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

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

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

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

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

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

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,

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

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 $1, \dots, 80$. Before the game starts, you choose 10 different numbers from amongst $1, \dots, 80$ 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

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

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 $A_r \cap A_s = A_{\text{lcm}(r,s)}$, 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, $A_1 = \{\text{all women sit together}\}$, $A_2 = \{\text{all men sit together}\}$, $A_1 \cap A_2 = \{\text{both women and men sit together}\}$, 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 $A_N = \{\text{Norwegians sit together}\}$ 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 $A_i = \{i\text{th employee gets his or her own gift}\}$. Then, what we are looking for is

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

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,

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

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, $i = 0, \ldots 5$?

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 $S1, S2, \ldots, S20$. 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 $1, \ldots 9$. 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. $P(\text{1st ball is white}) = \frac{1}{2}$.
2. $P(\text{2nd ball is white}|\text{1st ball is white}) = \frac{9}{19}$.
3. $P(\text{3rd ball is white}|\text{1st two picked are white}) = \frac{8}{18}$.

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 $F_1, \ldots, F_n$ are pairwise disjoint and that $F_1 \cup \ldots \cup F_n = \Omega$, that is, exactly one of them always happens. Then, for an event A,*

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

Proof.

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

□

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 $S \subset \{1, \dots, n\}$ 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, $i = 0, \dots, 5$, 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 $A_1, \ldots, A_r \subset U$. 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.

Solution:

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, $i = 1, 2 \dots$. 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:

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

Example 5.6. Previous example, continued. Compute 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 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. $\text{Var}(X) = np(1-p)$.

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, $\text{Var}(X)$ and $P(X \le 10)$? As X is Binomial($50, \frac{1}{2}$), so $EX = 25$, $\text{Var}(X) = 12.5$, 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. $\text{Var}(X) = \lambda$.

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

Proof.

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

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,

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

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,

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

(c) We have

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

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

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

For (a), $n = 300$ and the above expression is $P(Z \ge \frac{10}{3})$, 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, $n \geq 3303$.