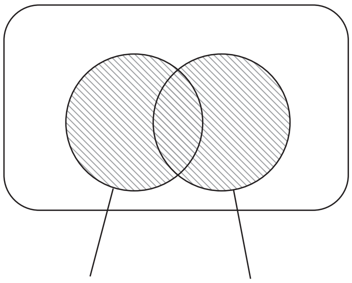
CHAPTER 1

1.1 Using the same Venn diagram for illustration, we want the probability of outcomes from the two events that lead to the cross-hatched area shown below:



This represents getting *A* in event 1 and not *B* in event 2, plus not getting *A* in event 1 but getting *B* in event 2 (these two are the common "or but not both" combination calculated in Problem 1.2) plus getting *A* in event 1 and *B* in event 2.

1.2 First the formula will be derived using equations, and then Venn diagrams will be compared with the steps in the equation. In terms of formulas and probabilities, there are two ways that the desired pair of outcomes can come about. One way is that we could get *A* on the first event and not *B* on the second ($A_1 \cap (\sim B_2)$). The probability of this is taken as the simple product, since events 1 and 2 are independent:

$$\begin{aligned} p_{A_1 \cap (\sim B_2)} &= p_A \times p_{\sim B} \\ &= p_A \times (1 - p_B) \\ &= p_A - p_A p_B \end{aligned} \tag{A.1.1}$$

The second way is that we could not get *A* on the first event and we could get *B* on the second ($(\sim A_1) \cap B_2$), with probability

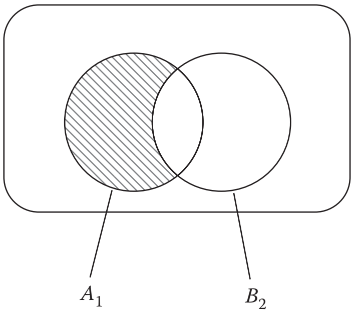
$$\begin{aligned} p_{(\sim A_1) \cap B_2} &= p_{\sim A} \times p_B \\ &= (1 - p_A) \times p_B \\ &= p_B - p_A p_B \end{aligned} \tag{A.1.2}$$

Since either one will work, we want the or combination. Because the two ways are mutually exclusive (having both would mean both A and $\sim A$ in the first outcome, and with equal impossibility, both B and $\sim B$), this or combination is equal to the union $\{A_1 \cap (\sim B_2)\} \cup \{(\sim A_1) \cap B_2\}$, and its probability is simply the sum of the probability of the two separate ways above (Equations A.1.1 and A.1.2):

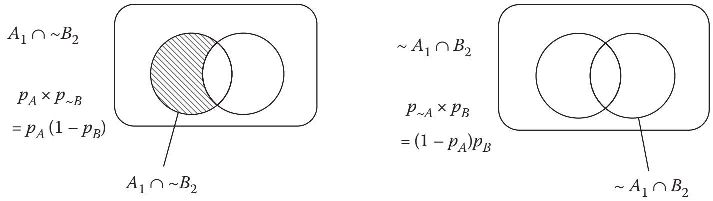
$$\begin{aligned} p_{\{A_1 \cap (\sim B_2)\} \cup \{(\sim A_1) \cap B_2\}} &= p_{A_1 \cap (\sim B_2)} + p_{(\sim A_1) \cap B_2} \\ &= p_A - p_A p_B + p_B - p_A p_B \\ &= p_A + p_B - 2 p_A p_B \end{aligned}$$

The connection to Venn diagrams is shown below. In this exercise we will work backward from the combination of outcomes we seek to the individual outcomes. The probability we are after is for the cross-hatched area below.

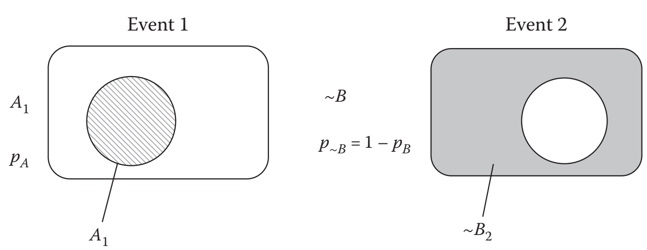
$\{A_1 \cap (\sim B_2)\} \cup \{(\sim A_1) \cap B_2\}$



As indicated, the circles correspond to getting the outcome A in event 1 (left) and outcome B in event 2. Even though the events are identical, the Venn diagram is constructed so that there is some overlap between these two (which we don't want to include in our "or but not both" combination. As described above, the two cross-hatched areas above don't overlap, thus the probability of their union is the simple sum of the two separate areas given below.



Adding these two probabilities gives the full "or but not both" expression above. The only thing remaining is to show that the probability of each of the crescents is equal to the product of the probabilities as shown in the top diagram. This will only be done for one of the two crescents, since the other follows in an exactly analogous way. Focusing on the gray crescent above, it represents the A outcomes of event 1 and not the B outcomes in event 2. Each of these outcomes is shown below:



Because Event 1 and Event 2 are independent, the “and” combination of these two outcomes is given by the intersection, and the probability of the intersection is given by the product of the two separate probabilities, leading to the expressions for probabilities for the gray cross-hatched crescent.

(a) These are two independent elementary events each with an outcome probability of 0.5. We are asked for the probability of the sequence $H_1 T_2$, which requires multiplication of the elementary probabilities:

$$p_{H_1H_2} = H_1 \cap T_2 = p_{H_1} \times p_{T_2} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

We can arrange this probability, along with the probability for the other three possible sequences, in a table:

	Toss 1	
Toss 2	*H* (0.5)	*T* (0.5)
H (0.5)	H_1H_2 (0.25)	T_1H_2 (0.25)
T (0.5)	H_1T_2 (0.25)	T_1T_2 (0.25)

Note: Probabilities are given in parentheses.

The probability of getting a head on the first toss or a tail on the second toss, but not both, is

$$\begin{aligned} p_{H_1 \text{ or } H_2} &= p_{H_1} + p_{H_2} - 2(p_{H_1} \times p_{H_2}) \\ &= \frac{1}{2} + \frac{1}{2} - 2\left(\frac{1}{2} \times \frac{1}{2}\right) \\ &= \frac{1}{2} \end{aligned}$$

In the table above, this combination corresponds to the sum of the two off-diagonal elements (the H_1T_2 and the T_1H_2 boxes).

(b) This is the "and" combination for independent events, so we multiply the elementary probability p_H for each of N tosses:

$$\begin{aligned} p_{H_1H_2H_3\ldots H_N} &= p_{H_1} \times p_{H_2} \times p_{H_3} \times \cdots \times p_{H_N} \\ &= \left(\frac{1}{2}\right)^N \end{aligned}$$

This is both a permutation and a composition (there is only one permutation for all-heads). And note that since both outcomes have equal probability (0.5), this gives the probability of any permutation of any number N_H of heads with any number $N - N_H$ of tails.

1.3 Two different approaches will be given for this problem. One is an approximation that is very close to being correct. The second is exact. By comparing the results, the reasonableness of the first approximation can be examined.

Whichever approach we use to solve this problem, we begin by representing the probability that you know a randomly selected person from the population. This is $p_k = 2000/300{,}000{,}000 = 2/300{,}000 = 6.67 \times 10^{-6}$. To avoid dealing with "or" combinations, we can greatly simplify the problem by calculating the probability that you do NOT know anyone on the plane, and then recognize that one minus this probability represents all the ways you could know at least one person (that is, you know only one, or you know only two, or… you know all of them). Working with the probability you do not know a given passenger (p_{nk}) converts a complicated "or" problem to a simple "and" problem. The probability you do not know an individual selected randomly from the population is $p_{nk} = p_{\sim k} = 1 - p_k = 0.9999933$ (nk for "not know"). The approximate to find the answer is to use the binomial distribution, and calculate the probability that you do not know randomly selected people. This is

$$\begin{aligned} p(200;200) &= W(200;200) \times p_{nk}{}^{200} \times (p_k)^{(200-200)} \\ &= \frac{200!}{200!(200-200)!} \times_A \left(1 - \frac{2000}{300{,}000{,}000}\right)^{200} \times (p_k)^{(0)} \\ &= \left(1 - \frac{2000}{300{,}000{,}000}\right)^{200} \\ &= 0.9986675507\ldots \end{aligned}$$

Thus, the chances you know somebody on the plane is about $1 - p(200;200) = 0.0013324492779\ldots$, a little better than one in one thousand chance (note, the rounding error in these calculations can be a big deal, especially when comparing to the exact result below). A bit of thought about the binomial distribution underscores the value of doing the calculation on people you don't know, rather than on the ones you do. There is only one composition the binomial distribution in which each person (besides you) is not known (200 nk's), and only one way to arrange the 200 nk's (hence the factorial term goes to one). If we had calculated probabilities for knowing people, we would have had calculate for 200 different compositions (1k, 2k, 3k,…200k) and add them all, and the numbers in the compositions in the middle of the distribution would involve calculating huge factorials, which is not straightforward.†

The reason the binomial approach is approximate is that it assumes that for each person, the probability is independent. Although this seems reasonable (and because of the large total population, the extent to which it is wrong is fairly small), it is not quite right. The probability of knowing the first one on the plane is still $p_k = 200/300{,}000{,}000$. But if you do not know the first person on the plane, the probability you know the second is 2000/299,999,999; you still know 2000 people outside the plane, but there are no longer 300,000,000 left. And if you do not know the first two people on the plane, the probability of knowing the third is 2000/299,999,998,… and if you didn't know the first 199 people on the plane, the probability you don't know the last person on (200th) is 2000/299,999,801. In general, this relationship is

$$p_{k,i} = \frac{2000}{300{,}000{,}000 - i + 1}$$

where $p_{k,i}$ is the probability that you know the ith person on the plane, given you did not know any of the previous passengers aboard. And the probability you don't know the ith person, given you didn't know any of the previous passengers ($p_{k,i}$), is one minus this:

$$p_{nk,i} = 1 - \frac{2000}{300{,}000{,}000 - i + 1}$$

The total probability you don't know anyone on the plane is the product of these dependent probabilities from the first person to the 200th, that is,

† This can be done, but it requires Stirling's approximation, to be introduced in Chapter 4.