Chapter 3

Conditional Probability

3.1 Definition

Suppose we are told that the event A with $P(A) > 0$ occurs. As explained in Section 1.3, then the sample space is reduced from Ω to A and by (1.6) and the probability that B will occur given that A has occurred is

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{3.1}$$

Example 3.1. 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. 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 $C = B^c$

$$P(C|A) = 1 - P(B|A) \tag{3.2}$$

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 3.2. 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 $1 + 2 + 3 + 4 + 5 = 21$ outcomes in A and if we condition on A they are all equally likely. $A \cap R_5$ has four outcomes, so $P(R_5|A) = 4/21$. In general, $P(R_i|A) = (i-1)/21$ for $1 \le i \le 6$.

Example 3.3. A person picks 13 cards out of a deck of 52. Let A_1 = "he hass 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 (3.2) 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.

Multiplying the definition of conditional probability in (2.1) on each side by $P(A)$ gives the **multiplication rule**

$$P(A)P(B|A) = P(B \cap A) \tag{3.3}$$

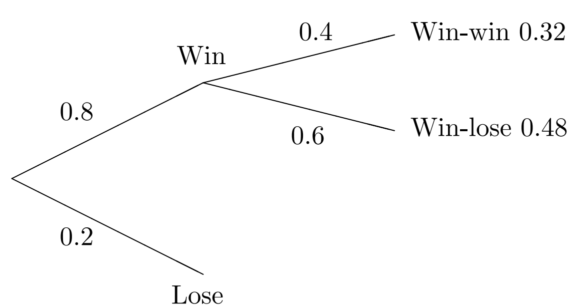
Example 3.4. Suppose we draw two cards out of a deck of 52. What is the probability both cards are spades?

Let A = "the first card is a spade," B = "the second card is a spade." $P(A) = 1/13$. To compute $P(B|A)$ we note that if A has occurred then only 12 of the remaining 51 cards are spades, so $P(B|A) = 12/51$ and

$$P(A \cap B) = P(A)P(B|A) = \frac{13}{52} \cdot \frac{12}{51}$$

Note that in this example we computed $P(B|A)$ by thinking about the situation that exists after A has occurred, rather than using the definition $P(B|A) = P(A \cap B)/P(A)$. Indeed, it is more common to use $P(A)$ and $P(B|A)$ to compute $P(A \cap B)$ than to use $P(A)$ and $P(A \cap B)$ to compute $P(B|A)$.

Example 3.5. The Cornell hockey 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

$$P(A \cap B) = P(A)P(B|A) = 0.8(0.4) = 0.32$$

The reasoning in the last two examples extends easily to three events:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

since the right-hand side is equal to

$$P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}$$

Example 3.6. In the town of Mythica 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?

Answer $= (0.9)(0.6)(0.2) = 0.108$.

The formula for three events generalizes to any number of events.

Example 3.7. 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?

$$1 \cdot \frac{12}{51} \cdot \frac{11}{50} \cdot \frac{10}{49} \cdot \frac{9}{48}$$

The first time we can draw anything. On the second draw we must pick one of the other 12 cards in that suit among the 51 that remain. If we succeed on the second draw then there are 11 good cards out of 50, etc.

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 3.8. 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.

The paradox actually predates the game show in the following form. Three prisoners, Al, Bob, and Charlie, are in a cell. At dawn two will be set free and one will be hanged, but they do not know who will be chosen. The guard offers to tell Al the name of one of the other two prisoners who will go free but Al stops him, screaming, "No, don't! That would increase my chances of being hanged to 1/2."

Example 3.9. 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 monkeys preferences by observing how quickly each monkey sought out different colors of M&Ms.

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.

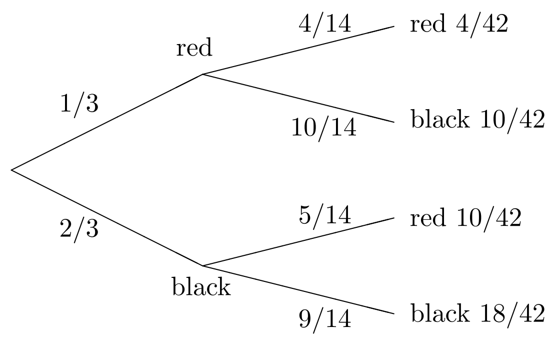
Putting aside this interpretation it is natural to ask: What would happen if monkeys were acting at random? The six orderings RGB, RBG, GRB, GBR, BGR, and BRG would have equal probability. In the first three cases red is preferred to blue, but in 2/3s of those cases green is preferred to blue. Just as 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.

3.2 Two-Stage Experiments

We begin with several examples and then describe the collection of problems we will treat in this section.

Example 3.10. 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: $4/42+10/42 = 1/3$.



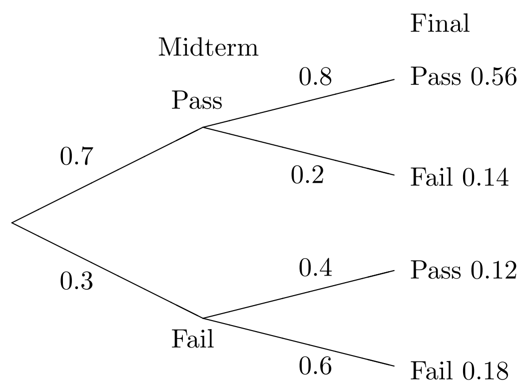
To do this with formulas, let R_i be the event of a red ball on the ith draw and let B_1 be the event of a black ball on the first draw. Breaking things down according to the outcome of the first test, then using the multiplication rule, we have

$$\begin{aligned} P(R_2) &= P(R_2 \cap R_1) + P(R_2 \cap B_1) \\ &= P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1) \\ &= (1/3)(4/14) + (2/3)(5/14) = 14/42 = 1/3 \end{aligned}$$

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 3.11. 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.

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B)P(B) + P(A|B^c)P(B^c) \\ &= (0.8)(0.7) + (0.4)(0.3) = 0.68 \end{aligned}$$

Example 3.12. 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?

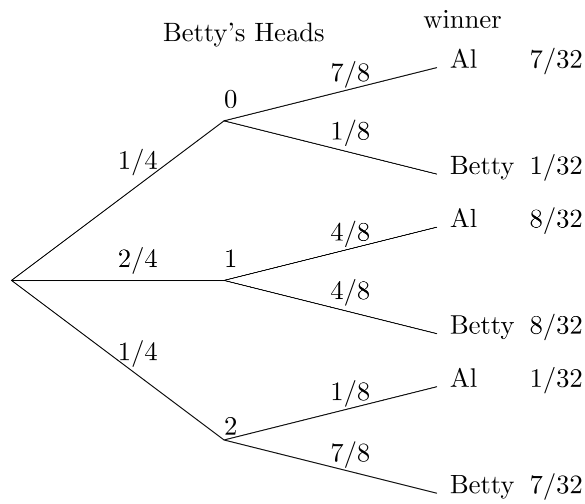
Let W be the event that Al wins. We will break things down according to the number of heads Betty gets. Let B_i be the event that Betty gets i Heads, and let A_j be the event that Al gets j Heads. By considering the four outcomes of flipping two coins it is easy to see that

$$P(B_0) = 1/4 \qquad P(B_1) = 1/2 \qquad P(B_2) = 1/4$$

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 (3.3). Plugging in the values we computed,

$$P(W) = \frac{1}{4} \cdot \frac{7}{8} + \frac{2}{4} \cdot \frac{4}{8} + \frac{1}{4} \cdot \frac{1}{8} = \frac{7 + 8 + 1}{32} = \frac{1}{2}$$

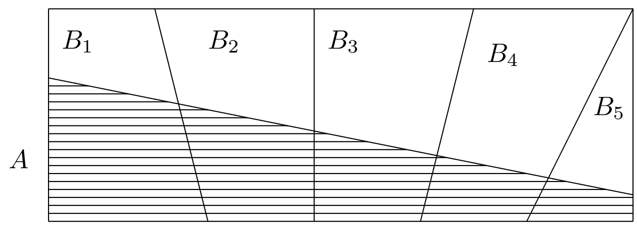
The previous analysis makes it look miraculous that we have a fair game. However it is true in general.

Example 3.13. 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

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i) \tag{3.4}$$

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 3.14. 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 (3.4),

$$\begin{aligned} P(A) &= \frac{1}{6}\left\{\frac{1}{8} + \frac{4}{16} + \frac{10}{32} + \frac{20}{64}\right\} \\ &= \frac{1}{6}\left\{\frac{8 + 16 + 20 + 20}{64}\right\} = \frac{1}{6} \end{aligned}$$

Example 3.15. Suppose we roll three dice. What is the probability that the sum is 9?

Let A = "the sum is 9," B_i = "the first die shows i," and C_j = "the sum of the second and third dice is j." Now $P(A|B_i) = P(C_{9-i})$ and we know the probabilities for the sum of two dice:

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

Using (3.4), now we have

$$
\begin{aligned}
P(A) = \sum_{i=1}^{6} P(B_i)P(A|B_i) &= \frac{1}{6}\left(P(C_8) + P(C_7) + \cdots + P(C_3)\right) \\
&= \frac{1}{6}\left(\frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36}\right) = \frac{25}{216}
\end{aligned}
$$

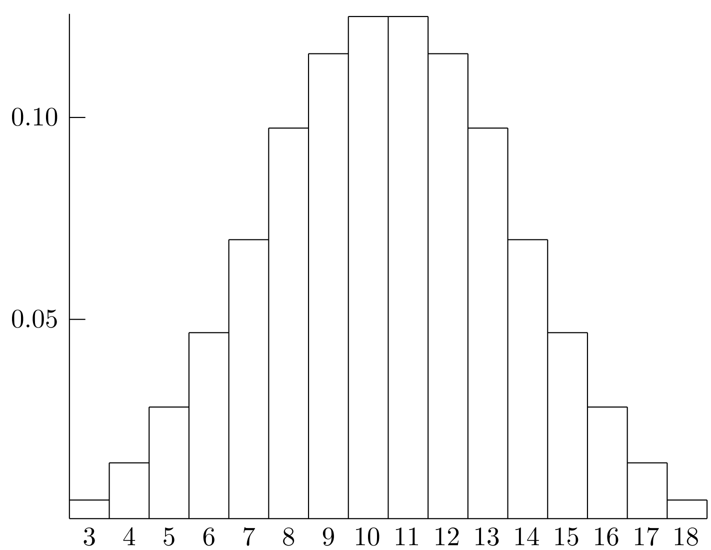


Figure 3.1: Distribution of the sum of three dice.

In the same way we can compute the probability of A_k = "The sum of three dice is k". To check the symmetry in the table, note that if the numbers on top are $i_1 + i_2 + i_3 = k$, then the sum of the numbers on the bottom are $(7 - i_1) + (7 - i_2) + (7 - i_3) = 21 - k$.

k	3,18	4,17	5,16	6,15	7,14	8,13	9,12	10,11
$P(A_k)$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$

The graph in Figure 3.1 shows the shape of the distribution. Note that the triangular shape of the sum of two dice has become a little more rounded.

Example 3.16. 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

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \tag{3.5}$$

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	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 (3.4) 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, $P(A|B_k) = P(k$ 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,

$$\begin{aligned} P(A) &= \frac{6}{36} + \frac{2}{36} + 2\left(\frac{1}{36} + \frac{4 \cdot 2}{36 \cdot 5} + \frac{5 \cdot 5}{36 \cdot 11}\right) \\ &= \frac{4}{18} + 2\left(\frac{55 + 88 + 125}{36 \cdot 11 \cdot 5}\right) = \frac{220 + 268}{18 \cdot 11 \cdot 5} = \frac{488}{990} = 0.4929 \end{aligned} \qquad (3.6)$$

which is not very much less than $1/2 = 495/990$.

Example 3.17. 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 $q = 4/10$.

Example 3.18. NCAA basketball tournament. Since 1985 the tournament has had 64 teams, four regions with 16 seeded teams. This is a knockout tournament, i.e., after each game the loser is eliminated. The table below will present data for 1985-2004, 20 seasons. Since there are four regions, this means that each seeding has had a total of 80 trials. The table describes relative success of the various seeds in advancing in the tournament to the rounds of 32, sweet 16, elite 8, the final 4, the 2 teams in the championship game, and to win the tournament. The numbers are decreasing across each row. For readability once a number becomes 0 the remaining entries are left blank.

For reasons that will become clear as you read the table we have listed the seeds in the order dictated by how the games are played. That is, in the first round the sum of the seeds of the two teams is always 17, and the number of times the teams advance will add up to 80. In the round of 16 statistics if we divide the 16 numbers into four groups of four, each will add up to 80, etc.

seed	32	16	8	4	2	winner
1	80	68	56	34	17	11
16	0					
8	37	9	6	3	1	1
9	43	3	1	0		
4	64	36	12	7	2	1
13	16	3	0			
5	54	28	4	3	2	0
12	26	13	1	0		
3	67	38	18	11	7	2
14	13	2	0			
6	56	30	11	3	2	1
11	24	10	3	1	0	
2	76	51	37	18	9	4
15	4	0				
7	48	13	5	0		
10	32	16	6	0		
total	640	320	160	80	40	20

From this table we can compute the probabilities for the first four seeds to win a game in each round, given that it reached that round.

	64	32	16	8	4	2
1	1.0	0.85	0.823	0.607	0.5	0.647
2	0.95	0.671	0.725	0.486	0.5	0.444
3	0.838	0.567	0.474	0.611	0.636	0.285
4	0.8	0.563	0.333	0.583	0.286	0.5

Here $68/80 = 0.85$, $56/68 = 0.823$, etc. As we should expect the conditional probabilities generally decrease from left to right and from top to bottom. We leave it to the reader to ponder the meaning of the exceptions, some of which may be due only to the small sample sizes.

3.3 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 3.19. 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 exti 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 (1.6),

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

To evaluate the two probabilities, we use the multiplication rule (3.3)

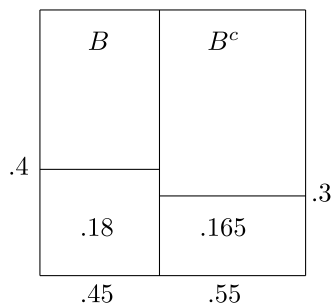
$$P(B \cap A) = P(B)P(A|B) = 0.45 \cdot 0.4 = 0.18$$
$$P(B^c \cap A) = P(B^c)P(A|B^c) = 0.55 \cdot 0.3 = 0.165$$

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 3.20. 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 (3.3)

$$\begin{aligned} P(B \cap A) &= P(B)P(A|B) = 0.01 \cdot 0.9 = 0.009 \\ P(B^c \cap A) &= P(B^c)P(A|B^c) = 0.99 \cdot 0.1 = 0.099 \end{aligned}$$

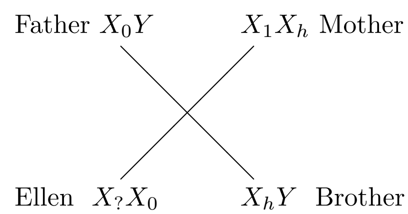
From this it follows that

$$P(B|A) = \frac{0.009}{0.009 + 0.099} = \frac{9}{108}$$

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 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 canner. 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 3.21. 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 (3.3). 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 3.22. 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

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A)}{\sum_{i=1}^{3} P(B_i \cap A)}$$

By the definition of conditional probability (1.6),

$$P(B_3|A) = P(B_3 \cap A)/P(A)$$

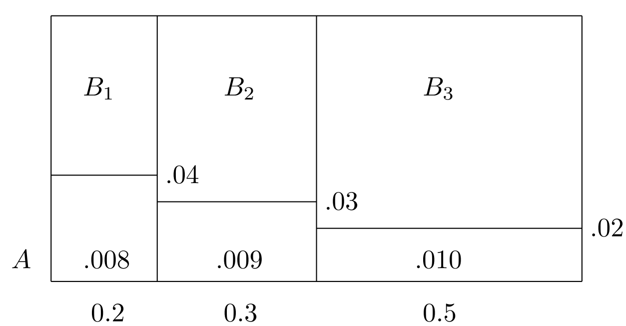
To evaluate the three probabilities, we use the multiplication rule (3.3)

$$\begin{aligned}
P(B_1 \cap A) &= P(B_1)P(A|B_1) = 0.2 \cdot (0.04) = 0.008 \\
P(B_2 \cap A) &= P(B_2)P(A|B_2) = 0.3 \cdot (0.03) = 0.009 \\
P(B_3 \cap A) &= P(B_3)P(A|B_3) = 0.5 \cdot (0.02) = 0.010
\end{aligned}$$

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

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(B_1)P(A|B_1)}{\sum_{i=1}^n P(B_i)P(A|B_i)} \tag{3.7}$$

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.

Our last two examples come from law.

Example 3.23. Paternity probabilities. Before there were sophisticated tests based on DNA samples, the testing of blood type and other hereditary factors was used in paternity cases to infer, using Bayes formula, the probability that a particular man is the father. For a concrete example suppose that the baby's blood type is B, the mother's is A, and that of the suspected father, whom for convenience we will call Bob, is B. Given this information what is the probability Bob is the father?

To explain how this could happen, we note that the genes that control blood type can be O, A, or B, with A and B dominant over O but neither A nor B dominating the other, so we get the following correspondence between genotypes (the genes on the two chromosomes) and phenotypes (observed blood type):

genotype	OO	AO	AA	BO	BB	AB
phenotype	O	A	A	B	B	AB
proportion	.479	.310	.050	.116	.007	.038

From this table, we see that if the baby's blood type is B then it must be the case that the mother's genotype is AO, she contributed an O gene, and the father contributed a B gene.

Let E (for "evidence") be the event that the baby's blood type is B, and F be the event that Bob is indeed the father. We cannot observe Bob's genotype, but using the proportions of the various genotypes from the table, we can compute that

$$P(\text{genotype is } BO | \text{ phenotype is } B) = 0.116/0.123$$

$$P(E|F) = \frac{(0.116)0.5 + 0.007}{0.123} = \frac{0.065}{0.123} = 0.528$$

There is not too much to argue about in the last computation. When we compute $P(E|F^c)$, we make the first of two questionable assumptions: If Bob is not the father, then the real father is someone chosen at random from the population, so

$$P(E|F^c) = (0.116)0.5 + 0.007 = 0.065$$

To evaluate $P(F|E)$ we thus need to evaluate the **prior probability** $P(F)$ that Bob is the father. It would be natural to make $P(F)$ equal to the fraction of times that the mother had intercourse with Bob near the time of conception. However, it is not unusual for the mother to claim this number is 1 and the alleged father to claim it is 0, so the common practice in these computations is to set $P(F) = 1/2$ (our second questionable assumption). If we do this, then $P(F) = P(F^c) = 1/2$ so

$$\begin{aligned} P(F|E) &= \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)} \\ &= \frac{0.065/0.123}{0.065/0.123 + 0.065} = \frac{1}{1.123} = 0.8904 \end{aligned}$$

Example 3.24. O.J. Simpson trial. DNA testing has considerably more power than blood tests. RFLP's (restriction fragment length polymorphisms) are typed by digesting DNA with "restriction" enzymes and then determining the lengths of the fragments. These lengths are highly variable (polymorphic) in humans, so the use of eight or nine such markers results in incredibly small probabilities. For example, Robin Cotton of Cellmark Diagnostics testified that blood found on a sock near Simpson's bed had the genetic type of Nicole Brown Simpson and the chances of another person having the exact same RFLP alleles were 1 in 9.7 billion.

For this problem we will concentrate on a less dramatic example that involves blood found at the murder scence (item #49). Three blood factors were recorded that matched Simpson's blood types. The next table which comes from p.10 of Vol. 7, No. 4 of *Chance* gives the frequencies estimated from the overall population.

System	Item #49	Frequency
ABO	A	0.347
EsD	1	0.79
PGM	2+,2−	0.016

Here 2+, 2− indicates that two alleles were present one inherited from each parent.

Multiplying the probabilities together gives 0.00438 or 1/227, a number that was approximated in the trial and quoted in the press as 1/200. Letting E denote this evidence and G the event that Simpson is guilty, and sticking with the simpler fraction, we see that $P(E|G^c) = 1/200$. It is an error known as the "Prosecutor's Fallacy" to think of this as $P(G^c|E)$, i.e., the probability that Simpson is innocent given this evidence is 1/200. A second error known as the "Defendant's Fallacy" is to note that 1/200 of the population of Los Angeles is 40,000, so the probability that it is O.J. Simpson's blood is 1/40,000.

Both of these fallacies are based on assuming that unknown probabilities are uniform on the set of possibilities. The correct way to compute $P(G|E)$ is

$$P(G|E) = \frac{P(E|G)P(G)}{P(E|G)P(G) + P(E|G^c)P(G^c)}$$

but of course this requires giving a value to $P(G)$. It is perhaps for this reason that Bruce Weir (Nature Genetics, Vol 11, pages 365–368) argues for the use of the likelihood ratio $P(E|G)/P(E|G^c) = 200$, i.e., the evidence is 200 times more likely if O.J. Simpson is guilty than if the murderer is a randomly chosen person.

3.4 Joint Distributions

In many situations we need to know the relationship between several random variables $X_1, \ldots, X_n$. Here, we will confine our attention to the case $n = 2$. Once this case is understood the extension to $n > 2$ is straightforward. For one random variable, the distribution is a list of the probabilities of all of the possible values. The joint distribution of a pair of random variables is a table of numbers that gives the probabilities for all the possible values of the pair.

Example 3.25. Roll two four-sided dice with the numbers $1, 2, 3, 4$ on their sides. Let X be the maximum of the two numbers that appear and Y be the sum. By considering the sixteen possible outcomes we find the following joint distribution of X and Y.

X	Y=2	3	4	5	6	7	8
1	1/16	0	0	0	0	0	0
2	0	2/16	1/16	0	0	0	0
3	0	0	2/16	2/16	1/16	0	0
4	0	0	0	2/16	2/16	2/16	1/16

Example 3.26. Suppose we draw 2 balls out of an urn with 6 red, 5 blue and 4 green balls. Let X be the number of red balls we get and Y the number of blue balls.

There are $C_{15,2} = 105$ outcomes. The number of outcomes with i red, j blue, and $2 - (i + j)$ green balls is $C_{6,i}C_{5,j}C_{4,2-(i+j)}$. Using this formula we have

X	Y=0	1	2
0	$\frac{1\cdot1\cdot6}{105} = 6/105$	$\frac{1\cdot5\cdot4}{105} = 20/105$	$\frac{1\cdot10\cdot1}{105} = 10/105$
1	$\frac{6\cdot1\cdot4}{105} = 24/105$	$\frac{6\cdot5\cdot1}{105} = 30/105$	0
2	$\frac{15\cdot1\cdot1}{105} = 15/105$	0	0

Example 3.27. Consider the following hypothetical joint distribution of X, a person's grade on the AP calculus exam (a number between 1 and 5), and their grade Y in their high school calculus course, which we assume was $A = 4$, $B = 3$, or $C = 2$.

X	Y= 4	3	2
5	.1	.05	0
4	.15	.15	0
3	.1	.15	.10
2	0	.05	.10
1	0	0	.05

Marginal Distributions

The next question to be addressed is: Given the joint distribution of (X, Y), how do we recover the distributions of X and Y? The answer is that the **marginal distributions** of X and Y are given by

$$\begin{aligned} P(X = x) &= \textstyle\sum_y P(X = x, Y = y) \\ P(Y = y) &= \textstyle\sum_x P(X = x, Y = y) \end{aligned} \tag{3.8}$$

To explain the first formula in words, if $X = x$ then Y will take on some value y, so to find $P(X = x)$ we sum the probabilities of the disjoint events $\{X = x, Y = y\}$ over all the values of y. To illustrate these formulas we return to Example 3.25, where we rolled a four sided die and let X be the larger number and Y the sum. Omitting the probabilities that are 0:

$$\begin{aligned} P(X = 1) &= P(X = 1, Y = 1) = 1/16 \\ P(X = 2) &= P(X = 2, Y = 3) + P(X = 2, Y = 4) = 2/16 + 1/16 = 3/16 \\ P(X = 3) &= P(X = 3, Y = 4) + P(X = 3, Y = 5) + P(X = 3, Y = 6) \\ &= 2/16 + 2/16 + 1/16 = 5/16 \\ P(X = 4) &= \sum_{y=5}^{8} P(X = 4, Y = y) = 2/16 + 2/16 + 2/16 + 1/16 = 7/16 \end{aligned}$$

In words we add the probabilities in each row to get the marginal distribution of X. Similarly we add up the probabilities in each column to get the distribution of Y.

X	Y=2	3	4	5	6	7	8	
1	1/16	0	0	0	0	0	0	1/16
2	0	2/16	1/16	0	0	0	0	3/16
3	0	0	2/16	2/16	1/16	0	0	5/16
4	0	0	0	2/16	2/16	2/16	1/16	7/16
	1/16	2/16	3/16	4/16	3/16	2/16	1/16	

In the urn example

X	Y=0	1	2	
0	6/105	20/105	10/105	36/105
1	24/105	30/105	0	54/104
2	15/105	0	0	15/105
	45/105	50/105	10/105	

To check the marginal distribution of Y note that when we draw from an urn with 6 red, 5 blue and 4 green balls, the probabilities for the number of blue balls are

$$0 : \frac{C_{10,2}}{105} = \frac{45}{105} \qquad 1 : \frac{C_{5,1}C_{10,1}}{105} = \frac{50}{105} \qquad 2 : \frac{C_{5,2}}{105} = \frac{10}{105}$$

Finally in the AP calculus example, the marginal distributions are

X	Y= 4	3	2	
5	.10	.05	0	.15
4	.15	.15	0	.30
3	.10	.15	.10	.35
2	0	.05	.10	.15
1	0	0	.05	.05
	.35	.40	.25	

Independence

Two random variables are **independent** if

$$P(X = x, Y = y) = P(X = x)P(Y = y) \tag{3.9}$$

In words, two random variables are independent if for each x and y the events $\{X = x\}$ and $\{Y = y\}$ are independent. To use two of our new terms, this occurs if their joint distribution is the product of the two marginal distributions. In the dice example

$$P(X = 1, Y = 4) = 0 < (1/16) \cdot (3/16) = P(X = 1)P(Y = 4)$$

so the random variables are not independent. In general independence fails if there is a 0 in the table where the row and column sums are positive. This simple observation takes care of all of our examples so to further explore the concept we need a new one.

X	Y=0	1	
0	0.4	0.3	0.7
1	0.2	0.1	0.3
	0.6	0.4	

There is no zero in the table but it fails the independence test:

$$P(X = 0, Y = 0) = 0.4 \neq 0.42 = (0.7)(0.6) = (X = 0)P(Y = 0)$$

If we want to have independent random variables with these marginal distributions there is only one way to fill in the table.

X	Y=0	1	
0	0.42	0.28	0.7
1	0.18	0.12	0.3
	0.6	0.4	

Example 3.28. This example gives a remarkable property of the Poisson distribution. Let $A_1, \ldots, A_k$ be disjoint events whose union $\cup_{i=1}^k A_i = \Omega$. Suppose we perform the experiment a random number of times N, where N has a Poisson distribution with mean λ, and let X_i be the number of times A_i occurs.

If $n = x_1 + \cdots + x_k$, then recalling the formula for the multinomial distribution (Example 1.9 in Chapter 2),

$$P(X_i = x_i \text{ for } 1 \le i \le k) = e^{-\lambda} \frac{\lambda^n}{n!} \frac{n!}{x_1! \cdots x_k!} P(A_1)^{x_1} \cdots P(A_k)^{x_k}$$
$$= e^{-\lambda P(A_1)} \frac{(\lambda P(A_1))^{x_1}}{x_1!} \cdots e^{-\lambda P(A_k)} \frac{(\lambda P(A_k))^{x_k}}{x_k!}$$

since $\sum_{i=1}^k P(A_i) = 1$. In words, $X_1, \ldots, X_k$ are independent Poissons with parameters $\lambda P(A_i)$.

To see why this is surprising, consider the special case $k = 2$, i.e., $A_2 = A_1^c$. If we performed our experiment a fixed number of times then N_1 and N_2 would not be independent since $N_2 = n - N_1$. It is remarkable that when we perform our experiment a Poisson number of times, the number of successes tells us nothing about the number of failures. This result is not only surprising but also useful. For a concrete example, suppose that a Poisson number of cars arrive at a fast food restaurant each hour and let A_i be the event that the car has i passengers. Then the number of cars with i passengers that arrive are independent Poissons.

Conditional Distribution

For discrete random variables, the definition of conditional probability implies

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(X = x, Y = y)}{\sum_u P(X = u, Y = y)} \tag{3.10}$$

If we fix y and look at $P(X = x|Y = y)$ as a function of x, what we have is the **conditional distribution of X given that** $Y = y$.

Example 3.29. To illustrate this formula we look at our AP calculus example:

X	Y= 4	3	2	
5	.10	.05	0	.15
4	.15	.15	0	.30
3	.10	.15	.10	.35
2	0	.05	.10	.15
1	0	0	.05	.05
	.35	.40	.25	

It follows from the definition of conditional probability that

$$P(X = 5|Y = 4) = P(X = 5, Y = 4)/P(Y = 4) = 0.10/0.35 = 2/7$$
$$P(X = 4|Y = 4) = P(X = 4, Y = 4)/P(Y = 4) = 0.15/0.35 = 3/7$$
$$P(X = 3|Y = 4) = P(X = 3, Y = 4)/P(Y = 4) = 0.10/0.35 = 2/7$$

In words, 2/7's of the students who get A's in the course get a 5 on the exam, 3/7's get a 4, and 2/7's get a 3. Operationally, we divide the entries in the second column by their sum to turn them into a probability distribution. We leave it to the reader to check

x	5	4	3	2	1
$P(X=x\|Y=3)$	1/8	3/8	3/8	1/8	0
$P(X=x\|Y=2)$	0	0	2/5	2/5	1/5

Conditional Expectation

The conditional expectation is mean of the conditional distribution

$$E(X|Y=y) = \sum_x xP(X=x|Y=y)$$

In the previous example

$$\begin{aligned} E(X|Y=4) &= (5(.10)+4(.15)+3(.10))/.35 = 4 \\ E(X|Y=3) &= (5(.05)+4(.15)+3(.15)+2(.05)/.40 = 3.5 \\ E(X|Y=2) &= (3(.10)+2(.10)+1(.05))/.25 = 2.2 \end{aligned}$$

Simpson's paradox is the phenomenon that means of subgroups can show much different patterns than the eman of the group as a whole. For a real life example, consider the average SAT verbal score. It was 504 in 1981 and 21 years later in 2002 it was again 504. However when we break things down by ethnic groups, we see that all of them increased their scores:

	1981	2002
non-hispanic whites	519	527
African Americans	412	431
Mexican Americans	438	446
Asian Americans	474	501

The explanation is simple: minorities made up a much larger portion of the testing population in 2002 than in 1981, and although they have shown significant improvement their averages are lower than non-hispanic whites which reduces the overall mean.

3.5 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. 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?

6. 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?

7. 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?

8. 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?

9. 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?

10. 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?

11. 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.

12. 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?

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

14. 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 $P(A \cup B \cup C)$.

Two-stage experiments

15. 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?

16. 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 more ($>$) than three?

17. 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?

18. Suppose we draw k cards out of a deck. What is the probability that we do not draw an Ace? Is the answer larger or smaller than $(3/4)^k$?

19. You and a friend each roll two dice. What is the probability you will both have the same two numbers?

20. 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?

21. What is the most likely total for the sum of four dice and what is its probability?

22. 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?

23. 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.

24. 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?

25. 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?

26. 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?

27. 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.

28. 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.

29. 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?

30. 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?

31. 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?

32. 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.

33. Change the second and third probabilities in the last problem so that each boy has an equal chance of winning.

Bayes' formula

34. 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?

35. 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?

36. 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?

37. 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?

38. Two hunters shoot at a deer, which is hit by exactly one bullet. If the first hunter hits his targets with probability 0.3 and the second with probability 0.6, what is the probability the second hunter killed the deer? The answer is not 2/3. Do you think the answer is larger or smaller?

39. A cab was involved in a hit and run accident at night. Two cab companies green and blue operate 85% and 15% of the cabs in the city respectively. A witness identified the cab as blue. However, in a test only 80% of witnesses were able to correctly identify the cab color. Given this what is the probability that the cab involved in the accident was blue?

40. 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?

41. 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?

42. 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?

43. 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?

44. 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?

45. 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?

46. 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?

47. A student is taking a multiple choice exam in which each question has four possible answers. She knows the answers to 60% of the questions and guesses at the others. What is the probability she guessed given that she got question #12 right?

48. 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 he 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?

49. 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?

50. 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?

51. A company gives a test to 100 salesmen, 80 with good sales records and 20 with poor records. 60% of the good salesmen pass the test, but only 30% of the poor salesmen do. A new applicant takes the test and passes. What is the probability he is a good salesman?

52. 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?

53. Two masked robbers try to rob a crowded bank during the lunch hour but the teller presses a button that sets off an alarm and locks the front door. The robbers, realizing they are trapped, throw away their masks and disappear into the chaotic crowd. Confronted with 40 people claiming they are innocent, the police give everyone a lie detector test. Suppose that guilty people are detected with probability 0.95, and innocent people appear to be guilty with probability

0.01. What is the probability Mr. Jones is guilty given that the lie detector says he is?

54. 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?

55. 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 she is Liberal?

56. An undergraduate student has asked a professor for a letter of recommendation. He estimates that the probability he will get the job is 0.8 with a strong letter, 0.4 with a medium letter, and 0.1 with a weak letter. He also believes that the probabilities that the letter will be strong, medium, or weak are 0.5, 0.3, and 0.2. What is the probability that the letter was strong given that he got the job.

57. 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?

58. 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?

59. 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.

Joint distributions

60. Suppose we draw two tickets from a hat that contains tickets numbered 1,2,3,4. Let X be the first number drawn and Y be the second. Find the joint distribution of X and Y.

61. Suppose we roll one die repeatedly and let N_i be the number of the roll on which i first appears. Find the joint distribution of N_1 and N_6.

62. Compute (a) $P(X = 1|Y = 1)$, (b) $P(X = 2|Y = 2)$ for the following joint distribution:

Y	X=1	2	3
1	.1	.2	.3
2	.15	.15	0
3	.05	0	.05

63. Compute (a) $P(X = 2|Y = 3)$, (b) $P(Y = 3|X = 3)$ for the following joint distribution

Y	X=1	2	3
1	.2	.15	.05
2	.10	0	.10
3	.05	.15	.20

64. Using the clues given below, fill in the rest of the joint distribution. There is only one answer.

Y	X=0	3	6
1	?	?	?
2	.1	.05	?

(a) $P(Y = 2|X = 0) = 1/4$, (b) X and Y are independent.

65. Using the clues given below, fill in the rest of the joint distribution. There is only one answer.

Y	X=1	2	3
1	?	?	?
2	?	0	?
3	0	?	0

For $k = 1, 2, 3$, (a) $P(Y = 1|X = k) = 2/3$, (b) $P(X = k|Y = 1) = k/6$.

66. Fill in the rest of the joint distribution so that X and Y are independent. There are two possible answers

Y	X=0	1
0	?	2/9
1	2/9	?