Probability Solutions

Emiliano Jimenez Guadarrama

September 26, 2023

Abstract

These are my solutions to the book *Introduction to Probability* by Joseph K. Blitzstein and Jessica Hwang. The solutions include the problems developed by these two authors just for reference; I do not claim ownership over the problems.

I decided to do a selection of the problems of this book due to my lack of understanding of this topic even after taking an introductory class to probability. I started to work on this at the end of May 2023 using overleaf, although chances are you are reading this using github.

1   1.9 Counting

1.1   1

How many ways are there to permute the letters in the word MISSISSIPPI?

   There are $(11!)/(4! \cdot 4! \cdot 2!) = 34650$

1.2   2

1. How many 7-digit phone numbers are possible, assuming that the first digit can't be a 0 or a 1?

    The solution is: $10^6 \times 8 = 8,000,000$

2. Re-solve (a), except now assume also that the phone number is not allowed to start with 911 (since this is reserved for emergency use, and it would not be desirable for the system to wait to see whether more digits were going to be dialed after someone has dialed 911).

    The solution is $10^7 - 10^4 = 9990000$

1.3   4

A round-robin tournament is being held with n tennis players; this means that every player will play against every other player exactly once.

1. How many possible outcomes are there for the tournament (the outcome lists out who won and who lost for each game)? **There are** $2^{\frac{n(n-1)}{2}}$ **possible outcomes.**

2. How many games are played in total? **There are** $[n(n-1)]/2$ **games being played.**

1.4   5

A knock-out tournament is being held with 2^n tennis players. This means that for each round, the winners move on to the next round and the losers are eliminated, until only one person remains. For example, if initially there are $2^4 = 16$ players, then there are 8 games in the first round, then the 8 winners move on to round 2, then the 4 winners move on to round 3, then the 2 winners move on to round 4, the winner of which is declared the winner of the tournament. (There are various systems for determining who plays whom within a round, but these do not matter for this problem.)

1. How many rounds are there?

   n

2. Count how many games in total are played, by adding up the numbers of games played in each round.

   $2^n/2 + 2^n/4 + \cdots + 2^n/2^n$ **or according to the geometric series** $1(2^n - 1)/(2 - 1) = 2^n - 1$

3. Count how many games in total are played, this time by directly thinking about it without doing almost any calculation.

   Since in each match you eliminate 1 person you need to do $2^n - 1$

Hint: How many players need to be eliminated?

1.5   6

There are 20 people at a chess club on a certain day. They each find opponents and start playing. How many possibilities are there for how they are matched up, assuming that in each game it does matter who has the white pieces (in a chess game, one player has the white pieces and the other player has the black pieces)?

   There are $10! \cdot \binom{20}{10}$ **possibilities. Because there are** $\binom{20}{10}$ **ways of choosing the white players and then there are** $10!$ **ways of choosing an opponent for the white pieces.**

1.6   10

To fulfill the requirements for a certain degree, a student can choose to take any 7 out of a list of 20 courses, with the constraint that at least 1 of the 7 courses must be a statistics course. Suppose that 5 of the 20 courses are statistics courses.

1. How many choices are there for which 7 courses to take?

   $\binom{5}{1} \times \binom{15}{6} + \binom{5}{2} \times \binom{15}{5} + \binom{5}{3} \times \binom{15}{4} + \binom{5}{3} \times \binom{15}{4} + \binom{5}{4} \times \binom{15}{3} + \binom{5}{5} \times \binom{15}{2}.$

2. Explain intuitively why the answer to (a) is not $\binom{5}{1} \times \binom{19}{6}$

   Because this means from the 5 statistics courses chose only 1 and then form the remaining 19 courses left chose 6. The second binomial coefficient might or might not add the remaining possibilities of statistic courses.

1.7   14

You are ordering two pizzas. A pizza can be small, medium, large, or extra large, with any combination of 8 possible toppings (getting no toppings is allowed, as is getting all 8). How many possibilities are there for your two pizzas?

   There should be 4×2^8 **possibilities for 1 pizza. For two pizzas we use Bose-Einstein which results in** $\binom{1024 + 2 - 1}{2} = \binom{1025}{2} = 524800$ **possibilities.**

1.8   15

Give a story proof that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

Using the binomial coefficient we are the number of subsets of all the sizes from zero up to n and 2^n is the number of subsets of a set of size n.

1.9   16

Show that for all positive integers n and k with $n \geq k, nk! + nk - 1! = n + 1k!$, doing this in two ways: (a) algebraically and (b) with a story, giving an interpretation for why both sides count the same thing. Hint for the story proof: Imagine an organization consisting of $n + 1$ people, with one of them pre-designated as the president of the organization.

1.10   17

Give a story proof that n X k=0 n k!2 = 2n n!, for all positive integers n.

1.11   24

A certain family has 6 children, consisting of 3 boys and 3 girls. Assuming that all birth orders are equally likely, what is the probability that the 3 eldest children are the 3 girls?
In three methods(just to test my understanding):
   As individual probabilities,

$$1/2 \cdot 2/5 \cdot 1/4 = 1/20$$

The probability of the first second and third being girls multiplied.
   As permutations,

$$[(3!)(3!)]/(6!)$$

Think of A,B,C as the girls and D,E,F as the boys. There are 3! permutations of the girls at first and then 3! of the boys and in total 6! permutations of the letters.
   As sets,

$$\frac{1}{\binom{6}{3}}$$

If we use the same analogy as before we just need to be able to chose (as in the committee example) a subset of tree specifically $\{A, B, C\}$ which is only one of all the possible subsets of size 3, since we don't care about the boys we don't need to count them.

1.12   25

A city with 6 districts has 6 robberies in a particular week. Assume the robberies are located randomly, with all possibilities for which robbery occurred where equally likely. What is the probability that some district had more than 1 robbery?
   The probability is

$$1 - \frac{6!}{6^6}$$

So think of a dice if you roll it 6 times there are 6^6 possible outcomes we want the outcomes where something repeats at least once that should be the complement to if it doesn't repeat. Now the question is how do we represent the unique outcomes? the answer is a permutation 6!; Putting it all together we have something like above.

1.13   26

A survey is being conducted in a city with 1 million residents. It would be far too expensive to survey all of the residents, so a random sample of size 1000 is chosen (in practice, there are many challenges with sampling, such as obtaining a complete list of everyone in the city, and dealing with people who refuse to participate). The survey is conducted by choosing people one at a time, with replacement and with equal probabilities.
(a) Explain how sampling with vs. without replacement here relates to the birthday problem.

   In this problem we use the same process as in the birthday problem. A sampling with replacement in the denominator. In the birthday problem the sampling with replacement corresponds to the ways in which we can choose birth dates. And in the denominator we use sampling without replacement, but in this case we use the binomial coefficient instead of n permutation k.
(b) Find the probability that at least one person will get chosen more than once.

   We will get it using the negation of the statement

$$1 - \frac{\binom{10^6}{10^3}}{(10^6)^{10^3}}$$

   The reasoning behind this is that we need to get 1 - the probability that no person will be chosen twice. So first we get the denominator from the 1 million people take 1 and repeat this process a 1000 times with replacement. For the numerator we take the million and we choose from there a 1000 without replacement.

1.14   28

A college has 10 time slots for its courses, and blithely assigns courses to completely random time slots, independently. The college offers exactly 3 statistics courses. What is the probability that 2 or more of the statistics courses are in the same time slot?

$$1 - \frac{10 \cdot 9 \cdot 8 \cdot 7^7}{10^{10}}$$

   Explanation: we will do it by complement so we need the probability of each one of these statistics courses to be in unique positions. So the total set has $10^{1}0$ possible outcomes and we have 3 statistics courses so we chose a unique position for the tree of them we get $10 \cdot 9 \cdot 8$ and for the remaining things we don't need to care if they repeat or not as long as we dont touch the 3 statistic courses so we have 7^7.

1.15   29

For each part, decide whether the blank should be filled in with $=, <, or >$, and give a clear explanation.
(a) (probability that the total after rolling 4 fair dice is 21)>(probability that the total after rolling 4 fair dice is 22)
(b) (probability that a random 2-letter word is a palindrome)= (probability that a random 3-letter word is a palindrome)

   Justification for part (a) there are 3 ways to get 21

$$6, 6, 6, 3 \quad 6, 6, 5, 4 \quad 6, 5, 5, 5$$

they can each be written in

$$\frac{4!}{3!} = 4 \quad \frac{4!}{2!} = 12 \quad \frac{4!}{3!} = 4$$

ways respectively and therefore the probability of getting 21 is

$$2\frac{4}{6^4} + \frac{6}{6^4} = 0.0108$$

and 22 can be written as

$$6, 6, 6, 4 \quad 6, 6, 5, 5$$

and these can be written in

$$\frac{4!}{3!} = 4 \qquad \frac{4!}{2! \cdot 2!} = \frac{6}{4}$$

ways respectively and therefore the probability of getting 22 is

$$\frac{4}{6^4} + \frac{\frac{6}{4}}{6^4} = 0.00424$$

Justification for part (b) we have for a 2-letter palindrome we have a probability of

$$\frac{26}{26 \cdot 26} = 0.0384$$

for a 3 letter palyndrome ~~we may think of this as aba or aaa~~ we may think of this as having the first and last letters being the same, so we have a probability of

$$\frac{26}{26} \cdot \frac{26}{26} \cdot \frac{1}{26} = 0.0384$$

same as in the two letter one because we are doing the same thing (think carefully).

1.16   30

With definitions as in the previous problem, find the probability that a random n-letter word is a palindrome for n = 7 and for n = 8.

Let's think about this as if we were doing an algorithm from trial and error I found that when $n = 7$ we can put the first 4 randomly and then select the last 3 ourselves, in the same way when $n = 8$ we may get 4 randomly and we can make a palindrome by selecting the last 4 ourselves.

Therefore

$$\frac{26^4}{26^7}$$

and

$$\frac{26^4}{26^8}$$

are the solutions.

1.17   31

Elk dwell in a certain forest. There are N elk, of which a simple random sample of size n are captured and tagged ("simple random sample" means that all $\binom{N}{n}$ sets of n elk are equally likely). The captured elk are returned to the population, and then a new sample is drawn, this time with size m. This is an important method that is widely used in ecology, known as capture-recapture. What is the probability that exactly k of the m elk in the new sample were previously tagged? (Assume that an elk that was captured before doesn't become more or less likely to be captured again.)

$$\frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}$$

The total possibilities are the product of drawing the first n out of N and drawing the other m out of N but we don't have to take into account the taking of the first n since we just need the number (they don't change the odds). And for the possibilities above first we have to take into account that we need to get the ways in which we can select the first n, $\binom{n}{k}$. Now that we have selected what we need, the **rest** of $m - k$ what we need to select we need to take it from outside of the first group so $N - n$.

1.18 34

A random 5-card poker hand is dealt from a standard deck of cards. Find the probability of each of the following possibilities (in terms of binomial coefficients).
(a) A flush (all 5 cards being of the same suit; do not count a royal flush, which is a flush with an ace, king, queen, jack, and 10).

$$\frac{\binom{4}{1} \cdot \binom{13}{5}}{\binom{52}{5}} - \frac{\binom{4}{1}}{\binom{52}{5}}$$

The first fraction is the probability of getting a flush counting the royal flush. First we count the 4 possible suits and then we multiply them by choosing 5 from the 13 cards in each suit; we divide the whole thing over choosing a 5 card hand. Now we subtract the probability of getting a royal flush. We notice that there are only 4 such hands in the entire deck.
(b) Two pair (e.g., two 3's, two 7's, and an ace).

$$\frac{\binom{13}{2} \cdot \binom{3}{1} \cdot \binom{3}{1} \cdot \binom{48}{1}}{\binom{52}{5}}$$

We choose the value of 2 card out of 13 and we search for their pair in the remaining 3 suits. We do it once and then (we have already chosen one) we do this same process again. And for the last card we choose it randomly out of the $52 - 8 = 44$ remaining, since we cannot choose anything that might turn the double into a triple. We divide the whole thing by the ways in which we can get 5 cards.

1.19 39

An organization with 2n people consists of n married couples. A committee of size k is selected, with all possibilities equally likely. Find the probability that there are exactly j married couples within the committee.

$$\frac{\binom{n}{j} \cdot \binom{n-2j}{k-2j}}{\binom{2n}{k}}$$

This can be thought as a combination of choosing a committee and how to get the probability of a certain arrangement of cards.

1.20 41

Each of n balls is independently placed into one of n boxes, with all boxes equally likely. What is the probability that exactly one box is empty?
Let's do the contrary to this problem so we need to search for $1-$ probability of each ball landing in a different box. Therefore the result is

$$1 - \frac{1}{n!}$$

1.21   42

A norepeatword is a sequence of at least one (and possibly all) of the usual 26 letters a,b,c,...,z, with repetitions not allowed. For example, "course" is a norepeatword, but "statistics" is not. Order matters, e.g., "course" is not the same as "source". A norepeatword is chosen randomly, with all norepeatwords equally likely. Show that the probability that it uses all 26 letters is very close to 1/e.

$$\frac{26!}{\sum_{i=1}^{26} \binom{26}{i} \cdot i!}$$

There are 26! ways of making a norepeatword with 26 letters and we divide this number by the total amount of norepeatwords possible which we get by grouping them using a binomial and then a permutation for the actual counting.

For the $1/e$ part

$$\frac{26!}{\sum_{i=1}^{26} \binom{26}{i} \cdot i!} = \frac{26!}{\sum_{i=1}^{26} \frac{26!}{\not{i!} \cdot (26-i)!} \cdot \not{i!}} =$$

$$\frac{\not{26!}}{\not{26!} \cdot \sum_{i=1}^{26} \frac{1}{(26-i)!}} =$$

$$\frac{1}{\sum_{i=1}^{26} \frac{1}{(26-i)!}} \approx \frac{1}{\sum_{k=0}^{\infty} \frac{1}{k!}} = \frac{1}{e}$$

1.22   44

Let A and B be events. The difference $B - A$ is defined to be the set of all elements of B that are not in A. Show that if $A \subseteq B$, then

$$P(B - A) = P(B) - P(A),$$

directly using the axioms of probability.

Proof. Suppose $A \subseteq B$ then we write B as the union of A and $B \cap A^C$, where $B \cap A^C$ is the part of B not also in A or in other words $B - A$.

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

$$P(B - A) = P(B \cap A^C) = P(B) - P(A)$$

this is congruent with the axioms of probability particularly because we know $P(A) \leq P(B)$ so $1 \leq P(B - A) \geq 0$, since $P(B)$ and $P(A)$ are smaller or equal to 1.

$\square$

1.23   46

Let $A_1, A_2, ..., A_n$ be events. Let Bk be the event exactly k of the A_i occur, and C_k be the event that at least k of the A_i occur, for $0 \leq k \leq n$. Find a simple expression for $P(B_k)$ in terms of $P(C_k)$ and $P(C_{k+1})$.
The expression would be:

$$P(B_k) = P(C_k) - P(C_{k+1})$$

Since the Probability of exactly k is the probability of at least k (meaning from k to n) minus the probability of at least k +1 (meaning from k+1 to n). By subtracting we get exactly k.

1.24   47

Events A and B are independent if $P(A \cap B) = P(A)P(B)$ (independence is explored in detail in the next chapter).

(a) Give an example of independent events A and B in a finite sample space S (with neither equal to ∅ or S), and illustrate it with a Pebble World diagram.

S is 2 dice throws (dice-1,dice-2).
A is the event of getting 2 in dice-1.
B is the event of getting 4 in dice-2.

(b) Consider the experiment of picking a random point in the rectangle

$$R = \{(x,y) : 0 < x < 1, 0 < y < 1\},$$

where the probability of the point being in any particular region contained within R is the area of that region. Let A1 and B1 be rectangles contained within R, with areas not equal to 0 or 1. Let A be the event that the random point is in A1, and B be the event that the random point is in B1. Give a geometric description of when it is true that A and B are independent. Also, give an example where they are independent and another example where they are not independent.

They are independent if they do not share a value of x or y at any point; They are dependent if they do share values.

(c) Show that if A and B are independent, then

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B) = 1 - P(A^c) \cdot P(B^c).$$

We know from the properties of probability that:

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

Then since $P(A \cap B) = P(A) \cdot P(B)$:

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

We also know that:

$$P(A \cup B) = 1 - P(A^C \cap B^C)$$

so:

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B) = 1 - P(A^c) \cdot P(B^c).$$

1.25   50

A card player is dealt a 13-card hand from a well-shuffled, standard deck of cards. What is the probability that the hand is void in at least one suit ("void in a suit" means having no cards of that suit)?

We will do it using the union:

$$P(\clubsuit \cup \diamondsuit \cup \spadesuit \cup \heartsuit) =$$

$$4 \cdot \frac{\binom{52-13}{13}}{\binom{52}{13}} - 6 \cdot \frac{\binom{52-26}{13}}{\binom{52}{13}} + 4 \cdot \frac{\binom{52-39}{13}}{\binom{52}{13}} - 0$$

Explanation: First 4 fractions are the lack of one suit, second 6 fractions are the lack of 2 suits, third 4 fractions are the lack of 3 suits and the 0 is the lack of all 4 suits.

1.26   53

Fred needs to choose a password for a certain website. Assume that he will choose an 8-character password, and that the legal characters are the lowercase letters a, b, c, ..., z, the uppercase letters A, B, C, ..., Z, and the numbers 0, 1, ..., 9.

(a) How many possibilities are there if he is required to have at least one lowercase letter in his password?

$P(\textbf{At least one letter}) =$

Incorrect version:

$$\binom{26}{1} \cdot \frac{\binom{62}{7}}{\binom{62}{8}} - \binom{26}{2} \cdot \frac{\binom{62}{6}}{\binom{62}{8}} + \binom{26}{3} \cdot \frac{\binom{62}{5}}{\binom{62}{8}} - \binom{26}{4} \cdot \frac{\binom{62}{4}}{\binom{62}{8}} + \binom{26}{5} \cdot \frac{\binom{62}{3}}{\binom{62}{8}} - \binom{26}{6} \cdot \frac{\binom{62}{2}}{\binom{62}{8}} + \binom{26}{7} \cdot \frac{\binom{62}{1}}{\binom{62}{8}} + \binom{26}{8}$$

Correct version:

$$\sum_{i=1}^{8} (-1)^{i+1} \cdot 26^i \cdot 62^{8-i}$$

(b) How many possibilities are there if he is required to have at least one lowercase letter and at least one uppercase letter in his password? **We will get the number of possibilities of this event using the contrary 1(in this case 62^8) - the possibilities of the password not having any letter.**

$$62^8 - 2 \cdot 36^8 + 10^8$$

Explanation: The sum of the lowercase plus the digits is 36 same with uppercase and digits in both this cases we might have the case in which everything in digits so as we are subtracting it two time we need to add it once (that's where 10^8 comes from).

(c) How many possibilities are there if he is required to have at least one lowercase letter, at least one uppercase letter, and at least one number in his password?

In the same way we did the last one using inclusion exclusion and the 1 - trick.

$$62^8 - (38^8 - 36^8 - 52^8) + (10^8 + 26^8 + 26^8) - 0^8$$

Explanation: We are subtracting first ud ul and ld and then we are adding d, u and l and then subtracting udl

2   55

A club consists of 10 seniors, 12 juniors, and 15 sophomores. An organizing committee of size 5 is chosen randomly (with all subsets of size 5 equally likely).

(a) Find the probability that there are exactly 3 sophomores in the committee.

$$\frac{\binom{15}{3} \cdot \binom{12}{2}}{\binom{37}{5}} + \frac{\binom{15}{3} \cdot \binom{10}{2}}{\binom{37}{5}} + \frac{\binom{15}{3} \cdot \binom{12}{1} \cdot \binom{10}{1}}{\binom{37}{5}}$$

Explanation: We are going to calculate the probability of having 3 sophomores and making the committe of the other 2 groups. There are 3 ways of doing this so we calculate the probability for each one.

(b) Find the probability that the committee has at least one representative from each of the senior, junior, and sophomore classes.

$$1 - (\frac{\binom{37-15}{5}}{\binom{37}{5}} + \frac{\binom{37-12}{5}}{\binom{37}{5}} + \frac{\binom{37-10}{5}}{\binom{37}{5}}) + (\frac{\binom{10}{5}}{\binom{37}{5}} + \frac{\binom{12}{5}}{\binom{37}{5}} + \frac{\binom{15}{5}}{\binom{37}{5}}) - 0$$

Explanation: We do inclusion exclusion with the 1- (the probability of not having at least one from one of the groups). We accomplish this by having in mind that our single tones exclude entirely one group.

2.1 58

A widget inspector inspects 12 widgets and finds that exactly 3 are defective. Unfortunately, the widgets then get all mixed up and the inspector has to find the 3 defective widgets again by testing widgets one by one.

(a) Find the probability that the inspector will now have to test at least 9 widgets.

$$1 - \frac{\binom{8}{3} \cdot 3! \cdot 9!}{12!}$$

Explanation: We will find the contrary. So there are 12! ways of arranging the widgets and you can choose 8 chose 3 locations for the defective widgets. But we are under-counting since we need to count for the different ways in which we can organize them so 3! for the defective and 9! for the rest

(b) Find the probability that the inspector will now have to test at least 10 widgets.

$$1 - \frac{\binom{9}{3} \cdot 3! \cdot 9!}{12!}$$

Explanation: Same as above

3   2.11 Conditioning on evidence

3.1 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?

We will use the following notation S for spam and F for it contains "free money"

$$\frac{P(F|S)P(S)}{P(F|S)P(S) + P(F|S^c)P(S^c)}$$

$$= \frac{0.1 \cdot 0.8}{0.1 \cdot 0.8 + 0.01 \cdot 0.2} = 0.975$$

3.2 2

A woman is pregnant with twin boys. Twins may be either identical or fraternal. Suppose that 1/3 of twins born are identical, that identical twins have a 50% chance of being both boys and a 50% chance of being both girls, and that for fraternal twins each twin independently has a 50% chance of being a boy and a 50% chance of being a girl. Given the above information, what is the probability that the woman's twins are identical?

We will use the following notation B for both boys and I for Identical twins

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

$$= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 1/4 \cdot 2/3} = \frac{1/6}{1/6 + 1/6} = 0.5$$

3.3 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?

We will use the following notation C for cancer S for smokes.

$$P(S|C) = \frac{P(C|S)P(S)}{P(C|S)P(S) + P(C|S^c)P(S^c)}$$

$$= \frac{0.216 \cdot 23 \cdot P(C|S^c)}{0.216 \cdot 23 \cdot P(C|S^c) + P(C|S^c) \cdot 0,784} = \frac{4,968 \cdot P(C|S^c)}{5.752 \cdot P(C|S^c)}$$

$$= 0,8633699$$

3.4 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).

$$P(K|R) = \frac{P(R|K)P(K)}{P(R|K)P(K) + P(R|K^c)P(K^c)}$$

$$= \frac{P(K)}{P(K) + 1/n \cdot P(K^c)}$$

$$= \frac{p}{p + 1/n \cdot (1-p)}$$

$$= \frac{p}{p(1 - 1/n) + 1/n}$$

(b) Show that $P(K|R) \geq p$, and explain why this makes sense intuitively. When (if ever) does $P(K|R)$ equal p?
   Since we know that $P(R) \leq 1$.

$$p(1 - 1/n) + 1/n \leq 1$$

$$= p \cdot (1 - 1/n) + 1/n \leq p$$

$$p \leq \frac{p}{p(1 - 1/n) + 1/n}$$

$$p \leq P(K|R)$$

Intuitively since we already know he got it right and we know that he will always select the correct option when he knows it, we are more certain that he knows the answer. $P(K|R)$ equal p when $P(K) = p = 1$ so when he knows the answer.

3.5 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.

3.6 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.)

We will use the following notation S for seven heads D for double headed coin.

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

$$= \frac{0,01}{0,01 + 1/128 \cdot 0.99} = \frac{0,01}{0,017734375} = 0,563876652$$

3.7 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?

We will fist calculate $P(S|D)$. Using F as the event of the coin being fair

$$P(S|D) = P(S|D \cap F)P(F|D) + P(S|D \cap F^c)P(F^c|D)$$

$$= 1/128 \cdot 99/100 + 1 \cdot 1/100 = 0,017734375$$

Now for the actual problem

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

$$= \frac{0,017734375 \cdot 0,5}{0,5 \cdot 0,017734375 + 1/128 \cdot 0.5} = \frac{0,0088671875}{0,0127734375} = 0,6941896024$$

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

We will define C as the event of choosing a double-headed coin. We will first calculate $P(C)$.

$$P(C) = P(C|D)P(D) + P(C|D^c)P(D^c) = 1/100 \cdot 1/2 + 0 \cdot 1/2 = 1/200$$

$$P(C|S) = \frac{P(S|C)P(C)}{P(S|C)P(C) + P(S|C^c)P(C^c)}$$

$$= \frac{1 \cdot 1/200}{1 \cdot 1/200 + 1/128 \cdot 1/198} = \frac{0,005}{0,0127734375} = 0,3914373089$$

3.8   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.

Proof. Since A_1 implies B and A_2 also implies B we know that $A_1 \subseteq B$ and $A_2 \subseteq B$ so

$$P(A_1 \cap B) = P(A_1)$$

$$P(A_2 \cap B) = P(A_2)$$

therefore

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

$\square$

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

If you want the conditional probability $P(A_1|B)$ and $P(A_2|B)$ and you know that A_1 implies B and A_2 implies B and $P(A_1) = P(A_2)$ so since in both cases we are looking for a probability of an event A happening given an event B, we are looking for the success scenarios over the total scenarios or

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

In our case the success scenarios are actually just the P of the As which are equal in both events, since both As are contained inside B their intersection is just itself. And since they are equal then it is done.

For the concrete example suppose that B is the event that it rains and A_1 is the event that it rains in the morning and A_2 is the event that it rains in the afternoon with $P(A_1) = P(A_2)$. Then $P(A_1|B) = P(A_2|B)$ as explained above.

3.9 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.

We first need to calculate what their claims affect.

$$P(+|D) = 0, P(+|D^c) = 0, P(-|D) = 1, P(-|D^c) = 1$$

Now we will calculate overall success rate C of company A and B assuming that company A is still with specificity and sensitivity as given.

Formula:

$$P(C) = P(C|D)P(D) + P(C|D^c)P(D^c) = P(+|D)P(D) + P(-|D^c)P(D^c) = Sensitivity \cdot P(D) + Specifitiy \cdot P(D^c)$$

Company A:
$$P(C) = 0,95 \cdot 0,01 + 0,95 \cdot 0,99 = 0,95$$

Company B (using the information we got):

$$P(C) = 0 \cdot 0,01 + 1 \cdot 0,99 = 0,99$$

As you may have noticed although Company B just labels every one as negative no matter if they actually have the disease or not, they have an overall higher success rate than A. Why? I believe because the disease is very rare so 99 percent of the time you are going to be right.

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

Because Company A's provides a test that will actually label (as best as it can) correctly those people that actually have the disease, while it might scare sometimes people that do not have the disease with false positive this test is the only one that will actually label true positives.

(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?

Just to remind ourselves here is the data for company A and company B:

Company A:
$$P(C) = 0,95 \cdot 0,01 + 0,95 \cdot 0,99 = 0,95$$

Company B (using the information we got):

$$P(C) = 0 \cdot 0,01 + 1 \cdot 0,99 = 0,99$$

Company A(With sensitivity equal to 1):

$$P(C) = 1 \cdot 0,01 + 0,95 \cdot 0,99 = 0,9505$$

For Company A to have a higher overalls success rate than company B specificity should be:

$$0,01 + x \cdot 0,95 > 0,99 = x \cdot 0,99 > 0,98$$

$$= x > 0,98/0,99$$

Their overall success rate would be bigger if their specificity is bigger than $\frac{0,98}{0,99}$
Company A(With specificity equal to 1, and sensitivity as it started):

$$P(C) = 0,95 \cdot 0,01 + 1 \cdot 0,99 = 0,9995$$

For Company A to have a higher overalls success rate than company B specificity should be:

$$x \cdot 0,01 + 1 \cdot 0,99 > 0,99 = x \cdot 0,01 > 00$$

$$= x > 0$$

Their overall success rate would be bigger if their specificity is bigger than 0

3.10   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?

That A occurred since:

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

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

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

We have the lowest denominator if A occurred therefore the highest probability.

3.11   16

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

$$P(A) \ge P(A|B)$$
$$= P(A|B)P(B) + P(A|B^c)P(B^c) \ge P(A|B)$$
$$= P(A|B^c)P(B^c) \ge P(A|B) - P(A|B)P(B)$$
$$= P(A|B^c)P(B^c) \ge P(A|B)P(B^c)$$
$$= P(A|B^c) \ge P(A|B)$$
$$= P(B)P(A|B^c) \ge P(A|B)P(B)$$
$$= P(A|B^c) \ge P(A|B)P(B) + P(A|B^c) - P(B)P(A|B^c)$$
$$= P(A|B^c) \ge P(A|B)P(B) + P(A|B^c)(1 - P(B))$$
$$= P(A|B^c) \ge P(A|B)P(B) + P(A|B^c)P(B^c)$$
$$= P(A|B^c) \ge P(A)$$

Intuitively, we are given that if you know B then you have a lower chance of knowing A then if you know the contrary of B you would get a higher chance than A.

3.12 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 prob- ability, 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. **Since**

$$P(B|A) = 1$$

We can get:

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

Then we can use this in:

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

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

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

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

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

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

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

$$1 = P(A^c|B^c)$$

(b) Show however that the result in (a) does not hold in general if = is replaced by ≈. 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?

   Say B is the event of you eating tomorrow and A is the event of the sun being in the sky the next year, just in case this does not seem independent enough, assume independence. Then $P(B|A) \approx 1$ but $P(A^c|B^c) \approx 0$. Why? Because they are independent and because of definition 2.5.1 where if A and B are independent $P(A|B) = P(A), P(B|A) = P(B)$.

3.13   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 $P(B \cap A^c) = 0$.

We will use the following information:

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

$$P(B|A^c) = \frac{P(B \cap A^c)}{1 - P(A)}$$

Since P(A)=1

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

$$P(B \cap A^c) = P(B|A^c)(1 - P(A)) = 0$$

Now for the actual operation:

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

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

Using the information we got from above:

$$P(A|B) = \frac{P(B \cap A)}{P(B \cap A) + 0} = 1$$

3.14   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.

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)}$$

Using the following information:

$$P(-|D^c) = \frac{P(D^c|-)P(-)}{P(D^c)} = \frac{P(D^c|-)(1 - P(+))}{P(D^c)}$$

$$1 - P(+) = \frac{P(D^c)P(-|D^c)}{P(D^c|-)}$$

$$P(+) = 1 - \frac{P(D^c)P(-|D^c)}{P(D^c|-)}$$

Therefore:

$$P(D|+) = \frac{P(+|D)P(D)}{1 - \frac{P(D^c)P(-|D^c)}{P(D^c|-)}}$$

(b) Not surprisingly, Fred is much more interested in $P(havedisease|testpositive)$, known as the positive predictive value, than in the sensitivity $P(testpositive|havedisease)$. 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 sen-sitivity 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.

Let's think about this like the exercise of company A and company B. If you have 10000 people and you know 1% is affected by the disease then there are 100 people that actually have the disease. Now say you have a test that has a sensitivity of 1 (therefore we can at least diagnose the 100 cases) now say you have a specificity of 0.95 then you will have 495 people that do not have the disease but got diagnosed as if they had it. If instead we could have a specificity of 1 then we would not have False positives, which are the largest proportion when a disease is rare.

3.15   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. Under the assumptions of Example 2.2.5, show that the probability that both children are girls given that at least one is a girl with characteristic C is $\frac{2-p}{4-p}$. Note that this 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).

We need to Calculate: $P($both girls$|$at least one girl$)$

$$= \frac{P(2girls \cap atLeastOneC)}{P(atLeastOneC)}$$

In the denominator we need to calculate the equivalent $1 - P(nogirlswithC)$ which is:

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

The numerator is since C is independent results in: Numerator=P(both girls)P(at least one Child with C) this would be equal to (1/4)(1-p(both Children don't have C))

$$(\frac{1}{4}) \cdot (1 - (1 - p)^2) = \frac{2p - p^2}{4}$$

Therefore: $P($both girls$|$at least one girl$)$ =

$$\frac{\frac{2p-p^2}{4}}{\frac{4p-p^2}{4}} = \frac{2p - p^2}{4p - p^2} = \frac{2 - p}{4 - p}$$

3.16 31

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

Yes, suppose the event is independent from itself then:

$$P(E) = P(E \cap E) = P(E)P(E) = (P(E))^2$$

Therefore this can only happen if $P(E) = 0$ or 1.

3.17 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 $P(D > A)$.

$$P(A > B) = 4/6$$

$$P(B > C) = 0$$

$$P(C > D) = 12/36 + 12/36 = 24/36$$

$$P(D > A) = 12/36 + 6/36 = 18/36$$

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

Both of them are not independent since if we know the first event it will give us some information about how big the numbers are, increasing or decreasing the overall probabiility of the second event.

3.18   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?

$$P(F) = P(F|M)P(M) + P(F|I)P(I) + P(F|B)P(B) = (.3)(1/3) + (.5)(1/3) + (.9)(1/3) = 0.566667 = \frac{17}{30}$$

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

$$P(S|F) = \frac{P(F \cap S)}{P(F)}$$

$$P(F \cap S) = P(S \cap F|B)P(B) + P(S \cap B|I)P(I) + P(S \cap F|M)P(M)$$

Since given the skill of your opponent F and S are independent:

$$= P(S|B)P(F|B)P(B) + P(S|I)P(F|I)P(I) + P(S|M)P(F|M)P(M)$$

$$= (0.9^2 + 0.5^2 + 0.3^2)(0.3) = \frac{23}{60}$$

Therefore:

$$P(S|F) = \frac{\frac{23}{60}}{\frac{17}{30}} = \frac{23}{34} = 0.676470588$$

Other solution

$$P(S) = P(S|B)P(B) + P(S|I)P(I) + P(S|M)P(M)$$
$$P(S) = (S|B)P(B|F) + P(S|I)P(I|F) + P(S|M)P(M|F)$$

Intermediate step:

$$P(B|F) = \frac{P(F|B)P(B)}{P(F)} = \frac{9/10 * 1/3}{17/30} = \frac{9}{17}$$

$$P(I|F) = \frac{P(F|I)P(I)}{P(F)} = \frac{5/10 * 1/3}{17/30} = \frac{5}{17}$$

$$P(M|F) = \frac{P(F|M)P(M)}{P(F)} = \frac{3/10 * 1/3}{17/30} = \frac{3}{17}$$

Continue:

$$P(S) = \frac{9}{10} \cdot \frac{9}{17} + \frac{5}{10} \cdot \frac{5}{17} + \frac{3}{10} \cdot \frac{3}{17} = \frac{23}{34} = 0.676470588$$

Since given the skill of your opponent F and S are independent.

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

If the outcome of the events were to be independent then winning the first round wouldn't tell us any relevant data as to if we are going to win or lose the second round (which is not the case in part b). Having conditional independence means that given that our adversary is the same the outcome of the first and second game is independent, so we still have the same chance to win against him during the individual isolated matches, but since we have the information that we won the first round we have an easier way to calculate if we will win the second.

3.19   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.

It would make sense because the university would not accept anyone that is bad at both. And since being good at math (say M) and being good at Baseball (say B) are independent then being both good at math and baseball is quite rare. If you are good at one it would be difficult to have the other in this context.

(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.

Proof. **We know A and B are independent therefore:**

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

And it is important to note that:

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

and:

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

Then:

$$P(A \cup B) < 1$$
$$= P(A)P(A \cup B) < P(A)$$
$$= P(A) < \frac{P(A)}{P(A \cup B)}$$
$$P(A|B \cap C) = P(A) < \frac{P(A)}{P(C)} = P(A|C)$$

$\square$

3.20   37

Two different diseases cause a certain weird symptom; anyone who has either or both of these diseases will experience the symptom. Let D_1 be the event of having the first disease, D_2 be the event of having the second disease, and W be the event of having the weird symptom. Suppose that D_1 and D_2 are independent with $P(D_j) = p_j$, and that a person with neither of these diseases will have the weird symptom with probability w_0. Let $q_j = 1 - p_j$,and assume that $0 < p_j < 1$.

(a) Find $P(W)$.

$$P(W) = P(W|D_1)P(D_1) + P(W|D_2)P(D_2) + P(W|D_1^C)P(D_1^C) + P(W|D_2^C)P(D_2^C)$$
$$= P(W|D_1)p_1 + P(W|D_2)p_2 + P(W|D_1^C)q_1 + P(W|D_2^C)q_2$$

or:

$$P(W) = P(W \cap D_1) + P(W \cap D_2) + P(W \cap D_1^C) + P(W \cap D_2^C)$$

(b) Find $P(D1|W)$, $P(D2|W)$, and $P(D1, D2|W)$.

(c) Determine algebraically whether or not D_1 and D_2 are conditionally independent given W.

(d) Suppose for this part only that $w_0 = 0$. Give a clear, convincing intuitive explanation in words of whether D_1 and D_2 are conditionally independent given W.