most

$$\frac{0.5 \cdot (p^2 + (1-p)^2)}{\sqrt{n\, p\, (1-p)}}.$$

A commonly cited rule of thumb is that this is a decent approximation when $np(1-p) \ge 10$; however, if we take $p = 1/3$ and $n = 45$, so that $np(1-p) = 10$, the bound above is about 0.0878, too large for many purposes. Various corrections have been developed to diminish the error, but they are, in my opinion, obsolete by now. In the situation when the above upper bound on the error is too high, we should simply compute directly with the Binomial distribution and not use the Normal approximation. (We will assume that the approximation is adequate in the examples below.) Remember that, when n is large and p is small, say $n = 100$ and $p = \frac{1}{100}$, the Poisson approximation (with $\lambda = np$) is much better!

Example 6.9. A roulette wheel has 38 slots: 18 red, 18 black, and 2 green. The ball ends at one of these at random. You are a player who plays a large number of games and makes an even bet of \$1 on red in every game. After n games, what is the probability that you are ahead? Answer this for $n = 100$ and $n = 1000$.

Let S_n be the number of times you win. This is a Binomial$(n, \frac{9}{19})$ random variable.

$$\begin{aligned}
P(\text{ahead}) &= P(\text{win more than half of the games}) \\
&= P\left(S_n > \frac{n}{2}\right) \\
&= P\left(\frac{S_n - np}{\sqrt{np(1-p)}} > \frac{\frac{1}{2}n - np}{\sqrt{np(1-p)}}\right) \\
&\approx P\left(Z > \frac{(\frac{1}{2} - p)\sqrt{n}}{\sqrt{p(1-p)}}\right)
\end{aligned}$$

For $n = 100$, we get

$$P\left(Z > \frac{5}{\sqrt{90}}\right) \approx 0.2990,$$

and for $n = 1000$, we get

$$P\left(Z > \frac{5}{3}\right) \approx 0.0478.$$

For comparison, the true probabilities are 0.2650 and 0.0448, respectively.

Example 6.10. What would the answer to the previous example be if the game were fair, i.e., you bet even money on the outcome of a fair coin toss each time.

Then, $p = \frac{1}{2}$ and

$$P(\text{ahead}) \to P(Z > 0) = 0.5,$$

as $n \to \infty$.

Example 6.11. How many times do you need to toss a fair coin to get 100 heads with probability 90%?

Let n be the number of tosses that we are looking for. For S_n, which is Binomial$(n, \frac{1}{2})$, we need to find n so that

$$P(S_n \geq 100) \approx 0.9.$$

We will use below that $n > 200$, as the probability would be approximately $\frac{1}{2}$ for $n = 200$ (see the previous example). Here is the computation:

$$
\begin{aligned}
P\left(\frac{S_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \geq \frac{100 - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}\right) &\approx P\left(Z \geq \frac{100 - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}\right) \\
&= P\left(Z \geq \frac{200 - n}{\sqrt{n}}\right) \\
&= P\left(Z \geq -\left(\frac{n - 200}{\sqrt{n}}\right)\right) \\
&= P\left(Z \leq \frac{n - 200}{\sqrt{n}}\right) \\
&= \Phi\left(\frac{n - 200}{\sqrt{n}}\right) \\
&= 0.9
\end{aligned}
$$

Now, according to the tables, $\Phi(1.28) \approx 0.9$, thus we need to solve $\frac{n-200}{\sqrt{n}} = 1.28$, that is,

$$n - 1.28\sqrt{n} - 200 = 0.$$

This is a quadratic equation in $\sqrt{n}$, with the only positive solution

$$\sqrt{n} = \frac{1.28 + \sqrt{1.28^2 + 800}}{2}.$$

Rounding up the number n we get from above, we conclude that $n = 219$.

Problems

1. A random variable X has the density function

$$f(x) = \begin{cases} c(x + \sqrt{x}) & x \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

(a) Determine c. (b) Compute $E(1/X)$. (c) Determine the probability density function of $Y = X^2$.

2. The density function of a random variable X is given by

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

We also know that $E(X) = 7/6$. (a) Compute a and b. (b) Compute $\mathrm{Var}(X)$.

3. After your complaint about their service, a representative of an insurance company promised to call you "between 7 and 9 this evening." Assume that this means that the time T of the call is uniformly distributed in the specified interval.

(a) Compute the probability that the call arrives between 8:00 and 8:20.

(b) At 8:30, the call still hasn't arrived. What is the probability that it arrives in the next 10 minutes?

(c) Assume that you know in advance that the call will last *exactly* 1 hour. From 9 to 9:30, there is a game show on TV that you wanted to watch. Let M be the amount of time of the show that you miss because of the call. Compute the expected value of M.

4. Toss a fair coin twice. You win \$1 if *at least one* of the two tosses comes out heads.

(a) Assume that you play this game 300 times. What is, approximately, the probability that you win at least \$250?

(b) Approximately how many times do you need to play so that you win at least \$250 with probability at least 0.99?

5. Roll a die n times and let M be the number of times you roll 6. Assume that n is large.

(a) Compute the expectation EM.

(b) Write down an approximation, in terms on n and Φ, of the probability that M differs from its expectation by less than 10%.

(c) How large should n be so that the probability in (b) is larger than 0.99?

Solutions

1. (a) As

$$1 = c\int_0^1 (x+\sqrt{x})\,dx = c\left(\frac{1}{2}+\frac{2}{3}\right) = \frac{7}{6}c,$$

it follows that $c = \frac{6}{7}$.

(b)

$$\frac{6}{7}\int_0^1 \frac{1}{x}(x+\sqrt{x})\,dx = \frac{18}{7}.$$

(c)

$$\begin{aligned} F_r(y) &= P(Y \le y) \\ &= P(X \le \sqrt{y}) \\ &= \frac{6}{7}\int_0^{\sqrt{y}} (x+\sqrt{x})\,dx, \end{aligned}$$

and so

$$f_Y(y) = \begin{cases} \frac{3}{7}(1 + y^{-\frac{1}{4}}) & \text{if } y \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

2. (a) From $\int_0^1 f(x)\,dx = 1$ we get $2a + 2b = 1$ and from $\int_0^1 xf(x)\,dx = \frac{7}{6}$ we get $2a + \frac{8}{3}b = \frac{7}{6}$. The two equations give $a = b = \frac{1}{4}$.

(b) $E(X^2) = \int_0^1 x^2 f(x)\,dx = \frac{5}{3}$ and so $\text{Var}(X) = \frac{5}{3} - (\frac{7}{6})^2 = \frac{11}{36}$.

3. (a) $\frac{1}{6}$.

(b) Let T be the time of the call, from 7pm, in minutes; T is uniform on $[0, 120]$. Thus,

$$P(T \leq 100 | T \geq 90) = \frac{1}{3}.$$

(c) We have

$$M = \begin{cases} 0 & \text{if } 0 \leq T \leq 60, \\ T - 60 & \text{if } 60 \leq T \leq 90, \\ 30 & \text{if } 90 \leq T. \end{cases}$$

Then,

$$EM = \frac{1}{120} \int_{60}^{90} (t - 60)\,dx + \frac{1}{120} \int_{90}^{120} 30\,dx = 11.25.$$

4. (a) $P(\text{win a single game}) = \frac{3}{4}$. If you play n times, the number X of games you win is $\text{Binomial}(n, \frac{3}{4})$. If Z is $N(0,1)$, then

$$P(X \geq 250) \approx P\left(Z \geq \frac{250 - n \cdot \frac{3}{4}}{\sqrt{n \cdot \frac{3}{4} \cdot \frac{1}{4}}}\right).$$

For (a), $n = 300$ and the above expression is $P(Z \geq \frac{10}{3})$, which is approximately $1 - \Phi(3.33) \approx 0.0004$.

For (b), you need to find n so that the above expression is 0.99 or so that

$$\Phi\left(\frac{250 - n \cdot \frac{3}{4}}{\sqrt{n \cdot \frac{3}{4} \cdot \frac{1}{4}}}\right) = 0.01.$$

The argument must be negative, hence

$$\frac{250 - n \cdot \frac{3}{4}}{\sqrt{n \cdot \frac{3}{4} \cdot \frac{1}{4}}} = -2.33.$$

If $x = \sqrt{3n}$, this yields

$$x^2 - 4.66x - 1000 = 0$$

and solving the quadratic equation gives $x \approx 34.04$, $n > (34.04)^2/3$, $n \geq 387$.

5. (a) M is Binomial$(n, \frac{1}{6})$, so $EM = \frac{n}{6}$.

(b)

$$P\left(\left|M - \frac{n}{6}\right| < \frac{n}{6} \cdot 0.1\right) \approx P\left(|Z| < \frac{\frac{n}{6} \cdot 0.1}{\sqrt{n \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = 2\Phi\left(\frac{0.1\sqrt{n}}{\sqrt{5}}\right) - 1.$$

(c) The above must be 0.99 and so $\Phi\left(\frac{0.1\sqrt{n}}{\sqrt{5}}\right) = 0.995$, $\frac{0.1\sqrt{n}}{\sqrt{5}} = 2.57$, and, finally, $n \geq 3303$.

7 Joint Distributions and Independence

Discrete Case

Assume that you have a pair (X,Y) of discrete random variables X and Y. Their *joint probability mass function* is given by

$$p(x,y) = P(X = x, Y = y)$$

so that

$$P((X,Y) \in A) = \sum_{(x,y)\in A} p(x,y).$$

The *marginal probability mass functions* are the p. m. f.'s of X and Y, given by

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_y p(x,y)$$
$$P(Y = y) = \sum_x P(X = x, Y = y) = \sum_x p(x,y)$$

Example 7.1. An urn has 2 red, 5 white, and 3 green balls. Select 3 balls at random and let X be the number of red balls and Y the number of white balls. Determine (a) joint p. m. f. of (X,Y), (b) marginal p. m. f.'s, (c) $P(X \geq Y)$, and (d) $P(X = 2|X \geq Y)$.

The joint p. m. f. is given by $P(X = x, Y = y)$ for all possible x and y. In our case, x can be $0, 1$, or 2 and y can be $0, 1, 2$, or 3. The values are given in the table.

$y \backslash x$	0	1	2	$P(Y = y)$
0	$1/120$	$2 \cdot 3/120$	$3/120$	$10/120$
1	$5 \cdot 3/120$	$2 \cdot 5 \cdot 3/120$	$5/120$	$50/120$
2	$10 \cdot 3/120$	$10 \cdot 2/120$	0	$50/120$
3	$10/120$	0	0	$10/120$
$P(X = x)$	$56/120$	$56/120$	$8/120$	1

The last row and column entries are the respective column and row sums and, therefore, determine the marginal p. m. f.'s. To answer (c) we merely add the relevant probabilities,

$$P(X \geq Y) = \frac{1 + 6 + 3 + 30 + 5}{120} = \frac{3}{8},$$

and, to answer (d), we compute

$$\frac{P(X = 2, X \geq Y)}{P(X \geq Y)} = \frac{\frac{8}{120}}{\frac{3}{8}} = \frac{8}{45}.$$

Two random variables X and Y are *independent* if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all intervals A and B. In the discrete case, X and Y are independent exactly when

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all possible values x and y of X and Y, that is, the joint p. m. f. is the product of the marginal p. m. f.'s.

Example 7.2. In the previous example, are X and Y independent?

No, the 0's in the table are dead giveaways. For example, $P(X = 2, Y = 2) = 0$, but neither $P(X = 2)$ nor $P(Y = 2)$ is 0.

Example 7.3. Most often, independence is an assumption. For example, roll a die twice and let X be the number on the first roll and let Y be the number on the second roll. Then, X and Y are independent: we are used to assuming that all 36 outcomes of the two rolls are equally likely, which is the same as assuming that the two random variables are discrete uniform (on $\{1, 2, \ldots, 6\}$) and independent.

Continuous Case

We say that (X, Y) is a *jointly continuous* pair of random variables if there exists a *joint density* $f(x, y) \geq 0$ so that

$$P((X, Y) \in S) = \iint_S f(x, y)\, dx\, dy,$$

where S is some nice (say, open or closed) subset of $\mathbb{R}^2$

Example 7.4. Let (X, Y) be a random point in S, where S is a compact (that is, closed and bounded) subset of $\mathbb{R}^2$. This means that

$$f(x, y) = \begin{cases} \frac{1}{\text{area}(S)} & \text{if } (x, y) \in S, \\ 0 & \text{otherwise.} \end{cases}$$

The simplest example is a square of side length 1 where

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

Example 7.5. Let

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

Determine (a) the constant c, (b) $P(X \ge Y)$, (c) $P(X = Y)$, and (d) $P(X = 2Y)$.

For (a),

$$\begin{aligned} \int_{-1}^{1} dx \int_{x^2}^{1} c\,x^2 y\,dy &= 1 \\ c \cdot \frac{4}{21} &= 1 \end{aligned}$$

and so

$$c = \frac{21}{4}.$$

For (b), let S be the region between the graphs $y = x^2$ and $y = x$, for $x \in (0,1)$. Then,

$$\begin{aligned} P(X \ge Y) &= P((X,Y) \in S) \\ &= \int_0^1 dx \int_{x^2}^{x} \frac{21}{4} \cdot x^2 y\,dy \\ &= \frac{3}{20} \end{aligned}$$

Both probabilities in (c) and (d) are 0 because a two-dimensional integral over a line is 0.

If f is the joint density of (X,Y), then the two *marginal densities*, which are the densities of X and Y, are computed by integrating out the other variable:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy, \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx. \end{aligned}$$

Indeed, for an interval A, $X \in A$ means that $(X,Y) \in S$, where $S = A \times \mathbb{R}$, and, therefore,

$$P(X \in A) = \int_A dx \int_{-\infty}^{\infty} f(x,y)\,dy.$$

The marginal densities formulas follow from the definition of density. With some advanced calculus expertise, the following can be checked.

Two jointly continuous random variables X and Y are independent exactly when the joint density is the product of the marginal ones:

$$f(x,y) = f_X(x) \cdot f_Y(y),$$

for all x and y.

Example 7.6. Previous example, continued. Compute marginal densities and determine whether X and Y are independent.

We have

$$f_X(x) = \int_{x^2}^{1} \frac{21}{4} \cdot x^2 y \, dy = \frac{21}{8} x^2(1 - x^4),$$

for $x \in [-1, 1]$, and 0 otherwise. Moreover,

$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} \cdot x^2 y \, dx = \frac{7}{2} y^{\frac{5}{2}},$$

where $y \in [0, 1]$, and 0 otherwise. The two random variables X and Y are clearly not independent, as $f(x, y) \neq f_X(x) f_Y(y)$.

Example 7.7. Let (X, Y) be a random point in a square of length 1 with the bottom left corner at the origin. Are X and Y independent?

$$f(x, y) = \begin{cases} 1 & (x, y) \in [0, 1] \times [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

The marginal densities are

$$f_X(x) = 1,$$

if $x \in [0, 1]$, and

$$f_Y(y) = 1,$$

if $y \in [0, 1]$, and 0 otherwise. Therefore, X and Y are independent.

Example 7.8. Let (X, Y) be a random point in the triangle $\{(x, y) : 0 \leq y \leq x \leq 1\}$. Are X and Y independent?

Now

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

The marginal densities are

$$f_X(x) = 2x,$$

if $x \in [0, 1]$, and

$$f_Y(y) = 2(1 - y),$$

if $y \in [0, 1]$, and 0 otherwise. So X and Y are no longer distributed uniformly and no longer independent.

We can make a more general conclusion from the last two examples. Assume that (X, Y) is a jointly continuous pair of random variables, uniform on a compact set $S \subset \mathbb{R}^2$. If they are to be independent, their marginal densities have to be constant, thus uniform on some sets, say A and B, and then $S = A \times B$. (If A and B are both intervals, then $S = A \times B$ is a rectangle, which is the most common example of independence.)

Example 7.9. Mr. and Mrs. Smith agree to meet at a specified location "between 5 and 6 p.m." Assume that they both arrive there at a random time between 5 and 6 and that their arrivals are independent. (a) Find the density for the time one of them will have to wait for the other. (b) Mrs. Smith later tells you she had to wait; given this information, compute the probability that Mr. Smith arrived before 5:30.

Let X be the time when Mr. Smith arrives and let Y be the time when Mrs. Smith arrives, with the time unit 1 hour. The assumptions imply that (X, Y) is uniform on $[0,1] \times [0,1]$.

For (a), let $T = |X - Y|$, which has possible values in $[0,1]$. So, fix $t \in [0,1]$ and compute (drawing a picture will also help)

$$\begin{aligned} P(T \le t) &= P(|X-Y| \le t) \\ &= P(-t \le X - Y \le t) \\ &= P(X - t \le Y \le X + t) \\ &= 1 - (1-t)^2 \\ &= 2t - t^2, \end{aligned}$$

and so

$$f_T(t) = 2 - 2t.$$

For (b), we need to compute

$$P(X \le 0.5 | X > Y) = \frac{P(X \le 0.5, X > Y)}{P(X > Y)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}.$$

Example 7.10. Assume that X and Y are independent, that X is uniform on $[0,1]$, and that Y has density $f_Y(y) = 2y$, for $y \in [0,1]$, and 0 elsewhere. Compute $P(X + Y \le 1)$.

The assumptions determine the joint density of (X, Y)

$$f(x,y) = \begin{cases} 2y & \text{if } (x,y) \in [0,1] \times [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

To compute the probability in question we compute

$$\int_0^1 dx \int_0^{1-x} 2y\, dy$$

or

$$\int_0^1 dy \int_0^{1-y} 2y\, dx,$$

whichever double integral is easier. The answer is $\frac{1}{3}$.

Example 7.11. Assume that you are waiting for two phone calls, from Alice and from Bob. The waiting time T_1 for Alice's call has expectation 10 minutes and the waiting time T_2 for

Bob's call has expectation 40 minutes. Assume T_1 and T_2 are independent exponential random variables. What is the probability that Alice's call will come first?

We need to compute $P(T_1 < T_2)$. Assuming our unit is 10 minutes, we have, for $t_1, t_2 > 0$,

$$f_{T_1}(t_1) = e^{-t_1}$$

and

$$f_{T_2}(t_2) = \frac{1}{4}e^{-t_2/4},$$

so that the joint density is

$$f(t_1, t_2) = \frac{1}{4}e^{-t_1 - t_2/4},$$

for $t_1, t_2 \geq 0$. Therefore,

$$\begin{aligned} P(T_1 < T_2) &= \int_0^\infty dt_1 \int_{t_1}^\infty \frac{1}{4}e^{-t_1-t_2/4}\,dt_2 \\ &= \int_0^\infty e^{-t_1}\,dt_1\,e^{-t_1/4} \\ &= \int_0^\infty e^{\frac{-5t_1}{4}}\,dt_1 \\ &= \frac{4}{5}. \end{aligned}$$

Example 7.12. *Buffon needle problem.* Parallel lines at a distance 1 are drawn on a large sheet of paper. Drop a needle of length ℓ onto the sheet. Compute the probability that it intersects one of the lines.

Let D be the distance from the center of the needle to the nearest line and let Θ be the acute angle relative to the lines. We will, reasonably, *assume* that D and Θ are independent and uniform on their respective intervals $0 \leq D \leq \frac{1}{2}$ and $0 \leq \Theta \leq \frac{\pi}{2}$. Then,

$$\begin{aligned} P(\text{the needle intersects a line}) &= P\left(\frac{D}{\sin\Theta} < \frac{\ell}{2}\right) \\ &= P\left(D < \frac{\ell}{2}\sin\Theta\right). \end{aligned}$$

Case 1: $\ell \leq 1$. Then, the probability equals

$$\frac{\int_0^{\pi/2} \frac{\ell}{2}\sin\theta\,d\theta}{\pi/4} = \frac{\ell/2}{\pi/4} = \frac{2\ell}{\pi}.$$

When $\ell = 1$, you famously get $\frac{2}{\pi}$, which can be used to get (very poor) approximations for π.

Case 2: $\ell > 1$. Now, the curve $d = \frac{\ell}{2}\sin\theta$ intersects $d = \frac{1}{2}$ at $\theta = \arcsin\frac{1}{\ell}$. The probability equals

$$\frac{4}{\pi}\left[\frac{\ell}{2}\int_0^{\arcsin\frac{1}{\ell}} \sin\theta\,d\theta + \left(\frac{\pi}{2} - \arcsin\frac{1}{\ell}\right)\cdot\frac{1}{2}\right] = \frac{4}{\pi}\left[\frac{\ell}{2} - \frac{1}{2}\sqrt{\ell^2-1} + \frac{\pi}{4} - \frac{1}{2}\arcsin\frac{1}{\ell}\right].$$

A similar approach works for cases with more than two random variables. Let us do an example for illustration.

Example 7.13. Assume X_1, X_2, X_3 are uniform on $[0,1]$ and independent. What is $P(X_1 + X_2 + X_3 \le 1)$?

The joint density is

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3) = \begin{cases} 1 & \text{if } (x_1,x_2,x_3) \in [0,1]^3, \\ 0 & \text{otherwise.} \end{cases}$$

Here is how we get the answer:

$$P(X_1 + X_2 + X_3 \le 1) = \int_0^1 dx_1 \int_0^{1-x_1} dx_2 \int_0^{1-x_1-x_2} dx_3 = \frac{1}{6}.$$

In general, if $X_1, \ldots, X_n$ are independent and uniform on $[0,1]$,

$$P(X_1 + \ldots + X_n \le 1) = \frac{1}{n!}$$

Conditional distributions

The conditional p. m. f. of X given $Y = y$ is, in the discrete case, given simply by

$$p_X(x|Y = y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

This is trickier in the continuous case, as we cannot divide by $P(Y = y) = 0$.

For a jointly continuous pair of random variables X and Y, we define the *conditional density of X given $Y = y$* as follows:

$$f_X(x|Y = y) = \frac{f(x,y)}{f_Y(y)},$$

where $f(x,y)$ is, of course, the joint density of (X,Y).

Observe that when $f_Y(y) = 0$, $\int_{-\infty}^{\infty} f(x,y)\,dx = 0$, and so $f(x,y) = 0$ for every x. So, we have a $\frac{0}{0}$ expression, which we define to be 0.

Here is a "physicist's proof" why this should be the conditional density formula:

$$\begin{aligned} P(X = x + dx|Y = y + dy) &= \frac{P(X = x + dx, Y = y + dy)}{P(Y = y + dy)} \\ &= \frac{f(x,y)\,dx\,dy}{f_Y(y)\,dy} \\ &= \frac{f(x,y)}{f_Y(y)}\,dx \\ &= f_X(x|Y = y)\,dx\ . \end{aligned}$$

Example 7.14. Let (X, Y) be a random point in the triangle $\{(x, y) : x, y \geq 0, x + y \leq 1\}$. Compute $f_X(x|Y = y)$.

The joint density $f(x, y)$ equals 2 on the triangle. For a given $y \in [0, 1]$, we know that, if $Y = y$, X is between 0 and $1 - y$. Moreover,

$$f_Y(y) = \int_0^{1-y} 2\, dx = 2(1 - y).$$

Therefore,

$$f_X(x|Y = y) = \begin{cases} \frac{1}{1-y} & 0 \leq x \leq 1 - y, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, given $Y = y$, X is distributed uniformly on $[0, 1 - y]$, which is hardly surprising.

Example 7.15. Suppose (X, Y) has joint density

$$f(x, y) = \begin{cases} \frac{21}{4} x^2 y & x^2 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Compute $f_X(x|Y = y)$.

We compute first

$$f_Y(y) = \frac{21}{4} y \int_{-\sqrt{y}}^{\sqrt{y}} x^2\, dx = \frac{7}{2}\, y^{\frac{5}{2}},$$

for $y \in [0, 1]$. Then,

$$f_X(x|Y = y) = \frac{\frac{21}{4} x^2 y}{\frac{7}{2} y^{5/2}} = \frac{3}{2}\, x^2 y^{-3/2},$$

where $-\sqrt{y} \leq x \leq \sqrt{y}$.

Suppose we are asked to compute $P(X \geq Y|Y = y)$. This makes no literal sense because the probability $P(Y = y)$ of the condition is 0. We reinterpret this expression as

$$P(X \geq y|Y = y) = \int_y^{\infty} f_X(x|Y = y)\, dx,$$

which equals

$$\int_y^{\sqrt{y}} \frac{3}{2}\, x^2 y^{-3/2}\, dx = \frac{1}{2}\, y^{-3/2} \left(y^{3/2} - y^3 \right) = \frac{1}{2} \left(1 - y^{3/2} \right).$$

Problems

1. Let (X, Y) be a random point in the square $\{(x, y) : -1 \leq x, y \leq 1\}$. Compute the conditional probability $P(X \geq 0 \,|\, Y \leq 2X)$. (It may be a good idea to draw a picture and use elementary geometry, rather than calculus.)

2. Roll a fair die 3 times. Let X be the number of 6's obtained and Y the number of 5's.

(a) Compute the joint probability mass function of X and Y.

(b) Are X and Y independent?

3. X and Y are independent random variables and they have the same density function

$$f(x) = \begin{cases} c(2-x) & x \in (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

(a) Determine c. (b) Compute $P(Y \le 2X)$ and $P(Y < 2X)$.

4. Let X and Y be independent random variables, both uniformly distributed on $[0,1]$. Let $Z = \min(X,Y)$ be the smaller value of the two.

(a) Compute the density function of Z.

(b) Compute $P(X \le 0.5 | Z \le 0.5)$.

(c) Are X and Z independent?

5. The joint density of (X,Y) is given by

$$\begin{cases} f(x,y) = 3x & \text{if } 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute the conditional density of Y given $X = x$.

(b) Are X and Y independent?

Solutions to problems

1. After noting the relevant areas,

$$\begin{aligned} P(X \ge 0 | Y \le 2X) &= \frac{P(X \ge 0, Y \le 2X)}{P(Y \le 2X)} \\ &= \frac{\frac{1}{4}\left(2 - \frac{1}{2}\cdot\frac{1}{2}\cdot 1\right)}{\frac{1}{2}} \\ &= \frac{7}{8} \end{aligned}$$

2. (a) The joint p. m. f. is given by the table

$y\backslash x$	0	1	2	3	$P(Y=y)$
0	$\frac{4^3}{216}$	$\frac{3\cdot 4^2}{216}$	$\frac{3\cdot 4}{216}$	$\frac{1}{216}$	$\frac{125}{216}$
1	$\frac{3\cdot 4^2}{216}$	$\frac{3\cdot 2\cdot 4}{216}$	$\frac{3}{216}$	0	$\frac{75}{216}$
2	$\frac{3\cdot 4}{216}$	$\frac{3}{216}$	0	0	$\frac{15}{216}$
3	$\frac{1}{216}$	0	0	0	$\frac{1}{216}$
$P(X=x)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$	1

Alternatively, for $x, y = 0, 1, 2, 3$ and $x + y \le 3$,

$$P(X = x, Y = y) = \binom{3}{x}\binom{3-x}{y}\left(\frac{1}{6}\right)^{x+y}\left(\frac{4}{6}\right)^{3-x-y}$$

(b) No. $P(X = 3, Y = 3) = 0$ and $P(X = 3)P(Y = 3) \neq 0$.

3. (a) From

$$c\int_0^1 (2-x)\,dx = 1,$$

it follows that $c = \frac{2}{3}$.

(b) We have

$$\begin{aligned}
P(Y \le 2X) &= P(Y < 2X) \\
&= \int_0^1 dy \int_{y/2}^1 \frac{4}{9}(2-x)(2-y)\,dx \\
&= \frac{4}{9}\int_0^1 (2-y)\,dy \int_{y/2}^1 (2-x)\,dx \\
&= \frac{4}{9}\int_0^1 (2-y)\,dy \left[2\left(1-\frac{y}{2}\right) - \frac{1}{2}\left(1-\frac{y^2}{4}\right)\right] \\
&= \frac{4}{9}\int_0^1 (2-y)\left[\frac{3}{2} - y + \frac{y^2}{8}\right]\,dy \\
&= \frac{4}{9}\int_0^1 \left[3 - \frac{7}{2}y + \frac{5}{4}y^2 - \frac{y^3}{8}\right]\,dy \\
&= \frac{4}{9}\left[3 - \frac{7}{4} + \frac{5}{12} - \frac{1}{32}\right].
\end{aligned}$$

4. (a) For $z \in [0,1]$

$$P(Z \leq z) = 1 - P(\text{both } X \text{ and } Y \text{ are above } z) = 1 - (1-z)^2 = 2z - z^2,$$

so that

$$f_Z(z) = 2(1-z),$$

for $z \in [0,1]$ and 0 otherwise.

(b) From (a), we conclude that $P(Z \leq 0.5) = \frac{3}{4}$ and $P(X \leq 0.5, Z \leq 0.5) = P(X \leq 0.5) = \frac{1}{2}$, so the answer is $\frac{2}{3}$.

(c) No: $Z \leq X$.

5. (a) Assume that $x \in [0,1]$. As

$$f_X(x) = \int_0^x 3x\, dy = 3x^2,$$

we have

$$f_Y(y|X = x) = \frac{f(x,y)}{f_X(x)} = \frac{3x}{3x^2} = \frac{1}{x},$$

for $0 \leq y \leq x$. In other words, Y is uniform on $[0,x]$.

(b) As the answer in (a) depends on x, the two random variables are not independent.

Interlude: Practice Midterm 2

This practice exam covers the material from chapters 5 through 7. Give yourself 50 minutes to solve the four problems, which you may assume have equal point score.

1. A random variable X has density function

$$f(x) = \begin{cases} c(x + x^2), & x \in [0, 1], \\ 0, & \text{otherwise.} \end{cases}$$

(a) Determine c.

(b) Compute $E(1/X)$.

(c) Determine the probability density function of $Y = X^2$.

2. A certain country holds a presidential election, with two candidates running for office. Not satisfied with their choice, each voter casts a vote independently at random, based on the outcome of a fair coin flip. At the end, there are 4,000,000 valid votes, as well as 20,000 invalid votes.

(a) Using a relevant approximation, compute the probability that, in the final count of *valid votes only*, the numbers for the two candidates will differ by less than 1000 votes.

(b) Each *invalid vote* is double-checked independently with probability 1/5000. Using a relevant approximation, compute the probability that at least 3 invalid votes are double-checked.

3. Toss a fair coin 5 times. Let X be the total number of Heads among the first three tosses and Y the total number of Heads among the last three tosses. (Note that, if the third toss comes out Heads, it is counted both into X and into Y).

(a) Write down the joint probability mass function of X and Y.

(b) Are X and Y independent? Explain.

(c) Compute the conditional probability $P(X \geq 2 \,|\, X \geq Y)$.

4. Every working day, John comes to the bus stop exactly at 7am. He takes the first bus that arrives. The arrival of the first bus is an exponential random variable with expectation 20 minutes.

Also, every working day, and *independently*, Mary comes to the same bus stop at a random time, uniformly distributed between 7 and 7:30.

(a) What is the probability that tomorrow John will wait for more than 30 minutes?

(b) Assume day-to-day independence. Consider Mary *late* if she comes after 7:20. What is the probability that Mary will be late on 2 or more working days among the next 10 working days?

(c) What is the probability that John and Mary will meet at the station tomorrow?

Solutions to Practice Midterm 2

1. A random variable X has density function

$$f(x) = \begin{cases} c(x + x^2), & x \in [0,1], \\ 0, & \text{otherwise.} \end{cases}$$

(a) Determine c.

Solution:
Since

$$\begin{aligned} 1 &= c\int_0^1 (x + x^2)\,dx, \\ 1 &= c\left(\frac{1}{2} + \frac{1}{3}\right), \\ 1 &= \frac{5}{6}c, \end{aligned}$$

and so $c = \frac{6}{5}$.

(b) Compute $E(1/X)$.

Solution:

$$\begin{aligned} E\left(\frac{1}{X}\right) &= \frac{6}{5}\int_0^1 \frac{1}{x}(x + x^2)\,dx \\ &= \frac{6}{5}\int_0^1 (1 + x)\,dx \\ &= \frac{6}{5}\left(1 + \frac{1}{2}\right) \\ &= \frac{9}{5}. \end{aligned}$$

(c) Determine the probability density function of $Y = X^2$.

Solution:
The values of Y are in $[0,1]$, so we will assume that $y \in [0,1]$. Then,

$$\begin{aligned} F(y) &= P(Y \le y) \\ &= P(X^2 \le y) \\ &= P(X \le \sqrt{y}) \\ &= \int_0^{\sqrt{y}} \frac{6}{5}(x + x^2)\, dx, \end{aligned}$$

and so

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{6}{5}(\sqrt{y} + y) \cdot \frac{1}{2\sqrt{y}} \\ &= \frac{3}{5}(1 + \sqrt{y}). \end{aligned}$$

2. A certain country holds a presidential election, with two candidates running for office. Not satisfied with their choice, each voter casts a vote independently at random, based on the outcome of a fair coin flip. At the end, there are 4,000,000 valid votes, as well as 20,000 invalid votes.

   (a) Using a relevant approximation, compute the probability that, in the final count of *valid votes only*, the numbers for the two candidates will differ by less than 1000 votes.

   Solution:

   Let S_n be the vote count for candidate 1. Thus, S_n is Binomial(n, p), where $n = 4,000,000$ and $p = \frac{1}{2}$. Then, $n - S_n$ is the vote count for candidate 2.

$$\begin{aligned} P(|S_n - (n - S_n)| \le 1000) &= P(-1000 \le 2S_n - n \le 1000) \\ &= P\left(\frac{-500}{\sqrt{n \cdot \frac{1}{2} \cdot \frac{1}{2}}} \le \frac{S_n - n/2}{\sqrt{n \cdot \frac{1}{2} \cdot \frac{1}{2}}} \le \frac{500}{\sqrt{n \cdot \frac{1}{2} \cdot \frac{1}{2}}} \right) \\ &\approx P(-0.5 \le Z \le 0.5) \\ &= P(Z \le 0.5) - P(Z \le -0.5) \\ &= P(Z \le 0.5) - (1 - P(Z \le 0.5)) \\ &= 2P(Z \le 0.5) - 1 \\ &= 2\Phi(0.5) - 1 \\ &\approx 2 \cdot 0.6915 - 1 \\ &= 0.383. \end{aligned}$$

(b) Each *invalid vote* is double-checked independently with probability 1/5000. Using a relevant approximation, compute the probability that at least 3 invalid votes are double-checked.

Solution:

Now, let S_n be the number of double-checked votes, which is $\text{Binomial}\left(20000, \frac{1}{5000}\right)$ and thus approximately Poisson(4). Then,

$$\begin{aligned} P(S_n \geq 3) &= 1 - P(S_n = 0) - P(S_n = 1) - P(S_n = 2) \\ &\approx 1 - e^{-4} - 4e^{-4} - \frac{4^2}{2}e^{-4} \\ &= 1 - 13e^{-4}. \end{aligned}$$

3. Toss a fair coin 5 times. Let X be the total number of Heads among the first three tosses and Y the total number of Heads among the last three tosses. (Note that, if the third toss comes out Heads, it is counted both into X and into Y).

(a) Write down the joint probability mass function of X and Y.

Solution:

$P(X = x, Y = y)$ is given by the table

$x\backslash y$	0	1	2	3
0	1/32	2/32	1/32	0
1	2/32	5/32	4/32	1/32
2	1/32	4/32	5/32	2/32
3	0	1/32	2/32	1/32

To compute these, observe that the number of outcomes is $2^5 = 32$. Then,

$$\begin{aligned} P(X = 2, Y = 1) &= P(X = 2, Y = 1, \text{ 3rd coin Heads}) \\ &\quad + P(X = 2, Y = 1, \text{ 3rd coin Tails}) \\ &= \frac{2}{32} + \frac{2}{32} = \frac{4}{32}, \\ P(X = 2, Y = 2) &= \frac{2 \cdot 2}{32} + \frac{1}{32} = \frac{5}{32}, \\ P(X = 1, Y = 1) &= \frac{2 \cdot 2}{32} + \frac{1}{32} = \frac{5}{32}, \end{aligned}$$

etc.

(b) Are X and Y independent? Explain.

Solution:

No,

$$P(X = 0, Y = 3) = 0 \neq \frac{1}{8} \cdot \frac{1}{8} = P(X = 0)P(Y = 3).$$

(c) Compute the conditional probability $P(X \geq 2 \,|\, X \geq Y)$.

Solution:

$$\begin{aligned} P(X \geq 2 | X \geq Y) &= \frac{P(X \geq 2, X \geq Y)}{P(X \geq Y)} \\ &= \frac{1+4+5+1+2+1}{1+2+1+5+4+1+5+2+1} \\ &= \frac{14}{22} \\ &= \frac{7}{11}. \end{aligned}$$

4. Every working day, John comes to the bus stop exactly at 7am. He takes the first bus that arrives. The arrival of the first bus is an exponential random variable with expectation 20 minutes.

   Also, every working day, and *independently*, Mary comes to the same bus stop at a random time, uniformly distributed between 7 and 7:30.

   (a) What is the probability that tomorrow John will wait for more than 30 minutes?

   Solution:

   Assume that the time unit is 10 minutes. Let T be the arrival time of the bus. It is Exponential with parameter $\lambda = \frac{1}{2}$. Then,

$$f_T(t) = \frac{1}{2} e^{-t/2},$$

   for $t \geq 0$, and

$$P(T \geq 3) = e^{-3/2}.$$

(b) Assume day-to-day independence. Consider Mary *late* if she comes after 7:20. What is the probability that Mary will be late on 2 or more working days among the next 10 working days?

Solution:

Let X be Mary's arrival time. It is uniform on $[0,3]$. Therefore,

$$P(X \geq 2) = \frac{1}{3}.$$

The number of late days among 10 days is Binomial$\left(10, \frac{1}{3}\right)$ and, therefore,

$$\begin{aligned}
& P(2 \text{ or more late working days among 10 working days}) \\
= \; & 1 - P(0 \text{ late}) - P(1 \text{ late}) \\
= \; & 1 - \left(\frac{2}{3}\right)^{10} - 10 \cdot \frac{1}{3}\left(\frac{2}{3}\right)^{9}.
\end{aligned}$$

(c) What is the probability that John and Mary will meet at the station tomorrow?

Solution:
We have

$$f_{(T,X)}(t,x) = \frac{1}{6}e^{-t/2},$$

for $x \in [0,3]$ and $t \geq 0$. Therefore,

$$\begin{aligned}
P(X \geq T) &= \frac{1}{3}\int_0^3 dx \int_x^\infty \frac{1}{2}e^{-t/2}\, dt \\
&= \frac{1}{3}\int_0^3 e^{-x/2}\, dx \\
&= \frac{2}{3}(1 - e^{-3/2}).
\end{aligned}$$

8 More on Expectation and Limit Theorems

Given a pair of random variables (X, Y) with joint density f and another function g of two variables,

$$Eg(X, Y) = \int\int g(x, y)f(x, y)\,dxdy;$$

if instead (X, Y) is a discrete pair with joint probability mass function p, then

$$Eg(X, Y) = \sum_{x,y} g(x, y)p(x, y).$$

Example 8.1. Assume that two among the 5 items are defective. Put the items in a random order and inspect them one by one. Let X be the number of inspections needed to find the first defective item and Y the number of additional inspections needed to find the second defective item. Compute $E|X - Y|$.

The joint p. m. f. of (X, Y) is given by the following table, which lists $P(X = i, Y = j)$, together with $|i - j|$ in parentheses, whenever the probability is nonzero:

$i\backslash j$	1	2	3	4
1	.1 (0)	.1 (1)	.1 (2)	.1 (3)
2	.1 (1)	.1 (0)	.1 (1)	0
3	.1 (2)	.1 (1)	0	0
4	.1 (3)	0	0	0

The answer is, therefore, $E|X - Y| = 1.4$.

Example 8.2. Assume that (X, Y) is a random point in the right triangle $\{(x, y) : x, y \geq 0, x + y \leq 1\}$. Compute EX, EY, and EXY.

Note that the density is 2 on the triangle, and so,

$$\begin{aligned}
EX &= \int_0^1 dx \int_0^{1-x} x\,2\,dy \\
&= \int_0^1 2x(1-x)\,dx \\
&= 2\left(\frac{1}{2} - \frac{1}{3}\right) \\
&= \frac{1}{3},
\end{aligned}$$

and, therefore, by symmetry, $EY = EX = \frac{1}{3}$. Furthermore,

$$\begin{aligned} E(XY) &= \int_0^1 dx \int_0^{1-x} xy2\,dy \\ &= \int_0^1 2x\frac{(1-x)^2}{2}\,dx \\ &= \int_0^1 (1-x)x^2\,dx \\ &= \frac{1}{3} - \frac{1}{4} \\ &= \frac{1}{12}. \end{aligned}$$

Linearity and monotonicity of expectation

Theorem 8.1. *Expectation is linear and monotone:*

1. *For constants a and b, $E(aX + b) = aE(X) + b$.*
2. *For arbitrary random variables $X_1, \dots, X_n$ whose expected values exist,*

$$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n).$$

3. *For two random variables $X \le Y$, we have $EX \le EY$.*

Proof. We will check the second property for $n = 2$ and the continuous case, that is,

$$E(X + Y) = EX + EY.$$

This is a consequence of the same property for two-dimensional integrals:

$$\iint (x+y)f(x,y)\,dx\,dy = \iint xf(x,y)\,dx\,dy + \iint yf(x,y)\,dx\,dy.$$

To prove this property for arbitrary n (and the continuous case), one can simply proceed by induction.

By the way we defined expectation, the third property is not immediately obvious. However, it is clear that $Z \ge 0$ implies $EZ \ge 0$ and, applying this to $Z = Y - X$, together with linearity, establishes monotonicity. □

We emphasize again that linearity holds for arbitrary random variables which *do not need to be independent*! This is very useful. For example, we can often write a random variable X as a sum of (possibly dependent) indicators, $X = I_1 + \cdots + I_n$. An instance of this method is called the *indicator trick*.

Example 8.3. Assume that an urn contains 10 black, 7 red, and 5 white balls. Select 5 balls (a) with and (b) without replacement and let X be the number of red balls selected. Compute EX.

Let I_i be the indicator of the event that the ith ball is red, that is,

$$I_i = I_{\{i\text{th ball is red}\}} = \begin{cases} 1 & \text{if } i\text{th ball is red,} \\ 0 & \text{otherwise.} \end{cases}$$

In both cases, $X = I_1 + I_2 + I_3 + I_4 + I_5$.

In (a), X is Binomial$(5, \frac{7}{22})$, so we know that $EX = 5 \cdot \frac{7}{22}$, but we will not use this knowledge. Instead, it is clear that

$$EI_1 = 1 \cdot P(\text{1st ball is red}) = \frac{7}{22} = EI_2 = \ldots = EI_5.$$

Therefore, by additivity, $EX = 5 \cdot \frac{7}{22}$.

For (b), one solution is to compute the p. m. f. of X,

$$P(X = i) = \frac{\binom{7}{i}\binom{15}{5-i}}{\binom{22}{5}},$$

where $i = 0, 1, \ldots, 5$, and then

$$EX = \sum_{i=0}^{5} i \, \frac{\binom{7}{i}\binom{15}{5-i}}{\binom{22}{5}}.$$

However, the indicator trick *works exactly as before* (the fact that I_i are now dependent does not matter) and so the answer is also exactly the same, $EX = 5 \cdot \frac{7}{22}$.

Example 8.4. *Matching problem*, revisited. Assume n people buy n gifts, which are then assigned at random, and let X be the number of people who receive their own gift. What is EX?

This is another problem very well suited for the indicator trick. Let

$$I_i = I_{\{\text{person } i \text{ receives own gift}\}}.$$

Then,

$$X = I_1 + I_2 + \ldots + I_n.$$

Moreover,

$$EI_i = \frac{1}{n},$$

for all i, and so

$$EX = 1.$$

Example 8.5. Five married couples are seated around a table at random. Let X be the number of wives who sit next to their husbands. What is EX?

Now, let

$$I_i = I_{\{\text{wife } i \text{ sits next to her husband}\}}.$$

Then,

$$X = I_1 + \ldots + I_5,$$

and

$$EI_i = \frac{2}{9},$$

so that

$$EX = \frac{10}{9}.$$

Example 8.6. *Coupon collector problem*, revisited. Sample from n cards, with replacement, indefinitely. Let N be the number of cards you need to sample for a complete collection, i.e., to get all different cards represented. What is EN?

Let N_i be the number the number of *additional* cards you need to get the ith *new* card, *after* you have received the $(i-1)$st new card.

Then, N_1, the number of cards needed to receive the first new card, is trivial, as the first card you buy is new: $N_1 = 1$. Afterward, N_2, the number of additional cards needed to get the second new card is Geometric with success probability $\frac{n-1}{n}$. After that, N_3, the number of additional cards needed to get the third new card is Geometric with success probability $\frac{n-2}{n}$. In general, N_i is geometric with success probability $\frac{n-i+1}{n}$, $i = 1, \ldots, n$, and

$$N = N_1 + \ldots + N_n,$$

so that

$$EN = n\left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\right).$$

Now, we have

$$\sum_{i=2}^{n} \frac{1}{i} \le \int_1^n \frac{1}{x}\, dx \le \sum_{i=1}^{n-1} \frac{1}{i},$$

by comparing the integral with the Riemman sum at the left and right endpoints in the division of $[1, n]$ into $[1, 2]$, $[2, 3], \ldots, [n-1, n]$, and so

$$\log n \le \sum_{i=1}^{n} \frac{1}{i} \le \log n + 1,$$

which establishes the limit

$$\lim_{n\to\infty} \frac{EN}{n \log n} = 1.$$

Example 8.7. Assume that an urn contains 10 black, 7 red, and 5 white balls. Select 5 balls (a) with replacement and (b) without replacement, and let W be the number of white balls selected, and Y the number of different colors. Compute EW and EY.

We already know that

$$EW = 5 \cdot \frac{5}{22}$$

in either case.

Let I_b, I_r, and I_w be the indicators of the event that, respectively, black, red, and white balls are represented. Clearly,

$$Y = I_b + I_r + I_w,$$

and so, in the case with replacement

$$EY = 1 - \frac{12^5}{22^5} + 1 - \frac{15^5}{22^5} + 1 - \frac{17^5}{22^5} \approx 2.5289,$$

while in the case without replacement

$$EY = 1 - \frac{\binom{12}{5}}{\binom{22}{5}} + 1 - \frac{\binom{15}{5}}{\binom{22}{5}} + 1 - \frac{\binom{17}{5}}{\binom{22}{5}} \approx 2.6209.$$

Expectation and independence

Theorem 8.2. *Multiplicativity of expectation for independent factors.*

The expectation of the product of independent random variables is the product of their expectations, i.e., if X and Y are independent,

$$E[g(X)h(Y)] = Eg(X) \cdot Eh(Y)$$

Proof. For the continuous case,

$$\begin{aligned}
E[g(X)h(Y)] &= \iint g(x)h(y)f(x,y)\,dx\,dy \\
&= \iint g(x)h(y)f_X(x)g_Y(y)\,dxdy \\
&= \int g(x)f_X(x)\,dx \int h(y)f_Y(y)\,dy \\
&= Eg(X) \cdot Eh(Y)
\end{aligned}$$

□

Example 8.8. Let us return to a random point (X, Y) in the triangle $\{(x, y) : x, y \geq 0, x + y \leq 1\}$. We computed that $E(XY) = \frac{1}{12}$ and that $EX = EY = \frac{1}{3}$. The two random variables have $E(XY) \neq EX \cdot EY$, thus, they cannot be independent. Of course, we already knew that they were not independent.

If, instead, we pick a random point (X, Y) in the square $\{(x, y) : 0 \le x, y \le 1\}$, X and Y are independent and, therefore, $E(XY) = EX \cdot EY = \frac{1}{4}$.

Finally, pick a random point (X, Y) in the diamond of radius 1, that is, in the square with corners at $(0, 1)$, $(1, 0)$, $(0, -1)$, and $(-1, 0)$. Clearly, we have, by symmetry,

$$EX = EY = 0,$$

but also

$$\begin{aligned}
E(XY) &= \frac{1}{2}\int_{-1}^{1} dx \int_{-1+|x|}^{1-|x|} xy\, dy \\
&= \frac{1}{2}\int_{-1}^{1} x\, dx \int_{-1+|x|}^{1-|x|} y\, dy \\
&= \frac{1}{2}\int_{-1}^{1} x\, dx \cdot 0 \\
&= 0.
\end{aligned}$$

This is an example where $E(XY) = EX \cdot EY$ even though X and Y are not independent.

Computing expectation by conditioning

For a pair of random variables (X, Y), we define the *conditional expectation of Y given $X = x$* by

$$\begin{aligned}
E(Y|X = x) &= \sum_y y\, P(Y = y|X = x) \text{ (discrete case)}, \\
&= \int_y y\, f_Y(y|X = x) \text{ (continuous case)}.
\end{aligned}$$

Observe that $E(Y|X = x)$ is a function of x; let us call it $g(x)$ for a moment. We denote $g(X)$ by $E(Y|X)$. This is the *expectation of Y provided the value X is known*; note again that this is an expression dependent on X and so we can compute *its* expectation. Here is what we get.

Theorem 8.3. *Tower property.*

The formula $E(E(Y|X)) = EY$ holds; less mysteriously, in the discrete case

$$EY = \sum_x E(Y|X = x) \cdot P(X = x),$$

and, in the continuous case,

$$EY = \int E(Y|X = x) f_X(x)\, dx.$$

Proof. To verify this in the discrete case, we write out the expectation inside the sum:

$$\begin{aligned}
\sum_x \sum_y yP(Y = y|X = x) \cdot P(X = x) &= \sum_x \sum_y y \frac{P(X = x, Y = y)}{P(X = x)} \cdot P(X = x) \\
&= \sum_x \sum_y yP(X = x, Y = y) \\
&= EY
\end{aligned}$$

□

Example 8.9. Once again, consider a random point (X, Y) in the triangle $\{(x, y) : x, y \geq 0, x + y \leq 1\}$. Given that $X = x$, Y is distributed uniformly on $[0, 1 - x]$ and so

$$E(Y|X = x) = \frac{1}{2}(1 - x).$$

By definition, $E(Y|X) = \frac{1}{2}(1 - X)$, and the expectation of $\frac{1}{2}(1 - X)$ must, therefore, equal the expectation of Y, and, indeed, it does as $EX = EY = \frac{1}{3}$, as we know.

Example 8.10. Roll a die and then toss as many coins as shown up on the die. Compute the expected number of Heads.

Let X be the number on the die and let Y be the number of Heads. Fix an $x \in \{1, 2, \ldots, 6\}$. Given that $X = x$, Y is Binomial$(x, \frac{1}{2})$. In particular,

$$E(Y|X = x) = x \cdot \frac{1}{2},$$

and, therefore,

$$\begin{aligned}
E(\text{number of Heads}) &= EY \\
&= \sum_{x=1}^{6} x \cdot \frac{1}{2} \cdot P(X = x) \\
&= \sum_{x=1}^{6} x \cdot \frac{1}{2} \cdot \frac{1}{6} \\
&= \frac{7}{4}.
\end{aligned}$$

Example 8.11. Here is another job interview question. You die and are presented with three doors. One of them leads to heaven, one leads to one day in purgatory, and one leads to two days in purgatory. After your stay in purgatory is over, you go back to the doors and pick again, but the doors are reshuffled each time you come back, so you must in fact choose a door at random each time. How long is your expected stay in purgatory?

Code the doors 0, 1, and 2, with the obvious meaning, and let N be the number of days in purgatory.

Then

$$\begin{aligned}
E(N|\text{your first pick is door } 0) &= 0,\\
E(N|\text{your first pick is door } 1) &= 1 + EN,\\
E(N|\text{your first pick is door } 2) &= 2 + EN.
\end{aligned}$$

Therefore,

$$EN = (1 + EN)\frac{1}{3} + (2 + EN)\frac{1}{3},$$

and solving this equation gives

$$EN = 3.$$

Covariance

Let X, Y be random variables. We define the *covariance* of (or between) X and Y as

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= E((X - EX)(Y - EY))\\
&= E(XY - (EX) \cdot Y - (EY) \cdot X + EX \cdot EY)\\
&= E(XY) - EX \cdot EY - EY \cdot EX + EX \cdot EY\\
&= E(XY) - EX \cdot EY.
\end{aligned}$$

To summarize, the most useful formula is

$$\boxed{\mathrm{Cov}(X, Y) = E(XY) - EX \cdot EY.}$$

Note immediately that, if X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$, but the converse is false.

Let X and Y be indicator random variables, so $X = I_A$ and $Y = I_B$, for two events A and B. Then, $EX = P(A)$, $EY = P(B)$, $E(XY) = E(I_{A\cap B}) = P(A \cap B)$, and so

$$\mathrm{Cov}(X, Y) = P(A \cap B) - P(A)P(B) = P(A)[P(B|A) - P(B)].$$

If $P(B|A) > P(B)$, we say the two events are *positively correlated* and, in this case, the covariance is positive; if the events are *negatively correlated* all inequalities are reversed. For general random variables X and Y, $\mathrm{Cov}(X, Y) > 0$ intuitively means that, "on the average," increasing X will result in larger Y.

Variance of sums of random variables

Theorem 8.4. *Variance-covariance formula:*

$$\boxed{\begin{aligned}
E(\sum_{i=1}^n X_i)^2 &= \sum_{i=1}^n EX_i^2 + \sum_{i\neq j} E(X_iX_j),\\
\mathrm{Var}(\sum_{i=1}^n X_i) &= \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j).
\end{aligned}}$$

Proof. The first formula follows from writing the sum

$$\left(\sum_{i=1}^{n} X_i\right)^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j$$

and linearity of expectation. The second formula follows from the first:

$$\begin{aligned}
\mathrm{Var}(\sum_{i=1}^{n} X_i) &= E\left[\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i)\right]^2 \\
&= E\left[\sum_{i=1}^{n} (X_i - EX_i)\right]^2 \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} E(X_i - EX_i)E(X_j - EX_j),
\end{aligned}$$

which is equivalent to the formula. □

Corollary 8.5. *Linearity of variance for independent summands.*

If $X_1, X_2, \ldots, X_n$ *are independent, then* $\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$.

The variance-covariance formula often makes computing variance possible even if the random variables are not independent, especially when they are indicators.

Example 8.12. Let S_n be Binomial(n, p). We will now fulfill our promise from Chapter 5 and compute its expectation and variance.

The crucial observation is that $S_n = \sum_{i=1}^{n} I_i$, where I_i is the indicator $I_{\{i\text{th trial is a success}\}}$. Therefore, I_i are independent. Then, $ES_n = np$ and

$$\begin{aligned}
\mathrm{Var}(S_n) &= \sum_{i=1}^{n} \mathrm{Var}(I_i) \\
&= \sum_{i=1}^{n} (EI_i - (EI_i)^2) \\
&= n(p - p^2) \\
&= np(1-p).
\end{aligned}$$

Example 8.13. *Matching problem*, revisited yet again. Recall that X is the number of people who get their own gift. We will compute $\mathrm{Var}(X)$.

Recall also that $X = \sum_{i=1}^n I_i$, where $I_i = I_{\{i\text{th person gets own gift}\}}$, so that $EI_i = \frac{1}{n}$ and

$$\begin{aligned} E(X^2) &= n \cdot \frac{1}{n} + \sum_{i \neq j} E(I_i I_j) \\ &= 1 + \sum_{i \neq j} \frac{1}{n(n-1)} \\ &= 1 + 1 \\ &= 2. \end{aligned}$$

The $E(I_i I_j)$ above is the probability that the ith person and jth person both get their own gifts and, thus, equals $\frac{1}{n(n-1)}$. We conclude that $\mathrm{Var}(X) = 1$. (In fact, X is, for large n, very close to Poisson with $\lambda = 1$.)

Example 8.14. Roll a die 10 times. Let X be the number of 6's rolled and Y be the number of 5's rolled. Compute $\mathrm{Cov}(X, Y)$.

Observe that $X = \sum_{i=1}^{10} I_i$, where $I_i = I_{\{i\text{th roll is } 6\}}$, and $Y = \sum_{i=1}^{10} J_i$, where $J_i = I_{\{i\text{th roll is } 5\}}$. Then, $EX = EY = \frac{10}{6} = \frac{5}{3}$. Moreover, $E(I_i J_j)$ equals 0 if $i = j$ (because both 5 and 6 cannot be rolled on the same roll), and $E(I_i J_j) = \frac{1}{6^2}$ if $i \neq j$ (by independence of different rolls). Therefore,

$$\begin{aligned} EXY &= \sum_{i=1}^{10} \sum_{j=1}^{10} E(I_i J_j) \\ &= \sum_{i \neq j} \frac{1}{6^2} \\ &= \frac{10 \cdot 9}{36} \\ &= \frac{5}{2} \end{aligned}$$

and

$$\mathrm{Cov}(X, Y) = \frac{5}{2} - \left(\frac{5}{3}\right)^2 = -\frac{5}{18}$$

is negative, as should be expected.

Weak Law of Large Numbers

Assume that an experiment is performed in which an event A happens with probability $P(A) = p$. At the beginning of the course, we promised to attach a precise meaning to the statement: "If you repeat the experiment, independently, a large number of times, the proportion of times A happens converges to p." We will now do so and actually prove the more general statement below.

Theorem 8.6. *Weak Law of Large Numbers.*

If $X, X_1, X_2, \ldots$ are independent and identically distributed random variables with finite expectation and variance, then $\frac{X_1+\ldots+X_n}{n}$ converges to EX in the sense that, for any fixed $\epsilon > 0$,

$$P\left(\left|\frac{X_1+\ldots+X_n}{n} - EX\right| \geq \epsilon\right) \to 0,$$

as $n \to \infty$.

In particular, if S_n is the number of successes in n independent trials, each of which is a success with probability p, then, as we have observed before, $S_n = I_1 + \ldots + I_n$, where $I_i = I_{\{\text{success at trial } i\}}$. So, for every $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \to 0,$$

as $n \to \infty$. Thus, the proportion of successes converges to p in this sense.

Theorem 8.7. *Markov Inequality. If $X \geq 0$ is a random variable and $a > 0$, then*

$$P(X \geq a) \leq \frac{1}{a} EX.$$

Example 8.15. If $EX = 1$ and $X \geq 0$, it must be that $P(X \geq 10) \leq 0.1$.

Proof. Here is the crucial observation:

$$I_{\{X \geq a\}} \leq \frac{1}{a} X.$$

Indeed, if $X < a$, the left-hand side is 0 and the right-hand side is nonnegative; if $X \geq a$, the left-hand side is 1 and the right-hand side is at least 1. Taking the expectation of both sides, we get

$$P(X \geq a) = E(I_{\{X \geq a\}}) \leq \frac{1}{a} EX.$$

□

Theorem 8.8. *Chebyshev inequality. If $EX = \mu$ and $\mathrm{Var}(X) = \sigma^2$ are both finite and $k > 0$, then*

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Example 8.16. If $EX = 1$ and $\mathrm{Var}(X) = 1$, $P(X \geq 10) \leq P(|X - 1| \geq 9) \leq \frac{1}{81}$.

Example 8.17. If $EX = 1$, $\mathrm{Var}(X) = 0.1$,

$$P(|X - 1| \geq 0.5) \leq \frac{0.1}{0.5^2} = \frac{2}{5}.$$

As the previous two examples show, the Chebyshev inequality is useful if either σ is small or k is large.

Proof. By the Markov inequality,

$$P(|X-\mu| \geq k) = P((X-\mu)^2 \geq k^2) \leq \frac{1}{k^2} E(X-\mu)^2 = \frac{1}{k^2}\mathrm{Var}(X).$$

□

We are now ready to prove the Weak Law of Large Numbers.

Proof. Denote $\mu = EX$, $\sigma^2 = \mathrm{Var}(X)$, and let $S_n = X_1 + \ldots + X_n$. Then,

$$ES_n = EX_1 + \ldots + EX_n = n\mu$$

and

$$\mathrm{Var}(S_n) = n\sigma^2.$$

Therefore, by the Chebyshev inequality,

$$P(|S_n - n\mu| \geq n\epsilon) \leq \frac{n\sigma^2}{n^2\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0,$$

as $n \to \infty$. □

A careful examination of the proof above will show that it remains valid if ϵ depends on n, but goes to 0 slower than $\frac{1}{\sqrt{n}}$. This suggests that $\frac{X_1+\ldots+X_n}{n}$ converges to EX at the rate of about $\frac{1}{\sqrt{n}}$. We will make this statement more precise below.

Central Limit Theorem

Theorem 8.9. *Central limit theorem.*

Assume that $X, X_1, X_2, \ldots$ are independent, identically distributed random variables, with finite $\mu = EX$ and $\sigma^2 = Var(X)$. Then,

$$P\left(\frac{X_1 + \ldots + X_n - \mu n}{\sigma\sqrt{n}} \leq x\right) \to P(Z \leq x),$$

as $n \to \infty$, where Z is standard Normal.

We will not prove this theorem in full detail, but will later give good indication as to why it holds. Observe, however, that it is a remarkable theorem: the random variables X_i have an arbitrary distribution (with given expectation and variance) and the theorem says that their sum approximates a very particular distribution, the normal one. Adding many independent copies of a random variable erases all information about its distribution other than expectation and variance!

On the other hand, the convergence is not very fast; the current version of the celebrated Berry-Esseen theorem states that an upper bound on the difference between the two probabilities in the Central limit theorem is

$$0.4785 \cdot \frac{E|X-\mu|^3}{\sigma^3\sqrt{n}}.$$

Example 8.18. Assume that X_n are independent and uniform on $[0,1]$. Let $S_n = X_1+\ldots+X_n$. (a) Compute approximately $P(S_{200} \le 90)$. (b) Using the approximation, find n so that $P(S_n \ge 50) \ge 0.99$.

We know that $EX_i = \frac{1}{2}$ and $\text{Var}(X_i) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$.

For (a),

$$\begin{aligned}
P(S_{200} \le 90) &= P\left(\frac{S_{200} - 200\cdot\frac{1}{2}}{\sqrt{200\cdot\frac{1}{12}}} \le \frac{90 - 200\cdot\frac{1}{2}}{\sqrt{200\cdot\frac{1}{12}}}\right) \\
&\approx P(Z \le -\sqrt{6}) \\
&= 1 - P(Z \le \sqrt{6}) \\
&\approx 1 - 0.993 \\
&= 0.007
\end{aligned}$$

For (b), we rewrite

$$P\left(\frac{S_n - n\cdot\frac{1}{2}}{\sqrt{n\cdot\frac{1}{12}}} \ge \frac{50 - n\cdot\frac{1}{2}}{\sqrt{n\cdot\frac{1}{12}}}\right) = 0.99,$$

and then approximate

$$P\left(Z \ge -\frac{\frac{n}{2} - 50}{\sqrt{n\cdot\frac{1}{12}}}\right) = 0.99,$$

or

$$P\left(Z \le \frac{\frac{n}{2} - 50}{\sqrt{n\cdot\frac{1}{12}}}\right) = \Phi\left(\frac{\frac{n}{2} - 50}{\sqrt{n\cdot\frac{1}{12}}}\right) = 0.99.$$

Using the fact that $\Phi(z) = 0.99$ for (approximately) $z = 2.326$, we get that

$$n - 1.345\sqrt{n} - 100 = 0$$

and that $n = 115$.

Example 8.19. A casino charges \$1 for entrance. For promotion, they offer to the first 30,000 "guests" the following game. Roll a fair die:

- if you roll 6, you get free entrance and \$2;

- if you roll 5, you get free entrance;
- otherwise, you pay the normal fee.

Compute the number s so that the revenue loss is at most s with probability 0.9.

In symbols, if L is lost revenue, we need to find s so that

$$P(L \leq s) = 0.9.$$

We have $L = X_1 + \cdots + X_n$, where $n = 30,000$, X_i are independent, and $P(X_i = 0) = \frac{4}{6}$, $P(X_i = 1) = \frac{1}{6}$, and $P(X_i = 3) = \frac{1}{6}$. Therefore,

$$EX_i = \frac{2}{3}$$

and

$$\mathrm{Var}(X_i) = \frac{1}{6} + \frac{9}{6} - \left(\frac{2}{3}\right)^2 = \frac{11}{9}.$$

Therefore,

$$\begin{aligned} P(L \leq s) &= P\left(\frac{L - \frac{2}{3} \cdot n}{\sqrt{n \cdot \frac{11}{9}}} \leq \frac{s - \frac{2}{3} \cdot n}{\sqrt{n \cdot \frac{11}{9}}}\right) \\ &\approx P\left(Z \leq \frac{s - \frac{2}{3} \cdot n}{\sqrt{n \cdot \frac{11}{9}}}\right) = 0.9, \end{aligned}$$

which gives

$$\frac{s - \frac{2}{3} \cdot n}{\sqrt{n \cdot \frac{11}{9}}} \approx 1.28,$$

and finally,

$$s \approx \frac{2}{3}n + 1.28\sqrt{n \cdot \frac{11}{9}} \approx 20,245.$$

Problems

1. An urn contains 2 white and 4 black balls. Select three balls in three successive steps without replacement. Let X be the total number of white balls selected and Y the step in which you selected the first black ball. For example, if the selected balls are white, black, black, then $X = 1$, $Y = 2$. Compute $E(XY)$.

2. The joint density of (X, Y) is given by

$$\begin{cases} f(x,y) = 3x & \text{if } 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Compute $\text{Cov}(X, Y)$.

3. Five married couples are seated at random in a row of 10 seats.

(a) Compute the expected number of women that sit next to their husbands.

(b) Compute the expected number of women that sit next to at least one man.

4. There are 20 birds that sit in a row on a wire. Each bird looks left or right with equal probability. Let N be the number of birds not seen by any neighboring bird. Compute EN.

5. Recall that a full deck of cards contains 52 cards, 13 cards of each of the four suits. Distribute the cards at random to 13 players, so that each gets 4 cards. Let N be the number of players whose four cards are of the same suit. Using the indicator trick, compute EN.

6. Roll a fair die 24 times. Compute, using a relevant approximation, the probability that the sum of numbers exceeds 100.

Solutions to problems

1. First we determine the joint p. m. f. of (X, Y). We have

$$\begin{aligned}
P(X = 0, Y = 1) &= P(\text{bbb}) = \frac{1}{5}, \\
P(X = 1, Y = 1) &= P(\text{bwb or bbw}) = \frac{2}{5}, \\
P(X = 1, Y = 2) &= P(\text{wbb}) = \frac{1}{5}, \\
P(X = 2, Y = 1) &= P(\text{bww}) = \frac{1}{15}, \\
P(X = 2, Y = 2) &= P(\text{wbw}) = \frac{1}{15}, \\
P(X = 2, Y = 3) &= P(\text{wwb}) = \frac{1}{15},
\end{aligned}$$

so that

$$E(XY) = 1 \cdot \frac{2}{5} + 2 \cdot \frac{1}{5} + 2 \cdot \frac{1}{15} + 4 \cdot \frac{1}{15} + 6 \cdot \frac{1}{15} = \frac{8}{5}.$$

2. We have

$$EX = \int_0^1 dx \int_0^x x \cdot 3x \, dy = \int_0^1 3x^3 \, dx = \frac{3}{4},$$
$$EX = \int_0^1 dx \int_0^x y \cdot 3x \, dy = \int_0^1 \frac{3}{2} x^3 \, dx = \frac{3}{8},$$
$$E(XY) = \int_0^1 dx \int_0^x xy \cdot 3x \, dy = \int_0^1 \frac{3}{2} x^4 \, dx = \frac{3}{10},$$

so that $\text{Cov}(X, Y) = \frac{3}{10} - \frac{3}{4} \cdot \frac{3}{8} = \frac{3}{160}$.

3. (a) Let the number be M. Let $I_i = I_{\{\text{couple } i \text{ sits together}\}}$. Then

$$EI_i = \frac{9!\,2!}{10!} = \frac{1}{5},$$

and so

$$EM = EI_1 + \ldots + EI_5 = 5\,EI_1 = 1.$$

(b) Let the number be N. Let $I_i = I_{\{\text{woman } i \text{ sits next to a man}\}}$. Then, by dividing into cases, whereby the woman either sits on one of the two end chairs or on one of the eight middle chairs,

$$EI_i = \frac{2}{10} \cdot \frac{5}{9} + \frac{8}{10} \cdot \left(1 - \frac{4}{9} \cdot \frac{3}{8}\right) = \frac{7}{9},$$

and so

$$EM = EI_1 + \ldots + EI_5 = 5\,EI_1 = \frac{35}{9}.$$

4. For the two birds at either end, the probability that it is not seen is $\frac{1}{2}$, while for any other bird this probability is $\frac{1}{4}$. By the indicator trick

$$EN = 2 \cdot \frac{1}{2} + 18 \cdot \frac{1}{4} = \frac{11}{2}.$$

5. Let

$$I_i = I_{\{\text{player } i \text{ has four cards of the same suit}\}},$$

so that $N = I_1 + ... + I_{13}$. Observe that:

- the number of ways to select 4 cards from a 52 card deck is $\binom{52}{4}$;
- the number of choices of a suit is 4; and
- after choosing a suit, the number of ways to select 4 cards of that suit is $\binom{13}{4}$.

Therefore, for all i,

$$EI_i = \frac{4 \cdot \binom{13}{4}}{\binom{52}{4}}$$

and

$$EN = \frac{4 \cdot \binom{13}{4}}{\binom{52}{4}} \cdot 13.$$

6. Let X_1, $X_2, \ldots$ be the numbers on successive rolls and $S_n = X_1 + \ldots + X_n$ the sum. We know that $EX_i = \frac{7}{2}$, and $\mathrm{Var}(X_i) = \frac{35}{12}$. So, we have

$$P(S_{24} \geq 100) = P\left(\frac{S_{24} - 24 \cdot \frac{7}{2}}{\sqrt{24 \cdot \frac{35}{12}}} \geq \frac{100 - 24 \cdot \frac{7}{2}}{\sqrt{24 \cdot \frac{35}{12}}}\right) \approx P(Z \geq 1.85) = 1 - \Phi(1.85) \approx 0.032.$$

Interlude: Practice Final

This practice exam covers the material from chapters 1 through 8. Give yourself 120 minutes to solve the six problems, which you may assume have equal point score.

1. Recall that a full deck of cards contains 13 cards of each of the four suits ($\clubsuit, \diamondsuit, \heartsuit, \spadesuit$). Select cards from the deck at random, one by one, without replacement.

(a) Compute the probability that the first four cards selected are all hearts ($\heartsuit$).

(b) Compute the probability that all suits are represented among the first four cards selected.

(c) Compute the expected number of different suits among the first four cards selected.

(d) Compute the expected number of cards you have to select to get the first hearts card.

2. Eleven Scandinavians: 2 Swedes, 4 Norwegians, and 5 Finns are seated in a *row* of 11 chairs at random.

(a) Compute the probability that *all* groups sit together (i.e., the Swedes occupy adjacent seats, as do the Norwegians and Finns).

(b) Compute the probability that *at least one* of the groups sits together.

(c) Compute the probability that the two Swedes have exactly one person sitting between them.

3. You have two fair coins. Toss the first coin three times and let X be the number of Heads. Then toss the second coin X times, that is, as many times as the number of Heads in the first coin toss. Let Y be the number of Heads in the second coin toss. (For example, if $X = 0$, Y is automatically 0; if $X = 2$, toss the second coin twice and count the number of Heads to get Y.)

(a) Determine the joint probability mass function of X and Y, that is, write down a formula for $P(X = i, Y = j)$ for all relevant i and j.

(b) Compute $P(X \geq 2 \,|\, Y = 1)$.

4. Assume that $2,000,000$ single male tourists visit Las Vegas every year. Assume also that each of these tourists, independently, gets married while drunk with probability $1/1,000,000$.

(a) Write down the *exact* probability that exactly 3 male tourists will get married while drunk next year.

(b) Compute the expected number of such drunk marriages in the next 10 years.

(c) Write down a relevant *approximate* expression for the probability in (a).

(d) Write down an approximate expression for the probability that there will be no such drunk marriage during at least one of the next 3 years.

5. Toss a fair coin twice. You *win* \$2 if both tosses comes out Heads, *lose* \$1 if no toss comes out Heads, and win or lose nothing otherwise.

(a) What is the expected number of games you need to play to win once?

(b) Assume that you play this game 500 times. What is, approximately, the probability that you win at least \$135?

(c) Again, assume that you play this game 500 times. Compute (approximately) the amount of money x such that your winnings will be at least x with probability 0.5. Then, do the same with probability 0.9.

6. Two random variables X and Y are independent and have the same probability density function

$$g(x) = \begin{cases} c(1+x) & x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of c. Here and in (b): use $\int_0^1 x^n\,dx = \frac{1}{n+1}$, for $n > -1$.

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

(c) Find $P(X+Y<1)$ and $P(X+Y\le 1)$. Here and in (d): when you get to a *single* integral involving powers, stop.

(d) Find $E|X-Y|$.

Solutions to Practice Final

1. Recall that a full deck of cards contains 13 cards of each of the four suits ($\clubsuit, \diamondsuit, \heartsuit, \spadesuit$). Select cards from the deck at random, one by one, without replacement.

   (a) Compute the probability that the first four cards selected are all hearts ($\heartsuit$).

   Solution:

   $$\frac{\binom{13}{4}}{\binom{52}{4}}$$

   (b) Compute the probability that all suits are represented among the first four cards selected.

   Solution:

   $$\frac{13^4}{\binom{52}{4}}$$

   (c) Compute the expected number of different suits among the first four cards selected.

   Solution:
   If X is the number of suits represented, then $X = I_\heartsuit + I_\diamondsuit + I_\clubsuit + I_\spadesuit$, where $I_\heartsuit = I_{\{\heartsuit \text{ is represented}\}}$, etc. Then,

   $$EI_\heartsuit = 1 - \frac{\binom{39}{4}}{\binom{52}{4}},$$

   which is the same for the other three indicators, so

   $$EX = 4\, EI_\heartsuit = 4\left(1 - \frac{\binom{39}{4}}{\binom{52}{4}}\right).$$

   (d) Compute the expected number of cards you have to select to get the first hearts card.

   Solution:
   Label non-$\heartsuit$ cards $1, \ldots, 39$ and let $I_i = I_{\{\text{card } i \text{ selected before any } \heartsuit \text{ card}\}}$. Then, $EI_i = \frac{1}{14}$ for any i. If N is the number of cards you have to select to get the first hearts card, then

   $$EN = E\left(I_1 + \cdots + EI_{39}\right) = \frac{39}{14}.$$

2. Eleven Scandinavians: 2 Swedes, 4 Norwegians, and 5 Finns are seated in a *row* of 11 chairs at random.

   (a) Compute the probability that *all* groups sit together (i.e., the Swedes occupy adjacent seats, as do the Norwegians and Finns).

   Solution:

$$\frac{2!\,4!\,5!\,3!}{11!}$$

   (b) Compute the probability that *at least one* of the groups sits together.

   Solution:
   Define $A_S = \{\text{Swedes sit together}\}$ and, similarly, A_N and A_F. Then,

$$\begin{aligned}
&P(A_S \cup A_N \cup A_F) \\
&= P(A_S) + P(A_N) + P(A_F) \\
&\quad - P(A_S \cap A_N) - P(A_S \cap A_F) - P(A_N \cap A_F) \\
&\quad + P(A_S \cap A_N \cap A_F) \\
&= \frac{2!\,10! + 4!\,8! + 5!\,7! - 2!\,4!\,7! - 2!\,5!\,6! - 4!\,5!\,3! + 2!\,4!\,5!\,3!}{11!}
\end{aligned}$$

   (c) Compute the probability that the two Swedes have exactly one person sitting between them.

   Solution:
   The two Swedes may occupy chairs $1, 3$; or $2, 4$; or $3, 5$; ...; or $9, 11$. There are exactly 9 possibilities, so the answer is

$$\frac{9}{\binom{11}{2}} = \frac{9}{55}.$$

3. You have two fair coins. Toss the first coin three times and let X be the number of Heads. Then, toss the second coin X times, that is, as many times as you got Heads in the first coin toss. Let Y be the number of Heads in the second coin toss. (For example, if $X = 0$, Y is automatically 0; if $X = 2$, toss the second coin twice and count the number of Heads to get Y.)

(a) Determine the joint probability mass function of X and Y, that is, write down a formula for $P(X = i, Y = j)$ for all relevant i and j.

Solution:

$$P(X = i, Y = j) = P(X = i)P(Y = j|X = i) = \binom{3}{i}\frac{1}{2^3} \cdot \binom{i}{j}\frac{1}{2^i},$$

for $0 \leq j \leq i \leq 3$.

(b) Compute $P(X \geq 2 \,|\, Y = 1)$.

Solution:
This equals

$$\frac{P(X = 2, Y = 1) + P(X = 3, Y = 1)}{P(X = 1, Y = 1) + P(X = 2, Y = 1) + P(X = 3, Y = 1)} = \frac{5}{9}.$$

4. Assume that $2,000,000$ single male tourists visit Las Vegas every year. Assume also that each of these tourists independently gets married while drunk with probability $1/1,000,000$.

(a) Write down the *exact* probability that exactly 3 male tourists will get married while drunk next year.

Solution:
With X equal to the number of such drunk marriages, X is Binomial(n, p) with $p = 1/1,000,000$ and $n = 2,000,000$, so we have

$$P(X = 3) = \binom{n}{3}p^3(1-p)^{n-3}$$

(b) Compute the expected number of such drunk marriages in the next 10 years.

Solution:
As X is binomial, its expected value is $np = 2$, so the answer is $10\,EX = 20$.

(c) Write down a relevant *approximate* expression for the probability in (a).

Solution:
We use that X is approximately Poisson with $\lambda = 2$, so the answer is

$$\frac{\lambda^3}{3!}\,e^{-\lambda} = \frac{4}{3}\,e^{-2}.$$

(d) Write down an approximate expression for the probability that there will be no such drunk marriage during at least one of the next 3 years.

Solution:
This equals $1-P(\text{at least one such marriage in each of the next 3 years})$, which equals

$$1-\left(1-e^{-2}\right)^3.$$

5. Toss a fair coin twice. You *win* \$2 if both tosses comes out Heads, *lose* \$1 if no toss comes out Heads, and win or lose nothing otherwise.

(a) What is the expected number of games you need to play to win once?

Solution:
The probability of winning in $\frac{1}{4}$. The answer, the expectation of a Geometric$\left(\frac{1}{4}\right)$ random variable, is 4.

(b) Assume that you play this game 500 times. What is, approximately, the probability that you win at least \$135?

Solution:
Let X be the winnings in one game and $X_1, X_2, \ldots, X_n$ the winnings in successive games, with $S_n = X_1 + \ldots + X_n$. Then, we have

$$EX = 2\cdot\frac{1}{4} - 1\cdot\frac{1}{4} = \frac{1}{4}$$

and

$$\mathrm{Var}(X) = 4\cdot\frac{1}{4} + 1\cdot\frac{1}{4} - \left(\frac{1}{4}\right)^2 = \frac{19}{16}.$$

Thus,

$$P(S_n \geq 135) = P\left(\frac{S_n - n\cdot\frac{1}{4}}{\sqrt{n\cdot\frac{19}{16}}} \geq \frac{135 - n\cdot\frac{1}{4}}{\sqrt{n\cdot\frac{19}{16}}}\right) \approx P\left(Z \geq \frac{135 - n\cdot\frac{1}{4}}{\sqrt{n\cdot\frac{19}{16}}}\right),$$

where Z is standard Normal. Using $n = 500$, we get the answer

$$1-\Phi\left(\frac{10}{\sqrt{500\cdot\frac{19}{16}}}\right).$$

(c) Again, assume that you play this game 500 times. Compute (approximately) the amount of money x such that your winnings will be at least x with probability 0.5. Then, do the same with probability 0.9.

Solution:

For probability 0.5, the answer is exactly $ES_n = n \cdot \frac{1}{4} = 125$. For probability 0.9, we approximate

$$P(S_n \geq x) \approx P\left(Z \geq \frac{x - 125}{\sqrt{500 \cdot \frac{19}{16}}}\right) = P\left(Z \leq \frac{125 - x}{\sqrt{500 \cdot \frac{19}{16}}}\right) = \Phi\left(\frac{125 - x}{\sqrt{500 \cdot \frac{19}{16}}}\right),$$

where we have used that $x < 125$. Then, we use that $\Phi(z) = 0.9$ at $z \approx 1.28$, leading to the equation

$$\frac{125 - x}{\sqrt{500 \cdot \frac{19}{16}}} = 1.28$$

and, therefore,

$$x = 125 - 1.28 \cdot \sqrt{500 \cdot \frac{19}{16}}.$$

(This gives $x = 93$.)

6. Two random variables X and Y are independent and have the same probability density function

$$g(x) = \begin{cases} c(1 + x) & x \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of c. Here and in (b): use $\int_0^1 x^n \, dx = \frac{1}{n+1}$, for $n > -1$.

Solution:
As

$$1 = c \cdot \int_0^1 (1 + x) \, dx = c \cdot \frac{3}{2},$$

we have $c = \frac{2}{3}$

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

Solution:

By independence, this equals $2\mathrm{Var}(X) = 2(E(X^2) - (EX)^2)$. Moreover,

$$E(X) = \frac{2}{3}\int_0^1 x(1+x)\,dx = \frac{5}{9},$$
$$E(X^2) = \frac{2}{3}\int_0^1 x^2(1+x)\,dx = \frac{7}{18},$$

and the answer is $\frac{13}{81}$.

(c) Find $P(X+Y<1)$ and $P(X+Y\le 1)$. Here and in (d): when you get to a *single* integral involving powers, stop.

Solution:
The two probabilities are both equal to

$$\left(\frac{2}{3}\right)^2 \int_0^1 dx \int_0^{1-x} (1+x)(1+y)\,dy =$$
$$\left(\frac{2}{3}\right)^2 \int_0^1 (1+x)\left((1-x) + \frac{(1-x)^2}{2}\right)\,dx.$$

(d) Find $E|X-Y|$.

Solution:
This equals

$$\left(\frac{2}{3}\right)^2 \cdot 2\int_0^1 dx \int_0^x (x-y)(1+x)(1+y)\,dy =$$
$$\left(\frac{2}{3}\right)^2 \cdot 2\int_0^1 \left[x(1+x)\left(x+\frac{x^2}{2}\right) - (1+x)\left(\frac{x^2}{2}+\frac{x^3}{3}\right)\right]\,dx.$$

9 Convergence in probability

One of the goals of probability theory is to extricate a useful deterministic quantity out of a random situation. This is typically possible when a large number of random effects cancel each other out, so some limit is involved. In this chapter we consider the following setting: given a sequence of random variables, $Y_1, Y_2, \ldots$, we want to show that, when n is large, Y_n is approximately $f(n)$, for some simple deterministic function $f(n)$. The meaning of "approximately" is what we now make clear.

A sequence $Y_1, Y_2, \ldots$ of random variables converges to a number a *in probability* if, as $n \to \infty$, $P(|Y_n - a| \le \epsilon)$ converges to 1, for any fixed $\epsilon > 0$. This is equivalent to $P(|Y_n - a| > \epsilon) \to 0$ as $n \to \infty$, for any fixed $\epsilon > 0$.

Example 9.1. Toss a fair coin n times, independently. Let R_n be the "longest run of Heads," i.e., the longest sequence of consecutive tosses of Heads. For example, if $n = 15$ and the tosses come out

HHTTHHHTHTHTHHH,

then $R_n = 3$. We will show that, as $n \to \infty$,

$$\frac{R_n}{\log_2 n} \to 1,$$

in probability. This means that, to a first approximation, one should expect about 20 consecutive Heads somewhere in a million tosses.

To solve a problem such as this, we need to find upper bounds for probabilities that R_n is large and that it is small, i.e., for $P(R_n \ge k)$ and $P(R_n \le k)$, for appropriately chosen k. Now, for arbitrary k,

$$\begin{aligned} P(R_n \ge k) &= P\left(k \text{ consecutive Heads start at some } i,\, 0 \le i \le n-k+1\right) \\ &= P\Big(\bigcup_{i=1}^{n-k+1} \{i \text{ is the first Heads in a succession of at least } k \text{ Heads}\}\Big) \\ &\le n \cdot \frac{1}{2^k}. \end{aligned}$$

For the lower bound, divide the string of size n into disjoint blocks of size k. There are $\lfloor \frac{n}{k} \rfloor$ such blocks (if n is not divisible by k, simply throw away the leftover smaller block at the end). Then, $R_n \ge k$ as soon as one of the blocks consists of Heads only; different blocks are independent. Therefore,

$$P(R_n < k) \le \left(1 - \frac{1}{2^k}\right)^{\lfloor \frac{n}{k} \rfloor} \le \exp\left(-\frac{1}{2^k}\left\lfloor \frac{n}{k} \right\rfloor\right),$$

using the famous inequality $1 - x \le e^{-x}$, valid for all x.

Below, we will use the following trivial inequalities, valid for any real number $x \ge 2$: $\lfloor x \rfloor \ge x - 1$, $\lceil x \rceil \le x + 1$, $x - 1 \ge \frac{x}{2}$, and $x + 1 \le 2x$.

To demonstrate that $\frac{R_n}{\log_2 n} \to 1$, in probability, we need to show that, for any $\epsilon > 0$,

$$P(R_n \ge (1+\epsilon)\log_2 n) \to 0, \tag{1}$$

$$P(R_n \le (1-\epsilon)\log_2 n) \to 0, \tag{2}$$

as

$$\begin{aligned} P\left(\left|\frac{R_n}{\log_2 n} - 1\right| \ge \epsilon\right) &= P\left(\frac{R_n}{\log_2 n} \ge 1+\epsilon \text{ or } \frac{R_n}{\log_2 n} \le 1-\epsilon\right) \\ &= P\left(\frac{R_n}{\log_2 n} \ge 1+\epsilon\right) + P\left(\frac{R_n}{\log_2 n} \le 1-\epsilon\right) \\ &= P\left(R_n \ge (1+\epsilon)\log_2 n\right) + P\left(R_n \le (1-\epsilon)\log_2 n\right). \end{aligned}$$

A little fussing in the proof comes from the fact that $(1 \pm \epsilon)\log_2 n$ are not integers. This is common in such problems. To prove (1), we plug $k = \lfloor (1+\epsilon)\log_2 n \rfloor$ into the upper bound to get

$$\begin{aligned} P(R_n \ge (1+\epsilon)\log_2 n) &\le n \cdot \frac{1}{2^{(1+\epsilon)\log_2 n - 1}} \\ &= n \cdot \frac{2}{n^{1+\epsilon}} \\ &= \frac{2}{n^\epsilon} \to 0, \end{aligned}$$

as $n \to \infty$. On the other hand, to prove (2) we need to plug $k = \lceil (1-\epsilon)\log_2 n \rceil + 1$ into the lower bound,

$$\begin{aligned} P(R_n \le (1-\epsilon)\log_2 n) &\le P(R_n < k) \\ &\le \exp\left(-\frac{1}{2^k}\left\lfloor \frac{n}{k} \right\rfloor\right) \\ &\le \exp\left(-\frac{1}{2^k}\left(\frac{n}{k} - 1\right)\right) \\ &\le \exp\left(-\frac{1}{32} \cdot \frac{1}{n^{1-\epsilon}} \cdot \frac{n}{(1-\epsilon)\log_2 n}\right) \\ &= \exp\left(-\frac{1}{32}\,\frac{n^\epsilon}{(1-\epsilon)\log_2 n}\right) \\ &\to 0, \end{aligned}$$

as $n \to \infty$, as n^ϵ is much larger than $\log_2 n$.

The most basic tool in proving convergence in probability is the *Chebyshev inequality*: if X is a random variable with $EX = \mu$ and $\mathrm{Var}(X) = \sigma^2$, then

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2},$$

for any $k > 0$. We proved this inequality in the previous chapter and we will use it to prove the next theorem.

Theorem 9.1. *Connection between variance and convergence in probability.*

> *Assume that Y_n are random variables and that a is a constant such that*
> $$EY_n \to a,$$
> $$\mathrm{Var}(Y_n) \to 0,$$
> *as $n \to \infty$. Then,*
> $$Y_n \to a,$$
> *as $n \to \infty$, in probability.*

Proof. Fix an $\epsilon > 0$. If n is so large that

$$|EY_n - a| < \epsilon/2,$$

then

$$\begin{aligned} P(|Y_n - a| > \epsilon) &\leq P(|Y_n - EY_n| > \epsilon/2) \\ &\leq 4\,\frac{\mathrm{Var}(Y_n)}{\epsilon^2} \\ &\to 0, \end{aligned}$$

as $n \to \infty$. Note that the second inequality in the computation is the Chebyshev inequality. □

This is most often applied to sums of random variables. Let

$$S_n = X_1 + \ldots + X_n,$$

where X_i are random variables with finite expectation and variance. Then, *without any independence assumption*,

$$ES_n = EX_1 + \ldots + EX_n$$

and

$$E(S_n^2) = \sum_{i=1}^{n} EX_i^2 + \sum_{i \neq j} E(X_i X_j),$$

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j).$$

Recall that

$$\mathrm{Cov}(X_1, X_j) = E(X_i X_j) - EX_i EX_j$$

and

$$\mathrm{Var}(aX) = a^2 \mathrm{Var}(X).$$

Moreover, if X_i are independent,

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n).$$

Continuing with the review, let us reformulate and prove again the most famous convergence in probability theorem. We will use the common abbreviation i. i. d. for independent identically distributed random variables.

Theorem 9.2. *Weak law of large numbers. Let $X, X_1, X_2, \ldots$ be i. i. d. random variables with with $EX = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. Let $S_n = X_1 + \ldots + X_n$. Then, as $n \to \infty$,*

$$\frac{S_n}{n} \to \mu$$

in probability.

Proof. Let $Y_n = \frac{S_n}{n}$. We have $EY_n = \mu$ and

$$\mathrm{Var}(Y_n) = \frac{1}{n^2}\,\mathrm{Var}(S_n) = \frac{1}{n^2}\, n\, \sigma^2 = \frac{\sigma^2}{n}.$$

Thus, we can simply apply the previous theorem. □

Example 9.2. We analyze a typical "investment" (the accepted euphemism for gambling on financial markets) problem. Assume that you have two investment choices at the beginning of each year:

- a risk-free "bond" which returns 6% per year; and
- a risky "stock" which increases your investment by 50% with probability 0.8 and wipes it out with probability 0.2.

Putting an amount s in the bond, then, gives you $1.06s$ after a year. The same amount in the stock gives you $1.5s$ with probability 0.8 and 0 with probability 0.2; note that the expected value is $0.8 \cdot 1.5s = 1.2s > 1.06s$. We will assume year-to-year independence of the stock's return.

We will try to maximize the return to our investment by "hedging." That is, we invest, at the beginning of each year, a fixed proportion x of our current capital into the stock and the remaining proportion $1 - x$ into the bond. We collect the resulting capital at the end of the year, which is simultaneously the beginning of next year, and reinvest with the same proportion x. Assume that our initial capital is x_0.

It is important to note that the *expected value* of the capital at the end of the year is maximized when $x = 1$, but by using this strategy you will eventually *lose everything*. Let X_n be your capital at the end of year n. Define the *average growth rate* of your investment as

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log \frac{X_n}{x_0},$$

so that

$$X_n \approx x_0 e^{\lambda n}.$$

We will express λ in terms of x; in particular, we will show that it is a nonrandom quantity.

Let $I_i = I_{\{\text{stock goes up in year } i\}}$. These are independent indicators with $EI_i = 0.8$.

$$\begin{aligned} X_n &= X_{n-1}(1-x)\cdot 1.06 + X_{n-1}\cdot x \cdot 1.5 \cdot I_n \\ &= X_{n-1}(1.06(1-x) + 1.5x \cdot I_n) \end{aligned}$$

and so we can unroll the recurrence to get

$$X_n = x_0(1.06(1-x) + 1.5x)^{S_n}((1-x)1.06)^{n-S_n},$$

where $S_n = I_1 + \ldots + I_n$. Therefore,

$$\begin{aligned} \frac{1}{n}\log\frac{X_n}{x_0} &= \frac{S_n}{n}\log(1.06+0.44x) + \left(1 - \frac{S_n}{n}\right)\log(1.06(1-x)) \\ &\to 0.8\log(1.06+0.44x) + 0.2\log(1.06(1-x)), \end{aligned}$$

in probability, as $n \to \infty$. The last expression defines λ as a function of x. To maximize this, we set $\frac{d\lambda}{dx} = 0$ to get

$$\frac{0.8\cdot 0.44}{1.06+0.44x} = \frac{0.2}{1-x}.$$

The solution is $x = \frac{7}{22}$, which gives $\lambda \approx 8.1\%$.

Example 9.3. Distribute n balls independently at random into n boxes. Let N_n be the number of empty boxes. Show that $\frac{1}{n} N_n$ converges in probability and identify the limit.

Note that

$$N_n = I_1 + \ldots + I_n,$$

where $I_i = I_{\{i\text{th box is empty}\}}$, but we cannot use the weak law of large numbers as I_i are not independent. Nevertheless,

$$EI_i = \left(\frac{n-1}{n}\right)^n = \left(1 - \frac{1}{n}\right)^n,$$

and so

$$EN_n = n \cdot \left(1 - \frac{1}{n}\right)^n.$$

Moreover,

$$E(N_n^2) = EN_n + \sum_{i \neq j} E(I_i I_j)$$

with

$$E(I_i I_j) = P(\text{box } i \text{ and } j \text{ are both empty}) = \left(\frac{n-2}{n}\right)^n,$$

so that

$$\mathrm{Var}(N_n) = E(N_n^2) - (EN_n)^2 = n\left(1 - \frac{1}{n}\right)^n + n(n-1)\left(1 - \frac{2}{n}\right)^n - n^2\left(1 - \frac{1}{n}\right)^{2n}.$$

Now, let $Y_n = \frac{1}{n} N_n$. We have

$$EY_n \to e^{-1},$$

as $n \to \infty$, and

$$\begin{aligned} \mathrm{Var}(Y_n) &= \frac{1}{n}\left(1 - \frac{1}{n}\right)^n + \frac{n-1}{n}\left(1 - \frac{2}{n}\right)^n - \left(1 - \frac{1}{n}\right)^{2n} \\ &\to 0 + e^{-2} - e^{-2} = 0, \end{aligned}$$

as $n \to \infty$. Therefore,

$$Y_n = \frac{N_n}{n} \to e^{-1},$$

as $n \to \infty$, in probability.

Problems

1. Assume that n married couples (amounting to $2n$ people) are seated at random on $2n$ seats around a round table. Let T be the number of couples that sit together. Determine ET and $\mathrm{Var}(T)$.

2. There are n birds that sit in a row on a wire. Each bird looks left or right with equal probability. Let N be the number of birds not seen by any neighboring bird. Determine, with proof, the constant c so that, as $n \to \infty$, $\frac{1}{n} N \to c$ in probability.

3. Recall the coupon collector problem: sample from n cards, with replacement, indefinitely, and let N be the number of cards you need to get so that each of n different cards are represented. Find a sequence a_n so that, as $n \to \infty$, N/a_n converges to 1 in probability.

4. Kings and Lakers are playing a "best of seven" playoff series, which means they play until one team wins four games. Assume Kings win every game independently with probability p.

(There is no difference between home and away games.) Let N be the number of games played. Compute EN and $\mathrm{Var}(N)$.

5. An urn contains n red and m black balls. Select balls from the urn one by one without replacement. Let X be the number of red balls selected before any black ball, and let Y be the number of red balls between the first and the second black one. Compute EX and EY.

Solutions to problems

1. Let I_i be the indicator of the event that the ith couple sits together. Then, $T = I_1 + \cdots + I_n$. Moreover,

$$EI_i = \frac{2}{2n-1}, \quad E(I_iI_j) = \frac{2^2(2n-3)!}{(2n-1)!} = \frac{4}{(2n-1)(2n-2)},$$

for any i and $j \neq i$. Thus,

$$ET = \frac{2n}{2n-1}$$

and

$$E(T^2) = ET + n(n-1)\,\frac{4}{(2n-1)(2n-2)} = \frac{4n}{2n-1},$$

so

$$\mathrm{Var}(T) = \frac{4n}{2n-1} - \frac{4n^2}{(2n-1)^2} = \frac{4n(n-1)}{(2n-1)^2}.$$

2. Let I_i indicate the event that bird i is not seen by any other bird. Then, EI_i is $\frac{1}{2}$ if $i = 1$ or $i = n$ and $\frac{1}{4}$ otherwise. It follows that

$$EN = 1 + \frac{n-2}{4} = \frac{n+2}{4}.$$

Furthermore, I_i and I_j are independent if $|i-j| \geq 3$ (two birds that have two or more birds between them are observed independently). Thus, $\mathrm{Cov}(I_i, I_j) = 0$ if $|i-j| \geq 3$. As I_i and I_j are indicators, $\mathrm{Cov}(I_i, I_j) \leq 1$ for any i and j. For the same reason, $\mathrm{Var}(I_i) \leq 1$. Therefore,

$$\mathrm{Var}(N) = \sum_i \mathrm{Var}(I_i) + \sum_{i \neq j} \mathrm{Cov}(I_i, I_j) \leq n + 4n = 5n.$$

Clearly, if $M = \frac{1}{n}N$, then $EM = \frac{1}{n}EN \to \frac{1}{4}$ and $\mathrm{Var}(M) = \frac{1}{n^2}\mathrm{Var}(N) \to 0$. It follows that $c = \frac{1}{4}$.

3. Let N_i be the number of coupons needed to get i different coupons after having $i-1$ different ones. Then $N = N_1 + \ldots + N_n$, and N_i are independent Geometric with success probability

$\frac{n-i+1}{n}$. So,

$$EN_i = \frac{n}{n-i+1}, \quad \mathrm{Var}(N_i) = \frac{n(i-1)}{(n-i+1)^2},$$

and, therefore,

$$EN = n\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right),$$

$$\mathrm{Var}(N) = \sum_{i=1}^{n} \frac{n(i-1)}{(n-i+1)^2} \le n^2 \left(1 + \frac{1}{2^2} + \ldots + \frac{1}{n^2}\right) \le n^2 \frac{\pi^2}{6} < 2n^2.$$

If $a_n = n \log n$, then

$$\frac{1}{a_n} EN \to 1, \quad \frac{1}{a_n^2} EN \to 0,$$

as $n \to \infty$, so that

$$\frac{1}{a_n} N \to 1$$

in probability.

4. Let I_i be the indicator of the event that the ith game is played. Then, $EI_1 = EI_2 = EI_3 = EI_4 = 1$,

$$EI_5 = 1 - p^4 - (1-p)^4,$$
$$EI_6 = 1 - p^5 - 5p^4(1-p) - 5p(1-p)^4 - (1-p)^5,$$
$$EI_7 = \binom{6}{3} p^3(1-p)^3.$$

Add the seven expectations to get EN. To compute $E(N^2)$, we use the fact that $I_iI_j = I_i$ if $i > j$, so that $E(I_iI_j) = EI_i$. So,

$$EN^2 = \sum_i EI_i + 2\sum_{i>j} E(I_iI_j) = \sum_i EI_i + 2\sum_i (i-1)EI_i = \sum_{i=1}^{7} (2i-1)EI_i,$$

and the final result can be obtained by plugging in EI_i and by the standard formula

$$\mathrm{Var}(N) = E(N^2) - (EN)^2.$$

5. Imagine the balls ordered in a row where the ordering specifies the sequence in which they are selected. Let I_i be the indicator of the event that the ith red ball is selected before any black ball. Then, $EI_i = \frac{1}{m+1}$, the probability that in a random ordering of the ith red ball and all m black balls, the red comes first. As $X = I_1 + \ldots + I_n$, $EX = \frac{n}{m+1}$.

Now, let J_i be the indicator of the event that the ith red ball is selected between the first and the second black one. Then, EJ_i is the probability that the red ball is second in the ordering of the above $m + 1$ balls, so $EJ_i = EI_i$, and $EY = EX$.

10 Moment generating functions

If X is a random variable, then its *moment generating function* is

$$\phi(t) = \phi_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} P(X = x) & \text{in the discrete case,} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx & \text{in the continuous case.} \end{cases}$$

Example 10.1. Assume that X is an Exponential(1) random variable, that is,

$$f_X(x) = \begin{cases} e^{-x} & x > 0, \\ 0 & x \le 0. \end{cases}$$

Then,

$$\phi(t) = \int_0^\infty e^{tx} e^{-x}\, dx = \frac{1}{1-t},$$

only when $t < 1$. Otherwise, the integral diverges and the moment generating function does not exist. Have in mind that the moment generating function is meaningful only when the integral (or the sum) converges.

Here is where the name comes from: by writing its Taylor expansion in place of e^{tX} and exchanging the sum and the integral (which can be done in many cases)

$$\begin{aligned} E(e^{tX}) &= E[1 + tX + \frac{1}{2}t^2X^2 + \frac{1}{3!}t^3X^3 + \ldots] \\ &= 1 + tE(X) + \frac{1}{2}t^2E(X^2) + \frac{1}{3!}t^3E(X^3) + \ldots \end{aligned}$$

The expectation of the k-th power of X, $m_k = E(X^k)$, is called the k-th *moment* of x. In combinatorial language, $\phi(t)$ is the exponential generating function of the sequence m_k. Note also that

$$\frac{d}{dt}E(e^{tX})|_{t=0} = EX,$$
$$\frac{d^2}{dt^2}E(e^{tX})|_{t=0} = EX^2,$$

which lets us compute the expectation and variance of a random variable once we know its moment generating function.

Example 10.2. Compute the moment generating function for a Poisson(λ) random variable.

By definition,

$$\begin{aligned}\phi(t) &= \sum_{n=0}^{\infty} e^{tn} \cdot \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(e^t \lambda)^n}{n!} \\ &= e^{-\lambda+\lambda e^t} \\ &= e^{\lambda(e^t-1)}.\end{aligned}$$

Example 10.3. Compute the moment generating function for a standard Normal random variable.

By definition,

$$\begin{aligned}\phi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2}\, dx \\ &= e^{\frac{1}{2}t^2},\end{aligned}$$

where, from the first to the second line, we have used, in the exponent,

$$tx - \frac{1}{2}x^2 = -\frac{1}{2}(-2tx + x^2) = \frac{1}{2}((x-t)^2 - t^2).$$

Lemma 10.1. *If $X_1, X_2, \ldots, X_n$ are independent and $S_n = X_1 + \ldots + X_n$, then*

$$\phi_{S_n}(t) = \phi_{X_1}(t) \ldots \phi_{X_n}(t).$$

If X_i is identically distributed as X, then

$$\phi_{S_n}(t) = (\phi_X(t))^n.$$

Proof. This follows from multiplicativity of expectation for independent random variables:

$$E[e^{tS_n}] = E[e^{tX_1} \cdot e^{tX_2} \cdot \ldots \cdot e^{tX_n}] = E[e^{tX_1}] \cdot E[e^{tX_2}] \cdot \ldots \cdot E[e^{tX_n}].$$

□

Example 10.4. Compute the moment generating function of a Binomial(n, p) random variable.

Here we have $S_n = \sum_{k=1}^{n} I_k$, where the indicators I_k are independent and $I_k = I_{\{\text{success on } k\text{th trial}\}}$, so that

$$\phi_{S_n}(t) = (e^t p + 1 - p)^n.$$

Why are moment generating functions useful? One reason is the computation of *large deviations*. Let $S_n = X_1 + \cdots + X_n$, where X_i are independent and identically distributed as X, with expectation $EX = \mu$ and moment generating function ϕ. At issue is the probability that S_n is far away from its expectation $n\mu$, more precisely, $P(S_n > an)$, where $a > \mu$. We can, of course, use the Chebyshev inequality to get an upper bound of order $\frac{1}{n}$. It turns out that this probability is, for large n, much smaller; the theorem below gives an upper bound that is a much better estimate.

Theorem 10.2. *Large deviation bound.*

> *Assume that $\phi(t)$ is finite for some $t > 0$. For any $a > \mu$,*
>
> $$P(S_n \geq an) \leq \exp(-n\, I(a)),$$
>
> *where*
>
> $$I(a) = \sup\{at - \log\phi(t) : t > 0\} > 0.$$

Proof. For any $t > 0$, using the Markov inequality,

$$P(S_n \geq an) = P(e^{tS_n - tan} \geq 1) \leq E[e^{tS_n - tan}] = e^{-tan}\phi(t)^n = \exp\left(-n(at - \log\phi(t))\right).$$

Note that $t > 0$ is arbitrary, so we can optimize over t to get what the theorem claims. We need to show that $I(a) > 0$ when $a > \mu$. For this, note that $\Phi(t) = at - \log\phi(t)$ satisfies $\Phi(0) = 0$ and, assuming that one can differentiate inside the integral sign (which one can in this case, but proving this requires abstract analysis beyond our scope),

$$\Phi'(t) = a - \frac{\phi'(t)}{\phi(t)} = a - \frac{E(Xe^{tX})}{\phi(t)},$$

and, then,

$$\Phi'(0) = a - \mu > 0,$$

so that $\Phi(t) > 0$ for some small enough positive t. □

Example 10.5. Roll a fair die n times and let S_n be the sum of the numbers you roll. Find an upper bound for the probability that S_n exceeds its expectation by at least n, for $n = 100$ and $n = 1000$.

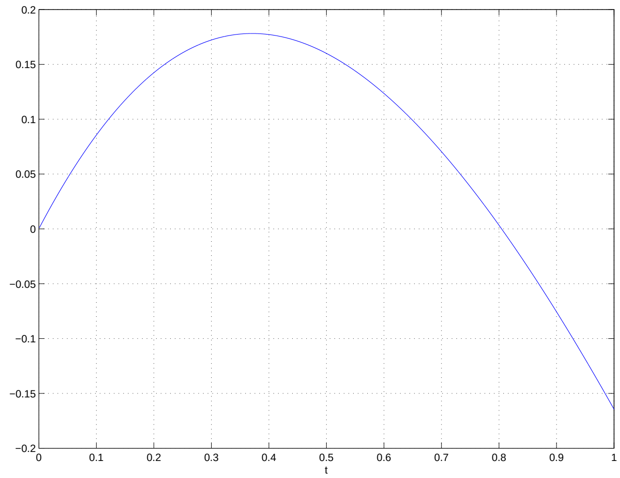
We fit this into the above theorem: observe that $\mu = 3.5$ and $ES_n = 3.5\,n$, and that we need to find an upper bound for $P(S_n \geq 4.5\,n)$, i.e., $a = 4.5$. Moreover,

$$\phi(t) = \frac{1}{6}\sum_{i=1}^{6} e^{it} = \frac{e^t(e^{6t} - 1)}{6(e^t - 1)}$$

and we need to compute $I(4.5)$, which, by definition, is the maximum, over $t > 0$, of the function

$$4.5\,t - \log\phi(t),$$

whose graph is in the figure below.



It would be nice if we could solve this problem by calculus, but unfortunately we cannot (which is very common in such problems), so we resort to numerical calculations. The maximum is at $t \approx 0.37105$ and, as a result, $I(4.5)$ is a little larger than 0.178. This gives the upper bound

$$P(S_n \geq 4.5\,n) \leq e^{-0.178\cdot n},$$

which is about 0.17 for $n = 10$, $1.83 \cdot 10^{-8}$ for $n = 100$, and $4.16 \cdot 10^{-78}$ for $n = 1000$. The bound $\frac{35}{12n}$ for the same probability, obtained by the Chebyshev inequality, is much too large for large n.

Another reason why moment generating functions are useful is that they characterize the distribution and convergence of distributions. We will state the following theorem without proof.

Theorem 10.3. *Assume that the moment generating functions for random variables X, Y, and X_n are finite for all t.*

1. *If $\phi_X(t) = \phi_Y(t)$ for all t, then $P(X \leq x) = P(Y \leq x)$ for all x.*
2. *If $\phi_{X_n}(t) \to \phi_X(t)$ for all t, and $P(X \leq x)$ is continuous in x, then $P(X_n \leq x) \to P(X \leq x)$ for all x.*

Example 10.6. Show that the sum of independent Poisson random variables is Poisson.

Here is the situation. We have n independent random variables $X_1, \ldots, X_n$, such that:

$$\begin{array}{lll}
X_1 & \text{is Poisson}(\lambda_1), & \phi_{X_1}(t) = e^{\lambda_1(e^t-1)}, \\
X_2 & \text{is Poisson}(\lambda_2), & \phi_{X_2}(t) = e^{\lambda_2(e^t-1)}, \\
 & \vdots & \\
X_n & \text{is Poisson}(\lambda_n), & \phi_{X_n}(t) = e^{\lambda_n(e^t-1)}.
\end{array}$$

Therefore,

$$\phi_{X_1+\ldots+X_n}(t) = e^{(\lambda_1+\ldots+\lambda_n)(e^t-1)}$$

and so $X_1 + \ldots + X_n$ is Poisson$(\lambda_1 + \ldots + \lambda_n)$. Similarly, one can also prove that the sum of independent Normal random variables is Normal.

We will now reformulate and prove the Central Limit Theorem in a special case when the moment generating function is finite. This assumption is not needed and the theorem may be applied it as it was in the previous chapter.

Theorem 10.4. *Assume that X is a random variable, with $EX = \mu$ and $\mathrm{Var}(X) = \sigma^2$, and assume that $\phi_X(t)$ is finite for all t. Let $S_n = X_1 + \ldots + X_n$, where $X_1, \ldots, X_n$ are i. i. d. and distributed as X. Let*

$$T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Then, for every x,

$$P(T_n \le x) \to P(Z \le x),$$

as $n \to \infty$, where Z is a standard Normal random variable.

Proof. Let $Y = \frac{X-\mu}{\sigma}$ and $Y_i = \frac{X_i-\mu}{\sigma}$. Then, Y_i are independent, distributed as Y, $E(Y_i) = 0$, $\mathrm{Var}(Y_i) = 1$, and

$$T_n = \frac{Y_1 + \ldots + Y_n}{\sqrt{n}}.$$

To finish the proof, we show that $\phi_{T_n}(t) \to \phi_Z(t) = \exp(t^2/2)$ as $n \to \infty$:

$$\begin{aligned}
\phi_{T_n}(t) &= E\left[e^{t\,T_n}\right] \\
&= E\left[e^{\frac{t}{\sqrt{n}}Y_1+\ldots+\frac{t}{\sqrt{n}}Y_n}\right] \\
&= E\left[e^{\frac{t}{\sqrt{n}}Y_1}\right]\cdots E\left[e^{\frac{t}{\sqrt{n}}Y_n}\right] \\
&= E\left[e^{\frac{t}{\sqrt{n}}Y}\right]^n \\
&= \left(1 + \frac{t}{\sqrt{n}}\,EY + \frac{1}{2}\,\frac{t^2}{n}\,E(Y^2) + \frac{1}{6}\,\frac{t^3}{n^{3/2}}\,E(Y^3) + \ldots\right)^n \\
&= \left(1 + 0 + \frac{1}{2}\,\frac{t^2}{n} + \frac{1}{6}\,\frac{t^3}{n^{3/2}}\,E(Y^3) + \ldots\right)^n \\
&\approx \left(1 + \frac{t^2}{2}\,\frac{1}{n}\right)^n \\
&\to e^{\frac{t^2}{2}}.
\end{aligned}$$

$\square$

Problems

1. A player selects three cards at random from a full deck and collects as many dollars as the number of red cards among the three. Assume 10 people play this game once and let X be the number of their combined winnings. Compute the moment generating function of X.

2. Compute the moment generating function of a uniform random variable on $[0, 1]$.

3. This exercise was in fact the original motivation for the study of large deviations by the Swedish probabilist Harald Cramèr, who was working as an insurance company consultant in the 1930's. Assume that the insurance company receives a steady stream of payments, amounting to (a deterministic number) λ per day. Also, every day, they receive a certain amount in claims; assume this amount is Normal with expectation μ and variance σ^2. Assume also day-to-day independence of the claims. Regulators require that, within a period of n days, the company must be able to cover its claims by the payments received in the same period, *or else*. Intimidated by the fierce regulators, the company wants to fail to satisfy their requirement with probability less than some small number ϵ. The parameters n, μ, σ and ϵ are fixed, but λ is a quantity the company controls. Determine λ.

4. Assume that S is Binomial(n, p). For every $a > p$, determine by calculus the large deviation bound for $P(S \geq an)$.

5. Using the central limit theorem for a sum of Poisson random variables, compute

$$\lim_{n\to\infty} e^{-n} \sum_{i=0}^{n} \frac{n^i}{i!}.$$

Solutions to problems

1. Compute the moment generating function for a single game, then raise it to the 10th power:

$$\phi(t) = \left(\frac{1}{\binom{52}{3}} \left(\binom{26}{3} + \binom{26}{1}\binom{26}{2} \cdot e^t + \binom{26}{2}\binom{26}{1} \cdot e^{2t} + \binom{26}{3} \cdot e^{3t} \right) \right)^{10}.$$

2. Answer: $\phi(t) = \int_0^1 e^{tx}\, dx = \frac{1}{t}(e^t - 1)$.

3. By the assumption, a claim Y is Normal $N(\mu, \sigma^2)$ and, so, $X = (Y - \mu)/\sigma$ is standard normal. Note that $Y = \sigma X + \mu$. Thus, he combined amount of claims is $\sigma(X_1 + \cdots + X_n) + n\mu$, where

X_i are i. i. d. standard Normal, so we need to bound

$$P(X_1 + \cdots + X_n \geq \frac{\lambda - \mu}{\sigma} n) \leq e^{-In}.$$

As $\log \phi(t) = \frac{1}{2}t^2$, we need to maximize, over $t > 0$,

$$\frac{\lambda - \mu}{\sigma} t - \frac{1}{2}t^2,$$

and the maximum equals

$$I = \frac{1}{2}\left(\frac{\lambda - \mu}{\sigma}\right)^2.$$

Finally, we solve the equation

$$e^{-In} = \epsilon,$$

to get

$$\lambda = \mu + \sigma \cdot \sqrt{\frac{-2\log \epsilon}{n}}.$$

4. After a computation, the answer we get is

$$I(a) = a \log \frac{a}{p} + (1-a)\log \frac{1-a}{1-p}.$$

5. Let S_n be the sum of i. i. d. Poisson(1) random variables. Thus, S_n is Poisson(n) and $ES_n = n$. By the Central Limit Theorem, $P(S_n \leq n) \to \frac{1}{2}$, but $P(S_n \leq n)$ is exactly the expression in question. So, the answer is $\frac{1}{2}$.

11 Computing probabilities and expectations by conditioning

Conditioning is the method we encountered before; to remind ourselves, it involves two-stage (or multistage) processes and conditions are appropriate events at the first stage. Recall also the basic definitions:

- *Conditional probability*: if A and B are two events, $P(A|B) = \frac{P(A\cap B)}{P(B)}$;
- *Conditional probability mass function*: if (X,Y) has probability mass function p, $p_X(x|Y = y) = \frac{p(x,y)}{p_Y(y)} = P(X = x|Y = y)$;
- *Conditional density*: if (X,Y) has joint density f, $f_X(x|Y = y) = \frac{f(x,y)}{f_Y(y)}$.
- *Conditional expectation*: $E(X|Y = y)$ is either $\sum_x x p_X(x|Y = y)$ or $\int x f_X(x|Y = y)\,dx$ depending on whether the pair (X,Y) is discrete or continuous.

Bayes' formula also applies to expectation. Assume that the distribution of a random variable X conditioned on $Y = y$ is *given*, and, consequently, its expectation $E(X|Y = y)$ is also known. Such is the case of a two-stage process, whereby the value of Y is chosen at the first stage, which then determines the distribution of X at the second stage. This situation is very common in applications. Then,

$$\boxed{E(X) = \begin{cases} \sum_y E(X|Y = y)P(Y = y) & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} E(X|Y = y) f_Y(y)\,dy & \text{if } Y \text{ is continuous.} \end{cases}}$$

Note that this applies to the probability of an event (which is nothing other than the expectation of its indicator) as well — if we know $P(A|Y = y) = E(I_A|Y = y)$, then we may compute $P(A) = EI_A$ by Bayes' formula above.

Example 11.1. Assume that X, Y are independent Poisson, with $EX = \lambda_1$, $EY = \lambda_2$. Compute the conditional probability mass function of $p_X(x|X + Y = n)$.

Recall that $X + Y$ is Poisson$(\lambda_1 + \lambda_2)$. By definition,

$$\begin{aligned} P(X = k|X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)} \\ &= \frac{P(X = k)P(Y = n - k)}{\frac{(\lambda_1+\lambda_2)^n}{n!}\, e^{-(\lambda_1+\lambda_2)}} \\ &= \frac{\frac{\lambda_1^k}{k!}\, e^{-\lambda_1} \cdot \frac{\lambda_2^{n-k}}{(n-k)!}\, e^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!}\, e^{-(\lambda_1+\lambda_2)}} \\ &= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}. \end{aligned}$$

Therefore, conditioned on $X+Y=n$, X is Binomial$(n, \frac{\lambda_1}{\lambda_1+\lambda_2})$.

Example 11.2. Let T_1, T_2 be two independent Exponential(λ) random variables and let $S_1 = T_1$, $S_2 = T_1 + T_2$. Compute $f_{S_1}(s_1|S_2 = s_2)$.

First,

$$\begin{aligned} P(S_1 \le s_1, S_2 \le s_2) &= P(T_1 \le s_1, T_1 + T_2 \le s_2) \\ &= \int_0^{s_1} dt_1 \int_0^{s_2 - t_1} f_{T_1,T_2}(t_1, t_2)\, dt_2. \end{aligned}$$

If $f = f_{S_1,S_2}$, then

$$\begin{aligned} f(s_1, s_2) &= \frac{\partial^2}{\partial s_1 \partial s_2} P(S_1 \le s_1, S_2 \le s_2) \\ &= \frac{\partial}{\partial s_2} \int_0^{s_2 - s_1} f_{T_1,T_2}(s_1, t_2)\, dt_2 \\ &= f_{T_1,T_2}(s_1, s_2 - s_1) \\ &= f_{T_1}(s_1) f_{T_2}(s_2 - s_1) \\ &= \lambda e^{-\lambda s_1} \lambda\, e^{-\lambda(s_2 - s_1)} \\ &= \lambda^2 e^{-\lambda s_2}. \end{aligned}$$

Therefore,

$$f(s_1, s_2) = \begin{cases} \lambda^2 e^{-\lambda s_2} & \text{if } 0 \le s_1 \le s_2, \\ 0 & \text{otherwise} \end{cases}$$

and, consequently, for $s_2 \ge 0$,

$$f_{S_2}(s_2) = \int_0^{s_2} f(s_1, s_2)\, ds_1 = \lambda^2 s_2\, e^{-\lambda s_2}.$$

Therefore,

$$f_{S_1}(s_1|S_2 = s_2) = \frac{\lambda^2 e^{-\lambda s_2}}{\lambda^2 s_2\, e^{-\lambda s_2}} = \frac{1}{s_2},$$

for $0 \le s_1 \le s_2$, and 0 otherwise. Therefore, conditioned on $T_1 + T_2 = s_2$, T_1 is uniform on $[0, s_2]$.

Imagine the following: a new lightbulb is put in and, after time T_1, it burns out. It is then replaced by a new lightbulb, identical to the first one, which also burns out after an additional time T_2. If we know the time when the second bulb burns out, the first bulb's failure time is uniform on the interval of its possible values.

Example 11.3. *Waiting to exceed the initial score*. For the first problem, roll a die once and assume that the number you rolled is U. Then, continue rolling the die until you either match or exceed U. What is the expected number of additional rolls?

Let N be the number of additional rolls. This number is Geometric, if we know U, so let us condition on the value of U. We know that

$$E(N|U = n) = \frac{6}{7-n},$$

and so, by Bayes' formula for expectation,

$$\begin{aligned} E(N) &= \sum_{n=1}^{6} E(N|U = n)\, P(U = n) \\ &= \frac{1}{6}\sum_{n=1}^{6} \frac{6}{7-n} \\ &= \sum_{n=1}^{6} \frac{1}{7-n} \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{6} = 2.45\,. \end{aligned}$$

Now, let U be a uniform random variable on $[0, 1]$, that is, the result of a call of a random number generator. Once we know U, generate additional independent uniform random variables (still on $[0, 1]$), $X_1, X_2, \ldots$, until we get one that equals or exceeds U. Let n be the number of additional calls of the generator, that is, the smallest n for which $X_n \geq U$. Determine the p. m. f. of N and EN.

Given that $U = u$, N is Geometric$(1-u)$. Thus, for $k = 0, 1, 2, \ldots$,

$$P(N = k|U = u) = u^{k-1}(1-u)$$

and so

$$P(N = k) = \int_0^1 P(N = k|U = u)\, du = \int_0^1 u^{k-1}(1-u)\, du = \frac{1}{k(k+1)}.$$

In fact, a slick alternate derivation shows that $P(N = k)$ does not depend on the distribution of random variables (which we assumed to be uniform), as soon as it is continuous, so that there are no "ties" (i.e., no two random variables are equal). Namely, the event $\{N = k\}$ happens exactly when X_k is the largest and U is the second largest among $X_1, X_2, \ldots, X_k, U$. All orders, by diminishing size, of these $k+1$ random numbers are equally likely, so the probability that X_k and U are the first and the second is $\frac{1}{k+1} \cdot \frac{1}{k}$.

It follows that

$$EN = \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty,$$

which can (in the uniform case) also be obtained by

$$EN = \int_0^1 E(N|U = u)\, du = \int_0^1 \frac{1}{1-u}\, du = \infty.$$

As we see from this example, random variables with infinite expectation are more common and natural than one might suppose.

Example 11.4. The number N of customers entering a store on a given day is Poisson(λ). Each of them buys something independently with probability p. Compute the probability that exactly k people buy something.

Let X be the number of people who buy something. Why should X be Poisson? Approximate: let n be the (large) number of people in the town and ϵ the probability that any particular one of them enters the store on a given day. Then, by the Poisson approximation, with $\lambda = n\epsilon$, $N \approx$ Binomial(n, ϵ) and $X \approx$ Binomial$(n, p\epsilon) \approx$ Poisson$(p\lambda)$. A more straightforward way to see this is as follows:

$$\begin{aligned}
P(X = k) &= \sum_{n=k}^{\infty} P(X = k | N = n) P(N = n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{\lambda^n e^{-\lambda}}{n!} \\
&= e^{-\lambda} \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} \\
&= \frac{e^{-\lambda} p^k \lambda^k}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k} \lambda^{n-k}}{(n-k)!} \\
&= \frac{e^{-\lambda} (p\lambda)^k}{k!} \sum_{\ell=0}^{\infty} \frac{((1-p)\lambda)^{\ell}}{\ell!} \\
&= \frac{e^{-\lambda} (p\lambda)^k}{k!} e^{(1-p)\lambda} \\
&= \frac{e^{-p\lambda} (p\lambda)^k}{k!}.
\end{aligned}$$

This is, indeed, the Poisson($p\lambda$) probability mass function.

Example 11.5. A coin with Heads probability p is tossed repeatedly. What is the expected number of tosses needed to get k successive heads?

Note: If we remove "successive," the answer is $\frac{k}{p}$, as it equals the expectation of the sum of k (independent) Geometric(p) random variables.

Let N_k be the number of needed tosses and $m_k = EN_k$. Let us condition on the value of N_{k-1}. If $N_{k-1} = n$, then observe the next toss; if it is Heads, then $N_k = n + 1$, but, if it is Tails, then we have to start from the beginning, with $n + 1$ tosses wasted. Here is how we translate this into mathematics:

$$\begin{aligned}
E[N_k | N_{k-1} = n] &= p(n+1) + (1-p)(n + 1 + E(N_k)) \\
&= pn + p + (1-p)n + 1 - p + m_k(1-p) \\
&= n + 1 + m_k(1-p).
\end{aligned}$$

Therefore,

$$\begin{aligned} m_k = E(N_k) &= \sum_{n=k-1}^{\infty} E[N_k|N_{k-1}=n]P(N_{k-1}=n) \\ &= \sum_{n=k-1}^{\infty} (n+1+m_k(1-p))\, P(N_{k-1}=n) \\ &= m_{k-1}+1+m_k(1-p) \\ &= \frac{1}{p}+\frac{m_{k-1}}{p}. \end{aligned}$$

This recursion can be unrolled,

$$\begin{aligned} m_1 &= \frac{1}{p} \\ m_2 &= \frac{1}{p}+\frac{1}{p^2} \\ &\vdots \\ m_k &= \frac{1}{p}+\frac{1}{p^2}+\ldots+\frac{1}{p^k}. \end{aligned}$$

In fact, we can even compute the moment generating function of N_k by different conditioning[1]. Let F_a, $a=0,\ldots,k-1$, be the event that the tosses begin with a Heads, followed by Tails, and let F_k the event that the first k tosses are Heads. One of $F_0,\ldots,F_k$ must happen, therefore, by Bayes' formula,

$$E[e^{tN_k}] = \sum_{a=0}^{k} E[e^{tN_k}|F_k]P(F_k).$$

If F_k happens, then $N_k=k$, otherwise $a+1$ tosses are wasted and one has to start over with the same conditions as at the beginning. Therefore,

$$E[e^{tN_k}] = \sum_{a=0}^{k-1} E[e^{t(N_k+a+1)}]p^a(1-p)+e^{tk}p^k = (1-p)E[e^{tN_k}]\sum_{a=0}^{k-1} e^{t(a+1)}p^a + e^{tk}p^k,$$

and this gives an equation for $E[e^{tN_k}]$ which can be solved:

$$E[e^{tN_k}] = \frac{p^k e^{tk}}{1-(1-p)\sum_{a=0}^{k-1} e^{t(a+1)}p^a} = \frac{p^k e^{tk}(1-pe^t)}{1-pe^t-(1-p)e^t(1-p^k e^{tk})}.$$

We can, then, get EN_k by differentiating and some algebra, by

$$EN_k = \frac{d}{dt}E[e^{tN_k}]|_{t=0}.$$

[1]Thanks to Travis Scrimshaw for pointing this out.

Example 11.6. *Gambler's ruin.* Fix a probability $p \in (0,1)$. Play a sequence of games; in each game you (independently) win \$1 with probability p and lose \$1 with probability $1-p$. Assume that your initial capital is i dollars and that you play until you either reach a predetermined amount N, or you lose all your money. For example, if you play a fair game, $p = \frac{1}{2}$, while, if you bet on Red at roulette, $p = \frac{9}{19}$. You are interested in the probability P_i that you leave the game happy with your desired amount N.

Another interpretation is that of a *simple random walk* on the integers. Start at i, $0 \le i \le N$, and make steps in discrete time units: each time (independently) move rightward by 1 (i.e., add 1 to your position) with probability p and move leftward by 1 (i.e., add -1 to your position) with probability $1-p$. In other words, if the position of the walker at time n is S_n, then

$$S_n = i + X_1 + \cdots + X_n,$$

where X_k are i. i. d. and $P(X_k = 1) = p$, $P(X_k = -1) = 1 - p$. This random walk is one of the very basic random (or, if you prefer a Greek word, stochastic) processes. The probability P_i is the probability that the walker visits N before a visit to 0.

We condition on the first step X_1 the walker makes, i.e., the outcome of the first bet. Then, by Bayes' formula,

$$\begin{aligned}
P_i &= P(\text{visit } N \text{ before } 0 | X_1 = 1) P(X_1 = 1) + P(\text{visit } N \text{ before } 0 | X_1 = -1) P(X_1 = -1) \\
&= P_{i+1} p + P_{i-1}(1-p),
\end{aligned}$$

which gives us a recurrence relation, which we can rewrite as

$$P_{i+1} - P_i = \frac{1-p}{p}\,(P_i - P_{i-1}).$$

We also have boundary conditions $P_0 = 0$, $P_N = 1$. This is a recurrence we can solve quite easily, as

$$\begin{aligned}
P_2 - P_1 &= \frac{1-p}{p}\,P_1 \\
P_3 - P_2 &= \frac{1-p}{p}\,(P_2 - P_1) = \left(\frac{1-p}{p}\right)^2 P_1 \\
&\vdots \\
P_i - P_{i-1} &= \left(\frac{1-p}{p}\right)^{i-1} P_1, \text{ for } i = 1, \ldots, N.
\end{aligned}$$

We conclude that

$$\begin{aligned}
P_i - P_1 &= \left(\left(\frac{1-p}{p}\right) + \left(\frac{1-p}{p}\right)^2 + \ldots + \left(\frac{1-p}{p}\right)^{i-1}\right) P_1, \\
P_i &= \left(1 + \left(\frac{1-p}{p}\right) + \left(\frac{1-p}{p}\right)^2 + \ldots + \left(\frac{1-p}{p}\right)^{i-1}\right) P_1.
\end{aligned}$$

Therefore,

$$P_i = \begin{cases} \frac{1-\left(\frac{1-p}{p}\right)^i}{1-\frac{1-p}{p}} P_1 & \text{if } p \neq \frac{1}{2}, \\ iP_1 & \text{if } p = \frac{1}{2}. \end{cases}$$

To determine the unknown P_1 we use that $P_N = 1$ to finally get

$$P_i = \begin{cases} \frac{1-\left(\frac{1-p}{p}\right)^i}{1-\left(\frac{1-p}{p}\right)^N} & \text{if } p \neq \frac{1}{2}, \\ \frac{i}{N} & \text{if } p = \frac{1}{2}. \end{cases}$$

For example, if $N = 10$ and $p = 0.6$, then $P_5 \approx 0.8836$; if $N = 1000$ and $p = \frac{9}{19}$, then $P_{900} \approx 2.6561 \cdot 10^{-5}$.

Example 11.7. *Bold Play.* Assume that the only game available to you is a game in which you can place even bets at any amount, and that you win each of these bets with probability p. Your initial capital is $x \in [0, N]$, a real number, and again you want to increase it to N before going broke. Your bold strategy (which can be proved to be the best) is to bet everything unless you are close enough to N that a smaller amount will do:

1. Bet x if $x \leq \frac{N}{2}$.
2. Bet $N - x$ if $x \geq \frac{N}{2}$.

We can, without loss of generality, fix our monetary unit so that $N = 1$. We now define

$$P(x) = P(\text{reach 1 before reaching 0}).$$

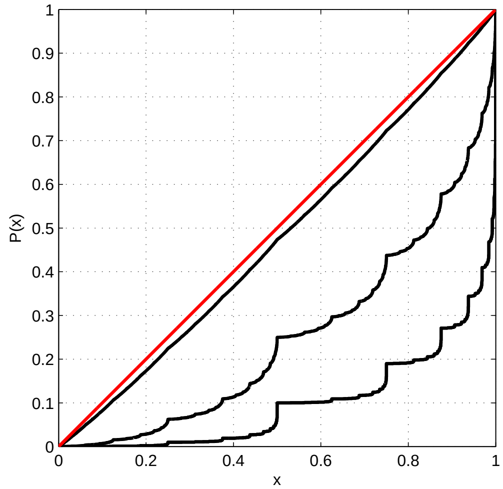
By conditioning on the outcome of your first bet,

$$P(x) = \begin{cases} p \cdot P(2x) & \text{if } x \in \left[0, \frac{1}{2}\right], \\ p \cdot 1 + (1-p) \cdot P(2x-1) & \text{if } x \in \left[\frac{1}{2}, 1\right]. \end{cases}$$

For each positive integer n, this is a linear system for $P(\frac{k}{2^n})$, $k = 0, \ldots, 2^n$, which can be solved. For example:

- When $n = 1$, $P\left(\frac{1}{2}\right) = p$.
- When $n = 2$, $P\left(\frac{1}{4}\right) = p^2$, $P\left(\frac{3}{4}\right) = p + (1-p)p$.
- When $n = 3$, $P\left(\frac{1}{8}\right) = p^3$, $P\left(\frac{3}{8}\right) = p \cdot P\left(\frac{3}{4}\right) = p^2 + p^2(1-p)$, $P\left(\frac{5}{8}\right) = p + p^2(1-p)$, $P\left(\frac{7}{8}\right) = p + p(1-p) + p(1-p)^2$.

It is easy to verify that $P(x) = x$, for all x, if $p = \frac{1}{2}$. Moreover, it can be computed that $P(0.9) \approx 0.8794$ for $p = \frac{9}{19}$, which is not too different from a fair game. The figure below displays the graphs of functions $P(x)$ for $p = 0.1$, 0.25, $\frac{9}{19}$, and $\frac{1}{2}$.



A few remarks for the more mathematically inclined: The function $P(x)$ is continuous, but nowhere differentiable on $[0, 1]$. It is thus a highly irregular function despite the fact that it is strictly increasing. In fact, $P(x)$ is the distribution function of a certain random variable Y, that is, $P(x) = P(Y \leq x)$. This random variable with values in $(0, 1)$ is defined by its binary expansion

$$Y = \sum_{j=1}^{\infty} D_j \frac{1}{2^j},$$

where its binary digits D_j are independent and equal to 1 with probability $1 - p$ and, thus, 0 with probability p.

Theorem 11.1. *Expectation and variance of sums with a random number of terms.*

Assume that $X, X_1, X_2, \ldots$ is an i. i. d. sequence of random variables with finite $EX = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Let N be a nonnegative integer random variable, independent of all X_i, and let

$$S = \sum_{i=1}^{N} X_i.$$

Then

$$ES = \mu\, EN,$$
$$\mathrm{Var}(S) = \sigma^2\, EN + \mu^2 \mathrm{Var}(N).$$

Proof. Let $S_n = X_1 + \ldots + X_n$. We have

$$E[S|N = n] = ES_n = nEX_1 = n\mu.$$

Then,

$$ES = \sum_{n=0}^{\infty} n\mu \, P(N = n) = \mu \, EN.$$

For variance, compute first

$$\begin{aligned}
E(S^2) &= \sum_{n=0}^{\infty} E[S^2|N = n]P(N = n) \\
&= \sum_{n=0}^{\infty} E(S_n^2)P(N = n) \\
&= \sum_{n=0}^{\infty}((\mathrm{Var}(S_n) + (ES_n)^2)P(N = n) \\
&= \sum_{n=0}^{\infty}(n\,\sigma^2 + n^2\mu^2)P(N = n) \\
&= \sigma^2 EN + \mu^2 E(N^2).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(S) &= E(S^2) - (ES)^2 \\
&= \sigma^2 EN + \mu^2 E(N^2) - \mu^2 (EN)^2 \\
&= \sigma^2 EN + \mu^2 \mathrm{Var}(N).
\end{aligned}$$

□

Example 11.8. Toss a fair coin until you toss Heads for the first time. Each time you toss Tails, roll a die and collect as many dollars as the number on the die. Let S be your total winnings. Compute ES and $\mathrm{Var}(S)$.

This fits into the above context, with X_i, the numbers rolled on the die, and N, the number of Tails tossed before first Heads. We know that

$$\begin{aligned}
EX_1 &= \frac{7}{2}, \\
\mathrm{Var}(X_1) &= \frac{35}{12}.
\end{aligned}$$

Moreover, $N + 1$ is a Geometric($\frac{1}{2}$) random variable, and so

$$\begin{aligned}
EN &= 2 - 1 = 1. \\
\mathrm{Var}(N) &= \frac{1 - \frac{1}{2}}{\left(\frac{1}{2}\right)^2} = 2
\end{aligned}$$

Plug in to get $ES = \frac{7}{2}$ and $\mathrm{Var}(S) = \frac{59}{12}$.

Example 11.9. We now take another look at Example 8.11. We will rename the number of days in purgatory as S, to fit it better into the present context, and call the three doors 0, 1, and 2. Let N be the number of times your choice of door is not door 0. This means that N is Geometric($\frac{1}{3}$)-1. Any time you do not pick door 0, you pick door 1 or 2 with equal probability. Therefore, each X_i is 1 or 2 with probability $\frac{1}{2}$ each. (Note that X_i are *not* 0, 1, or 2 with probability $\frac{1}{3}$ each!)

It follows that

$$EN = 3 - 1 = 2,$$

$$\mathrm{Var}(N) = \frac{1-\frac{1}{3}}{\left(\frac{1}{3}\right)^2} = 6$$

and

$$EX_1 = \frac{3}{2},$$

$$\mathrm{Var}(X_1) = \frac{1^2+2^2}{2} - \frac{9}{4} = \frac{1}{4}.$$

Therefore, $ES = EN \cdot EX_1 = 3$, which, of course, agrees with the answer in Example 8.11. Moreover,

$$\mathrm{Var}(S) = \frac{1}{4} \cdot 2 + \frac{9}{4} \cdot 6 = 14.$$

Problems

1. Toss an unfair coin with probability $p \in (0,1)$ of Heads n times. By conditioning on the outcome of the last toss, compute the probability that you get an even number of Heads.

2. Let X_1 and X_2 be independent Geometric(p) random variables. Compute the conditional p. m. f. of X_1 given $X_1 + X_2 = n$, $n = 2, 3, \dots$

3. Assume that the joint density of (X, Y) is

$$f(x,y) = \frac{1}{y}e^{-y}, \text{ for } 0 < x < y,$$

and 0 otherwise. Compute $E(X^2|Y = y)$.

4. You are trapped in a dark room. In front off you are two buttons, A and B. If you press A,

- with probability 1/3 you will be released in two minutes;
- with probability 2/3 you will have to wait five minutes and then you will be able to press one of the buttons again.

If you press B,

- you will have to wait three minutes and then be able to press one of the buttons again.

Assume that you cannot see the buttons, so each time you press one of them at random. Compute the expected time of your confinement.

5. Assume that a Poisson number with expectation 10 of customers enters the store. For promotion, each of them receives an in-store credit uniformly distributed between 0 and 100 dollars. Compute the expectation and variance of the amount of credit the store will give.

6. Generate a random number Λ uniformly on $[0, 1]$; once you observe the value of Λ, say $\Lambda = \lambda$, generate a Poisson random variable N with expectation λ. Before you start the random experiment, what is the probability that $N \geq 3$?

7. A coin has probability p of Heads. Alice flips it first, then Bob, then Alice, etc., and the winner is the first to flip Heads. Compute the probability that Alice wins.

Solutions to problems

1. Let p_n be the probability of an even number of Heads in n tosses. We have

$$p_n = p \cdot (1 - p_{n-1}) + (1 - p)p_{n-1} = p + (1 - 2p)p_{n-1},$$

and so

$$p_n - \frac{1}{2} = (1 - 2p)(p_{n-1} - \frac{1}{2}),$$

and then

$$p_n = \frac{1}{2} + C(1 - 2p)^n.$$

As $p_0 = 1$, we get $C = \frac{1}{2}$ and, finally,

$$p_n = \frac{1}{2} + \frac{1}{2}(1 - 2p)^n.$$

2. We have, for $i = 1, \dots, n - 1$,

$$P(X_1 = i|X_1+X_2 = n) = \frac{P(X_1 = i)P(X_2 = n - i)}{P(X_1 + X_2 = n)} = \frac{p(1-p)^{i-1}p(1-p)^{n-i-1}}{\sum_{k=1}^{n-1} p(1-p)^{k-1}p(1-p)^{n-k-1}} = \frac{1}{n-1},$$

so X_1 is uniform over its possible values.

3. The conditional density of X given $Y = y$ is $f_X(x|Y = y) = \frac{1}{y}$, for $0 < x < y$ (i.e., uniform on $[0, y]$), and so the answer is $\frac{y^2}{3}$.

4. Let I be the indicator of the event that you press A, and X the time of your confinement in minutes. Then,

$$EX = E(X|I = 0)P(I = 0) + E(X|I = 1)P(I = 1) = (3 + EX)\frac{1}{2} + (\frac{1}{3} \cdot 2 + \frac{2}{3} \cdot (5 + EX))\frac{1}{2}$$

and the answer is $EX = 21$.

5. Let N be the number of customers and X the amount of credit, while X_i are independent uniform on $[0, 100]$. So, $EX_i = 50$ and $\mathrm{Var}(X_i) = \frac{100^2}{12}$. Then, $X = \sum_{i=1}^N X_i$, so $EX = 50 \cdot EN = 500$ and $\mathrm{Var}(X) = \frac{100^2}{12} \cdot 10 + 50^2 \cdot 10$.

6. The answer is

$$P(N \geq 3) = \int_0^1 P(N \geq 3|\Lambda = \lambda)\, d\lambda = \int_0^1 (1 - (1 + \lambda + \frac{\lambda^2}{2})e^{-\lambda})d\lambda.$$

7. Let $f(p)$ be the probability. Then,

$$f(p) = p + (1 - p)(1 - f(p))$$

which gives

$$f(p) = \frac{1}{2 - p}.$$

Interlude: Practice Midterm 1

This practice exam covers the material from chapters 9 through 11. Give yourself 50 minutes to solve the four problems, which you may assume have equal point score.

1. Assume that a deck of $4n$ cards has n cards of each of the four suits. The cards are shuffled and dealt to n players, four cards per player. Let D_n be the number of people whose four cards are of four different suits.

(a) Find ED_n.

(b) Find $\text{Var}(D_n)$.

(c) Find a constant c so that $\frac{1}{n}D_n$ converges to c in probability, as $n \to \infty$.

2. Consider the following game, which will also appear in problem 4. Toss a coin with probability p of Heads. If you toss Heads, you win \$2, if you toss Tails, you win \$1.

(a) Assume that you play this game n times and let S_n be your combined winnings. Compute the moment generating function of S_n, that is, $E(e^{tS_n})$.

(b) Keep the assumptions from (a). Explain how you would find an upper bound for the probability that S_n is more than 10% larger than its expectation. Do not compute.

(c) Now you roll a fair die and you play the game as many times as the number you roll. Let Y be your total winnings. Compute $E(Y)$ and $\text{Var}(Y)$.

3. The joint density of X and Y is

$$f(x, y) = \frac{e^{-x/y}e^{-y}}{y},$$

for $x > 0$ and $y > 0$, and 0 otherwise. Compute $E(X|Y = y)$.

4. Consider the following game again: Toss a coin with probability p of Heads. If you toss Heads, you win \$2, if you toss Tails, you win \$1. Assume that you start with no money and you have to quit the game when your winnings match or exceed the dollar amount n. (For example, assume $n = 5$ and you have \$3: if your next toss is Heads, you collect \$5 and quit; if your next toss is Tails, you play once more. Note that, at the amount you quit, your winnings will be either n or $n + 1$.) Let p_n be the probability that you will quit with winnings exactly n.

(a) What is p_1? What is p_2?

(b) Write down the recursive equation which expresses p_n in terms of p_{n-1} and p_{n-2}.

(c) Solve the recursion.

Solutions to Practice Midterm 1

1. Assume that a deck of $4n$ cards has n cards of each of the four suits. The cards are shuffled and dealt to n players, four cards per player. Let D_n be the number of people whose four cards are of four different suits.

   (a) Find ED_n.

   Solution:

   As
   $$D_n = \sum_{i=1}^{n} I_i,$$
   where
   $$I_i = I_{\{i\text{th player gets four different suits}\}},$$
   and
   $$EI_i = \frac{n^4}{\binom{4n}{4}},$$
   the answer is
   $$ED_n = \frac{n^5}{\binom{4n}{4}}.$$

   (b) Find $\mathrm{Var}(D_n)$.

   Solution:

   We also have, for $i \neq j$,
   $$E(I_i I_j) = \frac{n^4(n-1)^4}{\binom{4n}{4}\binom{4n-4}{4}}.$$
   Therefore,
   $$E(D_n^2) = \sum_{i\neq j} E(I_i I_j) + \sum_i EI_i = \frac{n^5(n-1)^5}{\binom{4n}{4}\binom{4n-4}{4}} + \frac{n^5}{\binom{4n}{4}}$$
   and
   $$\mathrm{Var}(D_n) = \frac{n^5(n-1)^5}{\binom{4n}{4}\binom{4n-4}{4}} + \frac{n^5}{\binom{4n}{4}} - \left(\frac{n^5}{\binom{4n}{4}}\right)^2.$$

(c) Find a constant c so that $\frac{1}{n}D_n$ converges to c in probability as $n \to \infty$.

Solution:
Let $Y_n = \frac{1}{n}D_n$. Then

$$EY_n = \frac{n^4}{\binom{4n}{4}} = \frac{6n^3}{(4n-1)(4n-2)(4n-3)} \to \frac{6}{4^3} = \frac{3}{32},$$

as $n \to \infty$. Moreover,

$$\mathrm{Var}(Y_n) = \frac{1}{n^2}\mathrm{Var}(D_n) = \frac{n^3(n-1)^5}{\binom{4n}{4}\binom{4n-4}{4}} + \frac{n^3}{\binom{4n}{4}} - \left(\frac{n^4}{\binom{4n}{4}}\right)^2 \to \frac{6^2}{4^6} + 0 - \left(\frac{3}{32}\right)^2 = 0,$$

as $n \to \infty$, so the statement holds with $c = \frac{3}{32}$.

2. Consider the following game, which will also appear in problem 4. Toss a coin with probability p of Heads. If you toss Heads, you win \$2, if you toss Tails, you win \$1.

(a) Assume that you play this game n times and let S_n be your combined winnings. Compute the moment generating function of S_n, that is, $E(e^{tS_n})$.

Solution:

$$E(e^{tS_n}) = (e^{2t} \cdot p + e^t \cdot (1-p))^n.$$

(b) Keep the assumptions from (a). Explain how you would find an upper bound for the probability that S_n is more than 10% larger than its expectation. Do not compute.

Solution:
As $EX_1 = 2p + 1 - p = 1 + p$, $ES_n = n(1+p)$, and we need to find an upper bound for $P(S_n > n(1.1 + 1.1p)$. When $(1.1 + 1.1p) \geq 2$, i.e., $p \geq \frac{9}{11}$, this is an impossible event, so the probability is 0. When $p < \frac{9}{11}$, the bound is

$$P(S_n > n(1.1 + 1.1p) \leq e^{-I(1.1+1.1p)n},$$

where $I(1.1 + 1.1p) > 0$ and is given by

$$I(1.1 + 1.1p) = \sup\{(1.1 + 1.1p)t - \log(pe^{2t} + (1-p)e^t) : t > 0\}.$$

(c) Now you roll a fair die and you play the game as many times as the number you roll. Let Y be your total winnings. Compute $E(Y)$ and $\mathrm{Var}(Y)$.

Solution:
Let $Y = X_1 + \ldots + X_N$, where X_i are independently and identically distributed with $P(X_1 = 2) = p$ and $P(X_1 = 1) = 1 - p$, and $P(N = k) = \frac{1}{6}$, for $k = 1, \ldots, 6$. We know that

$$\begin{aligned} EY &= EN \cdot EX_1, \\ \mathrm{Var}(Y) &= \mathrm{Var}(X_1) \cdot EN + (EX_1)^2 \cdot \mathrm{Var}(N). \end{aligned}$$

We have

$$EN = \frac{7}{2}, \ \mathrm{Var}(N) = \frac{35}{12}.$$

Moreover,

$$EX_1 = 2p + 1 - p = 1 + p,$$

and

$$EX_1^2 = 4p + 1 - p = 1 + 3p,$$

so that

$$\mathrm{Var}(X_1) = 1 + 3p - (1 + p)^2 = p - p^2.$$

The answer is

$$\begin{aligned} EY &= \frac{7}{2} \cdot (1 + p), \\ \mathrm{Var}(Y) &= (p - p^2)\frac{7}{2} + (1 + p)^2 \frac{35}{12}. \end{aligned}$$

3. The joint density of X and Y is

$$f(x, y) = \frac{e^{-x/y} e^{-y}}{y},$$

for $x > 0$ and $y > 0$, and 0 otherwise. Compute $E(X|Y = y)$.

Solution:
We have

$$E(X|Y = y) = \int_0^\infty x f_X(x|Y = y)\, dx.$$

As

$$\begin{aligned} f_Y(y) &= \int_0^\infty \frac{e^{-x/y} e^{-y}}{y}\, dx \\ &= \frac{e^{-y}}{y} \int_0^\infty e^{-x/y}\, dx \\ &= \frac{e^{-y}}{y} \left[y e^{-x/y} \right]_{x=0}^{x=\infty} \\ &= e^{-y} \text{ , for } y > 0, \end{aligned}$$

we have

$$\begin{aligned} f_X(x|Y=y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{e^{-x/y}}{y}\text{, for } x,y>0, \end{aligned}$$

and so

$$\begin{aligned} E(X|Y=y) &= \int_0^\infty \frac{x}{y} e^{-x/y}\,dx \\ &= y\int_0^\infty ze^{-z}\,dz \\ &= y. \end{aligned}$$

4. Consider the following game again: Toss a coin with probability p of Heads. If you toss Heads, you win \$2, if you toss Tails, you win \$1. Assume that you start with no money and you have to quit the game when your winnings match or exceed the dollar amount n. (For example, assume $n = 5$ and you have \$3: if your next toss is Heads, you collect \$5 and quit; if your next toss is Tails, you play once more. Note that, at the amount you quit, your winnings will be either n or $n+1$.) Let p_n be the probability that you will quit with winnings exactly n.

   (a) What is p_1? What is p_2?

   Solution:
   We have
   $$p_1 = 1 - p$$
   and
   $$p_2 = (1-p)^2 + p.$$
   Also, $p_0 = 1$.

   (b) Write down the recursive equation which expresses p_n in terms of p_{n-1} and p_{n-2}.

   Solution:
   We have
   $$p_n = p \cdot p_{n-2} + (1-p)p_{n-1}.$$

(c) Solve the recursion.

Solution:
We can use

$$\begin{aligned} p_n - p_{n-1} &= (-p)(p_{n-1} - p_{n-2}) \\ &= (-p)^{n-1}(p_1 - p_0). \end{aligned}$$

Another possibility is to use the characteristic equation $\lambda^2 - (1-p)\lambda - p = 0$ to get

$$\lambda = \frac{1-p \pm \sqrt{(1-p)^2 + 4p}}{2} = \frac{1-p \pm (1+p)}{2} = \begin{cases} 1 \\ -p \end{cases}.$$

This gives

$$p_n = a + b(-p)^n,$$

with

$$a + b = 1,\ a - bp = 1 - p.$$

We get

$$a = \frac{1}{1+p},\ b = \frac{p}{1+p},$$

and then

$$p_n = \frac{1}{1+p} + \frac{p}{1+p}(-p)^n.$$

12 Markov Chains: Introduction

Example 12.1. Take your favorite book. Start, at step 0, by choosing a random letter. Pick one of the five random procedures described below and perform it at each time step $n = 1, 2, \dots$

1. Pick another random letter.
2. Choose a random occurrence of the letter obtained at the previous step ($(n-1)$st), then pick the letter following it in the text. Use the convention that the letter that follows the last letter in the text is the first letter in the text.
3. At step 1 use procedure (2), while for $n \geq 2$ choose a random occurrence of the two letters obtained, in order, in the previous two steps, then pick the following letter.
4. Choose a random occurrence of all previously chosen letters, in order, then pick the following letter.
5. At step n, perform procedure (1) with probability $\frac{1}{n}$ and perform procedure (2) with probability $1 - \frac{1}{n}$.

Repeated iteration of procedure (1) merely gives the familiar independent experiments — selection of letters is done with replacement and, thus, the letters at different steps are independent.

Procedure (2), however, is different: the probability mass function for the letter at the next time step depends on the letter at this step and nothing else. If we call our current letter our *state*, then we *transition* into a new state chosen with the p. m. f. that depends only on our current state. Such processes are called *Markov*.

Procedure (3) is not Markov at first glance. However, it becomes such via a natural redefinition of state: keep track of the last two letters; call an ordered pair of two letters a state.

Procedure (4) can be made Markov in a contrived fashion, that is, by keeping track, at the current state, of the entire history of the process. There is, however, no natural way of making this process Markov and, indeed, there is something different about this scheme: it ceases being random after many steps are performed, as the sequence of the chosen letters occurs just once in the book.

Procedure (5) is Markov, but what distinguishes it from (2) is that the p. m. f. is dependent not only on the current step, but also on time. That is, the process is Markov, but not *time-homogeneous*. We will only consider time-homogeneous Markov processes.

In general, a *Markov chain* is given by

- a *state space*, a countable set S of states, which are often labeled by the positive integers $1, 2, \dots$;
- *transition probabilities*, a (possibly infinite) matrix of numbers P_{ij}, where i and j range over all states in S; and

- an initial distribution α, a probability mass function on the states.

Here is how these three ingredients determine a sequence $X_0, X_1, \dots$ of random variables (with values in S): Use the initial distribution as your random procedure to pick X_0. Subsequently, given that you are at state $i \in S$ at any time n, make the transition to state $j \in S$ with probability P_{ij}, that is

$$P_{ij} = P(X_{n+1} = j | X_n = i). \tag{12.1}$$

The transition probabilities are collected into the *transition matrix*:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \dots \\ P_{21} & P_{22} & P_{23} & \dots \\ & \vdots & & \end{bmatrix}.$$

A *stochastic matrix* is a (possibly infinite) square matrix with nonnegative entries, such that all rows sum to 1, that is

$$\sum_j P_{ij} = 1,$$

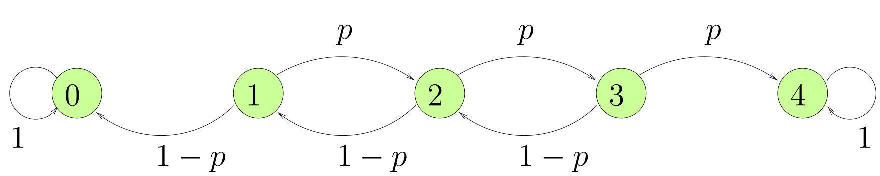
for all i. In other words, every row of the matrix is a p. m. f. Clearly, by (12.1), every transition matrix is a stochastic matrix (as X_{n+1} must be in some state). The opposite also holds: given any stochastic matrix, one can construct a Markov chain on positive integer states with the same transition matrix, by using the entries as transition probabilities as in (12.1).

Geometrically, a Markov chain is often represented as oriented graph on S (possibly with self-loops) with an oriented edge going from i to j whenever a transition from i to j is possible, i.e., whenever $P_{ij} > 0$; such an edge is labeled by P_{ij}.

Example 12.2. A random walker moves on the set $\{0, 1, 2, 3, 4\}$. She moves to the right (by 1) with probability, p, and to the left with probability $1 - p$, except when she is at 0 or at 4. These two states are *absorbing*: once there, the walker does not move. The transition matrix is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

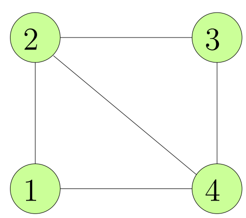
and the matching graphical representation is below.



Example 12.3. Same as the previous example except that now 0 and 4 are *reflecting*. From 0, the walker always moves to 1, while from 4 she always moves to 3. The transition matrix changes to

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Example 12.4. *Random walk on a graph.* Assume that a graph with undirected edges is given by its adjacency matrix, which is a binary matrix with the i, jth entry 1 exactly when i is connected to j. At every step, a random walker moves to a randomly chosen neighbor. For example, the adjacency matrix of the graph



is

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix},$$

and the transition matrix is

$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}.$$

Example 12.5. *The general two-state Markov chain.* There are two states 1 and 2 with transitions:

- $1 \to 1$ with probability α;
- $1 \to 2$ with probability $1 - \alpha$;
- $2 \to 1$ with probability β;
- $2 \to 2$ with probability $1 - \beta$.

The transition matrix has two parameters $\alpha, \beta \in [0, 1]$:

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}.$$

Example 12.6. *Changeovers*. Keep track of two-toss blocks in an infinite sequence of independent coin tosses with probability p of Heads. The states represent (previous flip, current flip) and are (in order) HH, HT, TH, and TT. The resulting transition matrix is

$$\begin{bmatrix} p & 1-p & 0 & 0 \\ 0 & 0 & p & 1-p \\ p & 1-p & 0 & 0 \\ 0 & 0 & p & 1-p \end{bmatrix}.$$

Example 12.7. *Simple random walk on $\mathbb{Z}$*. The walker moves left or right by 1, with probabilities p and $1-p$, respectively. The state space is doubly infinite and so is the transition matrix:

$$\begin{bmatrix} \ddots & & & & & & \\ \dots & 1-p & 0 & p & 0 & 0 & \dots \\ \dots & 0 & 1-p & 0 & p & 0 & \dots \\ \dots & 0 & 0 & 1-p & 0 & p & \dots \\ & & & & & & \ddots \end{bmatrix}$$

Example 12.8. *Birth-death chain*. This is a general model in which a population may change by at most 1 at each time step. Assume that the size of a population is x. *Birth* probability p_x is the transition probability to $x+1$, *death* probability q_x is the transition to $x-1$. and $r_x = 1 - p_x - q_x$ is the transition to x. Clearly, $q_0 = 0$. The transition matrix is

$$\begin{bmatrix} r_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & r_1 & p_1 & 0 & 0 & \dots \\ 0 & q_2 & r_2 & p_2 & 0 & \dots \\ & & & & & \ddots \end{bmatrix}$$

We begin our theory by studying *n-step probabilities*

$$P_{ij}^n = P(X_n = j | X_0 = i) = P(X_{n+m} = j | X_m = i).$$

Note that $P_{ij}^0 = I$, the identity matrix, and $P_{ij}^1 = P_{ij}$. Note also that the condition $X_0 = i$ simply specifies a particular non-random initial state.

Consider an oriented path of length n from i to j, that is $i, k_1, \dots, k_{n-1}, j$, for some states $k_1, \dots, k_{n-1}$. One can compute the probability of following this path by multiplying all transition probabilities, i.e., $P_{ik_1} P_{k_1 k_2} \cdots P_{k_{n-1} j}$. To compute P_{ij}^n, one has to sum these products over all paths of length n from i to j. The next theorem writes this in a familiar and neater fashion.

Theorem 12.1. *Connection between n-step probabilities and matrix powers:*

$$\boxed{P_{ij}^n \textit{ is the } i,j\textit{th entry of the } n\textit{th power of the transition matrix.}}$$

Proof. Call the transition matrix P and temporarily denote the n-step transition matrix by $P^{(n)}$. Then, for $m, n \geq 0$,

$$\begin{aligned}
P_{ij}^{(n+m)} &= P(X_{n+m} = j | X_0 = i) \\
&= \sum_k P(X_{n+m} = j, X_n = k | X_0 = i) \\
&= \sum_k P(X_{n+m} = j | X_n = k) \cdot P(X_n = k | X_0 = i) \\
&= \sum_k P(X_m = j | X_0 = k) \cdot P(X_n = k | X_0 = i) \\
&= \sum_k P_{kj}^{(m)} P_{ik}^{(n)}.
\end{aligned}$$

The first equality decomposes the probability according to where the chain is at time n, the second uses the Markov property and the third time-homogeneity. Thus,

$$P^{(m+n)} = P^{(n)} P^{(m)},$$

and, then, by induction

$$P^{(n)} = P^{(1)} P^{(1)} \cdots P^{(1)} = P^n.$$

□

The fact that the matrix powers of the transition matrix give the n-step probabilities makes linear algebra useful in the study of finite-state Markov chains.

Example 12.9. For the two state Markov Chain

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix},$$

and

$$P^2 = \begin{bmatrix} \alpha^2 + (1-\alpha)\beta & \alpha(1-\alpha) + (1-\alpha)(1-\beta) \\ \alpha\beta + (1-\beta)\beta & \beta(1-\alpha) + (1-\beta)^2 \end{bmatrix}$$

gives all P_{ij}^2.

Assume now that the initial distribution is given by

$$\alpha_i = P(X_0 = i),$$

for all states i (again, for notational purposes, we assume that $i = 1, 2, \ldots$). As this must determine a p. m. f., we have $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. Then,

$$\begin{aligned}
P(X_n = j) &= \sum_i P(X_n = j | X_0 = i) P(X_0 = i) \\
&= \sum_i \alpha_i P_{ij}^n.
\end{aligned}$$

Then, the *row* of probabilities at time n is given by $[P(X_n = i), i \in S] = [\alpha_1, \alpha_2, \ldots] \cdot P^n$.

Example 12.10. Consider the random walk on the graph from Example 12.4. Choose a starting vertex at random. (a) What is the probability mass function at time 2? (b) Compute $P(X_2 = 2, X_6 = 3, X_{12} = 4)$.

As

$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix},$$

we have

$$\begin{bmatrix} P(X_2 = 1) & P(X_2 = 2) & P(X_2 = 3) & P(X_2 = 4) \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \cdot P^2 = \begin{bmatrix} \frac{2}{9} & \frac{5}{18} & \frac{2}{9} & \frac{5}{18} \end{bmatrix}.$$

The probability in (b) equals

$$P(X_2 = 2) \cdot P_{23}^4 \cdot P_{34}^6 = \frac{8645}{708588} \approx 0.0122.$$

Problems

1. Three white and three black balls are distributed in two urns, with three balls per urn. The state of the system is the number of white balls in the first urn. At each step, we draw at random a ball from each of the two urns and exchange their places (the ball that was in the first urn is put into the second and vice versa). (a) Determine the transition matrix for this Markov chain. (b) Assume that initially all white balls are in the first urn. Determine the probability that this is also the case after 6 steps.

2. A Markov chain on states 0, 1, 2, has the transition matrix

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{5}{6} & 0 & \frac{1}{6} \end{bmatrix}$$

Assume that $P(X_0 = 0) = P(X_0 = 1) = \frac{1}{4}$. Determine EX_3.

3. We have two coins: coin 1 has probability 0.7 of Heads and coin 2 probability 0.6 of Heads. You flip a coin once per day, starting today (day 0), when you pick one of the two coins with equal probability and toss it. On any day, if you flip Heads, you flip coin 1 the next day,

otherwise you flip coin 2 the next day. (a) Compute the probability that you flip coin 1 on day 3. (b) Compute the probability that you flip coin 1 on days 3, 6, and 14. (c) Compute the probability that you flip Heads on days 3 and 6.

4. A walker moves on two positions a and b. She begins at a at time 0 and is at a the next time as well. Subsequently, if she is at the same position for two consecutive time steps, she changes position with probability 0.8 and remains at the same position with probability 0.2; in all other cases she decides the next position by a flip of a fair coin. (a) Interpret this as a Markov chain on a suitable state space and write down the transition matrix P. (b) Determine the probability that the walker is at position a at time 10.

Solutions to problems

1. The states are $0, 1, 2, 3$. For (a),

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} & 0 \\ 0 & \frac{4}{9} & \frac{4}{9} & \frac{1}{9} \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

For (b), compute the fourth entry of

$$\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \cdot P^6,$$

that is, the 4, 4th entry of P^6.

2. The answer is given by

$$EX_3 = \begin{bmatrix} P(X_3 = 0) & P(X_3 = 1) & P(X_3 = 2) \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \cdot P^3 \cdot \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

3. The state X_n of our Markov chain, 1 or 2, is the coin we flip on day n. (a) Let

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}.$$

Then,

$$\begin{bmatrix} P(X_3 = 1) & P(X_3 = 2) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \cdot P^3$$

and the answer to (a) is the first entry. (b) Answer: $P(X_3 = 1) \cdot P_{11}^3 \cdot P_{11}^8$. (c) You toss Heads on days 3 and 6 if and only if you toss Coin 1 on days 4 and 7, so the answer is $P(X_4 = 1) \cdot P_{11}^3$.

4. (a) The states are the four ordered pairs aa, ab, ba, and bb, which we will code as 1, 2, 3, and 4. Then,

$$P = \begin{bmatrix} 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 \end{bmatrix}.$$

The answer to (b) is the sum of the first and the third entries of

$$\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} P^9.$$

The power is 9 instead of 10 because the initial time for the chain (when it is at state aa) is time 1 for the walker.

13 Markov Chains: Classification of States

We say that a state j is *accessible from state* i, $i \to j$, if $P_{ij}^n > 0$ for some $n \geq 0$. This means that there is a possibility of reaching j from i in some number of steps. If j is not accessible from i, $P_{ij}^n = 0$ for all $n \geq 0$, and thus the chain started from i never visits j:

$$\begin{aligned} P(\text{ever visit } j|X_0 = i) &= P(\bigcup_{n=0}^{\infty}\{X_n = j\}\,|X_0 = i) \\ &\leq \sum_{n=0}^{\infty} P(X_n = j|X_0 = i) = 0. \end{aligned}$$

Also, note that for accessibility the size of entries of P does not matter, all that matters is which are positive and which are 0. For computational purposes, one should also observe that, if the chain has m states, then j is accessible from i if and only if $(P + P^2 + \ldots + P^m)_{ij} > 0$.

If i is accessible from j and j is accessible from i, then we say that i and j *communicate*, $i \leftrightarrow j$. It is easy to check that this is an *equivalence relation*:

1. $i \leftrightarrow i$;
2. $i \leftrightarrow j$ implies $j \leftrightarrow i$; and
3. $i \leftrightarrow j$ and $j \leftrightarrow k$ together imply $i \leftrightarrow k$.

The only nontrivial part is (3) and, to prove it, let us assume $i \to j$ and $j \to k$. This means that there exists an $n \geq 0$ so that $P_{ij}^n > 0$ and an $m \geq 0$ so that $P_{jk}^m > 0$. Now, one can get from i to j in $m + n$ steps by going first to j in n steps and then from j to k in m steps, so that

$$P_{ik}^{n+m} \geq P_{ij}^n P_{jk}^m > 0.$$

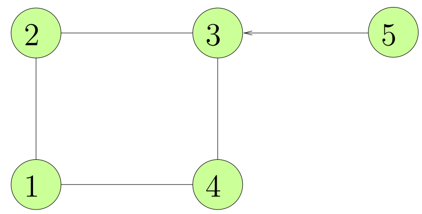
(Alternatively, one can use that $P^{m+n} = P^n \cdot P^m$ and then

$$P_{ik}^{n+m} = \sum_{\ell} P_{i\ell}^n P_{\ell k}^m \geq P_{ij}^n P_{jk}^m,$$

as the sum of nonnegative numbers is at least as large as one of its terms.)

The accessibility relation divides states into *classes*. Within each class, all states communicate with each other, but no pair of states in different classes communicates. The chain is *irreducible* if there is only one class. If the chain has m states, irreducibility means that all entries of $I + P + \ldots + P^m$ are nonzero.

Example 13.1. To determine the classes we may present the Markov chain as a graph, in which we only need to depict the edges that signify nonzero transition probabilities (their precise value is irrelevant for this purpose); by convention, we draw an undirected edge when probabilities in both directions are nonzero. Here is an example:



Any state $1, 2, 3, 4$ is accessible from any of the five states, but 5 is not accessible from $1, 2, 3, 4$. So, we have two classes: $\{1, 2, 3, 4\}$, and $\{5\}$. The chain is not irreducible.

Example 13.2. Consider the chain on states 1, 2, 3 and

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

As $1 \leftrightarrow 2$ and $2 \leftrightarrow 3$, this is an irreducible chain.

Example 13.3. Consider the chain on states 1, 2, 3, 4, and

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This chain has three classes $\{1, 2\}$, $\{3\}$ and $\{4\}$, hence, it is not irreducible.

For any state i, denote

$$f_i = P(\text{ever reenter } i | X_0 = i).$$

We call a state i *recurrent* if $f_i = 1$, and *transient* if $f_i < 1$.

Example 13.4. Back to the previous example. Obviously, 4 is recurrent, as it is an absorbing state. The only possibility of returning to 3 is to do so in one step, so we have $f_3 = \frac{1}{4}$, and 3 is transient. Moreover, $f_1 = 1$ because in order to never return to 1 we need to go to state 2 and stay there forever. We stay at 2 for n steps with probability

$$\left(\frac{1}{2}\right)^n \to 0,$$

as $n \to \infty$, so the probability of staying at 1 forever is 0 and, consequently, $f_1 = 1$. By similar logic, $f_2 = 1$. We will soon develop better methods to determine recurrence and transience.

Starting from any state, a Markov Chain visits a recurrent state infinitely many times or not at all. Let us now compute, in two different ways, the expected number of visits to i (i.e., the times, including time 0, when the chain is at i). First, we observe that, at every visit to i, the probability of never visiting i again is $1 - f_i$, therefore,

$$P(\text{exactly } n \text{ visits to } i | X_0 = i) = f_i^{n-1}(1 - f_i).$$

This formula says that the number of visits to i is a Geometric$(1 - f_i)$ random variable and so its expectation is

$$E(\text{number of visits to } i|X_0 = i) = \frac{1}{1 - f_i}.$$

A second way to compute this expectation is by using the indicator trick:

$$E(\text{number of visits to } i|X_0 = i) = E(\sum_{n=0}^{\infty} I_n|X_0 = i),$$

where $I_n = I_{\{X_n=i\}}$, $n = 0, 1, 2, \ldots$. Then,

$$\begin{aligned} E(\sum_{n=0}^{\infty} I_n|X_0 = i) &= \sum_{n=0}^{\infty} P(X_n = i|X_0 = i) \\ &= \sum_{n=0}^{\infty} P_{ii}^n. \end{aligned}$$

Thus,

$$\frac{1}{1 - f_i} = \sum_{n=0}^{\infty} P_{ii}^n$$

and we have proved the following theorem.

Theorem 13.1. *Characterization of recurrence via n step return probabilities:*

$$\boxed{\textit{A state } i \textit{ is recurrent if and only if } \sum_{n=1}^{\infty} P_{ii}^n = \infty.}$$

We call a subset $S_0 \subset S$ of states *closed* if $P_{ij} = 0$ for each $i \in S_0$ and $j \notin S_0$. In plain language, once entered, a closed set cannot be exited.

Proposition 13.2. *If a closed subset S_0 has only finitely many states, then there must be at least one recurrent state. In particular, any finite Markov chain must contain at least one recurrent state.*

Proof. Start from any state from S_0. By definition, the chain stays in S_0 forever. If all states in S_0 are transient, then each of them is visited either not at all or only finitely many times. This is impossible. □

Proposition 13.3. *If i is recurrent and $i \to j$, then also $j \to i$.*

Proof. There is an n_0 such that $P_{ij}^{n_0} > 0$, i.e., starting from i, the chain can reach j in n_0 steps. Thus, every time it is at i, there is a fixed positive probability that it will be at j n_0 steps later. Starting from i, the chain returns to i infinitely many times and, every time it does so, it has an independent chance to reach j n_0 steps later; thus, eventually the chain does reach j. Now assume that it is not true that $j \to i$. Then, once the chain reaches j, it never returns to i, but then, i is not recurrent. This contradiction ends the proof. □

Proposition 13.4. *If i is recurrent and $i \to j$, then j is also recurrent. Therefore, in any class, either all states are recurrent or all are transient. In particular, if the chain is irreducible, then either all states are recurrent or all are transient.*

In light of this proposition, we can classify each class, as well as an irreducible Markov chain, as recurrent or transient.

Proof. By the previous proposition, we know that also $j \to i$. We will now give two arguments for the recurrence of j.

We could use the same logic as before: starting from j, the chain must visit i with probability 1 (or else the chain starting at i has a positive probability of no return to i, by visiting j), then it returns to i infinitely many times and, at each of those times, it has an independent chance of getting to j at a later time — so it must do so infinitely often.

For another argument, we know that there exist $k, m \geq 0$ so that $P_{ij}^k > 0, P_{ji}^m > 0$. Furthermore, for any $n \geq 0$, one way to get from j to j in $m+n+k$ steps is by going from j to i in m steps, then from i to i in n steps, and then from i to j in k steps; thus,

$$P_{jj}^{m+n+k} \geq P_{ji}^m P_{ii}^n P_{ij}^k.$$

If $\sum_{n=0}^{\infty} P_{ii} = \infty$, then $\sum_{n=0}^{\infty} P_{jj}^{m+n+k} = \infty$ and, finally, $\sum_{\ell=0}^{\infty} P_{jj}^{\ell} = \infty$. In short, if i is recurrent, then so is j. □

Proposition 13.5. *Any recurrent class is a closed subset of states.*

Proof. Let S_0 be a recurrent class, $i \in S_0$ and $j \notin S_0$. We need to show that $P_{ij} = 0$. Assume the converse, $P_{ij} > 0$. As j does not communicate with i, the chain never reaches i from j, i.e., i is not accessible from j. But this is a contradiction to Proposition 13.3. □

For finite Markov chains, these propositions make it easy to determine recurrence and transience: if a class is closed, it is recurrent, but if it is not closed, it is transient.

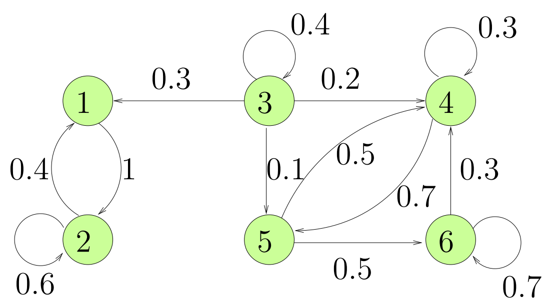
Example 13.5. Assume that the states are 1, 2, 3, 4 and that the transition matrix is

$$P = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

By inspection, every state is accessible from every other state and so this chain is irreducible. Therefore, every state is recurrent.

Example 13.6. Assume now that the states are $1, \dots, 6$ and

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0.4 & 0.2 & 0.1 & 0 \\ 0 & 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.3 & 0 & 0.7 \end{bmatrix}.$$



We observe that 3 can only be reached from 3, therefore, 3 is in a class of its own. States 1 and 2 can reach each other and no other state, so they form a class together. Furthermore, 4, 5, 6 all communicate with each other. Thus, the division into classes is $\{1, 2\}$, $\{3\}$, and $\{4, 5, 6\}$. As it is not closed, $\{3\}$ is a transient class (in fact, it is clear that $f_3 = 0.4$). On the other hand, $\{1, 2\}$ and $\{4, 5, 6\}$ both are closed and, therefore, recurrent.

Example 13.7. *Recurrence of a simple random walk on $\mathbb{Z}$.* Recall that such a walker moves from x to $x + 1$ with probability p and to $x - 1$ with probability $1 - p$. We will assume that $p \in (0, 1)$ and denote the chain $S_n = S_n^{(1)}$. (The superscript indicates the dimension. We will make use of this in subsequent examples in which the walker will move in higher dimensions.) As such a walk is irreducible, we only have to check whether state 0 is recurrent or transient, so we assume that the walker begins at 0. First, we observe that the walker will be at 0 at a later time only if she makes an equal number of left and right moves. Thus, for $n = 1, 2, \dots$,

$$P_{00}^{2n-1} = 0$$

and

$$P_{00}^{2n} = \binom{2n}{n} p^n (1 - p)^n.$$

Now, we recall Stirling's formula:

$$n! \sim n^n e^{-n} \sqrt{2\pi n}$$

(the symbol "$\sim$" means that the quotient of the two quantities converges to 1 as $n \to \infty$).

Therefore,

$$\begin{aligned}\binom{2n}{n} &= \frac{(2n)!}{(n!)^2} \\ &\sim \frac{(2n)^{2n}e^{-2n}\sqrt{2\pi 2n}}{n^{2n}e^{-2n}2\pi n} \\ &= \frac{2^{2n}}{\sqrt{n\pi}},\end{aligned}$$

and, therefore,

$$\begin{aligned}P_{00}^{2n} &= \frac{2^{2n}}{\sqrt{n\pi}}p^n(1-p)^n \\ &\sim \frac{1}{\sqrt{n\pi}}(4p(1-p))^n\end{aligned}$$

In the symmetric case, when $p = \frac{1}{2}$,

$$P_{00}^{2n} \sim \frac{1}{\sqrt{n\pi}},$$

therefore,

$$\sum_{n=0}^{\infty} P_{00}^{2n} = \infty,$$

and the random walk is recurrent.

When $p \neq \frac{1}{2}$, $4p(1-p) < 1$, so that P_{00}^{2n} goes to 0 faster than the terms of a convergent geometric series,

$$\sum_{n=0}^{\infty} P_{00}^{2n} < \infty,$$

and the random walk is transient. In this case, what is the probability f_0 that the chain ever reenters 0? We need to recall the Gambler's ruin probabilities,

$$P(S_n \text{ reaches } N \text{ before } 0|S_0 = 1) = \frac{1 - \frac{1-p}{p}}{1 - \left(\frac{1-p}{p}\right)^N}.$$

As $N \to \infty$, the probability

$$P(S_n \text{ reaches } 0 \text{ before } N|S_0 = 1) = 1 - P(S_n \text{ reaches } N \text{ before } 0|S_0 = 1)$$

converges to

$$P(S_n \text{ ever reaches } 0|S_0 = 1) = \begin{cases} 1 & \text{if } p < \frac{1}{2}, \\ \frac{1-p}{p} & \text{if } p > \frac{1}{2}. \end{cases}$$

Assume that $p > \frac{1}{2}$. Then,

$$\begin{aligned} f_0 &= P(S_1 = 1,\, S_n \text{ returns to 0 eventually}) + P(S_1 = -1,\, S_n \text{ returns to 0 eventually}) \\ &= p \cdot \frac{1-p}{p} + (1-p)\cdot 1 \\ &= 2(1-p). \end{aligned}$$

If $p < 1/2$, we may use the fact that replacing the walker's position with its mirror image replaces p by $1-p$; this gives $f_0 = 2p$ when $p < \frac{1}{2}$.

Example 13.8. *Is the simple symmetric random walk on $\mathbb{Z}^2$ recurrent?* A walker now moves on integer points in two dimensions: each step is a distance 1 jump in one of the four directions (N, S, E, or W). We denote this Markov chain by $S_n^{(2)}$ and imagine a drunk wandering at random through the rectangular grid of streets of a large city. (Chicago would be a good example.) The question is whether the drunk will eventually return to her home at $(0,0)$. All starting positions in this and in the next example will be the appropriate origins. Note again that the walker can only return in an even number of steps and, in fact, both the number of steps in the x direction (E or W) and in the y direction (N or S) must be even (otherwise, the respective coordinate cannot be 0).

We condition on the number N of times the walker moves in the x-direction:

$$\begin{aligned} P(S_{2n}^{(2)} = (0,0)) &= \sum_{k=0}^{n} P(N=2k)P(S_{2n}^{(2)} = (0,0)|N=2k) \\ &= \sum_{k=0}^{n} P(N=2k)P(S_{2k}^{(1)} = 0)P(S_{2(n-k)}^{(1)} = 0). \end{aligned}$$

In order not to obscure the computation, we will not show the full details from now on; filling in the missing pieces is an excellent computational exercise.

First, as the walker chooses to go horizontally or vertically with equal probability, $N \sim \frac{2n}{2} = n$ with overwhelming probability and so we can assume that $k \sim \frac{n}{2}$. Taking this into account,

$$\begin{aligned} P(S_{2k}^{(1)} = 0) &\sim \frac{\sqrt{2}}{\sqrt{n\pi}}, \\ P(S_{2(n-k)}^{(1)} = 0) &\sim \frac{\sqrt{2}}{\sqrt{n\pi}}. \end{aligned}$$

Therefore,

$$\begin{aligned} P(S_{2n}^{(2)} = (0,0)) &\sim \frac{2}{n\pi}\sum_{k=0}^{n} P(N=2k) \\ &\sim \frac{2}{n\pi} P(N \text{ is even}) \\ &\sim \frac{1}{n\pi}, \end{aligned}$$

as we know that (see Problem 1 in Chapter 11)

$$P(N \text{ is even}) = \frac{1}{2}.$$

Therefore,

$$\sum_{n=0}^{\infty} P(S_{2n}^{(2)} = (0,0)) = \infty$$

and we have demonstrated that this chain is still recurrent, albeit barely. In fact, there is an easier slick proof that does not generalize to higher dimensions, which demonstrates that

$$P(S_{2n}^{(2)} = 0) \quad = \quad P(S_{2n}^{(1)} = 0)^2.$$

Here is how it goes. If we let each coordinate of a two-dimensional random walker move *independently*, then the above is certainly true. Such a walker makes *diagonal* moves, from (x, y) to $(x+1, y+1)$, $(x-1, y+1)$, $(x+1, y-1)$, or $(x-1, y-1)$ with equal probability. At first, this appears to be a different walk, but if we rotate the lattice by 45 degrees, scale by $\frac{1}{\sqrt{2}}$, and ignore half of the points that are never visited, this becomes the same walk as $S_n^{(2)}$. In particular, it is at the origin exactly when $S_n^{(2)}$ is.

Example 13.9. *Is the simple symmetric random walk on $\mathbb{Z}^3$ recurrent?* Now, imagine a squirrel running around in a 3 dimensional maze. The process $S_n^{(3)}$ moves from a point (x, y, z) to one of the six neighbors $(x \pm 1, y, z)$, $(x, y \pm 1, z)$, $(x, y, z \pm 1)$ with equal probability. To return to $(0,0,0)$, it has to make an even number number of steps in each of the three directions. We will condition on the number N of steps in the z direction. This time $N \sim \frac{2n}{3}$ and, thus,

$$\begin{aligned}
P(S_{2n}^{(3)} = (0,0,0)) &= \sum_{k=0}^{n} P(N = 2k) P(S_{2k}^{(1)} = 0) P(S_{2(n-k)}^{(2)} = (0,0)) \\
&\sim \sum_{k=0}^{n} P(N = 2k) \frac{\sqrt{3}}{\sqrt{\pi n}} \frac{3}{2\pi n} \\
&= \frac{3\sqrt{3}}{2\pi^{3/2} n^{3/2}} P(N \text{ is even}) \\
&\sim \frac{3\sqrt{3}}{4\pi^{3/2} n^{3/2}}.
\end{aligned}$$

Therefore,

$$\sum_n P(S_{2n}^{(3)} = (0,0,0)) < \infty$$

and the three-dimensional random walk is transient, so the squirrel may never return home. The probability $f_0 = P(\text{return to } 0)$, thus, is not 1, but can we compute it? One approximation is obtained by using

$$\frac{1}{1 - f_0} = \sum_{n=0}^{\infty} P(S_{2n}^{(3)} = (0,0,0)) = 1 + \frac{1}{6} + \dots,$$

but this series converges slowly and its terms are difficult to compute. Instead, one can use the remarkable formula, derived by Fourier analysis,

$$\frac{1}{1-f_0} = \frac{1}{(2\pi)^3} \iiint_{(-\pi,\pi)^3} \frac{dx\,dy\,dz}{1-\frac{1}{3}(\cos(x)+\cos(y)+\cos(z))},$$

which gives, to four decimal places,

$$f_0 \approx 0.3405.$$

Problems

1. For the following transition matrices, determine the classes and specify which are recurrent and which transient.

$$P_1 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \end{bmatrix} \qquad P_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad P_4 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

2. Assume that a Markov chain X_n has states $0, 1, 2\ldots$ and transitions from each $i > 0$ to $i+1$ with probability $1 - \frac{1}{2\cdot i^\alpha}$ and to 0 with probability $\frac{1}{2\cdot i^\alpha}$. Moreover, from 0 it transitions to 1 with probability 1. (a) Is this chain irreducible? (b) Assume that $X_0 = 0$ and let R be the first return time to 0 (i.e., the first time after the initial time the chain is back at the origin). Determine α for which

$$1 - f_0 = P(\text{no return}) = \lim_{n\to\infty} P(R > n) = 0.$$

(c) Depending on α, determine which classes are recurrent.

3. Consider the one-dimensional simple symmetric random walk $S_n = S_n^{(1)}$ with $p = \frac{1}{2}$. As in the Gambler's ruin problem, fix an N and start at some $0 \le i \le N$. Let E_i be the expected time at which the walk first hits either 0 or N. (a) By conditioning on the first step, determine the recursive equation for E_i. Also, determine the boundary conditions E_0 and E_N. (b) Solve the recursion. (c) Assume that the chain starts at 0 and let R be the first time (after time 0) that it revisits 0. By recurrence, we know that $P(R < \infty) = 1$; use (b) to show that $ER = \infty$. The walk will eventually return to 0, but the expected waiting time is infinite!

Solutions to problems

1. Assume that the states are $1, \dots, 5$. For P_1: $\{1, 2, 3\}$ recurrent, $\{4, 5\}$ transient. For P_2: irreducible, so all states are recurrent. For P_3: $\{1, 2, 3\}$ recurrent, $\{4, 5\}$ recurrent. For P_4: $\{1, 2\}$ recurrent, $\{3\}$ recurrent (absorbing), $\{4\}$ transient, $\{5\}$ transient.

2. (a) The chain is irreducible. (b) If $R > n$, then the chain, after moving to 1, makes $n - 1$ consecutive steps to the right, so

$$P(R > n) = \prod_{i=1}^{n-1} \left(1 - \frac{1}{2 \cdot i^\alpha}\right).$$

The product converges to 0 if and only if its logarithm converges to $-\infty$ and that holds if and only if the series

$$\sum_{i=1}^{\infty} \frac{1}{2 \cdot i^\alpha}$$

diverges, which is when $\alpha \le 1$. (c) For $\alpha \le 1$, the chain is recurrent, otherwise, it is transient.

3. For (a), the walker makes one step and then proceeds from $i+1$ or $i-1$ with equal probability, so that

$$E_i = 1 + \frac{1}{2}(E_{i+1} + E_i),$$

with $E_0 = E_N = 0$. For (b), the homogeneous equation is the same as the one in the Gambler's ruin, so its general solution is linear: $Ci + D$. We look for a particular solution of the form Bi^2 and we get $Bi^2 = 1 + \frac{1}{2}(B(i^2 + 2i + 1) + B(i^2 - 2i + 1)) = 1 + Bi^2 + B$, so $B = -1$. By plugging in the boundary conditions we can solve for C and D to get $D = 0$, $C = N$. Therefore,

$$E_i = i(N - i).$$

For (c), after a step the walker proceeds either from 1 or -1 and, by symmetry, the expected time to get to 0 is the same for both. So, for every N,

$$ER \ge 1 + E_1 = 1 + 1 \cdot (N - 1) = N,$$

and so $ER = \infty$.

14 Branching processes

In this chapter we will consider a random model for population growth in the absence of spatial or any other resource constraints. So, consider a population of individuals which evolves according to the following rule: in every generation $n = 0, 1, 2, \ldots$, each individual produces a random number of offspring in the next generation, independently of other individuals. The probability mass function for offspring is often called the *offspring distribution* and is given by

$$p_i = P(\text{number of offspring} = i),$$

for $i = 0, 1, 2, \ldots$. We will *assume* that $p_0 < 1$ and $p_1 < 1$ to eliminate the trivial cases. This model was introduced by F. Galton in the late 1800s to study the disappearance of family names; in this case p_i is the probability that a man has i sons.

We will start with a single individual in generation 0 and generate the resulting random *family tree*. This tree is either finite (when some generation produces no offspring at all) or infinite — in the former case, we say that the branching process *dies out* and, in the latter case, that it *survives*.

We can look at this process as a Markov chain where X_n is the number of individuals in generation n. Let us start with the following observations:

- If X_n reaches 0, it stays there, so 0 is an absorbing state.
- If $p_0 > 0$, $P(X_{n+1} = 0 | X_n = k) > 0$, for all k.
- Therefore, by Proposition 13.5, all states other than 0 are transient if $p_0 > 0$; the population must either die out or increase to infinity. If $p_0 = 0$, then the population cannot decrease and each generation increases with probability at least $1 - p_1$, therefore it must increase to infinity.

It is possible to write down the transition probabilities for this chain, but they have a rather complicated explicit form, as

$$P(X_{n+1} = i | X_n = k) \quad = \quad P(W_1 + W_2 + \ldots + W_k = i),$$

where $W_1, \ldots, W_k$ are independent random variables, each with the offspring distribution. This suggests using moment generating functions, which we will indeed do. Recall that we are assuming that $X_0 = 1$.

Let

$$\delta_n = P(X_n = 0)$$

be the probability that the population is extinct by generation (which we also think of as time) n. The probability π_0 that the branching process dies out is, then, the limit of these probabilities:

$$\pi_0 = P(\text{the process dies out}) = P(X_n = 0 \text{ for some } n) = \lim_{n \to \infty} P(X_n = 0) = \lim_{n \to \infty} \delta_n.$$

Note that $\pi_0 = 0$ if $p_0 = 0$. Our main task will be to compute π_0 for general probabilities p_k. We start, however, with computing the expectation and variance of the population at generation n.

Let μ and σ^2 be the expectation and the variance of the offspring distribution, that is,

$$\mu = EX_1 = \sum_{k=0}^{\infty} kp_k,$$

and

$$\sigma^2 = \text{Var}(X_1).$$

Let $m_n = E(X_n)$ and $v_n = \text{Var}(X_n)$. Now, X_{n+1} is the sum of a random number X_n of independent random variables, each with the offspring distribution. Thus, we have by Theorem 11.1,

$$m_{n+1} = m_n \mu,$$

and

$$v_{n+1} = m_n \sigma^2 + v_n \mu^2.$$

Together with initial conditions $m_0 = 1$, $v_0 = 0$, the two recursive equations determine m_n and v_n. We can very quickly solve the first recursion to get $m_n = \mu^n$ and so

$$v_{n+1} = \mu^n \sigma^2 + v_n \mu^2.$$

When $\mu = 1$, $m_n = 1$ and $v_n = n\sigma^2$. When $\mu \neq 1$, the recursion has the general solution $v_n = A\mu^n + B\mu^{2n}$. The constant A must satisfy

$$A\mu^{n+1} = \sigma^2 \mu^n + A\mu^{n+2},$$

so that,

$$A = \frac{\sigma^2}{\mu(1-\mu)}.$$

From $v_0 = 0$ we get $A + B = 0$ and the solution is given in the next theorem.

Theorem 14.1. *Expectation m_n and variance v_n of the nth generation count.*

We have

$$m_n = \mu^n$$

and

$$v_n = \begin{cases} \frac{\sigma^2 \mu^n (1-\mu^n)}{\mu(1-\mu)} & \text{if } \mu \neq 1, \\ n\sigma^2 & \text{if } \mu = 1. \end{cases}$$

We can immediately conclude that $\mu < 1$ implies $\pi_0 = 1$, as

$$P(X_n \neq 0) = P(X_n \geq 1) \leq EX_n = \mu^n \to 0;$$

if the individuals have less than one offspring on average, the branching process dies out.

Now, let ϕ be the moment generating function of the offspring distribution. It is more convenient to replace e^t in our original definition with s, so that

$$\phi(s) = \phi_{X_1}(s) = E(s^{X_1}) = \sum_{k=0}^{\infty} p_k s^k.$$

In combinatorics, this would be exactly the generating function of the sequence p_k. Then, the moment generating function of X_n is

$$\phi_{X_n}(s) = E[s^{X_n}] = \sum_{k=0}^{\infty} P(X_n = k) s^k.$$

We will assume that $0 \le s \le 1$ and observe that, for such s, this power series converges. Let us get a recursive equation for ϕ_{X_n} by conditioning on the population count in generation $n-1$:

$$\begin{aligned}
\phi_{X_n}(s) &= E[s^{X_n}] \\
&= \sum_{k=0}^{\infty} E[s^{X_n} | X_{n-1} = k] P(X_{n-1} = k) \\
&= \sum_{k=0}^{\infty} E[s^{W_1 + \ldots + W_k}] P(X_{n-1} = k) \\
&= \sum_{k=0}^{\infty} (E(s^{W_1}) E(s^{W_2}) \cdots E(s^{W_k}) P(X_{n-1} = k) \\
&= \sum_{k=0}^{\infty} \phi(s)^k P(X_{n-1} = k) \\
&= \phi_{X_{n-1}}(\phi(s)).
\end{aligned}$$

So, ϕ_{X_n} is the nth iterate of ϕ,

$$\phi_{X_2}(s) = \phi(\phi(s)),\ \phi_{X_3}(s) = \phi(\phi(\phi(s))),\ \ldots$$

and we can also write

$$\phi_{X_n}(s) = \phi(\phi_{X_{n-1}}(s)).$$

Next, we take a closer look at the properties of ϕ. Clearly,

$$\phi(0) = p_0 > 0$$

and

$$\phi(1) = \sum_{k=0}^{\infty} p_k = 1.$$

Moreover, for $s > 0$,

$$\phi'(s) = \sum_{k=0}^{\infty} k p_k s^{k-1} > 0,$$

so ϕ is strictly increasing, with

$$\phi'(1) = \mu.$$

Finally,

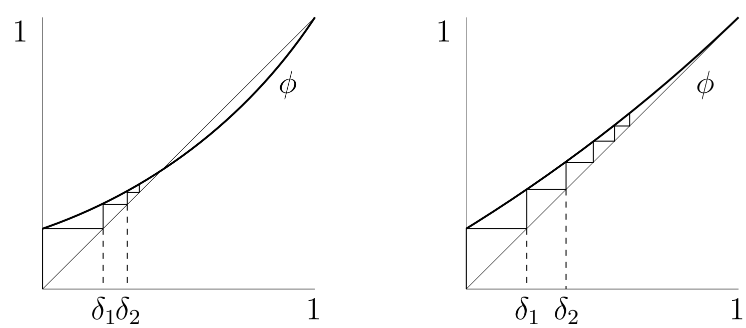
$$\phi''(s) = \sum_{k=1}^{\infty} k(k-1)p_k s^{k-2} \geq 0,$$

so ϕ is also convex. The crucial observation is that

$$\delta_n = \phi_{X_n}(0),$$

and so δ_n is obtained by starting at 0 and computing the nth iteration of ϕ. It is also clear that δ_n is a nondecreasing sequence (because $X_{n-1} = 0$ implies that $X_n = 0$). We now consider separately two cases:

- Assume that ϕ is always above the diagonal, that is, $\phi(s) \geq s$ for all $s \in [0, 1]$. This happens exactly when $\mu = \phi'(1) \leq 1$. In this case, δ_n converges to 1, and so $\pi_0 = 1$. This is shown in the right graph of the figure below.
- Now, assume that ϕ is not always above the diagonal, which happens when $\mu > 1$. In this case, there exists exactly one $s < 1$ which solves $s = \phi(s)$. As δ_n converges to this solution, we conclude that $\pi_0 < 1$ is the smallest solution to $s = \phi(s)$. This is shown in the left graph of the figure below.



The following theorem is a summary of our findings.

Theorem 14.2. *Probability that the branching process dies out.*

If $\mu \leq 1$, $\pi_0 = 1$. If $\mu > 1$, then π_0 is the smallest solution on $[0, 1]$ to $s = \phi(s)$.

Example 14.1. Assume that a branching process is started with $X_0 = k$ instead of $X_0 = 1$. How does this change the survival probability? The k individuals all evolve independent family trees, so that the probability of eventual death is π_0^k. It also follows that

$$P(\text{the process ever dies out}|X_n = k) = \pi_0^k$$

for every n.

If μ is barely larger than 1, the probability π_0 of extinction is quite close to 1. In the context of family names, this means that the ones with already a large number of representatives in the population are at a distinct advantage, as the probability that they die out by chance is much lower than that of those with only a few representatives. Thus, common family names become ever more common, especially in societies that have used family names for a long time. The most famous example of this phenomenon is in Korea, where three family names (Kim, Lee, and Park in English transcriptions) account for about 45% of the population.

Example 14.2. Assume that

$$p_k = p^k(1-p),\; k = 0, 1, 2, \ldots.$$

This means that the offspring distribution is Geometric$(1-p)$ minus 1. Thus,

$$\mu = \frac{1}{1-p} - 1 = \frac{p}{1-p}$$

and, if $p \le \frac{1}{2}$, $\pi_0 = 1$. Now suppose that $p > \frac{1}{2}$. Then, we have to compute

$$\begin{aligned} \phi(s) &= \sum_{k=0}^{\infty} s^k p^k (1-p) \\ &= \frac{1-p}{1-ps}. \end{aligned}$$

The equation $\phi(s) = s$ has two solutions, $s = 1$ and $s = \frac{1-p}{p}$. Thus, when $p > \frac{1}{2}$,

$$\pi_0 = \frac{1-p}{p}.$$

Example 14.3. Assume that the offspring distribution is Binomial$(3, \frac{1}{2})$. Compute π_0.

As $\mu = \frac{3}{2} > 1$, π_0 is given by

$$\phi(s) = \frac{1}{8} + \frac{3}{8}s + \frac{3}{8}s^2 + \frac{1}{8}s^3 = s,$$

with solutions $s = 1$, $-\sqrt{5} - 2$, and $\sqrt{5} - 2$. The one that lies in $(0, 1)$, $\sqrt{5} - 2 \approx 0.2361$, is the probability π_0.

Problems

1. For a branching process with offspring distribution given by $p_0 = \frac{1}{6}, p_1 = \frac{1}{2}, p_3 = \frac{1}{3}$, determine (a) the expectation and variance of X_9, the population at generation 9, (b) the probability that the branching process dies by generation 3, but not by generation 2, and (c) the probability that

the process ever dies out. Then, assume that you start 5 independent copies of this branching process at the same time (equivalently, change X_0 to 5), and (d) compute the probability that the process ever dies out.

2. Assume that the offspring distribution of a branching process is Poisson with parameter λ. (a) Determine the expected *combined* population through generation 10. (b) Determine, with the aid of computer if necessary, the probability that the process ever dies out for $\lambda = \frac{1}{2}$, $\lambda = 1$, and $\lambda = 2$.

3. Assume that the offspring distribution of a branching process is given by $p_1 = p_2 = p_3 = \frac{1}{3}$. Note that $p_0 = 0$. Solve the following problem for $a = 1, 2, 3$. Let Y_n be the proportion of individuals in generation n (out of the total number of X_n individuals) from families of size a. (A family consists of individuals that are offspring of the same parent from the previous generation.) Compute the limit of Y_n as $n \to \infty$.

Solutions to problems

1. For (a), compute $\mu = \frac{3}{2}$, $\sigma^2 = \frac{7}{2} - \frac{9}{4} = \frac{5}{4}$, and plug into the formula. Then compute

$$\phi(s) = \frac{1}{6} + \frac{1}{2}s + \frac{1}{3}s^3.$$

For (b),

$$P(X_3 = 0) - P(X_2 = 0) = \phi(\phi(\phi(0))) - \phi(\phi(0)) \approx 0.0462.$$

For (c), we solve $\phi(s) = s$, $0 = 2s^3 - 3s + 1 = (s-1)(2s^2 + 2s - 1)$, and so $\pi_0 = \frac{\sqrt{3}-1}{2} \approx 0.3660$. For (d), the answer is π_0^5.

2. For (a), $\mu = \lambda$ and

$$E(X_0 + X_1 + \ldots + X_{10}) = EX_0 + EX_1 + \ldots + EX_{10} = 1 + \lambda + \cdots + \lambda^{10} = \frac{\lambda^{11} - 1}{\lambda - 1},$$

if $\lambda \neq 1$, and 11 if $\lambda = 1$. For (b), if $\lambda \leq 1$ then $\pi_0 = 1$, but if $\lambda > 1$, then π_0 is the solution for $s \in (0, 1)$ to

$$e^{\lambda(s-1)} = s.$$

This equation cannot be solved analytically, but we can numerically obtain the solution for $\lambda = 2$ to get $\pi_0 \approx 0.2032$.

3. Assuming $X_{n-1} = k$, the number of families at time n is also k. Each of these has, independently, a members with probability p_a. If k is large — which it will be for large n, as the branching process cannot die out — then, with overwhelming probability, the number of children in such families is about ap_ak, while X_n is about μk. Then, the proportion Y_n is about $\frac{ap_a}{\mu}$, which works out to be $\frac{1}{6}$, $\frac{1}{3}$, and $\frac{1}{2}$, for $a = 1$, 2, and 3.

15 Markov Chains: Limiting Probabilities

Example 15.1. Assume that the transition matrix is given by

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.6 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Recall that the n-step transition probabilities are given by powers of P. Let us look at some large powers of P, beginning with

$$P^4 = \begin{bmatrix} 0.5401 & 0.4056 & 0.0543 \\ 0.5412 & 0.4048 & 0.054 \\ 0.54 & 0.408 & 0.052 \end{bmatrix}.$$

Then, to four decimal places,

$$P^8 \approx \begin{bmatrix} 0.5405 & 0.4054 & 0.0541 \\ 0.5405 & 0.4054 & 0.0541 \\ 0.5405 & 0.4054 & 0.0541 \end{bmatrix}.$$

and subsequent powers are the same to this precision.

The matrix elements appear to converge and the rows become almost identical. Why? What determines the limit? These questions will be answered in this chapter.

We say that a state $i \in S$ has *period* $d \geq 1$ if (1) $P_{ii}^n > 0$ implies that $d|n$ and (2) d is the largest positive integer that satisfies (1).

Example 15.2. Simple random walk on $\mathbb{Z}$, with $p \in (0, 1)$. The period of any state is 2 because the walker can return to her original position in any even number of steps, but in no odd number of steps.

Example 15.3. Random walk on the vertices of a square. Again, the period of any state is 2 for the same reason.

Example 15.4. Random walk on the vertices of a triangle. The period of any state is 1 because the walker can return in two steps (one step out and then back) or three steps (around the triangle).

Example 15.5. *Deterministic cycle*. If a chain has n states $0, 1, \ldots, n-1$ and transitions from i to $(i + 1) \mod n$ with probability 1, then it has period n. So, any period is possible.

However, if the following two transition probabilities are changed: $P_{01} = 0.9$ and $P_{00} = 0.1$, then the chain has period 1. In fact, the period of any state i with $P_{ii} > 0$ is trivially 1.

It can be shown that the period is the same for all states in the same class. If a state, and therefore its class, has period 1, it is called *aperiodic*. If the chain is irreducible, we call the entire chain aperiodic if all states have period 1.

For a state $i \in S$, let

$$R_i = \inf\{n \geq 1 : X_n = i\}$$

be the first time, after time 0, that the chain is at $i \in S$. Also, let

$$f_i^{(n)} = P(R_i = n | X_0 = i)$$

be the p. m. f. of R_i when the starting state is i itself (in which case we may call R_i the *return time*). We can connect these to the familiar quantity

$$f_i = P(\text{ever reenter } i | X_0 = i) = \sum_{n=1}^{\infty} f_i^{(n)},$$

so that the state i is recurrent exactly when $\sum_{n=1}^{\infty} f_i^{(n)} = 1$. Then, we define

$$m_i = E[R_i | X_0 = i] = \sum_{n=1}^{\infty} n f_i^{(n)}.$$

If the above series converges, i.e., $m_i < \infty$, then we say that i is *positive recurrent*. It can be shown that positive recurrence is also a class property: a state shares it with all members of its class. Thus, an irreducible chain is positive recurrent if each of its states is.

It is not hard to show that a finite irreducible chain is positive recurrent. In this case, there must exist an $m \geq 1$ and an $\epsilon > 0$ so that i can be reached from any j in at most m steps with probability at least ϵ. Then, $P(R_i \geq n) \leq (1-\epsilon)^{\lfloor n/m \rfloor}$, which goes to 0 geometrically fast.

We now state the key theorems. Some of these have rather involved proofs (although none is exceptionally difficult), which we will merely sketch or omit altogether.

Theorem 15.1. *Proportion of the time spent at i.*

Assume that the chain is irreducible and positive recurrent. Let $N_n(i)$ be the number of visits to i in the time interval from 0 through n. Then,

$$\lim_{n \to \infty} \frac{N_n(i)}{n} = \frac{1}{m_i},$$

in probability.

Proof. The idea is quite simple: once the chain visits i, it returns, on average, once per m_i time steps, hence the proportion of time spent there is $1/m_i$. We skip a more detailed proof. □

A vector of probabilities, π_i, $i \in S$, such that $\sum_{i \in S} \pi_i = 1$ is called an *invariant*, or *stationary*, *distribution* for a Markov chain with transition matrix P if

$$\sum_{i \in S} \pi_i P_{ij} = \pi_j, \text{ for all } j \in S.$$

In matrix form, if we put π into a row vector $[\pi_1, \pi_2, \ldots]$, then

$$[\pi_1, \pi_2, \ldots] \cdot P = [\pi_1, \pi_2, \ldots].$$

Thus $[\pi_1, \pi_2, \ldots]$ is a *left eigenvector* of P, for eigenvalue 1. More important for us is the following probabilistic interpretation. If π_i is the p. m. f. for X_0, that is, $P(X_0 = i) = \pi_i$, for all $i \in S$, it is also the p. m. f. for X_1 and hence for all other X_n, that is, $P(X_n = i) = \pi_i$, for all n.

Theorem 15.2. *Existence and uniqueness of invariant distributions.*

An irreducible positive recurrent Markov chain has a unique invariant distribution, which is given by

$$\pi_i = \frac{1}{m_i}.$$

In fact, an irreducible chain is positive recurrent if and only if a stationary distribution exists.

The formula for π should not be a surprise: if the probability that the chain is at i is always π_i, then one should expect that the proportion of time spent at i, which we already know to be $1/m_i$, to be equal to π_i. We will not, however, go deeper into the proof.

Theorem 15.3. *Convergence to invariant distribution.*

If a Markov chain is irreducible, aperiodic, and positive recurrent, then, for every $i, j \in S$,

$$\lim_{n\to\infty} P_{ij}^n = \pi_j.$$

Recall that $P_{ij}^n = P(X_n = j | X_0 = i)$ and note that the limit is independent of the initial state. Thus, the rows of P^n are more and more similar to the row vector π as n becomes large.

The most elegant proof of this theorem uses *coupling*, an important idea first developed by a young French probabilist Wolfgang Doeblin in the late 1930s. (Doeblin's life is a romantic, and quite tragic, story. An immigrant from Germany, he died as a soldier in the French army in 1940, at the age of 25. He made significant mathematical contributions during his army service.) Start with two *independent* copies of the chain — two particles moving from state to state according to the transition probabilities — one started from i, the other using the initial distribution π. Under the stated assumptions, they will eventually meet. Afterwards, the two particles move together in unison, that is, they are *coupled*. Thus, the difference between the two probabilities at time n is bounded above by twice the probability that coupling does not happen by time n, which goes to 0. We will not go into greater detail, but, as we will see in the next example, periodicity is necessary.

Example 15.6. A deterministic cycle with $a = 3$ has the transition matrix

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

This is an irreducible chain with the invariant distribution $\pi_0 = \pi_1 = \pi_2 = \frac{1}{3}$ (as it is easy to check). Moreover,

$$P^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

$P^3 = I$, $P^4 = P$, etc. Although the chain does spend $1/3$ of the time at each state, the transition probabilities are a periodic sequence of 0's and 1's and do not converge.

Our final theorem is mostly a summary of the results for the special, and for us the most common, case.

Theorem 15.4. *Convergence theorem for a finite state space S.*

Assume that s Markov chain with a finite state space is irreducible.

1. *There exists a unique invariant distribution given by $\pi_i = \frac{1}{m_i}$.*
2. *For every i, irrespective of the initial state,*
$$\frac{1}{n} N_n(i) \to \pi_i,$$
in probability.
3. *If the chain is also aperiodic, then, for all i and j,*
$$P_{ij}^n \to \pi_j.$$
4. *If the chain is periodic with period d, then, for every pair $i, j \in S$, there exists an integer r, $0 \le r \le d-1$, so that*
$$\lim_{m\to\infty} P_{ij}^{md+r} = d\pi_j$$
and so that $P_{ij}^n = 0$ for all n such that $n \ne r \mod d$.

Example 15.7. We begin by our first example, Example 15.1. That was clearly an irreducible and aperiodic (note that $P_{00} > 0$) chain. The invariant distribution $[\pi_1, \pi_2, \pi_3]$ is given by

$$\begin{aligned} 0.7\pi_1 + 0.4\pi_2 \quad & = \pi_1 \\ 0.2\pi_1 + 0.6\pi_2 \quad + \pi_3 & = \pi_2 \\ 0.1\pi_1 \quad & = \pi_3 \end{aligned}$$

This system has infinitely many solutions and we need to use

$$\pi_1 + \pi_2 + \pi_3 = 1$$

to get the unique solution

$$\pi_1 = \frac{20}{37} \approx 0.5405,\ \pi_2 = \frac{15}{37} \approx 0.4054,\ \pi_3 = \frac{2}{37} \approx 0.0541.$$

Example 15.8. *The general two-state Markov chain.* Here $S = \{1, 2\}$ and

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

and we assume that $0 < \alpha, \beta < 1$.

$$\begin{aligned} \alpha\pi_1 + \beta\pi_2 &= \pi_1 \\ (1-\alpha)\pi_1 + (1-\beta)\pi_2 &= \pi_2 \\ \pi_1 + \pi_2 &= 1 \end{aligned}$$

and, after some algebra,

$$\begin{aligned} \pi_1 &= \frac{\beta}{1+\beta-\alpha}, \\ \pi_2 &= \frac{1-\alpha}{1-\beta+\alpha}. \end{aligned}$$

Here are a few common follow-up questions:

- Start the chain at 1. In the long run, what proportion of time does the chain spend at 2? Answer: π_2 (and this does not depend on the starting state).
- Start the chain at 2. What is the expected return time to 2? Answer: $\frac{1}{\pi_2}$.
- In the long run, what proportion of time is the chain at 2, while at the previous time it was at 1? Answer: $\pi_1 P_{12}$, as it needs to be at 1 at the previous time and then make a transition to 2 (again, the answer does not depend on the starting state).

Example 15.9. In this example we will see how to compute the average length of time a chain remains in a subset of states, once the subset is entered. Assume that a machine can be in 4 states labeled 1, 2, 3, and 4. In states 1 and 2 the machine is up, working properly. In states 3 and 4 the machine is down, out of order. Suppose that the transition matrix is

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix}.$$

(a) Compute the average length of time the machine remains up after it goes up. (b) Compute the proportion of time that the system is up, but down at the next time step (this is called the *breakdown rate*).

We begin with computing the invariant distribution, which works out to be $\pi_1 = \frac{9}{48}$, $\pi_2 = \frac{12}{48}$, $\pi_3 = \frac{14}{48}$, $\pi_4 = \frac{13}{48}$. Then, the breakdown rate is

$$\pi_1(P_{13} + P_{14}) + \pi_2(P_{23} + P_{24}) = \frac{9}{32},$$

the answer to (b).

Now, let u be the average stretch of time the machine remains up and let d be the average stretch of time it is down. We need to compute u to answer (a). We will achieve this by figuring out the two equations that u and d must satisfy. For the first equation, we use that the proportion of time the system is up is

$$\frac{u}{u+d} = \pi_1 + \pi_2 = \frac{21}{48}.$$

For the second equation, we use that there is a single breakdown in every interval of time consisting of the stretch of up time followed by the stretch of down time, i.e., from one repair to the next repair. This means

$$\frac{1}{u+d} = \frac{9}{32},$$

the breakdown rate from (b). The system of two equations gives $d = 2$ and, to answer (a), $u = \frac{14}{9}$.

Computing the invariant distribution amounts to solving a system of linear equations. Nowadays this is not difficult to do, even for enormous systems; still, it is worthwhile to observe that there are cases when the invariant distribution is easy to identify.

We call a square matrix with nonnegative entries *doubly stochastic* if the sum of the entries in each row and in each column is 1.

Proposition 15.5. *Invariant distribution in a doubly stochastic case.*

If the transition matrix for an irreducible Markov chain with a finite state space S is doubly stochastic, its (unique) invariant measure is uniform over S.

Proof. Assume that $S = \{1, \ldots, m\}$, as usual. If $[1, \ldots, 1]$ is the row vector with m 1's, then $[1, \ldots, 1]P$ is exactly the vector of column sums, thus $[1, \ldots, 1]$. This vector is preserved by right multiplication by P, as is $\frac{1}{m}[1, \ldots, 1]$. This vector specifies the uniform p. m. f. on S. □

Example 15.10. *Simple random walk on a circle.* Pick a probability $p \in (0, 1)$. Assume that a points labeled $0, 1, \ldots, a-1$ are arranged on a circle clockwise. From i, the walker moves to $i+1$ (with a identified with 0) with probability p and to $i-1$ (with -1 identified with $a-1$)

with probability $1-p$. The transition matrix is

$$P = \begin{bmatrix} 0 & p & 0 & 0 & \dots & 0 & 0 & 1-p \\ 1-p & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & \dots & 0 & 0 & 0 \\ & & & \dots & & & & \\ 0 & 0 & 0 & 0 & \dots & 1-p & 0 & p \\ p & 0 & 0 & 0 & \dots & 0 & 1-p & 0 \end{bmatrix}$$

and is doubly stochastic. Moreover, the chain is aperiodic if a is odd and otherwise periodic with period 2. Therefore, the proportion of time the walker spends at any state is $\frac{1}{a}$, which is also the limit of P_{ij}^n for all i and j if a is odd. If a is even, then $P_{ij}^n = 0$ if $(i-j)$ and n have a different parity, while if they are of the same parity, $P_{ij}^n \to \frac{2}{a}$.

Assume that we change the transition probabilities a little: assume that, only when the walker is at 0, she stays at 0 with probability $r \in (0,1)$, moves to 1 with probability $(1-r)p$, and to $a-1$ with probability $(1-r)(1-p)$. The other transition probabilities are unchanged. Clearly, now the chain is aperiodic for any a, but the transition matrix is no longer doubly stochastic. What happens to the invariant distribution?

The walker spends a longer time at 0; if we stop the clock while she stays at 0, the chain is the same as before and spends an equal an proportion of time at all states. It follows that our perturbed chain spends the same proportion of time at all states except 0, where it spends a Geometric$(1-r)$ time at every visit. Therefore, π_0 is larger by the factor $\frac{1}{1-r}$ than other π_i. Thus, the row vector with invariant distributions is

$$\frac{1}{\frac{1}{1-r}+a-1}\begin{bmatrix}\frac{1}{1-r} & 1 & \dots & 1\end{bmatrix} = \begin{bmatrix}\frac{1}{1+(1-r)(a-1)} & \frac{1-r}{1+(1-r)(a-1)} & \dots & \frac{1-r}{1+(1-r)(a-1)}\end{bmatrix}.$$

Thus, we can still determine the invariant distribution if only the self-transition probabilities P_{ii} are changed.

Problems

1. Consider the chain in Problem 2 of Chapter 12. (a) Determine the invariant distribution. (b) Determine $\lim_{n\to\infty} P_{10}^n$. Why does it exist?

2. Consider the chain in Problem 4 of Chapter 12 with the same initial state. Determine the proportion of time the walker spends at a.

3. Roll a fair die n times and let S_n be the sum of the numbers you roll. Determine, with proof, $\lim_{n\to\infty} P(S_n \mod 13 = 0)$.

4. Peter owns two pairs of running shoes. Each morning he goes running. He is equally likely

to leave from his front or back door. Upon leaving the house, he chooses a pair of running shoes at the door from which he leaves or goes running barefoot if there are no shoes there. On his return, he is equally likely to enter at either door and leaves his shoes (if any) there. (a) What proportion of days does he run barefoot? (b) What proportion of days is there at least one pair of shoes at the *front* door (before he goes running)? (c) Now, assume also that the pairs of shoes are green and red and that he chooses a pair at random if he has a choice. What proportion of mornings does he run in green shoes?

5. Prof. Messi does one of three service tasks every year, coded as 1, 2, and 3. The assignment changes randomly from year to year as a Markov chain with transition matrix

$$\begin{bmatrix} \frac{7}{10} & \frac{1}{5} & \frac{1}{10} \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \\ \frac{1}{10} & \frac{2}{5} & \frac{1}{5} \end{bmatrix}.$$

Determine the proportion of years that Messi has the same assignment as the previous two years.

6. Consider the Markov chain with states $0, 1, 2, 3, 4$, which transition from state $i > 0$ to one of the states $0, \dots, i-1$ with equal probability, and transition from 0 to 4 with probability 1. Show that all P_{ij}^n converge as $n \to \infty$ and determine the limits.

Solutions to problems

1. The chain is irreducible and aperiodic. Moreover, (a) $\pi = \left[\frac{10}{21}, \frac{5}{21}, \frac{6}{21}\right]$ and (b) the limit is $\pi_0 = \frac{10}{21}$.

2. The chain is irreducible and aperiodic. Moreover, $\pi = \left[\frac{5}{41}, \frac{8}{41}, \frac{8}{41}, \frac{20}{41}\right]$ and the answer is $\pi_1 + \pi_3 = \frac{13}{41}$.

3. Consider $S_n \mod 13$. This is a Markov chain with states $0, 1 \dots, 12$ and transition matrix is

$$\begin{bmatrix} 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \dots & 0 \\ \dots \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & \dots & 0 \end{bmatrix}.$$

(To get the next row, shift cyclically to the right.) This is a doubly stochastic matrix with $\pi_i = \frac{1}{13}$, for all i. So the answer is $\frac{1}{13}$.

4. Consider the Markov chain with states given by the number of shoes at the front door. Then

$$P = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{bmatrix}.$$

This is a doubly stochastic matrix with $\pi_0 = \pi_1 = \pi_2 = \frac{1}{3}$. Answers: (a) $\pi_0 \cdot \frac{1}{2} + \pi_2 \cdot \frac{1}{2} = \frac{1}{3}$; (b) $\pi_1 + \pi_2 = \frac{2}{3}$; (c) $\pi_0 \cdot \frac{1}{4} + \pi_2 \cdot \frac{1}{4} + \pi_1 \cdot \frac{1}{2} = \frac{1}{3}$.

5. Solve for the invariant distribution: $\pi = \left[\frac{6}{17}, \frac{7}{17}, \frac{4}{17}\right]$. The answer is $\pi_1 \cdot P_{11}^2 + \pi_2 \cdot P_{22}^2 + \pi_3 \cdot P_{33}^2 = \frac{19}{50}$.

6. As the chain is irreducible and aperiodic, P_{ij}^n converges to π_j, $j = 0, 1, 2, 3, 4$, where π is given by $\pi = \left[\frac{12}{37}, \frac{6}{37}, \frac{4}{37}, \frac{3}{37}, \frac{12}{37}\right]$.

Interlude: Practice Midterm 2

This practice exam covers the material from chapters 12 through 15. Give yourself 50 minutes to solve the four problems, which you may assume have equal point score.

1. Suppose that whether it rains on a day or not depends on whether it did so on the previous two days. If it rained yesterday and today, then it will rain tomorrow with probability 0.7; if it rained yesterday but not today, it will rain tomorrow with probability 0.4; if it did not rain yesterday, but rained today, it will rain tomorrow with probability 0.5; if it did not rain yesterday nor today, it will rain tomorrow with probability 0.2.

(a) Let X_n be the Markov chain with 4 states, $(R, R), (N, R), (R, N), (N, N)$, which code (weather yesterday, weather today) with R = rain and N = no rain. Write down the transition probability matrix for this chain. (b) Today is Wednesday and it is raining. It also rained yesterday. Explain how you would compute the probability that it will rain on Saturday. Do not carry out the computation.

(c) Under the same assumption as in (b), explain how you would approximate the probability of rain on a day exactly a year from today. Carefully justify your answer, but do not carry out the computation.

2. Consider the Markov chain with states 1, 2, 3, 4, 5, given by the following transition matrix:

$$P = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Specify the classes and determine whether they are transient or recurrent.

3. In a branching process the number of descendants is determined as follows. An individual first tosses a coin that comes out Heads with probability p. If this coin comes out Tails, the individual has no descendants. If the coin comes Heads, the individual has 1 or 2 descendants, each with probability $\frac{1}{2}$.

(a) Compute π_0, the probability that the branching process eventually dies out. Your answer will, of course, depend on the parameter p.

(b) Write down the expression for the probability that the branching process is still alive at generation 3. Do not simplify.

4. A random walker is in one of the four states, 0, 1, 2, or 3. If she is at i at some time, she makes the following transition. With probability $\frac{1}{2}$ she moves from i to $(i+1) \mod 4$ (that is, if she is at 0 she moves to 1, from 1 she moves to 2, from 2 to 3, and from 3 to 0). With probability $\frac{1}{2}$, she moves to a random state among the four states, each chosen with equal probability.

(a) Show that this chain has a unique invariant distribution and compute it. (Take a good look at the transition matrix before you start solving this).

(b) After the walker makes many steps, compute the proportion of time she spends at 1. Does the answer depend on the chain's starting point?

(c) After the walker makes many steps, compute the proportion of times she is at the same state as at the previous time.

Solutions to Practice Midterm 2

1. Suppose that whether it rains on a day or not depends on whether it did so on the previous two days. If it rained yesterday and today, then it will rain tomorrow with probability 0.7; if it rained yesterday but not today, it will rain tomorrow with probability 0.4; if it did not rain yesterday, but rained today, it will rain tomorrow with probability 0.5; if it did not rain yesterday nor today, it will rain tomorrow with probability 0.2.

   (a) Let X_n be the Markov chain with 4 states, $(R,R),(N,R),(R,N),(N,N)$, which code (weather yesterday, weather today) with $R=$ rain and $N=$ no rain. Write down the transition probability matrix for this chain.

   Solution:
   Let (R,R) be state 1, (N,R) state 2, (R,N) state 3, and (N,N) state 4. The transition matrix is
$$P=\begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix}.$$

   (b) Today is Wednesday and it is raining. It also rained yesterday. Explain how you would compute the probability that it will rain on Saturday. Do not carry out the computation.

   Solution: If Wednesday is time 0, then Saturday is time 3. The initial state is given by the row $[1,0,0,0]$ and it will rain on Saturday if we end up at state 1 or 2. Therefore, our solution is
$$\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}\cdot P^3\cdot \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$
   that is, the sum of the first two entries of $\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}\cdot P^3$.

   (c) Under the same assumption as in (b), explain how you would approximate the probability of rain on a day exactly a year from today. Carefully justify your answer, but do not carry out the computation.

   Solution:
   The matrix P is irreducible since the chain makes the following transitions with positive probability: $(R,R)\to(R,N)\to(N,N)\to(N,R)\to(R,R)$. It is also

aperiodic because the transition $(R, R) \to (R, R)$ has positive probability. Therefore, the probability can be approximated by $\pi_1 + \pi_2$, where $[\pi_1, \pi_2, \pi_3, \pi_4]$ is the unique solution to $[\pi_1, \pi_2, \pi_3, \pi_4] \cdot P = [\pi_1, \pi_2, \pi_3, \pi_4]$ and $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$.

2. Consider the Markov chain with states 1, 2, 3, 4, 5, given by the following transition matrix:

$$P = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Specify all classes and determine whether they are transient or recurrent.

Solution:
Answer:

- $\{2\}$ is transient;
- $\{1, 3\}$ is recurrent;
- $\{4, 5\}$ is recurrent.

3. In a branching process, the number of descendants is determined as follows. An individual first tosses a coin that comes out Heads with probability p. If this coin comes out Tails, the individual has no descendants. If the coin comes Heads, the individual has 1 or 2 descendants, each with probability $\frac{1}{2}$.

   (a) Compute π_0, the probability that the branching process eventually dies out. Your answer will, of course, depend on the parameter p.

   Solution:
   The probability mass function for the number of descendants is

$$\begin{pmatrix} 0 & 1 & 2 \\ 1-p & \frac{p}{2} & \frac{p}{2} \end{pmatrix},$$

   and so

$$E(\text{number of descendants}) = \frac{p}{2} + p = \frac{3p}{2}.$$

   If $\frac{3p}{2} \le 1$, i.e., $p \le \frac{2}{3}$, then $\pi_0 = 1$. Otherwise, we need to compute $\phi(s)$ and solve $\phi(s) = s$. We have

$$\phi(s) = 1 - p + \frac{p}{2}s + \frac{p}{2}s^2.$$

   Then,

$$\begin{aligned} s &= 1 - p + \frac{p}{2}s + \frac{p}{2}s^2, \\ 0 &= ps^2 + (p-2)s + 2(1-p), \\ 0 &= (s-1)(ps - 2(1-p)). \end{aligned}$$

We conclude that $\pi_0 = \frac{2(1-p)}{p}$, if $p > \frac{2}{3}$.

(b) Write down the expression for the probability that the branching process is still alive at generation 3. Do not simplify.

Solution:
The answer is $1 - \phi(\phi(\phi(0)))$ and we compute

$$\begin{aligned}
\phi(0) &= 1 - p, \\
\phi(\phi(0)) &= 1 - p + \frac{p}{2}(1-p) + \frac{p}{2}(1-p)^2, \\
1 - \phi(\phi(\phi(0))) &= 1 - \Big(1 - p + \frac{p}{2}(1 - p + \frac{p}{2}(1-p) + \\
&\quad \frac{p}{2}(1-p)^2) + \frac{p}{2}(1 - p + \frac{p}{2}(1-p) + \frac{p}{2}(1-p)^2)^2\Big).
\end{aligned}$$

4. A random walker is at one of the four states, 0, 1, 2, or 3. If she at i at some time, she makes the following transition. With probability $\frac{1}{2}$ she moves from i to $(i+1) \mod 4$ (that is, if she is at 0 she moves to 1, from 1 she moves to 2, from 2 to 3, and from 3 to 0). With probability $\frac{1}{2}$, she moves to a random state among the four states, each chosen with equal probability.

(a) Show that this chain has a unique invariant distribution and compute it. (Take a good look at the transition matrix before you start solving this).

Solution:
The transition matrix is

$$P = \begin{bmatrix} \frac{1}{8} & \frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{5}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{5}{8} \\ \frac{5}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}.$$

As P is a doubly stochastic irreducible matrix, $\pi = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$ is the unique invariant distribution. (Note that irreducibility is trivial, as all entries are positive).

(b) After the walker makes many steps, compute the proportion of time she spends at 1. Does the answer depend on the chain's starting point?

Solution:
The proportion equals $\pi_1 = \frac{1}{4}$, independently of the starting point.

(c) After the walker makes many steps, compute the proportion of time she is at the same state as at the previous time.

Solution:
The probability of staying at the same state is always $\frac{1}{8}$, which is the answer.

16 Markov Chains: Reversibility

Assume that you have an irreducible and positive recurrent chain, started at its unique invariant distribution π. Recall that this means that π is the p. m. f. of X_0 and of all other X_n as well. Now suppose that, for every n, $X_0, X_1, \dots, X_n$ have the same joint p. m. f. as their time-reversal $X_n, X_{n-1}, \dots, X_0$. Then, we call the chain *reversible* — sometimes it is, equivalently, also said that its invariant distribution π is reversible. This means that a recorded simulation of a reversible chain looks the same if the "movie" is run backwards.

Is there a condition for reversibility that can be easily checked? The first thing to observe is that for the chain started at π, reversible or not, the time-reversed chain has the Markov property. This is not completely intuitively clear, but can be checked:

$$\begin{aligned}
&P(X_k = i | X_{k+1} = j, X_{k+2} = i_{k+2}, \dots, X_n = i_n) \\
&= \frac{P(X_k = i, X_{k+1} = j, X_{k+2} = i_{k+2}, \dots, X_n = i_n)}{P(X_{k+1} = j, X_{k+2} = i_{k+2}, \dots, X_n = i_n)} \\
&= \frac{\pi_i P_{ij} P_{j i_{k+2}} \cdots P_{i_{n-1} i_n}}{\pi_j P_{j i_{k+2}} \cdots P_{i_{n-1} i_n}} \\
&= \frac{\pi_i P_{ij}}{\pi_j},
\end{aligned}$$

which is an expression dependent only on i and j. For reversibility, this expression must be the same as the forward transition probability $P(X_{k+1} = i | X_k = j) = P_{ji}$. Conversely, if both the original and the time-reversed chain have the same transition probabilities (and we already know that the two start at the same invariant distribution and that both are Markov), then their p. m. f.'s must agree. We have proved the following useful result.

Theorem 16.1. *Reversibility condition.*

> *A Markov chain with invariant measure π is reversible if and only if*
>
> $$\pi_i P_{ij} = \pi_j P_{ji},$$
>
> *for all states i and j.*

Another useful fact is that once reversibility is checked, invariance is automatic.

Proposition 16.2. *Reversibility implies invariance. If a probability mass function π_i satisfies the condition in the previous theorem, then it is invariant.*

Proof. We need to check that, for every j, $\pi_j = \sum_i \pi_i P_{ij}$, and here is how we do it:

$$\sum_i \pi_i P_{ij} = \sum_i \pi_j P_{ji} = \pi_j \sum_i P_{ji} = \pi_j.$$

□

We now proceed to describe *random walks on weighted graphs*, the most easily recognizable examples of reversible chains. Assume that every undirected edge between vertices i and j in a complete graph has weight $w_{ij} = w_{ji}$; we think of edges with zero weight as not present at all. When at i, the walker goes to j with probability proportional to w_{ij}, so that

$$P_{ij} = \frac{w_{ij}}{\sum_k w_{ik}}.$$

What makes such random walks easy to analyze is the existence of a simple reversible measure. Let

$$s = \sum_{i,k} w_{ik}$$

be the sum of all weights and let

$$\pi_i = \frac{\sum_k w_{ik}}{s}.$$

To see why this is a reversible distribution, compute

$$\pi_i P_{ij} = \frac{\sum_k w_{ik}}{s} \cdot \frac{w_{ij}}{\sum_k w_{ik}} = \frac{w_{ij}}{s},$$

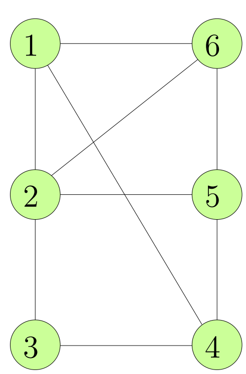
which clearly remains the same if we switch i and j.

We should observe that this chain is irreducible exactly when the graph with present edges (those with $w_{ij} > 0$) is connected. The graph can only be periodic and the period can only be 2 (because the walker can always return in two steps) when it is *bipartite*: the set of vertices V is divided into two sets V_1 and V_2 with every edge connecting a vertex from V_1 to a vertex from V_2. Finally, we note that there is no reason to forbid self-edges: some of the weights w_{ii} may be nonzero. (However, each w_{ii} appears only once in s, while each w_{ij}, with $i \neq j$, appears there twice.)

By far the most common examples have no self-edges and all nonzero weights equal to 1 — we already have a name for these cases: random walks on graphs. The number of neighbors of a vertex is commonly called its *degree*. Then, the invariant distribution is

$$\pi_i = \frac{\text{degree of } i}{2 \cdot (\text{number of all edges})}.$$

Example 16.1. Consider the random walk on the graph below.



What is the proportion of time the walk spends at vertex 2?

The reversible distribution is

$$\pi_1 = \frac{3}{18},\ \pi_2 = \frac{4}{18},\ \pi_3 = \frac{2}{18},\ \pi_4 = \frac{3}{18},\ \pi_5 = \frac{3}{18},\ \pi_6 = \frac{3}{18},$$

and, thus, the answer is $\frac{2}{9}$.

Assume now that the walker may stay at a vertex with probability P_{ii}, but when she does move she moves to a random neighbor as before. How can we choose P_{ii} so that π becomes uniform, $\pi_i = \frac{1}{6}$ for all i?

We should choose the weights of self-edges so that the sum of the weights of all edges emanating from any vertex is the same. Thus, $w_{22} = 0$, $w_{33} = 2$, and $w_{ii} = 1$, for all other i.

Example 16.2. *Ehrenfest chain.* We have M balls distributed in two urns. Each time, pick a ball at random, move it from the urn where it currently resides to the other urn. Let X_n be the number of balls in urn 1. Prove that this chain has a reversible distribution.

The nonzero transition probabilities are

$$\begin{aligned}
P_{0,1} &= 1,\\
P_{M,M-1} &= 1,\\
P_{i,i-1} &= \frac{i}{M},\\
P_{i,i+1} &= \frac{M-i}{M}.
\end{aligned}$$

Some inspiration: the invariant measure puts each ball at random into one of the two urns, as switching any ball between the two urns does not alter this assignment. Thus, π is Binomial$(M, \frac{1}{2})$,

$$\pi_i = \binom{M}{i}\frac{1}{2^M}.$$

Let us check that this is a reversible measure. The following equalities need to be verified:

$$\begin{aligned}
\pi_0 P_{01} &= \pi_1 P_{10},\\
\pi_i P_{i,i+1} &= \pi_{i+1} P_{i+1,i},\\
\pi_i P_{i,i-1} &= \pi_{i-1} P_{i-1,i},\\
\pi_M P_{M,M-1} &= \pi_{M-1} P_{M-1,M},
\end{aligned}$$

and it is straightforward to do so. Note that this chain is irreducible, but not aperiodic (it has period 2).

Example 16.3. *Markov chain Monte Carlo.* Assume that we have a very large probability space, say some subset of $S = \{0,1\}^V$, where V is a large set of n *sites*. Assume also that we have a probability measure on S given via the *energy* (sometimes called the *Hamiltonian*) function $E : S \to \mathbb{R}$. The probability of any configuration $\omega \in S$ is

$$\pi(\omega) = \frac{1}{Z}\epsilon^{-\frac{1}{T}E(\omega)}.$$

Here, $T > 0$ is the *temperature*, a parameter, and Z is the normalizing constant that makes $\sum_{\omega\in S}\pi(\omega) = 1$. Such distributions frequently occur in statistical physics and are often called *Maxwell-Boltzmann* distributions. They have numerous other applications, however, especially in optimization problems, and have yielded an optimization technique called *simulated annealing*.

If T is very large, the role of the energy is diminished and the states are almost equally likely. On the other hand, if T is very small, the large energy states have a much lower probability than the small energy ones, thus the system is much more likely to be found in the close to minimal energy states. If we want to find states with small energy, we merely choose some small T and generate at random, according to P, some states, and we have a reasonable answer. The only problem is that, although E is typically a simple function, π is very difficult to evaluate exactly, as Z is some enormous sum. (There are a few celebrated cases, called *exactly solvable systems*, in which exact computations are difficult, but possible.)

Instead of generating a random state directly, we design a Markov chain, which has π as its invariant distribution. It is common that the convergence to π is quite fast and that the necessary number of steps of the chain to get close to π is some small power of n. This is in startling contrast to the size of S, which is typically exponential in n. However, the convergence slows down at a rate exponential in T^{-1} when T is small.

We will illustrate this on the *Knapsack problem*. Assume that you are a burglar and have just broken into a jewelry store. You see a large number n of items, with weights w_i and values v_i. Your backpack (knapsack) has a weight limit b. You are faced with a question of how to fill your backpack, that is, you have to maximize the combined value of the items you will carry out

$$V = V(\omega_1, \dots, \omega_n) = \sum_{i=1}^{n} v_i\omega_i,$$

subject to the constraints that $\omega_i \in \{0, 1\}$ and that the combined weight does not exceed the backpack capacity,

$$\sum_{i=1}^{n} w_i\omega_i \le b.$$

This problem is known to be NP-hard; there is no known algorithm to solve it quickly.

The set S of *feasible solutions* $\omega = (\omega_1, \dots, \omega_n)$ that satisfy the constraints above will be our state space and the energy function E on S is given as $E = -V$, as we want to maximize V. The temperature T measures how good a solution we are happy with — the idea of simulated annealing is, in fact, a gradual lowering of the temperature to improve the solution. There is give and take: higher temperature improves the speed of convergence and lower temperature improves the quality of the result.

Finally, we are ready to specify the Markov chain (sometimes called a *Metropolis algorithm*, in honor of N. Metropolis, a pioneer in computational physics). Assume that the chain is at state ω at time t, i.e., $X_t = \omega$. Pick a coordinate i, uniformly at random. Let ω^i be the same as ω except that its ith coordinate is flipped: $\omega^i_i = 1 - \omega_i$. (This means that the status of the ith item is changed from in to out or from out to in.) If ω^i is not feasible, then $X_{t+1} = \omega$ and the state is unchanged. Otherwise, evaluate the difference in energy $E(\omega^i) - E(\omega)$ and proceed as follows:

- if $E(\omega^i) - E(\omega) \leq 0$, then make the transition to ω^i, $X_{t+1} = \omega^i$;
- if $E(\omega^i) - E(\omega) > 0$, then make the transition to ω^i with probability $e^{\frac{1}{T}(E(\omega)-E(\omega^i))}$, or else stay at ω.

Note that, in the second case, the new state has higher energy, but, in physicist's terms, we tolerate the transition because of temperature, which corresponds to the energy input from the environment.

We need to check that this chain is irreducible on S: to see this, note that we can get from any feasible solution to an empty backpack by removing object one by one, and then back by reversing the steps. Thus, the chain has a unique invariant measure, but is it the right one, that is, π? In fact, the measure π on S is reversible. We need to show that, for any pair $\omega, \omega' \in S$,

$$\pi(\omega)P(\omega, \omega') = \pi(\omega')P(\omega', \omega),$$

and this is enough to do with $\omega' = \omega^i$, for arbitrary i, and assume that both are feasible (as only such transitions are possible). Note first that the normalizing constant Z cancels out (the key feature of this method) and so does the probability $\frac{1}{n}$ that i is chosen. If $E(\omega^i) - E(\omega) \leq 0$, then the equality reduces to

$$e^{-\frac{1}{T}E(\omega)} = e^{-\frac{1}{T}E(\omega^i)} e^{\frac{1}{T}(E(\omega^i)-E(\omega))},$$

and similarly in the other case.

Problems

1. Determine the invariant distribution for the random walk in Examples 12.4 and 12.10.

2. A total of m white and m black balls are distributed into two urns, with m balls per urn. At each step, a ball is randomly selected from each urn and the two balls are interchanged. The state of this Markov chain can, thus, be described by the number of black balls in urn 1. Guess the invariant measure for this chain and prove that it is reversible.

3. Each day, your opinion on a particular political issue is either positive, neutral, or negative. If it is positive today, then it is neutral or negative tomorrow with equal probability. If it is

neutral or negative, it stays the same with probability 0.5, and, otherwise, it is equally likely to be either of the other two possibilities. Is this a reversible Markov chain?

4. A king moves on a standard 8×8 chessboard. Each time, it makes one of the available legal moves (to a horizontally, vertically or diagonally adjacent square) at random. (a) Assuming that the king starts at one of the four corner squares of the chessboard, compute the expected number of steps before it returns to the starting position. (b) Now you have two kings, they both start at the same corner square and move independently. What is, now, the expected number of steps before they simultaneously occupy the starting position?

Solutions to problems

1. Answer: $\pi = \left[\frac{1}{5}, \quad \frac{3}{10}, \quad \frac{1}{5}, \quad \frac{3}{10}\right]$.

2. If you choose m balls to put into urn 1 at random, you get

$$\pi_i = \frac{\binom{m}{i}\binom{m}{m-i}}{\binom{2m}{m}},$$

and the transition probabilities are

$$P_{i,i-1} = \frac{i^2}{m^2},\ P_{i,i+1} = \frac{(m-i)^2}{m^2},\ P_{i,i} = \frac{2i(m-i)}{m^2}.$$

Reversibility check is routine.

3. If the three states are labeled in the order given, 1, 2, and 3, then we have

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

The only way to check reversibility is to compute the invariant distribution π_1, π_2, π_3, form the diagonal matrix D with π_1, π_2, π_3 on the diagonal and to check that DP is symmetric. We get $\pi_1 = \frac{1}{5}$, $\pi_2 = \frac{2}{5}$, $\pi_3 = \frac{2}{5}$, and DP is, indeed, symmetric, so the chain is reversible.

4. This is a random walk on a graph with 64 vertices (squares) and degrees 3 (4 corner squares), 5 (24 side squares), and 8 (36 remaining squares). If i is a corner square, $\pi_i = \frac{3}{3 \cdot 4 + 5 \cdot 24 + 8 \cdot 36}$, so the answer to (a) is $\frac{420}{3}$. In (b), you have two independent chains, so $\pi_{(i,j)} = \pi_i \pi_j$ and the answer is $\left(\frac{420}{3}\right)^2$.

17 Three Applications

Parrondo's Paradox

This famous paradox was constructed by the Spanish physicist J. Parrondo. We will consider three games A, B and C with five parameters: probabilities p, p_1, p_2, and γ, and an integer period $M \geq 2$. These parameters are, for now, general so that the description of the games is more transparent. We will choose particular values once we are finished with the analysis.

We will call a game *losing* if, after playing it for a long time, a player's capital becomes more and more negative, i.e., the player loses more and more money.

Game A is very simple; in fact it is an asymmetric one-dimensional simple random walk. Win \$1, i.e., add $+1$ to your capital, with probability p, and lose a dollar, i.e., add -1 to your capital, with probability $1-p$. This is clearly a losing game if $p < \frac{1}{2}$.

In game B, the winning probabilities depend on whether your current capital is divisible by M. If it is, you add $+1$ with probability p_1, and -1 with probability $1-p_1$, and, if it is not, you add $+1$ with probability p_2 and -1 with probability $1-p_2$. We will determine below when this is a losing game.

Now consider game C, in which you, at every step, play A with probability γ and B with probability $1-\gamma$. Is it possible that A and B are losing games, while C is winning?

The surprising answer is yes! However, this should not be *so* surprising as in game B your winning probabilities depend on the capital you have and you can manipulate the proportion of time your capital spends at "unfavorable" amounts by playing the combination of the two games.

We now provide a detailed analysis. As mentioned, game A is easy. To analyze game B, take a simple random walk which makes a $+1$ step with probability p_2 and -1 step with probability $1-p_2$. Assume that you start this walk at some x, $0 < x < M$. Then, by the Gambler's ruin computation (Example 11.6),

$$P(\text{the walk hits } M \text{ before } 0) = \frac{1-\left(\frac{1-p_2}{p_2}\right)^x}{1-\left(\frac{1-p_2}{p_2}\right)^M}. \tag{17.1}$$

Starting from a multiple of M, the probability that you increase your capital by M before either decreasing it by M or returning to the starting point is

$$p_1 \cdot \frac{1-\left(\frac{1-p_2}{p_2}\right)}{1-\left(\frac{1-p_2}{p_2}\right)^M}. \tag{17.2}$$

(You have to make a step to the right and then use the formula (17.1) with $x = 1$.) Similarly, from a multiple of M, the probability that you decrease your capital by M before either increasing it by M or returning to the starting point is

$$(17.3) \qquad (1-p_1)\cdot \frac{\left(\frac{1-p_2}{p_2}\right)^{M-1} - \left(\frac{1-p_2}{p_2}\right)^{M}}{1-\left(\frac{1-p_2}{p_2}\right)^{M}}.$$

(Now you have to move one step to the left and then use $1-$(probability in (17.1) with $x = M-1$).)

The main trick is to observe that game B is losing if (17.2)<(17.3). Why? Observe your capital at multiples of M: if, starting from kM, the probability that the next (different) multiple of M you visit is $(k-1)M$ exceeds the probability that it is $(k+1)M$, then the game is losing and that is exactly when (17.2)<(17.3). After some algebra, this condition reduces to

$$(17.4) \qquad \frac{(1-p_1)(1-p_2)^{M-1}}{p_1 p_2^{M-1}} > 1.$$

Now, game C is the same as game B with p_1 and p_2 replaced by $q_1 = \gamma p + (1-\gamma)p_1$ and $q_2 = \gamma p + (1-\gamma)p_2$, yielding a winning game if

$$(17.5) \qquad \frac{(1-q_1)(1-q_2)^{M-1}}{q_1 q_2^{M-1}} < 1.$$

This is easily achieved with large enough M as soon as $p_2 < \frac{1}{2}$ and $q_2 > \frac{1}{2}$, but even for $M = 3$, one can choose $p = \frac{5}{11}$, $p_1 = \frac{1}{11^2}$, $p_2 = \frac{10}{11}$, $\gamma = \frac{1}{2}$, to get $\frac{6}{5}$ in (17.4) and $\frac{217}{300}$ in (17.5).

A Discrete Renewal Theorem

Theorem 17.1. *Assume that $f_1, \ldots, f_N \geq 0$ are given numbers with $\sum_{k=1}^N f_k = 1$. Let $\mu = \sum_{k=1}^N k\, f_k$. Define a sequence u_n as follows:*

$$\begin{aligned} u_n &= 0 && \text{if } n < 0, \\ u_0 &= 1, \\ u_n &= \sum_{k=1}^{N} f_k u_{n-k} && \text{if } n > 0. \end{aligned}$$

Assume that the greatest common divisor of the set $\{k : f_k > 0\}$ is 1. Then,

$$\lim_{n\to\infty} u_n = \frac{1}{\mu}.$$

Example 17.1. Roll a fair die forever and let S_m be the sum of outcomes of the first m rolls. Let $p_n = P(S_m \text{ ever equals } n)$. Estimate $p_{10{,}000}$.

One can write a linear recursion

$$\begin{aligned} p_0 &= 1, \\ p_n &= \frac{1}{6}\left(p_{n-1} + \cdots + p_{n-6}\right), \end{aligned}$$

and then solve it, but this is a lot of work! (Note that one should either modify the recursion for $n \le 5$ or, more easily, define $p_n = 0$ for $n < 0$.) By the above theorem, however, we can immediately conclude that p_n converges to $\frac{2}{7}$.

Example 17.2. Assume that a random walk starts from 0 and jumps from x either to $x+1$ or to $x+2$, with probability p and $1-p$, respectively. What is, approximately, the probability that the walk ever hits $10,000$? The recursion is now much simpler:

$$\begin{aligned} p_0 &= 1, \\ p_n &= p \cdot p_{n-1} + (1-p) \cdot p_{n-2}, \end{aligned}$$

and we can solve it, but again we can avoid the work by applying the theorem to get that p_n converges to $\frac{1}{2-p}$.

Proof. We can assume, without loss of generality, that $f_N > 0$ (or else reduce N).

Define a Markov chain with state space $S = \{0, 1, \ldots, N-1\}$ by

$$\begin{bmatrix} f_1 & 1-f_1 & 0 & 0 & \ldots \\ \frac{f_2}{1-f_1} & 0 & \frac{1-f_1-f_2}{1-f_1} & 0 & \ldots \\ \frac{f_3}{1-f_1-f_2} & 0 & 0 & \frac{1-f_1-f_2-f_3}{1-f_1-f_2} & \ldots \\ \ldots & & & & \\ \frac{f_N}{1-f_1-\cdots-f_{N-1}} & 0 & 0 & 0 & \ldots \end{bmatrix}.$$

This is called a *renewal chain*: it moves to the right (from x to $x+1$) on the nonnegative integers, except for *renewals*, i.e., jumps to 0. At $N-1$, the jump to 0 is certain (note that the matrix entry $P_{N-1,0}$ is 1, since the sum of f_k's is 1).

The chain is irreducible (you can get to $N-1$ from anywhere, from $N-1$ to 0, and from 0 anywhere) and we will see shortly that is also aperiodic. If $X_0 = 0$ and R_0 is the first return time to 0, then

$$P(R_0 = k)$$

clearly equals f_1, if $k = 1$. Then, for $k = 2$ it equals

$$(1-f_1) \cdot \frac{f_2}{1-f_1} = f_2,$$

and, then, for $k = 3$ it equals

$$(1-f_1) \cdot \frac{1-f_1-f_2}{1-f_1} \cdot \frac{f_3}{1-f_1-f_2} = f_3,$$

and so on. We conclude that (recall again that $X_0 = 0$)

$$P(R_0 = k) = f_k \quad \text{for all } k \geq 1.$$

In particular, the promised aperiodicity follows, as the chain can return to 0 in k steps if $f_k > 0$. Moreover, the expected return time to 0 is

$$m_{00} = \sum_{k=1}^{N} k f_k = \mu.$$

The next observation is that the probability P_{00}^n that the chain is at 0 in n steps is given by the recursion

$$P_{00}^n = \sum_{k=1}^{n} P(R_0 = k) P_{00}^{n-k}. \tag{17.6}$$

To see this, observe that you must return to 0 at some time not exceeding n in order to end up at 0; either you return for the first time at time n or you return at some previous time k and, then, you have to be back at 0 in $n - k$ steps.

The above formula (17.6) is true for every Markov chain. In this case, however, we note that the first return time to 0 is, certainly, at most N, so we can always sum to N with the proviso that $P_{00}^{n-k} = 0$ when $k > n$. So, from (17.6) we get

$$P_{00}^n = \sum_{k=1}^{N} f_k P_{00}^{n-k}.$$

The recursion for P_{00}^n is the same as the recursion for u_n. The initial conditions are also the same and we conclude that $u_n = P_{00}^n$. It follows from the convergence theorem (Theorem 15.3) that

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} P_{00}^n = \frac{1}{m_{00}} = \frac{1}{\mu},$$

which ends the proof. □

Patterns in coin tosses

Assume that you repeatedly toss a coin, with Heads represented by 1 and Tails represented by 0. On any toss, 1 occurs with probability p. Assume also that you have a *pattern* of outcomes, say 1011101. What is the expected number of tosses needed to obtain this pattern? It should be about $2^7 = 128$ when $p = \frac{1}{2}$, but what is it exactly? One can compare two patterns by this *waiting game*, saying that the one with the smaller expected value wins.

Another way to compare two patterns is the *horse race*: you and your adversary each choose a pattern, say 1001 and 0100, and the person whose pattern appears first wins.

Here are the natural questions. How do we compute the expectations in the waiting game and the probabilities in the horse race? Is the pattern that wins in the waiting game more likely to win in the horse race? There are several ways of solving these problems (a particularly elegant one uses the so called Optional Stopping Theorem for martingales), but we will use Markov chains.

The Markov chain X_n we will utilize has as the state space all patterns of length ℓ. Each time, the chain transitions into the pattern obtained by appending 1 (with probability p) or 0 (with probability $1-p$) at the right end of the current pattern and by deleting the symbol at the left end of the current pattern. That is, the chain simply keeps track of the last ℓ symbols in a sequence of tosses.

There is a slight problem before we have ℓ tosses. For now, assume that the chain starts with some particular sequence of ℓ tosses, chosen in some way.

We can immediately figure out the invariant distribution for this chain. At any time $n \geq 2\ell$ and for any pattern A with k 1's and $\ell - k$ 0's,

$$P(X_n = A) = p^k(1-p)^{\ell-k},$$

as the chain is generated by independent coin tosses! Therefore, the invariant distribution of X_n assigns to A the probability

$$\pi_A = p^k(1-p)^{\ell-k}.$$

Now, if we have two patterns B and A, denote by $N_{B\to A}$ the expected number of *additional* tosses we need to get A provided that the first tosses ended in B. Here, if A is a subpattern of B, this *does not count*, we have to actually make A in the additional tosses, although we can use *a part* of B. For example, if $B = 111001$ and $A = 110$, and the next tosses are 10, then $N_{B\to A} = 2$, and, if the next tosses are 001110, then $N_{B\to A} = 6$.

Also denote

$$E(B \to A) = E(N_{B\to A}).$$

Our initial example can, therefore, be formulated as follows: compute

$$E(\emptyset \to 1011101).$$

The convergence theorem for Markov chains guarantees that, for every A,

$$E(A \to A) = \frac{1}{\pi_A}.$$

The hard part of our problem is over. We now show how to analyze the waiting game by the example.

We know that

$$E(1011101 \to 1011101) = \frac{1}{\pi_{1011101}}.$$

However, starting with 1011101, we can only use the *overlap* 101 to help us get back to 1011101, so that

$$E(1011101 \to 1011101) = E(101 \to 1011101).$$

To get from $\emptyset$ to 1011101, we have to get first to 101 and then from there to 1011101, so that

$$E(\emptyset \to 1011101) = E(\emptyset \to 101) + E(101 \to 1011101).$$

We have reduced the problem to 101 and we iterate our method:

$$\begin{aligned} E(\emptyset \to 101) &= E(\emptyset \to 1) + E(1 \to 101) \\ &= E(\emptyset \to 1) + E(101 \to 101) \\ &= E(1 \to 1) + E(101 \to 101) \\ &= \frac{1}{\pi_1} + \frac{1}{\pi_{101}}. \end{aligned}$$

The final result is

$$\begin{aligned} E(\emptyset \to 1011101) &= \frac{1}{\pi_{1011101}} + \frac{1}{\pi_{101}} + \frac{1}{\pi_1} \\ &= \frac{1}{p^5(1-p)^2} + \frac{1}{p^2(1-p)} + \frac{1}{p}, \end{aligned}$$

which is equal to $2^7 + 2^3 + 2 = 138$ when $p = \frac{1}{2}$.

In general, the expected time $E(\emptyset \to A)$ can be computed by adding to $1/\pi_A$ all the overlaps between A and its shifts, that is, all the patterns by which A begins and ends. In the example, the overlaps are 101 and 1. The more overlaps A has, the larger $E(\emptyset \to A)$ is. Accordingly, for $p = \frac{1}{2}$, of all patterns of length ℓ, the largest expectation is $2^\ell + 2^{\ell-1} + \cdots + 2 = 2^{\ell+1} - 2$ (for constant patterns $11\ldots1$ and $00\ldots0$) and the smallest is 2^ℓ when there is no overlap at all (for example, for $100\ldots0$).

Now that we know how to compute the expectations in the waiting game, we will look at the horse race. Fix two patterns A and B and let $p_A = P(A \text{ wins})$ and $p_B = P(B \text{ wins})$. The trick is to consider the time N, the first time *one* of the two appears. Then, we can write

$$N_{\emptyset \to A} = N + I_{\{B \text{ appears before } A\}} N'_{B \to A},$$

where $N'_{B\to A}$ is the additional number of tosses we need to get to A *after* we reach B for the first time. In words, to get to A we either stop at N or go further starting from B, but the second case occurs only when B occurs before A. It is clear that $N'_{B\to A}$ has the same distribution as $N_{B\to A}$ and is independent of the event that B appears before A. (At the time B appears for the first time, what matters for $N'_{B\to A}$ is that we are at B and not whether we have seen A earlier.) Taking expectations,

$$\begin{aligned} &E(\emptyset \to A) = E(N) + p_B \cdot E(B \to A), \\ &E(\emptyset \to B) = E(N) + p_A \cdot E(A \to B), \\ &p_A + p_B = 1. \end{aligned}$$

We already know how to compute $E(\emptyset \to A)$, $E(\emptyset \to B)$, $E(B \to A)$, and $E(A \to B)$, so this is a system of three equations with three unknowns: p_A, p_B and N.

Example 17.3. Let us return to the patterns $A = 1001$ and $B = 0100$, and $p = \frac{1}{2}$, and compute the winning probabilities in the horse race.

We compute $E(\emptyset \to A) = 16 + 2 = 18$ and $E(\emptyset \to B) = 16 + 2 = 18$. Next, we compute $E(B \to A) = E(0100 \to 1001)$. First, we note that $E(0100 \to 1001) = E(100 \to 1001)$ and, then, $E(\emptyset \to 1001) = E(\emptyset \to 100) + E(100 \to 1001)$, so that $E(0100 \to 1001) = E(\emptyset \to 1001) - E(\emptyset \to 100) = 18 - 8 = 10$. Similarly, $E(A \to B) = 18 - 4 = 14$, and, then, the above three equations with three unknowns give $p_A = \frac{5}{12}$, $p_B = \frac{7}{12}$, $E(N) = \frac{73}{6}$.

We conclude with two examples, each somewhat paradoxical and thus illuminating.

Example 17.4. Consider sequences $A = 1010$ and $B = 0100$. It is straightforward to verify that $E(\emptyset \to A) = 20$, $E(\emptyset \to B) = 18$, while $p_A = \frac{9}{14}$. So, A loses in the waiting game, but wins in the horse race! What is going on? Simply, when A loses in the horse race, it loses by a lot, thereby tipping the waiting game towards B.

Example 17.5. This example concerns the horse race only. Consider the relation $\geq$ given by $A \geq B$ if $P(A \text{ beats } B) \geq 0.5$. Naively, one would expect that this relation is transitive, but this is not true! The simplest example are triples $011 \geq 100 \geq 001 \geq 011$, with probabilities $\frac{1}{2}$, $\frac{3}{4}$ and $\frac{2}{3}$.

Problems

1. Start at 0 and perform the following random walk on the integers. At each step, flip 3 fair coins and make a jump forward equal to the number of Heads (you stay where you are if you flip no Heads). Let p_n be the probability that you ever hit n. Compute $\lim_{n\to\infty} p_n$. (It is not $\frac{2}{3}$!)

2. Suppose that you have three patterns $A = 0110$, $B = 1010$, $C = 0010$. Compute the probability that A appears first among the three in a sequence of fair coin tosses.

Solutions to problems

1. The size S of the step has the p. m. f. given by $P(X = 0) = \frac{1}{8}$, $P(X = 1) = \frac{3}{8}$, $P(X = 2) = \frac{3}{8}$, $P(X = 3) = \frac{1}{8}$. Thus,

$$p_n = \frac{1}{8}p_n + \frac{3}{8}p_{n-1} + \frac{3}{8}p_{n-2} + \frac{1}{8}p_{n-3},$$

and so

$$p_n = \frac{8}{7}\left(\frac{3}{8}p_{n-1} + \frac{3}{8}p_{n-2} + \frac{1}{8}p_{n-3}\right).$$

It follows that p_n converges to the reciprocal of

$$\frac{8}{7}\left(\frac{3}{8}\cdot 1+\frac{3}{8}\cdot 2+\frac{1}{8}\cdot 3\right),$$

that is, to $E(S|S>0)^{-1}$. The answer is

$$\lim_{n\to\infty} p_n = \frac{7}{12}.$$

2. If N is the first time one of the three appears, we have

$$\begin{aligned}
&E(\emptyset\to A) = EN + p_B E(B\to A) + p_C E(C\to A)\\
&E(\emptyset\to B) = EN + p_A E(A\to B) + p_C E(C\to B)\\
&E(\emptyset\to C) = EN + p_A E(A\to C) + p_B E(B\to C)\\
&p_A + p_B + p_C = 1
\end{aligned}$$

and

$$\begin{aligned}
E(\emptyset\to A) &= 18\\
E(\emptyset\to B) &= 20\\
E(\emptyset\to C) &= 18\\
E(B\to A) &= 16\\
E(C\to A) &= 16\\
E(A\to B) &= 16\\
E(C\to B) &= 16\\
E(A\to C) &= 16\\
E(B\to C) &= 16
\end{aligned}$$

The solution is $EN=8$, $p_A=\frac{3}{8}$, $p_B=\frac{1}{4}$, and $p_C=\frac{3}{8}$. The answer is $\frac{3}{8}$.

18 Poisson Process

A *counting process* is a random process $N(t)$, $t \geq 0$, such that

1. $N(t)$ is a nonnegative integer for each t;
2. $N(t)$ is nondecreasing in t; and
3. $N(t)$ is right-continuous.

The third condition is merely a convention: if the first two events happen at $t = 2$ and $t = 3$, then $N(2) = 1$, $N(3) = 2$, $N(t) = 1$ for $t \in (2, 3)$, and $N(t) = 0$ for $t < 2$. Thus, $N(t) - N(s)$ represents the *number of events in* $(s, t]$.

A *Poisson process* with *rate* (or *intensity*) $\lambda > 0$ is a counting process $N(t)$ such that

1. $N(0) = 0$;
2. it has independent increments: if $(s_1, t_1] \bigcap (s_2, t_2] = \emptyset$, then $N(t_1) - N(s_1)$ and $N(t_2) - N(s_2)$ are independent; and
3. the number of events in any interval of length t is Poisson(λt).

In particular,

$$
\begin{aligned}
&P(N(t+s) - N(s) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!},\ k = 0, 1, 2, \ldots,\\
&E(N(t+s) - N(s)) = \lambda t.
\end{aligned}
$$

Moreover, as $h \to 0$,

$$
\begin{aligned}
&P(N(h) = 1) = e^{-\lambda h}\lambda h \sim \lambda h,\\
&P(N(h) \geq 2) = \mathcal{O}(h^2) \ll \lambda h.
\end{aligned}
$$

Thus, in small time intervals, a single event happens with probability proportional to the length of the interval; this is why λ is called the rate.

A definition as the above should be followed by the question whether the object in question exists — we may be wishing for contradictory properties. To demonstrate the existence, we will outline two constructions of *the* Poisson process. Yes, it is unique, but it would require some probabilistic sophistication to prove this, as would the proof (or even the formulation) of convergence in the first construction we are about to give. Nevertheless, it is very useful, as it makes many properties of the Poisson process almost instantly understandable.

Construction by tossing a low-probability coin very fast. Pick a large n and assume that you have a coin with (low) Heads probability $\frac{\lambda}{n}$. Toss the coin at times which are positive integer multiples of $\frac{1}{n}$ (that is, very fast) and let $N_n(t)$ be the number of Heads in $[0, t]$. Clearly, as $n \to \infty$, the number of Heads in any interval $(s, t]$ is Binomial with the number of trials $n(t-s) \pm 2$ and success probability $\frac{\lambda}{n}$; thus, it converges to Poisson($t - s$), as $n \to \infty$. Moreover, N_n has independent increments for any n and hence the same holds in the limit. We should note that the Heads probability does not need to be exactly $\frac{\lambda}{n}$, instead, it suffices that this probability converges to λ when multiplied by n. Similarly, we do not need all integer multiples of $\frac{1}{n}$; it is enough that their number in $[0, t]$, divided by n, converges to t in probability for any fixed t.

An example of a property that follows immediately is the following. Let S_k be the time of the kth (say, 3rd) event (which is a *random* time) and let $N_k(t)$ be the number of additional events within time t after time S_k. Then, $N_k(t)$ is another Poisson process, with the same rate λ, as starting to count the Heads afresh after the kth Heads gives us the same process as if we counted them from the beginning — we can *restart* a Poisson process at the time of the kth event. In fact, we can do so at any *stopping time*, a random time T with the property that $T = t$ depends only on the behavior of the Poisson process up to time t (i.e., depends on the past, but not on the future). The Poisson process, restarted at a stopping time, has the same properties as the original process started at time 0; this is called the *strong Markov property*.

As each N_k is a Poisson process, $N_k(0) = 0$, so two events in the original Poisson $N(t)$ process do not happen at the same time.

Let $T_1, T_2, \ldots$ be the *interarrival times*, where T_n is the time elapsed between $(n-1)$st and nth event. A typical example would be the times between consecutive buses arriving at a station.

Proposition 18.1. *Distribution of interarrival times:*

$$\boxed{T_1, T_2, \ldots \textit{ are independent and Exponential}(\lambda).}$$

Proof. We have

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t},$$

which proves that T_1 is Exponential(λ). Moreover, for any $s > 0$ and any $t > 0$,

$$P(T_2 > t | T_1 = s) = P(\text{no events in } (s, s+t] | T_1 = s) = P(N(t) = 0) = e^{-\lambda t},$$

as events in $(s, s+t]$ are not influenced by what happens in $[0, s]$. So, T_2 is independent of T_1 and Exponential(λ). Similarly, we can establish that T_3 is independent of T_1 and T_2 with the same distribution, and so on. $\square$

Construction by exponential interarrival times. We can use the above Proposition 18.1 for another construction of a Poisson process, which is convenient for simulations. Let $T_1, T_2, \ldots$ be i. i. d. Exponential(λ) random variables and let $S_n = T_1 + \ldots + T_n$ be the waiting time for the nth event. We define $N(t)$ to be the largest n so that $S_n \leq t$.

We know that $ES_n = \frac{n}{\lambda}$, but we can derive its density; the distribution is called Gamma(n, λ). We start with

$$P(S_n > t) = P(N(t) < n) = \sum_{j=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^j}{j!},$$

and then we differentiate to get

$$\begin{aligned}
-f_{S_n}(t) &= \sum_{j=0}^{n-1} \frac{1}{j!}(-\lambda e^{-\lambda t}(\lambda t)^j + e^{-\lambda t} j(\lambda t)^{j-1}\lambda) \\
&= \lambda e^{-\lambda t} \sum_{j=0}^{k-1} \left(-\frac{(\lambda t)^j}{j!} + \frac{(\lambda t)^{j-1}}{(j-1)!} \right) \\
&= -\lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!},
\end{aligned}$$

and so

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}.$$

Example 18.1. Consider a Poisson process with rate λ. Compute (a) E(time of the 10th event), (b) P(the 10th event occurs 2 or more time units after the 9th event), (c) P(the 10th event occurs later than time 20), and (d) P(2 events in $[1,4]$ and 3 events in $[3,5]$).

The answer to (a) is $\frac{10}{\lambda}$ by Proposition 18.1. The answer to (b) is $e^{-2\lambda}$, as one can restart the Poisson process at any event. The answer to (c) is $P(S_{10} > 20) = P(N(20) < 10)$, so we can either write the integral

$$P(S_{10} > 20) = \int_{20}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^9}{9!}\, dt,$$

or use

$$P(N(20) < 10) = \sum_{j=0}^{9} e^{-20\lambda} \frac{(20\lambda)^j}{9!}.$$

To answer (d), we condition on the number of events in $[3,4]$:

$$\begin{aligned}
&\sum_{k=0}^{2} P(2 \text{ events in } [1,4] \text{ and } 3 \text{ events in } [3,5] \,|\, k \text{ events in } [3,4]) \cdot P(k \text{ events in } [3,4]) \\
=&\sum_{k=0}^{2} P(2-k \text{ events in } [1,3] \text{ and } 3-k \text{ events in } [4,5]) \cdot P(k \text{ events in } [3,4]) \\
=&\sum_{k=0}^{2} e^{-2\lambda} \frac{(2\lambda)^{2-k}}{(2-k)!} \cdot e^{-\lambda} \frac{\lambda^{3-k}}{(3-k)!} \cdot e^{-\lambda} \frac{\lambda^k}{k!} \\
=&\, e^{-4\lambda} \left(\frac{1}{3}\lambda^5 + \lambda^4 + \frac{1}{2}\lambda^3 \right).
\end{aligned}$$

Theorem 18.2. *Superposition of independent Poisson processes.*

Assume that $N_1(t)$ and $N_2(t)$ are independent Poisson processes with rates λ_1 and λ_2. Combine them into a single process by taking the union of both sets of events or, equivalently, $N(t) = N_1(t) + N_2(t)$. This is a Poisson process with rate $\lambda_1 + \lambda_2$.

Proof. This is a consequence of the same property for Poisson random variables. □

Theorem 18.3. *Thinning of a Poisson process.*

Each event in a Poisson process $N(t)$ with rate λ is independently a Type I event with probability p; the remaining events are Type II. Let $N_1(t)$ and $N_2(t)$ be the numbers of Type I and Type II events in $[0,t]$. These are independent Poisson processes with rates λp and $\lambda(1-p)$.

The most substantial part of this theorem is independence, as the other claims follow from the thinning properties of Poisson random variables (Example 11.4).

Proof. We argue by discrete approximation. At each integer multiple of $\frac{1}{n}$, we toss two independent coins: coin A has Heads probability $\frac{p\lambda}{n}$; and coin B has Heads probability $\frac{(1-p)\lambda}{n} / \left(1 - \frac{p\lambda}{n}\right)$. Then call *discrete Type I events* the locations with coin A Heads; *discrete Type II(a) events* the locations with coin A Tails and coin B Heads; and *discrete Type II(b) events* the locations with coin B Heads. A location is a *discrete event* if it is either a Type I or a Type II(a) event.

One can easily compute that a location $\frac{k}{n}$ is a discrete event with probability $\frac{\lambda}{n}$. Moreover, given that a location is a discrete event, it is Type I with probability p. Therefore, the process of discrete events and its division into Type I and Type II(a) events determines the discrete versions of the processes in the statement of the theorem. Now, discrete Type I and Type II(a) events are not independent (for example, both cannot occur at the same location), but discrete Type I and Type II(b) events are independent (as they depend on different coins). The proof will be concluded by showing that discrete Type II(a) and Type II(b) events have the same limit as $n \to \infty$. The Type I and Type II events will then be independent as limits of independent discrete processes.

To prove the claimed asymptotic equality, observe first that the discrete schemes (a) and (b) result in a different outcome at a location $\frac{k}{n}$ exactly when *two* events occur there: a discrete Type I event and a discrete Type II(b) event. The probability that the two discrete Type II schemes differ at $\frac{k}{n}$ is thus at most $\frac{C}{n^2}$, for some constant C. This causes the expected number of such "double points" in $[0,t]$ to be at most $\frac{Ct}{n}$. Therefore, by the Markov inequality, an upper bound for the probability that there is at least one double point in $[0,t]$ is also $\frac{Ct}{n}$. This probability goes to zero, as $n \to \infty$, for any fixed t and, consequently, discrete (a) and (b) schemes indeed result in the same limit. □

Example 18.2. Customers arrive at a store at a rate of 10 per hour. Each is either male or female with probability $\frac{1}{2}$. Assume that you know that exactly 10 women entered within some hour (say, 10 to 11am). (a) Compute the probability that exactly 10 men also entered. (b) Compute the probability that at least 20 customers have entered.

Male and female arrivals are independent Poisson processes, with parameter $\frac{1}{2} \cdot 10 = 5$, so the answer to (a) is

$$e^{-5}\frac{5^{10}}{10!}.$$

The answer to (b) is

$$\sum_{k=10}^{\infty} P(k \text{ men entered}) = \sum_{k=10}^{\infty} e^{-5}\frac{5^k}{k!} = 1 - \sum_{k=0}^{9} e^{-5}\frac{5^k}{k!}.$$

Example 18.3. Assume that cars arrive at a rate of 10 per hour. Assume that each car will pick up a hitchhiker with probability $\frac{1}{10}$. You are second in line. What is the probability that you will have to wait for more than 2 hours?

Cars that pick up hitchhikers are a Poisson process with rate $10 \cdot \frac{1}{10} = 1$. For this process,

$$P(T_1 + T_2 > 2) = P(N(2) \le 1) = e^{-2}(1+2) = 3e^{-2}.$$

Proposition 18.4. *Order of events in independent Poisson processes.*

Assume that we have two independent Poisson processes, $N_1(t)$ with rate λ_1 and $N_2(t)$ with rate λ_2. The probability that n events occur in the first process before m events occur in the second process is

$$\sum_{k=n}^{n+m-1} \binom{n+m-1}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n+m-1-k}.$$

We can easily extend this idea to more than two independent Poisson processes; we will not make a formal statement, but instead illustrate by the few examples below.

Proof. Start with a Poisson process with $\lambda_1 + \lambda_2$, then independently decide for each event whether it belongs to the first process, with probability $\frac{\lambda_1}{\lambda_1+\lambda_2}$, or to the second process, with probability $\frac{\lambda_2}{\lambda_1+\lambda_2}$. The obtained processes are independent and have the correct rates. The probability we are interested in is the probability that among the first $m+n-1$ events in the combined process, n or more events belong to the first process, which is the binomial probability in the statement. □

Example 18.4. Assume that $\lambda_1 = 5$, $\lambda_2 = 1$. Then,

$$P(5 \text{ events in the first process before 1 in the second}) = \left(\frac{5}{6}\right)^5$$

and

$$P(5 \text{ events in the first process before 2 in the second}) = \sum_{k=5}^{6} \binom{6}{k} \left(\frac{5}{6}\right)^k \left(\frac{1}{6}\right)^{6-k} = \frac{11 \cdot 5^5}{6^6}.$$

Example 18.5. You have three friends, A, B, and C. Each will call you after an Exponential amount of time with expectation 30 minutes, 1 hour, and 2.5 hours, respectively. You will go out with the first friend that calls. What is the probability that you go out with A?

We could evaluate the triple integral, but we will avoid that. Interpret each call as the first event in the appropriate one of three Poisson processes with rates 2, 1, and $\frac{2}{5}$, assuming the time unit to be one hour. (Recall that the rates are inverses of the expectations.)

We will solve the general problem with rates λ_1, λ_2, and λ_3. Start with rate $\lambda_1 + \lambda_2 + \lambda_3$ Poisson process, distribute the events with probability $\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}$, $\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}$, and $\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$, respectively. The probability of A calling first is clearly $\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}$, which in our case works out to be

$$\frac{2}{2+1+\frac{2}{5}} = \frac{10}{17}.$$

Our next theorem illustrates what we can say about previous event times if we either know that their number by time t is k or we know that the kth one happens exactly at time t.

Theorem 18.5. *Uniformity of previous event times.*

1. *Given that $N(t) = k$, the conditional distribution of the interarrival times, $S_1, \dots, S_k$, is distributed as order statistics of k independent uniform variables: the set $\{S_1, \dots, S_k\}$ is equal, in distribution, to $\{U_1, \dots, U_k\}$, where U_i are independent and uniform on $[0, t]$.*
2. *Given that $S_k = t$, $S_1, \dots, S_{k-1}$ are distributed as order statistics of $k-1$ independent uniform random variables on $[0, t]$.*

Proof. Again, we discretize and the discrete counterpart is as follows. Assume that we toss a coin, with arbitrary fixed Heads probability, N times in succession and that we know that the number of Heads in these N tosses is k. Then, these Heads occur in any of the $\binom{N}{k}$ subsets (of k tosses out of a total of N) with equal probability, simply by symmetry. This is exactly the statement of the theorem, in the appropriate limit. $\square$

Example 18.6. Passengers arrive at a bus station as a Poisson process with rate λ.

(a) The only bus departs after a deterministic time T. Let W be the combined waiting time for all passengers. Compute EW.

If $S_1, S_2, \dots$ are the arrival times in $[0, T]$, then the combined waiting time is $W = T - S_1 + T - S_3 + \dots$. Recall that we denote by $N(t)$ the number of arrivals in $[0, t]$. We obtain the answer by conditioning on the value of $N(T)$: if we know that $N(T) = k$, then W is the sum of k i. i. d. uniform random variables on $[0, T]$. Hence, $W = \sum_{k=1}^{N(T)} U_k$, where $U_1, U_2, \dots$ are i. i. d. uniform on $[0, T]$ and independent of $N(T)$. By Theorem 11.1,

$$EW = \frac{\lambda T^2}{2}$$

and $\mathrm{Var}(W) = \frac{\lambda T^3}{3}$.

(b) Now two buses depart, one at T and one at $S < T$. What is EW now?

We have two independent Poisson processes in time intervals $[0, S]$ and $[S, T]$, so the answer is

$$\lambda \frac{S^2}{2} + \lambda \frac{(T-S)^2}{2}.$$

(c) Now assume T, the only bus departure time, is Exponential(μ), independent of the passengers' arrivals.

This time,

$$\begin{aligned}
EW &= \int_0^\infty E(W|T=t) f_T(t)\,dt = \int_0^\infty \lambda \frac{t^2}{2} f_T(t)\,dt = \frac{\lambda}{2} E(T^2) \\
&= \frac{\lambda}{2}(\mathrm{Var}(T) + (ET)^2) = \frac{\lambda}{2}\frac{2}{\mu^2} = \frac{\lambda}{\mu^2}.
\end{aligned}$$

(d) Finally, two buses depart as the first two events in a rate μ Poisson process.

This makes

$$EW = 2\frac{\lambda}{\mu^2}.$$

Example 18.7. You have two machines. Machine 1 has lifetime T_1, which is Exponential(λ_1), and Machine 2 has lifetime T_2, which is Exponential(λ_2). Machine 1 starts at time 0 and Machine 2 starts at a time T.

(a) Assume that T is deterministic. Compute the probability that M_1 is the first to fail.

We could compute this via a double integral (which is a good exercise!), but instead we proceed thus:

$$\begin{aligned}
P(T_1 < T_2 + T) &= P(T_1 < T) + P(T_1 \ge T, T_1 < T_2 + T) \\
&= P(T_1 < T) + P(T_1 < T_2 + T | T_1 \ge T) P(T_1 \ge T) \\
&= 1 - e^{-\lambda_1 T} + P(T_1 - T < T_2 | T_1 \ge T) e^{-\lambda_1 T} \\
&= 1 - e^{-\lambda_1 T} + P(T_1 < T_2) e^{-\lambda_1 T} \\
&= 1 - e^{-\lambda_1 T} + \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-\lambda_1 T}.
\end{aligned}$$

The key observation above is that $P(T_1 - T < T_2 | T_1 \ge T) = P(T_1 < T_2)$. Why does this hold? We can simply quote the memoryless property of the Exponential distribution, but it is instructive to make a short argument using Poisson processes. Embed the failure times into appropriate Poisson processes. Then, $T_1 \ge T$ means that no events in the first process occur during time $[0, T]$. Under this condition, $T_1 - T$ is the time of the first event of the same process restarted at T, but this restarted process is not influenced by what happened before T, so the condition (which in addition does not influence T_2) drops out.

(b) Answer the same question when T is Exponential(μ) (and, of course, independent of the machines). Now, by the same logic,

$$\begin{aligned}
P(T_1 < T_2 + T) &= P(T_1 < T) + P(T_1 \geq T, T_1 < T_2 + T) \\
&= \frac{\lambda_1}{\lambda_1 + \mu} + \frac{\lambda_1}{\lambda_1 + \lambda_2} \frac{\mu}{\lambda_1 + \lambda_2}.
\end{aligned}$$

Example 18.8. *Impatient hitchhikers*. Two people, Alice and Bob, are hitchhiking. Cars that would pick up a hitchhiker arrive as a Poisson process with rate λ_C. Alice is first in line for a ride. Moreover, after Exponential(λ_A) time, Alice quits, and after Exponential(λ_B) time, Bob quits. Compute the probability that Alice is picked up before she quits and compute the same for Bob.

Embed each quitting time into an appropriate Poisson process, call these A and B processes, and call the car arrivals C process. Clearly, Alice gets picked if the first event in the combined A and C process is a C event:

$$P(\text{Alice gets picked}) = \frac{\lambda_C}{\lambda_A + \lambda_C}.$$

Moreover,

$$\begin{aligned}
&P(\text{Bob gets picked}) \\
&= P(\{\text{at least 2 } C \text{ events before a } B \text{ event}\} \\
&\qquad \cup \{\text{at least one } A \text{ event before either a } B \text{ or a } C \text{ event,} \\
&\qquad\quad \text{and then at least one } C \text{ event before a } B \text{ event}\}) \\
&= P(\text{at least 2 } C \text{ events before a } B \text{ event}) \\
&\quad + P(\text{at least one } A \text{ event before either a } B \text{ or a } C \text{ event,} \\
&\qquad \text{and then at least one } C \text{ event before a } B \text{ event}) \\
&\quad - P(\text{at least one } A \text{ event before either a } B \text{ or a } C \text{ event,} \\
&\qquad \text{and then at least two } C \text{ events before a } B \text{ event}) \\
&= \left(\frac{\lambda_C}{\lambda_B + \lambda_C}\right)^2 \\
&\quad + \left(\frac{\lambda_A}{\lambda_A + \lambda_B + \lambda_C}\right)\left(\frac{\lambda_C}{\lambda_B + \lambda_C}\right) \\
&\quad - \left(\frac{\lambda_A}{\lambda_A + \lambda_B + \lambda_C}\right)\left(\frac{\lambda_C}{\lambda_B + \lambda_C}\right)^2 \\
&= \frac{\lambda_A + \lambda_C}{\lambda_A + \lambda_B + \lambda_C} \cdot \frac{\lambda_C}{\lambda_B + \lambda_C}.
\end{aligned}$$

This leaves us with an excellent hint that there may be a shorter way and, indeed, there is:

$$\begin{aligned}
&P(\text{Bob gets picked}) \\
&= P(\text{first event is either } A \text{ or } C\text{, and then the next event among } B \text{ and } C \text{ is } C).
\end{aligned}$$

Problems

1. An office has two clerks. Three people, A, B, and C enter simultaneously. A and B begin service with the two clerks, while C waits for the first available clerk. Assume that the service time is Exponential(λ). (a) Compute the probability that A is the last to finish the service. (b) Compute the expected time before C is finished (i.e., C's combined waiting and service time).

2. A car wash has two stations, 1 and 2, with Exponential(λ_1) and Exponential(λ_2) service times. A car enters at station 1. Upon completing the service at station 1, the car proceeds to station 2, provided station 2 is free; otherwise, the car has to wait at station 1, blocking the entrance of other cars. The car exits the wash after the service at station 2 is completed. When you arrive at the wash there is a single car at station 1. Compute the expected time before you exit.

3. A system has two server stations, 1 and 2, with Exponential(λ_1) and Exponential(λ_2) service times. Whenever a new customer arrives, any customer in the system immediately departs. Customer arrivals are a rate μ Poisson process, and a new arrival enters the system at station 1, then goes to station 2. (a) What proportion of customers complete their service? (b) What proportion of customers stay in the system for more than 1 time unit, but do not complete the service?

4. A machine needs frequent maintenance to stay on. The maintenance times occur as a Poisson process with rate μ. Once the machine receives no maintenance for a time interval of length h, it breaks down. It then needs to be repaired, which takes an Exponential(λ) time, after which it goes back on. (a) After the machine is started, find the probability that the machine will break down before receiving its first maintenance. (b) Find the expected time for the first breakdown. (c) Find the proportion of time the machine is on.

5. Assume that certain events (say, power surges) occur as a Poisson process with rate 3 per hour. These events cause damage to a certain system (say, a computer), thus, a special protecting unit has been designed. That unit now has to be removed from the system for 10 minutes for service.

(a) Assume that a single event occurring in the service period will cause the system to crash. What is the probability that the system will crash?

(b) Assume that the system will survive a single event, but two events occurring in the service period will cause it to crash. What is, now, the probability that the system will crash?

(c) Assume that a crash will not happen unless there are two events within 5 minutes of each other. Compute the probability that the system will crash.

(d) Solve (b) by assuming that the protective unit will be out of the system for a time which is exponentially distributed with expectation 10 minutes.

Solutions to problems

1. (a) This is the probability that two events happen in a rate λ Poisson process before a single event in an independent rate λ process, that is, $\frac{1}{4}$. (b) First, C has to wait for the first event in two combined Poisson processes, which is a single process with rate 2λ, and then for the service time; the answer is $\frac{1}{2\lambda} + \frac{1}{\lambda} = \frac{3}{2\lambda}$.

2. Your total time is (the time the other car spends at station 1) + (the time you spend at station 2)+(maximum of the time the other car spends at station 2 and the time you spend at station 1). If T_1 and T_2 are Exponential(λ_1) and Exponential(λ_2), then you need to compute

$$E(T_1) + E(T_2) + E(\max\{T_1, T_2\}).$$

Now use that

$$\max\{T_1, T_2\} = T_1 + T_2 - \min\{T_1, T_2\}$$

and that $\min\{T_1, T_2\}$ is Exponential($\lambda_1 + \lambda_2$), to get

$$\frac{2}{\lambda_1} + \frac{2}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}.$$

3. (a) A customer needs to complete the service at both stations before a new one arrives, thus the answer is

$$\frac{\lambda_1}{\lambda_1 + \mu} \cdot \frac{\lambda_2}{\lambda_2 + \mu}.$$

(b) Let T_1 and T_2 be the customer's times at stations 1 and 2. The event will happen if either:

- $T_1 > 1$, no newcomers during time 1, and a newcomer during time $[1, T_1]$; or
- $T_1 < 1$, $T_1 + T_2 > 1$, no newcomers by time 1, and a newcomer during time $[1, T_1 + T_2]$.

For the first case, nothing will happen by time 1, which has probability $e^{-(\mu+\lambda_1)}$. Then, after time 1, a newcomer has to appear before the service time at station 1, which has probability $\frac{\mu}{\lambda_1+\mu}$.

For the second case, conditioned on $T_1 = t < 1$, the probability is

$$e^{-\mu} e^{-\lambda_2(1-t)} \frac{\mu}{\lambda_2 + \mu}.$$

Therefore, the probability of the second case is

$$e^{-\mu} \frac{\mu}{\lambda_2 + \mu} \int_0^1 e^{-\lambda_2(1-t)} \lambda_1 e^{-\lambda_1 t}\, dt = e^{-\mu} \frac{\lambda_1 \mu}{\lambda_2 + \mu} e^{-\lambda_2} \frac{e^{\lambda_2 - \lambda_1} - 1}{\lambda_2 - \lambda_1},$$

where the last factor is 1 when $\lambda_1 = \lambda_2$. The answer is

$$e^{-(\mu+\lambda_1)}\frac{\mu}{\lambda_1+\mu} + \frac{\lambda_1\mu}{\lambda_2+\mu}e^{-(\mu+\lambda_2)}\frac{e^{\lambda_2-\lambda_1}-1}{\lambda_2-\lambda_1}.$$

4. (a) The answer is $e^{-\mu h}$. (b) Let W be the waiting time for maintenance such that the next maintenance is at least time h in the future, and let T_1 be the time of the first maintenance. Then, provided $t < h$,

$$E(W|T_1 = t) = t + EW,$$

as the process is restarted at time t. Therefore,

$$EW = \int_0^h (t + EW)\mu\, e^{-\mu t}\, dt = \int_0^h t\mu\, e^{-\mu t}\, dt + EW \int_0^h \mu\, e^{-\mu t}\, dt.$$

Computing the two integrals and solving for EW gives

$$EW = \frac{1 - \mu h e^{-\mu h} - e^{-\mu h}}{e^{-\mu h}}.$$

The answer to (b) is $EW + h$ (the machine waits for h more units before it breaks down). The answer to (c) is

$$\frac{EW + h}{EW + h + \frac{1}{\lambda}}.$$

5. Assume the time unit is 10 minutes, $\frac{1}{6}$ of an hour. The answer to (a) is

$$P(N(1/6) \geq 1) = 1 - e^{-\frac{1}{2}},$$

and to (b)

$$P(N(1/6) \geq 2) = 1 - \frac{3}{2}e^{-\frac{1}{2}}.$$

For (c), if there are 0 or 1 events in the 10 minutes, there will be no crash, but 3 or more events in the 10 minutes will cause a crash. The final possibility is exactly two events, in which case the crash will happen with probability

$$P\left(|U_1 - U_2| < \frac{1}{2}\right),$$

where U_1 and U_2 are independent uniform random variables on $[0,1]$. By drawing a picture, this probability can be computed to be $\frac{3}{4}$. Therefore,

$$\begin{aligned} P(\text{crash}) &= P(X > 2) + P(\text{crash}|X = 2)P(X = 2) \\ &= 1 - e^{-\frac{1}{2}} - \frac{1}{2}e^{-\frac{1}{2}} - e^{-\frac{1}{2}}\frac{\left(\frac{1}{2}\right)^2}{2} + \frac{3}{4}\cdot\frac{\left(\frac{1}{2}\right)^2}{2}e^{-\frac{1}{2}} \\ &= 1 - \frac{49}{32}e^{-\frac{1}{2}} \end{aligned}$$

Finally, for (d), we need to calculate the probability that two events in a rate 3 Poisson process occur before an event occurs in a rate 6 Poisson process. This probability is

$$\left(\frac{3}{3+6}\right)^2 = \frac{1}{9}.$$

Interlude: Practice Final

This practice exam covers the material from chapters 9 through 18. Give yourself 120 minutes to solve the six problems, which you may assume have equal point score.

1. You select a random number X in $[0, 1]$ (uniformly). Your friend then keeps selecting random numbers $U_1, U_2, \ldots$ in $[0, 1]$ (uniformly and independently) until he gets a number larger than $X/2$, then he stops.

(a) Compute the expected number N of times your friend selects a number.

(b) Compute the expected sum S of the numbers your friend selects.

2. You are a member of a sports team and your coach has instituted the following policy. You begin with zero warnings. After every game the coach evaluates whether you've had a discipline problem during the game; if so, he gives you a warning. After you receive two warnings (not necessarily in consecutive games), you are suspended for the next game and your warnings count goes back to zero. After the suspension, the rules are the same as at the beginning. You figure you will receive a warning after each game you play independently with probability $p \in (0, 1)$.

(a) Let the state of the Markov chain be your warning count after a game. Write down the transition matrix and determine whether this chain is irreducible and aperiodic. Compute its invariant distribution.

(b) Write down an expression for the probability that you are suspended for both games 10 and 15. Do not evaluate.

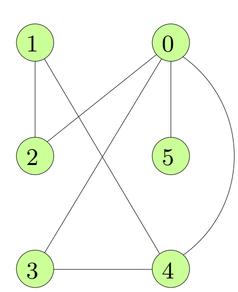
(c) Let s_n be the probability that you are suspended in the nth game. Compute $\lim_{n\to\infty} s_n$.

3. A random walker on the nonnegative integers starts at 0 and then at each step adds either 2 or 3 to her position, each with probability $\frac{1}{2}$.

(a) Compute the probability that the walker is at $2n + 3$ after making n steps.

(b) Let p_n be the probability that the walker ever hits n. Compute $\lim_{n\to\infty} p_n$.

4. A random walker is at one of the six vertices, labeled $0, 1, 2, 3, 4$, and 5, of the graph in the picture. At each time, she moves to a randomly chosen vertex connected to her current position by an edge. (All choices are equally likely and she never stays at the same position for two successive steps.)



(a) Compute the proportion of time the walker spends at 0, after she makes many steps. Does this proportion depend on the walker's starting vertex?

(b) Compute the proportion of time the walker is at an odd state (1, 3, or 5) while, previously, she was at even state (0, 2, or 4).

(c) Now assume that the walker starts at 0. What is expected number of steps she will take before she is back at 0?

5. In a branching process, an individual has two descendants with probability $\frac{3}{4}$ and no descendants with probability $\frac{1}{4}$. The process starts with a single individual in generation 0.

(a) Compute the expected number of individuals in generation 2.

(b) Compute the probability that the process ever becomes extinct.

6. Customers arrive at two service stations, labeled 1 and 2, as a Poisson process with rate λ. Assume that the time unit is one hour. Whenever a new customer arrives, any previous customer is immediately ejected from the system. A new arrival enters the service at station 1, then goes to station 2.

(a) Assume that the service time at each station is exactly 2 hours. What proportion of entering customers will complete the service (before they are ejected)?

(b) Assume that the service time at each station now is exponential with expectation 2 hours. What proportion of entering customers will now complete the service?

(c) Keep the service time assumption from (b). A customer arrives, but he is now given special treatment: he will not be ejected unless at least three or more new customers arrive during his service. Compute the probability that this special customer is allowed to complete his service.

Solutions to Practice Final

1. You select a random number X in $[0, 1]$ (uniformly). Your friend then keeps selecting random numbers $U_1, U_2, \ldots$ in $[0, 1]$ (uniformly and independently) until he gets a number larger than $X/2$, then he stops.

   (a) Compute the expected number N of times your friend selects a number.

   Solution:
   Given $X = x$, N is distributed geometrically with success probability $1 - \frac{x}{2}$, so

   $$E[N|X = x] = \frac{1}{1 - \frac{x}{2}},$$

   and so,

   $$\begin{aligned} EN &= \int_0^1 \frac{dx}{1 - \frac{x}{2}} \\ &= -2\log(1 - \frac{x}{2})|_0^1 \\ &= 2\log 2. \end{aligned}$$

   (b) Compute the expected sum S of the numbers your friend selects.

   Solution:
   Given $X = x$ and $N = n$, your friend selects $n - 1$ numbers uniformly in $[0, \frac{x}{2}]$ and one number uniformly in $[\frac{x}{2}, 1]$. Therefore,

   $$E[S|X = x, N = n] = (n-1)\frac{x}{4} + \frac{1}{2}\left(1 + \frac{x}{2}\right) = \frac{1}{4}nx + \frac{1}{2},$$

   $$E[S|X = x] = \frac{1}{4} \cdot \frac{1}{1 - \frac{x}{2}} \cdot x + \frac{1}{2} = \frac{1}{2 - x},$$

   $$ES = \int_0^1 \frac{1}{2 - x}\,dx = \log 2.$$

2. You are a member of a sports team and your coach has instituted the following policy. You begin with zero warnings. After every game the coach evaluates whether you've had a discipline problem during the game; if so, he gives you a warning. After you receive two warnings (not necessarily in consecutive games), you are suspended for the next game and your warnings count goes back to zero. After the suspension, the rules are the same as at the beginning. You figure you will receive a warning after each game you play independently with probability $p \in (0, 1)$.

(a) Let the state of the Markov chain be your warning count after a game. Write down the transition matrix and determine whether this chain is irreducible and aperiodic. Compute its invariant distribution.

Solution:
The transition matrix is

$$P = \begin{bmatrix} 1-p & p & 0 \\ 0 & 1-p & p \\ 1 & 0 & 0 \end{bmatrix},$$

and the chain is clearly irreducible (the transitions $0 \to 1 \to 2 \to 0$ happen with positive probability) and aperiodic ($0 \to 0$ happens with positive probability). The invariant distribution is given by

$$\begin{aligned} \pi_0(1-p) + \pi_2 &= \pi_0 \\ \pi_0 p + \pi_1(1-p) &= \pi_1 \\ \pi_1 p &= \pi_2 \end{aligned}$$

and

$$\pi_0 + \pi_1 + \pi_2 = 1,$$

which gives

$$\pi = \left[\frac{1}{p+2}, \frac{1}{p+2}, \frac{p}{p+2}\right].$$

(b) Write down an expression for the probability that you are suspended for both games 10 and 15. Do not evaluate.

Solution:
You must have 2 warnings after game 9 and then again 2 warnings after game 14:

$$P^9_{02} \cdot P^4_{02}.$$

(c) Let s_n be the probability that you are suspended in the nth game. Compute $\lim_{n\to\infty} s_n$.

Solution:
As the chain is irreducible and aperiodic,

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} P(X_{n-1} = 2) = \frac{p}{2+p}.$$

3. A random walker on the nonnegative integers starts at 0 and then at each step adds either 2 or 3 to her position, each with probability $\frac{1}{2}$.

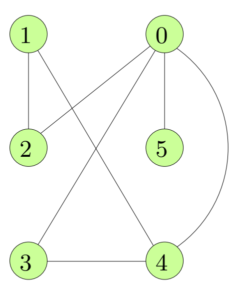
   (a) Compute the probability that the walker is at $2n+3$ after making n steps.

   Solution:
   The walker has to make $(n-3)$ 2-steps and 3 3-steps, so the answer is
   $$\binom{n}{3}\frac{1}{2^n}.$$

   (b) Let p_n be the probability that the walker ever hits n. Compute $\lim_{n\to\infty} p_n$.

   Solution:
   The step distribution is aperiodic, as the greatest common divisor of 2 and 3 is 1, so
   $$\lim_{n\to\infty} p_n = \frac{1}{\frac{1}{2}\cdot 2 + \frac{1}{2}\cdot 3} = \frac{2}{5}.$$

4. A random walker is at one of six vertices, labeled $0, 1, 2, 3, 4$, and 5, of the graph in the picture. At each time, she moves to a randomly chosen vertex connected to her current position by an edge. (All choices are equally likely and she never stays at the same position for two successive steps.)



   (a) Compute the proportion of time the walker spends at 0, after she makes many steps. Does this proportion depend on the walker's starting vertex?

   Solution:
   Independently of the starting vertex, the proportion is π_0, where $[\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5]$ is the unique invariant distribution. (Unique because of irreducibility.) This chain

is reversible with the invariant distribution given by $\frac{1}{14}[4,2,2,2,3,1]$. Therefore, the answer is $\frac{2}{7}$.

(b) Compute the proportion of times the walker is at an odd state (1, 3, or 5) while, previously, she was at even state (0, 2, or 4).

Solution:
The answer is

$$\pi_0(p_{03}+p_{05})+\pi_2\cdot p_{21}+\pi_4(p_{43}+p_{41})=\frac{2}{7}\cdot\frac{2}{4}+\frac{1}{7}\cdot\frac{1}{2}+\frac{3}{14}\cdot\frac{2}{3}=\frac{5}{14}.$$

(c) Now assume that the walker starts at 0. What is expected number of steps she will take before she is back at 0?

Solution:
The answer is

$$\frac{1}{\pi_0}=\frac{7}{2}$$

5. In a branching process, an individual has two descendants with probability $\frac{3}{4}$ and no descendants with probability $\frac{1}{4}$. The process starts with a single individual in generation 0.

(a) Compute the expected number of individuals in generation 2.

Solution:
As

$$\mu=2\cdot\frac{3}{4}+0\cdot\frac{1}{4}=\frac{3}{2},$$

the answer is

$$\mu^2=\frac{9}{4}.$$

(b) Compute the probability that the process ever goes extinct.

Solution:
As

$$\phi(s)=\frac{1}{4}+\frac{3}{4}s^2$$

the solution to $\phi(s) = s$ is given by $3s^2 - 4s + 1 = 0$, i.e., $(3s-1)(s-1) = 0$. The answer is

$$\pi_0 = \frac{1}{3}.$$

6. Customers arrive at two service stations, labeled 1 and 2, as a Poisson process with rate λ. Assume that the time unit is one hour. Whenever a new customer arrives, any previous customer is immediately ejected from the system. A new arrival enters the service at station 1, then goes to station 2.

   (a) Assume that the service time at each station is exactly 2 hours. What proportion of entering customers will complete the service (before they are ejected)?

   Solution:
   The answer is

   $$P(\text{customer served}) = P(\text{no arrival in 4 hours}) = e^{-4\lambda}.$$

   (b) Assume that the service time at each station now is exponential with expectation 2 hours. What proportion of entering customers will now complete the service?

   Solution:
   Now,

   $$\begin{aligned} &P(\text{customer served}) \\ &= P(2 \text{ or more arrivals in rate } \frac{1}{2} \text{ Poisson process} \\ &\qquad \text{before one arrival in rate } \lambda \text{ Poisson process}) \\ &= \left(\frac{\frac{1}{2}}{\frac{1}{2}+\lambda}\right)^2 = \frac{1}{(1+2\lambda)^2}. \end{aligned}$$

   (c) Keep the service time assumption from (b). A customer arrives, but he is now given special treatment: he will not be ejected unless at least three or more new customers arrive during his service. Compute the probability that this special customer is allowed to complete his service.

   Solution:
   The special customer is served exactly when 2 or more arrivals in rate $\frac{1}{2}$ Poisson

before three arrivals in rate λ Poisson process happen. Equivalently, among the first 4 arrivals in rate $\lambda+\frac{1}{2}$ Poisson process, 2 or more belong to the rate $\frac{1}{2}$ Poisson process. The answer is

$$1-\left(\frac{\lambda}{\lambda+\frac{1}{2}}\right)^4-4\left(\frac{\lambda}{\lambda+\frac{1}{2}}\right)^3\cdot\frac{\frac{1}{2}}{\lambda+\frac{1}{2}}=1-\frac{\lambda^4+2\lambda^3}{\left(\lambda+\frac{1}{2}\right)^4}.$$