Elementary Probability for Applications

Rick Durrett, Duke U

Version 2.4

August 20, 2021

Contents

Chapter 1

Combinatorial probability

1.1 Basic definitions

The subject of probability can be traced back to the 17th century when it arose out of the study of gambling games. As we will see, the range of applications extends beyond games into business decisions, insurance, law, medical tests, and the social sciences. The stock market, "the largest casino in the world," cannot do without it. The telephone network, call centers, and airline companies with their randomly fluctuating loads could not have been economically designed without probability theory. To quote Pierre Simon, Marquis de Laplace from several hundred years ago:

> "It is remarkable that this science, which originated in the consideration of games of chance, should become the most important object of human knowledge . . . The most important questions of life are, for the most part, really only problems of probability."

In order to address these applications, we need to develop a language for discussing them. Euclidean geometry begins with the undefined notions of point and line. The corresponding basic object of probability is an **experiment**: an activity or procedure that produces distinct, well-defined possibilities called **outcomes**. (Here and throughout the book **bold face type** indicates a term that is being defined.)

Example 1.1. If our experiment is to roll one die, then there are six outcomes corresponding to the number that shows on the top. The set of all outcomes in this case is $\{1, 2, 3, 4, 5, 6\}$. It is called the **sample space** and is usually denoted by Ω (capital Omega). Symmetry dictates that all outcomes are equally likely so each has probability $1/6$.

Example 1.2. Things get a little more interesting when we roll two dice. If we suppose, for convenience, that they are red and green, then we can write the outcomes of this experiment as (m, n), where m is the number on the red die and n is the number on the green die. To visualize the set of outcomes it is useful to make a little table:

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

There are $36 = 6 \cdot 6$ outcomes since there are 6 possible numbers to write in the first slot and for each number written in the first slot there are 6 possibilities for the second.

The goal of probability theory is to compute the probability of various events of interest. Intuitively, an event is a statement about the outcome of an experiment. The formal definition is: An **event** is a subset of the sample space. For example, "the sum of the two dice is 8" translates into the set $A = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$. Since this event contains 5 of the 36 possible outcomes its probability $P(A) = 5/36$.

For a second example, consider $B =$ "there is at least one six." B consists of the last row and last column of the table, so it contains 11 outcomes and hence has probability $P(B) = 11/36$. In general the probability of an event C concerning the roll of two dice is the number of outcomes in C divided by 36.

1.1.1 Axioms of Probability Theory

Let $\emptyset$ be the **empty set**, i.e., the event with no outcomes. We assume that the reader is familiar with the basic concepts of set theory such as **union** ($A \cup B$ the outcomes in either A or B) and **intersection** ($A \cap B$, the outcomes in both A and B).

Abstractly, a **probability** is a function that assigns numbers to events, which satisfies the following assumptions:

(i) For any event A, $0 \le P(A) \le 1$.
(ii) If Ω is the sample space then $P(\Omega) = 1$.
(iii) If A and B are **disjoint**, i.e., the intersection $A \cap B = \emptyset$ then

$$P(A \cup B) = P(A) + P(B)$$

(iv) If $A_1, A_2, \ldots$ is an infinite sequence of **pairwise disjoint events** (i.e., $A_i \cap A_j = \emptyset$ when $i \neq j$) then

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

These assumptions are motivated by the **frequency interpretation of probability**, which states that if we repeat an experiment a large number of times then the fraction of times the event A occurs will be close to $P(A)$. To be precise, if we let $N(A,n)$ be the number of times A occurs in the first n trials then

$$P(A) = \lim_{n \to \infty} \frac{N(A,n)}{n} \tag{1.1}$$

In Chapter 2 we will see this result is a theorem called the **law of large numbers**. For the moment, we will use this interpretation of $P(A)$ to explain the definition.

Given (1.1), properties (i) and (ii) are clear: the fraction of times that a given event A occurs must be between 0 and 1, and if Ω has been defined properly (recall that it is the set of ALL possible outcomes), the fraction of times something in Ω happens is 1. To explain (iii), note that if the events A and B are disjoint then

$$N(A \cup B, n) = N(A, n) + N(B, n)$$

since $A \cup B$ occurs if either A or B occurs but it is impossible for both to happen. Dividing by n and letting $n \to \infty$, we arrive at (iii).

Property (iii) implies that (iv) holds for a finite number of events, but for infinitely many events, the last argument breaks down, and this is a new assumption. Not everyone believes that assumption (iv) should be used. However, without (iv) the theory of probability becomes much more difficult and less useful, so we will impose this assumption and not apologize further for it. In many cases the sample space is finite so (iv) is not relevant anyway.

Example 1.3. Suppose we pick a letter at random from the word TENNESSEE. What is the sample space Ω and what probabilities should be assigned to the outcomes?

The sample space $\Omega = \{T, E, N, S\}$. To describe the probability it is enough to give the values for the individual outcomes, since (iii) implies that $P(A)$ is the sum of the probabilities of the outcomes in A. Since there are nine letters in TENNESSEE the probabilities are $P(\{T\}) = 1/9$, $P(\{E\}) = 4/9$, $P(\{N\}) = 2/9$, and $P(\{S\}) = 2/9$.

1.1.2 Basic Properties of P(A)

Having introduced a number of definitions, we will now derive some basic properties of probabilities and illustrate their use. The first one is an obvious consequence of the frequency interpretation: a large set will occur more often.

Property 1. *Monotonicity. If $A \subset B$, i.e., any outcome in A is also in B, then*

$$P(A) \leq P(B) \tag{1.2}$$

Proof. A and $A^c \cap B$ are disjoint, with union B, so assumption (iii) implies $P(B) = P(A) + P(A^c \cap B) \geq P(A)$ by (i). □

Property 2. *Let A^c be the* **complement** *of A, i.e., the set of outcomes not in A, then*

$$P(A) = 1 - P(A^c) \tag{1.3}$$

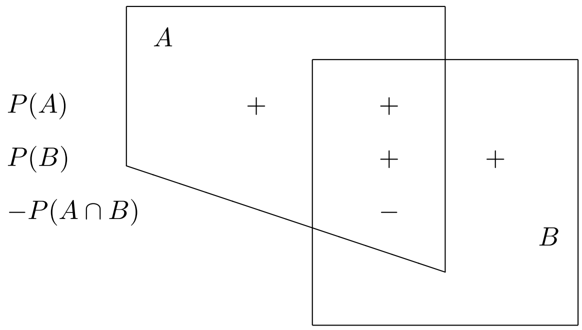
Proof. Let $A_1 = A$ and $A_2 = A^c$. Then $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 = \Omega$ so (iii) implies $P(A) + P(A^c) = P(\Omega) = 1$ by (ii). Subtracting $P(A)$ from each side of the equation gives the result. □

This formula is useful because sometimes it is easier to compute the probability of A^c. For an example, consider A = "at least one six." In this case A^c = "no six." There are $5 \cdot 5$ outcomes with no six, so $P(A^c) = 25/36$ and $P(A) = 1 - 25/36 = 11/36$, as we computed before.

Property 3. *For any events A and B,*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.4}$$

Proof by picture.



Intuitively, $P(A) + P(B)$ counts $A \cap B$ twice so we have to subtract $P(A \cap B)$ to make the net number of times $A \cap B$ is counted equal to 1. □

Proof. To prove this result we note that by assumption (ii)

$$P(A) = P(A \cap B) + P(A \cap B^c)$$
$$P(B) = P(B \cap A) + P(B \cap A^c)$$

Adding the two equations and subtracting $P(A \cap B)$:

$$P(A) + P(B) - P(A \cap B)$$
$$= P(A \cap B) + P(A \cap B^c) + P(B \cap A^c) = P(A \cup B)$$

which gives the desired equality. □

To illustrate Property 3, let $A =$ "the red die shows six," and $B =$ "the green die shows six." In this case $A \cup B =$ "at least one 6" and $A \cap B = \{(6, 6)\}$, so we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$$

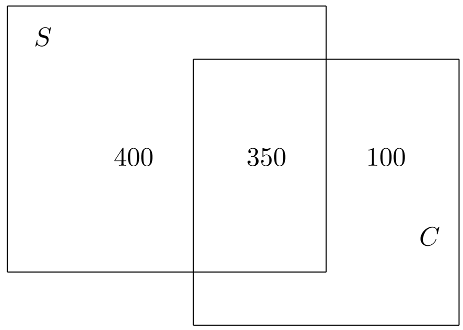
The same principle applies to counting outcomes in events.

Example 1.4. A survey of 1000 students revealed that 750 owned stereos, 450 owned cars, and 350 owned both. How many own either a car or a stereo?

Given a set A, we use to $|A|$ denote the number of points in A. The reasoning that led to (1.4) tells us that

$$|S \cup C| = |S| + |C| - |S \cap C| = 750 + 450 - 350 = 850$$

We can confirm this by drawing a picture:



Property 4. Monotone limits. We write $A_n \uparrow A$ if $A_1 \subset A_2 \subset \cdots$ and $\cup_{i=1}^{\infty} A_i = A$. We write $A_n \downarrow A$ if $A_1 \supset A_2 \supset \cdots$ and $\cap_{i=1}^{\infty} A_i = A$.

$$\text{IF } A_n \uparrow A \text{ or } A_n \downarrow A \text{ then } P(A_n) \to P(A). \tag{1.5}$$

Proof. Let $B_1 = A_1$ and for $i \geq 2$ let $B_i = A_i \cap A_{i-1}^c$. The events B_i are disjoint with $\cup_{i=1}^{\infty} B_i = A$ so

$$P(A) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(A_n)$$

To prove the second result let $C_i = A_i^c$. $C_i \uparrow C = A^c$ so $P(C_i) \to P(C)$ and hence

$$P(A_i) = 1 - P(C_i) \to 1 - P(C) = P(A)$$

which proves the second result. □

1.2 Permutations and Combinations

As usual we begin with a question:

Example 1.5. The New York State Lottery picks 6 numbers out of 59, or more precisely, a machine picks 6 numbered ping pong balls out of a set of 59. How many outcomes are there? The set of numbers chosen is all that is important. The order in which they were chosen is irrelevant.

To work up to the solution we begin with something that is obvious but is a key step in some of the reasoning to follow.

Example 1.6. A man has 4 pair of pants, 6 shirts, 8 pairs of socks, and 3 pairs of shoes. Ignoring the fact that some of the combinations may look ridiculous, in how many ways can he get dressed?

We begin by noting that there are $4 \cdot 6 = 24$ possible combinations of pants and shirts. Each of these can be paired with one of 8 choices of socks, so there are $192 = 24 \cdot 8$ ways of putting on pants, shirt, and socks. Repeating the last argument one more time, we see that for each of these 192 combinations there are 3 choices of shoes, so the answer is

$$4 \cdot 6 \cdot 8 \cdot 3 = 576 \text{ ways}$$

The reasoning in the last solution can clearly be extended to more than four experiments, and does not depend on the number of choices at each stage, so we have

The multiplication rule. Suppose that m experiments are performed in order and that, no matter what the outcomes of experiments $1, \ldots, k-1$ are, experiment k has n_k possible outcomes. Then the total number of outcomes is $n_1 \cdot n_2 \cdots n_m$.

Example 1.7. How many ways can 5 people stand in line?

To answer this question, we think about building the line up one person at a time starting from the front. There are 5 people we can choose to put at the front of the line. Having made the first choice, we have 4 possible choices for the second position. (The set of people we have to choose from depends upon who was chosen first, but there are always 4 people to choose from.) Continuing, there are 3 choices for the third position, 2 for the fourth, and finally 1 for the last. Invoking the multiplication rule, we see that the answer must be

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

Generalizing from the last example we define n **factorial** to be

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 \tag{1.6}$$

To see that this gives the number of ways n people can stand in line, notice that there are n choices for the first person, $n-1$ for the second, and each subsequent choice reduces the number of people by 1 until finally there is only 1 person who can be the last in line.

Note that $n!$ grows very quickly since $n! = n \cdot (n-1)!$.

1!	1	7!	5,040
2!	2	8!	40,320
3!	6	9!	362,880
4!	24	10!	3,628,800
5!	120	11!	39,916,800
6!	720	12!	479,001,600

The number of ways we can put the 22 volumes of an encyclopedia on a shelf is

$$22! = 1.24000728 \times 10^{21}$$

Here we have used our TI-83. We typed in 22 then used the MATH button to get to the PRB menu and scroll down to the fourth entry to get the ! which gives us 22! after we press ENTER.

The number of ways that cards in a deck of 52 can be arranged is

$$52! = 8.065817517 \times 10^{67}$$

Before there were calculators, people used Stirling's formula

$$n! \approx (n/e)^n \sqrt{2\pi n} \tag{1.7}$$

When $n = 52$, $52/e = 19.12973094$ and $\sqrt{2\pi n} = 18.07554591$ so

$$52! \approx (19.12973094)^{52} \cdot 18.07554591 = 8.0529 \times 10^{67}$$

Example 1.8. Twelve people belong to a club. How many ways can they pick a president, vice-president, secretary, and treasurer?

Again we think of filling the offices one at a time in the order in which they were given in the last sentence. There are 12 people we can pick for president. Having made the first choice, there are always 11 possibilities for vice-president, 10 for secretary, and 9 for treasurer. So by the multiplication rule, the answer is

$$\frac{12}{P}\,\frac{11}{V}\,\frac{10}{S}\,\frac{9}{T} = 11,800$$

To compute $P_{12,4}$ with the TI-83 calculator: type 12, push the MATH button, move the cursor across to the PRB submenu, scroll down to nPr on the second row, and press ENTER. nPr appears on the screen after the 12. Now type 4 and press ENTER.

Passing to the general situation, if we have k offices and n club members then the answer is

$$n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)$$

To see this, note that there are n choices for the first office, $n-1$ for the second, and so on until there are $n-k+1$ choices for the last, since after the last person is chosen there will be $n-k$ left. Products like the last one come up so often that they have a name: the **number of permutations of** k **objects from a set of size** n, or $P_{n,k}$ for short. Multiplying and dividing by $(n-k)!$ we have

$$n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) \cdot \frac{(n-k)!}{(n-k)!} = \frac{n!}{(n-k)!}$$

which gives us a short formula,

$$P_{n,k} = \frac{n!}{(n-k)!} \tag{1.8}$$

The last formula would give us the answer to the lottery problem if the order in which the numbers drawn was important. Our last step is to consider a related but slightly simpler problem.

Example 1.9. A club has 23 members. How many ways can they pick 4 people to be on a committee to plan a party?

To reduce this question to the previous situation, we imagine making the committee members stand in line, which by (1.8) can be done in $23 \cdot 22 \cdot 21 \cdot 20$ ways. To get from this to the number of committees, we note that each committee can stand in line 4! ways, so the number of committees is the number of lineups divided by 4! or

$$\frac{23 \cdot 22 \cdot 21 \cdot 20}{1 \cdot 2 \cdot 3 \cdot 4} = 23 \cdot 11 \cdot 7 \cdot 5 = 8,855$$

To compute $C_{23,4}$ with the TI-83 calculator: type 23, push the MATH button, move the cursor across to the PRB submenu, scroll down to nCr on the third row, and press ENTER. nCr appears on the screen after the 23, now type 4 and press ENTER.

Passing to the general situation, suppose we want to pick k people out of a group of n. Our first step is to make the k people stand in line, which can be done in $P_{n,k}$ ways, and then to realize that each set of k people can stand in line $k!$ ways, so the number of ways to choose k people out of n is

$$C_{n,k} = \frac{P_{n,k}}{k!} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{1 \cdot 2 \cdots k} \tag{1.9}$$

by (1.9) and (1.6). Here, $C_{n,k}$ is short for the **number of combinations of k things taken from a set of n.** $C_{n,k}$ is often written as $\binom{n}{k}$, a symbol that is read as "n choose k." We are now ready for the

Answer to the Lottery Problem, Example 1.5. We are choosing $k = 6$ objects out of a total of $n = 59$ when order is not important, so the number of possibilities is

$$\begin{aligned} C_{59,6} = \frac{59!}{6!53!} &= \frac{59 \cdot 58 \cdot 57 \cdot 56 \cdot 55 \cdot 54}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} \\ &= 59 \cdot 58 \cdot 19 \cdot 7 \cdot 11 \cdot 9 = 45,057,474 \end{aligned}$$

You should consider this the next time you think about spending \$1 for two chances to win a jackpot that starts at \$3 million and increases by \$1 million each week there is no winner.

Pascal's triangle. The number of outcomes for coin tossing problems fit together in a nice pattern:

```
                    1
                 1     1
              1     2     1
           1     3     3     1
        1     4     6     4     1
     1     5    10    10     5     1
  1     6    15    20    15     6     1
1     7    21    35    35    21     7     1
```

Each number is the sum of the ones on the row above on its immediate left and right. To get the 1's on the edges to work correctly we consider the blanks to be zeros. In symbols

$$C_{n,k} = C_{n-1,k-1} + C_{n-1,k} \tag{1.10}$$

Verbal proof. In picking k things out of n, which can be done in $C_{n,k}$ ways, we may or may not pick the last object. If we pick the last object then we must complete our set of k by picking $k-1$ objects from the first $n-1$, which can be done in $C_{n-1,k-1}$ ways. If we do not pick the last object then we must pick all k objects from the first $n-1$, which can be done in $C_{n-1,k}$ ways. □

Analytic proof. Using the definition (1.9)

$$C_{n-1,k-1} + C_{n-1,k} = \frac{(n-1)!}{(n-k)!(k-1)!} + \frac{(n-1)!}{(n-k-1)!k!}$$

Factoring out the parts common to the two fractions

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

which proves (1.10). □

Binomial theorem. The numbers in Pascal's triangle also arise if we take powers of $(x+y)$:

$$(x+y)^2 = x^2 + 2xy + y^2$$

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

$$\begin{aligned}(x+y)^4 &= (x+y)(x^3+3x^2y+3xy^2+y^3)\\ &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4\end{aligned}$$

or in general

$$(x+y)^n = \sum_{m=0}^{n} C_{n,m} x^m y^{n-m} \tag{1.11}$$

To see this consider $(x+y)^5$ and write it as

$$(x+y)(x+y)(x+y)(x+y)(x+y)$$

Since we can choose x or y from each parenthesis, there are 2^5 terms in all. If we want a term of the form x^3y^2 then in 3 of the 5 cases we must pick x, so there are $C_{5,3} = (5 \cdot 4)/2 = 10$ ways to do this. The same reasoning applies to the other terms, so we have

$$\begin{aligned}(x+y)^5 &= C_{5,5}x^5 + C_{5,4}x^4y + C_{5,3}x^3y^2 + C_{5,2}x^2y^3 + C_{5,1}xy^4 + C_{5,0}y^5\\ &= x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5\end{aligned}$$

1.2.1 More than two categories

We defined $C_{n,k}$ as the number of ways of picking k objects out of n. To motivate the next generalization we would like to observe that $C_{n,k}$ is also the number of ways we can divide n objects into two groups, the first one with k objects and the second with $n-k$. To connect this observation with the next problem, think of it as asking: "How many ways can we divide 12 objects into three numbered groups of sizes 4, 3, and 5?"

Example 1.10. A house has 12 rooms. We want to paint 4 yellow, 3 purple, and 5 red. In how many ways can this be done?

This problem can be solved using what we know already. We first pick 4 of the 12 rooms to be painted yellow, which can be done in $C_{12,4}$ ways, and then pick 3 of the remaining 8 rooms to be painted purple, which can be done in $C_{8,3}$ ways. (The 5 unchosen rooms will be painted red.) The answer is:

$$C_{12,4}C_{8,3} = \frac{12!}{4!\,8!} \cdot \frac{8!}{3!\,5!} = \frac{12!}{4!\,3!\,5!} = 27,720$$

A second way of looking at the problem, which gives the last answer directly, is to first decide the order in which the rooms will be painted, which can be done in 12! ways, then paint the first 4 on the list yellow, the next 3 purple, and the last 5 red. One example is

$$\frac{9}{Y}\,\frac{6}{Y}\,\frac{11}{Y}\,\frac{1}{Y}\,\frac{8}{P}\,\frac{2}{P}\,\frac{10}{P}\,\frac{5}{R}\,\frac{3}{R}\,\frac{7}{R}\,\frac{12}{R}\,\frac{4}{R}$$

Now, the first four choices can be rearranged in 4! ways without affecting the outcome, the middle three in 3! ways, and the last five in 5! ways. Invoking the multiplication rule, we see that in a list of the 12! possible permutations each possible painting thus appears 4! 3! 5! times. Hence the number of possible paintings is

$$\frac{12!}{4!\,3!\,5!}$$

The second computation is a little more complicated than the first, but makes it easier to see

Theorem 1.1. *The number of ways a group of n objects to be divided into m groups of size $n_1, \ldots, n_m$ with $n_1 + \cdots + n_m = n$ is*

$$\frac{n!}{n_1!\,n_2! \cdots n_m!} \tag{1.12}$$

The formula may look complicated but it is easy to use.

Example 1.11. There are 39 students in a class. In how many ways can a professor give out 9 A's, 13 B's, 12 C's, and 5 F's?

$$\frac{39!}{9!\,13!\,12!\,5!} = 1.57 \times 10^{22}$$

1.3 Flipping Coins, the World Series, Birthdays

Even simpler than rolling a die is flipping a coin, which produces one of two outcomes, called "Heads" (H) or "Tails" (T). If we flip two coins there are four outcomes

	HH	HT TH	TT
heads	2	1	0
probability	1/4	2/4	1/4

Flipping three coins there are eight possibilities:

	HHH	HHT HTH THH	TTH THT HTT	TTT
heads	3	2	1	0
probability	1/8	3/8	3/8	1/8

If we flip a large number of coins it will be tedious to write out all the otucomes so we need to derive some fomrulas.

Example 1.12. Suppose we flip seven coins. Compute the probability that we get 0, 1, 2, or 3 heads.

There are $2^7 = 128$ total outcomes. There is only 1, $TTTTT$ that gives 0 heads, so that probability is 1/32. There are 7 outcomes that have one heads. We could write them out but it is better to reason that we can pick the toss for the heads to occur in

$$C_{7,1} = \frac{7!}{6!\,1!} = 7$$

Extending the last reasoning to two heads, the number of outcomes is the number of ways of picking 2 tosses for the heads to occur or

$$C_{7,2} = \frac{7!}{5!\,2!} = \frac{7 \cdot 6}{2} = 21$$

Likewise the number of outcomes with 3 heads is

$$C_{7,3} = \frac{7!}{4!\,3!} = \frac{7 \cdot 6 \cdot 5}{3!} = 35$$

By symmetry the numbers of outcomes for 4, 5, 6, and 7 heads are 35, 21, 7, and 1. In terms of binomial coefficients this says

$$C_{n,m} = \frac{n!}{m!(n-m)!} = C_{n,n-m} \tag{1.13}$$

To prove this in words: The number of ways of picking m objects out of n to take is the same as the number of ways of choosing $n-m$ to leave behind. Of course, one can also check this directly from the formula in (1.9).

Our next problem concerns flipping four to seven coins:

Example 1.13. World Series. In this baseball event, the first team to win four games wins the championship. Obviously, the series may last 4, 5, 6, or 7 games. Here we will compute the probabilities of each of these outcomes. To do this, we will assume that the two teams are equally matched and ignoring potential complicating factors like the advantage of playing at home or psychological factors that make the outcome of one game affect the next one. In short, we suppose that the games are decided by tossing a fair coin to determine if team A or team B wins.

Four games. There are two possible ways this can happen: A wins all four games or B wins all four games. There are $2 \cdot 2 \cdot 2 \cdot 2 = 16$ possible outcomes and these are 2 of them so $P(4) = 2/16 = 1/8$.

Five games. Here and in the next two cases we will compute the probability that A wins in the specified number of games and then multiply by 2. There are four possible outcomes

$$BAAAA, \quad ABAAA, \quad AABAA, \quad AAABA$$

$AAAAB$ is not possible since in that case the series would have ended in four games. There are $2^5 = 32$ outcomes so $P(5) = 2 \cdot 4/32 = 1/4$.

Six games. As we noticed in the last example, A has to win the last game. That means B must win exactly two of the first five games, so there are

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

outcomes. Each has probability $1/2^6$, so the probability A wins in 6 games is $10/2^6$ and the probability the series lasts 6 games is $2 \cdot 10/64 = 5/16$.

Seven games. In this case B must win exactly three of the first six games, so there are

$$C_{6.3} = \frac{6!}{3!3!} = \frac{6 \cdot 5 \cdot 4}{3!} = 20$$

outcomes. Each has probability $1/2^7$, so the probability A wins in 6 games is $20/2^7$ and the probability the series lasts 7 games is $2 \cdot 20/128 = 5/16$.

As mentioned earlier, we are ignoring things that many fans think are important to determining the outcomes of the games, so our next step is to compare the probabilities just calculated with the observed frequencies of the duration of best of seven series in three sports. The numbers in parentheses give the number of series in our sample.

Games	4	5	6	7
Probabilities	0.125	0.25	0.3125	0.3125
World Series (94)	0.181	0.224	0.224	0.372
Stanley Cup (74)	0.270	0.216	0.230	0.284
NBA finals (57)	0.122	0.228	0.386	0.263

In statistics class you will learn that the **chi-squared statistic** to see if the observations are consistent with the probabilities we have computed. The details of the test are beyond the scope of this book so we just quote the results: the Stanley Cup data is very unusual due to the larger than expected number of four game series. The World Series data does not fit the model well, but is not very unusual. On the other hand, the NBA finals data looks like what we should expect to see. The excess of six game series can be due just to chance.

Example 1.14. Birthday problem. There are 30 people at a party. Someone wants to bet you \$10 that there are two people with exactly the same birthday. Should you take the bet?

To pose a mathematical problem, we ignore Feb. 29 which only comes in leap years, and suppose that each person at the party picks their birthday at random from the calendar. There are 365^{30} possible outcomes for that experiment. The number of outcomes in which all the birthdays are different is

$$365 \cdot 364 \cdot 363 \cdots 336$$

since the second person must avoid the first person's birthday, the third the first two birthdays and so on until the 30th person must avoid the 29 previous birthdays. Let D be the event that all birthdays are different. Dividing the number of outcomes in which all the birthdays are different by the total number of outcomes, we have

$$P(D) = \frac{365 \cdot 364 \cdot 363 \cdots 336}{365^{30}} = 0.293684$$

In words, only about 29% of the time all the birthdays are different, so you will lose the bet 71% of the time.

At first glance it is surprising that the probability of two people having the same birthday is so large, since there are only 30 people compared with 365 days on the calendar. Some of the surprise disappears if you realize that there are $(30 \cdot 29)/2 = 435$ pairs of people who are going to compare their birthdays. Let p_k be the probability that k people all have different birthdays. Clearly, $p_1 = 1$ and $p_{k+1} = p_k(365 - k)/365$. Using this recursion it is easy to generate the values of p_k.

The graph shows the trends, but to get precise values a table is better:

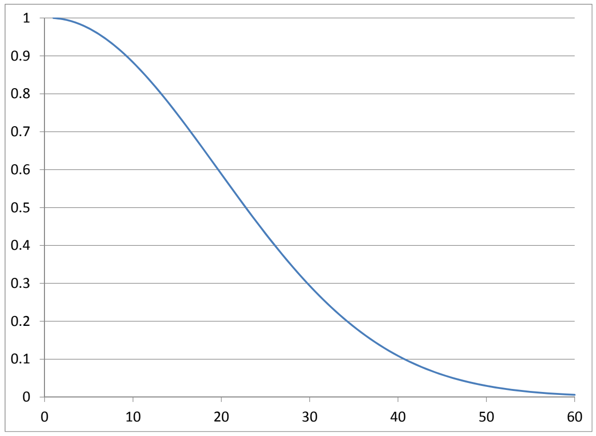


Figure 1.1: Probability of no birthday match in a group of size k

1	1.00000	11	0.85886	21	0.55631	31	0.26955
2	0.99726	12	0.83298	22	0.52430	32	0.24665
3	0.99180	13	0.80559	23	0.49270	33	0.22503
4	0.98364	14	0.77690	24	0.46166	34	0.20468
5	0.97286	15	0.74710	25	0.43130	35	0.18562
6	0.95954	16	0.71640	26	0.40176	36	0.16782
7	0.94376	17	0.68499	27	0.37314	37	0.15127
8	0.92566	18	0.65309	28	0.34554	38	0.13593
9	0.90538	19	0.62088	29	0.31903	39	0.12178
10	0.88305	20	0.58856	30	0.29368	40	0.10877

1.4 Random variables, Expected value

A **random variable** is a real valued function defined on a probability space. Three examples:

- Roll two dice and let X = the sum of the two numbers that appear.
- Flip a coin 10 times and let X = the number of Heads we get.
- Draw 13 cards out of a deck and let X = the number of Hearts we get.

The **distribution** of a discrete random variable is described by listing its possible values and giving their probabilities. In the previous section we computed the distribution of the number of games N to finish the world series. The distribution of the random variable is given by

x	4	5	6	7
$P(X = x)$	2/16	4/16	5/16	5/16

The **expected value** of a discrete random value X is obtained form the formula

$$EX = \sum_x xP(X = x) \tag{1.14}$$

In words we sum over the set of possible values the value times the probability it occurs. EX is also called the mean of X. It should be clear from the definitions that

$$E(X + a) = EX + a \qquad E(bX) = bEX \tag{1.15}$$

In words if we add 3 to a random variable we add 3 to its mean. If we multiply a random variable by 5 we multiply its mean by 5.

Example 1.15. Roulette. If you play roulette and bet \$1 on black then you win \$1 with probability 18/38 and you lose \$1 with probability 20/38, so the expected value of your winnings X is

$$EX = 1 \cdot \frac{18}{38} + (-1) \cdot \frac{20}{38} = \frac{-2}{38} = -0.0526$$

The expected value has a meaning much like the frequency interpretation of probability: in the long run you will lose about 5.26 cents per play.

Example 1.16. In the case of the world series

$$EX = \frac{2 \cdot 4 + 4 \cdot 5 + 5 \cdot 6 + 5 \cdot 7}{16} = \frac{93}{16} = 5.8125$$

The interpretation of the mean is similar to the frequency interpretation of probability.If we repeat our little coin flipping World Series a large number of times, the average number of games will be close to the mean.

Example 1.17. Roll one die. Let X be the number that appears on the die. $P(X = x) = 1/6$ for $x = 1, 2, 3, 4, 5, 6$, so

$$EX = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3\frac{1}{2}$$

Example 1.18. Suppose we roll two dice and let S_2 be the sum of the two numbers. Consulting a table of the possible outcomes

(1,1)	(2,1)	(3,1)	(4,1)	**(5,1)**	(6,1)
(1,2)	(2,2)	(3,2)	**(4,2)**	(5,2)	(6,2)
(1,3)	(2,3)	**(3,3)**	(4,3)	(5,3)	(6,3)
(1,4)	**(2,4)**	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

we see that if S_2 is the sum of the two dice then $P(S_2 = 6) = 5/36$. Realizing that the outcomes with a given sum lie on a diagonal line we see that the distribution of the sum is given by

n	2	3	4	5	6	7	8	9	10	11	12
$P(S_2 = n)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

We could compute ES_2 by using the definition

$$ES_2 = \frac{2\cdot 1 + 3\cdot 2 + 4\cdot 3 + 5\cdot 4 + 6\cdot 5 + 7\cdot 6}{36} + \frac{8\cdot 5 + 9\cdot 4 + 10\cdot 3 + 11\cdot 2 + 12\cdot 1}{36} = \frac{252}{36} = 7$$

However, there is a simpler way. Let X_1 be the number on the first die, and let X_2 be the number on the second die. Since $P(X_i = j) = 1/6$ for $j = 1, 2, \dots 6$ it is clear that

$$EX_i = \frac{1+2+3+4+5+6}{6} = 3.5$$

Theorem 1.2. *If $X_1, X_2, \dots X_n$ are any discrete random variables then*

$$E(X_1 + \cdots + X_n) = EX_1 + \cdots + EX_n$$

This is true for the exact same reason that

$$\int f(x) + g(x)\, dx = \int f(x)\, dx + \int g(x)\, dx$$

so we will not prove this.

From this it follows that if S_2 is the sum of two dice then $ES_2 = EX_1 + EX_2 = 7$, or more generally if S_k is the total of k dice then $ES_k = 3.5k$. The same reasoning applies to H_k the number of heads when we flip k coins. Let $Y_i = 1$ if the ith coin is heads and $Y_i = 0$ if the ith coin is tails. Since $P(Y_i = 1) = P(Y_i = 0) = 1/2$ we have $EY_i = (1/2)\cdot 1 + (1/2)\cdot 01/2$ so we have

$$EH_k = Y_1 + \cdot EY_k = k/2$$

Example 1.19. The birthday problem. While it is a little tedious to compute the probability of no birthday match in a group of n people, it is easy to compute the expected value of M_n the total number of birthday matches. For $1 \le i < j \le n$ let $X_{i,j} = 1$ if the birthdays of i and j match. The total number of birthday matches

$$M_n = \sum_{1 \le i < j \le n} X_{ij}$$

where the sum is over all i and j with $i < j$. The number of such pairs is $C_{n,2}$ so

$$EM_n = \frac{C_{n,2}}{365}$$

If $n = 30$ this is

$$EM_{30} = \frac{30\cdot 29}{2}\cdot\frac{1}{365} = 0.78082.$$

In the previous section we computed that $P(M_{30} = 0) = 0.29368$ so $P(M_{30} > 0) = 0.70632$.

Example 1.20. The birthday problem in the Senate. Since there are 100 senators it is very likely that there are two with the same birthday. By the calculations in the previous example

$$EM_{100} = \frac{100\cdot 99}{2}\cdot\frac{1}{365} = \frac{4950}{365} = 13.56$$

With a significant number of birthday matches it is natural to ask if there are any triple birthdays. Let $T_{ijk} = 1$ senators i, j, and k all have the same birthday.

$$ET_{100} = \frac{100 \cdot 99 \cdot 98}{3!} \cdot \frac{1}{365^2} = \frac{161,700}{133,225} = 1.213$$

In the 2018 Senate there were 10 double birthdays and one triple birthdays: May 3: Jim Risch (Idaho), David Vitter (Louisiana), Ron Wyder (Oregon). It is difficult to compute the exact probability of no triple birthday. In Example 2.22 in Chapter 2 we will find a simple approximation, which suggests teh answer is ≈ 0.614.

Example 1.21. Fair division of a bet on an interrupted game. Pascal and Fermat were sitting in a café in Paris playing the simplest of all games, flipping a coin. Each had put up a bet of 50 francs and the first to get 10 points wins. Fermat was winning 8 points to 7 when he received an urgent message that a friend was sick, and he must rush to his home town of Toulouse. The carriage man who delivered the message offered to take him, but only if he would leave immediately. Later in correspondence between the two men, the problem arose: how should the money bet (100 francs) be divided?

In the case under consideration, it is easier to calculate the probability that Pascal (P) wins. In order for Pascal to win by a score of 10-8 he must win three flips in a row: PPP, an event of probability 1/8. To win 10-9 he can lose one flip but not the last one: $FPPP$, $PFPP$, $PPFP$, which has probability 3/16. Adding the two we see that Pascal wins with probability 5/16 and should get that fraction of the total bet, i.e., $(5/16)(100) = 31.25$.Fermat came up with the reasonable idea that the fraction of the stakes that each receives should be equal to the probability it would have won the match.

1.5 Card Games and Other Urn Problems

A number of problems in probability have the following form.

Example 1.22. Suppose we pick 3 balls out of an urn with 12 red balls and 8 black balls. What is the probability of B_2 = "We get two red balls and one black ball."?

Almost by definition, there are

$$C_{20,3} = \frac{20 \cdot 19 \cdot 18}{1 \cdot 2 \cdot 3} = 5 \cdot 19 \cdot 3 \cdot 17 = 1,140$$

ways of picking 4 balls out of the 20. To count the number of outcomes in B_2, we note that there are $C_{12,2} = (12 \cdot 11)/2 = 66$ ways to choose the red balls and $C_{8,1} = 8$ ways to choose the black balls, so the multiplication rule implies

$$|B_2| = C_{12,2}C_{8,1} = 66 \cdot 8 = 528$$

It follows that $P(B_2) = 528/1140 = 0.4632$.

If we let B_k be the event that we draw k red balls then we have

k	3	2	1	0
$\|B_k\|$	$C_{12,3} = 220$	$C_{12,2}C_{8,1} = 528$	$C_{12,1}C_{8,7} = 336$	$C_{8,3} = 56$
$P(B_K)$	0.1930	0.4632	0.2947	0.0491

Example 1.23. A deck of cards has four suits: spades ♠, hearts ♡, diamonds ♢ and clubs ♣. Each suit has an ace A, king K, queen Q, jack J, 10, 9, 8, 7, 6, 5, 4, 3, 2. Suppose we draw five cards out the deck. What is the probbability we have exactly three psaces.

The number of possible outcomes is

$$C_{52,5} = 2,598,960$$

The number of outcomes with 3 spades and 2 nonspades is

$$C_{13,3} \cdot C_{39,2} = \frac{13 \cdot 12 \cdot 11}{3!} \cdot \frac{39 \cdot 38}{2} = 211,926$$

so the answer is $211,926/2,598,960 = 0.8154$

Poker. In the game of poker the following hands are possible; they are listed in increasing order of desirability. In the definitions the word *value* refers to A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, or 2. This sequence also describes the relative ranks of the cards, with one exception: an Ace may be regarded as a 1 for the purposes of making a straight. (See the example in (d), below.)

(a) *one pair:* two cards of equal value plus three cards with different values
J♠ J♢ 9♡ Q♣ 3♠
(b) *two pair:* two pairs plus another card with a different value
J♠ J♢ 9♡ 9♣ 3♠
(c) *three of a kind:* three cards of the same value and two with different values
J♠ J♢ J♡ 9♣ 3♠
(d) *straight:* five cards with consecutive values
5♡ 4♠ 3♠ 2♡ A♣
(e) *flush:* five cards of the same suit
K♣ 9♣ 7♣ 6♣ 3♣
(f) *full house:* a three of a kind and a pair
J♠ J♢ J♡ 9♣ 9♠
(g) *four of a kind:* four cards of the same value plus another card
J♠ J♢ J♡ J♣ 9♠
(h) *straight flush:* five cards of the same suit with consecutive values
A♣ K♣ Q♣ J♣ 10♣
This example is called a *royal flush.*

To compute the probabilities of these poker hands we begin by observing that there are

$$C_{52,5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 2,598,960$$

ways of picking 5 cards out of a deck of 52, so it suffices to compute the number of ways each hand can occur. We will do three cases to illustrate the main ideas and then leave the rest to the reader.

(d) *straight:* $10 \cdot 4^5$
A straight must start with a card that is 5 or higher, 10 possibilities. Once the values are decided on, suits can be assigned in 4^5 ways. This counting regards a straight flush as a straight. If you want to exclude straight flushes, suits can be assigned in $4^5 - 4$ ways.

(f) *full house:* $13 \cdot C_{4,3} \cdot 12 \cdot C_{4,2}$
We first pick the value for the three of a kind (which can be done in 13 ways), then assign suits to those three cards ($C_{4,3}$ ways), then pick the value for the pair (12 ways), then we assign suits to the last two cards ($C_{4,2}$ ways).

(b)*two pair:* $C_{13,2} \cdot C_{4,2} \cdot C_{4,2} \cdot 44$
We first pick the values for the two pair then assign suits to each of them, and pick one final card with a value that does not match the first two. A common incorrect answer to this question is

$$13 \cdot C_{4,2} \cdot 12 \cdot C_{4,2} \cdot 44$$

Referring to the example above,choosing j and then 9 is counted as being different from 9 and then J so this is 2 times the correct answer. Similar reasoning leads to

(a) *one pair:* $13 \cdot C_{4,2} \cdot C_{12,3} \cdot 4^3$

The numerical values of the probabilities of all poker hands are given in the next table.

(a) *one pair*	.422569
(b) *two pair*	.047539
(c) *three of a kind*	.021128
(d) *straight*	.003940
(e) *flush*	.001981
(f) *full house*	.001441
(g) *four of a kind*	.000240
(h) *straight flush*	.000015

The probability of getting none of these hands can be computed by summing the values for (a) through (g) (recall that (d) includes (h)) and subtracting the result from 1. However, it is much simpler to observe that we have nothing if we have five different values that do not make a straight or a flush. So the number of nothing hands is $(C_{13,5} - 10) \cdot (4^5 - 4)$ and the probability of a nothing hand is 0.501177.

Example 1.24. Suppose we draw 13 cards from a deck. How many outcomes are there? How many lead to hands with 4 spades, 3 hearts, 3 diamonds, and 3 clubs? 3 spades, 5 hearts, 2 diamonds, and 3 clubs?

$$C_{52,13} = 6.350135596 \times 10^{11}$$
$$C_{13,4}C_{13,3}C_{13,3}C_{13,3} = 715 \cdot (286)^3 = 16,726,464,040$$
$$C_{13,3}C_{13,5}C_{13,2}C_{13,3} = 286 \cdot 1287 \cdot 78 \cdot 286 = 8,211,173,256$$

Example 1.25. Suit distributions. The last bridge hand in the previous example is said to have a 5-3-3-2 distribution. Here, we have listed the number cards in the longest suit first and continued in decreasing order. Permuting the four numbers we see that the example 3 spades, 5 hearts, 2 diamonds, and 3 clubs is just one of 4!/2! possible ways of assigning the numbers to suits, so the probability of a 5-3-3-2 distribution is

$$\frac{12 \cdot 8,211,173,256}{6.350135596 \times 10^{11}} = 0.155$$

Similar computations lead to the results in the next table. We have included the number of different permutations of the pattern to help explain why slightly unbalanced distributions have larger probability than 4-3-3-3.

Distribution	Probability	Permutations
4-4-3-2	0.216	12
5-3-3-2	0.155	12
5-4-3-1	0.129	24
5-4-2-2	0.106	12
4-3-3-3	0.105	4
6-3-2-2	0.056	12

Example 1.26. Keno. In this game the casino picks 20 balls out of 80 numbered balls. Before the draw you may, for example pick 10 numbers and bet \$1. In this case you win \$1 if 4 of your numbers are chosen; \$2 for 5; \$20 for 6; \$105 for 7; \$500 for 8; \$5000 for 9; and \$12,000 if all ten are chosen. We want to compute the expected value of the bet.

The number of possible draws is astronomically large:

$$C_{80,20} = 3.5353 \times 10^{18}$$

The probability that k of your numbers are chosen is

$$p_k = \frac{C_{10,k} C_{70,20-k}}{C_{80,20}}$$

When $k = 0$ this is

$$\frac{C_{70,20}}{C_{80,20}} = \frac{70!60!}{80!50!} = \frac{60 \cdot 59 \cdots 51}{80 \cdot 79 \cdots 71} = 0.045791$$

To compute the other probabilities it is useful to note that for $1 \le m \le n$

$$\begin{aligned} C_{n,m} = \frac{n!}{m!(n-m)!} &= \frac{n+1-m}{m} \cdot \frac{n!}{(m-1)!(n+1-m)!} \\ &= \frac{n+1-m}{m} \cdot C_{n,m-1} \end{aligned}$$

so we have

$$p_k = p_{k-1} \cdot \frac{11-k}{k} \cdot \frac{21-k}{50+k}$$

Writing w_k for the winning when k of our numbers are drawn, using this recursion and the result for p_0 gives

k	p_k	w_k	$w_k p_k$
0	0.045791		0
1	0.179571		0
2	0.295257		0
3	0.267402		0
4	0.147319	1	0.147319
5	0.051428	2	0.102855
6	0.011479	20	0.229588
7	0.001611	105	0.169701
8	0.000135	500	0.067710
9	6.12×10^{-6}	5000	0.030603
10	1.12×10^{-7}	12,000	0.001347
4-10	.2120		.7486

Thus, we win something about 21.2% of the time and our average winning is a little less than 75 cents, a typical expected value for Keno bets. The last column shows the contribution of the different payoffs to the expected value.

Example 1.27. Disputed elections. In a close election in a small town, 2,656 people (50.6%) voted for candidate A compared to 2,594 who voted for candidate B, a margin of victory of 62 votes. An investigation of the election found that 136 of the people who voted in the election should not have. Since this is more than the margin of victory, should the election results be thrown out even though there was no evidence of fraud on the part of the winner's supporters?

Like many problems that come from the real world (a court case *De Martini v. Power*), this one is not precisely formulated. To turn this into a probability problem we suppose that all the votes were equally likely to be one of the 136 erroneously cast and we investigate what happens when we remove 136 balls from an urn with 2,656 white balls and 2,594 black balls.

In order to reverse the outcome of the election, we must have

$$2,656 - m \leq 2,594 - (136 - m) \quad \text{or} \quad m \geq 99$$

This arithmetic was enough for the judge to make his decision. " Although the successful candidate was elected on the strength of only 50.6% of the vote, her majority would not be lost unless 99 votes i.e., 72.8% of the irregularities were cast in her favor. It taxes credulity to assume that, in so close a contest, such an extreme percentage of invalid votes would be cast in one direction. In the conceded absence of fraud, a valid determination is not rendered impossible (Election Law, by the remote possibility of a changed result."

To see how remote the possibility is, we note that the probability of removing exactly m white and $136 - m$ black balls is

$$p_m = \frac{C_{2656,m} C_{2594,136-m}}{C_{5250,136}}$$

If one tries to evaluate this when $m = 99$ with your calculator you get an overflow message. Expanding the factorials we have

$$p_{99} = \frac{\frac{2656 \cdot 2558}{99!} \frac{2594 \cdots 2558}{37!}}{\frac{5250 \cdots 5115}{136!}} = \frac{136!}{99!37!} \cdot \frac{2656 \cdots 2558 \cdot 2594 \cdots 2558}{5250 \cdots 5152 \cdot 5151 \cdots 5115}$$

In the last product the ratio of the term in the numerator to the one under it in the denomainator is always close to 1/2 so the above is

$$\approx \frac{136!}{99!37!} (1/2)^{136} \tag{1.16}$$

which corresponds to flipping 136 coins to determine to candidate who loses the votes. The probability in (1.16) is 3.27×10^{-8}. In the next section we will learn that this is the binomial distribution and we will learn how to compute the probability B loses ≤ 37 votes is 5.13×10^{-8} computer program shows the exact probability is 7.492×10^{-8}.

Example 1.28. Quality control. A shipment of 50 precision parts including 4 that are defective is sent to an assembly plant. The quality control division selects 10 at random for testing and rejects the entire shipment if 1 or more are found defective. What is the probability this shipment passes inspection?

There are $C_{50,10}$ ways of choosing the test sample, and $C_{46,10}$ ways of choosing all good parts so the probability is

$$\frac{C_{46,10}}{C_{50,10}} = \frac{46!/36!10!}{50!/40!10!} = \frac{46 \cdot 45 \cdots 37}{50 \cdot 49 \cdots 41}$$
$$= \frac{40 \cdot 39 \cdot 38 \cdot 37}{50 \cdot 49 \cdot 48 \cdot 47} = 0.396$$

Using almost identical calculations a company can decide on how many bad units they will alow in a shipment and design a testing program with a given probability of success.

Example 1.29. Capture-recapture experiments. An ecology graduate student goes to a pond and captures $k = 60$ beetles, marks each with a dot of paint, and then releases them. A few days later she goes back and captures another sample of $r = 50$, finding $m = 12$ marked beetles and $r - m = 38$ unmarked. What is her best guess about the size of the population of beetles?

To turn this into a precisely formulated problem, we will suppose that no beetles enter or leave the population between the two visits. With this assumption, if there were N beetles in the pond, then the probability of getting m marked and $r - m$ unmarked in a sample of r would be

$$p_N = \frac{C_{k,m}\, C_{N-k,r-m}}{C_{N,r}}$$

To estimate the population we will pick N to maximize p_N, the so-called **maximum likelihood estimate**. To find the maximizing N, we note that

$$C_{j-1,i} = \frac{(j-1)!}{(j-i-1)!i!} \quad \text{so} \quad C_{j,i} = \frac{j!}{(j-i)!i!} = \frac{jC_{j-1,i}}{(j-i)}$$

and it follows that

$$p_N = p_{N-1} \cdot \frac{N-k}{N-k-(r-m)} \cdot \frac{N-r}{N}$$

Now $p_N/p_{N-1} \geq 1$ if and only if

$$(N-k)(N-r) \geq N(N-k-r+m)$$

that is,

$$N^2 - kN - rN + kr \geq N^2 - kN - rN + mN$$

or equivalently if $N \leq kr/m$. Thus the value of N that maximizes the probability p_N is the largest integer $\leq kr/m$. This choice is reasonable since when $N = kr/m$ the proportion of marked beetles in the population, k/N, equals the proportion of marked beetles in the sample, m/r. Plugging in the numbers from our example, $kr/m = (60 \cdot 50)/12 = 250$, so the probability is maximized when $N = 250$.

1.6 Exercises

Basic definitions

1. A man receives presents from his three children, Allison, Betty, and Chelsea. To avoid disputes he opens the presents in a random order. What are the possible outcomes?

2. Suppose we pick a number at random from the phone book and look at the last digit. (a) What is the set of outcomes and what probability should be assigned to each outcome? (b) Would this model be appropriate if we were looking at the first digit?

3. Two students arrive late for a math final exam with the excuse that their car had a flat tire. Suspicious, the professor says "each one of you write down on a piece of paper which tire was flat. What is the probability that both students pick the same tire?

4. Suppose we roll a red die and a green die. What is the probability the number on the red die is larger ($>$) than the number on the green die?

5. Two dice are rolled. What is the probability (a) the two numbers will differ by 1 or less, (b) the maximum of the two numbers will be 5 or larger?

6. If we flip a coin 5 times, what is the probability that the number of heads is an even number (i.e., divisible by 2)?

7. The 1987 World Series was tied at two games a piece before the St. Louis Cardinals won the fifth game. According to the Associated Press, "The numbers of history support the Cardinals and the momentum they carry. Whenever the series has been tied 2-2 the team that won the fifth game won the series 71% of the time." If momentum is not a factor and each team has a 50% chance of winning each game, what is the probaiblity that the game 5 winner will win the series?

8. Two boys are repeatedly playing a game that they each have probability 1/2 of winning. The first person to win five games wins the match. What is the probability that Al will win if (a) he has won 4 games and Bobby has won 3; (b) he leads by a score of 3 games to 2?

9. 20 families live in a neighborhood. 4 have 1 child, 8 have 2 children, 5 have 3 children, and 3 have 4 children. If we pick a child at random what is the probability they come from a family with 1, 2, 3, 4 children?

10. In Galileo's time people thought that when three dice were rolled, a sum of 9 and a sum of 10 had the same probability since each could be obtained in 6 ways:

$$9 : 1+2+6,\ 1+3+5,\ 1+4+4,\ 2+2+5,\ 2+3+4,\ 3+3+3$$
$$10 : 1+3+6,\ 1+4+5,\ 2+4+4,\ 2+3+5,\ 2+2+6,\ 3+3+4$$

Compute the probabilities of these sums and show that 10 is a more likely total than 9.

11. Suppose we roll three dice. Compute the probability that the sum is (a) 3, (b) 4, (c) 5, (d) 6, (e) 7, (f) 8 (g) 9 (h) 10.

12. In a group of students, 25% smoke cigarettes, 60% drink alcohol, and 15% do both. What fraction of students have at least one of these bad habits?

13. In a group of 320 high school graduates, only 160 went to college but 100 of the 170 men did. How many women did not go to college?

14. In the freshman class, 62% of the students take math, 49% take science, and 38% take both science and math. What percentage takes at least one science or math course?

15. 24% of people have American Express cards, 61% have Visa cards and 8% have both. What percentage of people have at least one credit card?

16. Suppose $\Omega = \{a, b, c\}$, $P(\{a, b\}) = 0.7$, and $P(\{b, c\}) = 0.6$. Compute the probabilities of $\{a\}$, $\{b\}$, and $\{c\}$.

17. Suppose A and B are disjoint with $P(A) = 0.3$ and $P(B) = 0.5$. What is $P(A^c \cap B^c)$?

18. Given two events A and B with $P(A) = 0.4$ and $P(B) = 0.7$, what are the maximum and minimum possible values for $P(A \cap B)$?

Permutations and Combinations

19. How many possible batting orders are there for nine baseball players?

20. A tire manufacturer wants to test four different types of tires on three different types of roads at five different speeds. How many tests are required?

21. 16 horses race in the Kentucky Derby. How many possible results are there for win, place, and show (first, second, and third)?

22. A school gives awards in five subjects to a class of 30 students but no one is allowed to win more than one award. How many outcomes are possible?

23. A tourist wants to visit six of America's ten largest cities. In how many ways can she do this if the order of her visits is (a) important, (b) not important?

24. Five businessmen meet at a convention. How many handshakes are exchanged if each shakes hands with all the others?

25. A commercial for Glade Plug-ins says that by inserting two of a choice of 11 scents into the device, you can make more than 50 combinations. If we exclude the boring choice of two of the same scent how many possibilities are there?

26. In a class of 19 students, 7 will get A's. In how many ways can this set of students be chosen?

27. (a) How many license plates are possible if the first three places are occupied by letters and the last three by numbers? (b) Assuming all combinations are equally likely, what is the probability the three letters and the three numbers are different?

28. How many four-letter "words" can you make if no letter is used twice and each word must contain at least one vowel (A, E, I, O or U)?

29. Assuming all phone numbers are equally likely, what is the probability that all the numbers in a seven-digit phone number are different?

30. A domino is an ordered pair (m, n) with $0 \le m \le n \le 6$. How many dominoes are in a set if there is only one of each?

31. A person has 12 friends and will invite 7 to a party. (a) How many choices are possible if Al and Bob are feuding and will not both go to the party? (b) How many choices are possible if Al and Betty insist that they both go or neither one goes?

32. A basketball team has 5 players over six feet tall and 6 who are under six feet. How many ways can they have their picture taken if the 5 taller players stand in a row behind the 6 shorter players who are sitting on a row of chairs?

33. The Duke basketball team has 10 women who can play guard and 12 tall women who can play the other three positions. At the start of the game, the coach gives the referee a starting line-up that lists who will play left guard, right guard, left forward, center, and right forward. In how many ways can this be done?

34. Six students, three boys and three girls, lineup in a random order for a photograph. What is the probability that the boys and girls alternate?

35. Seven people sit at a round table. How many ways can this be done if Mr. Jones and Miss Smith (a) must sit next to each other, (b) must not sit next to each other? (Two seating patterns that differ only by a rotation of the table are considered the same).

36. How many ways can eight rooks be put on a chess board so that no rook can capture any other rook? Or, what is the same: How many ways can eight markers be placed on an 8×8 grid of squares so that there is at most one in each row or column?

37. A BINGO card is a 5 by 5 grid. The center square is a free space and has no number. The first column is filled with five distinct numbers from 1 to 15, the second with five from from 16 to 30, the middle column with four numbers from 31 to 45, the fourth with five numbers from 46 to 60, and the fifth with five numbers from 61 to 75. Since the object of the game is to get five in a row horizontally, vertically or diagonally the order is important. How many BINGO cards are there?

38. Continuing with the set-up from the previous problem, in BINGO numbers are drawn from 1 to 75 without replacement. When a number is called you put a marker on that square. If you have five in a row horizontally, vertically or diagonally, you have a BINGO. What is the probability you will have a BINGO after (a) four numbers are called? (b) after five?

Multinomial Counting Problems

39. How many different ways can the letters in the following words be arranged: (a) money, (b) tattoo, (c) banana, (d) redredged, (e) statistics, (f) mississippi?

40. Ernie has 10 days to go before his mommy picks him up to take him home from college. He has 10 pairs of clean socks: 5 white, 3 brown and 2 black. How many different ways can he wear his socks over the next 10 days assuming he wears one pair per day.

41. Twelve different toys are to be divided among 3 children so that each one gets 4 toys. How many ways can this be done?

42. A club with 50 members is going to form two committees, one with 8 members and the other with 7. How many ways can this be done (a) if the committees must be disjoint? (b) if they can overlap?

43. (a) If six dice are rolled, what is the probability that each of the six numbers will appear at least once? (b) Answer the last question for seven dice.

44. How many ways can 5 history books, 3 math books, and 4 novels be arranged on a shelf if the books of each type must be together?

45. Suppose three runners from team A and three runners from team B have a race. If all six runners have equal ability, what is the probability that the three runners from team A will finish first, second, and fourth?

46. 9 boys and 7 girls are at a picnic. How many ways can we pick 4 coed teams (one boy and one girl) for a competition.

47. Eight girls show up for a ballroom dance class, but no boys do. How many ways can the eight girls divide themselves into four pairs to practice dancing. Note: in some dances one partner leads and the other follows, but here we assume there is no difference between the roles the two dancers play.

48. Four men and four women are shipwrecked on a tropical island. How many ways can they (a) form four male-female couples, (b) get married if we keep track of the order in which the weddings occur, (c) divide themselves into four unnumbered pairs, (d) split up into four groups of two to search the North, East, West, and South shores of the island, (e) walk single-file up the ramp to the ship when they are rescued, (f) take a picture to remember their ordeal if all eight stand in a line but each man stands next to his wife?

Expected Value

49. You want to invent a gambling game in which a person rolls two dice and is paid some money if the sum is 7, but otherwise he loses his money. How much should you pay him for winning a $1 bet if you want this to be a fair game, that is, to have expected value 0?

50. A bet is said to carry 3 to 1 odds if you win $3 for each $1 you bet. What must the probability of winning be for this to be a fair bet?

51. A lottery has one $100 prize, two $25 prizes, and five $10 prizes. What should you be willing to pay for a ticket if 100 tickets are sold?

52. In a popular gambling game, three dice are rolled. For a $1 bet you win $1 for each six that appears (plus your dollar back). If no six appears you lose your dollar. What is your expected value?

53. In a dice game the "dealer" rolls two dice, the player rolls two dice, and the player wins if his total is larger ($>$) than the dealer's. What is the probability the player wins?

54. A roulette wheel has slots numbered 1 to 36 and two labeled with 0 and 00. Suppose that all 38 outcomes have equal probability. Compute the expected values of the following bets. In each case you bet one dollar and when you win you get your dollar back in addition to your winnings. (a) You win $1 if one of the numbers 1 through 18 comes up. (b) You win $2 if the number that comes up is divisible by 3 (0 and 00 do not count). (c) You win $35 if the number 7 comes up.

55. In the Las Vegas game Wheel of Fortune, there are 54 possible outcomes. One is labeled "Joker," one "Flag," two "20," four "10," seven "5," fifteen "2," and twenty-four "1." If you bet $1 on a number you win that amount of money if the number comes up (plus your dollar back). If you bet $1 on Flag or Joker you win $40 if that symbol comes up (plus your dollar back). What bets have the best and worst expected value here?

56. Sic Bo is an ancient Chinese dice game played with 3 dice. One of the possibilities for betting in the game is to bet "big." For this bet you win if the total X is 11, 12, 13, 14, 15, 16, or 17, except when there are three 4's or three 5's. On a $1 bet on big you win $1 plus your dollar back if it happens. What is your expected value?

57. Five people play a game of "odd man out" to determine who will pay for the pizza they ordered. Each flips a coin. If only one person gets Heads (or Tails) while the other four get

Tails (or Heads) then he is the odd man and has to pay. Otherwise they flip again. What is the expected number of tosses needed to determine who will pay?

Urn Problems

58. How can 5 black and 5 white balls be put into two urns to maximize the probability a white ball is drawn when we draw from a randomly chosen urn?

59. Two red cards and two black cards are lying face down on the table. You pick two cards and turn them over. What is the probability that the two cards are different colors?

60. Four people are chosen at random from 5 couples. What is the probability two men and two women are selected?

61. You pick 5 cards out of a deck of 52. What is the probability you get exactly 2 spades?

62. Seven students are chosen at random from a class with 17 boys and 13 girls. What is the probability that 4 boys and 3 girls are selected?

63. In a carton of 12 eggs, 2 are rotten. If we pick 4 eggs to make an omelet, what is the probability we do not get a rotten egg?

64. An electronics store receives a shipment of 30 calculators of which 4 are defective. Six of these calculators are selected to be sent to a local high school. What is the probability that exactly one is defective?

65. A scrabble set contains 54 consonants, 44 vowels, and 2 blank tiles. Find the probability that your initial draw contains 5 consonants and 2 vowels.

66. (a) How many ways can can we pick 4 students from a group of 30 to be on the math team? (b) Suppose there are 18 boys and 12 girls. What is the probability the team will have 2 boys and 2 girls.

67. The following probability problem arose in a court case concerning possible discrimination against black nurses. 26 white nurses and 9 black nurses took an exam. All the white nurses passed but only 4 of the black nurses did. What is the probability we would get this outcome if the five nurses who failed were chosen at random?

68. A closet contains 8 pairs of shoes. You pick out 5. What is the probability of (a) no pair, (b) exactly one pair, (c) two pairs?

69. A drawer contains 10 black, 8 brown, and 6 blue socks. If we pick two socks at random, what is the probability they match?

70. A dance class consists of 12 men and 10 women. Five men and five women are chosen and paired up to dance. In how many ways can this be done?

71. Suppose we pick 5 cards out of a deck of 52. What is the probability we get at least one card of each suit?

72. A bridge hand in which there is no card higher than a 9 is called a *Yarborough* after the Earl who liked to bet at 1000 to 1 that your bridge hand would have a card that was 10 or higher. What is the probability of a Yarborough when you draw 13 cards out of a deck of 52?

73. Two cards are a blackjack if one is an A and the other is a K, Q, J, or 10. (a) If you pick two cards out of a deck, what is the probability you will get a blackjack? (b) Suppose you

are playing blackjack against the dealer with a freshly shuffled deck. What is the probability that you or the dealer will get a blackjack?

74. A student studies 12 problems from which the professor will randomly choose 6 for a test. If the student can solve 9 of the problems, what is the probability she can solve at least 5 of the problems on the test?

75. A football team has 16 seniors, 12 juniors, 8 sophomores, and 4 freshmen. (a) If we pick 5 players at random, what is the probability we will get 2 seniors and 1 from each of the other 3 classes? (b) Find an outcome with larger probability than $2 - 1 - 1 - 1$.

76. In a kindergarten class of 20 students, one child is picked each day to help serve the morning snack. What is the probability that in one week five different children are chosen?

77. An investor picks 3 stocks out of 10 recommended by his broker. Of these, six will show a profit in the next year. What is the probability the investor will pick (a) 3 (b) 2 (c) 1 (d) 0 profitable stocks?

78. Four red cards (i.e., hearts and diamonds) and four black cards are face down on the table. A psychic who claims to be able to locate the four black cards turns over 4 cards and gets 3 black cards and 1 red card. What is the probability he would get 3 or 4 black cards if he were guessing?

79. A town council considers the question of closing down an "adult" theatre. The five men on the council all vote against this and the three women vote in favor. What is the probability we would get this result (a) if the council members determined their votes by flipping a coin? (b) if we assigned the five "no" votes to council members chosen at random?

80. An urn contains white balls numbered 1 to 15 and black balls also numbered 1 to 15. Suppose you draw 4 balls. What is the probability that (a) no two have the same number? (b) you get exactly one pair with the same number? (c) you get two pair with the same numbers?

81. A town has four TV repairmen. In the first week of September four TV sets break and their owners call repairmen chosen at random. Find the probability that the number of repairmen who have jobs is 1, 2, 3, 4.

82. **Powerball lottery.** In this lottery five balls are picked from white balls numbered $1, \ldots 69$ and a red powerball from another urn with balls numbered $1, \ldots 26$. If three of the white numbers you picked are drawn and you get the correct power ball then we call the result $3 + 1$. Compute the probabilities of the following results for a $2 bet:

result	prize	odds
5+1	split jackpot	1/292,201,338
5+0	1 million	1/11,688,053
4+1	$50,000	1/913,129
4+0	$100	1/36,525
3+1	$100	1/14,494
3+0	$ 7	1/597.76
2+1	$ 7	1/701.33
1+1	$4	1/91.98
0+1	$4	1/38.22

83. Compute the probabilities of the following poker hands when we roll five six sided dice.

(a) five of a kind	0.000771
(b) four of a kind,	0.019290
(c) a full house,	0.038580
(d) three of a kind	0.154320
(e) two pair	0.231481
(f) one pair	0.462962
(g) no pair	0.092592

The probabilities in the list above add to 1. When we have no pair, we could have a straight: 65432 or 54321. (h) compute the probabiility of a straight.

84. In seven-card stud you receive seven cards and use them to make the best poker hand you can. Ignoring the possibility of a straight or a flush the probability that the best hand you can make with your cards is

	seven cards	five cards
(a) four of a kind,	0.001680	0.000240
(b) a full house,	0.025968	0.001441
(c) three of a kind	0.049254	0.021128
(d) two pair	0.240113	0.047539
(e) one pair	0.472839	0.422569
(f) no pair	0.210150	0.507082

Verify the probabilities for seven card stud. Hint: For full house you need to consider hand patterns: 3-3-1 and 3-2-2 in addition to the more likely 3-2-1-1. For two pair you also have to consider the possibility of three pair.

85. If we get five cards then the probability of a flush is 0.001981. The goal of this problem is to calculate the probability of a flush in seven card stud. We begin by noting that

$$C_{52,7} = 133,484,560$$

For $5 \leq k \leq 7$, let p_k be the probability we get exactly k cards of one suit when we draw 7 from a deck of 52. Find p_5, p_6, p_7. Then find the answer to our question $p_5 + p_6 + p_7$.

Chapter 2

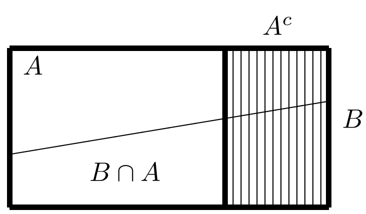
Independence

2.1 Conditional Probability

Suppose we are told that the event A with $P(A) > 0$ occurs. Then (i) only the part of B that lies in A can possibly occur, and (ii) since the sample space is now A, we have to divide by $P(A)$ to make $P(A|A) = 1$. Thus the probability that B will occur given that A has occurred is

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{2.1}$$

Then (i) only the part of B that lies in A can possibly occur, and (ii) since the sample space is now A, we have to divide by $P(A)$ to make $P(A|A) = 1$.



Example 2.1. Suppose we roll two dice. Let A = "the sum is 8," and B = "the first die is 3." $A = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$, so $P(A) = 5/36$. $A \cap B = \{(3,5)\}$, so

$$P(B|A) = \frac{1/36}{5/36} = \frac{1}{5}$$

The same result holds if B = "The first die is k" and $2 \le k \le 6$. Carrying this reasoning further, we see that given the outcome lies in A, all five possibilities have the same probability. This should not be surprising. The original probability is uniform over the 36 possibilities, so when we condition on the occurrence of A, its five outcomes are equally likely.

As the last example may have suggested, the mapping $B \to P(B|A)$ is a probability. That is, it is a way of assigning numbers to events that satisfies the axioms introduced in Chapter 1.

Intuitively, two events A and B are independent if the occurrence of A has no influence on the probability of occurrence of B. Expressed in terms of conditional probability this says

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

Rearranging we get the formal definition: A and B are **independent** if

$$P(A \cap B) = P(A)P(B)$$

Turning to concrete examples, in each case it should be clear that the intuitive definition is satisfied, so we will only check the formal one.

- Flip two coins. Let $A =$ "the first coin shows Heads," and $B =$ "the second coin shows Heads." $P(A) = 1/2$, $P(B) = 1/2$, $P(A \cap B) = 1/4$.
- Roll two dice. Let $A =$ "the first die shows 5," and $B =$ "the second die shows 2." $P(A) = 1/6$, $P(B) = 1/6$, $P(A \cap B) = 1/36$.
- Pick a card from a standard deck of 52 cards. Let $A =$ "the card is an ace," and $B =$ "the card is a spade." $P(A) = 1/13$, $P(B) = 1/4$, $P(A \cap B) = 1/52$.

A trivial example of events that are not independent is

Example 2.2. If A and B are disjoint events that have positive probability, they are not independent since $P(A)P(B) > 0 = P(A \cap B)$.

A concrete example is

Example 2.3. Draw two cards from a deck. Let $A =$ "the first card is a spade," and $B =$ "the second card is a spade." Then $P(A) = 1/4$ and $P(B) = 1/4$, but

$$P(A \cap B) = \frac{13 \cdot 12}{52 \cdot 51} < \left(\frac{1}{4}\right)^2$$

Sometimes it is not intuitively clear if events are independent. In this case we just have to check the definition.

Example 2.4. Roll two dice. Let $A =$ "the sum of the two dice is 7," and $B =$ "the first die is 2." $A = \{(6,1),(5,2),(4,3),(3,4),(2,5),(1,6)\}$, so $P(A) = 6/36$. $P(B) = 1/6$, Since $A \cap B = \{(2,5)\}$ we do have $P(A \cap B) = 1/36 = P(A) \cdot P(B)$. This only works for a sum of 7.

Independence for more than two events

There are two ways of extending the definition of independence to more than two events. $A_1, \ldots, A_n$ are said to be **pairwise independent** if for each $i \neq j$, $P(A_i \cap A_j) = P(A_i)P(A_j)$, that is, each pair is independent. $A_1, \ldots, A_n$ are said to be **independent** if for any $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ we have

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

The definition may look intimidating but it is easy to check in most situations. If we flip n coins and let A_i = "the ith coin shows Heads," then the events A_i are independent since $P(A_i) = 1/2$ and $P(A_{i_1} \cap \ldots \cap A_{i_k}) = 1/2^k$. Likewise if we roll n dice and let A_i be the event that the ith roll is a four, then the A_i are independent.

We have already seen an example of events that are pairwise independent but not independent:

Example 2.5. Birthdays. Let A = "Alice and Betty have the same birthday," B = "Betty and Carol have the same birthday," C = "Carol and Alice have the same birthday." Each pair of events is independent but the three are not.

Since there are 365 ways two girls can have the same birthday out of 365^2 possibilities (as in Example 1.14, we are assuming that leap year does not exist and that all the birthdays are equally likely), $P(A) = P(B) = P(C) = 1/365$. Likewise, there are 365 ways all three girls can have the same birthday out of 365^3 possibilities, so

$$P(A \cap B) = \frac{1}{365^2} = P(A)P(B)$$

i.e., A and B are independent. Similarly, B and C are independent and C and A are independent, so A, B, and C are pairwise independent. The three events A, B, and C are not independent, however, since $A \cap B = A \cap B \cap C$ and hence

$$P(A \cap B \cap C) = \frac{1}{365^2} \neq \left(\frac{1}{365}\right)^3 = P(A)P(B)P(C)$$

The last examples is somewhat unusual. However, the moral of the story is that to show several events are independent, you have to check more than just that each pair is independent.

2.2 Geometric, Binomial and Multinomial

All three distributions involve independent trials. The **Geometric(p) distribuion** is the distribution of the number of trials N to get a success when success has probability p. For $N = n$ we need $n - 1$ failures followed by a success so

$$P(N = n) = (1-p)^{n-1}p \qquad \text{for } n \geq 1.$$

Theorem 2.1. *The geometric(p) distribution has mean $1/p$.*

Proof. To get this from the definition, we begin with the sum of the geometric series

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \tag{2.2}$$

and differentiate with respect to x to get

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

We dropped the $k = 0$ term from the left since it is 0. Setting $x = 1 - p$

$$\sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2}$$

Multiplying each side by p we have $\sum_{k=1}^{\infty} kP(N = k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = 1/p$. □

The mean of the geometric is easier to calculate if we know

Theorem 2.2. *If X is a nonnegative integer valued random variable then*

$$EX = \sum_{n=1}^{\infty} P(X \geq n) \tag{2.3}$$

Proof Doing some algebra and then interchanging the order of summation:

$$\begin{aligned}\sum_{n=1}^{\infty} P(X \geq n) &= \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} P(X = m) \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{m} P(X = m) = \sum_{m=1}^{\infty} mP(X = m) = EX \qquad \square\end{aligned}$$

Applying (2.3) to the previous example we get

$$EX = \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{1}{1-(1-p)} = \frac{1}{p}$$

2.2.1 Binomial

We begin with a concrete example before we discuss the general case.

Example 2.6. Suppose we roll 6 dice. What is the probability of $A =$"we get exactly two 4's"?

One way that A can occur is

$$\frac{\times}{1}\,\frac{4}{2}\,\frac{\times}{3}\,\frac{4}{4}\,\frac{\times}{5}\,\frac{\times}{6}$$

where $\times$ stands for "not a 4." Since the six events "die one shows $\times$," "die two shows 4," ..., "die six shows $\times$" are independent, the indicated pattern has probability

$$\frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^4$$

Here, we have been careful to say "pattern" rather than "outcome" since the given pattern corresponds to 5^4 outcomes in the sample space of 6^6 possible outcomes for 6 dice. Each pattern that results in A corresponds to a choice of 2 of the 6 trials on which a 4 will occur, so the number of patterns is $C_{6,2}$. When we write out the probability of each pattern, there will be two 1/6's and four 5/6's so each pattern has the same probability and

$$P(A) = C_{6,2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^4$$

The **binomial(n,p) distribution**. Generalizing from the last example, suppose we perform an experiment n times and on each trial an event we call "success" has probability p. (Here and in what follows, when we repeat an experiment, we assume that the outcomes of the various trials are independent.) Then the probability of k successes is

$$C_{n,k} p^k (1-p)^{n-k} \tag{2.4}$$

since there are $C_{n,k}$ ways of picking k of the n trials for successes to occur, and each pattern of k successes and $n-k$ failures has probability $p^k(1-p)^{n-k}$.

Theorem 2.3. *The* **binomial(n,p) distribution** *has mean np.*

Proof. Using the definition of expected value

$$EN = \sum_{m=0}^{n} m \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}$$

The $m = 0$ term contributes nothing so we can cancel m's and rearrange to get

$$= np \sum_{m=1}^{n} \frac{(n-1)!}{(m-1)!(n-m)!} p^{m-1}(1-p)^{n-m} = np$$

since the sum computes the total probability for the binomial$(n-1,p)$ distribution. □

Example 2.7. A student takes a test with 20 multiple-choice questions, each of which has four possible answers. She has missed several classes so she knows 14 of the answers but must she has to choose at random from the 4 possible answers on the other 6. What is the probability she will get an A, i.e., at least 18 right?, a B, i.e., 17 or 16 right?

To begin to compute the answer, we recall

$$C_{6,1} = 6, \quad C_{6,2} = \frac{6 \cdot 5}{2} = 15 = C_{6,4}, \quad C_{6,3} = \frac{6 \cdot 5 \cdot 4}{6} = 20$$

20	$(1/4)^6$	1/4096
19	$6(1/4)^5(3/4)$	18/4096
18	$15(1/4)^4(3/4)^2$	135/4096
17	$20(1/3)^3(3/4)^3$	540/4096
16	$15(1/3)^2(3/4)^4$	1251/4096

so an A has probability $(1 + 18 + 135)/4096 = 0.0375$ while B hs probability $1691/4096 = 0.4128$. Thus with probability $1 - 0.0375 - 0.4218 = 0.5497$ she will get a C or worse.

Example 2.8. A football team wins each week with probability 0.6 and loses with probability 0.4. Suppose that the outcomes of their 12 games are independent. What is the probability (a) they will win exactly 7 games? (b) Win at least six games and hence be "bowl eligible."

By (2.4) the probability of $k = 7$ successes is

$$C_{12,7}(0.6)^7(0.4)^5 = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}(0.6)^7(0.4)^5 = 0.22703$$

With a TI-83 calculator this probability can be found by going to the DISTR menu and using binompdf(12,0.6,7). Here pdf is for probability density function.

(b) Let W be the number of games the team will win in the season. It would be tedious to calculate the probability $P(W \geq 6)$ by adding up the probability of $k = 6, 7, 8, \ldots 12$ games and adding up the answers. With a TI-83 calculator the probability of interest can be found by going by using binomcdf(12,0.6,5) which gives the $P(W \leq 5) = 0.1582$, and hence the answer is

$$P(W \geq 6) = 1 - P(W \leq 5) = 1 - 0.1582 = 0.8418.$$

Here cdf, is for cumulative distribution function, the name for the function $P(W \leq x)$, which will be defined precisely and studied in Section 3.2.

Example 2.9. Tennis. In men's tennis the winner is the first to win 3 out of 5 sets. If Roger Federer wins a set against his opponent with probability 2/3, what is the probability w that he will win the match?

He can win in three sets, four or five, but he must win the last set, so

$$\begin{aligned} w &= \left(\frac{2}{3}\right)^3 + C_{3,2}\left(\frac{2}{3}\right)^3 \frac{1}{3} + C_{4,2}\left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^2 \\ &= \left(\frac{2}{3}\right)^3 (1 + 3(1/3) + 6(1/9)) = \frac{8}{27} \cdot \frac{8}{3} = 0.790 \end{aligned}$$

Example 2.10. Aces at Bridge. When we draw 13 cards out of a deck of 52, each ace has a probability 1/4 of being chosen, but the four events are not independent. How does the probability of drawing two aces compare with that of the binomial distribution with $n = 4$ and $p = 1/4$?

The probability of drawing two aces:

$$\frac{C_{4,2}C_{48,11}}{C_{52,13}} = \frac{6 \cdot \dfrac{48 \cdots 38}{11!}}{\dfrac{52 \cdots 40}{13!}} = 6 \cdot \frac{13 \cdot 12 \cdot 39 \cdot 38}{52 \cdot 51 \cdot 50 \cdot 49} = 0.2135$$

In contrast, the probability for the binomial is

$$C_{4,2}(1/4)^2(3/4)^2 = 0.2109$$

Similar computations show:

	aces	binomial
0	$\dfrac{39 \cdot 38 \cdot 37 \cdot 36}{52 \cdot 51 \cdot 50 \cdot 49}$	$(3/4)^4$
1	$4 \cdot \dfrac{13 \cdot 39 \cdot 38 \cdot 37}{52 \cdot 51 \cdot 50 \cdot 49}$	$4 \cdot (1/4)(3/4)^3$
2	$6 \cdot \dfrac{13 \cdot 12 \cdot 39 \cdot 38}{52 \cdot 51 \cdot 50 \cdot 49}$	$6 \cdot (1/4)^2(3/4)^2$
3	$4 \cdot \dfrac{13 \cdot 12 \cdot 11 \cdot 39}{52 \cdot 51 \cdot 50 \cdot 49}$	$4 \cdot (1/4)^3(3/4)$
4	$\dfrac{13 \cdot 12 \cdot 11 \cdot 10}{52 \cdot 51 \cdot 50 \cdot 49}$	$(1/4)^4$

Evaluating these expressions leads to the following probabilities:

	aces	binomial
0	0.3038	0.3164
1	0.4388	0.4218
2	0.2134	0.2109
3	0.0412	0.0468
4	0.00264	0.00390

Example 2.11. In Bridge, each of the four players gets 13 cards from a deck of 52. In the last 8 games, Harry had 5 hands without an ace. Should he doubt that the cards are being shuffled properly?

By the previous problem, the number of hands in which Harry has no ace has a binomial distribution with $n = 8$ and $p = 0.3038$. The probability of exactly 5 hands without an ace is

$$C_{8,5}(0.3038)^5(0.6962)^3 = 0.0489$$

Of course if Harry had 6,7, or 8 hands without an ace he would complain, so the more relevant thing to compute is the probability of ≥ 5 which is

$$1 - \text{binomcdf}(8, 0.3038, 4) = 0.06097$$

Thus 6.1% of the time he will have five or more hands without an ace.

2.2.2 Multinomial Distribution

The arguments above generalize easily to independent events with more than two possible outcomes. We begin with an example.

Example 2.12. A student makes an A with probability 0.5, a B with probability 0.4, and a C with probability 0.1. In a year where she takes eight courses what is the probability of 4 A's, 3 B's, and 1 C.

The answer is

$$\frac{8!}{4!\,3!\,1!}(0.5)^4(0.4)^3(0.1)$$

The first factor, by (1.12), gives the number of ways to pick four courses for A's, three courses for B's, and one course for C. The second factor gives the probability of any outcome with four A's, three B's, and one C. The combinatorial coefficient is $8 \cdot 7 \cdot 5 = 280$ so the answer is 0.112.

Multinomial distribution. Generalizing from this example, we see that if we have k possible outcomes for our experiment with probabilities $p_1, \ldots, p_k$ then the probability of getting exactly n_i outcomes of type i in $n = n_1 + \cdots + n_k$ trials is

$$\frac{n!}{n_1! \cdots n_k!}\, p_1^{n_1} \cdots p_k^{n_k} \tag{2.5}$$

since the first factor gives the number of outcomes and the second the probability of each one.

Example 2.13. A baseball player gets a hit with probability 0.3, a walk with probability 0.1, and an out with probability 0.6. If he bats 4 times during a game and we assume that the outcomes are independent, what is the probability he will get 1 hit, 1 walk, and 2 outs?

The total number of trials $n = 4$. There are $k = 3$ categories hit, walk, and out. $n_1 = 1$, $n_2 = 1$, and $n_3 = 2$. Plugging in to our formula the answer is

$$\frac{4!}{1!1!2!}(0.3)^1(0.1)^1(0.6)^2 = 0.1296$$

Example 2.14. The output of a machine is graded excellent 70% of the time, good 20% of the time, and defective 10% of the time. What is the probability a sample of size 15 has 10 excellent, 3 good, and 2 defective items?

The total number of trials $n = 15$. There are $k = 3$ categories: excellent, good, and defective. We are interested in outcomes with $n_1 = 10$, $n_2 = 3$, and $n_3 = 2$. Plugging in to our formula the answer is

$$\frac{15!}{10!\,3!\,2!} \cdot (0.7)^{10}(0.2)^3(0.1)^2$$

2.3 Poisson Approximation to the Binomial

X is said to have a Poisson distribution with parameter λ, or Poisson(λ) if

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

Here $\lambda > 0$ is a parameter. To see that the probabilities sum to 1 we recall

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{2.6}$$

The next figure shows the Poisson distribution with $\lambda = 4$.

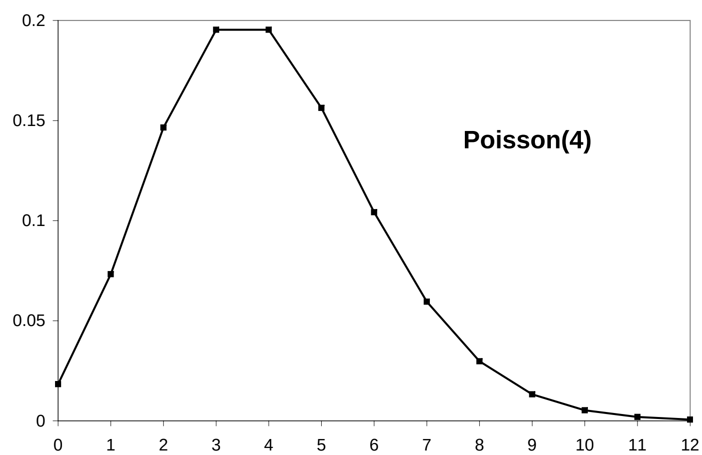


Figure 2.1: Poisson distribution $\lambda = 4$

As we will now show λ is the mean of the Poisson(λ) distribution.

Proof. Dropping the $k = 0$ term since it is zero

$$\begin{aligned} EX &= \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \lambda \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \end{aligned}$$

since the probabilitie sin the Poisson distribution sum to 1. □

Our next result explains why the Poisson distribution arises in a number of situations.

Theorem 2.4. *Suppose S_n has a binomial distribution with parameters n and p_n. If $p_n \to 0$ and $np_n \to \lambda$ as $n \to \infty$ then*

$$P(S_n = k) \to e^{-\lambda} \frac{\lambda^k}{k!} \tag{2.7}$$

In words, if we have a large number, n, of independent events with small probability p_n then the number that occur has approximately a Poisson distribution with mean $\lambda = np_n$. The key to the proof is the following fact:

Lemma. *If $p_n \to 0$ and $np_n \to \lambda$ then as $n \to \infty$*

$$(1 - p_n)^n \to e^{-\lambda} \tag{2.8}$$

Proof. Calculus tells us that if x is small then

$$\ln(1 - x) = -x - \frac{x^2}{2} - \dots$$

Using this we have

$$\begin{aligned} (1 - p_n)^n &= \exp(n \ln(1 - p_n)) \\ &\approx \exp(-np_n - np_n^2/2) \approx \exp(-\lambda) \end{aligned}$$

In the last step we used the assumption that $np_n \to \lambda$ and $p_n \to 0$ to conclude that $np_n \cdot p_n/2 \to 0$. □

Proof of (2.7) . Since $P(S_n = 0) = (1 - p_n)^n$ (2.8) gives the result for $k = 0$. To prove the result for $k > 0$, we let $\lambda_n = np_n$ and observe that

$$\begin{aligned} P(S_n = k) &= C_{n,k} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{\lambda_n}{n}\right)^{-k} \\ &\to 1 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 \end{aligned}$$

Here $n(n-1)\cdots(n-k+1)/n^k \to 1$ since there are k factors in the numerator and for each fixed j, $(n-j)/n = 1 - (j/n) \to 1$. The last term $(1 - \{\lambda_n/n\})^{-k} \to 1$ since k is fixed and $1 - \{\lambda_n/n\} \to 1$. □

In some situations it is not reasonable to make the assumption that all people have equal probabilities. For example, the number of people who go to a fast-food restaurant between 12 and 1, and the number of traffic accidents in a day. With these examples in mind we formulate a more general result.

Theorem 2.5. *Consider independent events A_i, $i = 1, 2, \ldots, n$ with probabilities $p_i = P(A_i)$. Let N be the number of events that occur, let $\lambda = p_1 + \cdots + p_n$, and let Z have a Poisson distribution with parameter λ. Then, for any set of integers B,*

$$|P(N \in B) - P(Z \in B)| \le \sum_{i=1}^{n} p_i^2 \le \lambda \max_i p_i$$

This result says that if all the p_i are small then the distribution of N is close to a Poisson with parameter λ. Taking $B = \{k\}$ we see that the individual probabilities $P(N = k)$ are all close to $P(Z = k)$, but this result says more. The probabilities of events such as $P(3 \le N \le 8)$ are close to $P(3 \le Z \le 8)$ and we have an explicit bound on the error.

When we apply (2.7) we think, "If $S_n = \text{binomial}(n, p)$ and p is small then S_n is approximately Poisson(np)." The next example illustrates the use of this approximation and shows that the number of trials does not have to be very large for us to get accurate answers.

Example 2.15. Suppose we roll two dice 12 times and we let D be the number of times a double 6 appears. Here $n = 12$ and $p = 1/36$, so $np = 1/3$. We will now compare $P(D = k)$ with the Poisson approximation for $k = 0, 1, 2$.

$k = 0$ exact answer:

$$P(D = 0) = \left(1 - \frac{1}{36}\right)^{12} = 0.7132$$

Poisson approximation: $P(D = 0) = e^{-1/3} = 0.7165$

$k = 1$ exact answer:

$$\begin{aligned} P(D = 1) &= C_{12,1} \frac{1}{36} \left(1 - \frac{1}{36}\right)^{11} \\ &= \left(1 - \frac{1}{36}\right)^{11} \cdot \frac{1}{3} = 0.2445 \end{aligned}$$

Poisson approximation: $P(D = 1) = e^{-1/3} \frac{1}{3} = 0.2388$

$k = 2$ exact answer:

$$\begin{aligned} P(D = 2) &= C_{12,2} \left(\frac{1}{36}\right)^2 \left(1 - \frac{1}{36}\right)^{10} \\ &= \left(1 - \frac{1}{36}\right)^{10} \cdot \frac{12 \cdot 11}{36^2} \cdot \frac{1}{2!} = 0.0384 \end{aligned}$$

Poisson approximation: $P(D = 2) = e^{-1/3} \left(\frac{1}{3}\right)^2 \frac{1}{2!} = 0.0398$

In this case the error bound in Theorem 2.5 is

$$\sum_{i=1}^{12} p_i^2 = 12 \left(\frac{1}{36}\right)^2 = \frac{1}{108} = 0.00926$$

while the error for the approximation for $k = 1$ is 0.0057.

When we apply the Poisson distribution to real data there is no underlying binomial distribution, so we use the fact that the parameter λ is the average value of the Poisson distribution.

Example 2.16. Shark Attacks. In the summer of 2001 there were 6 shark attacks in Florida, while the yearly average is 2. Is this unusual?

In an article in the September 7, 2001 National Post, Professor David Kelton of Penn State University argued that this was a random event. “Just because you see events happening in a rash this does not imply that there is some physical driver causing them to happen. It is characteristic of random processes that they have bursty behavior.” He did not seem to realize that the probability of six shark attacks under the Poisson distribution is

$$e^{-2}\frac{2^6}{6!} = 0.01203$$

This probability can be found with the TI-83 by using Poissonpdf(2,6) on DISTR menu. If we want the probability of at least six we would use $1 - \text{Poissoncdf}(2, 5) = 0.01656$.

Example 2.17. Death by horse kick. Ladislaus Bortkiewicz published a book about the Poisson distribution titled the *Law of Small Numbers* in 1898. In this book he analyzed the number of German soldiers kicked to death by cavalry horses between 1875 and 1894 in each of 14 cavalry corps, arguing that it fit the Poisson distribution. In the 20 years of data on 14 units there were 280 data points with a total of 196 deaths = 0.7 per unit per year. The next table gives the number of observed as well as the number expected if the number had a Poisson distribution with $\lambda = 0.7$

number	0	1	2	≥ 3
observed	144	91	32	13
expected	139.04	97.33	34.46	9.55

Example 2.18. Wayne Gretsky. He scored a remarkable 1669 points in 696 games as an Edmonton Oiler, for a rate of $1669/696 = 2.39$ points per game. From the Poisson formula with $k = 0$ the probability of Gretzky having a pointless game is $e^{-2.39} = 0.090$. The next table compares that actual number of games with the numbers predicted by the Poisson approximation.

points	games	Poisson
0	69	63.27
1	155	151.71
2	171	181.90
3	143	145.40
4	79	87.17
5	57	41.81
6	14	16.71
7	6	5.72
8	2	1.72
9	0	0.46

Example 2.19. March 30, 2005 powerball lottery. At that point in time the lottery picked 5 numbers out of $1, 2, \ldots 53$ and a powerball out of $1, 2, \ldots 42$. You won \$100,000 if

you matched the five white ball and missed the powerball, an event of probability

$$\frac{1}{C_{53,5}} \cdot \frac{41}{42} = \frac{1}{2,939,677}$$

To simplify the arithmetic let's round the denominator to 3 million. On the date in question about 12,000,000 people played the lottery, so we would expect the number of $100,000 winners to be approximately Poisson with mean 4. Hoever, the lottery officials were shocked to see 110 winners.

While at first criminal activity was suspected there was a simple explanation. Most of the people who played these numbers did so because they were in a fortune cookie manufactured by the Wonton Food Company of Long Island City in Queens, NY. This theory is confirmed by the fact that most winners had 40 as their powerball, which was what the fortune cookie recommended.

Example 2.20. Births in Ithaca. The Poisson distribution can be used for births as well as for deaths. There were 63 births in Ithaca, NY between March 1 and April 8, 2005, a total of 39 days, or 1.615 per day. The next table gives the observed number of births per day and compares with the prediction from the Poisson distribution. The agreement is good even though it is not sensible to assume that all women in the county have an equal probability of giving birth.

	0	1	2	3	4	5	6
observed	9	12	9	5	3	0	1
Poisson	7.75	12.52	10.11	5.44	2.19	.71	.19

Example 2.21. Birthday problem. In the birthday problem the events A_{ij} that i and j have the same birthday are not independent. Despite this we will use the Poisson approximation and compare with the exact result. When there are k people the mean is $C_{k,2}/365 = k(k-1)/700$ and hence the Poisson approximation to the probability of no birthday match is

$$\exp(-k(k-1)/700)$$

To see that this is reasonable note that by calculations in Example 1.14 the probability that k people all have different birthdays is

$$1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \cdot \left(1 - \frac{k-1}{365}\right)$$

using the fact that $1 - x \approx e^{-x}$ we can write this as

$$\approx \exp\left(-\sum_{j=1}^{k-1} j/365\right) = \exp(-k(k-1)/700)$$

since $\sum_{j=1}^{k} j = k(k=1)/2$. When $k = 23$ the approximation is 0.49999 versus the exact answer of 0.49270.

Example 2.22. Triple birthdays in the Senate. Using the Poisson approximation, the number of senators born on a given day is Binomial(100,1/365) and hence approximately

Poisson with mean 100/365. Using the Poisson approximation the probability that three senators were born on July 4 is

$$e^{-100/365}(100/365)^3/3! = 0.002606$$

Using the Poisson approximation again (assuming this event for different days are independent), the number of triple birthdays is approximately Poisson($365 \cdot 0.002606$) or Poisson(0.95119), so the probability of no triple birthday is ≈ 0.3862

2.4 Probabilities of Unions

In Section 1.1, we learned that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. In this section we will extend this formula to $n > 2$ events. We begin with $n = 3$ events:

$$\begin{aligned} P(A \cup B \cup C) = & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C) \end{aligned} \tag{2.9}$$

Proof. As in the proof of the formula for two events, we have to convince ourselves that the net number of times each part of $A \cup B \cup C$ is counted is 1. To do this, we make a table that identifies the areas counted by each term and note that the net number of pluses in each row is 1:

	A	B	C	$A \cap B$	$A \cap C$	$B \cap C$	$A \cap B \cap C$
$A \cap B \cap C$	+	+	+	−	−	−	+
$A \cap B \cap C^c$	+	+		−			
$A \cap B^c \cap C$	+		+		−		
$A^c \cap B \cap C$		+	+			−	
$A \cap B^c \cap C^c$	+						
$A^c \cap B \cap C^c$		+					
$A^c \cap B^c \cap C$			+				

Example 2.23. Suppose we roll three dice. What is the probability that we get at least one 6?

Let A_i = "we get a 6 on the ith die." Clearly,

$$\begin{aligned} & P(A_1) = P(A_2) = P(A_3) = 1/6 \\ & P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = 1/36 \\ & P(A_1 \cap A_2 \cap A_3) = 1/216 \end{aligned}$$

So plugging into (2.9) gives

$$P(A_1 \cup A_2 \cup A_3) = 3 \cdot \frac{1}{6} - 3 \cdot \frac{1}{36} + \frac{1}{216} = \frac{108 - 18 + 1}{216} = \frac{91}{216}$$

To check this answer, we note that $(A_1 \cup A_2 \cup A_3)^c$ = "no 6" = $A_1^c \cap A_2^c \cap A_3^c$ and $|A_1^c \cap A_2^c \cap A_3^c| = 5 \cdot 5 \cdot 5 = 125$ since there are five "non-6's" that we can get on each roll. Since there are $6^3 = 216$ outcomes for rolling three dice, it follows that $P(A_1^c \cap A_2^c \cap A_3^c) = 125/216$ and $P(A_1 \cup A_2 \cup A_3) = 1 - P(A_1^c \cap A_2^c \cap A_3^c) = 91/216$.

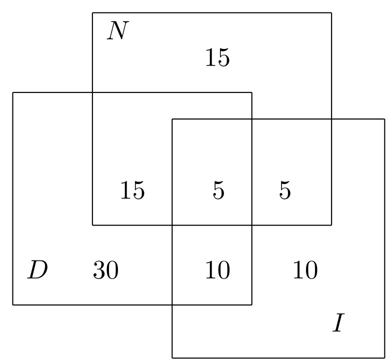
The same reasoning applies to sets.

Example 2.24. In a freshman dorm, 60 students read the Cornell Daily Sun, 40 read the New York Times and 30 read the Ithaca Journal. 20 read the Cornell Daily Sun and the New York Times, 15 read the Cornell Daily Sun and the Ithaca Journal, 10 read the New York Times and the Ithaca Journal, and 5 read all three. How many read at least one newspaper?

Using our formula the answer is

$$60 + 40 + 30 - 20 - 15 - 10 + 5 = 90$$

To check this we can draw picture using D, N, and I for the three newspapers



To figure out the number of students in each category we work out from the middle. $D \cap N \cap I$ has 5 students and $D \cap N$ has 20, so $D \cap N \cap I^c$ has 15. In the same way we compute that $D \cap N^c \cap I$ has $15 - 5 = 10$ students and $D^c \cap N \cap I$ has $10 - 5 = 5$ students. Having found that 30 of students in D read at least one other newspaper, the number who read only D is $60 - 30 = 30$. In a similar way, we compute that there are $40 - 25 = 15$ students who only read N and $30 - 20 = 10$ students who only read I. Adding up the numbers in the seven regions gives a total of 90, as we found before.

If we want to compute the number of students who read only the New York Times this can be done more easily using

$$\begin{aligned} |N \cap D^c \cap I^c| &= |N| - |N \cap D| - |N \cap I| + |N \cap D \cap I| \\ &= 40 - 20 - 10 + 5 = 15 \end{aligned}$$

To check this we see how many times each region is counted by the right-hand side

	$\|N\|$	$-\|N \cap D\|$	$-\|N \cap I\|$	$\|N \cap D \cap I\|$
$N \cap D^c \cap I^c$	$+$			
$N \cap D^c \cap I$	$+$		$-$	
$N \cap D \cap I^c$	$+$	$-$		
$N \cap D \cap I$	$+$	$-$	$-$	$+$

If we add up the $+1$ and -1 we get 1 on the first row and 0 on the others.

2.4.1 Inclusion-exclusion formula

Formula (2.9) generalizes to n events:

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k)$$
$$\cdots + (-1)^{n+1} P(A_1 \cap \ldots \cap A_n) \tag{2.10}$$

In words, we take all possible intersections of one, two, ... n events and the signs of the sums alternate. As in the case of three sets there is ana analogous formula for countint the number of elements in a union.

Proof. A point that is in exactly k sets is counted k times by the first sum, $C_{k,2}$ times by the second, $C_{k,3}$ times by the third, and so on until it is counted $C_{k,k} = 1$ time by the kth term. The net result is

$$C_{k,1} - C_{k,2} + C_{k,3} \ldots + (-1)^{k+1} 1$$

To show that this adds up to 1, we recall the Binomial theorem

$$(a+b)^k = a^k + C_{k,1} a^{k-1} b + C_{k,2} a^{k-2} b^2 + \cdots + b^k$$

Setting $a = 1$ and $b = -1$ we have

$$0 = 1 - C_{k,1} + C_{k,2} - C_{k,3} \ldots - (-1)^{k+1}$$

which proves the desired result. □

Example 2.25. We pick 13 cards out of a deck of 52, What is the probability that we get exactly five cards of one suit?

Let A_i be the event that we get five cards of suit i. Since it is impossible to have five cards in more than two suits (2.10) becomes

$$P(\cup_{i=1}^4 A_i) = \sum_{i=1}^4 P(A_i) - \sum_{1 \le i < j \le 4} P(A_i \cap A_j)$$

To find the desired probability we note that if $1 \le i < j \le 4$

$$P(A_i) = \frac{C_{13,5} C_{39,8}}{C_{52,13}} = \frac{7.918 \times 10^{10}}{6.350 \times 10^{11}} = 0.12469$$
$$P(A_i \cap A_j) = \frac{C_{13,5} C_{13,5} C_{26,3}}{C_{52,13}} = \frac{4.306 \times 10^9}{6.350 \times 10^{11}} = 0.00678$$

so taking into account the number of terms of each type we have

$$P\left(\cup_{i=1}^4 A_i\right) = 4(0.12469) - 6(0.00678) = 0.4588$$

Example 2.26. We now describe a situation which is perhaps the most famous application of inclusion-exclusion. An absent minded professor hands tests back at random. What is the probability of B, the event that no one of the n students gets their own test back.

Let A_i be the event that the ith student gets their test back. $P(A_i) = 1/n$ so

$$\sum_{i=1}^{n} P(A_i) = n \cdot \frac{1}{n}$$

As we compute the higher order terms a pattern develops:

$$\begin{aligned}
-\sum_{1\le i<j\le n} P(A_i \cap A_j) &= \frac{n(n-1)}{2!} \cdot \frac{1}{n(n-1)} = \frac{1}{2!} \\
\sum_{1\le i<j<k\le n} P(A_i \cap A_j \cap A_k) &= \frac{n(n-1)(n-2)}{3!} \cdot \frac{1}{n(n-1)(n-2)} = \frac{1}{3!} \\
-\sum_{1\le i<j<k<\ell\le n} P(A_i \cap A_j \cap A_k \cap A_\ell) & \\
&= \frac{n(n-1)(n-2)(n-3)}{4!} \cdot \frac{1}{n(n-1)(n-2)(n-3} = -\frac{1}{4!}
\end{aligned}$$

From this we see that

$$P(B_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} \ldots + \frac{(-1)^{n+1}}{n!} \approx 1 - e^{-1}$$

where the last equality follows from the fact that $e^x = 1 + x + x^2/2! + x^3/3! + \cdots$. The approximation is very good. Since the series alternates in sign and the magnitude of the terms are decreasing,

$$|P(B_n) - (1 - e^{-1})| \le 1/(n+1)!$$

To motivate the next topic we consider

Example 2.27. Suppose we roll a die 15 times. What is the probability that we do not see each of the 6 numbers at least once?

Let A_i be the event that we never see i. Since $P(A_1 \cap A_2 \ldots A_6$ is 0, the inclusion exclusion formula tells us that

$$\begin{aligned}
P(\cup_{i=1}^{6} A_i) &= \sum_{i=1}^{6} P(A_i) - \sum_{1\le i<j\le 6} P(A_i \cap A_j) + \sum_{1\le i<j<k\le 6} P(A_i \cap A_j \cap A_k) \\
&+ \sum_{1\le i<j<k<\ell\le 6} P(A_i \cap A_j \cap A_k \cap A_\ell) + \sum_{1\le i<j<k<\ell<m\le 6} P(A_i \cap A_j \cap A_k \cap A_\ell \cap A_m)
\end{aligned}$$

Taking into account the number of terms in each sum the above

$$\begin{aligned}
&= 6(5/6)^{15} - 15(4/6)^{15} + 20(3/6)^{15} - 15(2/6)^{15} + 6(1/6)^{15} \\
&= 0.389433 - 0.03426 + 3.05 \times 10^{-5} - 1.045 \times 10^{-6} + 1.27 \times 10^{-11}
\end{aligned}$$

The last two terms are very small so itis waste of effort to compute them.

2.4.2 Bonferroni inequalities

In brief, if you stop the inclusion-exclusion formula with a + term you get an upper bound; if you stop with a − term you get a lower bound.

$$P(\cup_{i=1}^n A_i) \le \sum_{i=1}^n P(A_i) \tag{2.11}$$

$$\ge \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) \tag{2.12}$$

$$\le \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \tag{2.13}$$

Proof. The first inequality is obvious since the right-hand side counts each outcome in $\cup_{i=1}^n A_i$ at least once. To prove the second, consider an outcome that is in exactly k sets. If $k = 1$, the first term will count it once and the second not at all. If $k = 2$ the first term counts it 2 times and the second once, with a net total of 1. If $k \ge 3$ the first term counts it k times and the second $C_{k,2} = k(k-1)/2 > k$ times so the net number of countings is < 0.

The third formula is similar.

in k sets	counted
1	1 - 0 + 0 =1
2	2 - 1 + 0 =1
3	3 - 3 + 1 =1

When ≥ 4 the number of countings is

$$C_{k,1} - C_{k,2} + C_{k,3} > k - \frac{k(k-1)}{2} + (k-1)(k-2) \ge 0 \qquad \square$$

Returning to the previous example

Example 2.28. Suppose we roll a die 15 times. What is the probability that we do not see each of the 6 numbers at least once?

Let A_i be the event that we never see i. Earlier we computed that

$$\sum_{i=1}^6 P(A_i) = 6(0.064905) = 0.389433$$

so we have

$$P\left(\cup_{i=1}^6 A_i\right) \le 0.389433$$

Using the next term

$$\sum_{i<j} P(A_i \cap A_j) = 15(0.002284) = 0.03426$$

and it follows that

$$P\left(\cup_{i=1}^6 A_i\right) \ge 0.389433 - 0.03426 = 0.355177$$

The third term is

$$\sum_{i<j<k} P(A_i \cap A_j \cap A_k) = 20(3.05 \times 10^{-5}) = 0.00061$$

and it follows that

$$P\left(\cup_{i=1}^{6} A_i\right) \leq 0.355177 + 0.00061 = 0.355738$$

The upper bound and the lower bound are very close so we can stop.

Example 2.29. The Streak. In the summer of 1941, Joe DiMaggio achieved what many people consider the greatest record in sports, in which he had at leastone hit in each of 56 games. What is the probability of this event?

A useful trick. Suppose for the moment that we know the probability p that Joe DiMaggio gets a hit in one game, and that successive games are independent. Assuming a 154-game season, we could let A_i be the probability that a player got hits in games $i+1, \ldots i+56$ for $0 \leq i \leq 98$. Using (2.11) it follows that the probability of the streak is

$$\leq 99p^{56}$$

The next sum is hard to evaluate because if $i < j$

$$P(A_i \cap A_j) = \begin{cases} p^{56}p^{j-i} & \text{if } j-i < 56 \\ p^{112} & \text{if } j-i < 56 \end{cases}$$

In words if A_i occurs, it becomes much easier for A_{i+1} and other "nearby" events to occur.

To avoid this problem, we will let B_i be the event the player gets no hit in game i but has hits in games $i+1, i+2, \ldots, i+56$ where $1 \leq i \leq 98$. Ignoring the probability of having hits in games $1, 2, \ldots 56$ the event of interest $S = \cup_{i=1}^{98} B_i$, so

$$P(S) \leq q \equiv 98p^{56}(1-p)$$

To compute the second bound we begin by noting $B_i \cap B_j = \emptyset$ if $i < j \leq i + 56$ since B_i requires a hit in game j while B_j requires no hit. If $56 + i < j \leq 98$ then $P(B_i \cap B_j) = P(B_i)P(B_j)$. To simplify the arithmetic we note that in either case $P(B_i \cap B_j) \leq P(B_i)P(B_j)$, so

$$\sum_{1 \leq i < j \leq 98} P(B_i \cap B_j) \leq C_{98,2}p^{112}(1-p)^2 \leq \frac{q^2}{2}$$

This is the number we have to subtract from the upper bound to get the lower bound, so we have

$$q \geq P(S) \geq q - \frac{q^2}{2} \tag{2.14}$$

Since q will end up being very small, the ratio of the two bounds is $1 - (q/2) \approx 1$.

To compute the probability p that Joe DiMaggio gets a hit in one game, we will introduce two somewhat questionable assumptions: (i) a player gets exactly four at bats per game (during the streak, DiMaggio averaged 3.98 at bats per game) and (ii) the outcomes of different at bats are independent with the probability of a hit being 0.325, Joe DiMaggio's lifetime batting average. From assumptions (i) and (ii) it follows that the probability

$$p = 1 - (0.675)^4 = 0.7924$$

and using (2.14) we have

$$P(S) \approx q = 98(0.7924)^{56}(0.2076) = 4.46 \times 10^{-5}$$

To interpret our result, note that the probability in (2.14) is roughly 1/22,000, so even if there were 220 players with 0.325 batting averages, it would take 100 years for this to occur again.

Example 2.30. A less famous streak. *Sports Illustrated* reports that a high school football team in Bloomington, Indiana lost 21 straight pre-game coin flips before finally winning one. Taking into account the fact that there are approximately 15,000 high school and college football teams, is this really surprising?

We will first compute the probability that this happens to one team some time in the decade 1995–2004 assuming that the team plays 10 games per year. Taking a lesson from the previous example, we let B_i be the event that the team won the coin flip in game i but lost it in games $i+1, \dots i+21$. Using the reasoning that led to (2.14)

$$P(S) \approx 79(1/2)^{22} = 1.883 \times 10^{-5}$$

What we have computed is the probability that one particular team will have this type of bad luck some time in the last decade. The probability that none of the 15,000 teams will do this is

$$(1 - 79(0.5)^{22})^{15{,}000} = 0.7539$$

i.e., with probability 0.2461 some team will have this happen to them. As a check on the last calculation, note that (2.11) gives an upper bound of

$$15,000 \times 1.883 \times 10^{-5} = 0.2825$$

2.5 Exercises

Conditional probability

1. A friend flips two coins and tells you that at least one is Heads. Given this information, what is the probability that the first coin is Heads?

2. A friend rolls two dice and tells you that there is at least one 6. What is the probability the sum is at least 9?

3. Suppose we roll two dice. What is the probability that the sum is 7 given that neither die showed a 6?

4. Suppose you draw five cards out of a deck of 52 and get 2 spades and 3 hearts. What is the probability the first card drawn was a spade?

5. Suppose 60% of the people subscribe to newspaper A, 40% to newspaper B, and 30% to both. If we pick a person at random who subscribes to at least one newspaper, what is the probability she subscribes to newspaper A?

6. In a town 40% of families have a dog and 30% have a cat. 25% of families with a dog also have a cat. (a) What fraction of people have a dog or cat? (b) What is the probability a family with a cat has a dog?

7. Two events have $P(A) = 1/4$, $P(B|A) = 1/2$, and $P(A|B) = 1/3$. Compute $P(A \cap B)$, $P(B)$, $P(A \cup B)$.

8. A, B, and C are events with $P(A) = 0.3$, $P(B) = 0.4$, $P(C) = 0.5$, A and B are disjoint, A and C are independent, and $P(B|C) = 0.1$. Find $P(A \cup B \cup C)$.

Independence

9. Suppose we draw two cards out of a deck of 52. Let A = "the first card is an Ace," and B = "the second card is a spade." Are A and B independent?

10. A family has three children, each of whom is a boy or a girl with probability 1/2. Let A = "there is at most 1 girl," B = "the family has children of both sexes." (a) Are A and B independent? (b) Are A and B independent if the family has four children?

11. Nine children are seated at random in three rows of three desks. Let A = "Al and Bobby sit in the same row," B = "Al and Bobby both sit at one of the four corner desks." Are A and B independent?

12. Suppose we roll a red and a green die. Let A = "the red die shows a 2 or a 5," B = "the sum of the two dice is at least 7." Are A and B independent?

13. Roll two dice. Let A = "the sum is even," and B = "the sum is divisible by 3," i.e., $B = \{3, 6, 9, 12\}$. Are A and B independent?

14. Roll two dice. Let A = the larger number is 5. B = the sum is 7 or 11. Are A and B independent?

15. We flip three coins with probability p of heads where $0 < p < 1$. Let A = the first and second flips are different. Let B = the second fand third flips are different. For what values of p are A and B independent.

16. Roll two dice. Let A = "the first die is odd," B = "the second die is odd," and C = "the sum is odd." Show that these events are pairwise independent but not independent.

17. A number X is picked at random from $\{1, 2, \ldots 100\}$. Are the following events independent (a) $A_1 = X$ is even, $B_1 = X$ is divisible by 5. (b) $A_2 = X$ is divisible by 4, $B_2 = X$ is divisible by 3. (c) What are the answers if X is picked at random from $\{1, 2, \ldots 36\}$.

18. Let A = "the sum of the two dice is a prime number," i.e., 2, 3, 5, 7, or 11. B = "the first die is 2 or 6." Are A and B independent?

19. Two students, Alice and Betty, are registered for a statistics class. Alice attends 80% of the time, Betty 60% of the time, and their absences are independent. On a given day, what is the probability (a) at least one of these students is in class, (b) exactly one of them is there?

20. Marilyn vos Savant claimed in her May 5, 2013 column that among the four digit numbers 0000 to 9999, exactly 4000 had at least one five. Compute the number of numbers with at least one 5 by computing the probability that a randomly chosen number has no 5.

21. At Duke, 20% of students like football (event F) and 60% like basketball (event B). Assuming these two events are independent calculate the fraction of students who like at least one of these sports.

22. Let A and B be two independent events with $P(A) = 0.4$ and $P(A \cup B) = 0.64$. What is $P(B)$?

23. A and B are independent events with $P(A) > P(B)$. $P(A \cap B) = 0.12$, $P(A \cup B) = 0.58$. Find $P(A)$ and $P(B)$.

24. Three students each have probability 1/3 of solving a problem. What is the probability at least one of them will solve the problem?

25. Three independent events have probabilities 1/4, 1/3, and 1/2. What is the probability exactly one will occur?

26. Three missiles are fired at a target. They will hit it with probabilities 0.2, 0.4, and 0.6. Find the probability that the target is hit by (a) three, (b) two, (c) one, (d) no missiles.

27. Three couples that were invited to dinner will independently show up with probabilities 0.9, 8/9, and 0.75. Let N be the number of couples that show up. Calculate the probability $N = 3, 2, 1, 0$.

28. In men's tennis the first person to win three sets wins the match. Suppose Rafael Nadal has probability of 0.6 of winning a set when he plays against Andy Murray. What is the probability that Nadal will win the match?

29. A college student takes 4 courses a semester for 8 semesters. In each course she has a probability 1/2 of getting an A. Assuming her grades in different courses are independent, what is the probability she will have at least one semester with all A's?

30. When Al and Bob play tennis, Al wins a set with probability 0.7 while Bob wins with probability 0.3. What is the probability Al will be the first to win (a) two sets, (b) three sets?

31. Chevalier de Mere made money betting that he could "roll at least one 6 in four tries." When people got tired of this wager he changed it to "roll at least one double 6 in 24 tries" but then he started losing money. Compute the probabilities of winning these two bets.

32. Samuel Pepys wrote to Isaac Newton: "What is more likely, (a) at least one 6 in 6 rolls of one die or (b) at least two 6's in 12 rolls?" Compute the probabilities of these events.

Binomial and Multinomial Distributions

33. A die is rolled 8 times. What is the probability we will get exactly two 3's?

34. Mary knows the answers to 20 of the 25 multiple choice questions on the Psychology 101 exam, but she has skipped several of the lectures, she must take random guesses for the other five. Assuming each question has four answers, what is the probability she will get exactly 3 of the last 5 questions right?

35. In 1997, 10.8% of female smokers smoked cigars. In a sample of size 10 female smokers What is the probability that (a) exactly 2 of the women smoke cigars? (b) at most 1 smokes cigars.

36. A 1994 report revealed that 32.6% of U.S. births were to unmarried women. A parenting magazine selected 30 women who gave birth in 1994 at random. (a) What is the probability

that exactly 10 of the women were unmarried? (b) Using your calculator determine the probability that in the sample at most 10 are unmarried.

37. 20% of all students are left-handed. A class of size 20 meets in a room with 5 left-handed and 18 right-handed chairs. Use your calculator to find the probability that each student will have a chair to match their needs.

38. David claims to be able to distinguish brand B beer from brand H but Alice claims that he just guesses. They set up a taste test with 10 small glasses of beer. David wins if he gets 8 or more right. What is the probability he will win (a) if he is just guessing? (b) if he gets the right answer with probability 0.9?

39. The following situation comes up the game of Yahtzee. We have three rolls of five dice and want to get three sixes or more. On each turn we reroll any dice that are not 6's. What is the probability we succeed?

40. A baseball pitcher throws a strike with probability 0.5 and a ball with probability 0.5. He is facing a batter who never swings at a pitch. What is the probability that he strikes out, i.e., gets three strikes before four balls?

41. A student is taking a multiple-choice test in which each question has four possible answers. She knows the answers to 5 of the questions, can narrow the choices down to 2 in 3 cases, and does not know anything about 2 of the questions. What is the probability she will correctly answer (a) 10, (b) 9, (c) 8, (d) 7, (e) 6, (f) 5 questions?

42. Consider a die with 1 painted on three sides, 2 painted on two sides, and 3 painted on one side. If we roll this die ten times what is the probability we get five 1's, three 2's and two 3's?

43. A game of bowling consists of 10 "frames" in which the bowler gets to roll the ball twice. The outcome of a frame may be a strike (all ten pins on the first try), a spare (all ten pins knocked down in two tries), or an "open frame." Bowler Bob gets a strike with probability 0.3, a spare with probability 0.6, and an open frame with probability 0.1. His recent game featured 2 strikes, 5 spares and 3 open frames. Find the probability of this outcome.

44. A baseball player gets a hit with probability 0.3, a walk with probability 0.1, and an out with probability 0.6. If he bats 5 times during a game and we assume that the outcomes are independent, what is the probability he will get (a) 2 hits, 1 walk, and 2 outs? (b) 2 hits and 3 outs?

45. A baseball player is said to "hit for the cycle" if he has a single, a double, a triple, and a home run all in one game. Suppose these four types of hits have probabilities 1/6, 1/20, 1/120, and 1/24. What is the probability of hitting for the cycle if he gets to bat (a) four times, (b) five times? (c) Using $P(\cup_i A_i) \le \sum_i P(A_i)$ shows that the answer to (b) is at most 5 times the answer to (a). What is the ratio of the two answers?

Poisson Approximation

46. Compare the Poisson approximation with the exact binomial probabilities when (a) $n = 10$, $p = 0.1$, (b) $n = 20$, $p = 0.05$, (c) $n = 40$, $p = 0.025$.

47. An NC State player hits 3 point shots with probability 1/6. In a game against UNC he attempted 12 and made 2. (a) What is the exact Binomial probability of this event? (b) What is the Poisson approximation?

48. Suppose that the probability of a defect in a foot of magnetic tape is 0.002. Use the Poisson approximation to compute the probability that a 1500 foot roll will have no defects.

49. Suppose 1% of a certain brand of Christmas lights is defective. Use the Poisson approximation to compute the probability that in a box of 25 there will be at most one defective bulb.

50. In February 2000, 2.8% of Colorado's labor force was unemployed. Calculate the probability that in a group of 50 workers exactly one is unemployed.

51. An insurance company insures 3000 people, each of whom has a 1/1000 chance of an accident in one year. Use the Poisson approximation to compute the probability there will be at most 2 accidents.

52. Suppose that 1% of people in the population are over 6 feet 3 inches tall. What is the chance that in a group of 200 people picked at random from the population at least four people will be over 6 feet 3 inches tall.

53. Use the Poisson approximation to compute the probability that you will roll at least one double 6 in 24 trials. How does this compare with the exact answer?

54. The probability of a three of a kind in poker is approximately 1/50. Use the Poisson approximation to compute the probability you will get at least one three of a kind if you play 20 hands of poker.

55. In one of the New York state lottery games, a number is chosen at random between 0 and 999. Suppose you play this game 250 times. Use the Poisson approximation to estimate the probability you will never win and compare this with the exact answer.

56. If you bet \$1 on number 13 at roulette (or on any other number) then you win \$35 if that number comes up, an event of probability 1/38, and you lose your dollar otherwise. Suppose you play 70 times. Use the Poisson approximation estimate the probability that (a) you have won 0 times and lost \$70, and (b) you have won 1 time and lost \$34. (c) if you win 2 times you have won \$2. Combine the results of (a) and (b) to conclude that the probability you will have won more money than you have lost is larger than 1/2.

57. In a particular Powerball drawing 210,850,582 tickets were sold. The chance of winning the lottery is 1 in 80,000,000. Use the Poisson approximation to estimate the probability that there is exactly one winner.

58. Books from a certain publisher contain an average of 1 misprint per page. What is the probability that on at least one page in a 300 page book there are five misprints?

59. Calls to a toll-free hotline service are made randomly at rate two per minute. The service has five operators, none of whom is currently busy. Use the Poisson distribution to estimate the probability that in the next minute there are < 5 calls.

60. In an average year in Ithaca there are 8 fires. Last year there were 12 fires. How likely is it to have 12 or more fires just by chance?

61. An airline company sells 160 tickets for a plane with 150 seats, knowing that the probability a passenger will not show up for the flight is 0.1. Use the Poisson approximation to compute the probability they will have enough seats for all the passengers who show up.

62. There are 100 senators. (a) Use the Poisson approximation to estimate the probability exactly one senator has the same birthday as you. (b) What is the exact probability? Calculate each answers to five (or more) decimal places so you can compare the approximation with the exact answer.

Expected value

63. In a popular gambling game, three dice are rolled. For a \$1 bet you win \$1 for each six that appears (plus your dollar back). If no six appears you lose your dollar. What is your expected value?

64. In blackjack the dealer gets two cards, one of which you can see and the other you cannot. When the dealer's visible card is an Ace, she offers you the chance to take out insurance. You can bet \$1 that the her other card counts 10 (K, Q, J, 10) and hence she has a "blackjack." The bet has $2:1$ odds, i.e., you win \$2 (plus your dollar back) if you bet a dollar. What is the expected of this bet? Since blackjack is now played dealing from a "shoe" that contains $4-6$ decks pretend that the new card is drawn from another deck of 52 cards.

65. Sic Bo is an ancient Chinese dice game played with 3 dice. One of the possibilities for betting in the game is to bet "big." For this bet you win if the total X is 11, 12, 13, 14, 15, 16, or 17, except when there are three 4's or three 5's. On a \$1 bet on big you win \$1 plus your dollar back if it happens. What is your expected value?

66. Five people play a game of "odd man out" to determine who will pay for the pizza they ordered. Each flips a coin. If only one person gets Heads (or Tails) while the other four get Tails (or Heads) then he is the odd man and has to pay. Otherwise they flip again. What is the expected number of tosses needed to determine who will pay?

67. A man and wife decide that they will keep having children until they have one of each sex. Ignoring the possibility of twins and supposing that each trial is independent and results in a boy or girl with probability 1/2, what is the expected value of the number of children they will have?

68. Every Sunday two children, Al and Betty take turns trying to fly their kites. Al is successful with probability 1/5, Betty with probability 1/3. What is the expected number of Sundays until at least one child is successful.

69. An unreliable clothes dryer dries your clothes and takes 20 minutes with probability 0.6, buzzes for 6 minutes and does nothing with probability 0.4. If we assume that successive trials are independent and that we patiently keep putting our money in to try to get it to work, what is the expected time we need to get our clothes dry?

70. Suppose we draw 13 cards out of a deck of 52. What is the expected value of the number of aces we get?

71. Suppose we pick 3 students at random from a class with 10 boys and 15 girls. Let X be the number of boys selected and Y be the number of girls selected. Find $E(X-Y)$.

72. Twelve ducks fly overhead. Each of 6 hunters picks one duck at random to aim at and kills it with probability 0.6. (a) What is the mean number of ducks that are killed? (b) What is the expected number of hunters who hit the duck they aim at?

73. 10 people get on an elevator on the first floor of a seven story building. Each gets off at one of the six higher floors chosen at random. What is the expected value of $S =$ thenumber of stops the elevator makes?

74. Suppose Noah started with 50 pairs of animals on the ark and 10 of them died. Suppose that fate chose the 10 animals at random. What is the expected number of complete pairs that are left? What is the probability all the deceased animals were from different species?

75. Suppose we draw 5 cards out of a deck of 52. (a) What is the expected number of different suits in our hand? For example, if we draw $K\spadesuit$ $3\spadesuit$ $10\heartsuit$ $8\heartsuit$ $6\clubsuit$ there are three different suits in our hand. (b) Find the probability of $A =$ we get at least one card from each of the four suits.

Probabilities of Unions

76. Six high school teams play each other in the Southern Tier division. Each team plays all of the other teams once. What is the probability some team has a perfect $5-0$ season?

77. Suppose you draw seven cards out of a deck of 52. What is the probability you will have (a) exactly five cards of one suit? (b) at least five cards of one suit?

78. In a certain city 60% of the people subscribe to newspaper A, 50% to B, 40% to C, 30% to A and B, 20% to B and C, and 10% to A and C, but no one subscribes to all three. What percentage subscribe to (a) at least one newspaper? (b) exactly one newspaper?, (c) only to A? (d) only to B?

79. Santa Claus has 45 drums, 50 cars, and 55 baseball bats in his sled. 15 boys will get a drum and a car, 20 a drum and a bat, 25 a bat and a car, and 5 will get three presents. (a) How many boys will receive presents? (b) How many boys will get just a drum?

80. Ten people call an electrician and ask him to come to their houses on randomly chosen days of the work week (Monday through Friday). What is the probability of $A =$ "he has at least one day with no jobs"?

81. We pick a number between 0 and 999, then a computer picks one at random from that range. Use (2.9) to compute the probability at least two of our digits will match the computer's number. (Note: We include any leading zeros, so 017 and 057 have two matching digits.)

82. You pick 13 cards out of a deck of 52. What is the probability that you will not get a card from every suit?

83. You pick 13 cards out of a deck of 52. Let $A =$ "you have exactly six cards in at least one suit" and $B =$ "you have exactly six spades." The first Bonferroni inequality says that $P(A) \le 4P(B)$. Compute $P(A)$ and $P(A)/P(B)$.

84. Use the first three Bonferroni inequalities to compute an upper and a lower bound on the probability of $B =$ there is at least one day of the year on which exactly three people were born, for a group of 60 people.

85. In the previous problem we ignored the possibility of $C =$ four people born on the same day. Use the first two Bonferroni inequalities to get upper and lower bounds on the probability of C for a group of size 60.

86. Suppose we roll two dice 6 times. Use the first three Bonferroni inequalities to compute bounds on the probability of $A =$ "we get at least one double 6." Compare the bounds with the exact answer $1 - (35/36)^6$.

87. Suppose we try 20 times for an event with probability 0.01. Use the first three Bonferroni inequalities to compute bounds on the probability of at least one success.

Chapter 3

Random Variables and their Distributions

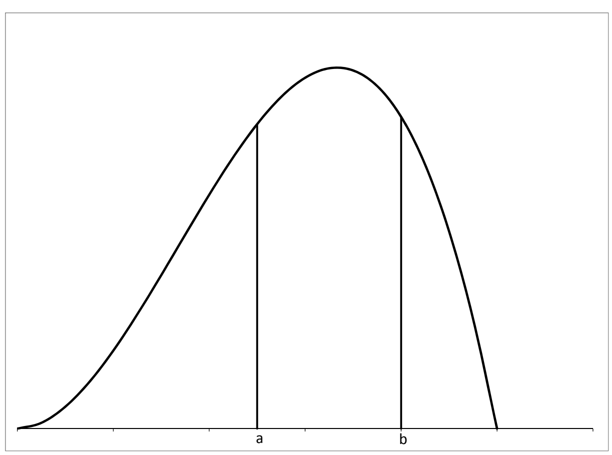
3.1 Distributions

We have already studied the case of discrete random variables, which take on a finite or countably infinite number of values. In this case the distribution is described by giving $P(X = x)$ for all the x for which this is positive.

In many situations random variables can take any value on the real line or in a subset of the real line, such as the nonnegative numbers. For concrete examples, consider the height or weight of a Duke student chosen at random, or the time it takes a person to drive from Durham to Asheville. A random variable X is said to have a **continuous distribution** with **density function** f if for all $a \le b$ we have

$$P(a \le X \le b) = \int_a^b f(x)\,dx \tag{3.1}$$

Geometrically, $P(a \le X \le b)$ is the area under the curve f between a and b.



For the purposes of understanding and remembering formulas, it is useful to think of $f(x)$ as $P(X = x)$ even though the last event has probability zero. To explain the last remark and to prove $P(X = x) = 0$, note that taking $a = x$ and $b = x + \Delta x$ in (3.1) we have

$$P(x \le X \le x + \Delta x) = \int_x^{x+\Delta x} f(y)\, dy \approx f(x)\Delta x$$

when Δx is small. Letting $\Delta x \to 0$, we see that $P(X = x) = 0$, but $f(x)$ tells us how likely it is for X to be near x. That is,

$$\frac{P(x \le X \le x + \Delta x)}{\Delta x} \approx f(x)$$

In order for $P(a \le X \le b)$ to be nonnegative for all a and b and for $P(-\infty < X < \infty) = 1$ we must have

$$f(x) \ge 0 \quad \text{and} \quad \int f(x)\, dx = 1 \tag{3.2}$$

Here, and in what follows, if the limits of integration are not specified, the integration is over all values of x from $-\infty$ to ∞. Any function f that satisfies (3.2) is said to be a **density function**.

We are now going to give four important examples of density functions. For simplicity, we will give the values of $f(x)$ where it is positive, and omit the phrase "0 otherwise."

Example 3.1. Uniform distribution. Given $a < b$ we define

$$f(x) = \frac{1}{b-a} \qquad a \le x \le b$$

The idea here is that we are picking a value "at random" from (a, b). That is, values outside the interval are impossible, and all those inside have the same probability (density).

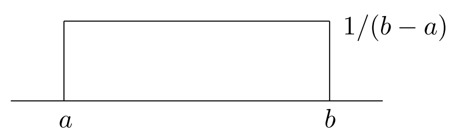


Figure 3.1: Uniform on (a, b) density function.

The area of the rectangle is $(b - a) \cdot 1/(b - a) = 1$ so this is a probability density. The most important special case occurs when $a = 0$ and $b = 1$. Random numbers generated by a computer are typically uniformly distributed on $(0, 1)$. Another case that comes up in applications is $a = -1/2$ and $b = 1/2$. If we take a measurement and round it off to the nearest integer then it is reasonable to assume that the "round-off error" is uniformly distributed on $(-1/2, 1/2)$.

Example 3.2. Exponential distribution. Given $\lambda > 0$ we define

$$f(x) = \lambda e^{-\lambda x} \qquad x \ge 0$$

To check that this is a density function, we note that

$$\int_0^\infty \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^\infty = 0 - (-1) = 1$$

Exponentially distributed random variables often come up as waiting times between events; for example, the arrival times of customers at a bank or ice cream shop. Sometimes we will indicate that X has an exponential distribution with parameter λ by writing $X =$ exponential(λ). Figure 3.2 shows the case $\lambda = 3/2$.

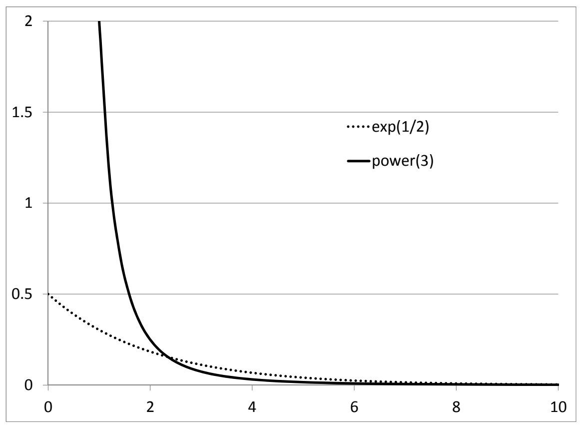


Figure 3.2: Exponential(1/2) and Power(3) density functions.

Example 3.3. Power laws.

$$f(x) = (\rho - 1)x^{-\rho} \qquad x \geq 1$$

Here $\rho > 1$ is a parameter that governs how fast the probabilities go to 0 at ∞.

To check that this is a density function, we note that

$$\int_1^\infty (\rho - 1)x^{-\rho}\,dx = -x^{-(\rho-1)}\Big|_1^\infty = 0 - (-1) = 1$$

These distributions are often used in situations where $P(X > x)$ does not go to 0 very fast as $x \to \infty$. For example, the Italian economist Pareto used them to describe the distribution of family incomes. Figure 3.2 gives a picture of the case $\rho = 3$.

Example 3.4. Normal distribution The standard normal distribution has density function.

$$n(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

Since there is no closed form expression for the anti-derivative of $e^{-x^2/2}$ it takes some ingenuity to check that

Theorem 3.1. *$n(x) = (2\pi)^{-1/2} e^{-x^2/2}$ is a probability density function.*

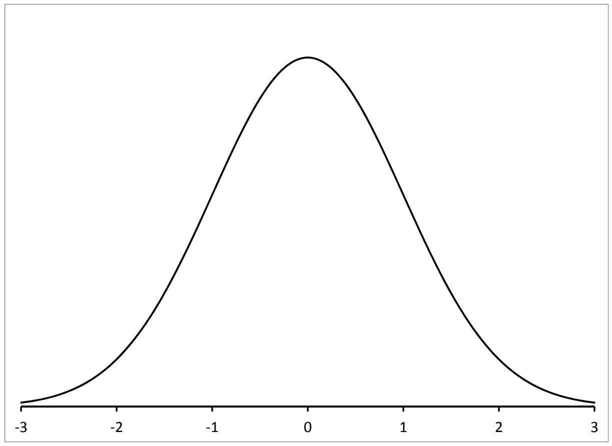


Figure 3.3: The standard normal density function.

Proof. Let $I = \int e^{-x^2/2}\,dx$. To show that $\int n(x)\,dx = 1$ we want to show that $I = \sqrt{2\pi}$.

$$I^2 = \int e^{-x^2/2}\,dx \int e^{-y^2/2}\,dy = \iint e^{-(x^2+y^2)/2}\,dx\,dy$$

Changing to polar coordinates the integral becomes

$$\int_0^\infty \int_0^{2\pi} e^{-r^2/2} r\,d\theta\,dr = 2\pi \int_0^\infty re^{-r^2/2}\,dr = 2\pi\left(-r^{-r^2/2}\right)\Big|_0^\infty$$

so $I^2 = 2\pi$ and we have proved the desired result. □

Example 3.5. Order statistics. Let $U_1, U_2, \ldots U_n$ be independent and uniform on $(0,1)$. Let $V_1 < V_2 \ldots V_n$ be the result of putting he values $U_1, U_2, \ldots U_n$ in increasing order. The V_k are called order statistics. Ignoring the possibility of two of the U_i landing in $(x-\epsilon, x+\epsilon)$,

$$P(V_k \in (x-\epsilon, x+\epsilon)) = n \cdot 2\epsilon \cdot C_{n-1,k-1}(x-\epsilon)^{k-1}(1-x-\epsilon)^{n-k}$$

In words there are n ways to pick the index of the U_i to land in $(x-\epsilon, x+\epsilon)$. There are $C_{n-1,k}$ to pick the values to be $< x-\epsilon$ and a probability $(x-\epsilon)^k(1-x-\epsilon)^{n-1-k}$ that they all land where we want them to. Dividing my 2ϵ and letting $\epsilon \to 0$ we see that the beta(n,k) density is

$$n \cdot C_{n-1,k-1}x^k(1-x)^{n-1-k} = \frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k}$$

This a special case of the beta distribution which has density

$$B_{\alpha,\beta}x^{\alpha-1}(1-x)^{\beta-1}$$

where $\alpha, \beta > 0$ are real numbers, and $B_{\alpha,\beta}$ is a constant to make the inegral one. The order statistics corrsepond to $\alpha = k$, $\beta = n-k+1$.

3.2 Distribution Functions

Any random variable (discrete or contnuous) has a **distribution function** defined by $F(x) = P(X \le x)$. However, these are the most useful in the continuous case.

3.2.1 For continuous random variables

If X has a density function $f(x)$ then

$$F(x) = P(-\infty < X \le x) = \int_{-\infty}^{x} f(y)\, dy$$

That is, F is an antiderivative of f, and a special one $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to \infty$.

One of the reasons for computing the distribution function is explained by the next formula. If $a < b$ then $\{X \le b\} = \{X \le a\} \cup \{a < X \le b\}$ with the two sets on the right-hand side disjoint so

$$P(X \le b) = P(X \le a) + P(a < X \le b)$$

or, rearranging,

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a) \tag{3.3}$$

The last formula is valid for any random variable. When X has density function f, it says that

$$\int_a^b f(x)\, dx = F(b) - F(a)$$

i.e., the integral can be evaluated by taking the difference of any antiderivative at the two endpoints.

To see what distribution functions look like, and to explain the use of (3.3), we return to our examples.

Example 3.6. Uniform distribution. $f(x) = 1/(b-a)$ for $a \le x \le b$.

$$F(x) = \begin{cases} 0 & x \le a \\ (x-a)/(b-a) & a \le x \le b \\ 1 & x \ge b \end{cases}$$

To check this, note that $P(a < X < b) = 1$ so $P(X \le x) = 1$ when $x \ge b$ and $P(X \le x) = 0$ when $x \le a$. For $a \le x \le b$ we compute

$$P(X \le x) = \int_{-\infty}^{x} f(y)\, dy = \int_a^x \frac{1}{b-a}\, dy = \frac{x-a}{b-a}$$

In the most important special case $a = 0$, $b = 1$ we have $F(x) = x$ for $0 \le x \le 1$.

Example 3.7. Exponential distribution. $f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$.

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

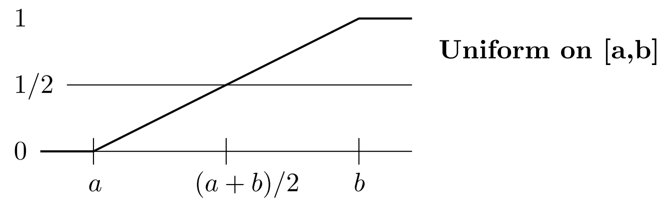


Figure 3.4: Distribution function for the unifomr on $[a, b]$.

The first line of the answer is easy to see. Since $P(X > 0) = 1$ we have $P(X \le x) = 0$ for $x \le 0$. For $x \ge 0$ we compute

$$\begin{aligned} P(X \le x) &= \int_{-\infty}^{x} f(y)\,dy = \int_{0}^{x} \lambda e^{-\lambda y}\,dy \\ &= -e^{-\lambda y}\Big|_0^x = -e^{-\lambda x} - (-1) \end{aligned}$$

Figure 3.5 shows the case $\lambda = 1/2$.

Suppose X has an exponential distribution with parameter λ. If $t \ge 0$ then $P(X > t) = 1 - P(X \le t) = e^{-\lambda t}$, so if $s \ge 0$ then

$$P(T > t + s | T > t) = \frac{P(T > t + s)}{P(T > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(T > s)$$

This is the **lack of memory property** of the exponential distribution. Given that you have been waiting t units of time, the probability you must wait an additional s units of time is the same as if you had not been waiting at all.

The following scaling property is useful.

$$\text{If } X = \text{exponential}(1) \text{ then } X/\lambda = \text{exponential}(\lambda). \qquad (3.4)$$

To prove this note that $P(X/\lambda > t) = P(X > \lambda t) e^{-\lambda t}$.

Example 3.8. Power laws. $f(x) = (\rho - 1)x^{-\rho}$ for $x \ge 1$ where $\rho > 1$.

$$F(x) = \begin{cases} 0 & x \le 1 \\ 1 - x^{-(\rho-1)} & x \ge 1 \end{cases}$$

The first line of the answer is easy to see. Since $P(X > 1) = 1$, we have $P(X \le x) = 0$ for $x \le 1$. For $x \ge 1$ we compute

$$\begin{aligned} P(X \le x) &= \int_{-\infty}^{x} f(y)\,dy = \int_{1}^{x} (\rho - 1)y^{-\rho}\,dy \\ &= -y^{-(\rho-1)}\Big|_1^x = 1 - x^{-(\rho-1)} \end{aligned}$$

Figure 3.5 gives a picture of the case $\rho = 3$. To illustrate the use of (3.3) we note that if $\rho = 3$ then

$$P(2 < X \le 4) = (1 - 4^{-2}) - (1 - 2^{-2}) = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$$

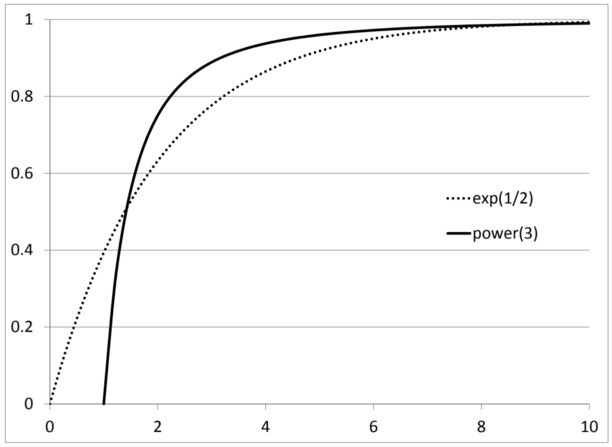


Figure 3.5: Distribution functions for exponential(1/2) and power(3).

Example 3.9. Normal distribution.

$$n(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

does not have a closed form antiderivative, so to find $P(X \le x)$ we have to use the table in the back of the book. We will talk more about this in Chapter 3.

3.2.2 For discrete random variables

Distribution functions are somewhat messier in the discrete case.

Example 3.10. Binomial(3,1/2). Flip three coins and let X be the number of heads that we see. The probability function is given by

x	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

In this case the distribution function can be written as

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \le x < 1 \\ 1/2 & 1 \le x < 2 \\ 7/8 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

or drawn as
To check this, note for example that for $1 \le x < 2$, $P(X \le x) = P(X \in \{0, 1\}) = 1/8 + 3/8$. The reader should note that F is discontinuous at each possible value of X and the height of the jump there is $P(X = x)$. The little black dots in the figure are there to indicate that at 0 the value is 1/8, at 1 it is 1/2, etc.

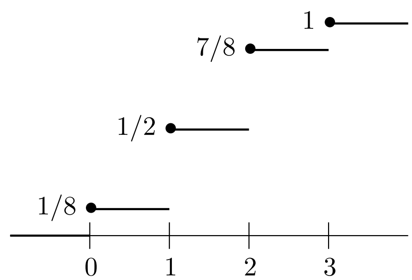


Figure 3.6: Distribution function for binomial(3,1/2)

Theorem 3.2. *All distribution functions have the following properties*

(i) If $x_1 < x_2$ *then* $F(x_1) \leq F(x_2)$ *i.e.,* F *is nondecreasing.*

(ii) $\lim_{x\to-\infty} F(x) = 0$

(iii) $\lim_{x\to\infty} F(x) = 1$

(iv) $\lim_{y\downarrow x} F(y) = F(x)$*, i.e.,* F *is continuous from the right.*

(v) $\lim_{y\uparrow x} F(y) = P(X < x)$

(vi) $\lim_{y\downarrow x} F(y) - \lim_{y\uparrow x} F(y) = P(X = x)$*,*
i.e., the jump in F *at* x *is equal to* $P(X = x)$*.*

Proof. To prove (i) we note that $\{X \leq x_1\} \subset \{X \leq x_2\}$, so (1.2) implies $F(x_1) = P(X \leq x_1) \leq P(X \leq x_2) = F(x_2)$.

For (ii), we note that $\{X \leq x\} \downarrow \emptyset$ as $x \downarrow -\infty$ (here $\downarrow$ is short for "decreases and converges to"), so (1.5) implies that $P(X \leq x) \downarrow P(\emptyset) = 0$.

The argument for (iii) is similar: $\{X \leq x\} \uparrow \Omega$ as $x \uparrow \infty$ (here $\uparrow$ is short for "increases and converges to"), so (1.5) implies that $P(X \leq x) \uparrow P(\Omega) = 1$.

For (iv), we note that if $y \downarrow x$ then $\{X \leq y\} \downarrow \{X \leq x\}$, so (1.5) implies that $P(X \leq y) \downarrow P(X \leq x)$.

The argument for (v) is similar. If $y \uparrow x$ then $\{X \leq y\} \uparrow \{X < x\}$ since $\{X = x\} \not\subset \{X \leq y\}$ when $y < x$. Using (1.5) now, (v) follows.

Subtracting (v) from (iv) gives (vi). □

The next example gives a more interesting discrete distribution function.

Example 3.11. New York Yankees 2004 salaries. Salaries are in units of M, millions of dollars per year and, for convenience, have been truncated at the thousands place.

A. Rodriguez	21.726	D. Jeter	18.6
M. Mussina	16	K. Brown	15.714
J. Giambi	12.428	B. Williams	12.357
G. Sheffield	12.029	M. Rivera	10.89
J. Posada	9	J. Vazquez	9
J. Contreras	9	J. Olerud	7.7
H. Matsui	7	S. Karsay	6
E. Loazia	4	T. Gordon	3.5
P. Quantrill	3	K. Lofton	2.985
J. Lieber	2.7	T. Lee	2
G. White	1.925	F. Heredia	1.8
R. Sierra	1	M. Cairo	.9
J. Falherty	.775	T. Clark	.75
E. Wilson	.7	O. Hernandez	.5
D. Osborne	.45	C.J. Nitowski	.35
J. DePaula	.302	B. Crosby	.301

Figure 3.7 shows the distribution function of Yankees salaries. The total team salary is 183.355 M. Dividing by 32 players gives a mean of 6.149 M dollars.

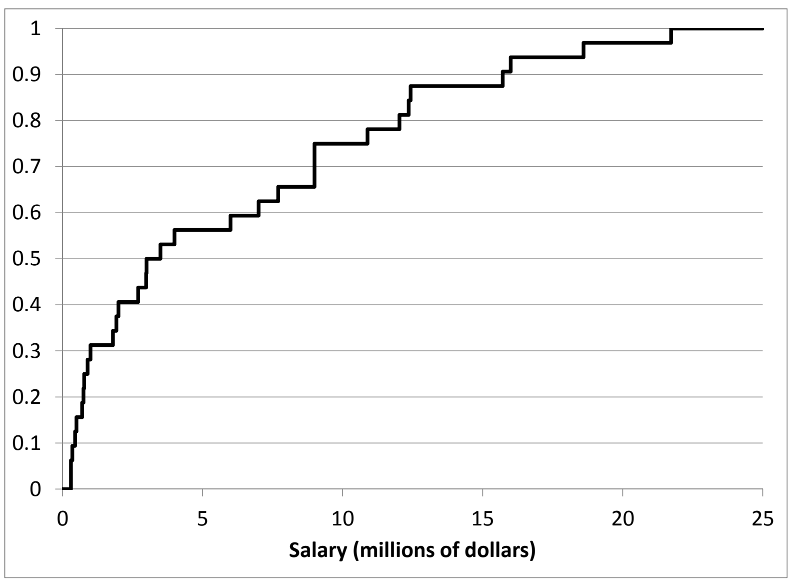


Figure 3.7: Distribution function for Yankees salaries

3.3 Means and medians

For a discrete random variable the mean is

$$EX = \sum_x xP(X = x)$$

To extend this to the continuous case we replace $P(X = x)$ by the density function

$$EX = \int x f_X(x)\,dx$$

Example 3.12. Uniform distribution. Suppose X has density function $f(x) = 1/(b-a)$ for $a \le x \le b$. Then

$$EX = \int_a^b \frac{x}{b-a}\,dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

Notice that $(a+b)/2$ is the midpoint of the interval and hence is the natural choice for the average value of X.

Before tackling the next example, we recall the **integration by parts formula**:

$$\int_a^b g(x)h'(x)\,dx = g(x)h(x)|_a^b - \int_a^b g'(x)h(x)\,dx \tag{3.5}$$

Example 3.13. Exponential distribution. Suppose X has density function

$$f(x) = \lambda e^{-\lambda x} \qquad \text{for } x \ge 0.$$

Integrating by parts with $g(x) = x$, $h'(x) = \lambda e^{-\lambda x}$, so $g'(x) = 1$ and $h(x) = -e^{-\lambda x}$.

$$EX = \int_0^\infty x\,\lambda e^{-\lambda x}\,dx = -xe^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx = 0 + 1/\lambda$$

To check the last step note that $-xe^{-\lambda x} = 0$ when $x = 0$ and when $x \to \infty$ $-xe^{-\lambda x} \to 0$ since $e^{-\lambda x} \to 0$ much faster than x grows. The evaluation of the integral follows from the definition of the exponential desnity, which implies $\int_0^\infty \lambda e^{-\lambda x}\,dx = 1$.

Example 3.14. Power laws. Let $\rho > 1$ and $f(x) = (\rho-1)x^{-\rho}$ for $x \ge 1$.

$$EX = \int_1^\infty x(\rho-1)x^{-\rho}\,dx = \left.\frac{\rho-1}{2-\rho}x^{2-\rho}\right|_1^\infty = \frac{\rho-1}{\rho-2}$$

if $\rho > 2$. If $1 < \rho \le 2$, $EX = \infty$.

Example 3.15. Normal distribution. To make this example more interesting we will look at the general normal with parameters μ and σ^2 or normal(μ, σ^2) introduced in (3.7):

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

By symmetry

$$\int (2\pi\sigma^2)^{-1/2} x e^{-(x-\mu)^2/2\sigma^2}\,dx = \mu$$

Medians. Intuitively, the median $m_{1/2}$ is the point where 50% of the values are $\ge m_{1/2}$ and 50% are $\le m_{1/2}$. It is a notion of the typical value of the distribution is not sensitive to large values. For example in the Yankee salaries data Alex Rodriguez's salary contributes 678,937 to the mean. Similar things would happen if we looked at salaries across the US or the sizes of cities in a state.

When the distribution function is nice the median is where it takes the value 1/2. The next table computes m for our four examples and compares with the mean

Exponential distribution	$m_{1/2} = (\log 2)/\lambda$.	$\mu = 1/\lambda$
Uniform distribution	$m_{1/2} = (a+b)/2 = \mu$	
Power laws	$m_{1/2} = 2^{1/(\rho-1)}$	$\mu = (\rho-1)/(\rho-2)$
Normal	$m_{1/2} = \mu$	

In the case of the exponential the mean and median differ by a constant factor. The mean and the median do not look very similar for the the power law. When $\rho < 2$ the mean is infinite but the median is finite. When $\rho = 4$, $\mu = 3/2$ while the median $m_{1/2} = 1.26$.

In the discrete case two unusual things can happen. In the case of the binomial$(3, 1/2)$ distribution $F(x) = 1/2$ when $1 \le x < 2$. A different type of problem exists for the uniform distribution on $\{1, 2, 3\}$. There is no place where $F(x) = 1/2$.

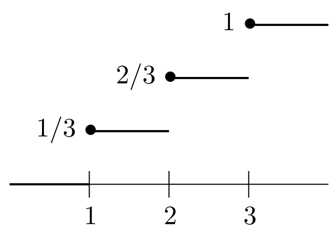


Figure 3.8: Distribution function for uniform on $\{1, 2, 3\}$

The next definition takes care of bad cases. Suppose X has distribution function F.

$$m_{1/2} \text{ is a median for } F \text{ if } P(X \le m_{1/2}) \ge 1/2 \text{ and } P(X \ge m_{1/2}) \ge 1/2$$

Using this definition

- 2 is the median for the uniform on $\{1, 2, 3\}$ since $P(X \le 2) = 2/3$ and $P(X \ge 2) = 2/3$.
- For the binomial$(3, 1/2)$ distribution any value $m \in [1, 2]$ is a median since if $m \in (1, 2)$ $P(X \le m) = P(X \ge m) = 1/2$ while $P(X \le 1) = 1/2$ and $P(X \ge 1) = 7/8$ and $P(X \le 2) = 7/8$ and $P(X \ge 2) = 1/2$
- For the Yankees salaries any number between 3 and 3.5 is a median. In contrast the mean is 6.49 million.

Generalizing the median, the α-percentile is the place m_α where $P(X \le m_\alpha) = \alpha$. Percentiles are commonly used when reporting values of standardized tests. For example, "my score was in the 72nd percentile on the SAT math test. These points also come up in solving optimization problems

Example 3.16. An enterprising young man sits in the lobby of a big office building selling newspapers. He buys papers for 15 cents and sells them for 25, but he cannot return any unsold papers. Suppose that each day X people by papers. Compute the difference between buying m papers and buy in $m - 1$ in order to determine the optimal number to buy.

If he buys m (or $m-1$) papers his profit is

$$-15m + \sum_{k=0}^{m-1} 25kP(X=k) + 25mP(X \geq m)$$
$$-15(m-1) + \sum_{k=0}^{m-1} 25kP(X=k) + 25(m-1)P(X \geq m-1)$$

In the first line the first term is his cost. The second term reflects the fact that if the demand is $k <, m$ he sells k papers but if $k \geq m$, he sells m. The second line is similar but we have split the two cases in a different way to make it easier to subtract the two to get $-15 + 25P(X \geq m)$. When this is positive buying m papers is better than $m-1$, so the optimum is the $\max\{m : P(X \geq m) \geq 0.6\}$. This is roughly the 0.4-percentile but the distribution is discrete that isn't quite the right answer.

3.4 Functions of Random Variables

In this section we answer the question "If X has density function f and $Y = r(X)$, then what is the density function for Y?" Before proving a general result, we will consider an example:

Example 3.17. Suppose X has an exponential distribution with parameter λ. What is the distribution of $Y = X^2$?

To solve this problem we will use the distribution function. First we recall from Example 3.7 that $P(X \leq x) = 1 - e^{-\lambda x}$ so if $y \geq 0$ then

$$P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = 1 - e^{-\lambda y^{1/2}}$$

Differentiating and using the chain rule $(g(h(x))' = g'(h(x))h'(x)$, we see that the density function of Y is given by

$$f_Y(y) = \frac{d}{dy} P(Y \leq y) = e^{-\lambda y^{1/2}} \frac{\lambda y^{-1/2}}{2} \qquad \text{for } y \geq 0$$

and 0 otherwise.

Intuitively, this is the same as changing variables in an integral. If we set $x = y^{1/2}$ and conseqeuntly $dx = (1/2)y^{-1/2}\,dy$ then

$$\int_a^b f(x)\,dx = \int_{a^2}^{b^2} f(y^{1/2})(1/2)y^{-1/2}\,dy$$

Generalizing from the last example, we get

Theorem 3.3. *Suppose X has density f and $P(a < X < b) = 1$. Let $Y = r(X)$. Suppose $r : (a,b) \to (\alpha, \beta)$ is continuous and strictly increasing, and let $s : (\alpha, \beta) \to (a, b)$ be the inverse of r, defined by $r(s(y)) = y$. Then Y has density*

$$f_Y(y) = f_X(s(y))s'(y) \qquad \textit{for } y \in (\alpha, \beta) \tag{3.6}$$

Before proving this, let's see how it applies to the last example. There X has density $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ so we can take $a = 0$ and $b = \infty$. The function $r(x) = x^2$ is indeed continuous and strictly increasing on $(0, \infty)$. To find the inverse function we set $y = x^2$ and solve to get $x = y^{1/2}$ so $s(y) = y^{1/2}$. Differentiating, we have $s'(y) = y^{-1/2}/2$ and plugging into the formula, we have

$$g(y) = \lambda e^{-\lambda y^{1/2}} \cdot y^{-1/2}/2 \qquad \text{for } y > 0$$

Proof of 3.3. If $y \in (\alpha, \beta)$ then

$$P(Y \leq y) = P(r(X) \leq y) = P(X \leq s(y))$$

since r is increasing and s is its inverse. Writing $F(x)$ for $P(X \leq x)$ and differentiating with respect to y now gives

$$g(y) = \frac{d}{dy} P(Y \leq y) = \frac{d}{dy} F(s(y)) = F'(s(y))s'(y) = f(s(y))s'(y)$$

by the chain rule. □

Remark. The same proof works if r is decreasing but in this case

$$P(Y \leq y) = P(r(X) \leq y) = P(X \geq s(y))$$

so the derivative is $f(s(y))(-s'(y))$. To combine the two cases into one we can write

$$f_Y(y) = f(s(y))|s'(y)|$$

Example 3.18. Linear transformations. Suppose X has density $f(x)$ and $Y = a + bX$ where $b > 0$. $r(x) = a + bx$ has inverse $s(y) = (y - a)/b$ so Y has density $f((y - a)/b)/b$. To illustrate the use of this formula suppose X has the the standard normal distribution with density

$$(2\pi)^{-1/2} e^{-x^2/2}$$

and let $Y = \mu + \sigma X$ where $\sigma > 0$. Our formula implies that Y has density

$$(2\pi\sigma^2)^{-1/2} e^{-(y-\mu)^2/2\sigma^2} \tag{3.7}$$

We call this the normal(μ, σ^2) density. We know that μ is the mean. In the next chapter, we will see that σ^2 is the variance.

Example 3.19. Suppose X has the standard normal distribution. Find the distribution of $Y = e^X$. It is called the **lognormal distribution** since $\log(Y)$ is normal.

$r(x) = e^x$ is increasing and maps $(-\infty, \infty)$ onto $(0, \infty)$. The inverse function is $s(y) = \log x$ so using (3.6) the answer is

$$(2\pi)^{-1/2} e^{-(\log x)^2/2} \cdot 1/x.$$

This is called the lognormal distribution because $\log(Y)$ is normally distributed

Example 3.20. Gumbel distribution. Suppose X is exponential(1). Find the desnity function of $Y = \log X$.

Here, $r(x) = \log x$ is increasing on $(0, \infty)$. Its inverse function is $s(y) = e^y$, so plugging into (3.6), the density function of Y is

$$f_Y(y) = f_X(s(y))s'(y) = \exp(-e^y)e^y \quad y \in (-\infty, \infty)$$

The distribution function is $G(y) = 1 - \exp(-e^y)$. For obvious reasons this is sometimes called the **double exponential distribution**.

Our finaL topic is creating random variables with a specified distribution by taking a function of the unifrom distribution. Most programming languages have built in functions for generating uniformly distributed random variables, so this allows us to create any distribution. To begin we define the inverse of a distribuiton function. Since $F(x)$ make take the value y at more than one point we let

$$F^{-1}(y) = \min\{x : F(x) \geq y\}$$

In most cases $F(x)$ will be strictly increasing when $0 < F(x) < 1$, so if $0 < y < 1$ then $F^{-1}(y)$ is the unique solution to $F(x) = y$

Theorem 3.4. *Suppose the distribution function is conitnuous ands let U have uniform distribution on (0,1). Then $Y = F^{-1}(U)$ has distribution function F.*

Proof. A distribution function is nondecreasing, i.e., if $x \leq y$ then $F(x) \leq F(y)$. Taking $x = F^{-1}U$ and using the fact that $F(F^{-1}(U) = U$

$$P(F^{-1}(U) \leq y) = P(F(F^{-1}(U) \leq F(y)) = P(U \leq F(x)) = F(x)$$

since $P(U \leq y) = y$ for $y \in (0, 1)$. □

Example 3.21. For a concrete example, suppose we want to construct an exponential distribution with parameter λ. Setting $1 - e^{-\lambda x} = y$ and solving gives $-\ln(1 - y)/\lambda = x$. So if U is uniform on $(0, 1)$ then $-\ln(1 - U)/\lambda$ has the desired exponential distribution. Of course since $1 - U$ is uniform on $(0, 1)$ we could also use

$$r(U) = -\ln(U)/\lambda$$

To check that we have achieved our goal we note that r is decreasing, and the inverse function $s(y) = e^{-\lambda y}$. Since the density function of U is $f(x) = 1$ for $0 < x < 1$. Using (3.6) the denity of $Y = r(U)$ is

$$g(y) = 1 \cdot |s'(y)| = \lambda e^{-\lambda y}.$$

Theorem 3.4 has a consequence thatis useful to reduce the general case to the uniform in nonparametric statistics.

Theorem 3.5. *If X has a continuous distribution function $F(x)$ then $F(X)$ is uniform on* $(0, 1)$.

Proof. By Theorem 3.4 $F^{-1}(U)$ has the same distribution as X so

$$P(F(X) \leq y) = P(F(F^{-1}(U)) \leq y) = P(U \leq y)$$

where we have used continuity to conclude $F(F^{-1}(y)) = y$. □

3.5 Joint distributions

In the next chapter we will study sums of random variables. However, before we can discuss that we need to introduce joint distributions.

3.5.1 Discrete distributions

We begin with some definitions. The **joint distribution** of two discrete random variables is given by

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

To recover the **marginal distribution** of X, we sum over the possible values of Y

$$P(X = x) = \sum_y P(X = x, Y = y) \tag{3.8}$$

Similarly, to recover the marginal distribution of Y, we sum over the possible values of X

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

Example 3.22. An urn has 6 red, 5 blue, and 4 green balls. We draw three balls out. Let X = the number of red balls drawn and Y = the number of blue balls. If $x, y \geq 0$ and $x + y \leq 3$ then $3 - (x + y)$ green balls were drawn so

$$P(X = x, Y = y) = \frac{C_{6,x} \cdot C_{5,y} \cdot C_{4,3-(x+y)}}{C_{15,3}}$$

The denominator is $C_{15,3} = (15 \cdot 14 \cdot 13)/6 = 455$. Using

$$\begin{array}{lll} C_{6,3} = 20 & C_{6,2} = 15 & C_{6,1} = 6 \\ C_{5,3} = 10 & C_{5,2} = 10 & C_{5,1} = 5 \\ C_{4,3} = 4 & C_{4,2} = 6 & C_{4,1} = 4 \end{array}$$

we can compute the joint distribution

Y	$X = 0$	1	2	3	
3	10/455				10/455
2	40/455	60/455			100/455
1	30/455	120/455	75/455	0	225/455
0	4/455	36/455	60/455	20/455	120/445
	84/455	216/455	135/455	20/455	

The marginal distribution of X given in the last row is obtained by adding up the columns. The marginal distribution of Y given in the last column is obtained by adding up the rows. To check the margnal distributions note that

$$P(X = 1) = \frac{C_{6,1} C_{9,3}}{C_{15,3}} = \frac{6 \cdot 9 \cdot 8/2}{455} = \frac{216}{455}$$

$$P(Y = 2) = \frac{C_{5,2} C_{10,1}}{C_{15,3}} = \frac{10 \cdot 5 \cdot 4/2}{455} = \frac{100}{455}$$

Example 3.23. Let X be the result of rolling a four sided die. Flip X coins and let Y be the number of heads that we observe. To begin to fill in the table, we note that

$$P(X=4,Y=0)=(1/4)\cdot 1/16, \qquad P(X=4,Y=1)=(1/4)\cdot 4/16$$

. By considering the different possibilities we see that the joint distribution is given by

X	$Y=0$	1	2	3	4	
4	1/64	4/64	6/64	4/64	1/64	1/4
3	1/32	3/32	3/32	1/32	0	1/4
2	1/16	2/16	1/16	0	0	1/4
1	1/8	1/8	0	0	0	1/4
	15/64	26/64	16/64	6/64	1/64	

It is comforting to note that in the marginal distribution of X, all four values have probability 1/4.

Two discrete random variables X an Y are said to be **independent** if for all x and y

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

In the two previous examples the variables are not independent since there are x and y so that $P(X=x,Y=y)=0$ but $P(X=x)>0$ and $P(Y=y)=0$. Most of the examples of independent random variables are boring: flip two coins, roll two dice, etc.

3.5.2 Continuous distributions

In the continuous case we give the joint distribution is described by the **joint density function** $f_{X,Y}(x,y)$ which has the interpretation that

$$\iint_A f_{X,Y}(x,y)\,dy\,dx = P((X,Y)\in A) \tag{3.9}$$

As in the one variable case, we must have $f_{X,Y}(x,y)\geq 0$ and

$$\iint f_{X,Y}(x,y)\,dy\,dx = 1.$$

Marginal densities

To find the marginal densities in the continuous case, we replace the sums in (3.8) by integrals

$$f_X(x)=\int f_{X,Y}(x,y)\,dy \qquad f_Y(y)=\int f_{X,Y}(x,y)\,dx \tag{3.10}$$

Example 3.24. *Suppose*$f(x,y)=(6/5)(x^2+y)$ when $0<x,y<1$.

The marginal densities are

$$f_X(x)=(6/5)\int_0^1 (x^2+y)\,dy = (6/5)(x^2y+y^2/2)\big|_{y=0}^1 = 6x^2/5+3/5$$
$$f_Y(y)=(6/5)\int_0^1 (x^2+y)\,dx = (6/5)(x^3/3+xy)\big|_{x=0}^1 = 2/5+6x/5$$

To check that 6/5's is the right constant to make the integral 1, we integrate the marginal densities

$$\int_0^1 6x^2/5 + 3/5\,dx = 2x^3/5 + 3/5\Big|_0^1 = 1$$
$$\int_0^1 2/5 + 6y/5\,dy = 2/5 + 3y^2/5\Big|_0^1 = 1$$

For another example of computing marginal densities.

Example 3.25. Suppose $f_{X,Y}(x,y) = 20(y-x)^3$ for $0 < x < y < 1$. Find the marginal density of X and the marginal density of Y

$$f_X(x) = \int_x^1 20(y-x)^3\,dy = 5(y-x)^4\Big|_x^1 = 5(1-x)^4$$
$$f_Y(y) = \int_0^y 20(y-x)^3\,dx = -5(y-x)^4\Big|_0^y = 5y^4$$

Independence

As in the discrete case, X and Y are independent if the joint distribution is the product of the marginal distributions

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

In order for the random variables in Example 3.24 to be independent we would need

$$(6/5)(x^2+y) = \left(\frac{2x^3}{5} + \frac{3}{4}\right)\left(\frac{2}{5} + \frac{3y^2}{5}\right)$$

This is false when $x = 0$ and $y = 0$ In the Example 3.25 the variables are not independent since $f_{X,Y}(0.6, 0.4) = 0$ while $f_X(0.6) > 0$ and $f_Y(0.4) > 0$.

For an example of random variables that are independent

Example 3.26. $f(x,y) = 6x^2y$ when $0 < x, y < 1$.

The marginal densities are

$$f_X(x) = \int_0^1 6x^2y\,dy = 3x^2y^2\Big|_0^1 = 3x^2$$
$$f_Y(y) = \int_0^1 6x^2y\,dx = 6(x^3y/3)\Big|_0^1 = 2y$$

$f_{X,Y}(x,y) = f_X(x)f_Y(y)$ so these random variables are independent.

The next lemma makes checking independence easier

Theorem 3.6. *If $f_{X,Y}(x,y)$ vanish outrside $[a,b]\times[c,d]$ and can be written as $h(x)k(y)$ then X and Y are independent.*

Proof. The marginal distributions are

$$f_X(x) = \int_c^d h(x)k(y)\,dy = h(x)\int_c^d k(y)\,dy$$
$$f_Y(y) = \int_a^b h(x)k(y)\,dx = k(y)\int_a^b h(x)\,dy$$

For these random variables to be independent we need

$$h(x)k(y) = \left(h(x)\int_c^d k(y)\,dy\right)\left(k(y)\int_a^b h(x)\,dy\right)$$

but this holds since

$$1 = \iint h(x)k(y)\,dy\,dx = \left(\int_a^b h(x)\,dy\right)\left(\int_c^d k(y)\,dy\right)$$

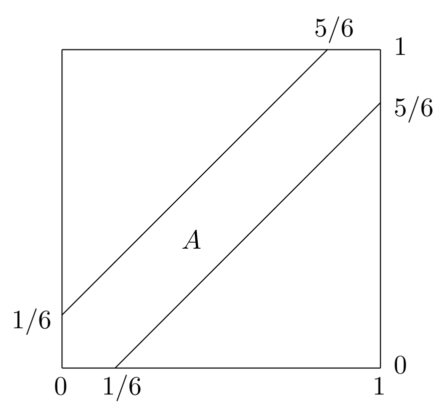
□

In words if we can write $f_{X,Y}(x,y)$ as a product $h(x)k(y)$ then there is a constant so that $f_X(x) = ch(x)$ and $f_Y(y) = k(y)/c$ and hence $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Computing probabilities

Example 3.27. Suppose two people agree to meet for coffee and each will arrive at independent randomly chosen times between 11:00 and 12:00. What is the probability of $A =$ the first one will not have to wait more than 10 minutes for the other to arrive?

Let X_1 and X_2 be uniform on $(0,1)$ so the joint desnity is $f_{X,Y}(x,y) = 1$ when $0 < x, y < 1$. The answer is $P(|X_1 - X_2| < 1/6)$. To compute the probability it is useful to draw a picture



The central diagonal strip is the event of interest. The complement of this strip consists of two right triangles with each side 5/6, so the answer is $1 - (5/6)^2 = 11/16$.

Example 3.28. Suppose X = exponential(λ) and Y = exponential(μ) are independent. Find $P(X < Y)$.

The joint density is $\lambda e^{-\lambda x} \cdot \mu e^{-\mu y}$. We want to integrate this over the region $0 < x < y < \infty$.

$$\begin{aligned}
&\int_0^\infty dy\, \mu e^{-\mu y} \int_0^y dx\, \lambda e^{-\lambda x} = \int_0^\infty dy \mu e^{-\mu y}(1 - e^{-\lambda y}0 \\
&= \int_0^\infty dy \mu e^{-\mu y} - \mu e^{-(\mu+\lambda)y} = 1 - \frac{\mu}{\mu+\lambda} = \frac{\lambda}{\mu+\lambda}
\end{aligned}$$

where we have used the observation that $\mu e^{-\mu y}$ and $(\mu+\lambda)e^{-(\mu+\lambda)y}$ integrate to 1 since they are exponential density functions.

Conditional densities

In the discrete case the joint distribution is $P(X = x, Y = y)$ while the marginal distributions for X and Y are given by

$$\begin{aligned}
P(X = x) &= \sum_y P(X = x, Y = y) \\
P(Y = y) &= \sum_x P(X = x, Y = y)
\end{aligned}$$

The purpose of this section is to define the conditional distributions

$$\begin{aligned}
P(X = x|Y = y) &= P(X = x, Y = y)/P(Y = y) \\
P(Y = y|X = x) &= P(X = x, Y = y)/P(X = x)
\end{aligned}$$

Example 3.29. In Example 3.23 we let X be the result of rolling a four sided die and let Y be the number of heads that we observe when we flip X coins. By considering the different possibilities we see that the joint distribution is given by

X	$Y = 0$	1	2	3	4	
4	1/64	4/64	6/64	4/64	1/64	1/4
3	1/32	3/32	3/32	1/32	0	1/4
2	1/16	2/16	1/16	0	0	1/4
1	1/8	1/8	0	0	0	1/4
	15/64	26/64	16/64	6/64	1/64	

The condition distribution of X given that Y has

$$P(X = x|Y = 0) = \frac{P(X = y, Y = 0)}{P(Y = 0)}$$

In words we divide the column by the sum of the probabilities to make it sum to 1.

x	0	1	2	3
$P(X = x\|Y = 0)$	8/15	4/15	2/15	1/15

The conditional distribution of Y given that $X = 3$ is iis obtained by dividing the $X = 3$ row by 1/4

y	0	1	2	3
$P(Y = y\|X = 3)$	1/8	3/8	3/8	1/8

which is the binomial distribution we used to compute the joint distribution.

The situation for continuous random variables is similar but with sums replaced by integrals.

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} \qquad f_{X|Y=y}(y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Example 3.30. In Example 3.24 $f(x,y) = (6/5)(x^2 + y)$ when $0 < x, y < 1$. We calculated that The marginal densities are

$$f_X(x) = 6x^2/5 + 3/5 \qquad f_Y(y) = 2/5 + 6y/5$$

The conditional distributions are

$$f_{Y|X=x}(y) = \frac{(6/5)(x^2 + y)}{6x^2/5 + 3/5} \qquad f_{X|Y=y}(y) = \frac{(6/5)(x^2 + y)}{6/5 + y}$$

3.6 Exercises

Discrete Distributions

1. Suppose we roll two dice and let X and Y be the two numbers that appear. Find the distribution of $|X - Y|$.

2. How many children should a family plan to have so that the probability of having at least one child of each sex is at least 0.95?

3. In a carton of six eggs two are bad. Find the distribution of the number of eggs N that we use before pick one that is bad.

4. We have an urn with m red balls and n blue balls. Let N be the number of balls that we have to draw until we get a red one. Find $P(N = k)$ for $1 \le m \le n + 1$.

5. Suppose we roll two four sided dice with the numbers $1, 2, 3, 4$ on them,and the outcomes are R_1 and R_2. Find the joint distribution of $X = \min\{R_1, R_2\}$ and $Y = R_1 + R_2$ and then compute the marginal distributions of X and Y.

6. Suppose we roll a six sided die repeatedly. Let N_i be the number of the roll on which we see i for the first time. (a) Find the joint distribution of N_1 and N_6. (b) Find the marginal distributions of N_1. (c) The conditional distribution of N_6 given $N_1 = i$

7. Show that if $X_1 = \text{geomteric}(p_1)$ and $X_2 = \text{geomteric}(p_2)$ are independent then $Y = \min\{X_1, X_2\}$ is geoemetric(q) where $q = 1 - (1 - p_1)(1 - p_2)$.

Density and distribution functions

8. Suppose X has density function $f(x) = c(1 - |x|)$ when $-1 < x < 1$.(a) What value of c makes this a density function? (b) Find the distribution function.

9. $F(x) = 3x^2 - 2x^3$ for $0 < x < 1$ (with $F(x) = 0$ if $x \le 0$ and $F(x) = 1$ if $x \ge 1$) defines a distribution function. (a) Find the corresponding density function. (b) Find $P(1/2 < X < 1)$

10. Let $F(x) = e^{-1/x}$ for $x \geq 0$, $F(x) = 0$ for $x \leq 0$. Is F a distribution function? If so, find its density function.

11. Let $F(x) = 3x - 2x^2$ for $0 \leq x \leq 1$, $F(x) = 0$ for $x \leq 0$, and $F(x) = 1$ for $x \geq 1$. Is F a distribution function? If so, find its density function.

12. Consider $f(x) = c(1 - x^2)$ for $-1 < x < 1$, 0 otherwise. (a) What value of c should we take to make f a density function? (b) Find the distribution function. (c) Find $P(1/2 < X < 1)$. (d) Find EX

13. Suppose X has density function $f(x) = 4x^3$ for $0 < x < 1$, 0 otherwise. Find (a) the distribution function, (b) $P(X < 1/2)$, (c) $P(1/3 < X < 2/3)$, (d) EX.

14. Suppose X has density function $x^{-1/2}/2$ for $0 < x < 1$, 0 otherwise. Find (a) the distribution function, (b) $P(X > 3/4)$, (c) $P(1/9 < X < 1/4)$, (d) EX.

15. Suppose X has the **Rayleigh** density function $xe^{-x^2/2}$ for $x \geq 0$. (a) Find the distribution function. (b) $P(X > \sqrt{2}$, (c) EX.

16. Suppose X has density $f(x) = 4x(1 - x^2)$ for $0 < x < 1$. (a) Find the distribution function. (b) Find the median. (b) Find the mean.

17. Suppose x has density function $f(x) = 12x^2(1 - x)$ for $0 < x < 1$. (a) Find the distribution function of X, (a) $P(0 < X < 1/2)$, and (c) EX.

18. Order statitistics were considered in Example 3.5. There it was shown that if $U_1, U_2, \ldots U_n$ are independent uniform on $(0, 1)$ and $V_1 < V_2 < \ldots V_n$ are the U_i put in increasing order then V_k has density

$$f_{k,n}(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}$$

Compute the mean EV_k.

Medians, etc.

19. Suppose $P(X = x) = x/21$ for $x = 1, 2, 3, 4, 5, 6$. Find all the medians of this distribution.

20. Suppose X has a Poisson distribution with $\lambda = \ln 2$. Find all the medians of X.

21. Suppose X has a geometric distribution with success probability 1/4, i.e., $P(X = k) = (3/4)^{k-1}(1/4)$. Find all the medians of X.

22. Suppose X has density function $f(x) = x/2$ for $0 < x < 2$, 0 otherwise. Find (a) the distribution function, (b) the median, (c) the mean.

23. Suppose X has density $f(x) = 4x(1 - x^2)$ for $0 < x < 1$. (a) Find the distribution function. (b) Find the median. (b) Find the mean.

24. Suppose X has the Rayleigh distribution $xe^{-x^2/2}$ for $0 < x < \infty$. Find (a) the distribution function, (b) the median, (c) the mean.

25. The owner of a dress store buys dresses for \$50 and sells them for \$125. Any unused dresses are sold to a second-hand store at the end of the season for \$25. Suppose to simplify things that the demand for dresses is X, a random varialbe with density function $f(x)$. How many dresses should the store owner buy to maximize profit.

Functions of Random Variables

26. Suppose X has density x^{-2} for $x \geq 1$. Find the density function for $Y = X^2$.

27. Suppose X is uniform on $(0,1)$. Find the density function for $Y = X^\alpha$ when $\alpha > 0$.

28. Suppose $X = \text{exponential}(\lambda)$. Show that $Y = X^{1/\alpha}$. has the **Weibull distribution** $\lambda\alpha y^{\alpha-1}e^{-\lambda y^\alpha}$.

29. Suppose that X has the standard normal distribution. Find the distribution of $Y = 1/X$.

30. Suppose that X has the standard normal distribution. Find the distribution of X^2.

31. Suppose X has the **Rayleigh** density function $xe^{-x^2/2}$ for $x \geq 0$. . Find the distribution of X^2.

32. Suppose $X_1, \ldots, X_n$ are independent and have distribution function $F(x)$. Find the distribution functions of (a) $Y = \max\{X_1, \ldots, X_n\}$ and (b) $Z = \min\{X_1, \ldots, X_n\}$

33. Suppose $X_1, \ldots, X_n$ are independent exponential(λ). Show that

$$\min\{X_1, \ldots, X_n\} = \text{exponential}(n\lambda)$$

34. Let $U_1 \ldots, U_n$ be independent and uniform on (0,1) and let $V = \max\{U_1, \ldots U_n\}$. Find the distribution function and density function for V.

Creating random variables with a given distribubtion

35. Suppose X has the powerlaw density function $(\rho - 1)x^{-\rho}$ for $x \geq 1$ where $\rho > 1$. Find a function g so that if U is uniform on $(0,1)$, $g(U)$ has this density function.

36. Suppose X has distribution function $x/(x+1)$ for $x \geq 0$, 0 otherwise. Find a function g so that if U is uniform on $[0,1]$ then $g(U)$ has this distribution function.

37. The Rayliegh distribution has density $f(x) = xe^{-x^2/2}$. Find a function g so that if U is uniform on $(0,1)$, $g(U)$ has this density function.

38. Suppose U uniform on $(0,1)$. Find a function g so that $g(U)$ has the Weibull distribution defined in problem 3.6.

Joint distributions: calculating probabilities

39. Suppose (X,Y) has joint density $6xy^2$ for $0 < x, y < 1$. Find $P(X + Y < 1)$.

40. Suppose X and Y are independent and uniform on $(0,1)$. Find $P(XY > z)$ and differentiate to find the density function.

41. Suppose X and Y are independent and uniform on $(0,1)$. Find $P(Y/X \leq z)$ and differentiate to find the density function.

42. Suppose X and Y have joint density $(y - x)$ for $0 < x < y < 1$. Compute $P(Y > 2X)$.

43. Suppose X and Y are independent standard normals. Show that $Z = \sqrt{X^2 + Y^2}$ has the Rayleigh distribution $re^{-r^2/2}$.

44. Let (X,Y) be uniform on $0 < x < y < 1$. Cut the unit interval at X and Y to make three pieces of size X, $Y - X$ and $1 - Y$. What is the probability of $\Delta =$ the three pieces can be used to make a triangle? Note: in order to be able to make a triangle the largest piece has to be shorter than the sum of the other two.

45. Let A, B, and C be independent and uniform on $(-1, 1)$. What is the probability of $E =$ the polynomial $Ax^2 + Bx + C = 0$ has no real roots. Note: there are no real roots if $B^2 - 4AC < 0$.

46. Suppose U and V are independent uniform on $(0, 1)$. Find $E(U - V)^2$.

47. Suppose X and Y are independent and exponential(λ). Find $E|X - Y|$.

Independence, marginal and conditional distributions

48. Suppose that the joint density of X and Y is given by $f(x, y) = xe^{-(x+y)}$ for $x, y > 0$ (a) Are X and Y independent? (b) Find the marginal densities of X and Y.

49. Suppose that the joint desnity of X and Y is given by $f(x, y) = 24xy$ when $x, y > 0$ and $x + y < 1$ (a) Are X and Y independent? (b) Find the marginal densities of X. (c) Thje conditional distribution of Y given X.

50. Suppose that the joint desnity of X and Y is given by $f(x, y) = xe^{-x(1+y)}$ when $x, y > 0$.(a) Are X and Y independent? (b) Find the marginal densities of X and Y.

51. Suppose X and Y have joint density $f_{X,Y}(x, y) = (6/5)(x^2 + xy/2)$ when $0 < x, y < 1$. (a) Find the marginal densities of X and Y. (b) Find $P(X > Y)$

52. Suppose that the joint density of X and Y is given by $f_{X,Y}(x, y) = (y - x)e^{-y}$ for $0 < x < y < \infty$. Find the marginal distributions of X and Y

53. Suppose (X, Y) has joint density $f(x, y) = 8xy$ for $0 < x < y < 1$. (a) Find the marginal distribution of Y. (b) Find the conditional distribution of Y given X.

54. Suppose $f_{X,Y}(x, y) = (x + y)$ for $0 < x, y < 1$. (a) Find the marginal distribution of X. (b) Find the conditional distribution of Y given X.

55. Suppose $f_{X,Y}(x, y) = (3/2)(xy^2 + y)$ for $0 < x, y < 1$. (a) Find the marginal distribution of X. (b) Find the conditional distribution of Y given X.

56. Suppose $U_1, \ldots U_n$ are independent and uniform on $(0, 1)$. Let $X = \min_i U_i$ and $Y = \max_i U_i$. Imitate the proof in Example 3.5 to show X and Y have joint density

$$f_{X,Y}(x, y) = n(n - 1)(y - x)^{n-2}$$

57. Taking $k = 3$ in the previous problem we see that if U_1, U_2, U_3 are independent and uniform on $(0, 1)$. The joint density of $X = \min\{U_1, U_2, U_3\}$ and $Y = \max\{U_1, U_2, U_3\}$ is given by $f_{X,Y}(x, y) = 6(y - x)$or $0 < x < y < 1$. (a) Find the marginal distributions of Y. (b) The conditional distribution of X given Y.

58. Let t_1 and t_2 be independent exponential with parameter 1. The joint density of $X = t_1$ and $Y = t_1 + t_2$ is e^{-y} for $0 < x < y$. (You do not have to prove this.) (a) Find the marginal densities of X and Y.(b) The conditonal density of X given Y and of Y given X.

59. Let X and Y have joint distribution $f_{X,Y}(x, y) = 1/x$ for $0 < y < x < 1$. (a) Find the marginal densities of X and Y.(b) The conditonal density of X given Y and of Y given X.

Chapter 4

Limit Theorems

4.1 Moments, Variance

In this section we will be interested in the expected values of various functions of random variables. The most important of these are the variance and the standard deviation which give an idea about how spread out the distribution is. The first basic fact we need in order to do computations is the following, which we state without proof.

Theorem 4.1. *If X has a discrete distribution then*

$$Er(X) = \sum_x r(x)P(X = x)$$

In the continuous case

$$Er(X) = \int r(x)f(x)\,dx$$

If $r(x) = x^k$, $E(X^k)$ is the k**th moment of** X. When $k = 1$ this is the first moment or mean of X. If $EX^2 < \infty$ then the **variance** of X is defined to be

$$\text{var}\,(X) = E(X - EX)^2$$

To illustrate this concept, we will consider some examples. But first, we need a formula that enables us to compute the variance more easily.

$$\text{var}\,(X) = EX^2 - (EX)^2 \tag{4.1}$$

Proof. Letting $\mu = EX$ to make the computations easier to see, we have

$$\text{var}\,(X) = E(X - \mu)^2 = E(X^2 - 2\mu X + \mu^2) = EX^2 - 2\mu EX + \mu^2$$

by (1.15) and the facts that $E(-2\mu X) = -2\mu EX$, $E(\mu^2) = \mu^2$. Substituting $\mu = EX$ now gives the result. □

The reader should note that EX^2 means the expected value of X^2 and in the proof $E(X-\mu)^2$ means the expected value of $(X-\mu)^2$. When we want the square of the expected value we will write $(EX)^2$. This convention is designed to cut down on parentheses.

The variance measures how spread-out the distribution of X is. To begin to explain this statement, we will show that

$$\text{var}\,(X+a) = \text{var}\,(X) \qquad \text{var}\,(bX) = b^2\,\text{var}\,(X) \tag{4.2}$$

In words, the variance is not changed by adding a constant to X, but multiplying X by a multiplies the variance by a^2.

Proof. If $Y = X + a$ then the mean of Y, $\mu_Y = \mu_X + a$ by (1.15), so

$$\text{var}\,(X+a) = E\{(X+a) - (\mu_X + a)\}^2 = E\{X - \mu_X\}^2 = \text{var}\,(X)$$

If $Y = bX$ then $\mu_Y = b\mu_X$ by (1.15), so

$$\text{var}\,(bX) = E\{(bX - b\mu_X)^2\} = b^2 E(X-\mu_X)^2 = b^2\,\text{var}\,(X) \qquad \square$$

The scaling relationship (4.2) shows that if X is measured in feet then the variance is measured in feet^2. This motivates the definition of the **standard deviation** $\sigma(X) = \sqrt{\text{var}\,(X)}$, which is measured in the same units as X and has a nicer scaling property:

$$\sigma(bX) = |b|\sigma(X) \tag{4.3}$$

We get the absolute value here since $\sqrt{b^2} = |b|$.

4.1.1 Discrete random variables

Example 4.1. Roll one die and let X be the resulting number.

$$EX^2 = \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6} = 15.1666$$
$$\text{var}\,(X) = EX^2 - (EX)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} = 2.91666$$

and $\sigma(X) = \sqrt{\text{var}\,(X)} = 1.7078$. The standard deviation $\sigma(X)$ gives the size of the "typical deviation from the mean." To explain this, we note that the deviation from the mean

$$|X - \mu| = \begin{cases} 0.5 & \text{when } X = 3,4 \\ 1.5 & \text{when } X = 2,5 \\ 2.5 & \text{when } X = 1,6 \end{cases}$$

so $E|X-\mu| = 1.5$. The standard deviation $\sigma(X) = \sqrt{E|X-\mu|^2}$ is a slightly less intuitive way of averaging the deviations from the mean but, as we will see later, is one that has nicer properties.

Example 4.2. The **geometric(p) distribution** has variance $(1-p)/p^2$.

Suppose N = geometric(p). That is, $P(N = n) = (1-p)^{n-1}p$ for $n = 1, 2, \ldots$ and 0 otherwise. To compute the variance, we begin, as in Example 2.1 by recalling the sum of the geometric series

$$\sum_{n=0}^{\infty} x^n = (1-x)^{-1}$$

Differentiating this identity and noticing that the $n = 0$ term in the first derivative is 0 gives

$$\sum_{n=1}^{\infty} nx^{n-1} = (1-x)^{-2}$$

Differentiating again

$$\sum_{n=2}^{\infty} n(n-1)x^{n-2} = 2(1-x)^{-3}$$

Setting $x = 1 - p$ in the second formula gives

$$\sum_{n=1}^{\infty} n(n-1)(1-p)^{n-2} = 2p^{-3}$$

Multiplying both sides by $p(1-p)$, we have

$$E\{N(N-1)\} = \sum_{n=1}^{\infty} n(n-1)(1-p)^{n-1}p = \frac{2p(1-p)}{p^2}$$

From this it follows that

$$EN^2 = E\{N(N-1)\} + EN = \frac{2-2p}{p^2} + \frac{1}{p} = \frac{(2-p)}{p^2}$$
$$\text{var}\,(N) = EN^2 - (EN)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{(1-p)}{p^2}$$

Taking the square root we see that $\sigma(X) = \sqrt{1-p}/p$.

Example 4.3. The **binomial(n, p) distribution** has variance np.

As in the case of the geometric, our next step is to compute

$$E(N(N-1)) = \sum_{m=2}^{n} m(m-1)\frac{n!}{m!(n-m)!}p^m(1-p)^{n-m}$$

We have dropped the first two terms which contribute nothing to the sum so we can cancel to get

$$= n(n-1)p^2 \sum_{m=2}^{n} \frac{(n-2)!}{(m-2)!(n-m)!}p^{m-2}(1-p)^{n-m} = n(n-1)p^2$$

since the sum computes the total probability for the binomial$(n-2, p)$ distribution.

To finish up we note that

$$\text{var}\,(N) = EN^2 - (EN)^2 = E(N(N-1)) + EN - (EN)^2$$
$$= n(n-1)p^2 + np - n^2p^2 = n(p-p^2) = np(1-p)$$

which completes the proof. In Section 4.4 we will see a much simpler proof.

Example 4.4. The **Poisson(λ) distribution** has variance λ

As in the case of the binomial and geometric we begin by computing

$$E(X(X-1)) = \sum_{k=2}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2$$

From this it follows that

$$\begin{aligned} \text{var}\,(X) &= EX^2 - (EX)^2 = E(X(X-1)) + EX - (EX)^2 \\ &= \lambda^2 + \lambda - \lambda^2 \end{aligned}$$

which completes the proof.

4.1.2 Continuous random variables

Example 4.5. Uniform distribution. Suppose X has density function $f(x) = 1/(b-a)$ for $a \le x \le b$. We begin with the case $a = 0$, $b = 1$. The result in Example 3.12 implies $EX = 1/2$.

$$\begin{aligned} E(X^2) &= \int_0^1 x^2\,dx = \left.\frac{x^3}{3}\right|_0^1 = 1/3 \\ \text{var}\,(X) &= E(X^2) - (EX)^2 = (1/3) - (1/2)^2 = 1/12 \end{aligned}$$

To extend to the general case we recall from that (4.2) that if $Y = c + dX$ then

$$\text{var}\,(Y) = d^2\,\text{var}\,(X)$$

Taking $c = a$ and $d = b - a$ it follows that $\text{var}\,(Y) = (b-a)^2/12$

Example 4.6. Exponential distribution. To compute $E(X^2)$ we integrate by parts with $g(x) = x^2$, $h'(x) = \lambda e^{-\lambda x}$, so $g'(x) = 2x$ and $h(x) = -e^{-\lambda x}$.

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2\,\lambda e^{-\lambda x}\,dx = -x^2 e^{-\lambda x}\Big|_0^\infty + \int_0^\infty 2xe^{-\lambda x}\,dx \\ &= 0 + \frac{2}{\lambda}\int_0^\infty x\lambda e^{-\lambda x}\,dx = \frac{2}{\lambda^2} \end{aligned}$$

by the result for EX. Combining the last two results:

$$\text{var}\,(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

To see that the answer must have the form c/λ^2 let Y be exponential(1) and $Z = Y/\lambda$.

$$P(Z > t) = P(Y/\lambda > t) = P(Y > t\lambda) = e^{-\lambda t}$$

so Z is exponential(λ). Taking expected va;lue $EZ^2 = EY^2/\lambda^2$. From this we can also see that $EZ^m = c_m/\lambda^m$.

Example 4.7. Power laws. Let $\rho > 1$ and $f(x) = (\rho-1)x^{-\rho}$ for $x \ge 1$. $E(X^2) = \infty$ if $1 < \rho \le 3$. If $\rho > 3$ we have

$$E(X^2) = \int_1^\infty x^2(\rho-1)x^{-\rho}\,dx = \frac{\rho-1}{3-\rho}x^{3-\rho}\bigg|_1^\infty = \frac{\rho-1}{\rho-3}$$

$$\begin{aligned}\text{var}\,(X) &= \frac{\rho-1}{\rho-3} - \left(\frac{\rho-1}{\rho-2}\right)^2 \\ &= (\rho-1)\left[\frac{(\rho-2)^2-(\rho-1)(\rho-3)}{(\rho-2)^2(\rho-3)}\right] = \frac{\rho-1}{(\rho-2)^2(\rho-3)}\end{aligned}$$

If, for example, $\rho = 4$, using the result in Example 3.14 and our new formula

$$EX = \frac{\rho-1}{\rho-2} = \frac{3}{2} \qquad \text{var}\,(X) = \frac{3}{4\cdot 1} = \frac{3}{4}.$$

Example 4.8. Normal distribution.

$$n(x) = (2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/2\sigma^2}$$

By (4.2) the variance is not changed if we set $\mu = 0$. To find $\text{var}\,(X) = EX^2$ we integrate by parts, see (3.5), with

$$\begin{aligned} g(x) &= (2\pi\sigma^2)^{-1/2}x, \qquad & h'(x) &= xe^{-x^2/2\sigma^2} \\ g'(x) &= (2\pi\sigma^2)^{-1/2} & h(x) &= -\sigma^2 e^{-x^2/2\sigma^2}\end{aligned}$$

to conclude

$$\begin{aligned}\int (2\pi\sigma^2)^{-1/2}x^2e^{-x^2/2\sigma^2}\,dx &= -(2\pi\sigma^2)^{-1/2}x\cdot\sigma^2e^{-x^2/2}\bigg|_{-\infty}^{\infty} \\ &\quad + \sigma^2\int(2\pi\sigma^2)^{-1/2}e^{-x^2/2\sigma^2}\,dx = 0+\sigma^2\end{aligned}$$

where we have evaluated the integral by noticing that the general normal density integrates to 1.

4.2 Sums of Independent Random Variables

4.2.1 Discrete distributions

In order for $X + Y = z$, X must take on some value x, and $Y = z - x$. Summing over the possibilities and using independence.

$$\begin{aligned} P(X+Y=z) &= \sum_x P(X=x, Y=z-x) \\ &= \sum_x P(X=x)P(Y=z-x) \end{aligned} \tag{4.4}$$

Example 4.9. Sum of three dice. Let X be the sum of two dice and Y be uniform on $\{1, 2, \ldots 6\}$. As we have seen in Example 1.18 the distribution of X is given by

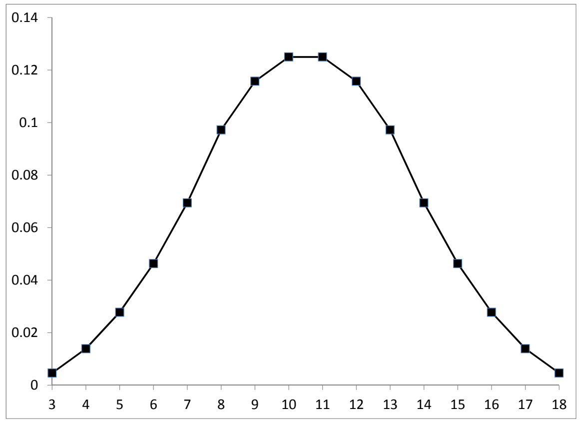


Figure 4.1: Distribution of the sum of three dice.

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Using this with the fact that $P(Y = y) = 1/6$ for $y = 1, 2, \ldots 6$ we have

$$\begin{aligned} P(X+Y=8) &= \sum_{x=2}^{7} P(X=x)P(Y=8-x) \\ &= \frac{1+2+3+4+5+6}{36} \cdot \frac{1}{6} = \frac{21}{216} \\ P(X+Y=9) &= \frac{2+3+4+5+6+5}{36} \cdot \frac{1}{6} = \frac{25}{216} \\ P(X+Y=10) &= \frac{3+4+5+6+5+4}{36} \cdot \frac{1}{6} = \frac{27}{216} \end{aligned}$$

With some patience you can compute the entire distribution:

3, 18	4, 17	5, 16	6, 15	7, 14	8, 13	9, 12	10, 11
$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$

The symmetry comes from the fact that if $i+j+k = m$ then $((7-i)+(7-j)+(7-k) = 21-m$, so m and $21-m$ must have the same probability.

Example 4.10. Geometric distribution. Let $X_1, \ldots X_n$ be independent geometric(p) and let $T_n = X_1 + \cdots + X_n$ be the time we have to wait for the nth success. Then if $m \ge n$

$$P(T_n = m) = C_{m-1,n-1} p^n (1-p)^{m-n}$$

This is called the **negative binomial distribution**. To prove the formula, consider the following outcome with $m = 10$, $n = 3$

$$F\,S\,F\,F\,F\,F\,S\,F\,F\,S$$

Any string with 3 S and 7 F has probability $p^3(1-p)^7$. For $T_3 = 10$ the last letter must be S. The other $n-1$ locations for S are chosen from the first $m-1$.

Example 4.11. If $X = \text{binomial}(n, p)$ and $Y = \text{binomial}(m, p)$ are independent then $X + Y = \text{binomial}(n + m, p)$.

Proof by thinking. If X is the number of successes in the first n trials and Y is the number of successes in the next m trials then $X + Y$ is the number of successes in the first $n + m$ trials. □

Proof by computation. Using (4.4)

$$\begin{aligned} P(X + Y = k) &= \sum_{j=0}^{k} P(X = j)P(Y = k - j) \\ &= \sum_{j=0}^{k} C_{n,j} p^j (1-p)^{n-j} \cdot C_{m,k-j} p^{k-j} (1 - p^{m-(k-j)}) \\ &= p^k (1 - p^{n+m-k}) \sum_{j=0}^{k} C_{n,j} C_{m,k-j} \\ &= C_{n+m,k} p^k (1-p)^{n+m-k} \end{aligned}$$

To see the last step note that if we pick k students from a class with n boys and m girls then for some j with $0 \le j \le k$ we will pick j boys (which can be done in $C_{n,j}$ ways) and $k - j$ girls (which can be done in $C_{m,k-j}$ ways). □

Since $\text{binomial}(n, \lambda/n)$ converges to $\text{Poisson}(\lambda)$ it should not be surprising that

Example 4.12. If $X = \text{Poisson}(\lambda)$ and $Y = \text{Poisson}(\mu)$ are independent then $X + Y = \text{Poisson}(\lambda + \mu)$.

Proof. Using (4.4)

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

by the binomial theorem (1.11) □

4.2.2 Continuous distributions

Replacing the sums in (4.4) by integrals,the density function for the sum of two independent continuous random variables is

$$f_{X+Y}(z) = \int f_X(x) f_Y(z - x)\, dx \tag{4.5}$$

Example 4.13. Uniform distribution. If X and Y are independent uniform on $(0,1)$ then

$$f_{X+Y}(z) = \begin{cases} z & 0 \le z \le 1 \\ 2-z & 1 \le z \le 2 \end{cases} \tag{4.6}$$

Since f_X and f_Y are 1 on $(0,1)$ and 0 otherwise, if $0 \le z \le 1$ then

$$f_{X+Y}(z) = \int_0^z 1\,dx = z$$

When $1 \le z \le 2$, we must have $X \ge z-1$ so

$$f_{X+Y}(z) = \int_{1-z}^1 1\,dx = 2-z$$

Example 4.14. Exponential distribution. We say X has a **gamma distribution** with parameters n and λ, and write $X = \text{gamma}(n,\lambda)$ if

$$f_X(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda z} \quad \text{for } x \ge 0 \tag{4.7}$$

Notice that when $n = 1$ this becomes the exponential density.

Theorem 4.2. *If $X = \text{gamma}(n,\lambda)$ and $Y = \text{exponential}(\lambda)$ are independent then $X+Y = \text{gamma}(n+1,\lambda)$*

It follows from this that

- If $X_1, \dots X_n$ are independent exponential(λ) then

$$X_1 + \cdots + X_n = \text{gamma}(n,\lambda).$$

- If $X = \text{gamma}(n,\lambda)$ and $Y = \text{gamma}(m,\lambda)$ are independent then

$$X + Y = \text{gamma}(n+m,\lambda).$$

Proof. It follows from (4.5) and (4.7) that

$$\begin{aligned} f_{X+Y}(z) &= \int_0^z \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} \cdot \lambda e^{-\lambda(z-x)}\,dx \\ &= \lambda^{n+1} e^{-z} \int_0^z \frac{x^{n-1}}{(n-1)!}\,dx = \frac{\lambda^{n+1} z^n}{n!} e^{-z} \end{aligned}$$

which proves the desired result. □

Example 4.15. *If $X = \text{normal}(\mu, a)$ and $Y = \text{normal}(\nu, b)$ are independent $X+Y = \text{normal}(\mu+\nu, a+b)$.*

Proof. By considering $X' = X - \mu$ and $Y' = Y - \nu$ and noting that $X+Y = X'+Y'+(\mu+\nu)$ it suffices to prove the result when $\mu = \nu = 0$ By (4.5) we need to evaluate

$$\int \frac{1}{\sqrt{2\pi a}} e^{-z^2/2a} \frac{1}{\sqrt{2\pi b}} e^{-(x-z)^2/2b} \, dz$$

The product of the two exponentials is

$$\exp\left(-\frac{z^2}{2a} - \frac{z^2}{2b} + \frac{2xz}{2b} - \frac{x^2}{2b}\right)$$

The quantity in the exponent is

$$\begin{aligned} \frac{-z^2(a+b) + 2axz - ax^2}{2ab} &= \frac{a+b}{2ab}\left(-z^2 + \frac{2a}{a+b}xz - \frac{a}{a+b}x^2\right) \\ &= \frac{a+b}{2ab}\left(-\left(z - \frac{ax}{a+b}x\right)^2 - \frac{ab}{(a+b)^2}x^2\right) \end{aligned}$$

since $(a^2+ab)/(a+b)^2 = a/(a+b)$. Working backwards through our steps we need to evaluate

$$\frac{1}{\sqrt{2\pi a}} \frac{1}{\sqrt{2\pi b}} e^{-x^2/2(a+b)} \cdot \int \exp\left(-\frac{a+b}{2ab}\left(z - \frac{ax}{a+b}\right)^2\right) dz$$

The integrand is a constant multiple of the density of the normal with mean $ax/(a+b)$ and variance $ab/(a+b)$ so the integral is $\sqrt{2\pi ab/(a+b)}$, and the expression we need to evaluate is

$$= \frac{1}{\sqrt{2\pi(a+b)}} e^{-x^2/2(a+b)}$$

which proves the desired result. □

4.3 Variance of sums of independent r.v.'s

Theorem 4.3. *If X and Y are independent then*

$$E(XY) = EX \cdot EY \tag{4.8}$$

Proof. In the discrete case,

$$\begin{aligned} EXY &= \sum_{x,y} xyP(X=x)P(Y=y) = \sum_y yP(Y=y) \sum_x xP(X=x) \\ &= \sum_y yP(Y=y) \cdot EX = EX \cdot EY \end{aligned}$$

In the continuous case

$$\begin{aligned} EXY &= \int\int xy f_X(x) f_Y(y) \, dy \, dx = \int y f_Y(y) \, dy \int x f_X(x) \, dx \\ &= \int y f_Y(y) \, dy \cdot EX = EX \cdot EY \end{aligned}$$ □

Too prepare for the next proof we need to show that $E|XY| < \infty$ if $EX^2 < \infty$ and $EY^2 < \infty$. The next result takes care of this detail.

Theorem 4.4. *The Cauchy-Schwarz inequality*

$$|EXY| \le E|XY| \le (EX^2)^{1/2}(EY^2)^{1/2} \tag{4.9}$$

Proof. The first inequality is trivial. To prove the second we begin by noting that $E(\theta|X| - |Y|)^2 \ge 0$, i.e.,

$$\theta^2 EX^2 - 2\theta E|XY| + EY^2 \ge 0$$

For this to hold for a quadratic $a\theta^2 + b\theta + c$ we must have

$$0 \ge b^2 - 4ac = 4(E|XY|)^2 - 4 \cdot EX^2 \cdot EY^2$$

which gives the desired result. □

Theorem 4.5. *If $X_1, \dots X_n$ are independent with $EX_i^2 < \infty$ then*

$$var(X_1 + \cdots + X_n) = var(X_1) + \cdots + var(X_n)$$

Proof. By (4.2) if $\mu_i = EX_i$ and $Y_i = X_i - \mu_i$ then

$$\begin{aligned} \text{var}\,(X_1 + \cdots + X_n) &= \text{var}\,(X_1 + \cdots + X_n - (\mu_1 + \cdots \mu_n)) \\ &= \text{var}\,(Y_1 + \cdots Y_n) \end{aligned}$$

so it suffices to prove the result when the random variables have mean 0. In this case the variance is the second moment, and expanding out the square gives

$$E(Y_1 + \cdots + Y_n)^2 = \sum_{i=1}^n EY_i^2 + 2 \sum_{1 \le i < j \le n} E(Y_iY_j) = \sum_{i=1}^n \text{var}\,(Y_i) \tag{4.10}$$

since $E(Y_iY_j) = EY_iEY_j = 0$. □

Theorem 4.5 allows us to easily compute

Example 4.16. The variance of the binoomial(n, p) distribution is np.

Proof. Let $X_i = 1$ if the ith trial is a success, and 0 otherwise. If $P(X_i = 1) = p$ then $S_n = X_1 + \cdot + X_n$, is the number of successes in n trials and has a binomial(n, p) distribution. Since the X_i are independent and have the same distribution Theorem 4.5 implies that $\text{var}\,(S_n) = n\,\text{var}\,(X_i)$.

To compute $\text{var}\,(X_i)$ we note that $EX_i = 1 \cdot p + 0 \cdot (1-p) = p$ and $EX_i^2 = 1 \cdot p + 0 \cdot (1-p) = p$ so $\text{var}\,(X_i) = EX_i^2 - (EX_i)^2 = p - p^2 = p(1-p)$ □

We will now give a formula for the variance of a sum that does not assume independence. Letting $\mu_i = EX_i$ and reintroducing $Y_i = X_i - \mu_i$ into (4.10) we have

$$\text{var}\,(X_1 + \cdots + X_n) = \sum_{i=1}^n \text{var}\,(X_i) + 2 \sum_{1 \le i < j \le n} E[(X_i - \mu_i)(X_j - \mu_j)] \tag{4.11}$$

Expanding out the product

$$\begin{aligned} E[(X_i - \mu_i)(X_j - \mu_j)] &= E[X_iX_j] - \mu_i EX_j - \mu_j EX_i + \mu_i\mu_j \\ &= E[X_iX_j] - \mu_i\mu_j \end{aligned}$$

Given two random variables X and Y with $EX^2, EY^2 < \infty$ the **covariance** is

$$\text{cov}\,(X,Y) = E(X-EX)(Y-EY) = E(XY) - EX \cdot EY$$

Adding a to X also adds a to its mean so $X - EX = (X+a) - E(X+a)$, and using the first definition it follows that

$$\text{cov}\,(X+a, Y+b) = \text{cov}\,(X,Y) \tag{4.12}$$

Thus the covariance, like the variance, is not effected by adding a constant. The similarity should not be surprising since

$$\text{cov}\,(X,X) = E(X^2) - (EX)^2 = \text{var}\,(X)$$

From (4.11) it follows immediately that we have

Theorem 4.6. *If $X_1, \dots X_n$ have $EX_i^2 < \infty$ then*

$$var\,(X_1 + \cdots + X_n) = \sum_{i=1}^n var\,(X_i) + 2 \sum_{1 \le i < j \le n} cov\,(X_i, X_j)$$

If $X_1, \dots X_n$ are uncorrelated, i.e., $cov\,(X_i, X_j) = 0$ for all $1 \le i < j \le n$ then

$$var\,(X_1 + \cdots + X_n) = \sum_{i=1}^n var\,(X_i)$$

If the X_i are independent then they are uncorrelated. To show that random variables can be uncorrelated without being independent we consider the following example

X	$Y = -1$	0	1	
1	0	0.2	0	0.2
0	0.3	0.2	0.3	0.8
	0.3	0.4	0.3	

The table gives the joint distribution of X and Y. The numbers on the edge give the marginal distributions of X and Y. In this example $EX = 0.2$, $EY = 0.3(-1) + 0.3(1) = 0$, and $EXY = 0$ since XY is always equal to 0. This implies

$$EXY = 0 = EX \cdot EY$$

so X and Y are uncorrelated.

Example 4.17. Birthday problem. Given a set of n people, let $X_{i,j} = 1$ if the ith and jth persons have the same oirthday. As we have noted earlier

$$P(X_{ij} = 1, X_{k\ell} = 1) = P(X_{ij} =)P(, X_{k\ell} = 1)$$

so we have

$$E(X_{ij}X_{k\ell}) = E(X_{ij})E(X_{k\ell})$$

Using Theorem 4.6 now we see that the number of borthday coincidencea $N = \sum_{i<j} X_{ij}$ has

$$\text{var}\,(N) = C_{n,2}\frac{1}{365} \cdot \frac{364}{365}$$

To illustrate the use of Theorem 4.6

Example 4.18. Hypergeometric distribution. Suppose we have an urn with n balls, m of which are red. Let $X_i = 1$ if we draw a red ball on the ith try. Let $R_k = X_1 + \cdots X_k$ be the number of red balls we get when we draw from the urn k times without replacement. By

$$ER_k = kEX_1 = kp \quad \text{where } p = m/n.$$

The X_i are Bernoulli(p) so $\text{var}\,(X_i) = p(1-p)$. To compute the variance using Theorem 4.6 we need to compute $\text{cov}\,(X_i, X_j)$. To do this we begin with

$$E(X_1X_2) = P(X_1 = 1, X_2 = 1) = \frac{m}{n} \cdot \frac{m-1}{n-1}$$

so we have

$$\begin{aligned} \text{cov}\,(X_1, X_2) &= \frac{m}{n} \cdot \frac{m-1}{n-1} - \frac{m}{n} \cdot \frac{m}{n} \\ &= \frac{m}{n}\left(\frac{(m-1)n - m(n-1)}{n(n-1)}\right) \\ &= -\frac{m}{n} \cdot \frac{n-m}{n(n-1)} = -\frac{p(1-p)}{n-1} \end{aligned}$$

so it follows from Theorem 4.6 that

$$\begin{aligned} \text{var}\,(R_k) &= kp(1-p) - k(k-1)\frac{p(1-p)}{n-1} \\ &= kp(1-p)\left[1 - \frac{k-1}{n-1}\right] \end{aligned} \tag{4.13}$$

Note that the variance of R_k is always less that the variance of binomial(k,p), and that $\text{var}\,(R_n)$ since at that point we have drawn all of the balls out of the urn and hence $R_n = m$.

Example 4.19. The elevator problem. We return to the siutation in which 10 passengers get on an elevator in a seven story building. Each indpendently chooses one of the six higher floors. Let N be the number of floors that the elevator does NOT stop at. We want to find the variance of N.

Let $X_i = 1$ if no one chooses floor i where $2 \le i \le 7$.

$$P(X_i = 1) = (5/6)^{10} = 0.1615$$

so $EN = 6(0.1615) = 0.969$.

$$\begin{aligned} \text{var}\,(X_i) &= \left(\frac{5}{6}\right)^{10}\left[1 - \left(\frac{5}{6}\right)^{10}\right] = 0.1354 \\ \text{cov}\,(X_2, X_3) &= P(X_2 = 1, X_3 = 1) - P(X_2 = 1)P(X_3 = 1) \\ &= (4/6)^{10} - (5/6)^{20} = -0.00874 \end{aligned}$$

Using Theorem 4.6 now and noting there are 6 terms in the first sum and $6 \cdot 5 = 3)$ in the second

$$\text{var}\,(N) = 6(0.1354) - 30(0.0874 = 0.5502$$

If the X_i were independent the answer would be the first term swhich is 0.8124

4.4 Laws of Large Numbers

Let $X_1, X_2, \ldots$ be **independent and identically distributed (i.i.d)** with $X_i = \mu$ and $\text{var}\,(X_i) = \sigma^2$. Let $S_n = X_1 + \cdots + X_n$. Using Theorems 1.2 and 4.5 we have

$$ES_n = n\mu \qquad \text{var}\,(S_n) = n\sigma^2 \quad \text{and} \quad \sigma(S_n) = \sigma\sqrt{n}. \tag{4.14}$$

If we let $\bar{X}_n = S_n/n$ be the **sample mean** then using (1.15) and (4.2).

$$E\bar{X}_n = \mu \qquad \text{var}\,(\bar{X}_n) = \sigma^2/n \quad \text{and} \quad \sigma(\bar{X}_n) = \sigma\sqrt{n}. \tag{4.15}$$

In statistical lingo, the first equation says that $\bar{X}_n$ is an **unbiased estimator** of μ. The second shows that the variance of $\bar{X}_n$ tends to 0 as $n \to \infty$. Thus we expect that if n is large $\bar{X}_n$ is close to μ with high probability.

Theorem 4.7. Weak law of Large Numbers. *Let $X_1, X_2, \ldots$ be i.i.d with $EX_i = \mu$ and var $(X_i) = \sigma^2$. For any $\epsilon > 0$, as $n \to \infty$*

$$P(|\bar{X}_n - \mu| > \epsilon) \to 0$$

To prove this we need some lemmas.

Lemma 4.1. *If $X \le Y$ and $E|X|, E|Y| < \infty$ then $EX \le EY$.*

This is similar to the calculus fact that if $f \le g$ then $\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx$.

Proof. If (X, Y) has a discrete distribution then using the fact that $X \le Y$

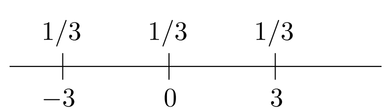
$$EX = \sum_{x,y} xP(X = x, Y = y) \le \sum_{x,y} yP(X = x, Y = y) = EY$$

The continuous case is the same with integrals instead of sums

$$EX = \iint x f_{X,Y}(x, y)\,dx\,dy \le \iint y f_{X,Y}(x, y)\,dx\,dy = EY$$ □

To motivate the next result we begin with a little puzzle. If $EX = 0$ and $\text{var}\,(X) = 6$ how large can $P(|X| \ge 3)$ be?

The solution is:



To explain: (i) if we have any distribution then we can make the mean 0 and not change the variance by making it symmetric about 0. (ii) there is no reason to put any mass on$\{x : |x| > 3\}$ because we could reduct the variance by moving all of the mass to 3. (iii)

Similarly, there is no reason to put any mass on $\{x : 0 < |x| < 3\}$ because we could reduct the variance by moving all of the mass to 0. This means that the optimal distribution has

$$P(X = 3) = P(X = -3) = a \qquad P(X = 0) = 1 - 2a$$

In order have $6 = \text{var}(X) = EX^2$ we want $2a \cdot 9 = 6$ or $a = 1/3$.

Lemma 4.2. Chebyshev's inequality. *Suppose $EZ = \mu$ and $E(Z^2) < \infty$. Then for any $z > 0$*

$$P(|Z - \mu| > z) \leq \text{var}(Z)/z^2$$

Before we do the proof we need some notation. Given an event A we define a random variable called **the indicator of** A by

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Proof. Let $Y = |Z - \mu|$ and $A = \{Y \geq z\}$. Since $Y \geq 0$ and $Y \geq z$ on A

$$Y^2 \geq Y^2 \cdot 1_A \geq z^2 \cdot 1_A$$

Taking expected value and using Lemma 4.1

$$\text{var}(Z) = EY^2 \geq z^2 P(A)$$

Diviging both sides by z^2 gives the desired result. □

.

Proof of Theorem 4.7. Let $Z = \bar{X}_n$. By (4.15), $EZ = \mu$ and $\text{var}(Z) = \sigma^2/n$. Using Chebyshev's inequality now with $z = \epsilon$

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon n} \to 0 \tag{4.16}$$

which proves the desired result. □

The weakness of Theorem 4.7 is that it shows that when n is large $\bar{X}_n$ is close to μ with high probability but it does not guarantee that with probability one the sequence of numbers $\bar{X}_n$ converges to μ. The next result does this. Note that it only assumes $E|X_i| < \infty$.

Theorem 4.8. Strong Law of Large Numbers. *Let $X_1, X_2, \ldots$ be i.i.d with $E|X_i| < \infty$ and $EX_i = \mu$. Then $\bar{X}_n \to \mu$ as $n \to \infty$.*

A corollary of this result is the frequency interpretation of probability. Let $X_i = 1$ if the event A occurs on the ith trial and 0 otherwise. $EX_i = P(A)$. Theorem 4.8 implies that $\bar{X}_n$ = the fraction of times A occurs in the first n trials converges to $P(A)$. While Theorem 4.8 is nice, for practical purposes Theorem 4.7 suffices since it says that if n is large the sample mean is close to the true mean with high probability.

4.5 Central Limit Theorem

At this point we know that if $X_1, X_2, \ldots$ are independent and have the same distribution with mean $EX_i = \mu$ and variance $\text{var}\,(X_i) = \sigma^2 \in (0, \infty)$ then, see (4.14), the sum $S_n = X_1 + \cdots + X_n$ has mean $ES_n = n\mu$ and $\text{var}\,(S_n) = n\sigma^2$. Using (1.15) and (4.2) we see that $S_n - n\mu$ has mean 0 and variance $n\sigma^2$ so

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \quad \text{has mean 0 and variance 1.}$$

The remarkable fact, called the central limit theorem is that as $n \to \infty$ this scaled variable converges to the standard normal distribution.

Theorem 4.9. *Suppose $X_1, X_2, \ldots$ are independent and have the same distribution with mean $EX_i = \mu$ and variance $\text{var}\,(X_i) = \sigma^2 \in (0, \infty)$. Then for all $a < b$*

$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$$

To apply this result we need to learn how to use the normal table. If we let $\Phi(x) = P(\chi \leq x)$ be the normal distribution function then

$$P(a \leq \chi \leq b) = \Phi(b) - \Phi(a)$$

The values of Φ for positive values of x are given in the table at the back of the book. By symmetry$\Phi(-x) = P(\chi \leq -x) = P(\chi \geq x) = 1 - \Phi(x)$ so

$$P(-x \leq \chi \leq x) = \Phi(x) - (1 - \Phi(x)) = 2\Phi(x) - 1 \tag{4.17}$$

To illustrate the use of Theorem 4.9 we will use a small part of the table

x	0	1	2	3
$\Phi(x)$	0.500	0.8413	0.9772	0.9986
$P(-x \leq \chi \leq x)$	0	0.6826	0.9544	0.9972

In words, for the normal distribution the probability of being within one standard deviation of the mean is 68%, within two standard deviations is 95%, and the probability of being more that three standard deviations away is ¡ 0.3%.
We begin by considering the situation $P(X_i = 1) = P(X_i = -1)$ which has $EX_i = 0$ and $\text{var}\,(X_i) = E(X_i^2) = 1$. Taking $a = -1$ and $b = 1$ and using our little table

$$P(-\sqrt{n} \leq S_n \leq \sqrt{n}) \approx P(-1 \leq \chi \leq 1) = 0.6826$$

Thus if $n = 2500 = 50^2$ our net winnings will be $\in [-50, 50]$ with probability ≈ 0.6826. Since $1275 - 1225 = 50$ this means that the number of heads will be $\in [1225, 1275]$ with that probability. $1275/2500 = 0.51$ so this corresponds to the fraction of heads $\in [0.49, 0.51]$.

Using the other two values in the table, if $n = 2500$ our net winnings will be in $[-100, 100]$ with probability ≈ 0.9544, and in $[-150, 150]$ with probability ≈ 0.9972. The last result shows that the bound from Chebyshev's inequality can be very crude. Chebyshev tells us that

$$P(|\chi| \geq 3) \leq \frac{\text{var}\,(\chi)}{3^2} = \frac{1}{9}$$

while the true value of $P(|\chi| \geq 3) = 1 - 0.9972 = 0.0028 \approx 1/357$.

Turning to a smaller value of n.

Example 4.20. Suppose we flip a coin 100 times. What is the probability we get at least 56 heads?

Let $X_i = 1$ if the ith flip is heads, 0 otherwise, and $S_{100} = X_1 + \cdots + X_{100}$. $EX_i = 1/2$ and $\text{var}(X_i) = 1/4$ so $ES_{100} = 50$ and $\sigma(S_{100}) = \sqrt{100/4} = 5$. To apply the central limit theorem we write

$$\begin{aligned} P(S_{100} \geq 56) &= P\left(\frac{S_{100} - 50}{5} \geq \frac{6}{5}\right) \approx P(\chi \geq 1.2) \\ &= 1 - P(\chi \leq 1.2) = 1 - 0.8849 = 0.1151 \end{aligned}$$

If the question has been “What is the probability of at most 55 heads?” we would have computed

$$P(S_{100} \leq 55) = P\left(\frac{S_{100} - 50}{5} \leq \frac{5}{5}\right) \approx P(\chi \leq 1.0) = 0.8413$$

To see that there is a problem with what we have done note that

$$1 = P(S_{100} \leq 55) + P(S_{100} \geq 56) \approx 0.8413 + 0.1151 = 0.9564$$

The solution is to regard each integer k as an interval $[k - 0.5, k + 0.5]$ so $P(S_{100} \leq 55)$ translates into $P(S_{100} \leq 55.5)$ and $P(S_{100} \geq 56)$ becomes $P(S_{100} \geq 55.5)$. Doing this we compute

$$\begin{aligned} P(S_{100} \geq 55.5) &= P\left(\frac{S_{100} - 50}{5} \geq \frac{5.5}{5}\right) \approx P(\chi \geq 1.1) \\ &= 1 - P(\chi \leq 1.1) = 1 - 0.8643 = 0.1357 \end{aligned}$$

which is a more accurate approximation of the exact probability 0.135627 than was out first answer 0.1151.

The correction used in the previous example is called the **histogram correction** because we think of drawing rectangles of hight $P(S_{100} = k)$ and base $[k - 0.5, k + 0.5]$ to get an approximation of the normal density.mSee Figure 4.2.

The next example shows this is not only a device to get more accurate approximations but it allows us to get answers when a naive application of the central limit theorem would give a nonesense answer.

Example 4.21. Suppose we flip 16 coins. What is the probability we get exactly 8 heads?

The mean number of heads in 16 tosses is $16/2 = 8$ and the standard deviation is $\sqrt{16/4} = 2$. If we write

$$P(S_{16} = 8) = P\left(\frac{S_{16-8}}{2} = 0\right) \approx P(\chi = 0)$$

then we get 0 since χ has a continuous distribution. If we consider 8 to be $[7.5, 8.5]$ then

$$\begin{aligned} P(7.5 \leq S_{16} \leq 8.5) &= P\left(\frac{0.5}{2} \leq \frac{S_{16} - 8}{2} \leq \frac{0.5}{2}\right) \approx P(-0.25 \leq \chi \leq 0.25) \\ &= 2P(\chi \leq 0.25) - 1 = 2(0.5987) - 1 = 0.1974 \end{aligned}$$

where we have used (4.17). The exact answer is

$$2^{-16} \frac{16!}{8!8!} = 0.1964$$

Similar calculations give the following results

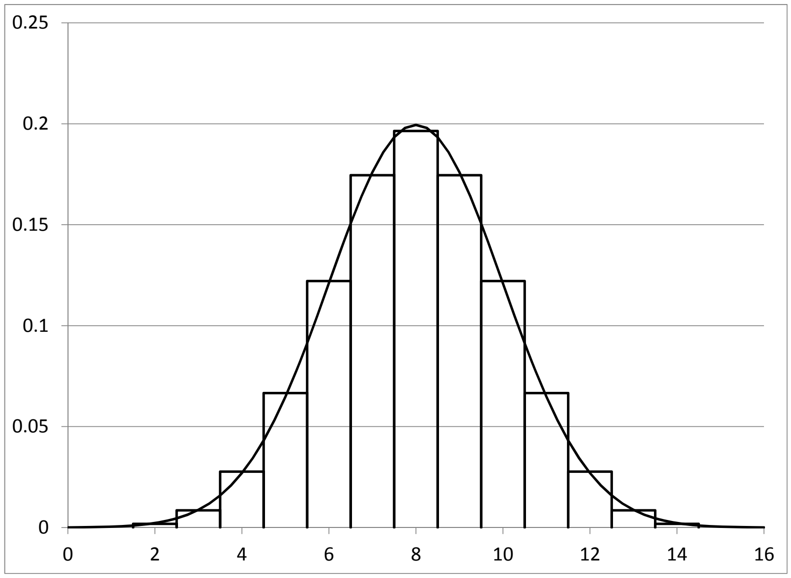


Figure 4.2: Binomial(16,0.5) compared with normal approximation.

k	8	7,9	6,10	5,11	4,12	3,13
normal approx.	.1974	.1747	.1209	.0656	.0279	.0092
exact ans.	.1964	.1746	.1221	.0666	.0277	.0085

Example 4.22. In 2007 the Boston Red Sox won 96 games and lost 66. How likely is this result if the games were decided by flipping coins?

If they were flipping coins the mean number of wins would be $162/2 = 81$ and the standard deviation would be $\sqrt{162(0.5)(0.5)} = 6.36396$. The event $W \geq 96$ translates into $W \geq 95.5$ when we apply the histogram correction. This is 14.5 above the mean or $14.5/6.36936 = 2.28$ standard deviations. The normal approximation is $P(\chi \geq 2.28) = 1 - 0.9887 = 0.0113$.

This probability is small but we must keep in mind that the Boston Red Sox are just one of 30 teams in baseball. This means that the expected number of teams with ≥ 96 wins is 0.339, so the Boston Red Sox record is not unusual for the best team in baseball. On the other hand Figure 4.3 shows that if we look at the records of all 30 teams, baseball game outcomes look different from flipping coins.

Example 4.23. Ipolito vs. Power. This court case considered an election where the winning margin was 1422 to 1405 but 101 votes had to be thrown out. To reverse the election the first candidate would have to lose at least m votes where

$$1422 - m = 1405 - (101 - m) \quad \text{or} \quad 2m = 118$$

Solving gives $m = 59$. The fraction of people voting for the first candidate was 0.5030, so we will simplifythe arithmetic by supposing that the votes lost by the first candidate has the same distribution as the number of heads when we flip 101 fair coins. Replacing ≥ 59

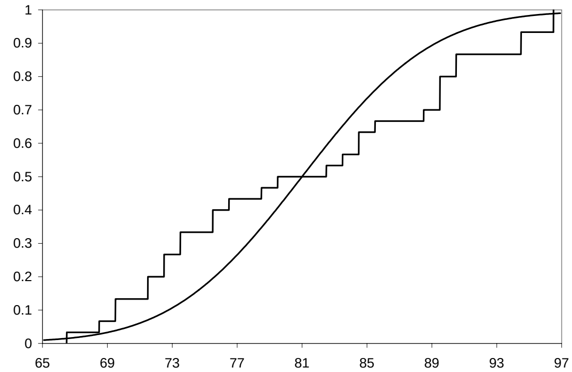


Figure 4.3: The number of teams with $\leq k$ wins plotted versus the normal distribution function with mean 81 and standard deviation 6.36396.

by ≥ 58.5

$$P(S_{101} \geq 58.5) = P\left(\frac{S_{101} - 50.5}{\sqrt{101/4}} \geq \frac{8}{5.025}\right)$$
$$\approx P(\chi \geq 1.59) = 1 - 0.9441 = 0.0559$$

The judge ruled that "it does not strain the probabilities to assume a likelihood that the questioned votes produced or could produce a change in the result."

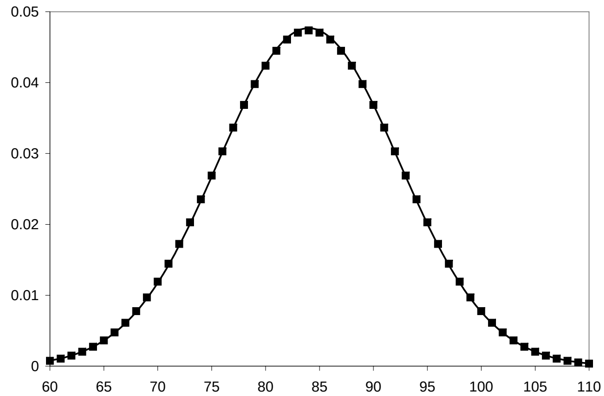
In Chapter 1 we considered *De Martini v. Power*). In a close election in a small town, 2,656 people (50.6%) voted for candidate A compared to 2,594 who voted for candidate B, a margin of victory of 62 votes. An investigation revealed that 136 people vted who should not have. Earlier we computed that In order to reverse the outcome of the election, candidate A must lose $m \geq 99$ votes. If we again suppose that the number of votes lost by candidate A has the same distribution as flipping 136 fair coins, then the mean number of votes lost is 68 and the standard deviation is $\sqrt{136/4} = 5.831$. Writing $m \geq 99$ as $m \geq 98.5$, $98.5 - 68 = 30.5$ is $30.5/5.831 = 5.23$ standard devations. Using normalcdf we see this probability is 8.49×10^{-8}, which is a little bit larger than the exact probability is 7.492×10^{-8} but clearly indicates that the outcome should not change if 136 randomly chosen votes are thrown out

Example 4.24. Dice. Suppose we roll a die 24 times. What is the probability that the sum of the numbers is ≥ 100.

As the figure shows the distribution of the sum of 24 dice is approximately normal (smooth curve). To apply (4.9) we note that $ES_{24} = 24(7/2)$ while $\sigma(S_{24}) = \sqrt{24 \cdot 35/12} = \sqrt{70} = 8.366$. Using the histogram correction we replace ≥ 100 by ≥ 99.5 and we have

$$P(S_{24} \geq 99.5) = P\left(\frac{S_{24} - 84}{\sqrt{70}} \geq \frac{15.5}{8.366}\right)$$
$$\approx P(\chi \geq 1.85) = 1 - P(\chi \leq 1.85) = 0.0322$$

compared with the exact answer 0.031760.



Example 4.25. Roulette. Consider a person who plays roulette 100 times and bets \$1 on black each time. What is the probability they will be ahead, i.e., their net winnings are ≥ 0?

The outcome of the ith play has $P(X_i = 1) = 18/38$ and $P(X_i = -1) = 20/38$. $EX_i = -1/19 = -0.05263$ while $EX_i^2 = 1$ and $\text{var}(X_i) = 1 - (1/19)^2 = 0.9972$ To simplify the arithetic we will use 1 instead of 0.9972. The mean of 100 plays $ES_{100} = -5.263$ while the $\sigma(S_{100}) = \sqrt{100} = 10$. To do the histogram correction we note that after 100 plays S_{100} can only take on even values, so we write $S_{100} \geq 0$ as $S_{100} \geq -1$. That is we place the boundary half-way between -2 and 0. The rest is the same

$$P(S_{100} \geq -1) = P\left(\frac{S_{100} - (-5.263)}{10} \geq \frac{4.263}{10}\right)$$
$$\approx P(\chi \geq 0.4263) = 1 - P(\chi \leq 0.4263) = 0.33494$$

The exact value is $1 - \text{binomcdf}(100, 18/38, 49) = 0.33431$.

Example 4.26. Normal approximation to Poisson. Suppose that in Durham each year an average of 25 postal workers are bitten by dogs. In a recent year there were 33 incidents. Is that number of unusually high?

Since dog bites are rare events and there are a large number of postmen the distribution of the number of dog bites should be approximately Poisson with mean 25. With modern calculators we don't have to think to solve this problem

$$\text{Poissoncdf}(25, 32) = 0.92854$$

so the probability there will be 33 or more people bitten by dogs is ≈ 0.07146. However we can also solve the problem by using the normal to approximate the Poisson. A Poisson(25) random variable is the sum of 25 independent Poisson(1) random variables. Since it is a sum of a large number of independent random variables the distribution should be approximately normal with mean 25 and variance 25. Using the histogram correction with the normal approximation

$$P(S_{25} \geq 32.5) = P\left(\frac{S_{25} - 25}{5} \leq \frac{7.5}{5}\right)$$
$$\approx P(\chi \geq 1.5) = 1 - 0.9332 = 0.0668.$$

Example 4.27. Let $Z = \text{Poisson}(16)$. Use the central limit theorem to approximate the the probability $Z = 16$?

The mean of Z is 16, the variance is 16 so the standard deviation is 4. Using the central limit theorem $(Z - 16)/4$ is approximatelystandard normal. Replacing 16 by the interval $[15.5, 16.5]$ the probability we want is $P(15.5 \le Z \le 16.5)$ or

$$\begin{aligned} P\left(-\frac{0.5}{4} \le \frac{Z-16}{4} \le \frac{0.5}{4}\right) &\approx P(-0.125 \le \chi 0.125) \\ &= nomralcdf(-0.125, 0.125) = 0.09945 \end{aligned}$$

The exact value of

$$P(Z = 16) = e^{-16}\frac{(16)^{16}}{16!} = 0.09922$$

Example 4.28. Suppose that the average weight of a man is 182 pounds with a standard deviation of 40 pounds. A large plane can hold 200 men. What is the probability that the total weight of the passengers will be more than 38,000 pounds.

The expected value of $S_{200} = 200 \cdot 182 = 36,400$. The standard deviation is $40\sqrt{200} = 565.68$. These are continuous random variables so there is no histogram correction.

$$\begin{aligned} P(S_{200} \ge 38,000) &= P\left(\frac{S_{200} - 36,400}{565.68} \ge \frac{1600}{565.68}\right) \\ &\approx P(\chi \ge 2.828) = 1 - P(\chi \le 2.828) = 0.00234 = 1/426.97 \end{aligned}$$

4.6 Confidence Intervals for Proportions

In many situations such as using a poll to forecast the outcome of an election we don't know the underlying value of p so we want to use the fraction of voters in the sample to estimate p. We begin with the observation that

$$P(-2 \le \chi \le 2) = 0.9544$$

If we have n independent observations that have $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$ then $\bar{X}_n$ has mean p and standard deviation $\sqrt{p(1-p)/n}$ so by the Central Limit Theorem

$$P\left(-2 \le \frac{\bar{X}_n - p}{\sqrt{p(1-p)/n}} \le 2\right) \approx 0.95$$

Changing notation so that $\hat{p} = \bar{X}_n$ is our estimate for p, and doing some algebra

$$P\left(|\hat{p} - p| \le 2\sqrt{\frac{p(1-p)}{n}}\right) \approx 0.95$$

Since we observe $\hat{p}$ and want to estimate p we rewrite this as

$$P\left(p \in \left[\hat{p} - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]\right) \approx 0.9544$$

where we have used $\hat{p}$ in the formula since we do not know the true value of p. We call

$$\left[\hat{p} - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \tag{4.18}$$

a **95% confidence interval** for the true value of p since if n is large the true value will lie in this interval $\approx 95\%$ of the time. By analogy with a circle we call $2\sqrt{\hat{p}(1-\hat{p})/n}$ the **radius** of the confidence interval since it is the distance from the center to the boundary. The radius represents the accuracy in using $\hat{p}$ to estimate p.

Remark. Many books replace 2 by 1.96 since $P(-1.96 \le \chi \le 1.96) = 0.95$. We will use 2 to make the arithmetic simpler and to acknowledge that this is only an approximation.

Example 4.29. During the 2015–2016 basketball season, Grayson Allen made 206 of his 245 free throw attempts. Find a 95% confidence interval for his true free throw shooting percentage.

$\hat{p} = 206/245 = 0.8408$ so using (4.18) the 95% confidence interval is

$$\left[.8408 \pm 2\sqrt{\frac{0.8408(0.1592)}{245}} = 0.8408 \pm 0.0234\right] = [0.8174, 0.8642]$$

Example 4.30. Of the first 10,000 votes cast in an election, 5,180 were for candidate A. Find a 95% confidence interval for the fraction of votes that candidate A will receive.

Plugging into eqref95CI

$$0.518 \pm 2\sqrt{(.518)(.482)/10,000} = [0.508, 0.528]$$

Note that this does not depend on the number of votes in the election. So if the sample is representative of the overall population the outcocme can be predicted from a small fraction of the votes cast.

Example 4.31. Weldon's dice data. An English botanist named Walter Frank Raphael Weldon was interested in the the "pip effect" in dice – the idea that the spots on pips which on some dice are produced by making small holes in the surface, would make the sides with more spots lighter and more likely to come up. As he wrote in a letter to Francis Galton on February 2, 1894, he threw 12 dice 26,036 times for a total of 315,672 throws. He observed a 5 or 6 come up on 106,602 throws or with probability 0.33770. A 95% confidence interval for the true value of p is

$$0.33770 \pm 2\sqrt{\frac{0.3377 \cdot 0.6623}{315,672}} = 0.33770 \pm 0.00168 = [0.33602, 0.33938]$$

Figure 4.4 compares the number of 5's and 6's in the 26,036 throws with the binomial(12,1/3) distribution. The discrepancy is small but visible. The difference from the uniform distribution is not enough to be noticeable by people who play games for amusement, but it is perhaps large enough to be of concern by a casino that entertains tens of thousands of gamblers a year. For this reason most casino use dice with no pips.

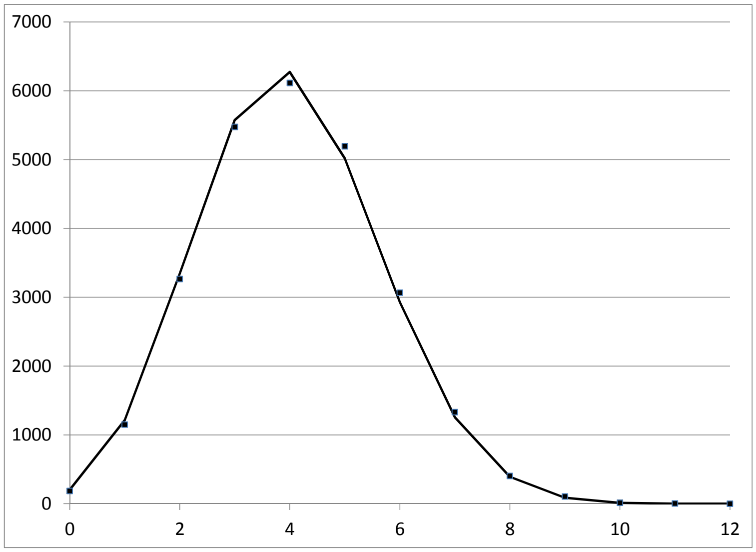


Figure 4.4: Comparison of Weldon's 26,306 rolls of 12 dice with the binomial(12,1/3) distribution.

Example 4.32. Sample size selection. Suppose that we want to forecast the result of a close election so that the radius of the confidence interval is less than 3%. How large a sample do we need?

The radius of the confidence interval is $2\sqrt{p(1-p)/n}$. Since we don't know p before we take the poll, we ask the question: what is the largest possible value of $f(p) = p(1-p) = p - p^2$ wehn $p \in [0,1]$. Differentiating twice $f'(p) = 1-2p$ and $f''(p) = -2$, so the maximum value occurs at 1/2. The value of $2\sqrt{p(1-p)/n}$ at $p = 1/2$ is $1/\sqrt{n}$. This is conservative but in the context of a fairly close election is not too bad. If $0.4 \le p \le 0.6$ then $2\sqrt{p(1-p)} \ge 2\sqrt{0.24} = 0.97979$, so without much error we can suppose that the radius of the confidence interval is $1/\sqrt{n}$. To have $1/\sqrt{n} = 0.03$ we need $\sqrt{n} = 1/0.03$ or $n = (1/0.03)^2 = 1,111$. To be accurate within 0.01 we would need $n = (1/0.01)^2 = 10,000$.

Example 4.33. The Literary Digest poll. In order to forecast the outcome of the 1936 election, the Literary Digest polled 2.4 million people and found that 57% would vote for Alf Landon while 43% were going to vote for Franklin Roosevelt. Use our simplified arithmetic again, a 95% confidence interval based on this sample would be

$$0.57 \pm \frac{1}{\sqrt{2.4 \times 10^6}} = 0.57 \pm 6.45 \times 10^{-4}$$

but Roosevelt won getting 62% of the vote.

To explain how they made such a large error we have to look at the methods Literary Digest used. They sent 101 million questionnaires to people whose names came from telephone books and club membership lists. Since many of the 9 million unemployed did not belong to clubs or have telephones the sample was not representative of the population as a whole.

It is interesting to note that George Gallup who was just getting started in the polling business, predicted based on a sample of size 50,000 that Roosevelt would get 56% of the vote. A 95% confidence interval for that sample would have radius

$$0.56 \pm \frac{1}{\sqrt{50,000}} = .56 \pm 0.0045$$

The true percentage of 0.62 is outside this interval but he got much closer to the actual answer with a much smaller sample.

4.7 Exercises

Moments, Variance

1. Can we have a random variable with $EX = 3$ and $EX^2 = 8$?

2. Suppose X and Y are independent with $EX = 1$, $EY = 2$, $\text{var}(X) = 3$ and $\text{var}(Y) = 1$. Find the mean and variance of $3X + 4Y - 5$.

3. The Elm Tree golf course in Cortland, NY is a par 70 layout with 3 par fives, 5 par threes, and 10 par fours. Find the mean and variance of par on this course.

4. Find the mean and variance of the number of games in the World Series. Recall that it is won by the first team to win four games and assume that the outcomes are determined by flipping a coin.

5. In a group of six items, two are defective. Find the distribution of N the number of draws we need to find the first defective item. Find the mean and variance of N.

6. A Scrabble set has 100 tiles. he point value of a randomly chosen scrabble tile, X has the following distribution:

value	0	1	2	3	4	5	8	10
prob.	.02	.68	.07	.08	.10	.01	.02	.02

Find the mean, variance, and standard deviation.

7. Roll two dice and let $Z = XY$ be the product of the two numbers obtained. Find the mean, variance, and standard deviation of Z.

Continuous distributions

8. Suppose X has density function $6x(1-x)$ for $0 < x < 1$ and 0 otherwise. Find the mean and variance of X.

9. Suppose X has density function $x^2/9$ for $0 < x < 3$ and 0 otherwise. Find the mean and variance of X.

10. Suppose X has density function $x^{-2/3}/21$ for $1 < x < 8$ and 0 otherwise. Find the mean and variance of X.

11. Suppose X has density function $30x^2(1-x)^2$ for $0 < x < 1$. Find the mean variance and standard deviation of X.

12. Let U_1, U_2, U_3, U_4 be independent uniform on $(0,1)$. Let $Z = \max_{1 \le i \le 4} U_i$. Find the density function for Z, EZ, $\text{var}(Z)$, and $\sigma(Z)$.

13. Let $U_1, U_2, \ldots U_n$ be independent uniform on $(0,1)$. Let $Z = \max_{1 \le i \le n} U_i$. Find the density function for Z, EZ, and $\text{var}(Z)$.

14. Suppose R has the Rayleigh distribution, which has density $re^{-r^2/2}$ for $0 < r < \infty$. Find the mean and the variance of R.

15. Suppose X is exponential(λ). Integrate by parts and use induction to conclude $EX^m = m!/\lambda^m$.

16. Suppose X has the standard normal distribution. By symmetry $EX^{2k-1} = 0$. Integrate by parts and use induction to conclude $EX^{2k} = (2k-1)(2k-3)\cdots 3cdot1$.

Distributions of sums of independent random variables

17. We roll two six-sided dice, one with sides 1,2,2,3,3,4 and the other with sides 1,3,4,5,6,8. What is the distribution of the sum?

18. Suppose we roll three tetrahedral dice that have 1, 2, 3, and 4 on their four sides. Find the distribution for the sum of the three numbers.

19. Let S_4 be the sum of the numbers when four six-sided dice are rolled. (a) Find the mean, variance, and standard deviation of S_4 (a) Find $P(S_4 = 9)$. (b) Find the value of m that maximizes $P(S_4 = m)$ and give the value of that probability. Hint for (b) and (c): $S_4 = S_2 + S_2'$ where S_2 and s_2' are independent rolls of two dice.

20. (a) Let X be uniform on $1, \ldots 6$ (i.e., a normal die) and Y be uniform on $1, \ldots, 10$ one of the unusual dice used in Dungeons and Dragons. Suppose X and Y are independent. Find the distribution of $X + Y$. (b) Solve the last problem when 6 is replaced by m and 10 is replaced by $n > m$.

21. Let U_1, U_2, and U_3 be independent and uniform on (0,1) (a) Find the density function f for $U_1 + U_2$. (b) Find the density function g for $U_1 + U_2 + U_3$. In the first case the density will be symmetric about 1, in the second syummetric about 3/2,

22. Let X and Y be independent and exponential(λ). Find the density function of $X - Y$.

23. Let X = exponential(λ) and Y = expoential(μ) be independent. Find the density function of $X + Y$.

24. (a) Let X = unifrom on $(0,1)$ and Y = exponential(λ) be independent. Find the density function of $X + Y$. (b) Generalize the last result to the case in which Y has density function f and distribution function F.

Variance of Sums

25. In a class with 18 boys and 12 girls, boys have probability 1/3 of knowing the answer and girls have probability 1/2 of knowing the answer to a typical question the teacher asks. Assuming that whether or not the students know the answer are independent events, find the mean and variance of the number of students who know the answer.

26. At a local high school, 12 boys and 4 girls are applying to MIT. Suppose that the boys have a 10% chance of getting in and the girls have a 20% chance. (a) Find the mean and variance of the number of students accepted. (b) What is more likely: 2 boys and no girls accepted or 1 boy and 1 girl?

27. A basketball player hits a three point shot with probability 0.1. (a) Find the mean and the standard deviation of N_3 the number of shots he needs to make 3 of them. (b) Find $P(N_3 = 29)$.

28. Suppose we roll a die repeatedly until we see each number at least once and let R be the number of rolls required. Find the ER, $\text{var}\,(R)$, and $\sigma(R)$.

29. There are 12 astrological signs: Aries, Taurus, ... Pisces. Suppose that each sign is present in 1/12. Let M be the number of people we have to meet to have seen all 12 signs. Find EM, $\text{var}\,(M)$, and $\sigma(M)$.

30. The two previous problems are instances of the coupon collector's problem. At various times people have cards with pictures of baseball players or Pokemon. Suppose that there are n cards and they are equally likely. Let N_n be the number we need to buy to get a complete set. (a) Find formulas for EN_n and $\text{var}\,(E_n)$. (b) Show that $EN_n \sim \log n$ where $a_n \sim b_n$ means $a_n/b_n \to 1$. (c) Show that $\text{var}\,(N_n)/n^2$ stays bounded. Combining (b) and (c) with Chebyshev's inequality tells us that $P(|N_n/\log n - 1| > \epsilon) \to 0$ as $n \to \infty$ but proving this is not part of the problem.

31. Four men and four women are randomly assigned to two person canoes. Find the mean and variance of N, the number of canoes with one person of each sex.

32. Suppose we draw 13 cards out of a deck of 52. What is the variance of $N =$ the number of aces we get? How does this compare to the variance of the binomial(4,1/2)?

33. Consider the elevator problem in which 10 people will choose randomly to get off at one of six floors. Compute the variance of the number of stops its makes by computing $\text{var}\,(6 - S)$.

34. In the Noah's ark problem there are 50 pairs of animals and 10 will die. Find the variance of $N =$ the number of complete pairs in the Noah's ark problem.

35. The U.S. senate has 100 members. (a) Isgnoring February 29, what is the probability at least one senator is born on April 1? (b) Let N be the number of days on which at least one student has a birthday. Find the expected value EN. (c) Find $\text{var}\,(N)$. by computing $\text{var}\,(365 - N)$.

36. Let $X_1, X_2, \ldots$ be independent continuous random variables. We say that a record occurs at time n if $V_n > \max\{X_1, X_2, \ldots X_{n-1}\}$. Let $Y_i = 1$ if a record occurs at time i. Let $R_n = Y_1 + \cdots + Y_n$ be the number of records that have occurred by time n.

Covariance

37. Suppose X takes on the values $-2, -1, 0, 1, 2$ with probability 1/5 each, and let $Y = X^2$. (a) Find $\text{cov}\,(X, Y)$. (b) Are X and Y independent?

38. Let Z have the standard normal distribution. Find $\text{cov}\,(Z, Z^2)$.

39. Let U be uniform on $(0, 1)$. Find $\text{cov}\,(U^2, U^3)$.

40. Suppose (X, Y) is uniformly distributed on $0 < x < y < 1$. Find the marginal densities of X and Y and $\text{cov}\,(X, Y)$.

Central Limit Theorem, I. Coin Flips

41. A person bets you that in 100 tosses of a fair coin the number of Heads will differ from 50 by 4 or more. What is the probability you will win this bet?

42. Suppose we toss a coin 100 times. Which is bigger, the probability of exactly 50 Heads or at least 60 Heads?

43. Bill is a student at Cornell. In any given course he gets an A with probability 1/2 and a B with probability 1/2. Suppose the outcome of his courses are independent. In his four years at Cornell he will take 33 courses. If he can get at least 22 A's he can graduate with a 3.666 average (or better). What is the probability he will do this?

44. For a 162 game baseball season find the approximate probability that a team with a 0.5 chance of winning will win at least 87 games.

45. British Airways and United offer identical service on two flights from New York to London that leave at the same time. Suppose that they are competing for the same pool of 400 customers who choose an airline at random. What is the probability United will have more ($>$) customers than its 230 seats?

46. A probability class has 30 students. As part of an assignment, each student tosses a coin 200 times and records the number of heads. What is the probability no student gets exactly 100 heads?

47. A fair coin is tossed 2500 times. Find a number m so that the chance that the number of heads is between 1250-m and 1250+m is approximately 2/3.

CLT, II. Biased Coins

48. Suppose we roll a die 600 times. What is the approximate probability that the number of 1's obtained lies between 90 and 110?

49. Suppose that each of 300 patients has a probability of 1/3 of being helped by a treatment. Find approximately the probability that more than 120 patients are helped by the treatment.

50. Suppose that 10% of a certain brand of jelly beans are red. Use the normal approximation to estimate the probability that in a bag of 400 jelly beans there are at least 45 red ones.

51. Basketball player Marshall Plumlee made roughly 60% of his free throws during the 2015-2016 season. Use the normal approximation to the binomial to compute the probability he makes 18 or more free throws if he attempts 24.

52. A casino owner is concerned based on past experience that his dice show 6 too often. He makes his employees roll a die 18,000 times and they observe 3,123 sixes. What is the probability this would occur if the die is not biased?

53. We suspect that a bridge player is cheating by putting an honor card (Ace, King, Queen, or Jack) at the bottom of the deck when he shuffles so that this card will end up in his hand. In 16 times when he dealt, the last card dealt to him was an honor on 9 occasions. Use the normal approximation to compute the probability that in 16 deals the last card will be an honor card at least 9 times.

54. If both parents carry one dominant (A) and one recessive gene (a) for a trait then Mendelian inheritance predicts that 1/4 of the offspring will have both recessive genes (aa) and show the recessive trait. What is the probability among 96 offspring of Aa parents we find 30 are aa

55. A softball player brags that he is a .300 hitter, yet at the end of the season he has gotten 21 hits in 84 at bats. What is the probability that he will have ≤ 21 hits if he gets a hit in each at bat with probability 0.3.

56. Suppose that we roll two dice 180 times and we are interested in the probability that we get exactly 5 double sixes. Find (a) the normal approximation, (b) the exact answer, (c) the Poisson approximation.

57. A gymnast has a difficult trick with a 10% chance of success. She tries the trick 25 times and wants to know the probability she will get exactly two successes. Compute the (a) exact answer, (b) Poisson approximation, (c) normal approximation.

58. A student is taking a true/false test with 48 questions. (a) Suppose she has a probability $p = 3/4$ of getting each question right. What is the probability she will get at least 38 right? (b) Answer the last question if she knows the answers to half the questions and flips a coin to answer the other half. Notice that in each case the expected number of questions she gets right is 36.

59. To estimate the percent of voters who oppose a certain ballot measure, a survey organization takes a random sample of 200 voters. If 45% of the voters oppose the measure, estimate the chance that (a) exactly 90 voters in the sample oppose the measure, (b) more than half the voters in the sample oppose the measure.

60. An airline knows that in the long run only 90% of passengers who book a seat show up for their flight. On a particular flight with 300 seats there are 324 reservations. Assuming passengers make independent decisions what is the chance that the flight will be over booked?

61. Suppose that 15% of people don't show up for a flight, and suppose that their decisions are independent. How many tickets can you sell for a plane with 144 seats and be 99% sure that not too many people will show up?

62. A seed manufacturer sells seeds in packets of 50. Assume that each seed germinates with probability 0.99 independently of all the others. The manufacturer promises to replace, at no cost to the buyer, any packet with 3 or more seeds that do not germinate. (a) Use the Poisson to estimate the probability a packet must be replaced. (b) Use the normal to estimate the probability that the manufacturer has to replace more than 70 of the last 4000 packets sold.

63. An electronics company produces devices that work properly 95% of the time. The new devices are shipped in boxes of 400. The company wants to guarantee that k or more devices per box work. What is the largest k so that at least 95% of the boxes meet the warranty?

CLT, III. General Distributions

64. The number of students who enroll in a psychology class is Poisson with mean 100. If the enrollment is > 120 then the class will be split into two sections. Estimate the probability that this will occur.

65. On each bet a gambler loses \$1 with probability 0.7, loses \$2 with probability 0.2, and wins \$10 with probability 0.1. Estimate the probability that the gambler will be losing after 100 bets.

66. Suppose we roll a die 10 times. What is the approximate probability that the sum of the numbers obtained lies between 30 and 40?

67. Members of the Beta Upsilon Zeta fraternity each drink a random number of beers with mean 6 and standard deviation 3. If there are 81 fraternity members, how much should they buy so that using the normal approximation they are 93.32% sure they will not run out?

68. An insurance company has 10,000 automobile policy holders. The expected yearly claim per policy holder is $240 with a standard deviation of $800. Approximate the probability that the yearly claim exceeds $2.7 million.

69. A die is rolled repeatedly until the sum of the numbers obtained is larger than 200. What is the probability that you need more than 66 rolls to do this?

70. Suppose that the checkout time at a grocery store has a mean of 5 minutes and a standard deviation of 2 minutes. Estimate the probability that a checker will serve at least 49 customers during her 4-hour shift.

71. A hot dog eating champion needs 10 seconds to eat one hot dog with a standard deviation of 3 seconds. What is the probability he can eat 65 hot dogs in 10 minutes.

Confidence Intervals

72. In the 2018 season Daniel Jones completed 237 out of 395 passes for a completion percentage of 0.605. Find a 95% confidence interval for his true completion percentaage.

73. Villanova's 95-51 victory over Oklahoma in the 2016 NCAA semi-finals, they made 35 out of 49 field goals (two point shots) or 71.43%. Find a 95% confidence interval for the probability they will make a field goal.

74. A bank examines the records of 150 patrons and finds that 63 have savings accounts. Find a 95% confidence interval for the fraction of people with savings accounts.

75. Among 625 randomly chosen Swedish citizens, it was found that 100 had previously been citizens of another country. Find a 95% confidence interval for the true proportion.

76. A sample of 2,809 hand-held video games revealed that 212 broke within the first three months of operation. Find a 95% confidence interval for the true proportion that break in the first three months.

77. Suppose we take a poll of 2,500 people. What percentage should the leader have for us to be 99% confident that the leader will be the winner?

78. For a class project, you are supposed to take a poll to forecast the outcome of an election. How many people do you have to ask so that with probability .95 your estimate will not differ from the true outcome by more than 5%?

Chapter 5

Conditional Probability

5.1 The multiplication rule

In Section 2.1., we introduced conditional probability primarily for the purpose of definitng independence. Here, we will study the concept in more deail. Firs,t we recall the definition. Suppose we are told that the event A with $P(A) > 0$ occurs. Then (i) only the part of B that lies in A can possibly occur, and (ii) since the sample space is now A, we have to divide by $P(A)$ to make $P(A|A) = 1$. Thus the probability that B will occur given that A has occurred is

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{5.1}$$

Multiplying the definition of conditional probability in (2.1) on each side by $P(A)$ gives the **multiplication rule**

$$P(A)P(B|A) = P(B \cap A) \tag{5.2}$$

Example 5.1. Alice and Bob are playing a gambling game. Each rolls one die and the person with the higher number wins. If they tie then they roll again. If Alice just won, what is the probability she rolled a 5?

Let $A =$ "Alice wins," and $R_i =$ "she rolls an i". If we write outcomes with Alice's roll first and Bob's second, the event A

(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
	(3,2)	(4,2)	(5,2)	(6,2)
		(4,3)	(5,3)	(6,3)
			(5,4)	(6,4)
				(6,5)

There are $1+2+3+4+5 = 15$ outcomes in A and if we condition on A they are all equally likely. $A \cap R_5$ has four outcomes, so $P(R_5|A) = 4/15$. In general, $P(R_i|A) = (i-1)/15$ for $1 \le i \le 6$.

As the last example may have suggested, the mapping $B \to P(B|A)$ is a probability. That is, it is a way of assigning numbers to events that satisfies the axioms introduced in Chapter 1. To prove this, we note that

(i) $0 \le P(B|A) \le 1$ since $0 \le P(B \cap A) \le P(A)$.

(ii) $P(\Omega|A) = P(\Omega \cap A)/P(A) = 1$

(iii) and (iv). If B_i are disjoint then $B_i \cap A$ are disjoint and $(\cup_i B_i) \cap A = \cup_i (B_i \cap A)$, so using the definition of conditional probability and parts (iii) and (iv) of the definition of probability we have

$$P(\cup_i B_i|A) = \frac{P(\cup_i (B_i \cap A))}{P(A)} = \frac{\sum_i P(B_i \cap A)}{P(A)} = \sum_i P(B_i|A)$$

From the last observation it follows that $P(\cdot|A)$ has the same properties that ordinary probabilities do, for example, if $C = B^c$

$$P(C|A) = 1 - P(B|A) \tag{5.3}$$

Actually for this to hold, it is enough that B and C complement each other inside A, i.e., $(B \cap C) \cap A = \emptyset$ and $(B \cup C) \supset A$.

Example 5.2. A woman has two children. If we are told that her family is in $A =$ at least one is a girl what is the probability the other one is a girl?

We assume that boys and girls are equally likely and the sexes of er two children are independent. $A = \{GG, GB, BG\}$ with all three outcomes equal so the answer is 2/3.

Example 5.3. Bridge. In the game of bridge there are four players called North, West, South, and East according to their positions at the table. Each player gets 13 cards. The game is somewhat complicated so we will content ourselves to analyze one situation that is important in the play of the game. Suppose that North and South have a total of eight Hearts. What is the probability that West will have 3 and East will have 2?

Even though this is not how the cards are usually dealt, we can imagine that West randomly draws 13 cards from the 26 that remain. This can be done in

$$C_{26,13} = \frac{26!}{13!\,13!} = 10,400,600 \text{ ways}$$

North and South have 8 hearts and 18 non-hearts so in the 26 cards that remain there are $13 - 8 = 5$ hearts and $39 - 18 = 21$ non-hearts. To construct a hand for West with 3 hearts and 10 non-hearts we must pick 3 of the 5 hearts, which can be done in $C_{5,3}$ ways and 10 of the 21 non-hearts in $C_{21,10}$. The multiplication rule then implies that the number of outcomes for West with 3 hearts is $C_{5,3} \cdot C_{21,10}$ and the probability of interest is

$$\frac{C_{5,3} \cdot C_{21,10}}{C_{26,13}} = 0.339$$

Multiplying by 2 gives the probability that one player will have 3 cards and the other 2, something called a $3 - 2$ split. Repeating the reasoning gives that an $i - j$ split $(i + j = 5)$ has probability

$$2 \cdot \frac{C_{5,i} \cdot C_{21,13-i}}{C_{26,13}}$$

This formula tells us that the probabilities are

3-2	0.678
4-1	0.282
5-0	0.039

Thus while a 3-2 split is the most common, one should not ignore the possibility of a 4-1 split.

Exercise Show that if North and South have 9 hearts then the probabilities are

2-2	0.406
3-1	0.497
4-0	0.095

In this case the uneven 3-1 split is more common than the 2-2 split since it can occur two ways, i.e., West might have 3 or 1.

Example 5.4. A person picks 13 cards out of a deck of 52. Let A_1 = "he has at least one Ace," H = "he has the Ace of hearts," and E_1 = "he receives exactly one Ace." Find $P(E_1|A_1)$ and $P(E_1|H)$. Do you think these will be equal? If not then which one is larger?

Let E_0 = "he has no Ace."

$$p_0 = P(E_0) = \frac{C_{48,13}}{C_{52,13}} \qquad p_1 = P(E_1) = \frac{4C_{48,12}}{C_{52,13}}$$

Since $E_1 \subset A_1$ and $A_1 = E_0^c$,

$$P(E_1|A_1) = \frac{P(E_1)}{P(A_1)} = \frac{p_1}{1-p_0}$$

Since $E_1 \cap H$ means you get the Ace of Hearts and no other ace

$$P(E_1|H) = \frac{P(E_1 \cap H)}{P(H)} = \frac{C_{48,12}/C_{52,13}}{1/4} = p_1$$

To compare the probabilities we observe

$$P(E_1|A_1) = \frac{p_1}{1-p_0} > p_1 = P(E_1|H)$$

Letting A_2 = "he has at least two Aces and using (5.3) we have

$$P(A_2|A_1) < P(A_2|H)$$

Intuitively, the event H is harder to achieve than A_1 so conditioning on it increases our chance of having other aces.

Conditional probabilities are the sources of many "paradoxes" in probability. One of these attracted worldwide attention in 1990 when Marilyn vos Savant discussed it in her weekly column in the Sunday *Parade* magazine.

Example 5.5. The Monty Hall problem. The problem is named for the host of the television show *Let's Make A Deal* in which contestants were often placed in situations like the following: Three curtains are numbered 1, 2, and 3. Behind one curtain is a car; behind the other two curtains are donkeys. You pick a curtain, say #1. To build some suspense the host opens up one of the two remaining curtains, say #3, to reveal a donkey. What is the probability you will win given that there is a donkey behind #3? Should you switch curtains and pick #2 if you are given the chance?

Many people argue that "the two unopened curtains are the same so they each will contain the car with probability 1/2, and hence there is no point in switching." As we will now show, this naive reasoning is incorrect. To compute the answer, we will suppose that the host always chooses to show you a donkey and picks at random if there are two unchosen curtains with donkeys. Assuming you pick curtain #1, there are three possibilities

	#1	#2	#3	host's action
case 1	donkey	donkey	car	opens #2
case 2	donkey	car	donkey	opens #3
case 3	car	donkey	donkey	opens #2 or #3

Now $P(\text{case 2}, \text{open door } \#3) = 1/3$ and

$$P(\text{case 3}, \text{open door } \#3) = P(\text{case 3})P(\text{open door } \#3|\text{case 3}) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

Adding the two ways door #3 can be opened gives $P(\text{open door } \#3) = 1/2$ and it follows that

$$P(\text{case 3}|\text{open door } \#3) = \frac{P(\text{case 3}, \text{open door } \#3)}{P(\text{open door } \#3)} = \frac{1/6}{1/2} = \frac{1}{3}$$

Although it took a number of steps to compute this answer, it is "obvious." When we picked one of the three doors initially we had probability 1/3 of picking the car, and since the host can always open a door with a donkey the new information does not change our chance of winning.

Example 5.6. Cognitive dissonance. An economist, M. Keith Chen, has recently uncovered a version of the Monty Hall problem in the theory of cognitive dissonance. For a half-century, experimenters have been using the so-called free choice paradigm to test our tendency to rationalize decisions. In an experiment typical of the genre, Yale psychologists measured monkey's preferences by observing how quickly each monkey sought out different colors of M&M's.

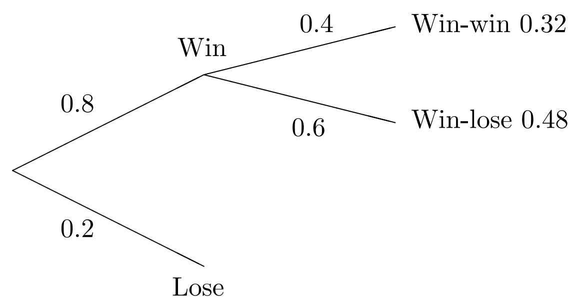
In the first step, the researchers gave the monkey a choice between say red and blue. If the monkey chose red, then it was given a choice between blue and green. Nearly two-thirds of the time it rejected blue in favor of green, which seemed to jibe with the theory of choice rationalization: once we reject something, we tell ourselves we never liked it anyway.

Putting aside this interpretation it is natural to ask: What would happen if monkeys were acting at random? If so, the six orderings RGB, RBG, GRB, GBR, BGR, and BRG would have equal probability. In the first three cases red is preferred to blue, but in 2/3's of those cases green is preferred to blue. Just as in the Monty Hall problem, we think that the probability of preferring blue to green is 1/2 due to symmetry, but the probability is 1/3. This time however conditioning on red being preferred to green reduced the original probability of 1/2 to 1/3, whereas in the Monty Hall problem the probability was initially 1/3 and did not change.

5.2 Two-Stage Experiments

We begin with several examples and then describe the collection of problems we will treat in this section.

Example 5.7. The Duke basketball team is playing in a four team tournament. In the first round they have any easy opponent that they will beat 80% of the time but if they win that game they will play against a tougher team where their probability of success is 0.4. What is the probability that they will win the tournament?

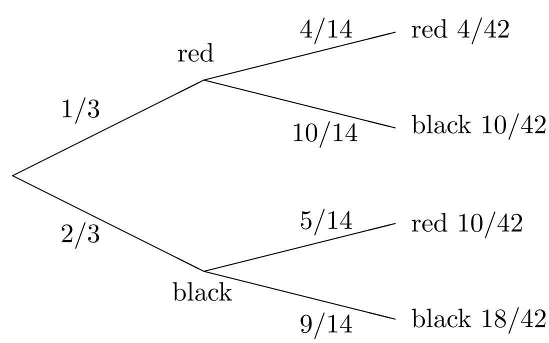


If A and B are the events of victory in the first and second games then $P(A) = 0.8$ and $P(B|A) = 0.4$, so the probability that they will win the tournament is

$$P(A \cap B) = P(A)P(B|A) = 0.8(0.4) = 0.32$$

Example 5.8. An urn contains 5 red and 10 black balls. We draw two balls from the urn without replacement. What is the probability that the second ball drawn is red?

This is easy to see if we draw a picture. The first split in the tree is based on the outcome of the first draw and the second on the outcome of the final. The outcome of the first draw dictates the probabilities for the second one. We multiply the probabilities on the edges to get probabilities of the four endpoints, and then sum the ones that correspond to Red to get the answer: $4/42 + 10/42 = 1/3$.



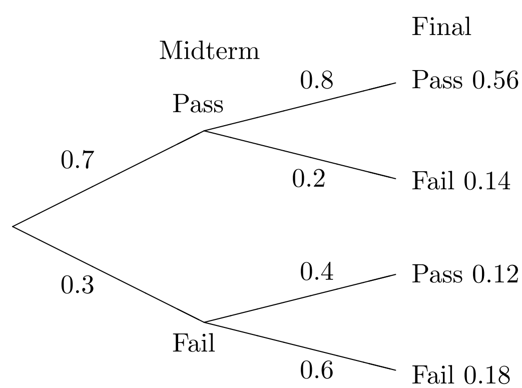
To do this with formulas, let R_i be the event of a red ball on the ith draw and let B_1 be the event of a black ball on the first draw. Breaking things down according to the outcome of the first draw, then using the multiplication rule, we have

$$\begin{aligned} P(R_2) &= P(R_2 \cap R_1) + P(R_2 \cap B_1) \\ &= P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1) \\ &= (1/3)(4/14) + (2/3)(5/14) = 14/42 = 1/3 \end{aligned}$$

From this we see that $P(R_2|R_1) < P(R_1) < P(R_2|B_1)$ but the two probabilities average to give $P(R_1)$. This calculation makes the result look like a miracle but it is not. If we number the 15 balls in the urn, then by symmetry each of them is equally likely to be the second ball chosen. Thus the probability of a red on the second, eighth, or fifteenth draw is always the same.

Example 5.9. Based on past experience, 70% of students in a certain course pass the midterm exam. The final exam is passed by 80% of those who passed the midterm, but only by 40% of those who fail the midterm. What fraction of students pass the final?

Drawing a tree as before with the first split based on the outcome of the midterm and the second on the outcome of the final, we get the answer: $0.56 + 0.12 = 0.68$



To do this with formulas, let A be the event that the student passes the final and let B be the event that the student passes the midterm. Breaking things down according to the outcome of the first test, then using the multiplication rule, we have

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B)P(B) + P(A|B^c)P(B^c) \\ &= (0.8)(0.7) + (0.4)(0.3) = 0.68 \end{aligned}$$

Example 5.10. Al flips 3 coins and Betty flips 2. Al wins if the number of Heads he gets is more than the number Betty gets. What is the probability Al will win?

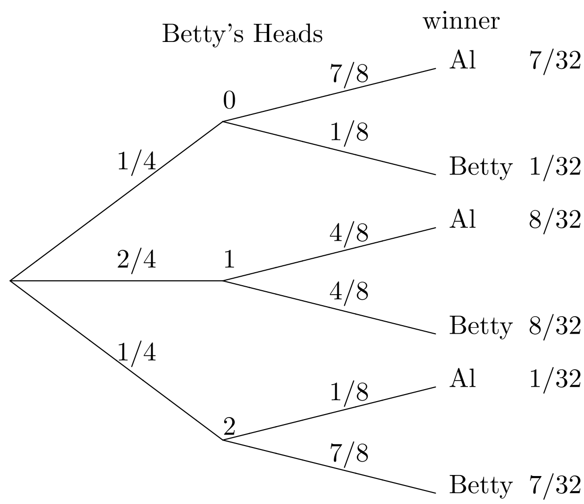
Let W be the event that Al wins. We will break things down according to the number of heads Betty gets. Let B_i be the event that Betty gets i Heads, and let A_j be the event that Al gets j Heads. By considering the four outcomes of flipping two coins it is easy to see that

$$P(B_0) = 1/4 \qquad P(B_1) = 1/2 \qquad P(B_2) = 1/4$$

while considering the eight outcomes for three coins leads to

$$\begin{aligned}
P(W|B_0) &= P(A_1 \cup A_2 \cup A_3) = 7/8 \\
P(W|B_1) &= P(A_2 \cup A_3) = 4/8 \\
P(W|B_2) &= P(A_3) = 1/8
\end{aligned}$$

This gives us the raw material for drawing our picture



Adding up the ways Al can win we get $7/32 + 8/32 + 1/32 = 1/2$. To check this draw a line through the middle of the picture and note the symmetry between top and bottom.

To do this with formulas, note that $W \cap B_i$, $i = 0, 1, 2$ are disjoint and their union is W, so

$$P(W) = \sum_{i=0}^{2} P(W \cap B_i) = \sum_{i=0}^{2} P(W|B_i)P(B_i)$$

since $P(W \cap B_i) = P(A|B_i)P(B_i)$ by the multiplication rule (5.2). Plugging in the values we computed,

$$P(W) = \frac{1}{4} \cdot \frac{7}{8} + \frac{2}{4} \cdot \frac{4}{8} + \frac{1}{4} \cdot \frac{1}{8} = \frac{7 + 8 + 1}{32} = \frac{1}{2}$$

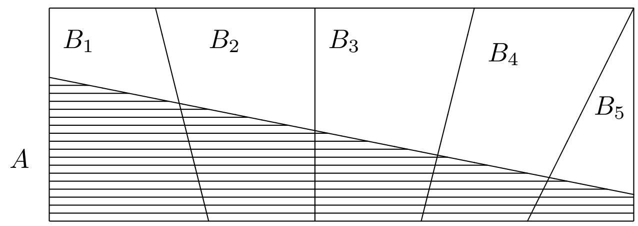
The previous analysis makes it look miraculous that we have a fair game. However it is true in general.

Example 5.11. Al flips $n + 1$ coins and Betty flips n. Al wins if the number of Heads he gets is more than the number Betty gets. What is the probability Al will win?

Consider the situation after Al has flipped n coins and Betty has flipped n. Using X and Y to denote the number of heads for Al and Betty at that time, there are the three possibilities $X > Y$, $X = Y$, $X < Y$. In the first case Al has already won. In the third he cannot win. In the second he wins with probability 1/2. Using symmetry if

$P(X > Y) = P(X < Y) = p$ then $P(X = Y) = 1 - 2p$, so the probability Al wins is $p + (1 - 2p)/2 = 1/2$.

Abstracting the structure of the last problem, let $B_1, \ldots, B_k$ be a **partition**, that is, a collection of disjoint events whose union is Ω.



Using the fact that the sets $A \cap B_i$ are disjoint, and the multiplication rule, we have

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i) \tag{5.4}$$

a formula that is sometimes called the **law of total probability**.

The name of this section comes from the fact that we think of our experiment as occurring in two stages. The first stage determines which of the B's occur, and when B_i occurs in the first stage, A occurs with probability $P(A|B_i)$ in the second. As the next example shows, the two stages are sometimes clearly visible in the problem itself.

Example 5.12. Roll a die and then flip that number of coins. What is the probability of $A =$ "we get exactly 3 Heads"?

Let $B_i =$ "the die shows i." $P(B_i) = 1/6$ for $i = 1, 2, \ldots, 6$ and

$$\begin{array}{lll} P(A|B_1) = 0 & P(A|B_2) = 0 & P(A|B_3) = 2^{-3} \\ P(A|B_4) = C_{4,3}\, 2^{-4} & P(A|B_5) = C_{5,3}\, 2^{-5} & P(A|B_6) = C_{6,3}\, 2^{-6} \end{array}$$

So plugging into (5.4),

$$\begin{aligned} P(A) &= \frac{1}{6}\left\{\frac{1}{8} + \frac{4}{16} + \frac{10}{32} + \frac{20}{64}\right\} \\ &= \frac{1}{6}\left\{\frac{8 + 16 + 20 + 20}{64}\right\} = \frac{1}{6} \end{aligned}$$

Example 5.13. Craps. In this game, if the sum of the two dice is 2, 3, or 12 on his first roll, the player loses; if the sum is 7 or 11, he wins; if the sum is 4, 5, 6, 8, 9, or 10, this number becomes his "point" and he wins if he "makes his point," i.e., his number comes up again before he throws a 7. What is the probability the player wins?

The first step in analyzing craps is to compute the probability that the player makes his point. Suppose his point is 5 and let E_k be the event that the sum is k. There are 4 outcomes in E_5 $((1,4), (2,3), (3,2), (4,1))$, 6 in E_7, and hence 26 not in $E_5 \cup E_7$. Letting $\times$ stand for "the sum is not 5 or 7," we see that

$$P(5) = \frac{4}{36} \quad P(\times\, 5) = \frac{26}{36} \cdot \frac{4}{36} \quad P(\times \times 5) = \left(\frac{26}{36}\right)^2 \frac{4}{36}$$

From the first three terms it is easy to see that for $k \geq 0$

$$P(\times \text{ on } k \text{ rolls then } 5) = \left(\frac{26}{36}\right)^k \frac{4}{36}$$

Summing over the possibilities, which represent disjoint ways of rolling 5 before 7, we have

$$P(5 \text{ before } 7) = \sum_{k=0}^{\infty} \left(\frac{26}{36}\right)^k \frac{4}{36} = \frac{4}{36} \cdot \frac{1}{1 - \frac{26}{36}}$$

since

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \tag{5.5}$$

Simplifying, we have $P(5 \text{ before } 7) = (4/36)/(10/36) = 4/10$. Such a simple answer should have a simple explanation, and it does. Consider an urn with four balls marked 5, six marked 7, and twenty-six marked with x. Drawing with replacement until we draw either a 5 or 7 is the same as drawing once from an urn with 10 balls with four balls marked 5 and six marked 7.

```
5 5 5 5 x x x x x x x x
7 7 7 7 7 7 x x x x x x
x x x x x x x x x x x x
```

Another way of saying this is that if we ignore the outcomes that result in a sum other than 5 or 7, we reduce the sample space from Ω to $E = E_5 \cup E_7$ and the distribution of the first outcome that lands in E follows the conditional probability $P(\cdot|E)$. Since $E_5 \cap E = E_5$ we have

$$P(E_5|E) = \frac{P(E_5)}{P(E)} = \frac{4/36}{10/36} = \frac{4}{10}$$

The last argument generalizes easily to give the probabilities of making any point:

k	4	5	6	8	9	10
$\lvert E_k \rvert$	3	4	5	5	4	3
P(k before 7)	3/9	4/10	5/11	5/11	4/10	3/9

To compute the probability of A = "he wins," we let B_k = "the first roll is k," and observe that (5.4) implies

$$P(A) = \sum_{k=2}^{12} P(A \cap B_k) = \sum_{k=2}^{12} P(B_k) P(A|B_k)$$

When $k = 2, 3$, or 12 comes up on the first roll we lose, so

$$P(A|B_k) = 0 \quad \text{and} \quad P(A \cap B_k) = 0$$

When $k = 7$ or 11 comes up on the first roll we win, so

$$P(A|B_k) = 1 \quad \text{and} \quad P(A \cap B_k) = P(B_k)$$

When the first roll is $k = 4, 5, 6, 8, 9$, or 10, $P(A|B_k) = P(k \text{ before } 7)$ and $P(A \cap B_k)$ is

$$\frac{3}{36} \cdot \frac{3}{9} \quad k = 4, 10 \qquad \frac{4}{36} \cdot \frac{4}{10} \quad k = 5, 9 \qquad \frac{5}{36} \cdot \frac{5}{11} \quad k = 6, 8$$

Adding up the terms in the sum in the order in which they were computed,

$$\begin{aligned} P(A) &= \frac{6}{36} + \frac{2}{36} + 2\left(\frac{1}{36} + \frac{4 \cdot 2}{36 \cdot 5} + \frac{5 \cdot 5}{36 \cdot 11}\right) \\ &= \frac{4}{18} + 2\left(\frac{55 + 88 + 125}{36 \cdot 11 \cdot 5}\right) = \frac{220 + 268}{18 \cdot 11 \cdot 5} = \frac{488}{990} = 0.4929 \end{aligned} \tag{5.6}$$

which is not very much less than $1/2 = 495/990$.

Example 5.14. Al and Bob take turns throwing one dart to try to hit a bullseye. Al hits with probability 1/4 while Bob hits with probability 1/3. If Al goes first what is the probability he will hit the first bullseye?

Let p be the answer. By considering one cycle of the game we see

$$p = 1/4 + (3/4)(1/3)(0) + (3/4)(2/3)p$$

In words, Al wins if he hits the bullseye on the first try. If he misses and Al hits then he loses. If they both miss then it is Al's turn and the game starts over, so Al's probability of success is p. Solving the equation we have $p/2 = 1/4$ or $p = 1/2$.

Back to craps. This reasoning in the last example can be used to compute the probability q that a player rolls a 5 before 7. By considering the outcome of the first roll $q = 4/36 + (6/36)0 + (26/36)q$ and solving we have $q = 4/10$.

5.3 Bayes' Formula

The title of the section is a little misleading since we will regard Bayes' formula as a method for computing conditional probabilities and will only reluctantly give the formula after we have done several examples to illustrate the method.

Example 5.15. Exit polls. In the California gubernatorial election in 1982, several TV stations predicted, on the basis of questioning people when they exited the polling place, that Tom Bradley, then mayor of Los Angeles, would win the election. When the votes were counted, however, he lost to George Deukmejian by a considerable margin. What caused the exit polls to be wrong?

To give our explanation we need some notation and some numbers. Suppose we choose a person at random, let $B =$ "the person votes for Bradley" and suppose that $P(B) = 0.45$. There were only two candidates, so this makes the probability of voting for Deukmejian $P(B^c) = 0.55$. Let $A =$ "the voter stops and answers a question about how they voted" and suppose that $P(A|B) = 0.4$, $P(A|B^c) = 0.3$. That is, 40% of Bradley voters will respond compared to 30% of the Deukmejian voters. We are interested in computing $P(B|A) =$ the fraction of voters in our sample that voted for Bradley. By the definition of conditional probability (5.1),

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A)}{P(B \cap A) + P(B^c \cap A)}$$

To evaluate the two probabilities, we use the multiplication rule (5.2)

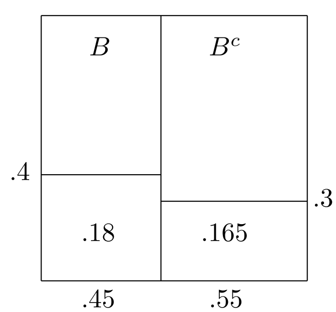
$$P(B \cap A) = P(B)P(A|B) = 0.45 \cdot 0.4 = 0.18$$
$$P(B^c \cap A) = P(B^c)P(A|B^c) = 0.55 \cdot 0.3 = 0.165$$

From this it follows that

$$P(B|A) = \frac{0.18}{0.18 + 0.165} = 0.5217$$

and from our sample it looks as if Bradley will win. The problem with the exit poll is that the difference in the response rates makes our sample not representative of the population as a whole.

Turning to the mechanics of the computation, note that 18% of the voters are for Bradley and respond, while 16.5% are for Deukmejian and respond, so the fraction of Bradley voters in our sample is $18/(18 + 16.5)$. In words, there are two ways an outcome can be in A – it can be in B or in B^c – and the conditional probability is the fraction of the total that comes from the first way.



Example 5.16. Mammogram posterior probabilities. Approximately 1% of women aged 40-50 have breast cancer. A woman with breast cancer has a 90% chance of a positive test from a mammogram, while a woman without has a 10% chance of a false positive result. What is the probability a woman has breast cancer given that she just had a positive test?

Let $B =$ "the woman has breast cancer" and $A =$ "a positive test." We want to calculate $P(B|A)$. Computing as in the previous example,

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A)}{P(B \cap A) + P(B^c \cap A)}$$

To evaluate the two probabilities, we use the multiplication rule (5.2)

$$P(B \cap A) = P(B)P(A|B) = 0.01 \cdot 0.9 = 0.009$$
$$P(B^c \cap A) = P(B^c)P(A|B^c) = 0.99 \cdot 0.1 = 0.099$$

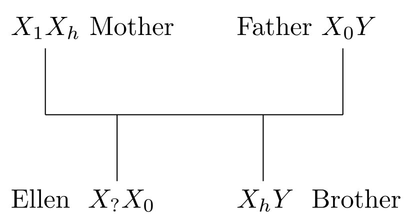
From this it follows that

$$P(B|A) = \frac{0.009}{0.009 + 0.099} = \frac{9}{108}$$

or a little less than 9%. This situation comes about because it is much easier to have a positive results from a false positive for a healthy woman which has probability 0.099, than from a woman with breast cancer having a positive test, which has probability 0.009.

This answer is somewhat surprising. Indeed, when ninety-five physicians were asked this question "What is the probability a woman has breast cancer given that she just had a positive test?" their average answer was 75%. The two statisticians who carried out this survey indicated that physicians were better able to see the answer when the data was presented in frequency format. 10 out of 1000 women have breast cancer. Of these 9 will have a positive mammogram. However, of the remaining 990 women without breast cancer 99 will have a positive test, and again we arrive at the answer $9/(9 + 99)$.

Example 5.17. Hemophilia. Ellen has a brother with hemophilia, but two parents who do not have the disease. Since hemophilia is caused by a recessive allele h on the X chromosome, we can infer that her mother is a carrier (that is, the mother has the hemophilia allele h on one of her X chromosomes and the healthy allele H on the other), while her father has the healthy allele on his one X chromosome. Since Ellen received one X chromosome from her father and one from her mother, there is a 50% chance that she is a carrier, and if so, there is a 50% chance that her sons will have the disease. If she has two sons without the disease, what is the probability she is a carrier?



Let B be the event that she is a carrier and A be the event that she has two healthy sons. Computing as in the two previous examples,

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A)}{P(B \cap A) + P(B^c \cap A)}$$

To evaluate the two probabilities we use the multiplication rule (5.2). Since the probability of having two healthy sons when she is a carrier is 1/4, and is 1 when she is not.

$$P(B \cap A) = P(B)P(A|B) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$$
$$P(B^c \cap A) = P(B^c)P(A|B^c) = \frac{1}{2} \cdot 1 = \frac{1}{2}$$

From this it follows that

$$P(B|A) = \frac{1/8}{1/8 + 1/2} = \frac{1}{5}$$

Example 5.18. Three factories make 20%, 30%, and 50% of the computer chips for a company. The probability of a defective chip is 0.04, 0.03, and 0.02 for the three factories. We have a defective chip. What is the probability it came from Factory 1?

Let B_i be the event that the chip came from factory i and let A be the event that the chip is defective. We want to compute $P(B_3|A)$. Adapting the computation from the two previous examples to the fact that there are now three B_i

$$P(B_3|A) = \frac{P(B_3 \cap A)}{P(A)} = \frac{P(B_3 \cap A)}{\sum_{i=1}^{3} P(B_i \cap A)}$$

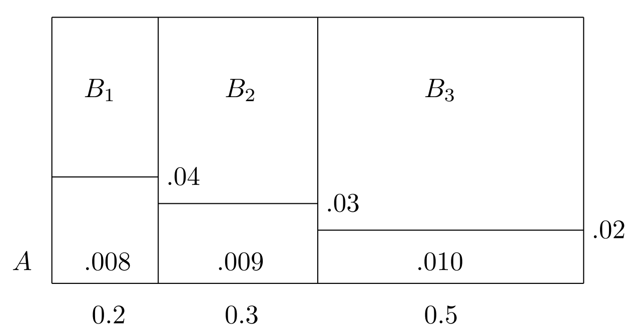
To evaluate the three probabilities, we use the multiplication rule (5.2)

$$\begin{aligned} P(B_1 \cap A) &= P(B_1)P(A|B_1) = 0.2 \cdot (0.04) = 0.008 \\ P(B_2 \cap A) &= P(B_2)P(A|B_2) = 0.3 \cdot (0.03) = 0.009 \\ P(B_3 \cap A) &= P(B_3)P(A|B_3) = 0.5 \cdot (0.02) = 0.010 \end{aligned}$$

From this it follows that

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{0.008}{0.008 + 0.009 + 0.010} = \frac{8}{27}$$

The calculation can be summarized by the following picture. The conditional probability $P(B_3|A)$ is the fraction of the event A that lies in B_3.



We are now ready to generalize from our examples and state **Bayes formula**. In each case, we have a **partition** of the probability space $B_1, \ldots, B_n$, i.e., a sequence of disjoint sets with $\cup_{i=1}^{n} B_i = \Omega$. (In the first three examples, $B_1 = B$ and $B_2 = B^c$.) We are given $P(B_i)$ and $P(A|B_i)$ for $1 \le i \le n$ and we want to compute $P(B_1|A)$. Reasoning as in the previous examples,

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(B_1 \cap A)}{\sum_i P(B_i \cap A)}$$

To evaluate the probabilities, we observe that

$$P(B_i \cap A) = P(B_i)P(A|B_i)$$

From this, it follows that

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(B_1)P(A|B_1)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)} \tag{5.7}$$

This is Bayes formula. Even though we have numbered it, we advise you not to memorize it. It is much better to remember the procedures we followed to compute the conditional probability.

5.4 Exercises

Conditional probability

1. There are two urns. Urn A has 4 red and 6 white balls, urn B has 7 red and 3 white balls. We pick an urn at random and pick a ball from it. Given that the ball is red what is the probability we picked from urn A. What is the answer if we pick two balls and they are both red?

2. An urn contains 8 red, 7 blue, and 5 green balls. You draw out two balls and they are different colors. Given this, what is the probability the two balls were red and blue?

3. Two people, whom we call South and North, draw 13 cards out of a deck of 52. South has two Aces. What is the probability that North has (a) none? (b) one? (c) the other two?

4. Suppose that the number of children in a family has the following distribution

number of children	0	1	2	3	4
probability	0.15	0.25	0.3	0.2	0.1

Assume that each child is independently a girl or a boy with probability 1/2 each. If a family is picked at random what is the chance it has exactly two girls.

5. A student is taking a multiple-choice test in which each question has four possible answers. She knows the answers to 50% of the questions, can narrow the choices down to two 30% of the time, and does not know anything about 20% of the questions. What is the probability she will correctly answer a question chosen at random from the test?

6. Plumber Bob does 40% of the plumbing jobs in a small town. 30% of the people in town are unhappy with their plumbers but 50% of Bob's customers are unhappy with his work. If your neighbor is not happy with his plumber, what is the probability it was Bob?

7. Suppose that the probability a married man votes is 0.45, the probability a married woman votes is 0.4, and the probability a woman votes given that her husband does is 0.6. What is the probability (a) both vote, (b) a man votes given that his wife does?

8. Suppose 60% of the people in a town will get exposed to flu in the next month. If you are exposed and not inoculated then the probability of your getting the flu is 80%, but if you are inoculated that probability drops to 15%. Of two executives at Beta Company, one is

inoculated and one is not. What is the probability at least one will not get the flu? Assume that the events that determine whether or not they get the flu are independent.

9. An ectopic preganancy is twice as likely if a woman smokes cigarettes. If 25% of women of childbearing age are smokers, what fraction of ectopic preganancies occur to smokers?

10. Brown eyes are dominant over blue. That is, there are two alleles B and b. bb individuals have blue eyes but other combinations has brown eyes. Your parents and you have brown eyes but your brother has blue. So you can infer that both of your parents are heterozygotes, i.e., have genetic type Bb. Given this information what is the probability you are a homozygote.

Two-stage Experiments

11. From a signpost that says MIAMI two letters fall off. A friendly drunk puts the two letters back into the two empty slots at random. What is the probability that the sign still says MIAMI?

12. Two balls are drawn from an urn with balls numbered from 1 up to 10. What is the probability that the two numbers will differ by at least three. (≥ 3) than three?

13. You and a friend each roll two dice. What is the probability you will both have the same two numbers?

14. Charlie draws five cards out of a deck of 52. If he gets at least three of one suit, he discards the cards not of that suit and then draws until he again has five cards. For example, if he gets three hearts, one club, and one spade, he throws the two nonhearts away and draws two more. What is the probability he will end up with five cards of the same suit?

15. John takes the bus with probability 0.3 and the subway with probability 0.7. He is late 40% of the time when he takes the bus but only 20% of the time when he takes the subway. What is the probability he is late for work?

16. The population of Cyprus is 70% Greek and 30% Turkish. 20% of the Greeks and 10% of the Turks speak English. What fraction of the people of Cyprus speak English?

17. You are going to meet a friend at the airport. Your experience tells you that the plane is late 70% of the time when it rains, but is late only 20% of the time when it does not rain. The weather forecast that morning calls for a 40% chance of rain. What is the probability the plane will be late?

18. Two boys have identical piggy banks. The older boy has 18 quarters and 12 dimes in his; the younger boy, 2 quarters and 8 dimes. One day the two banks get mixed up. You pick up a bank at random and shake it until a coin comes out. What is the probability you get a quarter? Note that there are 20 quarters and 20 dimes in all.

19. Two boys, Charlie and Doug, take turns rolling two dice with Charlie going first. If Charlie rolls a 6 before Doug rolls a 7 he wins. What is the probability Charlie wins?

20. Three boys take turns shooting a basketball and have probabilities 0.2, 0.3, and 0.5 of scoring a basket. Compute the probabilities for each boy to get the first basket.

21. Change the second and third probabilities in the last problem so that each boy has an equal chance of winning.

22. In the town of Carboro 90% of students graduate high school, 60% of high school graduates complete college, and 20% of college graduates get graduate or professional degrees. What fraction of students get advanced degrees?

23. What is the probability of a flush, i.e., all cards of the same suit when we draw 5 cards out of a deck of 52?

Bayes' Formula

24. Three bags lie on the table. One has two gold coins, one has two silver coins, and one has one silver and one gold. You pick a bag at random, and pick out one coin. If this coin is gold, what is the probability you picked from the bag with two gold coins?

25. 1 out of 1000 births results in fraternal twins; 1 out of 1500 births results in identical twins. Identical twins must be the same sex but the sexes of fraternal twins are independent. If two girls are twins, what is the probability they are fraternal twins?

26. Five pennies are sitting on a table. One is a trick coin that has Heads on both sides, but the other four are normal. You pick up a penny at random and flip it four times, getting Heads each time. Given this, what is the probability you picked up the two-headed penny?

27. 5% of men and 0.25% of women are color blind. Assuming that there are an equal number of men and women, what is the probability a color blind person is a man?

28. The alpha fetal protein test is meant to detect spina bifida in unborn babies, a condition that affects 1 out of 1000 children who are born. The literature on the test indicates that 5% of the time a healthy baby will cause a positive reaction. We will assume that the test is positive 100% of the time when spina bifida is present. Your doctor has just told you that your alpha fetal protein test was positive. What is the probability that your baby has spina bifida?

29. Binary digits, i.e., 0's and 1's, are sent down a noisy communications channel. They are received as sent with probability 0.9 but errors occur with probability 0.1. Assuming that 0's and 1's are equally likely, what is the probability that a 1 was sent given that we received a 1?

30. To improve the reliability of the channel described in the last example, we repeat each digit in the message three times. What is the probability that 111 was sent given that (a) we received 101? (b) we received 000?

31. A student goes to class on a snowy day with probability 0.4, but on a nonsnowy day attends with probability 0.7. Suppose that 20% of the days in February are snowy. What is the probability it snowed on February 7th given that the student was in class on that day?

32. A company gave a test to 100 salesman, 80 with good sales records and 20 with poor sales records. 60% of the good salesman passed the test but only 30% of the poor salesmen did. Andy passed the test. Given this, what is the probability that he is a good salesman?

33. A company rates 80% of its employees as satisfactory and 20% as unsatisfactory. Personnel records indicate that 70% of the satisfactory workers had prior experience but only 40% of the unsatisfactory workers did. If a person with previous work experience is hired, what is the probability they will be a satisfactory worker?

34. A golfer hits his drive in the fairway with probability 0.7. When he hits his drive in the fairway he makes par 80% of the time. When he doesn't he makes par only 30% of the time. He just made par on a hole. What is the probability he hit his drive in the fairway?

35. You are about to have an interview for Harvard Law School. 60% of the interviewers are conservative and 40% are liberal. 50% of the conservatives smoke cigars but only 25% of the liberals do. Your interviewer lights up a cigar. What is the probability he is a liberal?

36. One slot machine pays off 1/2 of the time, while another pays off 1/4 of the time. We pick one of the machines and play it six times, winning 3 times. What is the probability we are playing the machine that pays off only 1/4 of the time?

37. 20% of people are “accident-prone” and have a probability 0.15 of having an accident in a one-year period in contrast to a probability of 0.05 for the other 80% of people. (a) If we pick a person at random, what is the probability they will have an accident this year? (b) What is the probability a person is accident-prone if they had an accident last year? (c) What is the probability they will have an accident this year if they had one last year?

38. One die has 4 red and 2 white sides; a second has 2 red and 4 white sides. (a) If we pick a die at random and roll it, what is the probability the result is a red side? (b) If the first result is a red side and we roll the same die again, what is the probability of a second red side?

39. A particular football team is known to run 40% of its plays to the left and 60% to the right. When the play goes to the right, the right tackle shifts his stance 80% of the time, but does so only 10% of the time when the play goes to the left. As the team sets up for the play the right tackle shifts his stance. What is the probability that the play will go to the right?

40. A professor gives a test to 90 calculus students, 60 who have had calculus before and 30 who have not. 30% of the students in the first group got an A, but only 10% in the second group did. Barbara got an A. What is the probability she had calculus before?

41. You are a serious student who studies on Friday nights but your roommate goes out and has a good time. 40% of the time he goes out with his girlfriend; 60% of the time he goes to a bar. 30% of the times when he goes out with his girlfriend he spends the night at her apartment. 40% of the times when he goes to a bar he gets in a fight and gets thrown in jail. You wake up on Saturday morning and your roomate is not home. What is the probability he is in jail?

42. In a certain city 30% of the people are Conservatives, 50% are Liberals, and 20% are Independents. In a given election, 2/3 of the Conservatives voted, 80% of the Liberals voted, and 50% of the Independents voted. If we pick a voter at random what is the probability they are Liberal?

43. A group of 20 people go out to dinner. 10 go to an Italian restaurant, 6 to a Japanese restaurant, and 4 to a French restaurant. The fractions of people satisfied with their meals were 0.8, 2/3, and 1/2 respectively. The next day the person you are talking to was satified with what they ate. What is the probability they went to the Italian restuarant? the Japanese restaurant?, the French restaurant?

44. Consider the following data on traffic accidents

age group	% of drivers	accident probability
16 to 25	15	.10
26 to 45	35	.04
46 to 65	35	.06
over 65	15	.08

Calculate (a) the probability a randomly chosen driver will have an accident this year, and (b) the probability a driver is between 46 and 65 given that they had an accident.

Chapter 6

Markov Chains

6.1 Definitions and Examples

The importance of Markov chains comes from two facts: (i) there are a large number of physical, biological, economic, and social phenomena that can be described in this way, and (ii) there is a well-developed theory that allows us to understand their behavior. We begin with a famous example, then describe the property that is the defining feature of Markov chains. Throughout this chapter we will only consider Markov chains with a finite state space.

Example 6.1. Gambler's ruin. Consider a gambling game in which on any turn you win \$1 with probability $p = 0.4$ or lose \$1 with probability $1-p = 0.6$. Suppose further that you adopt the rule that you quit playing if your fortune reaches \$$N$. Of course, if your fortune reaches \$0 the casino makes you stop.

Let X_n be the amount of money you have after n plays. I claim that your fortune, X_n has the "Markov property." In words, this means that given the current state, any other information about the past is irrelevant for predicting the next state X_{n+1}. To check this, we note that if you are still playing at time n, i.e., your fortune $X_n = i$ with $0 < i < N$, then for any possible history of your wealth $i_{n-1}, i_{n-2}, \dots i_1, i_0$

$$P(X_{n+1} = i+1 | X_n = i, X_{n-1} = i_{n-1}, \dots X_0 = i_0) = 0.4$$

since to increase your wealth by one unit you have to win your next bet and the outcome of the previous bets has no useful information for predicting the next outcome.

Turning now to the formal definition, we say that X_n is a discrete time **Markov chain** with **transition matrix** $p(i,j)$ if for any $j, i, i_{n-1}, \dots i_0$

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p(i,j) \tag{6.1}$$

Equation (6.1), also called the "Markov property" says that the conditional probability $X_{n+1} = j$ given the entire history $X_n = i, X_{n-1} = i_{n-1}, \dots X_1 = i_1, X_0 = i_0$ is the same as the conditional probability $X_{n+1} = j$ given only the previous state $X_n = i$.

In formulating (6.1) we have restricted our attention to the **temporally homogeneous** case in which the **transition probability**

$$p(i,j) = P(X_{n+1} = j | X_n = i)$$

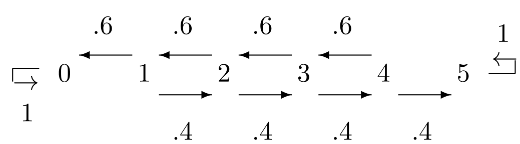
does not depend on the time n. Intuitively, the transition probability gives the rules of the game. It is the basic information needed to describe a Markov chain. In the case of the gambler's ruin chain, the transition probability has

$$p(i, i+1) = 0.4, \quad p(i, i-1) = 0.6, \quad \text{if } 0 < i < N$$
$$p(0,0) = 1 \qquad p(N,N) = 1$$

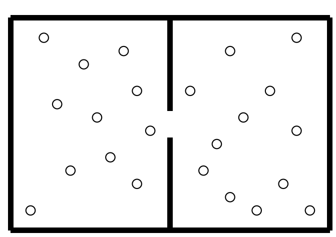
When $N = 5$ the matrix is

	0	**1**	**2**	**3**	**4**	**5**
0	1.0	0	0	0	0	0
1	0.6	0	0.4	0	0	0
2	0	0.6	0	0.4	0	0
3	0	0	0.6	0	0.4	0
4	0	0	0	0.6	0	0.4
5	0	0	0	0	0	1.0

but picture is perhaps more informative:



Example 6.2. Ehrenfest chain. We imagine two cubical volumes of air connected by a small hole. In the mathematical version, we have two urns in which there are a total of N balls. At time n we pick one of the N balls at random and move it to the other urn to create the state at time $n + 1$.



Let X_n be the number of balls in the left urn at time n. It should be clear that X_n has the Markov property; i.e., if we want to guess the state at time $n+1$, then the current number of balls in the left urn X_n, is the only relevant information from the observed sequence of states $X_n, X_{n-1}, \ldots X_1, X_0$. To check this we note that

$$P(X_{n+1} = i + 1 | X_n = i, X_{n-1} = i_{n-1}, \ldots X_0 = i_0) = \frac{N - i}{N}$$

since to increase the number we have to pick one of the $N-i$ balls in the other urn. The number can also decrease by 1 with probability i/N. Thus the transition probability is given by

$$p(i,i+1) = \frac{N-i}{N}, \quad p(i,i-1) = \frac{i}{N} \qquad \text{for } 0 \le i \le N$$

with $p(i,j) = 0$ otherwise. When $N = 5$, for example, the matrix is

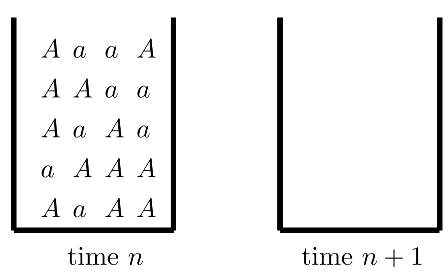
	0	**1**	**2**	**3**	**4**	**5**
0	0	5/5	0	0	0	0
1	1/5	0	4/5	0	0	0
2	0	2/5	0	3/5	0	0
3	0	0	3/5	0	2/5	0
4	0	0	0	4/5	0	1/5
5	0	0	0	0	5/5	0

Here we have written 1 as 5/5 to emphasize the pattern in the diagonals of the matrix.

Example 6.3. Wright–Fisher model. We consider a fixed population of N genes that can be one of two types: A or a. These types are called alleles. In the simplest version of this model the population at time $n+1$ is obtained by drawing with replacement from the population at time n. In this case if we let X_n be the number of A alleles at time n, then X_n is a Markov chain with transition probability

$$p(i,j) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j} \qquad 0 \le i,j \le N$$

since the right-hand side is the binomial distribution for N independent trials with success probability i/N. Note that when $i = 0$, $p(0,0) = 1$, and when $i = N$, $p(N,N) = 1$.



In the Gambler's Ruin chain and the Wright-Fisher model the states 0 and N are **absorbing states**. Once we enter these states we can never leave. The long run behavior of these models is not very interesting, they will eventually enter one of the absorbing states and stay there forever. To make the Wright-Fisher model more interesting and more realistic, we can introduce the possibility of mutations: an A that is drawn ends up being an a in the next generation with probability u, while an a that is drawn ends up being an A in the next generation with probability v. In this case the probability an A is produced by a given draw is

$$\rho_i = \frac{i}{N}(1-u) + \frac{N-i}{N}v$$

i.e., we can get an A by drawing an A and not having a mutation or by drawing an a and having a mutation. Since the draws are independent the transition probability still has the binomial form

$$p(i,j) = \binom{N}{j} (\rho_i)^j (1-\rho_i)^{N-j} \tag{6.2}$$

Example 6.4. Two-stage Markov chains. In a Markov chain the distribution of X_{n+1} only depends on X_n. This can easily be generalized to case in which the distribution of X_{n+1} only depends on (X_n, X_{n-1}). For a concrete example consider a basketball player who makes a shot with the following probabilities:

1/2 if he has missed the last two times

2/3 if he has hit one of his last two shots

3/4 if he has hit both of his last two shots

To formulate a Markov chain to model his shooting, we let the states of the process be the outcomes of his last two shots: $\{HH, HM, MH, MM\}$ where M is short for miss and H for hit. Noting that the first letter is X_n and the second is X_{n-1} the transition probability is

	HH	**HM**	**MH**	**MM**
HH	3/4	0	1/4	0
HM	2/3	0	1/3	0
MH	0	2/3	0	1/3
MM	0	1/2	0	1/2

To explain suppose the state is HM, i.e., $X_n = H$ and $X_{n-1} = M$. In this case the next outcome will be H with probability 2/3. When this occurs the next state will be $(X_{n+1}, X_n) = (H, H)$. If he misses, an event of probability 1/3, $(X_{n+1}, X_n) = (M, H)$. To check the placement of 0's note that since the first letter gets shifted to the right, if the first letter is H then after a jump the second letter will be H.

The Hot Hand is a phenomenon known to everyone who plays or watches basketball. After making a couple of shots, players are thought to "get into a groove" so that subsequent successes are more likely. Purvis Short of the Golden State Warriors describes this more poetically as "You're in a world all your own. It's hard to describe. But the basket seems to be so wide. No matter what you do, you know the ball is going to go in."

Unfortunately for basketball players, data collected by Tversky and Gliovich (*Chance* vol. 2 (1989), No. 1, pages 16–21) shows that this is a misconception. The next table gives data for the conditional probability of hitting a shot after missing the last three, missing the last two, ... hitting the last three, for nine players of the Philadelphia 76ers: Darryl Dawkins (403), Maurice Cheeks (339), Steve Mix (351), Bobby Jones (433), Clint Richardson (248), Julius Erving (884), Andrew Toney (451), Caldwell Jones (272), and Lionel Hollins (419). The numbers in parentheses are the number of shots for each player.

$P(H\|3M)$	$P(H\|2M)$	$P(H\|1M)$	$P(H\|1H)$	$P(H\|2H)$	$P(H\|3H)$
.88	.73	.71	.57	.58	.51
.77	.60	.60	.55	.54	.59
.70	.56	.52	.51	.48	.36
.61	.58	.58	.53	.47	.53
.52	.51	.51	.53	.52	.48
.50	.47	.56	.49	.50	.48
.50	.48	.47	.45	.43	.27
.52	.53	.51	.43	.40	.34
.50	.49	.46	.46	.46	.32

In fact, the data supports the opposite assertion: after missing a player is more likely to hit a shot

The first four example are meant to motivate the study. We will now introduce some simple numerical examples that will be useful as we develop the theory. Markov chains are described by giving their transition probabilities. To create a chain, we can write down any $n \times n$ matrix, provided that the entries satisfy:

(i) $p(i,j) \geq 0$, since they are probabilities.

(ii) $\sum_j p(i,j) = 1$, since when $X_n = i$, X_{n+1} will be in some state j.

The equation in (ii) is read "sum $p(i,j)$ over all possible values of j." In words the last two conditions say: the entries of the matrix are nonnegative and each **row** of the matrix sums to 1.

Any matrix with properties (i) and (ii) gives rise to a Markov chain, X_n. To construct the chain we can think of playing a board game. When we are in state i, we roll a die (or generate a random number on a computer) to pick the next state, going to j with probability $p(i,j)$. To illustrate this we will now introduce some simple examples.

Example 6.5. Weather chain. Let X_n be the weather on day n, which we assume is either: 1 = *rainy*, or 2 = *sunny*. Even though the weather is not exactly a Markov chain, we can propose a Markov chain model for the weather by writing down a transition probability

	1	**2**
1	.6	.4
2	.2	.8

The table says, for example, the probability a rainy day (state 1) is followed by a sunny day (state 2) is $p(1,2) = 0.4$.

Example 6.6. Social mobility. Let X_n be a family's social class in the nth generation, which we assume is either 1 = *lower*, 2 = *middle*, or 3 = *upper*. In our simple version of sociology, changes of status are a Markov chain with the following transition probability

	1	**2**	**3**
1	.7	.2	.1
2	.3	.5	.2
3	.2	.4	.4

Example 6.7. Brand preference. Suppose there are three types of laundry detergent, 1, 2, and 3, and let X_n be the brand chosen on the nth purchase. Customers who try these

brands are satisfied and choose the same thing again with probabilities 0.8, 0.6, and 0.4 respectively. When they change they pick one of the other two brands at random. The transition probability is

	1	**2**	**3**
1	.8	.1	.1
2	.2	.6	.2
3	.3	.3	.4

6.2 Multistep Transition Probabilities

The previous section introduced several examples for us to think about. The basic question concerning Markov chains is: what happens in the long run? In the case of the weather chain, that question is: does the probability that day n is sunny converge to a limit? In the case of the social mobility chain: do the fractions of the population in the three income classes stabilize as time goes on? The first step in answering these questions is to figure out what happens in the Markov chain after two or more steps.

The transition probability $p(i,j) = P(X_{n+1} = j|X_n = i)$ gives the probability of going from i to j in one step. Our goal in this section is to compute the probability of going from i to j in $m > 1$ steps:

$$p^m(i,j) = P(X_{n+m} = j|X_n = i)$$

For a concrete example, we start with the transition probability of the social mobility chain:

	1	**2**	**3**
1	.7	.2	.1
2	.3	.5	.2
3	.2	.4	.4

To warm-up we consider:

Example 6.8. Suppose the family starts in the middle class (state 2) in generation 0. What is the probability that the generation 1 rises to the upper class (state 3) and generation 2 falls to the lower class (state 1)?

Intuitively, the Markov property implies that starting from state 2 the probability of jumping to 1 and then to 3 is given by $p(2,3)p(3,1)$. To get this conclusion from the definitions, we note that using the definition of conditional probability,

$$P(X_2 = 1, X_1 = 3|X_0 = 2) = \frac{P(X_2 = 1, X_1 = 3, X_0 = 2)}{P(X_0 = 2)}$$

Multiplying and dividing by $P(X_1 = 3, X_0 = 2)$:

$$= \frac{P(X_2 = 1, X_1 = 3, X_0 = 2)}{P(X_1 = 3, X_0 = 2)} \cdot \frac{P(X_1 = 3, X_0 = 2)}{P(X_0 = 2)}$$

Using the definition of conditional probability:

$$= P(X_2 = 1|X_1 = 3, X_0 = 2) \cdot P(X_1 = 3|X_0 = 2)$$

By the Markov property (6.1) the last expression is

$$P(X_2 = 1|X_1 = 3) \cdot P(X_1 = 3|X_0 = 2) = p(2,3)p(3,1) \qquad \square$$

Moving on to the real question:

Example 6.9. Suppose the family starts in the middle class (state 2) in generation 0. What is the probability that generation 2 will be in the lower class (state 1)?

To do this we simply have to consider the three possible states for generation 1 and use the previous computation.

$$P(X_2 = 1|X_0 = 2) = \sum_{k=1}^{3} P(X_2 = 1, X_1 = k|X_0 = 2) = \sum_{k=1}^{3} p(2,k)p(k,1)$$
$$= (.3)(.7) + (.5)(.3) + (.2)(.2) = .21 + .15 + .04 = .40$$

There is nothing special here about the states 2 and 1 here. By the same reasoning,

$$P(X_2 = j|X_0 = i) = \sum_{k=1}^{3} p(i,k)\, p(k,j)$$

The right-hand side of the last equation gives the (i,j)th entry of the matrix p is multiplied by itself.

To explain this, we note that to compute $p^2(2,1)$ we multiplied the entries of the second row by those in the first column:

$$\begin{pmatrix} . & . & . \\ .3 & .5 & .2 \\ . & . & . \end{pmatrix} \begin{pmatrix} .7 & . & . \\ .3 & . & . \\ .2 & . & . \end{pmatrix} = \begin{pmatrix} . & . & . \\ .40 & . & . \\ . & . & . \end{pmatrix}$$

If we wanted $p^2(1,3)$ we would multiply the first row by the third column:

$$\begin{pmatrix} .7 & .2 & .1 \\ . & . & . \\ . & . & . \end{pmatrix} \begin{pmatrix} . & . & .1 \\ . & . & .2 \\ . & . & .4 \end{pmatrix} = \begin{pmatrix} . & . & .15 \\ . & . & . \\ . & . & . \end{pmatrix}$$

When all of the computations are done we have

$$\begin{pmatrix} .7 & .2 & .1 \\ .3 & .5 & .2 \\ .2 & .4 & .4 \end{pmatrix} \begin{pmatrix} .7 & .2 & .1 \\ .3 & .5 & .2 \\ .2 & .4 & .4 \end{pmatrix} = \begin{pmatrix} .57 & .28 & .15 \\ .40 & .39 & .21 \\ .34 & .40 & .26 \end{pmatrix}$$

The two step transition probability $p^2 = p \cdot p$. Based on this you can probably leap to the next conclusion:

Theorem 6.1. *The m-step transition probability*

$$p^m(i,j) = P(X_{n+m} = j|X_n = i) \tag{6.3}$$

is the mth power of the transition matrix p, i.e., $p \cdot p \cdots p$, where there are m terms in the product.

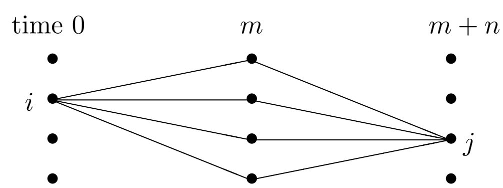
The key ingredient in proving this is the:

Chapman–Kolmogorov equation

$$p^{m+n}(i,j) = \sum_k p^m(i,k)\, p^n(k,j) \tag{6.4}$$

Once this is proved, (6.3) follows, since taking $n = 1$ in (6.4), we see that $p^{m+1} = p^m \cdot p$.

Why is (6.4) true? To go from i to j in $m + n$ steps, we have to go from i to some state k in m steps and then from k to j in n steps. The Markov property implies that the two parts of our journey are independent. □



Proof of (6.4). The *independence* in the second sentence of the previous explanation is the mysterious part. To show this, we combine Examples 6.8 and 6.9. Breaking things down according to the state at time m,

$$P(X_{m+n} = j | X_0 = i) = \sum_k P(X_{m+n} = j, X_m = k | X_0 = i)$$

Repeating the computation in Example 6.8, the definition of conditional probability implies:

$$P(X_{m+n} = j, X_m = k | X_0 = i) = \frac{P(X_{m+n} = j, X_m = k, X_0 = i)}{P(X_0 = i)}$$

Multiplying and dividing by $P(X_m = k, X_0 = i)$ gives:

$$= \frac{P(X_{m+n} = j, X_m = k, X_0 = i)}{P(X_m = k, X_0 = i)} \cdot \frac{P(X_m = k, X_0 = i)}{P(X_0 = i)}$$

Using the definition of conditional probability we have:

$$= P(X_{m+n} = j | X_m = k, X_0 = i) \cdot P(X_m = k | X_0 = i)$$

By the Markov property (6.1) the last expression is

$$= P(X_{m+n} = j | X_m = k) \cdot P(X_m = k | X_0 = i) = p^m(i,k)p^n(k,j)$$

and we have proved (6.4). □

Having established (6.4), we now return to computations. We begin with the weather chain

$$p^2 = \begin{pmatrix} .6 & .4 \\ .2 & .8 \end{pmatrix} \begin{pmatrix} .6 & .4 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} .44 & .56 \\ .28 & .72 \end{pmatrix}$$

Multiplying again

$$p^3 = p^2 \cdot p \begin{pmatrix} .44 & .56 \\ .28 & .72 \end{pmatrix} \begin{pmatrix} .6 & .4 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} .376 & .624 \\ .312 & .688 \end{pmatrix}$$

Taking higher powers

$$p^{10} = \begin{pmatrix} .333403 & .666597 \\ .333298 & .666702 \end{pmatrix} \qquad p^{20} = \begin{pmatrix} .3333333407 & .6666666597 \\ .3333333297 & .6666666703 \end{pmatrix}$$

Based on the last calculation, one might guess that as n gets large the matrix becomes closer and closer to

$$\begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}$$

This is true and will be explained in the next section.

6.3 Stationary Distributions

Our first step is to consider

What happens when the initial state is random?

Breaking things down according to the value of the initial state and using the definition of conditional probability

$$\begin{aligned} P(X_n = j) &= \sum_i P(X_0 = i, X_n = j) \\ &= \sum_i P(X_0 = i) P(X_n = j | X_0 = i) \end{aligned}$$

If we introduce $q(i) = P(X_0 = i)$, then the last equation can be written as

$$P(X_n = j) = \sum_i q(i) p^n(i,j) \tag{6.5}$$

In words, we multiply the transition matrix on the left by the vector q of initial probabilities. If there are k states, then $p^n(x,y)$ is a $k \times k$ matrix. So to make the matrix multiplication work out right, we should take q as a $1 \times k$ matrix or a "row vector."

For a concrete example consider the weather chain (Example 6.5) and suppose that the initial distribution is $q(1) = 0.3$ and $q(2) = 0.7$. In this case

$$\begin{pmatrix} .3 & .7 \end{pmatrix} \begin{pmatrix} .6 & .4 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} .32 & .68 \end{pmatrix}$$

since

$$\begin{aligned} .3(.6) + .7(.2) &= .32 \\ .3(.4) + .7(.8) &= .68 \end{aligned}$$

For a second example consider the social mobility chain (Example 6.6) and suppose that the initial distribution: $q(1) = .5$, $q(2) = .2$, and $q(3) = .3$. Multiplying the vector q by the transition probability gives the vector of probabilities at time 1.

$$\begin{pmatrix} .5 & .2 & .3 \end{pmatrix} \begin{pmatrix} .7 & .2 & .1 \\ .3 & .5 & .2 \\ .2 & .4 & .4 \end{pmatrix} = \begin{pmatrix} .47 & .32 & .21 \end{pmatrix}$$

To check the arithmetic note that the three entries on the right-hand side are

$$\begin{aligned} .5(.7) + .2(.3) + .3(.2) &= .47 \\ .5(.2) + .2(.5) + .3(.4) &= .32 \\ .5(.1) + .2(.2) + .3(.4) &= .21 \end{aligned}$$

If the distribution at time 0 is the same as the distribution at time 1, then by the Markov property it will be the distribution at all times $n \geq 1$. Because of this q is called a **stationary distribution**. Stationary distributions have a special importance in the theory of Markov chains, so we will use a special letter π to denote solutions of the equation

$$\pi \cdot p = \pi.$$

with $\pi_i \geq 0$ and $\sum_j \pi_j = 1$.

To have a mental picture of what happens to the distribution of probability when one step of the Markov chain is taken, it is useful to think that we have $q(i)$ pounds of sand at state i, with the total amount of sand $\sum_i q(i)$ being one pound. When a step is taken in the Markov chain, a fraction $p(i,j)$ of the sand at i is moved to j. The distribution of sand when this has been done is

$$q \cdot p = \sum_i q(i) p(i,j)$$

If the distribution of sand is not changed by this procedure q is a stationary distribution.

Example 6.10. General two state transition probability.

	1	**2**
1	$1-a$	a
2	b	$1-b$

We have written the chain in this way so the stationary distribution has a simple formula

$$\pi(1) = \frac{b}{a+b} \qquad \pi(2) = \frac{a}{a+b} \tag{6.6}$$

As a first check on this formula we note that in the weather chain $a = 0.4$ and $b = 0.2$ which gives $(1/3, 2/3)$ as we found before. We can prove this works in general by drawing a picture:

$$\frac{b}{a+b} \ \overset{\mathbf{1}}{\bullet} \ \underset{b}{\overset{a}{\rightleftarrows}} \ \overset{\mathbf{2}}{\bullet} \ \frac{a}{a+b}$$

In words, the amount of sand that flows from 1 to 2 is the same as the amount that flows from 2 to 1 so the amount of sand at each site stays constant. To check algebraically that this is true:

$$\begin{aligned}\frac{b}{a+b}(1-a)+\frac{a}{a+b}b &= \frac{b-ba+ab}{a+b} = \frac{b}{a+b}\\ \frac{b}{a+b}a+\frac{a}{a+b}(1-b) &= \frac{ba+a-ab}{a+b} = \frac{a}{a+b}\end{aligned} \tag{6.7}$$

Formula (6.6) gives the stationary distribution for any two state chain, so we progress now to the three state case and consider the brand preference chain (Example 6.7). The equation $\pi p = \pi$ says

$$\begin{pmatrix}\pi_1 & \pi_2 & \pi_3\end{pmatrix}\begin{pmatrix}.8 & .1 & .1\\ .2 & .6 & .2\\ .3 & .3 & .4\end{pmatrix} = \begin{pmatrix}\pi_1 & \pi_2 & \pi_3\end{pmatrix}$$

which translates into three equations

$$\begin{aligned}.8\pi_1 + .2\pi_2 + .3\pi_3 &= \pi_1\\ .1\pi_1 + .6\pi_2 + .3\pi_3 &= \pi_2\\ .1\pi_1 + .2\pi_2 + .4\pi_3 &= \pi_3\end{aligned}$$

Note that the columns of the matrix give the numbers in the rows of the equations. The third equation is redundant since if we add up the three equations we get

$$\pi_1 + \pi_2 + \pi_3 = \pi_1 + \pi_2 + \pi_3$$

If we replace the third equation by $\pi_1 + \pi_2 + \pi_3 = 1$ and subtract π_1 from each side of the first equation and π_2 from each side of the second equation we get

$$\begin{aligned}-.2\pi_1 + .2\pi_2 + .3\pi_3 &= 0\\ .1\pi_1 - .4\pi_2 + .3\pi_3 &= 0\\ \pi_1 + \pi_2 + \pi_3 &= 1\end{aligned} \tag{6.8}$$

At this point we can solve the equations by hand or using a calcualtor.

By hand. We note that the third equation implies $\pi_3 = 1 - \pi_1 - \pi_2$ and substituting this in the first two gives

$$\begin{aligned}.3 &= .5\pi_1 + .1\pi_2\\ .3 &= .2\pi_1 + .7\pi_2\end{aligned}$$

Multiplying the first equation by .7 and adding $-.1$ times the second gives

$$1.8 = (0.35 - 0.02)\pi_1 \qquad \text{or} \qquad \pi_1 = 18/33 = 6/11$$

Multiplying the first equation by .2 and adding $-.5$ times the second gives

$$-0.09 = (0.02 - 0.35)\pi_2 \qquad \text{or} \qquad \pi_2 = 9/33 = 3/11$$

Since the three probabilities add up to 1, $\pi_3 = 2/11$.

Using the TI83 calculator is easier. To begin we write (6.8) in matrix form as

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \begin{pmatrix} -.2 & .1 & 1 \\ .2 & -.4 & 1 \\ .3 & .3 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$$

If we let A be the 3×3 matrix in the middle this can be written as $\pi A = (0, 0, 1)$. To solve the equation we use the inverse matrix A^{-1} which is the unique matrix which has $AA^{-1} = I$, where I is the identity matrix that has 1's on the diagonal and 0's otherwise. Multiplying on each side by A^{-1} we see that

$$\pi = (0, 0, 1)A^{-1}$$

which is the third row of A^{-1}.

We do not need to know linear algebra to do this because we can compute A^{-1} by using our TI-83 calcular. We enter A into our calculator (using the MATRX menu and the EDIT submenu. We then use the MATRX menu to put $[A]$ on the computation line and then press the x^{-1} button. Reading the third row we find that the stationary distribution is

$$(0.545454, 0.272727, 0.181818)$$

Converting the answer to fractions using the first entry in the MATH menu gives

$$(6/11, 3/11, 2/11)$$

Example 6.11. Social Mobility (continuation of Example 6.6).

	1	**2**	**3**
1	.7	.2	.1
2	.3	.5	.2
3	.2	.4	.4

Using the first two equations and the fact that the sum of the π's is 1

$$\begin{aligned} .7\pi_1 + .3\pi_2 + .2\pi_3 &= \pi_1 \\ .2\pi_1 + .5\pi_2 + .4\pi_3 &= \pi_2 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

Taking the π's from the right hand side to the left we have

$$\begin{aligned} -0.3\pi_1 + .3\pi_2 + .2\pi_3 &= 0 \\ .2\pi_1 + -0.5\pi_2 + .4\pi_3 &= 0 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}$$

This translates into $\pi A = (0, 0, 1)$ with

$$A = \begin{pmatrix} -.3 & .2 & 1 \\ .3 & -.5 & 1 \\ .2 & .4 & 1 \end{pmatrix}$$

Note that here and in the previous example, A can be obtained from p by

- Subtracting 1 from the diagonal
- then replacing the final column

The first step corresponds to moving π_1 and π_2 from the right-hand side to the left. The second to replacing the third equation by $\pi_1 + \pi_2 + \pi_3 = 1$.

Computing the inverse and reading the last row gives

$$(0.468085, 0.340425, 0.191489)$$

Converting the answer to fractions using the first entry in the math menu gives

$$(22/47, 16/47, 9/47)$$

Example 6.12. Ehrenfest chain (continuation of Example 6.2). Consider first the case $N = 5$. As in the three previous examples, A consists of the first five columns of the transition matrix with 1 subtracted from the diagonal, and a final column of all 1's.

$$\begin{array}{cccccc} -1 & 1 & 0 & 0 & 0 & 1 \\ .2 & -1 & .8 & 0 & 0 & 1 \\ 0 & .4 & -1 & .6 & 0 & 1 \\ 0 & 0 & .6 & -1 & .4 & 1 \\ 0 & 0 & 0 & .8 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{array}$$

The answer is given by the sixth row of A^{-1}:

$$(0.03125, 0.15625, 0.3125, 0.3125, 0.15625, 0.03125)$$

Even without a calculator we can recognize these as

$$(1/32, 5/32, 10/32, 10/32, 5/32, 1/32)$$

the probabilities of 0 to 5 heads when we flip five coins.

Based on this we can guess that in general

$$\pi(k) = \binom{N}{k} / 2^N$$

Proof by thinking. To determine the initial state, (a) flip N coins, with heads = in the left urn, and tails = in the right. A transition of the chain corresponds to (b) picking a coin at random and turning it over. It is clear that the end result of (a) and (b) has all 2^N outcomes equally likely, so the state at time 1 is the same as the state at time 0. □

Proof by computation. For $0 < k < N$ we can end up in state k only by coming up from $k-1$ or down from $k+1$ so

$$\begin{aligned} &\pi(k-1)p(k-1,k) + \pi(k+1)p(k+1,k) \\ &= \frac{\binom{N}{k-1}}{2^N} \cdot \frac{N-k+1}{N} + \frac{\binom{N}{k+1}}{2^N} \cdot \frac{k+1}{N} \\ &= \frac{1}{2^N} \left(\frac{(N-1)!}{(k-1)!(N-k)!} + \frac{(N-1)!}{(k)!(N-k+1)!} \right) \\ &= \frac{1}{2^N} \binom{N}{k} \left(\frac{k}{N} + \frac{N-k}{N} \right) = \pi(k) \end{aligned}$$

The only way to end up at 0 is by coming down from 1 so

$$\pi(1)p(1,0) = \frac{N}{2^N} \cdot \frac{1}{N} = \pi(0)$$

Similarly, the only way to end up at N is by coming up from $N-1$ so

$$\pi(N-1)p(N-1,N) = \frac{N}{2^N} \cdot \frac{1}{N} = \pi(0)$$ □

Example 6.13. Basketball (continuation of Example 6.4). To find the stationary distribution in this case we can follow the same procedure. A consists of the first three columns of the transition matrix with 1 subtracted from the diagonal, and a final column of all 1's.

$-1/4$	0	$1/4$	1
$2/3$	-1	$1/3$	1
0	$2/3$	-1	1
0	$1/2$	0	1

The answer is given by the fourth row of A^{-1}:

$$(0.5, 0.1875, 0.1875, 0.125) = (1/2, 3/16, 3/16, 1/8)$$

Thus the long run fraction of time the player hits a shot is

$$\pi(HH) + \pi(MH) = 11/36 = 0.6875.$$

Notice that $\pi(MH) = \pi(HM)$. To see why this is true look at the following simulation

$$H\ H\ M\ M\ M\ H\ M\ M\ H\ H\ H\ M\ H\ M\ M$$

In this string the state is HM four times and MH three times, since we switch from H to M four times and from M to H three times. In the long run the number of switches in te two directions will either be the same or differ by 1, so the limiting frequencies of HM and MH must be the same.

6.4 Limit Behavior

A stationary distribution is an equilibrium state. That is, if we start in that distribution we stay in it for all time. In this section we will give conditions that guarantee that as $n \to \infty$

$$p^n(i,j) \to \pi(j) \quad \text{as } n \to \infty,$$

where π is a stationary distribution. This says that after a long time the probability of being in state j is $\approx \pi(j)$ independent of the starting point. It turns out that this holds if we eliminate two problems. The first problem is illustrated by the following

Example 6.14. Reducible chain.

	0	**1**	**2**	**3**
0	0.2	0.8	0	0
1	0.4	0.6	0	0
2	0	0	0.4	0.6
3	0	0	0.2	0.8

Using our formula for the stationary distribution of two state chains we see that $(1/3, 2/3, 0, 0)$ is a stationary distribution and so is $(0, 0, 1/4, 3/4)$. If we take the 10th power of the transition matrix then to six decimal places the matrix is

	0	**1**	**2**	**3**
0	1/3	2/3	0	0
1	1/3	2/3	0	0
2	0	0	1/4	3/4
3	0	0	1/4	3/4

The problem with this example is that is not really a four-state chain but two two-state chains in one matrix. To rule out examples like this, we need a condition.

Definition. A Markov chain is **irreducible** if for any x and y there is k so that $p^k(x, y) > 0$.

In words for any states x and y it is possible to get from x to y. In the example above it is impossible to get from 1 or 2 to 3 or 4, and vice versa. A nice consequence of irreducibility is:

Theorem 6.2. *If p is irreducible then there is a unique stationary distribution $\pi(x)$ and we have $\pi(y) > 0$ for all y.*

Our next step is to consider the two state case.

Theorem 6.3. *Let p_n be the probability of being in state 1 after n steps. For a two state Markov chain with transition probability*

$$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

we have

$$\left| p_n - \frac{b}{a+b} \right| = \left| p_0 - \frac{b}{a+b} \right| \cdot |1 - a - b|^n \tag{6.9}$$

If $0 < a + b < 2$ this implies that $p_n \to b/(a+b)$ exponentially fast.

In the case of the weather chain $|1 - a - b| = 0.4$, so the difference between p_n and the limit $b/(a+b)$ goes to 0 faster than $(0.4)^n$.

Proof. Using the Markov property we have for any $n \geq 1$ that

$$p_n = p_{n-1}(1-a) + (1 - p_{n-1})b$$

In words, the chain is in state 1 at time n if it was in state 1 at time $n-1$ (with probability p_{n-1}) and stays there (with probability $1-a$), or if it was in state 2 (with probability $1-p_{n-1}$) and jumps from 2 to 1 (with probability b). Since the probability of being in state 1 is constant when we start in the stationary distribution, see the first equation in (6.7):

$$\frac{b}{a+b} = \frac{b}{a+b}(1-a) + \left(1 - \frac{b}{a+b}\right) b$$

Subtracting this equation from the one for p_n we have

$$\begin{aligned} p_n - \frac{b}{a+b} &= \left(p_{n-1} - \frac{b}{a+b}\right)(1-a) + \left(\frac{b}{a+b} - p_{n-1}\right) b \\ &= \left(p_{n-1} - \frac{b}{a+b}\right)(1-a-b) \end{aligned}$$

If $0 < a + b < 2$ then $|1 - a - b| < 1$ and we have

$$\left| p_n - \frac{b}{a+b} \right| = \left| p_{n-1} - \frac{b}{a+b} \right| \cdot |1 - a - b|$$

Iterating the last equation gives the desired result. □

There are two cases $a = b = 0$ and $a = b = 1$ in which $|1 - a - b| = 1$ and hence $p^n(i, j)$ does not converge to $\pi(i)$. In the first case the matrix is

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

so the state never changes. In this case the chain is not irreducible. In the second case the matrix is

$$p = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

so the chain always jumps. In this case $p^2 = I$, $p^3 = p$, $p^4 = I$, etc. To see that something similar can happen in a "real example," we consider

Example 6.15. Ehrenfest chain. Consider the chain defined in Example 6.2 and for simplicty, suppose there are three balls. In this case the transition probability is

	0	**1**	**2**	**3**
0	0	3/3	0	0
1	1/3	0	2/3	0
2	0	2/3	0	1/3
3	0	0	3/3	0

In the second power of p the zero pattern is shifted:

	0	**1**	**2**	**3**
0	1/3	0	2/3	0
1	0	7/9	0	2/9
2	2/9	0	7/9	0
3	0	2/3	0	1/3

To see that the zeros will persist, note that if initially we have an odd number of balls in the left urn, then no matter whether we add or subtract one ball the result will be an even number. Thus X_n alternates between being odd and even. To see why this prevents convergence note that $p^{2n+1}(i, i) = 0$, so we cannot have $p^m(i, i) \to \pi(i)$, which is positive.

The problem with the Ehrenfest chain is called **periodicity**. The definition of the period of a state is a little complicated:

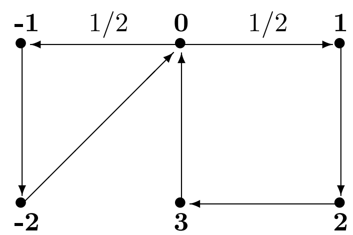
Definition. The period of a state i, $d(i)$ is the greatest common divisor of $I_i = \{n \geq 1$ that have $p^n(i, i) > 0\}$ is 1.

The next example explains why the definition is formulated in terms of the greatest common divisor.

Example 6.16. Triangle and Square. The state space is $\{-2,-1,0,1,2,3\}$ and the transition probability is

	−2	**−1**	**0**	**1**	**2**	**3**
−2	0	0	1	0	0	0
−1	1	0	0	0	0	0
0	0	1/2	0	1/2	0	0
1	0	0	0	0	1	0
2	0	0	0	0	0	1
3	0	0	1	0	0	0

In words, from 0 we are equally likely to go to 1 or -1. From -1 we go with probability one to -2 and then back to 0, from 1 we go to 2 then to 3 and back to 0. The name refers to the fact that $0 \to -1 \to -2 \to 0$ is a triangle and $0 \to 1 \to 2 \to 3 \to 0$ is a square.



Noting that $3, 4 \in I_0$, we see that the period is 1. The next result shows that once we find the period of one state we know the periods of all the states since

Lemma 6.1. *If p is irreducible then all states have the same period.*

If all states have period 1 then we call the chain **aperiodic**. From the last result we see that

Lemma 6.2. *If p is irreducible and $p(x,x) > 0$ then x has period 1, and hence by the previous lemma all states have period 1.*

With the key definitions made we can now state the

Convergence Theorem. *If p is irreducible and aperiodic then there is a unique stationary distribution π and for any i and j*

$$p^n(i,j) \to \pi(j) \quad \text{as } n \to \infty \tag{6.10}$$

An easy, but important, special case is if $p(i,j) > 0$ for all i and j then the chain is irreducible and aperiodic. This shows that the convergence theorem applies to the Wright-Fisher model with mutation, weather chain, social mobility, and brand preference chains that are Examples 6.3, 6.5, 6.6, and 6.7. The Convergence Theorem does not apply to the Gambler's Ruin chain (Example 6.1) or the Wright-Fisher model with no mutations since they have absorbing states and hence are not irreducible. We have already noted that the Ehrenfest chain (Example 6.2) does not converge since all states have period 2. The remaining example from Section 7.1.

Example 6.17. Basketball chain.

	HH	**HM**	**MH**	**MM**
HH	3/4	0	1/4	0
HM	2/3	0	1/3	0
MH	0	2/3	0	1/3
MM	0	1/2	0	1/2

The second power of the matrix is

	HH	**HM**	**MH**	**MM**
HH	9/16	1/6	3/16	1/12
HM	1/2	2/9	1/6	1/9
MH	4/5	1/6	2/9	1/6
MM	1/3	1/4	1/6	1/4

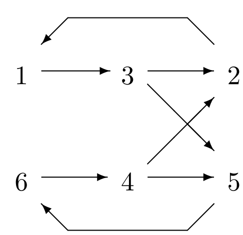
so the chain is irreducible. To check aperiodicity we note that $p(HH, HH) > 0$ and use Lemma 6.2.

Example 6.18. Period three. For another example that does not converge consider 4. Consider the following transition matrix:

	1	**2**	**3**	**4**	**5**	**6**
1	0	0	1	0	0	0
2	1	0	0	0	0	0
3	0	.5	0	0	.5	0
4	0	.5	0	0	.5	0.
5	0	0	0	0	0	1
6	0	0	0	1	0	0

(a) Find the stationary distribution. (b) Does the chain converge to it?

The stationary distribution is $\pi = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$ which is easily verified by checking that $\pi p = \pi$. If we draw a picture of the possible transitions of the chain we see that if we are in $\{1, 6\}$ at time 0 then we will be in $\{3, 4\}$ at time 1, in $\{2, 5\}$ at time 2, back in $\{1, 6\}$ at time 3, etc., so all states have period three.



6.5 Chains with absorbing states

Recall that x is an absorbing state if if $p(x, x) = 1$. When a chain has one or more of these states, the asymptotic behavior is not interesting. It eventually gets stuck in one of them. The only two interesting questions are: where does it get stuck? how long does it take?

6.5.1 Exit distribution

We begin with a simple example.

Example 6.19. Two year college. At a local two year college, 60% of freshmen become sophomores, 25% remain freshmen, and 15% drop out. 70% of sophomores graduate, 20% remain sophomores and 10% drop out. What fraction of new students eventually graduate?

We use a Markov chain with state space 1 = freshman, 2 = sophomore, G = graduate, D = dropout. The transition probability is

	1	**2**	**G**	**D**
1	0.25	0.6	0	0.15
2	0	0.2	0.7	0.1
G	0	0	1	0
D	0	0	0	1

Let $h(x)$ be the probability that a student currently in state x eventually graduates. By considering what happens on one step

$$h(1) = 0.25h(1) + 0.6h(2)$$
$$h(2) = 0.2h(2) + 0.7$$

so $h(2) = 0.7/0.8 = 0.875$ and
$h(1) = \frac{0.6}{0.75}h(2) = 0.7$. To check these answers we will look at the sixth power powers of the transition probability

$p^6 =$

	1	**2**	**G**	**D**
1	0.000244	0.002161	0.697937	0.283875
2	0	0.000064	0.874944	0.124492
G	0	0	1	0
D	0	0	0	1

Note that $p^6(1, D) \approx 0.7$ and $p^6(2, D) \approx 0.875$. The probability a student is still in college after six years is only $0.000244 + 0.002161 = 2.405 \times 10^{-3}$.

Example 6.20. Tennis. In tennis the winner of a game is the first player to win four points, unless the score is $4 - 3$, in which case the game must continue until one player wins by two points. Suppose that the game has reached the point where each player is trying to get two points ahead to win and that the server will independently win the point with probability 0.6. What is the probability the server will win the game if the score is tied? if she is ahead by one point? Behind by one point?

We formulate the game as a Markov chain in which the state is the difference of the scores. The state space is $2, 1, 0, -1, -2$ with 2 (win for server) and -2 (win for opponent). The transition probability is

	2	**1**	**0**	**-1**	**-2**
2	1	0	0	0	0
1	.6	0	.4	0	0
0	0	.6	0	.4	0
-1	0	0	.6	0	.4
-2	0	0	0	0	1

If we let $h(x)$ be the probability of the server eventually winning when the score is x then $h(2) = 1$, $h(-2) = 0$, and

$$\begin{aligned} h(1) &= 0.6 + 0.4h(0) \\ h(0) &= 0.6h(1) + 0.4h(-1) \\ h(-1) &= 0.6h(0) \end{aligned}$$

Rearranging we have

$$\begin{aligned} h(1) - 0.4h(0) + 0 \cdot h(-1) &= 0.6 \\ -0.6h(1) + h(0) - 0.4h(-1) &= 0 \\ 0 \cdot h(1) - 0.6h(0) + h(-1) &= 0 \end{aligned}$$

or in matrix form

$$\begin{pmatrix} 1 & -0.4 & 0 \\ -0.6 & 1 & -0.4 \\ 0 & -0.6 & 1 \end{pmatrix} \begin{pmatrix} h(1) \\ h(0) \\ h(-1) \end{pmatrix} = \begin{pmatrix} 0.6 \\ 0 \\ 0 \end{pmatrix}$$

If we let r be the rows and columns of transition matrix corresponding to the non-absorbing states $1, 0, -1$, then the matrix on the left-hand side is $(I - r)$ and the answer we want is

$$\begin{pmatrix} h(1) \\ h(0) \\ h(-1) \end{pmatrix} = (I - r)^{-1} \begin{pmatrix} 0.6 \\ 0 \\ 0 \end{pmatrix}$$

To prepare for later appliications note that the column vector is $p(i, 2)$, i.e., the probability of jumping to the absorbing state of interest. Inverting $(I - r)$ we have:

$$\begin{pmatrix} 19/13 & 10/13 & 4/13 \\ 15/13 & 25/13 & 10/13 \\ 9/13 & 15/13 & 19/13 \end{pmatrix} \begin{pmatrix} 0.6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 114/130 \\ 90/130 \\ 54/130 \end{pmatrix}$$

General solution. Suppose that the server wins each point with probability p. If the game is tied then after two points, the server will have won with probability p^2, lost with probability $(1 - p)^2$, and the game will have returned to being with probability $2p(1 - p)$, so $h(0) = p^2 + 2p(1 - p)h(0)$. Since $1 - 2p(1 - p) = p^2 + (1 - p)^2$, solving gives

$$h(0) = \frac{p^2}{p^2 + (1 - p)^2}$$

Figure 6.1 gives the probability of winning a tied game as a function of the probability of winning a point.

Example 6.21. Two year college example, revisited. We will now solve the problem using the machinery developed in the previous example. In this case

$$(I - r) = \begin{pmatrix} 0.75 & -0.6 \\ 0 & 0.8 \end{pmatrix} \qquad p(i, G) = \begin{pmatrix} 0 \\ 0.7 \end{pmatrix}$$

so the solution is

$$\begin{pmatrix} h(1) \\ h(2) \end{pmatrix} = (I - r)^{-1} \begin{pmatrix} 0 \\ 0.7 \end{pmatrix} = \begin{pmatrix} 4/3 & 1 \\ 0 & 5/4 \end{pmatrix} \begin{pmatrix} 0 \\ 0.7 \end{pmatrix} = \begin{pmatrix} 0.7 \\ 7/8 \end{pmatrix}$$

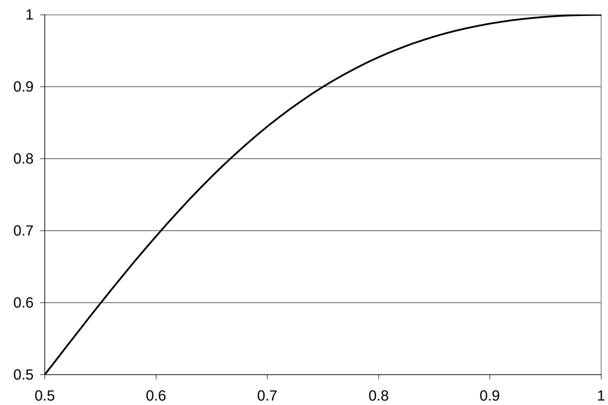


Figure 6.1: Probability of winning in tennis as a function of the probability of winning a point

Example 6.22. Wright-Fisher model. As described in Example 6.3, if we let X_n be the number of A alleles at time n, then X_n is a Markov chain with transition probability

$$p(x,y) = \binom{N}{y} \left(\frac{x}{N}\right)^y \left(1 - \frac{x}{N}\right)^{N-y}$$

0 and N are absorbing states. What is the probability the chain ends up in N starting from x?

As in the two previous examples if $h(x)$ is the probability of getting absorbed in state N when we start in state x then $h(0) = 0$, $h(N) = 1$, and for $0 < x < N$

$$h(x) = \sum_y p(x,y)h(y) \tag{6.11}$$

In the Wright-Fisher model $p(x,y)$ is the $P(\text{binomial}(N, x/N) = y)$, so the expected number of A's after the transition is x, i.e., the expected number of A's remains constant in time. Taking a leap of faith we assert that the average number of A's at the end is the same as the number in the starting state. Since we will have N with probability $h(y)$ and 0 with probability $1 - h(y)$ we must have

$$h(y) = \frac{y}{N}$$

Clearly, this function has $h(0) = 0$ and $h(N) = 1$. To check (6.11) we note that

$$\sum_y p(x,y)\frac{y}{N} = \frac{x}{N}$$

since the mean of the binomial is x.

The formula $h(y) = y/N$ says that if we start with y A's in the population then the probability we will end with a population of all A's (an event called "fixation" in genetics) is y/N, the fraction of the population that is A. The case $y = 1$ gives a famous result due to Kimura. If we suppose that each individual experiences mutations at rate μ, then since there are N individuals, new mutations occur at a total rate $N\mu$. Since each mutation achieves fixation with probability $1/N$, the rate at which mutations become fixed is μ independent of the size of population.

6.5.2 Exit Times

Returning to our first two examples:

Example 6.23. Two year college. Consider the transition probability in Example 6.19. How many years on the average does it take for a freshman to graduate or dropout?

Let $g(x)$ be the expected number of years to finish starting from x. $g(G) = g(D) = 0$. By considering what happens on the first step

$$\begin{aligned} g(1) &= 1 + 0.25g(1) + 0.6g(2) \\ g(2) &= 1 + 0.2g(2) \end{aligned}$$

Here the 1 takes into account the one year that elapsed during the transition. We can solve the equations to get

$$g(2) = \frac{1}{0.8} = \frac{5}{4} \qquad g(1) = \frac{1 + 0.6(5/4)}{3/4} = \frac{7}{3}$$

so, on the average, a freshman takes $7/3 = 2.33$ years to either graduate or drop out.

A second approach is to rearrange our initial equation to get

$$\begin{aligned} 0.75g(1) - 0.6g(2) &= 1 \\ 0 \cdot g(1) + 0.8g(2) &= 1 \end{aligned}$$

The matrix on the left-hand side is the one we saw in Example 6.21, so the solution is

$$\begin{pmatrix} g(1) \\ g(2) \end{pmatrix} = (I - r)^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Computing the inverse the answer is

$$\begin{pmatrix} 4/3 & 1 \\ 0 & 5/4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7/3 \\ 5/4 \end{pmatrix}$$

To explain why this is the answer note that

$$(I - r)(I + r + r^2 + \cdots) = (I + r + r^2 + \cdots) - (r + r^2 + r^3 + \cdots) = I$$

so we have $(I - r)^{-1} = I + r + r^2 + \cdots$. $r^n(i, j)$ is the probability of going from i to j in n steps so we have the

Theorem 6.4. $(I - r)^{-1}(i, j)$ *is the expected number of visits to* j *starting from* i*, including the visit at time 0 if* $j = i$.

When we multiply $(I - r)^{-1}$ by a column vector of all 1's we add the entries in each row, and end up with the expected time in the state space for each starting point.

Example 6.24. Tennis. Consider the transition probability in Example 6.20. Suppose that game is tied. How many more points do we expect to see before the game ends?

Let $g(x)$ be the expected number of points until the game ends when the score is x. Clearly, $g(2) = g(-2) = 0$. Writing equations as in the previous example

$$\begin{aligned} g(1) &= 1 + 0.4g(0) \\ g(0) &= 1 + 0.6g(1) + 0.4g(0) \\ g(-1) &= 1 + 0.6g(0) \end{aligned}$$

Rearranging we have

$$\begin{aligned} g(1) - 0.4g(0) + 0 \cdot g(-1) &= 1 \\ -0.6g(1) + g(0) - 0.4g(-1) &= 1 \\ 0 \cdot g(1) - 0.6g(0) + g(-1) &= 1 \end{aligned}$$

As in the previous example we can get to these equations by removing rows and columns for the absorbing states from the transition probability, to define

$$r = \begin{pmatrix} 0 & .4 & 0 \\ .6 & 0 & .4 \\ 0 & .6 & 0 \end{pmatrix} \quad \text{so that} \quad I - r = \begin{pmatrix} 1 & -0.4 & 0 \\ -0.6 & 1 & -0.4 \\ 0 & -0.6 & 1 \end{pmatrix}$$

and the equaitons are $(I - r)g = \mathbf{1}$ where $\mathbf{1}$ is a column vector with three 1's.

$$g = (I - r)^{-1}\mathbf{1} = \begin{pmatrix} 19/13 & 10/13 & 4/13 \\ 15/13 & 25/13 & 10/13 \\ 9/13 & 15/13 & 19/13 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 33/13 \\ 50/13 \\ 43/13 \end{pmatrix}$$

so, on the average, a tied game requires $50/13 = 3.846$ points to be completed.

General solution. To find the duration of a game starting from a tie. Divide the game into pairs of points. In each pair we will have a win by one player with probability $p^2 + (1 - p)^2$ or return to a tie otherwise. Thus the number of pairs of points we need to decide the game is 2 times a geometric with success probability $p^2 + (1 - p)^2$ and the mean is $2/(p^2 + (1 - p)^2)$. When $p = 0.6$ this is

$$\frac{2}{0.36 + 0.16} = \frac{50}{9 + 4}$$

Example 6.25. Credit ratings. At the end of a month, a large retail store classifies each of its customer's accounts according to current (0), 30–60 days overdue (1), 60–90 days overdue (2), more than 90 days overdue (3). Their experience indicates that from onth to month the accounts move from state to state according to a Markov chain with transition probability matrix:

	0	**1**	**2**	**3**
0	.9	.1	0	0
1	.8	0.2	0	
2	.5	0	0	.5
3	.1	0	0	.9

(a) In the long run what fraction of the accounts are in each category? (b) If the customer is in state 1 what is the expected amount of time until they return to state 0?

(a) The stationary distribution is the last row of

$$\begin{pmatrix} -.1 & .1 & 0 & 1 \\ .8 & -1 & .2 & 1 \\ .5 & 0 & -1 & 1 \\ .1 & 0 & 0 & 1 \end{pmatrix}^{-1}$$

which is $\pi(0) = 100/122$, $\pi(1) = 10/122$, $\pi(2) = 2/122$ and $\pi(3) = 10/122$.

If we let $g(x)$ be the expected time to get to 0 starting from x and r be the 3×3 matrix obtained by deleting the row and column for state 0, then as we have seen in the two previous examples $(I - r)g = \mathbf{1}$ where $\mathbf{1}$ is a column vector with three 1's. In matrix form this is

$$\begin{pmatrix} 1 & -.2 & 0 \\ 0 & 1 & -.5 \\ 0 & 0 & .1 \end{pmatrix} \begin{pmatrix} g(1) \\ g(2) \\ g(3) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Inverting the matrix $I - r$ we have

$$\begin{pmatrix} g(0) \\ g(1) \\ g(2) \end{pmatrix} = \begin{pmatrix} 1 & .2 & 1 \\ 90 & 1 & 5 \\ 50 & 0 & 10 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2.2 \\ 6 \\ 10 \end{pmatrix}$$

To connect the two answers note that we stay in state 0 for a geometric(0.1) number of step, which has mean 10, and then it takes us an average of 2.2 units of time to return to 0, so the fraction of time we spend in 0 in equilibrium is

$$\frac{10}{10 + 2.2} = \frac{100}{122}$$

Example 6.26. Waiting time for coin patterns. Suppose we repeatedly flip a coin until we have gotten three tails in a row. What is the expected number of tosses we need?

To solve this question we define a Markov chain with states 0, 1, 2, and 3 corresponding to the number of consecutive tails we have at that point. The transition probablity is

	0	**1**	**2**	**3**
0	1/2	1/2	0	0
1	1/2	0	1/2	0
2	1/2	0	0	1/2
3	0	0	0	1

As in the previous problems if $g(x)$ is the expected number of tosses to get to 3 starting from x, and r is the 3×3 matrix obtained by deleting the row and coumn for state 3 then $(I - r)g = \mathbf{1}$ where $\mathbf{1}$ is a column vector consisting for three 1's. Inverting the matrix

$$\begin{pmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ -1/2 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 8 & 4 & 2 \\ 6 & 4 & 2 \\ 4 & 2 & 2 \end{pmatrix}$$

so the waiting time is $8 + 4 + 2 = 14$.

It is somewhat surprising that the answer is 14 rather than 8. The values of the last three flips (X_{n-2}, X_{n-1}, X_n) are an 8 state Markov chain in which each state has equilibrium probability 1/8. The problem with TTT is that once we have one instance of TTT then we will have another if the next coin is tails. Overlaps cannot happen for a pattern like TTH so as we will now show the expected waiting time is 8. Consider the Markov chain.

	∅	**T**	**TT**	**TTH**
∅	1/2	1/2	0	0
T	1/2	0	1/2	0
TT	0	0	1/2	1/2
TTH	0	0	0	1

Inverting the matrix $I - r$ gives

$$\begin{pmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ -1/2 & 0 & 1/2 \end{pmatrix}^{-1} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}$$

so the waiting time is 8.

6.5.3 Gambler's ruin*

Example 6.27. Consider a fair game in which you win a \$1 with probability 1/2 or lose a \$1 with probability 1/2. In this case your fortune is a Markov chain X_n with transition probability $p(x, x+1) = 1/2$ and $p(x, x-1) = 1/2$. Find the probability $h(x)$ that we reach \$N before hitting \$0 if we start with \$x?

Clearly $h(0) = 0$ and $h(N) = 1$. By considering what happens on one step we see that if $0 < x < N$

$$h(x) = (1/2)h(x+1) + (1/2)h(x-1)$$

Rearranging we see that

$$h(x+1) - h(x) = h(x) - h(x-1)$$

or h is a line with constant slope. Since $h(0) = 0$ and $h(N) = 1$ that slope must be $1/N$ and $h(x) = x/N$. The final can be seen by using the reasoning we used for the Wright-Fisher model. Since game is fair the average amount of money we have at the end is the same as what we have at the beginning so the probability of reaching N before 0 is x/N.

Example 6.28. Duration of fair games. Continuing to consider the fair game introduced in the last problem. Let $\tau = \min\{n : X_n \notin (0, N)\}$ be the amount of time the game lasts. We claim that

$$E_x \tau = x(N - x) \tag{6.12}$$

There are two ways to prove this.

Verify the guess. Let $g(x) = x(N-x)$. Clearly, $g(0) = g(N) = 0$. If $0 < x < N$ then by considering what happens on the first step we have

$$g(x) = 1 + \frac{1}{2}g(x+1) + \frac{1}{2}g(x-1)$$

If $g(x) = x(N-x)$ then the right-hand side is

$$\begin{aligned} &= 1 + \frac{1}{2}(x+1)(N-x-1) + \frac{1}{2}(x-1)(N-x+1) \\ &= 1 + \frac{1}{2}[x(N-x) - x + N - x - 1] + \frac{1}{2}[x(N-x) + x - (N-x+1)] \\ &= 1 + x(N-x) - 1 = x(N-x) \end{aligned}$$

Derive the answer. Again we begin with

$$g(x) = 1 + (1/2)g(x+1) + (1/2)g(x-1)$$

Rearranging gives

$$g(x+1) - g(x) = -2 + g(x) - g(x-1).$$

Setting $g(1) - g(0) = c$ we have $g(2) - g(1) = c - 2$, $g(3) - g(2) = c - 4$ and in general that

$$g(k) - g(k-1) = c - 2(k-1).$$

Using $g(0) = 0$ and summing we have

$$0 = g(N) = \sum_{k=1}^{N} c - 2(k-1) = cN - 2 \cdot \frac{N(N-1)}{2}$$

since, as one can easily check by induction, $\sum_{j=1}^{m} j = m(m+1)/2$. Solving gives $c = (N-1)$. Summing again, we see that

$$g(x) = \sum_{k=1}^{x} (N-1) - 2(k-1) = x(N-1) - x(x+1) = x(N-x)$$

Example 6.29. Unfair games. Now consider the sitautaion in which on any turn you win \$1 with probability $p < 1/2$ or lose \$1 with probability $1-p$. In this case your fortune is a Markov chain X_n with transition probability $p(x, x+1) = p$ and $p(x, x-1) = 1-p$. To analyze this example it is convenient to introduce some notation. $T_k = \min\{n \geq 0 : X_n = x\}$ is the time of the first visit to x. Using the new notation we want to compute

$$h(x) = P_x(T_N < T_0)$$

be the probability that our gambler reaches the goal of \$$N$ before going bankrupt when starting with \$$x$.

As in the case of a fair game, $h(0) = 0$ and $h(1) = 1$, and we consider what happens on the first step to arrive at

$$h(x) = ph(x+1) + qh(x-1) \tag{6.13}$$

To solve this we rearrange to get $p(h(x+1) - h(x)) = q(h(x) - h(x-1))$ and conclude

$$h(x+1) - h(x) = \frac{q}{p} \cdot (h(x) - h(x-1)) \tag{6.14}$$

If we set $c = h(1) - h(0)$ then (6.14) implies that for $x \geq 1$

$$h(x) - h(x-1) = c\left(\frac{q}{p}\right)^{x-1}$$

Summing from $x = 1$ to N, we have

$$1 = h(N) - h(0) = \sum_{x=1}^{N} h(x) - h(x-1) = c\sum_{x=1}^{N}\left(\frac{q}{p}\right)^{x-1}$$

Now for $\theta \neq 1$ the partial sum of the geometric series is

$$\sum_{j=0}^{N-1} \theta^j = \frac{1-\theta^N}{1-\theta} \tag{6.15}$$

To check this note that

$$(1-\theta)(1+\theta+\cdots\theta^{N-1}) = (1+\theta+\cdots\theta^{N-1}) \\ -(\theta+\theta^2+\cdots\theta^N) = 1-\theta^N$$

Using (6.15) we see that $c = (1-\theta)/(1-\theta^N)$ with $\theta = q/p$. Summing and using the fact that $h(0) = 0$, we have

$$h(x) = h(x) - h(0) = c\sum_{i=0}^{x-1}\theta^i = c\cdot\frac{1-\theta^x}{1-\theta} = \frac{1-\theta^x}{1-\theta^N}$$

Recalling the definition of $h(x)$ and rearranging the fraction we have

$$P_x(T_N < T_0) = \frac{\theta^x - 1}{\theta^N - 1} \qquad \text{where } \theta = \tfrac{1-p}{p} \tag{6.16}$$

To see what (6.16) says in a concrete example, we consider:

Example 6.30. Roulette. If we bet \$1 on red on a roulette wheel with 18 red, 18 black, and 2 green (0 and 00) holes, we win \$1 with probability 18/38 = 0.4737 and lose \$1 with probability 20/38. Suppose we bring \$50 to the casino with the hope of reaching \$100 before going bankrupt. What is the probability we will succeed?

Here $\theta = q/p = 20/18$, so (6.16) implies

$$P_{50}(T_{100} < T_0) = \frac{\left(\frac{20}{18}\right)^{50} - 1}{\left(\frac{20}{18}\right)^{100} - 1}$$

Using $(20/18)^{50} = 194$, we have

$$P_{50}(T_{100} < T_0) = \frac{194-1}{(194)^2-1} = \frac{1}{194+1} = 0.005128$$

6.6 Exercises

Two state chains

1. Market research suggests that in a five year period 8% of people with cable television will get rid of it, and 26% of those without it will sign up for it. Compare the predictions of the Markov chain model with the following data on the fraction of people with cable TV: 56.4% in 1990, 63.4% in 1995, and 68.0% in 2000.

2. A sociology professor postulates that in each decade 8% of women in the work force leave it and 20% of the women not in it begin to work. Compare the predictions of his model with the following data on the percentage of women working: 43.3% in 1970, 51.5% in 1980, 57.5% in 1990, and 59.8% in 2000.

3. A car rental company has rental offices at both Kennedy and LaGuardia airports. Assume that a car rented at one airport must be returned to one of the two airports. If the car was rented at LaGuardia the probability it will be returned there is 0.8; for Kennedy the probability is 0.7. Suppose that we start with 1/2 of the cars at each airport and that each

week all of the cars are rented once. (a) What is the fraction of cars at LaGuardia ariport at the end of the first week? (b) at the end of the second? (c) in the long run?

4. The 1990 census showed that 36% of the households in the District of Columbia were homeowners while the reaminder were renters. During the next decade 6% of the homeowners became renters and 12% of the renters became homeowners. (a) What percentage were homeowners in 2000? in 2010? (b) If these trends continue what will be the long run fraction of homeowners?

5. A rapid transit system has just started operating. In the first month of operation, it was found that 25% of commuters are using the system while 75% are traveling by automobile. Suppose that each month 10% of transit users go back to using their cars, while 30% of automobile users switch to the transit system. (a) Compute the three step transition probability p^3. (b) What will be the fractions using rapid transit in the fourth month? (c) In the long run?

6. A regional health study indicates that from one year to the next, 75% percent of smokers will continue to smoke while 25% will quit. 8% of those who stopped smoking will resume smoking while 92% will not. If 70% of the population were smokers in 1995, what fraction will be smokers in 1998? in 2005? in the long run?

7. Census results reveal that in the United States 80% of the daughters of working women work and that 30% of the daughters of nonworking women work. In the long run what fraction of women will be working?

8. Three of every four trucks on the road are followed by a car, while only one of every five cars is followed by a truck. What fraction of vehicles on the road are trucks?

9. In a test paper the questions are arranged so that 3/4's of the time a True answer is followed by a True, while 2/3's of the time a False answer is followed by a False. You are confronted with a 100 question test paper. Approximately what fraction of the answers will be True?

10. When a basketball player makes a shot then he tries a harder shot the next time and hits (H) with probability 0.4, misses (M) with probability 0.6. When he misses he is more conservative the next time and hits (H) with probability 0.7, misses (M) with probability 0.3. Find the long-run fraction of time he hits a shot.

11. A university computer room has 30 terminals. Each day there is a 3% chance that a given terminal will break and a 72% chance that a given broken terminal will be repaired. Assume that the fates of the various terminals are independent. In the long run what is the distribution of the number of terminals that are broken.

Three or more states

12. Consider a gambler's ruin chain with $N = 4$. That is, if $1 \le i \le 3$, $p(i, i+1) = 0.4$, and $p(i, i-1) = 0.6$, but the endpoints are absorbing states: $p(0,0) = 1$ and $p(4,4) = 1$ Compute $p^3(1,4)$ and $p^3(1,0)$.

13. A taxicab driver moves between the airport A and two hotels B and C according to the following rules. If he is at the airport, he will go to one of the two hotels next with equal probability. If at a hotel then he returns to the airport with probability 3/4 and goes to the other hotel with probability 1/4. (a) Find the transition matrix for the chain. (b) Suppose

the driver begins at the airport at time 0. Find the probability for each of his three possible locations at time 2 and the probability he is at hotel B at time 3.

14. A person is flipping a coin repeatedly. Let X_n be the outcome of the two previous coin flips at time n, for example the state might be HT to indicate that the last flip was T and the one before that was H. (a) compute the transition probability for the chain. (b) Find p^2.

15. The town of Ithaca has a "free bikes for the people program." You can pick up bikes at the library (L), the coffee shop (C) or the cooperative grocery store (G) and drop them off at any of these locations. The movement of bikes can be modeled by a Markov chain

	L	**C**	**G**
L	0.5	0.1	.0.4
C	0.3	0.5	0.2
G	0.2	0.4	0.4

(a) What is the probability a bike from the library is at the three locations on Wednesday? (b) In the long run what fraction are at the three locations?

16. Bob eats lunch at the campus food court every week day. He either eats Chinese food, Quesadila, or Salad. His transition matrix is

	C	**Q**	**S**
C	.15	.6	.25
Q	.4	.1	.5
S	.1	.3	.6

(a) He had Chinese food on Monday. What are the probabilities for his three meal choices on Friday (four days later). (b) What are the lun run frequencies for his three meal choices.

17. A midwestern university has three types of health plans: a health maintenance organization (HMO), a preferred provider organization (PPO), and a traditional fee for service plan (FFS). In 2000, the prcentages for the three plans were HMO:30%, PPO:25%, and FFS:45%. Experience dictates that people change plans according to the following transition matrix

	HMO	**PPO**	**FFS**
HMO	.85	.1	.05
PPO	.2	.7	.1
FFS	.1	.3	.6

(a) What will be the precentages for the three plans in 2001? (b) What is the long run fraction choosing each of the three plans?

18. A plant species has red, pink, or white flowers according to the genotypes RR, RW, and WW, respectively. If each of these genotypes is crossed with a pink (RW) plant then the offspring fractions are

	RR	**RW**	**WW**
RR	.5	.5	0
RW	.25	.5	.25
WW	0	.5	.5

What is the long run fraction of plants of the three types?

19. A sociologist studying living patterns in a certain region determines that the pattern of movement between urban (U), suburban (S), and rural areas (R) is given by the following transition matrix.

	U	**S**	**R**
U	.86	.08	.06
S	.05	.88	.07
R	.03	.05	.92

In the long run what fraction of the population will live in the three areas?

20. In a large metropolitan area, commuters either drive alone (A), carpool (C), or take public transportation (T). The mode of transportation used changes from year to year according to the transition probability

	A	C	T
A	.8	.15	.05
C	.05	.9	.05
S	.05	.1	.85

In the long run what fraction of commuters will use the three types of transportation?

21. In a particular county voters declare themselves as members of the Republican, Democrat, or Green party. Voters change parties according to the transition probability:

	R	**D**	**G**
R	.85	.15	0
D	.05	.85	.10
G	0	.05	.95

In the long run what fraction of voters will belong to the three parties?

22. An auto insurance company classifies its customers in three categories: poor (P), satisfactory (S) and excellent (E). Ratings change over time according to the following transition matrix

	.6	.4	0
a.	.1	.6	.3
	0	.2	.8

What is the limiting fraction of drivers in each of these categories?

23. An individual has three umbrellas, some at her office, and some at home. If she is leaving home in the morning (or leaving work at night) and it is raining, she will take an umbrella, if one is there. Otherwise, she gets wet. Assume that independent of the past, it rains on each trip with probability 0.2. To formulate a Markov chain, let X_n be the number of umbrellas at her current location. (a) Find the transition probability for this Markov chain. (b) Calculate the limiting fraction of time she gets wet.

24. At the beginning of each day, a piece of equipment is inspected to determine its working condition, which is classified as state 1 = new, 2, 3, or 4 = broken. We assume the state is a Markov chain with the following transition matrix:

	1	**2**	**3**	**4**
1	.95	.05	0	0
2	0	.9	.1	0
3	0	0	.875	.125

(a) Suppose that a broken machine requires three days to fix. To incorporate this into the Markov chain we add states 5 and 6 and suppose that $p(4,5) = 1$, $p(5,6) = 1$, and $p(6,1) = 1$. Find the fraction of time that the machine is working. (b) Suppose now that we have the option of performing preventative maintenance when the machine is in state 3, and that this maintenance takes one day and returns the machine to state 1. This changes the transition probability to

	1	**2**	**3**
1	.95	.05	0
2	0	.9	.1
3	1	0	0

Find the fraction of time the machine is working under this new policy.

25. Five white balls and five black balls are distributed in two urns in such a way that each urn contains five balls. At each step we draw one ball from each urn and exchange them. Let X_n be the number of white balls in the left urn at time n. (a) Compute the transition probability for X_n. (b) Find the stationary distribution and show that it corresponds to picking five balls at random to be in the left urn.

Exit distributions and times

26. The Macrosoft company gives each of its employees the title of programmer (P) or project manager (M). In any given year 70% of programmers remain in that position 20% are promoted to project manager and 10% are fired (state X). 95% of project managers remain in that position while 5% are fired. How long on the average does a programmer work before they are fired?

27. At a nationwide travel agency, newly hired employees are classified as beginners (B). Every six months the performance of each agent is reviewed. Past records indicate that transitions through the ranks to intermediate (I) and qualified (Q) are according to the following Markov chain, where F indicates workers that were fired:

	B	**I**	**Q**	**F**
B	.45	.4	0	.15
I	0	.6	.3	.1
Q	0	0	1	0
F	0	0	0	1

(a) What fraction of beginners eventually become qualified? (b) What is the expected time until a beginner is fired or becomes qualified?

28. At a manufacturing plant, employees are classified as a recruit (R), technician (T) or supervisor (S). Writing Q for an employee who quits we model their progress through the ranks as a Markov chain with transition probability

	R	**T**	**S**	**Q**
R	.2	.6	0	.2
T	0	.55	.15	.3
S	0	0	1	0
Q	0	0	0	1

(a) What fraction of recruits eventually become a supervisor? (b) What is the expected time until a recruit quits or becomes supervisor?

29. The Markov chain associated with a manufacturing process may be described as follows: A part to be manufactured will begin the process by entering step 1. After step 1, 20% of the parts must be reworked, i.e., returned to step 1, 10% of the parts are thrown away, and 70% proceed to step 2. After step 2, 5% of the parts must be returned to the step 1, 10% to step 2, 5% are scrapped, and 80% emerge to be sold for a profit. In short, the transition probability is

	1	2	3	4
1	.2	.7	.1	0
2	.05	.1	.05	.8
3	0	0	1	0
4	0	0	0	1

(b) Compute the probability a part is scrapped in the production process. (c) Find the expected time until the part is either scrapped or finished.

30. The Duke football team can Pass, Run, throw an Interception, or Fumble. Suppose the sequence of outcomes is Markov chain with the following transition matrix.

	P	**R**	**I**	**F**
P	0.7	0.2	0.1	0
R	0.35	0.6	0	0.05
I	0	0	1	0
F	0	0	0	1

The first play is a pass. (a) What is the expected number of plays until a fumble or interception? (b) What is the probability the sequence of plays ends in an interception.

31. A professor has two light bulbs in his garage. When both are burned out, they are replaced, and the next day starts with two working light bulbs. Suppose that when both are working, one of the two will go out with probability .02 (each has probability .01 and we ignore the possibility of losing two on the same day). However, when only one is there, it will burn out with probability .05. Letting the state be the number of burnt out light bulbs we get a Markov chain.

	0	1	2
0	.98	.02	0
1	0	.95	.05
2	1	0	0

(a) What is the long-run fraction of time that there is exactly one bulb working? (b) Write and solve the usual equations to find the expected time to reach state 2 starting from state 0. (c) Find the answers to (a) and (b) by noting that the amount of time spent in state 0 is geometric(0.02) and the time in state 1 is geometric(0.05).

32. In Example 6.26 we considered the expected time to see TTT and TTH. Without loss of generality we can suppose that the first letter is a T. Find the expected time to see (a) THT, (b) THH.

33. A California surfer dude only says $A =$ awesome, $B =$ bogus, $C =$ chill and $D =$ dig it.

He changes words according to the following transition matrix

	A	**B**	**C**	**D**
A	.1	.3	.2	.4
B	.25	.35	.3	.1
C	.25	.45	.1	.2
D	.4	.12	.16	.32

(a) Find the stationary distribution for the words he uses.
(b) The surfer dude just said Bogus. What is the probability he will say Dig it before he says Awesome.

34. To make a crude model of a forest we might introduce states 0 = grass, 1 = bushes, 2 = small trees, 3 = large trees, and write down a transition matrix like the following:

	0	**1**	**2**	**3**
0	1/2	1/2	0	0
1	1/24	7/8	1/12	0
2	1/36	0	8/9	1/12
3	1/8	0	0	7/8

The idea behind this matrix is that if left undisturbed a grassy area will see bushes grow, then small trees, which of course grow into large trees. However, disturbances such as tree falls or fires can reset the system to state 0. (a) Find the limiting fraction of land in each of the states. (b) Compute the probability that starting from state 1, the system reaches state 3 before hitting 0. (c) Find the expected amount to time to get to 3 starting from 0.

35. A barber has a shop with two chairs for waiting customers. We assume that customers who come when there are three people in the shop go away. If we assume that the barber can complete a haircut with probability 0.6 before the next person comes we arrive at the following Markov chain.

	0	**1**	**2**	**3**
0	0	1	0	0
1	0.6	0	0.4	0
2	0	0.6	0	0.4
3	0	0	0.6	0.4

(a) Find the stationary distribution. (b) Find the probability that starting from 1 the shop becomes full before it becomes empty. (c) Find the expected time to return to state 0 starting from state 1. (d) What is the relationship between the answer to (c) and $\pi(0)$.

36. A warehouse has a capacity to hold four items. If the warehouse is neither full nor empty, the number of items in the warehouse changes whenever a new item is produced or an item is sold. Suppose that the probability that the next event is “a new item is produced” is 2/3 and that the new event is a “sale” is 1/3.

	0	**1**	**2**	**3**	**4**
0	1/3	2/3	0	0	0
1	1/3	0	2/3	0	0
2	0	1/3	0	2/3	0
3	0	0	1/3	0	2/3
4	0	0	0	1/3	2/3

Suppose there is currently one item in the warehouse. (a) What is the probability that the warehouse will become full before it becomes empty? (b) What is the expected time to reach 0 or 4?

37. **Brother-sister mating.** A pair of an animals are mated and then one male and one female offspring are chosen to be the next pair. Ignoring the genetics details this leads to the following Markov chain. 4 and 0 are absorbing states.

	4	**3**	**2A**	**2B**	**1**	**0**
4 (**AA**, **AA**)	1	0	0	0	0	0
3 (**AA**, **Aa**)	1/4	1/2	1/4	0	0	0
2A (**Aa**, **Aa**)	1/16	1/4	1/4	1/8	1/4	1/16
2B (**AA**, **aa**)	0	0	1	0	0	0
1 (**Aa**, **aa**)	0	0	1/4	0	1/2	1/4
0 (**aa**, **aa**)	0	0	0	0	0	1

To explain the most complicated row, note that when Aa and Aa mate to create one individual then the result is AA with probability 1/4, Aa with probability 1/2, and aa with probability 1/4. Since the two individuals are independent AA, AA and aa, aa have probability 1/16, Aa, Aa has probability 1/4, AA, aa has probability $2 \cdot 1/16$, AA, Aa and Aa, aa have probability $2 \cdot 1/8$. (a) Find $h(x)$ the probability of ending in state 4 starting from x. (b) Let τ be the time that the process first enters 0 or 4. Find $g(x) = E_x \tau$.

38. Six children (Dick, Helen, Joni, Mark, Sam, and Tony) play catch. If Dick has the ball he is equally likely to throw it to Helen, Mark, Sam, and Tony. If Helen has the ball she is equally likely to throw it to Dick, Joni, Sam, and Tony. If Sam has the ball he is equally likely to throw it to Dick, Helen, Mark, and Tony. If either Joni or Tony gets the ball, they keep throwing it to each other. If Mark gets the ball he runs away with it. (a) Find the transition probability. (b) Suppose Dick has the ball at the beginning of the game. What is the probability Mark will end up with it? (c) What is expected time until the ball ends up in the hands of Mark, or Joni and Tony?