2

Conditional probability

Conditioning on evidence

1. ⓢ A spam filter is designed by looking at commonly occurring phrases in spam. Suppose that 80% of email is spam. In 10% of the spam emails, the phrase "free money" is used, whereas this phrase is only used in 1% of non-spam emails. A new email has just arrived, which does mention "free money". What is the probability that it is spam?

   Solution: Let S be the event that an email is spam and F be the event that an email has the "free money" phrase. By Bayes' rule,

   $$P(S|F) = \frac{P(F|S)P(S)}{P(F)} = \frac{0.1 \cdot 0.8}{0.1 \cdot 0.8 + 0.01 \cdot 0.2} = \frac{80/1000}{82/1000} = \frac{80}{82} \approx 0.9756.$$

2. ⓢ A woman is pregnant with twin boys. Twins may be either identical or fraternal (non-identical). In general, 1/3 of twins born are identical. Obviously, identical twins must be of the same sex; fraternal twins may or may not be. Assume that identical twins are equally likely to be both boys or both girls, while for fraternal twins all possibilities are equally likely. Given the above information, what is the probability that the woman's twins are identical?

   Solution: By Bayes' rule,

   $$P(\text{identical}|BB) = \frac{P(BB|\text{identical})P(\text{identical})}{P(BB)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3}} = 1/2.$$

3. According to the CDC (Centers for Disease Control and Prevention), men who smoke are 23 times more likely to develop lung cancer than men who don't smoke. Also according to the CDC, 21.6% of men in the U.S. smoke. What is the probability that a man in the U.S. is a smoker, given that he develops lung cancer?

   Solution: Let S be the event that a man in the U.S. smokes and L be the event that he gets lung cancer. We are given that $P(S) = 0.216$ and $P(L|S) = 23P(L|S^c)$. By Bayes' rule and the law of total probability, we have

   $$P(S|L) = \frac{P(L|S)P(S)}{P(L|S)P(S) + P(L|S^c)P(S^c)} = \frac{P(L|S)P(S)}{P(L|S)P(S) + \frac{1}{23}P(L|S)P(S^c)}.$$

   We don't know $P(L|S)$, but it cancels out! Thus,

   $$P(S|L) = \frac{0.216}{0.216 + (1 - 0.216)/23} \approx 0.864.$$

4. Fred is answering a multiple-choice problem on an exam, and has to choose one of n options (exactly one of which is correct). Let K be the event that he knows the answer, and R be the event that he gets the problem right (either through knowledge or through luck). Suppose that if he knows the right answer he will definitely get the problem right, but if he does not know then he will guess completely randomly. Let $P(K) = p$.

(a) Find $P(K|R)$ (in terms of p and n).

(b) Show that $P(K|R) \geq p$, and explain why this makes sense intuitively. When (if ever) does $P(K|R)$ equal p?

Solution:

(a) By Bayes' rule and the law of total probability,

$$P(K|R) = \frac{P(R|K)P(K)}{P(R|K)P(K) + P(R|K^c)P(K^c)} = \frac{p}{p + (1-p)/n}.$$

(b) For the extreme case $p = 0$, we have $P(K|R) = 0 = p$. So assume $p > 0$. By the result of (a), $P(K|R) \geq p$ is equivalent to $p + (1-p)/n \leq 1$, which is a true statement since $p + (1-p)/n \leq p + 1 - p = 1$. This makes sense intuitively since getting the question right should increase our confidence that Fred knows the answer. Equality holds if and only if one of the extreme cases $n = 1$, $p = 0$, or $p = 1$ holds. If $n = 1$, it's not really a multiple-choice problem, and Fred getting the problem right is completely uninformative; if $p = 0$ or $p = 1$, then whether Fred knows the answer is a foregone conclusion, and no evidence will make us more (or less) sure that Fred knows the answer.

5. Three cards are dealt from a standard, well-shuffled deck. The first two cards are flipped over, revealing the Ace of Spades as the first card and the 8 of Clubs as the second card. Given this information, find the probability that the third card is an ace in two ways: using the definition of conditional probability, and by symmetry.

Solution: Let A be the event that the first card is the Ace of Spades, B be the event that the second card is the 8 of Clubs, and C be the event that the third card is an ace. By definition of conditional probability,

$$P(C|A, B) = \frac{P(C, A, B)}{P(A, B)} = \frac{P(A, B, C)}{P(A, B)}.$$

By the naive definition of probability,

$$P(A, B) = \frac{50!}{52!} = \frac{1}{51 \cdot 52}$$

and

$$P(A, B, C) = \frac{3 \cdot 49!}{52!} = \frac{3}{50 \cdot 51 \cdot 52}.$$

So $P(C|A, B) = 3/50$.

A simpler way is to see this is to use symmetry directly. Given the evidence, the third card is equally likely to be any card other than the Ace of Spades or 8 of Clubs, so it has probability $3/50$ of being an ace.

6. A hat contains 100 coins, where 99 are fair but one is double-headed (always landing Heads). A coin is chosen uniformly at random. The chosen coin is flipped 7 times, and it lands Heads all 7 times. Given this information, what is the probability that the chosen coin is double-headed? (Of course, another approach here would be to *look at both sides of the coin*—but this is a metaphorical coin.)

Solution: Let A be the event that the chosen coin lands Heads all 7 times, and B be the event that the chosen coin is double-headed. Then

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \frac{0.01}{0.01 + (1/2)^7 \cdot 0.99} = \frac{128}{227} \approx 0.564.$$

7. A hat contains 100 coins, where *at least* 99 are fair, but there may be one that is double-headed (always landing Heads); if there is no such coin, then all 100 are fair. Let D be the event that there is such a coin, and suppose that $P(D) = 1/2$. A coin is chosen uniformly at random. The chosen coin is flipped 7 times, and it lands Heads all 7 times.

   (a) Given this information, what is the probability that one of the coins is double-headed?

   (b) Given this information, what is the probability that the chosen coin is double-headed?

   Solution:

   (a) Let A be the event that the chosen coin lands Heads all 7 times, and C be the event that the chosen coin is double-headed. By Bayes' rule and LOTP,

   $$P(D|A) = \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)}.$$

   We have $P(D) = P(D^c) = 1/2$ and $P(A|D^c) = 1/2^7$, so the only remaining ingredient that we need to find is $P(A|D)$. We can do this using LOTP with extra conditioning (it would be useful to know whether the *chosen* coin is double-headed, not just whether *somewhere* there is a double-headed coin, so we condition on whether or not C occurs):

   $$P(A|D) = P(A|D,C)P(C|D) + P(A|D,C^c)P(C^c|D) = \frac{1}{100} + \frac{1}{2^7} \cdot \frac{99}{100}.$$

   Plugging in these results, we have

   $$P(D|A) = \frac{227}{327} = 0.694.$$

   (b) By LOTP with extra conditioning (it would be useful to know whether there *is* a double-headed coin),

   $$P(C|A) = P(C|A,D)P(D|A) + P(C|A,D^c)P(D^c|A),$$

   with notation as in (a). But $P(C|A,D^c) = 0$, and we already found $P(D|A)$ in (a). Also, $P(C|A,D) = \frac{128}{227}$, as shown in Exercise 6 (conditioning on D and A puts us exactly in the setup of that exercise). Thus,

   $$P(C|A) = \frac{128}{227} \cdot \frac{227}{327} = \frac{128}{327} \approx 0.391.$$

8. The screens used for a certain type of cell phone are manufactured by 3 companies, A, B, and C. The proportions of screens supplied by A, B, and C are 0.5, 0.3, and 0.2, respectively, and their screens are defective with probabilities 0.01, 0.02, and 0.03, respectively. Given that the screen on such a phone is defective, what is the probability that Company A manufactured it?

   Solution: Let $A, B,$ and C be the events that the screen was manufactured by Company A, B, and C, respectively, and let D be the event that the screen is defective. By Bayes' rule and LOTP,

   $$\begin{aligned} P(A|D) &= \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)} \\ &= \frac{0.01 \cdot 0.5}{0.01 \cdot 0.5 + 0.02 \cdot 0.3 + 0.03 \cdot 0.2} \\ &\approx 0.294. \end{aligned}$$

9. (a) Show that if events A_1 and A_2 have the same *prior* probability $P(A_1) = P(A_2)$, A_1 implies B, and A_2 implies B, then A_1 and A_2 have the same *posterior* probability $P(A_1|B) = P(A_2|B)$ if it is observed that B occurred.

(b) Explain why (a) makes sense intuitively, and give a concrete example.

Solution:

(a) Suppose that $P(A_1) = P(A_2)$, A_1 implies B, and A_2 implies B. Then

$$P(A_1|B) = \frac{P(A_1, B)}{P(B)} = \frac{P(A_1)}{P(B)} = \frac{P(A_2)}{P(B)} = \frac{P(A_2, B)}{P(B)} = P(A_2|B).$$

(b) The result in (a) makes sense intuitively since, thinking in terms of Pebble World, observing that B occurred entails restricting the sample space by removing the pebbles in B^c. But none of the removed pebbles are in A_1 or in A_2, so the updated probabilities for A_1 and A_2 are just rescaled versions of the original probabilities, scaled by a constant chosen to make the total mass 1.

For a simple example, let A_1 be the event that the top card in a well-shuffled standard deck is a diamond, let A_2 be the event that it is a heart, and let B be the event that it is a red card. Then $P(A_1) = P(A_2) = 1/4$ and $P(A_1|B) = P(A_2|B) = 1/2$.

10. Fred is working on a major project. In planning the project, two milestones are set up, with dates by which they should be accomplished. This serves as a way to track Fred's progress. Let A_1 be the event that Fred completes the first milestone on time, A_2 be the event that he completes the second milestone on time, and A_3 be the event that he completes the project on time.

Suppose that $P(A_{j+1}|A_j) = 0.8$ but $P(A_{j+1}|A_j^c) = 0.3$ for $j = 1, 2$, since if Fred falls behind on his schedule it will be hard for him to get caught up. Also, assume that the second milestone supersedes the first, in the sense that once we know whether he is on time in completing the second milestone, it no longer matters what happened with the first milestone. We can express this by saying that A_1 and A_3 are conditionally independent given A_2 and they're also conditionally independent given A_2^c.

(a) Find the probability that Fred will finish the project on time, given that he completes the first milestone on time. Also find the probability that Fred will finish the project on time, given that he is late for the first milestone.

(b) Suppose that $P(A_1) = 0.75$. Find the probability that Fred will finish the project on time.

Solution:

(a) We need to find $P(A_3|A_1)$ and $P(A_3|A_1^c)$. To do so, let's use LOTP to condition on whether or not A_2 occurs:

$$P(A_3|A_1) = P(A_3|A_1, A_2)P(A_2|A_1) + P(A_3|A_1, A_2^c)P(A_2^c|A_1).$$

Using the conditional independence assumptions, this becomes

$$P(A_3|A_2)P(A_2|A_1) + P(A_3|A_2^c)P(A_2^c|A_1) = (0.8)(0.8) + (0.3)(0.2) = 0.7.$$

Similarly,

$$P(A_3|A_1^c) = P(A_3|A_2)P(A_2|A_1^c) + P(A_3|A_2^c)P(A_2^c|A_1^c) = (0.8)(0.3) + (0.3)(0.7) = 0.45.$$

(b) By LOTP and Part (a),

$$P(A_3) = P(A_3|A_1)P(A_1) + P(A_3|A_1^c)P(A_1^c) = (0.7)(0.75) + (0.45)(0.25) = 0.6375.$$

11. An *exit poll* in an election is a survey taken of voters just after they have voted. One major use of exit polls has been so that news organizations can try to figure out as soon as possible who won the election, before the votes are officially counted. This has been notoriously inaccurate in various elections, sometimes because of *selection bias*: the sample of people who are invited to and agree to participate in the survey may not be similar enough to the overall population of voters.

Consider an election with two candidates, Candidate A and Candidate B. Every voter is invited to participate in an exit poll, where they are asked whom they voted for; some accept and some refuse. For a randomly selected voter, let A be the event that they voted for A, and W be the event that they are willing to participate in the exit poll. Suppose that $P(W|A) = 0.7$ but $P(W|A^c) = 0.3$. In the exit poll, 60% of the respondents say they voted for A (assume that they are all honest), suggesting a comfortable victory for A. Find $P(A)$, the true proportion of people who voted for A.

Solution: We have $P(A|W) = 0.6$ since 60% of the respondents voted for A. Let $p = P(A)$. Then

$$0.6 = P(A|W) = \frac{P(W|A)P(A)}{P(W|A)P(A) + P(W|A^c)P(A^c)} = \frac{0.7p}{0.7p + 0.3(1-p)}.$$

Solving for p, we obtain

$$P(A) = \frac{9}{23} \approx 0.391.$$

So actually A received fewer than half of the votes!

12. Alice is trying to communicate with Bob, by sending a message (encoded in binary) across a channel.

(a) Suppose for this part that she sends only one bit (a 0 or 1), with equal probabilities. If she sends a 0, there is a 5% chance of an error occurring, resulting in Bob receiving a 1; if she sends a 1, there is a 10% chance of an error occurring, resulting in Bob receiving a 0. Given that Bob receives a 1, what is the probability that Alice actually sent a 1?

(b) To reduce the chance of miscommunication, Alice and Bob decide to use a *repetition code*. Again Alice wants to convey a 0 or a 1, but this time she repeats it two more times, so that she sends 000 to convey 0 and 111 to convey 1. Bob will decode the message by going with what the majority of the bits were. Assume that the error probabilities are as in (a), with error events for different bits independent of each other. Given that Bob receives 110, what is the probability that Alice intended to convey a 1?

Solution:

(a) Let A_1 be the event that Alice sent a 1, and B_1 be the event that Bob receives a 1. Then

$$P(A_1|B_1) = \frac{P(B_1|A_1)P(A_1)}{P(B_1|A_1)P(A_1) + P(B_1|A_1^c)P(A_1^c)} = \frac{(0.9)(0.5)}{(0.9)(0.5) + (0.05)(0.5)} \approx 0.9474.$$

(b) Now let A_1 be the event that Alice intended to convey a 1, and B_{110} be the event that Bob receives 110. Then

$$
\begin{aligned}
P(A_1|B_{110}) &= \frac{P(B_{110}|A_1)P(A_1)}{P(B_{110}|A_1)P(A_1) + P(B_{110}|A_1^c)P(A_1^c)} \\
&= \frac{(0.9 \cdot 0.9 \cdot 0.1)(0.5)}{(0.9 \cdot 0.9 \cdot 0.1)(0.5) + (0.05 \cdot 0.05 \cdot 0.95)(0.5)} \\
&\approx 0.9715.
\end{aligned}
$$

13. Company A has just developed a diagnostic test for a certain disease. The disease afflicts 1% of the population. As defined in Example 2.3.9, the *sensitivity* of the test is the probability of someone testing positive, given that they have the disease, and the *specificity* of the test is the probability that of someone testing negative, given that they don't have the disease. Assume that, as in Example 2.3.9, the sensitivity and specificity are both 0.95.

Company B, which is a rival of Company A, offers a competing test for the disease. Company B claims that their test is faster and less expensive to perform than Company A's test, is less painful (Company A's test requires an incision), and yet has a higher overall success rate, where overall success rate is defined as the probability that a random person gets diagnosed correctly.

(a) It turns out that Company B's test can be described and performed very simply: no matter who the patient is, diagnose that they do not have the disease. Check whether Company B's claim about overall success rates is true.

(b) Explain why Company A's test may still be useful.

(c) Company A wants to develop a new test such that the overall success rate is higher than that of Company B's test. If the sensitivity and specificity are equal, how high does the sensitivity have to be to achieve their goal? If (amazingly) they can get the sensitivity equal to 1, how high does the specificity have to be to achieve their goal? If (amazingly) they can get the specificity equal to 1, how high does the sensitivity have to be to achieve their goal?

Solution:

(a) For Company B's test, the probability that a random person in the population is diagnosed correctly is 0.99, since 99% of the people do not have the disease. For a random member of the population, let C be the event that Company A's test yields the correct result, T be the event of testing positive in Company A's test, and D be the event of having the disease. Then

$$\begin{aligned} P(C) &= P(C|D)P(D) + P(C|D^c)P(D^c) \\ &= P(T|D)P(D) + P(T^c|D^c)P(D^c) \\ &= (0.95)(0.01) + (0.95)(0.99) \\ &= 0.95, \end{aligned}$$

which makes sense intuitively since the sensitivity and specificity of Company A's test are both 0.95. So Company B is correct about having a higher overall success rate.

(b) Despite the result of (a), Company A's test may still provide very useful information, whereas Company B's test is uninformative. If Fred tests positive on Company A's test, Example 2.3.9 shows that his probability of having the disease increases from 0.01 to 0.16 (so it is still fairly unlikely that he has the disease, but it is much more likely than it was before the test result; further testing may well be advisable). In contrast, Fred's probability of having the disease does not change after undergoing Company's B test, since the test result is a foregone conclusion.

(c) Let s be the sensitivity and p be the specificity of A's new test. With notation as in the solution to (a), we have

$$P(C) = 0.01s + 0.99p.$$

If $s = p$, then $P(C) = s$, so Company A needs $s > 0.99$.

If $s = 1$, then $P(C) = 0.01 + 0.99p > 0.99$ if $p > 98/99 \approx 0.9899$.

If $p = 1$, then $P(C) = 0.01s + 0.99$ is automatically greater than 0.99 (unless $s = 0$, in which case both companies have tests with sensitivity 0 and specificity 1).

14. Consider the following scenario, from Tversky and Kahneman:

> Let A be the event that before the end of next year, Peter will have installed a burglar alarm system in his home. Let B denote the event that Peter's home will be burglarized before the end of next year.

(a) Intuitively, which do you think is bigger, $P(A|B)$ or $P(A|B^c)$? Explain your intuition.

(b) Intuitively, which do you think is bigger, $P(B|A)$ or $P(B|A^c)$? Explain your intuition.

(c) Show that for *any* events A and B (with probabilities not equal to 0 or 1), $P(A|B) > P(A|B^c)$ is equivalent to $P(B|A) > P(B|A^c)$.

(d) Tversky and Kahneman report that 131 out of 162 people whom they posed (a) and (b) to said that $P(A|B) > P(A|B^c)$ and $P(B|A) < P(B|A^c)$. What is a plausible explanation for why this was such a popular opinion despite (c) showing that it is impossible for these inequalities both to hold?

Solution:

(a) Intuitively, $P(A|B)$ seems larger than $P(A|B^c)$ since if Peter's home is burglarized, he is likely to take increased precautions (such as installing an alarm) against future attempted burglaries.

(b) Intuitively, $P(B|A^c)$ seems larger than $P(B|A)$, since presumably having an alarm system in place deters prospective burglars from attempting a burglary and hampers their chances of being able to burglarize the home. However, this is in conflict with (a), according to (c). Alternatively, we could argue that $P(B|A)$ should be larger than $P(B|A^c)$, since observing that an alarm system is in place could be evidence that the neighborhood has frequent burglaries.

(c) First note that $P(A|B) > P(A|B^c)$ is equivalent to $P(A|B) > P(A)$, since LOTP says that $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$ is between $P(A|B)$ and $P(A|B^c)$ (in words, $P(A)$ is a weighted average of $P(A|B)$ and $P(A|B^c)$). But $P(A|B) > P(A)$ is equivalent to $P(A, B) > P(A)P(B)$, by definition of conditional probability. Likewise, $P(B|A) > P(B|A^c)$ is equivalent to $P(B|A) > P(B)$, which in turn is equivalent to $P(A, B) > P(A)P(B)$.

(d) It is reasonable to assume that a burglary at his home might cause Peter to install an alarm system and that having an alarm systems might reduce the chance of a future burglary. People with inconsistent beliefs about (a) and (b) may be thinking intuitively in causal terms, interpreting a probability $P(D|C)$ in terms of C causing D. But the definition of $P(D|C)$ does not invoke causality and does not require C's occurrence to precede D's occurrence or non-occurrence temporally.

15. Let A and B be events with $0 < P(A \cap B) < P(A) < P(B) < P(A \cup B) < 1$. You are hoping that *both* A and B occurred. Which of the following pieces of information would you be happiest to observe: that A occurred, that B occurred, or that $A \cup B$ occurred?

Solution: If C is one of the events $A, B, A \cup B$, then

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

So among the three options for C, $P(A \cap B|C)$ is maximized when C is the event A.

16. Show that $P(A|B) \leq P(A)$ implies $P(A|B^c) \geq P(A)$, and give an intuitive explanation of why this makes sense.

Solution: By LOTP,

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

So $P(A)$ is between $P(A|B)$ and $P(A|B^c)$; it is a weighted average of these two conditional probabilities. To see this in more detail, let $x = \min(P(A|B), P(A|B^c)), y = \max(P(A|B), P(A|B^c))$. Then

$$P(A) \geq xP(B) + xP(B^c) = x$$

and

$$P(A) \leq yP(B) + yP(B^c) = y,$$

so $x \leq P(A) \leq y$. Therefore, if $P(A|B) \leq P(A)$, then $P(A) \leq P(A|B^c)$.

It makes sense intuitively that B and B^c should work in opposite directions as evidence regarding A. If both B and B^c were evidence in favor of A, then $P(A)$ should have already reflected this.

17. In deterministic logic, the statement "A implies B" is equivalent to its *contrapositive*, "not B implies not A". In this problem we will consider analogous statements in probability, the logic of uncertainty. Let A and B be events with probabilities not equal to 0 or 1.

(a) Show that if $P(B|A) = 1$, then $P(A^c|B^c) = 1$.

Hint: Apply Bayes' rule and LOTP.

(b) Show however that the result in (a) does not hold in general if $=$ is replaced by $\approx$. In particular, find an example where $P(B|A)$ is very close to 1 but $P(A^c|B^c)$ is very close to 0.

Hint: What happens if A and B are independent?

Solution:

(a) Let $P(B|A) = 1$. Then $P(B^c|A) = 0$. So by Bayes' rule and LOTP,

$$P(A^c|B^c) = \frac{P(B^c|A^c)P(A^c)}{P(B^c|A^c)P(A^c) + P(B^c|A)P(A)} = \frac{P(B^c|A^c)P(A^c)}{P(B^c|A^c)P(A^c)} = 1.$$

(b) For a simple counterexample if $=$ is replaced by $\approx$ in (a), let A and B be independent events with $P(A)$ and $P(B)$ both extremely close to 1. For example, this can be done in the context of flipping a coin 1000 times, where A is an extremely likely (but not certain) event based on the first 500 tosses and B is an extremely likely (but not certain) event based on the last 500 tosses. Then $P(B|A) = P(B) \approx 1$, but $P(A^c|B^c) = P(A^c) \approx 0$.

18. Show that if $P(A) = 1$, then $P(A|B) = 1$ for any B with $P(B) > 0$. Intuitively, this says that if someone dogmatically believes something with absolute certainty, then no amount of evidence will change their mind. The principle of avoiding assigning probabilities of 0 or 1 to any event (except for mathematical certainties) was named *Cromwell's rule* by the statistician Dennis Lindley, due to Cromwell saying to the Church of Scotland, "think it possible you may be mistaken".

Hint: Write $P(B) = P(B \cap A) + P(B \cap A^c)$, and then show that $P(B \cap A^c) = 0$.

Solution: Let $P(A) = 1$. Then $P(B \cap A^c) \leq P(A^c) = 0$ since $B \cap A^c \subseteq A^c$, which shows that $P(B \cap A^c) = 0$. So

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

Thus,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A \cap B)} = 1.$$

19. Explain the following Sherlock Holmes saying in terms of conditional probability, carefully distinguishing between prior and posterior probabilities: "It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth."

Solution: Let E be the observed evidence after a crime has taken place, and let $A_1, A_2, \ldots, A_n$ be an exhaustive list of events, any one of which (if it occurred) would serve as an explanation of how the crime occurred. Assuming that the list $A_1, \ldots, A_n$ exhausts all possible explanations for the crime, we have

$$P(A_1 \cup A_2 \cup \cdots \cup A_n | E) = 1.$$

Sherlock's maxim says that

$$P(A_n | E, A_1^c, A_1^c, \ldots, A_{n-1}^c) = 1,$$

i.e., if we have determined that all explanations other than A_n can be ruled out, then the remaining explanation, A_n, must be the truth, even if $P(A_n)$ and $P(A_n|E)$ are small. To prove Sherlock's maxim, note that

$$P(A_1^c, \ldots, A_{n-1}^c | E) = P(A_1^c, \ldots, A_{n-1}^c, A_n^c | E) + P(A_1^c, \ldots, A_{n-1}^c, A_n | E),$$

where the first term on the right-hand side is 0 by De Morgan's laws. So

$$P(A_n | E, A_1^c, A_1^c, \ldots, A_{n-1}^c) = \frac{P(A_1^c, A_1^c, \ldots, A_{n-1}^c, A_n | E)}{P(A_1^c, A_1^c, \ldots, A_{n-1}^c | E)} = 1.$$

20. The Jack of Spades (with cider), Jack of Hearts (with tarts), Queen of Spades (with a wink), and Queen of Hearts (without tarts) are taken from a deck of cards. These four cards are shuffled, and then two are dealt.

(a) Find the probability that both of these two cards are queens, given that the first card dealt is a queen.

(b) Find the probability that both are queens, given that at least one is a queen.

(c) Find the probability that both are queens, given that one is the Queen of Hearts.

Solution:

(a) Let Q_i be the event that the ith card dealt is a queen, for $i = 1, 2$. Then $P(Q_i) = 1/2$ since the ith card dealt is equally likely to be any of the cards. Also,

$$P(Q_1, Q_2) = P(Q_1)P(Q_2|Q_1) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

As a check, note that by the naive definition of probability,

$$P(Q_1, Q_2) = \frac{1}{\binom{4}{2}} = \frac{1}{6}.$$

Thus,

$$P(Q_1 \cap Q_2 | Q_1) = \frac{P(Q_1 \cap Q_2)}{P(Q_1)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

(b) Continuing as in (a),

$$P(Q_1 \cap Q_2 | Q_1 \cup Q_2) = \frac{P(Q_1 \cap Q_2)}{P(Q_1 \cup Q_2)} = \frac{P(Q_1 \cap Q_2)}{P(Q_1) + P(Q_2) - P(Q_1 \cap Q_2)} = \frac{\frac{1}{6}}{\frac{1}{2} + \frac{1}{2} - \frac{1}{6}} = \frac{1}{5}.$$

Another way to see this is to note that there are 6 possible 2-card hands, all equally

likely, of which 1 (the "double-jack pebble") is eliminated by our conditioning; then by definition of conditional probability, we are left with 5 "pebbles" of equal mass.

(c) Let H_i be the event that the ith card dealt is a heart, for $i = 1, 2$. Then

$$P(Q_1 \cap Q_2|(Q_1 \cap H_1) \cup (Q_2 \cap H_2)) = \frac{P(Q_1 \cap H_1 \cap Q_2) + P(Q_1 \cap Q_2 \cap H_2)}{P(Q_1 \cap H_1) + P(Q_2 \cap H_2)}$$
$$= \frac{\frac{1}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{3}}{\frac{1}{4} + \frac{1}{4}}$$
$$= \frac{1}{3},$$

using the fact that $Q_1 \cap H_1$ and $Q_2 \cap H_2$ are disjoint. Alternatively, note that the conditioning reduces the sample space down to 3 possibilities, which are equally likely, and 1 of the 3 has both cards queens.

21. A fair coin is flipped 3 times. The toss results are recorded on separate slips of paper (writing "H" if Heads and "T" if Tails), and the 3 slips of paper are thrown into a hat.

(a) Find the probability that all 3 tosses landed Heads, given that at least 2 were Heads.

(b) Two of the slips of paper are randomly drawn from the hat, and both show the letter H. Given this information, what is the probability that all 3 tosses landed Heads?

Solution:

(a) Let A be the event that all 3 tosses landed Heads, and B be the event that at least 2 landed Heads. Then

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(A)}{P(2 \text{ or } 3 \text{ Heads})} = \frac{1/8}{4/8} = \frac{1}{4}.$$

(b) Let C be the event that the two randomly chosen slips of paper show Heads. Then

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)}$$
$$= \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|2 \text{ Heads})P(2 \text{ Heads}) + P(C|1 \text{ or } 0 \text{ Heads})P(1 \text{ or } 0 \text{ Heads})}$$
$$= \frac{\frac{1}{8}}{\frac{1}{8} + \frac{1}{3} \cdot \frac{3}{8} + 0 \cdot \frac{1}{2}}$$
$$= \frac{1}{2}.$$

Alternatively, let A_i be the event that the ith toss was Heads. Note that

$$P(A|A_i, A_j) = \frac{P(A)}{P(A_i, A_j)} = \frac{1/8}{1/4} = \frac{1}{2}$$

for any $i \neq j$. Since this probability is $1/2$ regardless of which 2 slips of paper were drawn, conditioning on which 2 slips were drawn gives

$$P(A|C) = \frac{1}{2}.$$

22. Ⓢ A bag contains one marble which is either green or blue, with equal probabilities. A green marble is put in the bag (so there are 2 marbles now), and then a random marble is taken out. The marble taken out is green. What is the probability that the remaining marble is also green?

Solution: Let A be the event that the initial marble is green, B be the event that the

removed marble is green, and C be the event that the remaining marble is green. We need to find $P(C|B)$. There are several ways to find this; one natural way is to condition on whether the initial marble is green:

$$P(C|B) = P(C|B, A)P(A|B) + P(C|B, A^c)P(A^c|B) = 1P(A|B) + 0P(A^c|B).$$

To find $P(A|B)$, use Bayes' rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{1/2}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{1/2}{1/2 + 1/4} = \frac{2}{3}.$$

So $P(C|B) = 2/3$.

Historical note: This problem was first posed by Lewis Carroll in 1893.

23. Ⓢ Let G be the event that a certain individual is guilty of a certain robbery. In gathering evidence, it is learned that an event E_1 occurred, and a little later it is also learned that another event E_2 also occurred. Is it possible that individually, these pieces of evidence increase the chance of guilt (so $P(G|E_1) > P(G)$ and $P(G|E_2) > P(G)$), but together they decrease the chance of guilt (so $P(G|E_1, E_2) < P(G)$)?

Solution: Yes, this is possible. In fact, it is possible to have two events which separately provide evidence in favor of G, yet which together preclude G! For example, suppose that the crime was committed between 1 pm and 3 pm on a certain day. Let E_1 be the event that the suspect was at a specific nearby coffeeshop from 1 pm to 2 pm that day, and let E_2 be the event that the suspect was at the nearby coffeeshop from 2 pm to 3 pm that day. Then $P(G|E_1) > P(G), P(G|E_2) > P(G)$ (assuming that being in the vicinity helps show that the suspect had the opportunity to commit the crime), yet $P(G|E_1 \cap E_2) < P(G)$ (as being in the coffeehouse from 1 pm to 3 pm gives the suspect an alibi for the full time).

24. Is it possible to have events A_1, A_2, B, C with $P(A_1|B) > P(A_1|C)$ and $P(A_2|B) > P(A_2|C)$, yet $P(A_1 \cup A_2|B) < P(A_1 \cup A_2|C)$? If so, find an example (with a "story" interpreting the events, as well as giving specific numbers); otherwise, show that it is impossible for this phenomenon to happen.

Solution: Yes, this is possible. First note that $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$, so it is *not* possible if A_1 and A_2 are disjoint, and that it is crucial to consider the intersection. So let's choose examples where $P(A_1 \cap A_2|B)$ is much larger than $P(A_1 \cap A_2|C)$, to offset the other inequalities.

Story 1: Consider two basketball players, one of whom is randomly chosen to shoot two free throws. The first player is very streaky, and always either makes both or misses both free throws, with probability 0.8 of making both (this is an extreme example chosen for simplicity, but we could also make it so the player has good days (on which there is a high chance of making both shots) and bad days (on which there is a high chance of missing both shots) without requiring *always* making both or missing both). The second player's free throws go in with probability 0.7, independently. Define the events as A_j: the jth free throw goes in; B: the free throw shooter is the first player; $C = B^c$. Then

$$P(A_1|B) = P(A_2|B) = P(A_1 \cap A_2|B) = P(A_1 \cup A_2|B) = 0.8,$$

$$P(A_1|C) = P(A_2|C) = 0.7, P(A_1 \cap A_2|C) = 0.49, P(A_1 \cup A_2|C) = 2 \cdot 0.7 - 0.49 = 0.91.$$

Story 2: Suppose that you can either take Good Class or Other Class, but not both. If you take Good Class, you'll attend lecture 70% of the time, and you will understand the material if and only if you attend lecture. If you take Other Class, you'll attend lecture 40% of the time and understand the material 40% of the time, but because the class is so poorly taught, the only way you understand the material is by studying on your own

and not attending lecture. Defining the events as A_1: attend lecture; A_2: understand material; B: take Good Class; C: take Other Class,

$$P(A_1|B) = P(A_2|B) = P(A_1 \cap A_2|B) = P(A_1 \cup A_2|B) = 0.7,$$

$$P(A_1|C) = P(A_2|C) = 0.4, P(A_1 \cap A_2|C) = 0, P(A_1 \cup A_2|C) = 2 \cdot 0.4 = 0.8.$$

25. Ⓢ A crime is committed by one of two suspects, A and B. Initially, there is equal evidence against both of them. In further investigation at the crime scene, it is found that the guilty party had a blood type found in 10% of the population. Suspect A does match this blood type, whereas the blood type of Suspect B is unknown.

(a) Given this new information, what is the probability that A is the guilty party?

(b) Given this new information, what is the probability that B's blood type matches that found at the crime scene?

Solution:

(a) Let M be the event that A's blood type matches the guilty party's and for brevity, write A for "A is guilty" and B for "B is guilty". By Bayes' rule,

$$P(A|M) = \frac{P(M|A)P(A)}{P(M|A)P(A) + P(M|B)P(B)} = \frac{1/2}{1/2 + (1/10)(1/2)} = \frac{10}{11}.$$

(We have $P(M|B) = 1/10$ since, given that B is guilty, the probability that A's blood type matches the guilty party's is the same probability as for the general population.)

(b) Let C be the event that B's blood type matches, and condition on whether B is guilty. This gives

$$P(C|M) = P(C|M, A)P(A|M) + P(C|M, B)P(B|M) = \frac{1}{10} \cdot \frac{10}{11} + \frac{1}{11} = \frac{2}{11}.$$

26. Ⓢ To battle against spam, Bob installs two anti-spam programs. An email arrives, which is either legitimate (event L) or spam (event L^c), and which program j marks as legitimate (event M_j) or marks as spam (event M_j^c) for $j \in \{1, 2\}$. Assume that 10% of Bob's email is legitimate and that the two programs are each "90% accurate" in the sense that $P(M_j|L) = P(M_j^c|L^c) = 9/10$. Also assume that given whether an email is spam, the two programs' outputs are conditionally independent.

(a) Find the probability that the email is legitimate, given that the 1st program marks it as legitimate (simplify).

(b) Find the probability that the email is legitimate, given that both programs mark it as legitimate (simplify).

(c) Bob runs the 1st program and M_1 occurs. He updates his probabilities and then runs the 2nd program. Let $\tilde{P}(A) = P(A|M_1)$ be the updated probability function after running the 1st program. Explain briefly in words whether or not $\tilde{P}(L|M_2) = P(L|M_1 \cap M_2)$: is conditioning on $M_1 \cap M_2$ in one step equivalent to first conditioning on M_1, then updating probabilities, and then conditioning on M_2?

Solution:

(a) By Bayes' rule,

$$P(L|M_1) = \frac{P(M_1|L)P(L)}{P(M_1)} = \frac{\frac{9}{10} \cdot \frac{1}{10}}{\frac{9}{10} \cdot \frac{1}{10} + \frac{1}{10} \cdot \frac{9}{10}} = \frac{1}{2}.$$

(b) By Bayes' rule,

$$P(L|M_1, M_2) = \frac{P(M_1, M_2|L)P(L)}{P(M_1, M_2)} = \frac{(\frac{9}{10})^2 \cdot \frac{1}{10}}{(\frac{9}{10})^2 \cdot \frac{1}{10} + (\frac{1}{10})^2 \cdot \frac{9}{10}} = \frac{9}{10}.$$

(c) Yes, they are the same, since Bayes' rule is coherent. The probability of an event given various pieces of evidence does not depend on the order in which the pieces of evidence are incorporated into the updated probabilities.

27. Suppose that there are 5 blood types in the population, named type 1 through type 5, with probabilities $p_1, p_2, \ldots, p_5$. A crime was committed by two individuals. A suspect, who has blood type 1, has prior probability p of being guilty. At the crime scene blood evidence is collected, which shows that one of the criminals has type 1 and the other has type 2.

   Find the posterior probability that the suspect is guilty, given the evidence. Does the evidence make it more likely or less likely that the suspect is guilty, or does this depend on the values of the parameters $p, p_1, \ldots, p_5$? If it depends, give a simple criterion for when the evidence makes it more likely that the suspect is guilty.

   Solution: Let B be the event that the criminals have blood types 1 and 2 and G be the event that the suspect is guilty, so $P(G) = p$. Then

   $$P(G|B) = \frac{P(B|G)P(G)}{P(B|G)P(G) + P(B|G^c)P(G^c)} = \frac{p_2 p}{p_2 p + 2p_1 p_2 (1-p)} = \frac{p}{p + 2p_1(1-p)},$$

   since given G, event B occurs if and only if the other criminal has blood type 2, while given G^c, the probability is $p_1 p_2$ that the elder criminal and the younger criminal have blood types 1 and 2 respectively, and also is $p_1 p_2$ for the other way around.

   Note that p_2 canceled out and p_3, p_4, p_5 are irrelevant. If $p_1 = 1/2$, then $P(G|B) = P(G)$. If $p_1 < 1/2$, then $P(G|B) > P(G)$, which means that the evidence increases the probability of guilt. But if $p_1 > 1/2$, then $P(G|B) < P(G)$, so the evidence decreases the probability of guilt, even though the evidence includes finding blood at the scene of the crime that matches the suspect's blood type!

28. Fred has just tested positive for a certain disease.

   (a) Given this information, find the posterior odds that he has the disease, in terms of the prior odds, the sensitivity of the test, and the specificity of the test.

   (b) Not surprisingly, Fred is much more interested in $P(\text{have disease}|\text{test positive})$, known as the *positive predictive value*, than in the sensitivity $P(\text{test positive}|\text{have disease})$. A handy rule of thumb in biostatistics and epidemiology is as follows:

   For a rare disease and a reasonably good test, specificity matters much more than sensitivity in determining the positive predictive value.

   Explain intuitively why this rule of thumb works. For this part you can make up some specific numbers and interpret probabilities in a frequentist way as proportions in a large population, e.g., assume the disease afflicts 1% of a population of 10000 people and then consider various possibilities for the sensitivity and specificity.

   Solution:

   (a) Let D be the event that Fred has the disease, and T be the event that he tests positive. Let sens $= P(T|D)$, spec $= P(T^c|D^c)$ be the sensitivity and specificity (respectively). By the odds form of Bayes' rule (or using Bayes' rule in the numerator and the denominator), the posterior odds of having the disease are

   $$\frac{P(D|T)}{P(D^c|T)} = \frac{P(D)}{P(D^c)} \frac{P(T|D)}{P(T|D^c)} = (\text{prior odds of } D) \left(\frac{\text{sens}}{1 - \text{spec}} \right).$$

(b) Let p be the prior probability of having the disease and $q = 1 - p$. Let PPV be the positive predictive value. By (a) or directly using Bayes' rule, we have

$$\text{PPV} = \frac{\text{sens}}{\text{sens} + \frac{q}{p}(1 - \text{spec})}.$$

For a concrete example to build intuition, let $p = 0.01$ and take sens $=$ spec $= 0.9$ as a baseline. Then PPV ≈ 0.083. In the calculations below, we describe what happens if sensitivity is changed while specificity is held constant at 0.9 or vice versa. If we can improve the sensitivity to 0.95, the PPV improves slightly, to 0.088. But if we can improve the specificity to 0.95, the PPV improves to 0.15, a much bigger improvement. If we can improve the sensitivity to 0.99, the PPV improves to 0.091, but the other way around the PPV improves drastically more, to 0.48. Even in the extreme case that we can make the sensitivity 1, the PPV only improves to 0.092. But in the extreme case that we can make the specificity 1, the PPV becomes 1, the best value possible!

To further the intuitive picture, imagine a population of 10000 people, in which 1% (i.e., 100 people) have the disease. Again take sens $=$ spec $= 0.9$ as a baseline. On average, there will be 90 true positives (correctly diagnosed diseased people), 10 false negatives (misdiagnosed diseased people), 8910 true negatives (correctly diagnosed healthy people), and 990 false positives (misdiagnosed healthy people). This is illustrated in the figure below (not to scale).
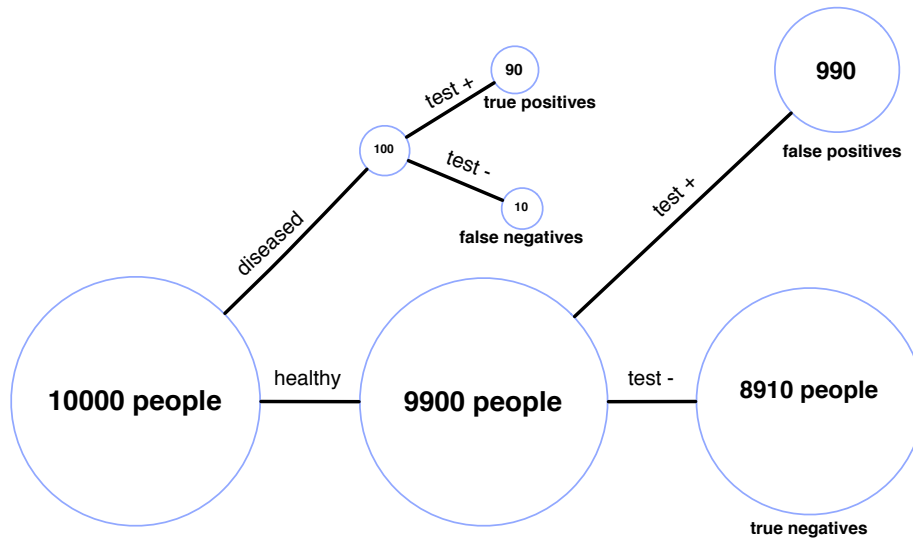
The PPV corresponds to the number of true positives over the number of positives, which is $90/(90+990) \approx 0.083$ in this example. Increasing specificity could dramatically decrease the number of false positives, replacing 990 by a much lower number; on the other hand, increasing sensitivity could at best increase the number of true positives from 90 to 100 here.

29. A family has two children. Let C be a characteristic that a child can have, and assume that each child has characteristic C with probability p, independently of each other and of gender. For example, C could be the characteristic "born in winter" as in Example 2.2.7. Show that the probability that both children are girls given that at least one is a girl with characteristic C is $\frac{2-p}{4-p}$, which is $1/3$ if $p = 1$ (agreeing with the first part of Example 2.2.5) and approaches $1/2$ from below as $p \to 0$ (agreeing with Example 2.2.7).

Solution: Let G be the event that both children are girls, A be the event that at least

one child is a girl with characteristic C, and B be the event that at least one child has characteristic C. Note that $G \cap A = G \cap B$, and G is independent of B. Then

$$
\begin{aligned}
P(G|A) &= \frac{P(G, A)}{P(A)} \\
&= \frac{P(G, B)}{P(A)} \\
&= \frac{P(G)P(B)}{P(A)} \\
&= \frac{\frac{1}{4}(1 - (1 - p)^2)}{1 - (1 - \frac{p}{2})^2} \\
&= \frac{1 - (1 - 2p + p^2)}{4 - (4 - 4p + p^2)} \\
&= \frac{2 - p}{4 - p}.
\end{aligned}
$$

This is $1/3$ if $p = 1$ and approaches $1/2$ as $p \to 0$, but is less than $1/2$ for all $p > 0$ since $\frac{2-p}{4-p} < \frac{1}{2}$ is equivalent to $4 - 2p < 4 - p$, which in turn is equivalent to $p > 0$.

Independence and conditional independence

30. Ⓢ A family has 3 children, creatively named A, B, and C.

    (a) Discuss intuitively (but clearly) whether the event "A is older than B" is independent of the event "A is older than C".

    (b) Find the probability that A is older than B, given that A is older than C.

    Solution:

    (a) They are not independent: knowing that A is older than B makes it more likely that A is older than C, as if A is older than B, then the only way that A can be younger than C is if the birth order is CAB, whereas the birth orders ABC and ACB are both compatible with A being older than B. To make this more intuitive, think of an extreme case where there are 100 children instead of 3, call them $A_1, \dots, A_{100}$. Given that A_1 is older than all of $A_2, A_3, \dots, A_{99}$, it's clear that A_1 is very old (relatively), whereas there isn't evidence about where A_{100} fits into the birth order.

    (b) Writing $x > y$ to mean that x is older than y,

    $$
    P(A > B | A > C) = \frac{P(A > B, A > C)}{P(A > C)} = \frac{1/3}{1/2} = \frac{2}{3}
    $$

    since $P(A > B, A > C) = P(A \text{ is the eldest child}) = 1/3$ (unconditionally, any of the 3 children is equally likely to be the eldest).

31. Ⓢ Is it possible that an event is independent of itself? If so, when is this the case?

    Solution: Let A be an event. If A is independent of itself, then $P(A) = P(A \cap A) = P(A)^2$, so $P(A)$ is 0 or 1. So this is only possible in the extreme cases that the event has probability 0 or 1.

32. Ⓢ Consider four nonstandard dice (the *Efron dice*), whose sides are labeled as follows (the 6 sides on each die are equally likely).

    A: $4, 4, 4, 4, 0, 0$
    B: $3, 3, 3, 3, 3, 3$

C: $6, 6, 2, 2, 2, 2$

D: $5, 5, 5, 1, 1, 1$

These four dice are each rolled once. Let A be the result for die A, B be the result for die B, etc.

(a) Find $P(A > B), P(B > C), P(C > D)$, and $P(D > A)$.

(b) Is the event $A > B$ independent of the event $B > C$? Is the event $B > C$ independent of the event $C > D$? Explain.

Solution:

(a) We have
$$P(A > B) = P(A = 4) = 2/3.$$
$$P(B > C) = P(C = 2) = 2/3.$$
$$P(C > D) = P(C = 6) + P(C = 2, D = 1) = 2/3.$$
$$P(D > A) = P(D = 5) + P(D = 1, A = 0) = 2/3.$$

(b) The event $A > B$ is independent of the event $B > C$ since $A > B$ is the same thing as $A = 4$, knowledge of which gives no information about $B > C$ (which is the same thing as $C = 2$). On the other hand, $B > C$ is *not* independent of $C > D$ since $P(C > D | C = 2) = 1/2 \neq 1 = P(C > D | C \neq 2)$.

33. Alice, Bob, and 100 other people live in a small town. Let C be the set consisting of the 100 other people, let A be the set of people in C who are friends with Alice, and let B be the set of people in C who are friends with Bob. Suppose that for each person in C, Alice is friends with that person with probability $1/2$, and likewise for Bob, with all of these friendship statuses independent.

    (a) Let $D \subseteq C$. Find $P(A = D)$.

    (b) Find $P(A \subseteq B)$.

    (c) Find $P(A \cup B = C)$.

    Solution:

    (a) More generally, let p be the probability of Alice being friends with any specific person in C (without assuming $p = 1/2$), and let $k = |D|$ (the size of D). Then
    $$P(A = D) = p^k (1 - p)^{100-k},$$
    by independence. For the case $p = 1/2$, this reduces to
    $$P(A = D) = 1/2^{100} \approx 7.89 \times 10^{-31}.$$
    That is, A is equally likely to be any subset of C.

    (b) The event $A \subseteq B$ says that for each person in C, it is not the case that they are friends with Alice but not with Bob. The event "friends with Alice but not Bob" has a probability of $1/4$ for each person in C, so by independence the overall probability is $(3/4)^{100} \approx 3.21 \times 10^{-13}$.

    (c) The event $A \cup B = C$ says that everyone in C is friends with Alice or Bob (inclusive of the possibility of being friends with both). For each person in C, there is a $3/4$ chance that they are friends with Alice or Bob, so by independence there is a $(3/4)^{100} \approx 3.21 \times 10^{-13}$ chance that everyone in C is friends with Alice or Bob.

34. Suppose that there are two types of drivers: good drivers and bad drivers. Let G be the event that a certain man is a good driver, A be the event that he gets into a car accident next year, and B be the event that he gets into a car accident the following year. Let $P(G) = g$ and $P(A|G) = P(B|G) = p_1, P(A|G^c) = P(B|G^c) = p_2$, with $p_1 < p_2$. Suppose that given the information of whether or not the man is a good driver, A and B are independent (for simplicity and to avoid being morbid, assume that the accidents being considered are minor and wouldn't make the man unable to drive).

    (a) Explain intuitively whether or not A and B are independent.

    (b) Find $P(G|A^c)$.

    (c) Find $P(B|A^c)$.

    Solution:

    (a) Intuitively, A and B are *not* independent, since learning that the man has a car accident next year makes it more likely that he is a bad driver, which in turn makes it more likely that he will have another accident the following year. We have that A and B are conditionally independent given G (and conditionally independent given G^c), but they are not independent since A gives information about whether the man is a good driver, and this information is very relevant for assessing how likely B is.

    (b) By Bayes's rule and LOTP,

    $$P(G|A^c) = \frac{P(A^c|G)P(G)}{P(A^c)} = \frac{(1 - p_1)g}{(1 - p_1)g + (1 - p_2)(1 - g)}.$$

    (c) Condition on whether or not the man is a good driver, using LOTP with extra conditioning:

    $$\begin{aligned} P(B|A^c) &= P(B|A^c, G)P(G|A^c) + P(B|A^c, G^c)P(G^c|A^c) \\ &= P(B|G)P(G|A^c) + P(B|G^c)P(G^c|A^c) \\ &= p_1 P(G|A^c) + p_2(1 - P(G|A^c)) \\ &= \frac{p_1(1 - p_1)g + p_2(1 - p_2)(1 - g)}{(1 - p_1)g + (1 - p_2)(1 - g)}. \end{aligned}$$

35. ⓢ You are going to play 2 games of chess with an opponent whom you have never played against before (for the sake of this problem). Your opponent is equally likely to be a beginner, intermediate, or a master. Depending on which, your chances of winning an individual game are 90%, 50%, or 30%, respectively.

    (a) What is your probability of winning the first game?

    (b) Congratulations: you won the first game! Given this information, what is the probability that you will also win the second game (assume that, given the skill level of your opponent, the outcomes of the games are independent)?

    (c) Explain the distinction between assuming that the outcomes of the games are independent and assuming that they are conditionally independent given the opponent's skill level. Which of these assumptions seems more reasonable, and why?

    Solution:

    (a) Let W_i be the event of winning the ith game. By the law of total probability,

    $$P(W_1) = (0.9 + 0.5 + 0.3)/3 = 17/30.$$

(b) We have $P(W_2|W_1) = P(W_2, W_1)/P(W_1)$. The denominator is known from (a), while the numerator can be found by conditioning on the skill level of the opponent:

$$P(W_1, W_2) = \frac{1}{3}P(W_1, W_2|\text{beginner}) + \frac{1}{3}P(W_1, W_2|\text{intermediate}) + \frac{1}{3}P(W_1, W_2|\text{expert}).$$

Since W_1 and W_2 are conditionally independent given the skill level of the opponent, this becomes

$$P(W_1, W_2) = (0.9^2 + 0.5^2 + 0.3^2)/3 = 23/60.$$

So

$$P(W_2|W_1) = \frac{23/60}{17/30} = 23/34.$$

(c) Independence here means that knowing one game's outcome gives no information about the other game's outcome, while conditional independence is the same statement where all probabilities are conditional on the opponent's skill level. Conditional independence given the opponent's skill level is a more reasonable assumption here. This is because winning the first game gives information about the opponent's skill level, which in turn gives information about the result of the second game.

That is, if the opponent's skill level is treated as fixed and known, then it may be reasonable to assume independence of games given this information; with the opponent's skill level random, earlier games can be used to help infer the opponent's skill level, which affects the probabilities for future games.

36. (a) Suppose that in the population of college applicants, being good at baseball is independent of having a good math score on a certain standardized test (with respect to some measure of "good"). A certain college has a simple admissions procedure: admit an applicant if and only if the applicant is good at baseball or has a good math score on the test.

Give an intuitive explanation of why it makes sense that among students that the college admits, having a good math score is *negatively associated* with being good at baseball, i.e., conditioning on having a good math score decreases the chance of being good at baseball.

(b) Show that if A and B are independent and $C = A \cup B$, then A and B are conditionally dependent given C (as long as $P(A \cap B) > 0$ and $P(A \cup B) < 1$), with

$$P(A|B, C) < P(A|C).$$

This phenomenon is known as *Berkson's paradox*, especially in the context of admissions to a school, hospital, etc.

Solution:

(a) Even though baseball skill and the math score are independent in the general population of applicants, it makes sense that they will become dependent (with a negative association) when restricting only to the students who are admitted. This is because within this sub-population, having a bad math score implies being good at baseball (else the student would not have been admitted). So having a good math score decreases the chance of being good in baseball (as shown in Exercise 16, if an event B is evidence in favor of an event A, then B^c is evidence against A).

As another explanation, note that 3 types of students are admitted: (i) good math score, good at baseball; (ii) good math score, bad at baseball; (iii) bad math score, good at baseball. Conditioning on having good math score removes students of type (iii) from consideration, which decreases the proportion of students who are good at baseball.

(b) Assume that A, B, C are as described. Then

$$P(A|B \cap C) = P(A|B) = P(A),$$

since A and B are independent and $B \cap C = B$. In contrast,

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A)}{P(C)} > P(A),$$

since $0 < P(C) < 1$. Therefore, $P(A|B, C) = P(A) < P(A|C)$.

37. We want to design a spam filter for email. As described in Exercise 1, a major strategy is to find phrases that are much more likely to appear in a spam email than in a non-spam email. In that exercise, we only consider one such phrase: "free money". More realistically, suppose that we have created a list of 100 words or phrases that are much more likely to be used in spam than in non-spam.

    Let W_j be the event that an email contains the jth word or phrase on the list. Let

    $$p = P(\text{spam}), p_j = P(W_j|\text{spam}), r_j = P(W_j|\text{not spam}),$$

    where "spam" is shorthand for the event that the email is spam.

    Assume that $W_1, \ldots, W_{100}$ are conditionally independent given that the email is spam, and conditionally independent given that it is not spam. A method for classifying emails (or other objects) based on this kind of assumption is called a *naive Bayes classifier*. (Here "naive" refers to the fact that the conditional independence is a strong assumption, not to Bayes being naive. The assumption may or may not be realistic, but naive Bayes classifiers sometimes work well in practice even if the assumption is not realistic.)

    Under this assumption we know, for example, that

    $$P(W_1, W_2, W_3^c, W_4^c, \ldots, W_{100}^c|\text{spam}) = p_1 p_2 (1 - p_3)(1 - p_4) \ldots (1 - p_{100}).$$

    Without the naive Bayes assumption, there would be vastly more statistical and computational difficulties since we would need to consider $2^{100} \approx 1.3 \times 10^{30}$ events of the form $A_1 \cap A_2 \cdots \cap A_{100}$ with each A_j equal to either W_j or W_j^c. A new email has just arrived, and it includes the 23rd, 64th, and 65th words or phrases on the list (but not the other 97). So we want to compute

    $$P(\text{spam}|W_1^c, \ldots, W_{22}^c, W_{23}, W_{24}^c, \ldots, W_{63}^c, W_{64}, W_{65}, W_{66}^c, \ldots, W_{100}^c).$$

    Note that we need to condition on *all* the evidence, not just the fact that $W_{23} \cap W_{64} \cap W_{65}$ occurred. Find the conditional probability that the new email is spam (in terms of p and the p_j and r_j).

    Solution:

    Let

    $$E = W_1^c \cap \cdots \cap W_{22}^c \cap W_{23} \cap W_{24}^c \cap \cdots \cap W_{63}^c \cap W_{64} \cap W_{65} \cap W_{66}^c \cap \cdots \cap W_{100}^c$$

    be the observed evidence. By Bayes' rule, LOTP, and conditional independence,

    $$P(\text{spam}|E) = \frac{P(E|\text{spam})P(\text{spam})}{P(E|\text{spam})P(\text{spam}) + P(E|\text{non-spam})P(\text{non-spam})}$$
    $$= \frac{ap}{ap + b(1 - p)},$$

    where

    $$a = (1 - p_1) \ldots (1 - p_{22})p_{23}(1 - p_{24}) \ldots (1 - p_{63})p_{64}p_{65}(1 - p_{66}) \ldots (1 - p_{100}),$$

    $$b = (1 - r_1) \ldots (1 - r_{22})r_{23}(1 - r_{24}) \ldots (1 - r_{63})r_{64}r_{65}(1 - r_{66}) \ldots (1 - r_{100}).$$

Monty Hall

38. ⓢ (a) Consider the following 7-door version of the Monty Hall problem. There are 7 doors, behind one of which there is a car (which you want), and behind the rest of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door. Monty Hall then opens 3 goat doors, and offers you the option of switching to any of the remaining 3 doors.

    Assume that Monty Hall knows which door has the car, will always open 3 goat doors and offer the option of switching, and that Monty chooses with equal probabilities from all his choices of which goat doors to open. Should you switch? What is your probability of success if you switch to one of the remaining 3 doors?

    (b) Generalize the above to a Monty Hall problem where there are $n \geq 3$ doors, of which Monty opens m goat doors, with $1 \leq m \leq n - 2$.

    Solution:

    (a) Assume the doors are labeled such that you choose door 1 (to simplify notation), and suppose first that you follow the "stick to your original choice" strategy. Let S be the event of success in getting the car, and let C_j be the event that the car is behind door j. Conditioning on which door has the car, we have

    $$P(S) = P(S|C_1)P(C_1) + \cdots + P(S|C_7)P(C_7) = P(C_1) = \frac{1}{7}.$$

    Let M_{ijk} be the event that Monty opens doors i, j, k. Then

    $$P(S) = \sum_{i,j,k} P(S|M_{ijk})P(M_{ijk})$$

    (summed over all i, j, k with $2 \leq i < j < k \leq 7$.) By symmetry, this gives

    $$P(S|M_{ijk}) = P(S) = \frac{1}{7}$$

    for all i, j, k with $2 \leq i < j < k \leq 7$. Thus, the conditional probability that the car is behind 1 of the remaining 3 doors is 6/7, which gives 2/7 for each. So you should switch, thus making your probability of success 2/7 rather than 1/7.

    (b) By the same reasoning, the probability of success for "stick to your original choice" is $\frac{1}{n}$, both unconditionally and conditionally. Each of the $n - m - 1$ remaining doors has conditional probability $\frac{n-1}{(n-m-1)n}$ of having the car. This value is greater than $\frac{1}{n}$, so you should switch, thus obtaining probability $\frac{n-1}{(n-m-1)n}$ of success (both conditionally and unconditionally).

39. ⓢ Consider the Monty Hall problem, except that Monty enjoys opening door 2 more than he enjoys opening door 3, and if he has a choice between opening these two doors, he opens door 2 with probability p, where $\frac{1}{2} \leq p \leq 1$.

    To recap: there are three doors, behind one of which there is a car (which you want), and behind the other two of which there are goats (which you don't want). Initially, all possibilities are equally likely for where the car is. You choose a door, which for concreteness we assume is door 1. Monty Hall then opens a door to reveal a goat, and offers you the option of switching. Assume that Monty Hall knows which door has the car, will always open a goat door and offer the option of switching, and as above assume that if Monty Hall has a choice between opening door 2 and door 3, he chooses door 2 with probability p (with $\frac{1}{2} \leq p \leq 1$).

    (a) Find the unconditional probability that the strategy of always switching succeeds (unconditional in the sense that we do not condition on which of doors 2 or 3 Monty opens).

(b) Find the probability that the strategy of always switching succeeds, given that Monty opens door 2.

(c) Find the probability that the strategy of always switching succeeds, given that Monty opens door 3.

Solution:

(a) Let C_j be the event that the car is hidden behind door j and let W be the event that we win using the switching strategy. Using the law of total probability, we can find the unconditional probability of winning:

$$\begin{aligned} P(W) &= P(W|C_1)P(C_1) + P(W|C_2)P(C_2) + P(W|C_3)P(C_3) \\ &= 0 \cdot 1/3 + 1 \cdot 1/3 + 1 \cdot 1/3 = 2/3. \end{aligned}$$

(b) A tree method works well here (delete the paths which are no longer relevant after the conditioning, and reweight the remaining values by dividing by their sum), or we can use Bayes' rule and the law of total probability (as below).

Let D_i be the event that Monty opens Door i. Note that we are looking for $P(W|D_2)$, which is the same as $P(C_3|D_2)$ as we first choose Door 1 and then switch to Door 3. By Bayes' rule and the law of total probability,

$$\begin{aligned} P(C_3|D_2) &= \frac{P(D_2|C_3)P(C_3)}{P(D_2)} \\ &= \frac{P(D_2|C_3)P(C_3)}{P(D_2|C_1)P(C_1) + P(D_2|C_2)P(C_2) + P(D_2|C_3)P(C_3)} \\ &= \frac{1 \cdot 1/3}{p \cdot 1/3 + 0 \cdot 1/3 + 1 \cdot 1/3} \\ &= \frac{1}{1+p}. \end{aligned}$$

(c) The structure of the problem is the same as Part (b) (except for the condition that $p \geq 1/2$, which was not needed above). Imagine repainting doors 2 and 3, reversing which is called which. By Part (b) with $1-p$ in place of p, $P(C_2|D_3) = \frac{1}{1+(1-p)} = \frac{1}{2-p}$.

40. The ratings of Monty Hall's show have dropped slightly, and a panicking executive producer complains to Monty that the part of the show where he opens a door lacks suspense: Monty always opens a door with a goat. Monty replies that the reason is so that the game is never spoiled by him revealing the car, but he agrees to update the game as follows.

Before each show, Monty secretly flips a coin with probability p of Heads. If the coin lands Heads, Monty resolves to open a goat door (with equal probabilities if there is a choice). Otherwise, Monty resolves to open a random unopened door, with equal probabilities. The contestant knows p but does not know the outcome of the coin flip. When the show starts, the contestant chooses a door. Monty (who knows where the car is) then opens a door. If the car is revealed, the game is over; if a goat is revealed, the contestant is offered the option of switching. Now suppose it turns out that the contestant opens door 1 and then Monty opens door 2, revealing a goat. What is the contestant's probability of success if he or she switches to door 3?

Solution: For $j = 1, 2, 3$, let C_j be the event that the car is behind door j, $G_j = C_j^c$, and M_j be the event that Monty opens door j. Let R be the event that Monty is in "random mode" (i.e., the coin lands Tails). By the law of total probability,

$$P(C_3|M_2, G_2) = P(R|M_2, G_2)P(C_3|M_2, G_2, R) + P(R^c|M_2, G_2)P(C_3|M_2, G_2, R^c),$$

where $P(C_3|M_2, G_2, R^c) = 2/3$ (since given R^c, Monty is operating as in the usual Monty Hall problem) and

$$P(C_3|M_2, G_2, R) = \frac{P(M_2, G_2|C_3, R)P(C_3|R)}{P(M_2, G_2|R)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{2}{3}} = \frac{1}{2}.$$

For the denominator above, note that M_2 and G_2 are conditionally independent given R; for the numerator, note also that $P(M_2, G_2|C_3, R) = P(M_2|C_3, R) = P(M_2|R)$. The posterior probability that Monty is in random mode is

$$P(R|M_2, G_2) = \frac{P(M_2, G_2|R)P(R)}{P(M_2, G_2|R)P(R) + P(M_2, G_2|R^c)P(R^c)} = \frac{\frac{1}{2} \cdot \frac{2}{3}(1-p)}{\frac{1}{2} \cdot \frac{2}{3}(1-p) + \frac{1}{2}p} = \frac{2(1-p)}{2+p},$$

again since M_2 and G_2 are conditionally independent given R, and since given R^c, M_2 implies G_2. Putting these results together, we have

$$P(C_3|M_2, G_2) = \frac{1}{2}\left(\frac{2(1-p)}{2+p}\right) + \frac{2}{3}\left(1 - \frac{2(1-p)}{2+p}\right) = \frac{1+p}{2+p}.$$

41. You are the contestant on the Monty Hall show. Monty is trying out a new version of his game, with rules as follows. You get to choose one of three doors. One door has a car behind it, another has a computer, and the other door has a goat (with all permutations equally likely). Monty, who knows which prize is behind each door, will open a door (but not the one you chose) and then let you choose whether to switch from your current choice to the other unopened door.

Assume that you prefer the car to the computer, the computer to the goat, and (by transitivity) the car to the goat.

(a) Suppose for this part only that Monty always opens the door that reveals your less preferred prize out of the two alternatives, e.g., if he is faced with the choice between revealing the goat or the computer, he will reveal the goat. Monty opens a door, revealing a goat (this is again for this part only). Given this information, should you switch? If you do switch, what is your probability of success in getting the car?

(b) Now suppose that Monty reveals your less preferred prize with probability p, and your more preferred prize with probability $q = 1 - p$. Monty opens a door, revealing a computer. Given this information, should you switch (your answer can depend on p)? If you do switch, what is your probability of success in getting the car (in terms of p)?

Solution:

(a) Let C be the event that the car is behind the door you originally chose, and let M_{goat} be the event that Monty reveals a goat when he opens a door. By Bayes' rule,

$$P(C|M_{\text{goat}}) = \frac{P(M_{\text{goat}}|C)P(C)}{P(M_{\text{goat}})} = \frac{(1)(1/3)}{2/3} = 1/2,$$

where the denominator comes from the fact that the two doors other than your initial choice are equally likely to have $\{\text{car}, \text{computer}\}$, $\{\text{computer}, \text{goat}\}$, or $\{\text{car}, \text{goat}\}$, and only in the first of these cases will Monty not reveal a goat.

So you should be indifferent between switching and not switching; either way, your conditional probability of getting the car is $1/2$. (Note though that the *unconditional* probability that switching would get you the car, before Monty revealed the goat, is $2/3$ since you will succeed by switching if and only if your initial door does not have the car.)

(b) Let C, R, G be the events that the car is behind the door you originally chosen is

a car, computer, goat, respectively, and let M_{comp} be the event that Monty reveals a computer when he opens a door. By Bayes' rule and LOTP,

$$P(C|M_{\text{comp}}) = \frac{P(M_{\text{comp}}|C)P(C)}{P(M_{\text{comp}})} = \frac{P(M_{\text{comp}}|C)P(C)}{P(M_{\text{comp}}|C)P(C) + P(M_{\text{comp}}|R)P(R) + P(M_{\text{comp}}|G)P(G)}.$$

We have $P(M_{\text{comp}}|C) = q, P(M_{\text{comp}}|R) = 0, P(M_{\text{comp}}|G) = p$, so

$$P(C|M_{\text{comp}}) = \frac{q/3}{q/3 + 0 + p/3} = q.$$

Thus, your conditional probability of success if you follow the switching strategy is p. For $p < 1/2$, you should not switch, for $p = 1/2$, you should be indifferent about switching, and for $p > 1/2$, you should switch.

First-step analysis and gambler's ruin

42. (S) A fair die is rolled repeatedly, and a running total is kept (which is, at each time, the total of all the rolls up until that time). Let p_n be the probability that the running total is ever *exactly* n (assume the die will always be rolled enough times so that the running total will eventually exceed n, but it may or may not ever equal n).

(a) Write down a recursive equation for p_n (relating p_n to earlier terms p_k in a simple way). Your equation should be true for all positive integers n, so give a definition of p_0 and p_k for $k < 0$ so that the recursive equation is true for small values of n.

(b) Find p_7.

(c) Give an intuitive explanation for the fact that $p_n \to 1/3.5 = 2/7$ as $n \to \infty$.

Solution:

(a) We will find something to condition on to reduce the case of interest to earlier, simpler cases. This is achieved by the useful strategy of *first step anaysis*. Let p_n be the probability that the running total is ever *exactly* n. Note that if, for example, the first throw is a 3, then the probability of reaching n exactly is p_{n-3} since starting from that point, we need to get a total of $n - 3$ exactly. So

$$p_n = \frac{1}{6}(p_{n-1} + p_{n-2} + p_{n-3} + p_{n-4} + p_{n-5} + p_{n-6}),$$

where we define $p_0 = 1$ (which makes sense anyway since the running total is 0 before the first toss) and $p_k = 0$ for $k < 0$.

(b) Using the recursive equation in (a), we have

$$p_1 = \frac{1}{6}, \quad p_2 = \frac{1}{6}(1 + \frac{1}{6}), \quad p_3 = \frac{1}{6}(1 + \frac{1}{6})^2,$$
$$p_4 = \frac{1}{6}(1 + \frac{1}{6})^3, \quad p_5 = \frac{1}{6}(1 + \frac{1}{6})^4, \quad p_6 = \frac{1}{6}(1 + \frac{1}{6})^5.$$

Hence,

$$p_7 = \frac{1}{6}(p_1 + p_2 + p_3 + p_4 + p_5 + p_6) = \frac{1}{6}\left((1 + \frac{1}{6})^6 - 1\right) \approx 0.2536.$$

(c) An intuitive explanation is as follows. The average number thrown by the die is (total of dots)/6, which is $21/6 = 7/2$, so that every throw adds on an average of $7/2$. We can therefore expect to land on 2 out of every 7 numbers, and the probability of

landing on any particular number is $2/7$. A mathematical derivation (which was not requested in the problem) can be given as follows:

$$p_{n+1} + 2p_{n+2} + 3p_{n+3} + 4p_{n+4} + 5p_{n+5} + 6p_{n+6}$$
$$= p_{n+1} + 2p_{n+2} + 3p_{n+3} + 4p_{n+4} + 5p_{n+5}$$
$$+ p_n + p_{n+1} + p_{n+2} + p_{n+3} + p_{n+4} + p_{n+5}$$
$$= p_n + 2p_{n+1} + 3p_{n+2} + 4p_{n+3} + 5p_{n+4} + 6p_{n+5}$$
$$= \cdots$$
$$= p_{-5} + 2p_{-4} + 3p_{-3} + 4p_{-2} + 5p_{-1} + 6p_0 = 6.$$

Taking the limit of the lefthand side as n goes to ∞, we have

$$(1 + 2 + 3 + 4 + 5 + 6)\lim_{n\to\infty} p_n = 6,$$

so $\lim_{n\to\infty} p_n = 2/7$.

43. A sequence of $n \geq 1$ independent trials is performed, where each trial ends in "success" or "failure" (but not both). Let p_i be the probability of success in the ith trial, $q_i = 1 - p_i$, and $b_i = q_i - 1/2$, for $i = 1, 2, \ldots, n$. Let A_n be the event that the number of successful trials is even.

(a) Show that for $n = 2$, $P(A_2) = 1/2 + 2b_1 b_2$.

(b) Show by induction that

$$P(A_n) = 1/2 + 2^{n-1} b_1 b_2 \ldots b_n.$$

(This result is very useful in cryptography. Also, note that it implies that if n coins are flipped, then the probability of an even number of Heads is $1/2$ if and only if at least one of the coins is fair.) Hint: Group some trials into a supertrial.

(c) Check directly that the result of (b) is true in the following simple cases: $p_i = 1/2$ for some i; $p_i = 0$ for all i; $p_i = 1$ for all i.

Solution:

(a) We have

$$P(A_2) = p_1 p_2 + q_1 q_2 = \left(\frac{1}{2} - b_1\right)\left(\frac{1}{2} - b_2\right) + \left(\frac{1}{2} + b_1\right)\left(\frac{1}{2} + b_2\right) = \frac{1}{2} + 2b_1 b_2.$$

(b) For $n = 1$, $P(A_1) = P(\text{1st trial is a failure}) = q_1 = 1/2 + b_1$. For $n = 2$, the result was shown in (a). Now assume the result is true for n, and prove it for $n + 1$, where $n \geq 2$ is fixed. Think of the first n trials as 1 supertrial, where "success" is defined as the number of successes being odd. By assumption, the probability of failure for the supertrial is $\tilde{q}_1 = \frac{1}{2} + 2^{n-1} b_1 \ldots b_n$. Let $\tilde{b}_1 = \tilde{q}_1 - \frac{1}{2}$. Then by (a),

$$P(A_{n+1}) = \frac{1}{2} + 2\tilde{b}_1 b_{n+1} = \frac{1}{2} + 2^n b_1 b_2 \ldots b_n b_{n+1},$$

which completes the induction. This result, called the *piling-up lemma*, is used in cryptography to compute or bound probabilities needed for some widely-used ciphers.

(c) Let $p_i = 1/2$ for some fixed i. Then $b_i = 0$ and we need to check directly that $P(A_n) = 1/2$. Let p be the probability that the number of successes in the $n - 1$ trials other than the ith is odd. Conditioning on the result of the ith trial, we have

$$P(A_n) = p \cdot \frac{1}{2} + (1 - p) \cdot \frac{1}{2} = \frac{1}{2}.$$

For the case that $p_i = 0$ for all i, the event A_n is guaranteed to occur (since there will be 0 successes), which agrees with $P(A_n) = 1/2 + 2^{n-1} \cdot 2^{-n} = 1$. For the case that $p_i = 1$ for all i, there will be n successes, so A_n occurs if and only if n is even. This agrees with (b), which simplifies in this case to $P(A_n) = (1 + (-1)^n)/2$.

44. Ⓢ Calvin and Hobbes play a match consisting of a series of games, where Calvin has probability p of winning each game (independently). They play with a "win by two"" rule: the first player to win two games more than his opponent wins the match. Find the probability that Calvin wins the match (in terms of p), in two different ways:

(a) by conditioning, using the law of total probability.

(b) by interpreting the problem as a gambler's ruin problem.

Solution:

(a) Let C be the event that Calvin wins the match, $X \sim \text{Bin}(2, p)$ be how many of the first 2 games he wins, and $q = 1 - p$. Then

$$P(C) = P(C|X = 0)q^2 + P(C|X = 1)(2pq) + P(C|X = 2)p^2 = 2pqP(C) + p^2,$$

so $P(C) = \frac{p^2}{1-2pq}$. This can also be written as $\frac{p^2}{p^2+q^2}$, since $p + q = 1$.

Sanity check: Note that this should (and does) reduce to 1 for $p = 1$, 0 for $p = 0$, and $\frac{1}{2}$ for $p = \frac{1}{2}$. Also, it makes sense that the probability of Hobbes winning, which is $1 - P(C) = \frac{q^2}{p^2+q^2}$, can also be obtained by swapping p and q.

(b) The problem can be thought of as a gambler's ruin where each player starts out with \$2. So the probability that Calvin wins the match is

$$\frac{1 - (q/p)^2}{1 - (q/p)^4} = \frac{(p^2 - q^2)/p^2}{(p^4 - q^4)/p^4} = \frac{(p^2 - q^2)/p^2}{(p^2 - q^2)(p^2 + q^2)/p^4} = \frac{p^2}{p^2 + q^2},$$

which agrees with the above.

45. A gambler repeatedly plays a game where in each round, he wins a dollar with probability $1/3$ and loses a dollar with probability $2/3$. His strategy is "quit when he is ahead by \$2", though some suspect he is a gambling addict anyway. Suppose that he starts with a million dollars. Show that the probability that he'll ever be ahead by \$2 is less than $1/4$.

Solution: This is a special case of the gambler's ruin problem. Let A_1 be the event that he is successful on the first play and let W be the event that he is ever ahead by \$2 before being ruined. Then by the law of total probability, we have

$$P(W) = P(W|A_1)P(A_1) + P(W|A_1^c)P(A_1^c).$$

Let a_i be the probability that the gambler achieves a profit of \$2 before being ruined, starting with a fortune of \$$i$. For our setup, $P(W) = a_i$, $P(W|A_1) = a_{i+1}$ and $P(W|A_1^c) = a_{i-1}$. Therefore,

$$a_i = a_{i+1}/3 + 2a_{i-1}/3,$$

with boundary conditions $a_0 = 0$ and $a_{i+2} = 1$. We can then solve this difference equation for a_i (directly or using the result of the gambler's ruin problem):

$$a_i = \frac{2^i - 1}{2^{2+i} - 1}.$$

This is always less than $1/4$ since $\frac{2^i-1}{2^{2+i}-1} < \frac{1}{4}$ is equivalent to $4(2^i - 1) < 2^{2+i} - 1$, which is equivalent to the true statement $2^{2+i} - 4 < 2^{2+i} - 1$.

46. As in the gambler's ruin problem, two gamblers, A and B, make a series of bets, until one of the gamblers goes bankrupt. Let A start out with i dollars and B start out with $N - i$ dollars, and let p be the probability of A winning a bet, with $0 < p < \frac{1}{2}$. Each bet is for $\frac{1}{k}$ dollars, with k a positive integer, e.g., $k = 1$ is the original gambler's ruin problem and $k = 20$ means they're betting nickels. Find the probability that A wins the game, and determine what happens to this as $k \to \infty$.

    Solution: Define 1 kidollar to be $\frac{1}{k}$ dollars. This problem is exactly the gambler's ruin problem if everything is written in terms of kidollars as the unit of currency, e.g., for $k = 20$ it's just the gambler's ruin problem, with nickels instead of dollars. Note that A starts out with ki kidollars and B starts out with $k(N - i)$ kidollars. Let $r = (1 - p)/p$. By the gambler's ruin problem, the probability that A wins the game is

$$P(\text{A wins}) = \frac{1 - r^{ki}}{1 - r^{kN}}.$$

    This is 0 if $i = 0$ and 1 if $i = N$. For $1 \le i \le N - 1$, since $r > 1$ we have

$$\lim_{k \to \infty} \frac{1 - r^{ki}}{1 - r^{kN}} = \lim_{k \to \infty} \frac{r^{ki}}{r^{kN}} = \lim_{k \to \infty} r^{k(i - N)} = 0.$$

47. There are 100 equally spaced points around a circle. At 99 of the points, there are sheep, and at 1 point, there is a wolf. At each time step, the wolf randomly moves either clockwise or counterclockwise by 1 point. If there is a sheep at that point, he eats it. The sheep don't move. What is the probability that the sheep who is initially opposite the wolf is the last one remaining?

    Solution: Call the sheep initially opposite the wolf Dolly. If Dolly is the last sheep surviving, then both of Dolly's neighbors get eaten before Dolly. But the converse is also true: if both of Dolly's neighbors get eaten before Dolly, then the wolf must have gone all the way around the long way after eating the first neighbor to get to the other neighbor.

    Now consider the moment just after the wolf has eaten the first of Dolly's neighbors. The question then becomes whether the wolf will reach Dolly first or the other neighbor first. This is the same as the gambler's ruin problem, viewed as a random walk started at 1 and ending when it reaches either 0 or 99 (reaching 0 corresponds to eating Dolly). Thus, the probability that Dolly is the last sheep surviving is $1/99$. (Similarly, it can be shown that the sheep are all equally likely to be the last sheep surviving!)

48. An immortal drunk man wanders around randomly on the integers. He starts at the origin, and at each step he moves 1 unit to the right or 1 unit to the left, with probabilities p and $q = 1 - p$ respectively, independently of all his previous steps. Let S_n be his position after n steps.

    (a) Let p_k be the probability that the drunk ever reaches the value k, for all $k \ge 0$. Write down a difference equation for p_k (you do not need to solve it for this part).

    (b) Find p_k, fully simplified; be sure to consider all 3 cases: $p < 1/2, p = 1/2$, and $p > 1/2$. Feel free to assume that if $A_1, A_2, \dots$ are events with $A_j \subseteq A_{j+1}$ for all j, then $P(A_n) \to P(\cup_{j=1}^{\infty} A_j)$ as $n \to \infty$ (because it is true; this is known as *continuity of probability*).

    Solution:

    (a) Conditioning on the first step,

$$p_k = pp_{k-1} + qp_{k+1}$$

    for all $k \ge 1$, with $p_0 = 1$.

(b) For fixed k and any positive integer j, let A_j be the event that the drunk reaches k before ever reaching $-j$. Then $A_j \subseteq A_{j+1}$ for all m since the drunk would have to walk past $-j$ to reach $-j-1$. Also, $\cup_{j=1}^{\infty} A_j$ is the event that the drunk ever reaches k, since if he reaches $-j$ before k for *all* j, then he will never have time to get to k. Now we just need to find $\lim_{j \to \infty} P(A_j)$, where we already know $P(A_j)$ from the result of the gambler's ruin problem!

For $p = 1/2$,

$$P(A_j) = \frac{j}{j+k} \to 1 \text{ as } j \to \infty.$$

For $p > 1/2$,

$$P(A_j) = \frac{1 - (\frac{q}{p})^j}{1 - (\frac{q}{p})^{j+k}} \to 1 \text{ as } j \to \infty,$$

since $(q/p)^j \to 0$. For $p < 1/2$,

$$P(A_j) = \frac{1 - (\frac{q}{p})^j}{1 - (\frac{q}{p})^{j+k}} \to \left(\frac{p}{q}\right)^k \text{ as } j \to \infty,$$

since $(q/p)^j \to \infty$ so the 1's in the numerator and denominator become negligible as $j \to \infty$.

Simpson's paradox

49. Ⓢ (a) Is it possible to have events A, B, C such that $P(A|C) < P(B|C)$ and $P(A|C^c) < P(B|C^c)$, yet $P(A) > P(B)$? That is, A is less likely than B given that C is true, and also less likely than B given that C is false, yet A is more likely than B if we're given no information about C. Show this is impossible (with a short proof) or find a counterexample (with a story interpreting A, B, C).

(b) If the scenario in (a) is possible, is it a special case of Simpson's paradox, equivalent to Simpson's paradox, or neither? If it is impossible, explain intuitively why it is impossible even though Simpson's paradox is possible.

Solution:

(a) It is *not* possible, as seen using the law of total probability:

$$P(A) = P(A|C)P(C) + P(A|C^c)P(C^c) < P(B|C)P(C) + P(B|C^c)P(C^c) = P(B).$$

(b) In Simpson's paradox, using the notation from the chapter, we can expand out $P(A|B)$ and $P(A|B^c)$ using LOTP to condition on C, but the inequality can flip because of the weights such as $P(C|B)$ on the terms (e.g., Dr. Nick performs a lot more Band-Aid removals than Dr. Hibbert). In this problem, the weights $P(C)$ and $P(C^c)$ are the same in both expansions, so the inequality is preserved.

50. Ⓢ Consider the following conversation from an episode of *The Simpsons*:

> Lisa: *Dad, I think he's an ivory dealer! His boots are ivory, his hat is ivory, and I'm pretty sure that check is ivory.*
> Homer: *Lisa, a guy who has lots of ivory is* less *likely to hurt Stampy than a guy whose ivory supplies are low.*

Here Homer and Lisa are debating the question of whether or not the man (named Blackheart) is likely to hurt Stampy the Elephant if they sell Stampy to him. They clearly disagree about how to use their observations about Blackheart to learn about the probability (conditional on the evidence) that Blackheart will hurt Stampy.

(a) Define clear notation for the various events of interest here.

(b) Express Lisa's and Homer's arguments (Lisa's is partly implicit) as conditional probability statements in terms of your notation from (a).

(c) Assume it is true that someone who has a lot of a commodity will have less desire to acquire more of the commodity. Explain what is wrong with Homer's reasoning that the evidence about Blackheart makes it less likely that he will harm Stampy.

Solution:

(a) Let H be the event that the man will hurt Stampy, let L be the event that a man has lots of ivory, and let D be the event that the man is an ivory dealer.

(b) Lisa observes that L is true. She suggests (reasonably) that this evidence makes D more likely, i.e., $P(D|L) > P(D)$. Implicitly, she suggests that this makes it likely that the man will hurt Stampy, i.e.,

$$P(H|L) > P(H|L^c).$$

Homer argues that

$$P(H|L) < P(H|L^c).$$

(c) Homer does not realize that observing that Blackheart has so much ivory makes it much more likely that Blackheart is an ivory dealer, which in turn makes it more likely that the man will hurt Stampy. This is an example of Simpson's paradox. It may be true that, *controlling for whether or not Blackheart is a dealer*, having high ivory supplies makes it less likely that he will harm Stampy: $P(H|L, D) < P(H|L^c, D)$ and $P(H|L, D^c) < P(H|L^c, D^c)$. However, this does not imply that $P(H|L) < P(H|L^c)$.

51. (a) There are two crimson jars (labeled C_1 and C_2) and two mauve jars (labeled M_1 and M_2). Each jar contains a mixture of green gummi bears and red gummi bears. Show by example that it is possible that C_1 has a much higher percentage of green gummi bears than M_1, and C_2 has a much higher percentage of green gummi bears than M_2, yet if the contents of C_1 and C_2 are merged into a new jar and likewise for M_1 and M_2, then the combination of C_1 and C_2 has a lower percentage of green gummi bears than the combination of M_1 and M_2.

(b) Explain how (a) relates to Simpson's paradox, both intuitively and by explicitly defining events A, B, C as in the statement of Simpson's paradox.

Solution:

(a) As an example, let C_1 have 9 green, 1 red; M_1 have 50 green, 50 red; C_2 have 30 green, 70 red; M_2 have 1 green, 9 red.

(b) This is a form of Simpson's paradox since which jar color is more likely to provide a green gummy bear flips depending on whether the jars get aggregated. To match this example up to the notation used in the statement of Simpson's paradox, let A be the event that a red gummi bear is chosen in the random draw, B be the event that it is drawn from a crimson jar, and C be the event that it is drawn from a jar with index 1. With the numbers from the solution to (a), $P(C|B) = 1/11$ is much less than $P(C|B^c) = 10/11$, which enables us to have $P(A|B) < P(A|B^c)$ even though the inequalities go the other way when we also condition on C or on C^c.

52. As explained in this chapter, Simpson's paradox says that it is possible to have events A, B, C such that $P(A|B, C) < P(A|B^c, C)$ and $P(A|B, C^c) < P(A|B^c, C^c)$, yet $P(A|B) > P(A|B^c)$.

(a) Can Simpson's paradox occur if A and B are independent? If so, give a concrete example (with both numbers and an interpretation); if not, prove that it is impossible.

(b) Can Simpson's paradox occur if A and C are independent? If so, give a concrete example (with both numbers and an interpretation); if not, prove that it is impossible.

(c) Can Simpson's paradox occur if B and C are independent? If so, give a concrete example (with both numbers and an interpretation); if not, prove that it is impossible.

Solution:

(a) No, since if A and B are independent, then $P(A|B) = P(A) = P(A|B^c)$, using the fact that A is also independent of B^c.

(b) No, as shown by the following. Suppose that A and C are independent and that the first two inequalities in Simpson's paradox hold. Then by LOTP,

$$\begin{aligned}
P(A) &= P(A|C) \\
&= P(A|C,B)P(B|C) + P(A|C,B^c)P(B^c|C) \\
&< P(A|C,B^c)P(B|C) + P(A|C,B^c)P(B^c|C) \\
&= P(A|C,B^c),
\end{aligned}$$

so

$$P(A) < P(A|C,B^c).$$

Similarly, $P(A) = P(A|C^c)$ gives

$$P(A) < P(A|C^c,B^c).$$

Thus,

$$P(A) < P(A|B^c),$$

since $P(A|B^c)$ is a weighted average of $P(A|C,B^c)$ and $P(A|C^c,B^c)$ (so is in between them). But then

$$P(A|B) < P(A|B^c),$$

since $P(A)$ is a weighted average of $P(A|B)$ and $P(A|B^c)$ (so is in between them).

(c) No, as shown by the following. Suppose that B and C are independent and that the first two inequalities in Simpson's paradox hold. Then LOTP (as on p. 60) yields

$$\begin{aligned}
P(A|B) &= P(A|C,B)P(C|B) + P(A|C^c,B)P(C^c|B) \\
&< P(A|C,B^c)P(C|B^c) + P(A|C^c,B^c)P(C^c|B^c) \\
&= P(A|B^c).
\end{aligned}$$

53. Ⓢ The book *Red State, Blue State, Rich State, Poor State* by Andrew Gelman [13] discusses the following election phenomenon: within any U.S. state, a wealthy voter is more likely to vote for a Republican than a poor voter, yet the wealthier states tend to favor Democratic candidates! In short: rich individuals (in any state) tend to vote for Republicans, while states with a higher percentage of rich people tend to favor Democrats.

(a) Assume for simplicity that there are only 2 states (called Red and Blue), each of which has 100 people, and that each person is either rich or poor, and either a Democrat or a Republican. Make up numbers consistent with the above, showing how this phenomenon is possible, by giving a 2×2 table for each state (listing how many people in each state are rich Democrats, etc.).

(b) In the setup of (a) (not necessarily with the numbers you made up there), let D be the event that a randomly chosen person is a Democrat (with all 200 people equally likely), and B be the event that the person lives in the Blue State. Suppose that 10 people move from the Blue State to the Red State. Write P_{old} and P_{new} for probabilities before and after they move. Assume that people do not change parties,

so we have $P_{\text{new}}(D) = P_{\text{old}}(D)$. Is it possible that *both* $P_{\text{new}}(D|B) > P_{\text{old}}(D|B)$ and $P_{\text{new}}(D|B^c) > P_{\text{old}}(D|B^c)$ are true? If so, explain how it is possible and why it does not contradict the law of total probability $P(D) = P(D|B)P(B) + P(D|B^c)P(B^c)$; if not, show that it is impossible.

Solution:

(a) Here are two tables that are as desired:

Red	Dem	Rep	**Total**
Rich	5	25	30
Poor	20	50	70
Total	25	75	100

Blue	Dem	Rep	**Total**
Rich	45	15	60
Poor	35	5	40
Total	80	20	100

In these tables, within each state a rich person is more likely to be a Republican than a poor person; but the richer state has a higher percentage of Democrats than the poorer state. Of course, there are many possible tables that work.

The above example is a form of Simpson's paradox: aggregating the two tables seems to give different conclusions than conditioning on which state a person is in. Letting D, W, B be the events that a randomly chosen person is a Democrat, wealthy, and from the Blue State (respectively), for the above numbers we have $P(D|W, B) < P(D|W^c, B)$ and $P(D|W, B^c) < P(D|W^c, B^c)$ (controlling for whether the person is in the Red State or the Blue State, a poor person is more likely to be a Democrat than a rich person), but $P(D|W) > P(D|W^c)$ (stemming from the fact that the Blue State is richer and more Democratic).

(b) Yes, it is possible. Suppose with the numbers from (a) that 10 people move from the Blue State to the Red State, of whom 5 are Democrats and 5 are Republicans. Then $P_{\text{new}}(D|B) = 75/90 > 80/100 = P_{\text{old}}(D|B)$ and $P_{\text{new}}(D|B^c) = 30/110 > 25/100 = P_{\text{old}}(D|B^c)$. Intuitively, this makes sense since the Blue State has a higher percentage of Democrats initially than the Red State, and the people who move have a percentage of Democrats which is between these two values.

This result does not contradict the law of total probability since the weights $P(B), P(B^c)$ also change: $P_{\text{new}}(B) = 90/200$, while $P_{\text{old}}(B) = 1/2$. The phenomenon could not occur if an equal number of people also move from the Red State to the Blue State (so that $P(B)$ is kept constant).

Mixed practice

54. Fred decides to take a series of n tests, to diagnose whether he has a certain disease (any individual test is not perfectly reliable, so he hopes to reduce his uncertainty by taking multiple tests). Let D be the event that he has the disease, $p = P(D)$ be the prior probability that he has the disease, and $q = 1 - p$. Let T_j be the event that he tests positive on the jth test.

(a) Assume for this part that the test results are conditionally independent given Fred's disease status. Let $a = P(T_j|D)$ and $b = P(T_j|D^c)$, where a and b don't depend on j. Find the posterior probability that Fred has the disease, given that he tests positive on all n of the n tests.

(b) Suppose that Fred tests positive on all n tests. However, some people have a certain gene that makes them *always* test positive. Let G be the event that Fred has the gene. Assume that $P(G) = 1/2$ and that D and G are independent. If Fred does *not* have the gene, then the test results are conditionally independent given his disease status. Let $a_0 = P(T_j|D, G^c)$ and $b_0 = P(T_j|D^c, G^c)$, where a_0 and b_0 don't depend on j. Find the

posterior probability that Fred has the disease, given that he tests positive on all n of the tests.

Solution:

(a) Let $T = T_1 \cap \cdots \cap T_n$ be the event that Fred tests positive on all the tests. By Bayes' rule and LOTP,

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{pa^n}{pa^n + qb^n}.$$

(b) Let T be the event that Fred tests positive on all n tests. Then

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{pP(T|D)}{pP(T|D) + qP(T|D^c)}.$$

Conditioning on whether or not he has the gene, we have

$$P(T|D) = P(T|D,G)P(G|D) + P(T|D,G^c)P(G^c|D) = \frac{1}{2} + \frac{a_0^n}{2},$$

$$P(T|D^c) = P(T|D^c,G)P(G|D^c) + P(T|D^c,G^c)P(G^c|D^c) = \frac{1}{2} + \frac{b_0^n}{2}.$$

Thus,

$$P(D|T) = \frac{p(1 + a_0^n)}{p(1 + a_0^n) + q(1 + b_0^n)}.$$

55. A certain hereditary disease can be passed from a mother to her children. Given that the mother has the disease, her children independently will have it with probability $1/2$. Given that she doesn't have the disease, her children won't have it either. A certain mother, who has probability $1/3$ of having the disease, has two children.

(a) Find the probability that neither child has the disease.

(b) Is whether the elder child has the disease independent of whether the younger child has the disease? Explain.

(c) The elder child is found not to have the disease. A week later, the younger child is also found not to have the disease. Given this information, find the probability that the mother has the disease.

Solution:

(a) Let M, A, B be the events that the mother, elder child, and younger child have the disease (respectively). Then

$$P(A^c, B^c) = P(A^c, B^c|M)P(M) + P(A^c, B^c|M^c)P(M^c) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} + \frac{2}{3} = \frac{3}{4}.$$

(b) These events are conditionally independent given the disease status of the mother, but they are not independent. Knowing whether the elder child has the disease gives information about whether the mother has the disease, which in turn gives information about whether the younger child has the disease.

(c) By Bayes' rule,

$$P(M|A^c, B^c) = \frac{P(A^c, B^c|M)P(M)}{P(A^c, B^c)} = \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3}}{\frac{3}{4}} = \frac{1}{9}.$$

Alternatively, we can do the conditioning in two steps: first condition on A^c, giving

$$P(M|A^c) = \frac{P(A^c|M)P(M)}{P(A^c)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + \frac{2}{3}} = \frac{1}{5}.$$

Then do further conditioning on B^c, giving

$$P(M|A^c, B^c) = \frac{P(B^c|M, A^c)P(M|A^c)}{P(B^c|A^c)} = \frac{\frac{1}{2} \cdot \frac{1}{5}}{\frac{1}{2} \cdot \frac{1}{5} + \frac{4}{5}} = \frac{1}{9},$$

which agrees with the result of the one-step method.

56. Three fair coins are tossed at the same time. Explain what is wrong with the following argument: "there is a 50% chance that the three coins all landed the same way, since obviously it is possible to find two coins that match, and then the other coin has a 50% chance of matching those two".

Solution: The probability of all three coins landing the same way is $1/8 + 1/8 = 1/4$, so clearly there is something wrong with the argument. What is wrong is that *which* two out of the three coins match is not specified, so "the other coin" is not well-defined. Given that the first two flips match, it is true that there is a 50% chance that the third flip matches those, but knowing that there is *at least one* pair of tosses that match does not provide a *specific* pair that matches. (The argument under discussion is known as *Galton's paradox*.)

57. An urn contains red, green, and blue balls. Let r, g, b be the proportions of red, green, blue balls, respectively ($r + g + b = 1$).

(a) Balls are drawn randomly *with replacement*. Find the probability that the first time a green ball is drawn is before the first time a blue ball is drawn.

Hint: Explain how this relates to finding the probability that a draw is green, given that it is either green or blue.

(b) Balls are drawn randomly *without replacement*. Find the probability that the first time a green ball is drawn is before the first time a blue ball is drawn. Is the answer the same or different than the answer in (a)?

Hint: Imagine the balls all lined up, in the order in which they will be drawn. Note that where the red balls are standing in this line is irrelevant.

(c) Generalize the result from (a) to the following setting. Independent trials are performed, and the outcome of each trial is classified as being exactly one of type 1, type 2, ..., or type n, with probabilities $p_1, p_2, \ldots, p_n$, respectively. Find the probability that the first trial to result in type i comes before the first trial to result in type j, for $i \neq j$.

Solution:

(a) Red balls are irrelevant here. What matters is whether the first non-red ball drawn is green or blue. The probability that this ball is green is

$$P(\text{green}|\text{green or blue}) = \frac{P(\text{green})}{P(\text{green or blue})} = \frac{g}{g + b}.$$

(b) Let N be the total number of balls, so there are rN, gN, bN red, green, and blue balls, respectively. Line the balls up in the order in which they will be drawn. Now look at the first non-red ball. By symmetry, it is equally like to be any of the $gN + bN$ non-red balls. The probability that it is green is

$$\frac{gN}{gN + bN} = \frac{g}{g + b},$$

which is the same as the answer in (a).

(c) Arguing as in (a), the probability of getting type i before type j (for $i \neq j$ is

$$P(\text{type } i|\text{type } i \text{ or type } j) = \frac{P(\text{type } i)}{P(\text{type } i \text{ or type } j)} = \frac{p_i}{p_i + p_j}.$$

58. Marilyn vos Savant was asked the following question for her column in *Parade*:

> You're at a party with 199 other guests when robbers break in and announce that they are going to rob one of you. They put 199 blank pieces of paper in a hat, plus one marked "you lose." Each guest must draw, and the person who draws "you lose" will get robbed. The robbers offer you the option of drawing first, last, or at any time in between. When would you take your turn?

The draws are made *without replacement*, and for (a) are uniformly random.

(a) Determine whether it is optimal to draw first, last, or somewhere in between (or whether it does not matter), to maximize the probability of not being robbed. Give a clear, concise, and compelling explanation.

(b) More generally, suppose that there is one "you lose" piece of paper, with "weight" v, and there are n blank pieces of paper, each with "weight" w. At each stage, draws are made with probability proportional to weight, i.e., the probability of drawing a particular piece of paper is its weight divided by the sum of the weights of all the remaining pieces of paper. Determine whether it is better to draw first or second (or whether it does not matter); here $v > 0, w > 0$, and $n \geq 1$ are known constants.

Solution:

(a) By symmetry, it does not matter: unconditionally, the jth draw is equally likely to be any of the 200 pieces of paper.

(b) Drawing first, the probability of being robbed is $v/(v + nw)$. Drawing second, by LOTP the probability of being robbed is

$$0 \cdot \frac{v}{v + nw} + \frac{v}{v + (n-1)w} \cdot \frac{nw}{v + nw} = \frac{vnw}{(v + (n-1)w)(v + nw)}.$$

This is greater than $v/(v + nw)$ if and only if $nw > v + (n-1)w$, which is equivalent to $v < w$. Similarly, it is less than $v/(v + nw)$ if $v > w$, and equal if $v = w$. So it is better to draw first if $v < w$, and draw second if $v > w$ (and it does not matter if $v = w$). Interestingly, the optimal choice does not depend on n. Note that this result is correct in the simple case $n = 1$, in the case $v = w$ (which reduces to the previous part), and in the extreme case where v is much, much larger than w.

59. Let D be the event that a person develops a certain disease, and C be the event that the person was exposed to a certain substance (e.g., D may correspond to lung cancer and C may correspond to smoking cigarettes). We are interested in whether exposure to the substance is related to developing the disease (and if so, how they are related). The *odds ratio* is a very widely used measure in epidemiology of the association between disease and exposure, defined as

$$\mathrm{OR} = \frac{\mathrm{odds}(D|C)}{\mathrm{odds}(D|C^c)},$$

where conditional odds are defined analogously to unconditional odds: $\mathrm{odds}(A|B) = \frac{P(A|B)}{P(A^c|B)}$. The *relative risk* of the disease for someone exposed to the substance, another widely used measure, is

$$\mathrm{RR} = \frac{P(D|C)}{P(D|C^c)}.$$

The relative risk is especially easy to interpret, e.g., $\mathrm{RR} = 2$ says that someone exposed to the substance is twice as likely to develop the disease as someone who isn't exposed (though this does not necessarily mean that the substance *causes* the increased chance of getting the disease, nor is there necessarily a causal interpretation for the odds ratio).

(a) Show that if the disease is rare, both for exposed people and for unexposed people, then the relative risk is approximately equal to the odds ratio.

(b) Let p_{ij} for $i = 0, 1$ and $j = 0, 1$ be the probabilities in the following 2×2 table.

	D	D^c
C	p_{11}	p_{10}
C^c	p_{01}	p_{00}

For example, $p_{10} = P(C, D^c)$. Show that the odds ratio can be expressed as a *cross-product ratio*, in the sense that

$$\text{OR} = \frac{p_{11}p_{00}}{p_{10}p_{01}}.$$

(c) Show that the odds ratio has the neat symmetry property that the roles of C and D can be swapped without changing the value:

$$\text{OR} = \frac{\text{odds}(C|D)}{\text{odds}(C|D^c)}.$$

This property is one of the main reasons why the odds ratio is so widely used, since it turns out that it allows the odds ratio to be estimated in a wide variety of problems where relative risk would be hard to estimate well.

Solution:

(a) The odds ratio is related to the relative risk by

$$\text{OR} = \frac{P(D|C)/P(D^c|C)}{P(D|C^c)/P(D^c|C^c)} = \text{RR} \cdot \frac{P(D^c|C^c)}{P(D^c|C)}.$$

So $\text{OR} \approx \text{RR}$ if both $P(D^c|C^c)$ and $P(D^c|C)$ are close to 1.

(b) By definition of conditional probability,

$$\text{OR} = \frac{P(D|C)P(D^c|C^c)}{P(D|C^c)P(D^c|C)} = \frac{P(D,C)P(D^c,C^c)}{P(D,C^c)P(D^c,C)} = \frac{p_{11}p_{00}}{p_{10}p_{01}}.$$

(c) We have

$$\frac{\text{odds}(C|D)}{\text{odds}(C|D^c)} = \frac{P(C|D)P(C^c|D^c)}{P(C|D^c)P(C^c|D)} = \frac{P(C,D)P(C^c,D^c)}{P(C,D^c)P(C^c,D)} = \text{OR}.$$

60. A researcher wants to estimate the percentage of people in some population who have used illegal drugs, by conducting a survey. Concerned that a lot of people would lie when asked a sensitive question like "Have you ever used illegal drugs?", the researcher uses a method known as *randomized response*. A hat is filled with slips of paper, each of which says either "I have used illegal drugs" or "I have not used illegal drugs". Let p be the proportion of slips of paper that say "I have used illegal drugs" (p is chosen by the researcher in advance).

Each participant chooses a random slip of paper from the hat and answers (truthfully) "yes" or "no" to whether the statement on that slip is true. The slip is then returned to the hat. The researcher does not know which type of slip the participant had. Let y be the probability that a participant will say "yes", and d be the probability that a participant has used illegal drugs.

(a) Find y, in terms of d and p.

(b) What would be the worst possible choice of p that the researcher could make in designing the survey? Explain.

(c) Now consider the following alternative system. Suppose that proportion p of the slips of paper say "I have used illegal drugs", but that now the remaining $1 - p$ say "I was

born in winter" rather than "I have not used illegal drugs". Assume that 1/4 of people are born in winter, and that a person's season of birth is independent of whether they have used illegal drugs. Find d, in terms of y and p.

Solution:

(a) Let A be the event that the participant will draw a "I have used illegal drugs" slip, and Y be the event that there is a "yes" response. By the law of total probability,

$$P(Y) = P(Y|A)P(A) + P(Y|A^c)P(A^c),$$

so

$$y = dp + (1 - d)(1 - p).$$

(b) The worst possible choice is $p = 1/2$, since then the survey gives no information about d, which is the main quantity of interest. Mathematically, this can be seen by trying to solve the equation from (a) for d in terms of y and p; this gives $d = (y + p - 1)/(2p - 1)$ for $p \neq 1/2$, but would involve dividing by 0 for $p = 1/2$. Intuitively, this makes sense since the $p = 1/2$ case is like dealing with someone who tells the truth half the time and lies half the time (someone who always tells the truth or always lies is much more informative!).

(c) LOTP gives

$$y = dp + \frac{1 - p}{4},$$

so $d = (y - 1/4)/p + 1/4$ (for $p \neq 0$). This makes sense since if $p = 0$, the survey is only asking about whether people were born in winter (and then $y = 1/4$, and the survey gives no information about d), while if $p = 1$ the survey is only asking about drug use (and then $d = y$, and the survey isn't actually using the randomized response idea).

61. At the beginning of the play *Rosencrantz and Guildenstern are Dead* by Tom Stoppard, Guildenstern is spinning coins and Rosencrantz is betting on the outcome for each. The coins have been landing Heads over and over again, prompting the following remark:

> *Guildenstern*: A weaker man might be moved to re-examine his faith, if in nothing else at least in the law of probability.

The coin spins have resulted in Heads 92 times in a row.

(a) Fred and his friend are watching the play. Upon seeing the events described above, they have the following conversation:

> *Fred*: That outcome would be incredibly unlikely with fair coins. They must be using trick coins (maybe with double-headed coins), or the experiment must have been rigged somehow (maybe with magnets).
> *Fred's friend:* It's true that the string HH...H of length 92 is very unlikely; the chance is $1/2^{92} \approx 2 \times 10^{-28}$ with fair coins. But *any* other specific string of H's and T's with length 92 has *exactly* the same probability! The reason the outcome seems extremely unlikely is that the number of possible outcomes grows exponentially as the number of spins grows, so *any* outcome would seem extremely unlikely. You could just as well have made the same argument even without looking at the results of their experiment, which means you really don't have evidence against the coins being fair.

Discuss these comments, to help Fred and his friend resolve their debate.

(b) Suppose there are only two possibilities: either the coins are all fair (and spun fairly), or double-headed coins are being used (in which case the probability of Heads is 1). Let p be the prior probability that the coins are fair. Find the posterior probability that the coins are fair, given that they landed Heads in 92 out of 92 trials.

(c) Continuing from (b), for which values of p is the posterior probability that the coins are fair greater than 0.5? For which values of p is it less than 0.05?

Solution:

(a) Fred is correct that the outcome HH...H is extremely suspicious, but Fred's friend is correct that with a fair coin, any specific string of H's and T's of length 92 has the same probability. To reconcile these statements, note that there is only 1 string of length 92 with 92 H's, whereas there are a vast number of strings with about half H's and about half T's. For example, there are $\binom{92}{46} \approx 4.1 \times 10^{26}$ strings of length 92 with 46 H's and 46 T's. It is enormously more likely to have 46 H's and 46 T's (with a fair coin) than 92 H's, even though any specific string with 46 H's and 46 T's has the same probability as HH...H. Furthermore, there are an even vaster number of possibilities where the number of Heads is roughly equal to the number of Tails.

We should check, however, that Fred is not just doing *data snooping* (also known as *data fishing*) after observing the data. By trying a lot of different calculations with the data that are not prescribed in advance, it is easy to find unusual patterns in a data set, but it is highly dangerous and misleading to present such results without reporting how many calculations were tried, and when and how it was decided which calculations to do. In this case though, looking at the number of Heads in a sequence of coin tosses is a very simple, common, natural summary of the data, and there is no reason to think that Fred was fishing for something suspicious.

There could also be a *selection bias* at play: maybe millions of such experiments are performed, by people all over the world and at different points in time, and Tom Stoppard wrote about this particular incident *because* the outcome was so interesting. Likewise, it's not surprising to read in the news that someone won the lottery, when millions of people are entering the lottery; the news article reports on the person who won the lottery *because* they won. But having no Heads in 92 trials is so staggeringly unlikely (for a fair coin) that there would be good reason to be suspicious even if several billion people were performing similar experiments and only the most extreme outcome was selected to include in the play.

(b) Let F be the event that the coins are fair and A be the event that they land Heads in 92 out of 92 trials. By Bayes' rule and LOTP,

$$P(F|A) = \frac{P(A|F)P(F)}{P(A)} = \frac{2^{-92} \cdot p}{2^{-92} \cdot p + 1 - p} = \frac{p}{p + 2^{92}(1-p)}.$$

(c) We have that

$$\frac{p}{p + 2^{92}(1-p)} > \frac{1}{2}$$

is equivalent to

$$p > \frac{2^{92}}{2^{92} + 1}.$$

This says that p must be incredibly close to 1, since

$$\frac{2^{92}}{2^{92} + 1} = 1 - \frac{1}{2^{92} + 1} \approx 1 - 2 \times 10^{-28}.$$

On the other hand,

$$\frac{p}{p + 2^{92}(1-p)} < 0.05$$

is equivalent to

$$p < \frac{2^{92}}{2^{92} + 19}.$$

Unless p is incredibly close to 1, the above inequality will hold.

62. There are n types of toys, which you are collecting one by one. Each time you buy a toy, it is randomly determined which type it has, with equal probabilities. Let p_{ij} be the probability that just after you have bought your ith toy, you have exactly j toy types in your collection, for $i \geq 1$ and $0 \leq j \leq n$. (This problem is in the setting of the *coupon collector* problem, a famous problem which we study in Example 4.3.11.)

    (a) Find a recursive equation expressing p_{ij} in terms of $p_{i-1,j}$ and $p_{i-1,j-1}$, for $i \geq 2$ and $1 \leq j \leq n$.

    (b) Describe how the recursion from (a) can be used to calculate p_{ij}.

    Solution:

    (a) There are two ways to have exactly j toy types just after buying your ith toy: either you have exactly $j-1$ toy types just after buying your $i-1$st toy and then the ith toy you buy is of a type you don't already have, or you already have exactly j toy types just after buying your $i-1$st toy and then the ith toy you buy is of a type you do already have. Conditioning on how many toy types you have just after buying your $i-1$st toy,

    $$p_{ij} = p_{i-1,j-1} \frac{n-j+1}{n} + p_{i-1,j} \frac{j}{n}.$$

    (b) First note that $p_{11} = 1$, and $p_{ij} = 0$ for $j = 0$ or $j > i$. Now suppose that we have computed $p_{i-1,1}, p_{i-1,2}, \ldots, p_{i-1,i-1}$ for some $i \geq 2$. Then we can compute $p_{i,1}, p_{i,2}, \ldots, p_{i,i}$ using the recursion from (a). We can then compute $p_{i+1,1}, p_{i+1,2}, \ldots, p_{i+1,i+1}$ using the recursion from (a), and so on. By induction, it follows that we can obtain any desired p_{ij} recursively by this method.

63. *A/B testing* is a form of randomized experiment that is used by many companies to learn about how customers will react to different treatments. For example, a company may want to see how users will respond to a new feature on their website (compared with how users respond to the current version of the website) or compare two different advertisements.

    As the name suggests, two different treatments, Treatment A and Treatment B, are being studied. Users arrive one by one, and upon arrival are randomly assigned to one of the two treatments. The trial for each user is classified as "success" (e.g., the user made a purchase) or "failure". The probability that the nth user receives Treatment A is allowed to depend on the outcomes for the previous users. This set-up is known as a *two-armed bandit*.

    Many algorithms for how to randomize the treatment assignments have been studied. Here is an especially simple (but fickle) algorithm, called a *stay-with-a-winner* procedure:

    (i) Randomly assign the first user to Treatment A or Treatment B, with equal probabilities.

    (ii) If the trial for the nth user is a success, stay with the same treatment for the $(n+1)$st user; otherwise, switch to the other treatment for the $(n+1)$st user.

    Let a be the probability of success for Treatment A, and b be the probability of success for Treatment B. Assume that $a \neq b$, but that a and b are unknown (which is why the test is needed). Let p_n be the probability of success on the nth trial and a_n be the probability that Treatment A is assigned on the nth trial (using the above algorithm).

    (a) Show that

    $$p_n = (a-b)a_n + b,$$
    $$a_{n+1} = (a+b-1)a_n + 1 - b.$$

    (b) Use the results from (a) to show that p_{n+1} satisfies the following recursive equation:

    $$p_{n+1} = (a+b-1)p_n + a + b - 2ab.$$

(c) Use the result from (b) to find the long-run probability of success for this algorithm, $\lim_{n\to\infty} p_n$, assuming that this limit exists.

Solution:

(a) Conditioning on which treatment is applied to the nth user, we have

$$p_n = a_n a + (1 - a_n)b = (a - b)a_n + b.$$

Treatment A will be applied to the $(n + 1)$st user if and only if (i) Treatment A is successfully applied to the nth user or (ii) Treatment B is unsuccessfully applied to the nth user. Again conditioning on which treatment is applied to the nth user, we have

$$a_{n+1} = a a_n + (1 - b)(1 - a_n) = (a + b - 1)a_n + 1 - b.$$

(b) Plugging the recursion for a_{n+1} into the recursion

$$p_{n+1} = (a - b)a_{n+1} + b,$$

we have

$$p_{n+1} = (a - b)((a + b - 1)a_n + 1 - b) + b.$$

Replacing a_n by $(p_n - b)/(a - b)$, we have

$$\begin{aligned}
p_{n+1} &= (a - b)\left((a + b - 1)\frac{p_n - b}{a - b} + 1 - b\right) + b \\
&= (a + b - 1)(p_n - b) + (a - b)(1 - b) + b \\
&= (a + b - 1)p_n + a + b - 2ab.
\end{aligned}$$

(c) Let $p = \lim_{n\to\infty} p_n$ (assuming that this exists). Taking the limit as $n \to \infty$ on both sides of

$$p_{n+1} = (a + b - 1)p_n + a + b - 2ab,$$

we have

$$p = (a + b - 1)p + a + b - 2ab.$$

Therefore,

$$p = \frac{a + b - 2ab}{2 - a - b}.$$

64. In humans (and many other organisms), genes come in pairs. A certain gene comes in two types (*alleles*): type a and type A. The *genotype* of a person for that gene is the types of the two genes in the pair: AA, Aa, or aa (aA is equivalent to Aa). Assume that the *Hardy-Weinberg law* applies here, which means that the frequencies of AA, Aa, aa in the population are $p^2, 2p(1 - p), (1 - p)^2$ respectively, for some p with $0 < p < 1$.

When a woman and a man have a child, the child's gene pair consists of one gene contributed by each parent. Suppose that the mother is equally likely to contribute either of the two genes in her gene pair, and likewise for the father, independently. Also suppose that the genotypes of the parents are independent of each other (with probabilities given by the Hardy-Weinberg law).

(a) Find the probabilities of each possible genotype (AA, Aa, aa) for a child of two random parents. Explain what this says about stability of the Hardy-Weinberg law from one generation to the next.

Hint: Condition on the genotypes of the parents.

(b) A person of type AA or aa is called *homozygous* (for the gene under consideration), and a person of type Aa is called *heterozygous* (for that gene). Find the probability

that a child is homozygous, given that both parents are homozygous. Also, find the probability that a child is heterozygous, given that both parents are heterozygous.

(c) Suppose that having genotype aa results in a distinctive physical characteristic, so it is easy to tell by looking at someone whether or not they have that genotype. A mother and father, neither of whom are of type aa, have a child. The child is also not of type aa. Given this information, find the probability that the child is heterozygous.

Hint: Use the definition of conditional probability. Then expand both the numerator and the denominator using LOTP, conditioning on the genotypes of the parents.

Solution:

(a) Let M_{AA}, F_{AA}, C_{AA} be the events that the mother, father, and child (respectively) have genotype AA, and likewise define M_{Aa} etc. Then

$$
\begin{aligned}
P(C_{AA}) &= P(C_{AA}|M_{AA}, F_{AA})P(M_{AA}, F_{AA}) + P(C_{AA}|M_{AA}, F_{Aa})P(M_{AA}, F_{Aa}) \\
&\quad + P(C_{AA}|M_{Aa}, F_{AA})P(M_{Aa}, F_{AA}) + P(C_{AA}|M_{Aa}, F_{Aa})P(M_{Aa}, F_{Aa}) \\
&= p^4 + \frac{1}{2}p^2(2p(1-p)) + \frac{1}{2}p^2(2p(1-p)) + \frac{1}{4}(2p(1-p))^2 \\
&= p^4 + 2p^3 - 2p^4 + p^2(1 - 2p + p^2) \\
&= p^2.
\end{aligned}
$$

Let $q = 1 - p$. Swapping A's and a's and swapping p's and q's in the above calculation,

$$
P(C_{aa}) = q^2 = (1-p)^2.
$$

It follows that

$$
P(C_{Aa}) = 1 - P(C_{AA}) - P(C_{aa}) = 1 - p^2 - (1-p)^2 = 2p - 2p^2 = 2p(1-p).
$$

So the Hardy-Weinberg law is stable, in the sense that the probabilities for the various genotypes are preserved from one generation to the next.

(b) Let H be the event that both parents are homozygous. Then

$$
P(\text{child homozygous}|H) = P(C_{AA}|H) + P(C_{aa}|H).
$$

To find $P(C_{AA}|H)$, we can do further conditioning on the exact genotypes of the parents:

$$
\begin{aligned}
P(C_{AA}|H) &= P(C_{AA}|M_{AA}, F_{AA})P(M_{AA}, F_{AA}|H) + P(C_{AA}|M_{AA}, F_{aa})P(M_{AA}, F_{aa}|H) \\
&\quad + P(C_{AA}|M_{aa}, F_{AA})P(M_{aa}, F_{AA}|H) + P(C_{AA}|M_{aa}, F_{aa})P(M_{aa}, F_{aa}|H) \\
&= P(M_{AA}, F_{AA}|H),
\end{aligned}
$$

since all the terms except the first are zero (because, for example, an AA mother and aa father can't produce an AA child). This result can also be seen directly by noting that, given H, the child being AA is equivalent to both parents being AA. Next, we have

$$
P(M_{AA}, F_{AA}|H) = \frac{P(M_{AA}, F_{AA}, H)}{P(H)} = \frac{P(M_{AA}, F_{AA})}{P(H)} = \frac{p^4}{(p^2 + (1-p)^2)^2}.
$$

By symmetry, we have

$$
P(C_{aa}|H) = P(M_{aa}, F_{aa}|H) = \frac{(1-p)^4}{((1-p)^2 + p^2)^2}.
$$

Hence,

$$
P(\text{child homozygous}|H) = \frac{p^4 + (1-p)^4}{(p^2 + (1-p)^2)^2}.
$$

Lastly,

$$P(\text{child heterozygous}|\text{both parents heterozygous}) = \frac{1}{2}$$

since if both parents are Aa, then a child will be heterozygous if and only if they receive the mother's A and the father's a or vice versa.

(c) We wish to find $P(C_{Aa}|C_{aa}^c \cap M_{aa}^c \cap F_{aa}^c)$. Let

$$G = C_{Aa} \cap M_{aa}^c \cap F_{aa}^c \text{ and } H = C_{aa}^c \cap M_{aa}^c \cap F_{aa}^c.$$

Then

$$P(C_{Aa}|C_{aa}^c \cap M_{aa}^c \cap F_{aa}^c) = \frac{P(G)}{P(H)},$$

and

$$\begin{aligned}
P(G) &= P(G|M_{AA}, F_{AA})P(M_{AA}, F_{AA}) + P(G|M_{AA}, F_{Aa})P(M_{AA}, F_{Aa}) \\
&\quad + P(G|M_{Aa}, F_{AA})P(M_{Aa}, F_{AA}) + P(G|M_{Aa}, F_{Aa})P(M_{Aa}, F_{Aa}) \\
&= 0 + \frac{1}{2}p^2(2p(1-p)) + \frac{1}{2}p^2(2p(1-p)) + \frac{1}{2}(2p(1-p))^2 \\
&= 2p^2(1-p),
\end{aligned}$$

while

$$\begin{aligned}
P(H) &= P(H|M_{AA}, F_{AA})P(M_{AA}, F_{AA}) + P(H|M_{AA}, F_{Aa})P(M_{AA}, F_{Aa}) \\
&\quad + P(H|M_{Aa}, F_{AA})P(M_{Aa}, F_{AA}) + P(H|M_{Aa}, F_{Aa})P(M_{Aa}, F_{Aa}) \\
&= p^4 + p^2(2p(1-p)) + p^2(2p(1-p)) + \frac{3}{4}(2p(1-p))^2 \\
&= p^2(3 - 2p).
\end{aligned}$$

Thus,

$$P(C_{Aa}|C_{aa}^c \cap M_{aa}^c \cap F_{aa}^c) = \frac{2p^2(1-p)}{p^2(3-2p)} = \frac{2-2p}{3-2p}.$$

65. A standard deck of cards will be shuffled and then the cards will be turned over one at a time until the first ace is revealed. Let B be the event that the *next* card in the deck will also be an ace.

(a) Intuitively, how do you think $P(B)$ compares in size with $1/13$ (the overall proportion of aces in a deck of cards)? Explain your intuition. (Give an intuitive discussion rather than a mathematical calculation; the goal here is to describe your intuition explicitly.)

(b) Let C_j be the event that the first ace is at position j in the deck. Find $P(B|C_j)$ in terms of j, fully simplified.

(c) Using the law of total probability, find an expression for $P(B)$ as a sum. (The sum can be left unsimplified, but it should be something that could easily be computed in software such as R that can calculate sums.)

(d) Find a fully simplified expression for $P(B)$ using a symmetry argument.

Hint: If you were deciding whether to bet on the next card after the first ace being an ace or to bet on the last card in the deck being an ace, would you have a preference?

Solution:

(a) Intuitively, it may seem that $P(B) < 1/13$ since we know an ace has been depleted from the deck and don't know anything about how many (if any) non-aces were depleted. On the other hand, it clearly matters how many cards are needed to reach the first ace (in the extreme case where the first ace is at the 49th card, we *know* the next 3 cards

are aces). But we are not given information about the number of cards needed to reach the first ace. Of course, we can condition on it, which brings us to the next two parts.

(b) There are $52 - j$ remaining cards in the deck, of which 3 are aces. By symmetry, the next card is equally likely to be any of the remaining cards. So

$$P(B|C_j) = \frac{3}{52 - j}.$$

(c) LOTP yields

$$P(B) = \sum_{j=1}^{49} P(B|C_j)P(C_j) = \sum_{j=1}^{49} \frac{\binom{48}{j-1}}{\binom{52}{j-1}} \cdot \frac{4}{52 - j + 1} \cdot \frac{3}{52 - j},$$

where the ratio of binomial coefficients is used to get the probability of the deck starting out with $j - 1$ non-aces. Computing the sum was not required for this part, but to do so in R we can type

```
j <- 1:49
sum(choose(48,j-1)/choose(52,j-1)*(4/(52-j+1))*(3/(52-j)))
```

(d) At any stage, by symmetry all unrevealed cards are completely interchangeable (in statistics, they are said to be *exchangeable*). Let L be the event that the last card in the deck is an ace. By symmetry, $P(B|C_j) = P(L|C_j)$ for any j. So $P(B) = P(L)$ unconditionally too (by LOTP). Thus,

$$P(B) = P(L) = \frac{1}{13}.$$