Introduction to Probability

Solution Manual

October 25, 2025

Contents

Chapter 1

Probability and Counting

1.1 Counting

1.1.1 problem 1

There are

$$\binom{11}{4}$$

ways to select 4 positions for I,

$$\binom{7}{4}$$

ways to select 4 postions for S,

$$\binom{3}{2}$$

ways to selection 2 positions for P leaving us with a single choice of position for M. In total, we get

$$\binom{11}{4}\binom{7}{4}\binom{3}{2}\binom{1}{1}$$

permutations.

1.1.2 problem 2

(a) If the first digit can't be 0 or 1, we have eight choices for the first digit. The remaining six digits can be anything from 0 to 9. Hence, the solution is

$$8 \times 10^6$$

(b) We can subtract the number of phone numbers that start with 911 from the total number of phone numbers we found in the previous part.

If a phone number starts with 911, it has ten choices for each of the remaining four digits.

$$8 \times 10^6 - 10^4$$

1.1.3 problem 3

(a) Fred has 10 choices for Monday, 9 choices for Tuesday, 8 choices for Wednesday, 7 choices for Thursday and 6 choices for Friday.

$$10 \times 9 \times 8 \times 7 \times 6$$

(b) For the first restaurant, Fred has 10 choices. For all subsequent days, Fred has 9 choices, since the only restriction is that he doesn't want to eat at the restaurant he ate at the previous day.

$$10 \times 9^4$$

1.1.4 problem 4

(a) There are $\binom{n}{2}$ matches.

For a given match, there are two outcomes. Each match has two possible outcomes. We can use the multiplication rule to count the total possible outcomes.

$$2^{\binom{n}{2}}$$

(b) Since every player plays every other player exactly once, the number of games is the number of ways to pair up n people.

$$\binom{n}{2}$$

1.1.5 problem 5

(a) By the end of each round, half of the players participating in the round are eliminated. So, the problem reduces to finding out how many times the number of players can be halved before a single player is left.

The number of times 2^N can be divided by two is $\log_2 2^N$ which means the total amount of rounds in the tournament is

$$N$$

(b) The number of games in a given round is $\frac{N_r}{2}$. We can sum up these values for all the rounds.

$$\begin{aligned} f(N) &= \frac{N}{2} + \frac{N}{4} + \frac{N}{8} + \cdots + \frac{N}{2^{\log_2 N}} \\ &= N \sum_{i=0}^{\log_2 N} \frac{1}{2^i} \\ &= N \times \frac{N-1}{N} \\ &= N - 1 \end{aligned} \tag{1.1}$$

(c) Tournament is over when a single player is left. Hece, $N-1$ players need to be eliminated. As a result of a match, exactly one player is eliminated. Hence, the number of matches needed to eliminate $N-1$ people is

$$N-1$$

1.1.6 problem 6

Line up the 20 players in some order then say the first two are a pair, the next two are a pair, etc. This overcounts by a factor of 10! because we don't care about the order of the games. So in total we have

$$\frac{20!}{10!}$$

ways for them to play. This correctly counts for the whether player A plays white or black. If we didn't care we would need to divide by 2^{10}.

Another way to look at it is to choose the 10 players who will play white then let each of them choose their opponent from the other 10 players. This gives a total of

$$\binom{20}{10} \times 10!$$

possibilities of how they are matched up. We don't care about the order of the players who play white but once we've chosen them the order of the players who play black matters since different orders mean different pairings.

1.1.7 problem 7

(a) There are $\binom{7}{3}$ ways to assign three wins to player A. For a specific combination of three games won by A, there are $\binom{4}{2}$ ways to assign two draws to A. There is only one way to assign two losses to A from the remaining two games, namely, A losses both games.

$$\binom{7}{3} \times \binom{4}{2} \times \binom{2}{2}$$

(b) If A were to draw every game, there would need to be at least 8 games for A to obtain 4 points, so A has to win at least 1 game. Similarly, if A wins more than 4 games, they will have more than 4 points.

Case 1: A wins 1 game and draws 6.

This case amounts to selecting 1 out of 7 for A to win and assigning a draw for the other 6 games. Hence, there are 7 possibilities.

Case 2: A wins 2 games and draws 4.

There are $\binom{7}{2}$ ways to assign 2 wins to A. For each of them, there are $\binom{5}{4}$ ways to assign four draws to A out of the remaining 5 games. Player B wins the remaining game. The total number of possibilities for this case is $\binom{7}{2} \times \binom{5}{4}$.

Case 3: A wins 3 games and draws 2.

There are $\binom{7}{3}$ ways to assign 3 wins to A. For each of them, there are $\binom{4}{2}$ ways to assign two draws to A out of the remaining 4 games. B wins the remaining 2 games. The total number of possibilities for this case is $\binom{7}{3} \times \binom{4}{2}$.

Case 4: A wins 4 games and loses 3.

There are $\binom{7}{4}$ ways to assign 4 wins to A. B wins the remaining 3 games. The total number of possibilities for this case is $\binom{7}{4}$.

Summing up the number of possibilities in each of the cases we get

$$\binom{7}{1} + \binom{7}{2} \times \binom{5}{4} + \binom{7}{3} \times \binom{4}{2} + \binom{7}{4}$$

(c) If B were to win the last game, that would mean that A had already obtained 4 points prior to the last game, so the last game would not be played at all. Hence, B could not have won the last game. The last game must have ended in either A winning (case 1) or a draw (case 2).

Case 1: A wins the last game. This means A had 3 points after 6 games.

There are four possibilities for A to earn 3 points in 6 games:

1.1. 6 draws
1.2. 3 wins and 3 losses
1.3. 2 wins, 2 draws, and 2 losses
1.4. 1 win, 4 draws, and 1 loss.

Let's calculate the number of possibilities for each of these subcases.

1.1. There is only one way to assign 6 draws to 6 games: The number of possibilities is 1.
1.2. There are $\binom{6}{3}$ ways to assign 3 wins to A out of the first 6 games. The remaining 3 games are losses for A. The number of possibilities is $\binom{6}{3}$.
1.3. There are $\binom{6}{2}$ ways to assign 2 wins to A out of the first 6 games. There are $\binom{4}{2}$ ways to assign 2 draws out of the remaining 4 games. The remaining 2 games are losses for A. The number of possibilities is $\binom{6}{2} \times \binom{4}{2}$.
1.4. There are $\binom{6}{1}$ ways to assign 1 win to A out of the first 6 games. There are $\binom{5}{4}$ ways to assign 4 draws out of the remaining 5 games. The remaining game is a loss for A. The number of possibilities is $\binom{6}{1} \times \binom{5}{4}$.

Case 2: The last game ends in a draw. This means A had 3.5 points after 6 games.

There are three possibilities for A to earn 3.5 points in 6 games:

2.1. 3 wins, 1 draw, and 2 losses
2.2. 2 wins, 3 draws, and 1 loss
2.3. 1 win, 5 draws.

Let's calculate the number of possibilities for each of these subcases.

2.1. There are $\binom{6}{3}$ ways to assign 3 wins to A out of the first 6 games. There are $\binom{3}{1}$ ways to assign 1 draw out of the remaining 3 games. The remaining 2 games are losses for A. The number of possibilities is $\binom{6}{3} \times \binom{3}{1}$.

2.2. There are $\binom{6}{2}$ ways to assign 2 wins to A out of the first 6 games. There are $\binom{4}{3}$ ways to assign 3 draws out of the remaining 4 games. The remaining game is a loss for A. The number of possibilities is $\binom{6}{2} \times \binom{4}{3}$.

2.3. There are $\binom{6}{1}$ ways to assign 1 win to A out of the first 6 games. The remaining 5 games are losses for A. The number of possibilities is $\binom{6}{1}$.

The total number of possibilities then is:

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

1.1.8 problem 10

(a) Case 1: Student takes exactly one statistics course.

There are 5 choices for the statistics course. There are $\binom{15}{6}$ choices of selecting 6 non-statistics courses.

Case 2: Student takes exactly two statistics courses.

There are $\binom{5}{2}$ choices for the two statistics course. There are $\binom{15}{5}$ choices of selecting 5 non-statistics courses.

Case 3: Student takes exactly three statistics courses.

There are $\binom{5}{3}$ choices for the three statistics course. There are $\binom{15}{4}$ choices of selecting 4 non-statistics courses.

Case 4: Student takes exactly four statistics courses.

There are $\binom{5}{4}$ choices for the four statistics course. There are $\binom{15}{3}$ choices of selecting 3 non-statistics courses.

Case 5: Student takes all the statistics courses.

There are $\binom{15}{2}$ choices of selecting 2 non-statistics courses.

So the total number of choices is

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

An Alternative Approach

There are $\binom{20}{7}$ choices of selecting 7 courses which is the maximum number of choices if there were no restriction as choosing at least one statistics course.

$\binom{15}{7}$ is the number of choices without any statistics course.

So the total number of choices with at least one statistics course is

$$\binom{20}{7} - \binom{15}{7}$$

(b) It is true that there are $\binom{5}{1}$ ways to select a statistics course, and $\binom{19}{6}$ ways to select 6 more courses from the remaining 19 courses, but this procedure results in overcounting.

For example, consider the following two choices.

(a) STAT110, STAT134, History 124, English 101, Calculus 102, Physics 101, Art 121

(b) STAT134, STAT110, History 124, English 101, Calculus 102, Physics 101, Art 121

Notice that both are selections of the same 7 courses.

1.1.9 problem 11

(a) Each of the n inputs has m choices for an output, resulting in

$$m^n$$

possible functions.

(b) If $n \geq m$, at least two inputs will be mapped to the same output, so no one-to-one function is possible.

If $n < m$, the first input has m choices, the second input has $m - 1$ choices, and so on. The total number of one-to-one functions then is

$$m(m-1)(m-2)\dots(m-n+1) = \frac{m!}{(m-n)!}$$

1.1.10 problem 12

(a)

$$\binom{52}{13}$$

(b) The number of ways to break 52 cards into 4 groups of size 13 is

$$\frac{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}}{4!}$$

.

The reason for dividision by 4! is that all permutations of specific 4 groups describe the same way to group 52 cards.

Since we do care about the order of the 4 groups, we should not divide by 4!. The final answer then is

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}$$

(c) The key is to notice that the sampling is done without replacement. $\binom{52}{13}^4$ assumes that all four players have $\binom{52}{13}$ choices of hands available to them. This would be true if sampling was done with replacement.

1.1.11 problem 13

The problem amounts to sampling with replacement where order does not matter, since having 10 copies of each card amounts to replacing the card. This is done using the Bose-Einstein method.

Thus, the answer is

$$\binom{52+10-1}{10} = \binom{61}{10}$$

1.1.12 problem 14

There are 4 choices for sizes and 8 choices for toppings, of which any combination (including no toppings) can be selected.

The total number of possible choices of toppings is $\sum_{i=0}^{8}\binom{8}{i} = 2^8 = 256$. Thus, the total number of possible size-topping combinations is $4 * 256 = 1024$.

We wish to sample two pizzas, with replacement, out of the 1024 possibilities. By Einstein-Bose, there are a total of $\binom{1025}{2}$ choices.

A common mistake is to use multiplication rule to get $(2^8) * (2^8)$ as total possible combinations for two pizzas, and try to adjust for overcounting by dividing the result with 2 (as order between pizzas doesn't matter). This fails because the possibilities with identical pizzas are counted only once.

1.2 Story Proofs

1.2.1 problem 17

$\binom{2n}{n}$ counts the number of ways to sample n objects from a set of $2n$. Instead of sampling from the whole set, we can break the set into two sets of size n each. Then, we have to sample n objects in total from both sets.

We can sample all n objects from the first set, or 1 object from the first set and $n-1$ objects from the second set, or 2 objects from the first set and $n-2$ objects from the second set and so on.

There are $\binom{n}{n}$ ways to sample all n objects from the first set, $\binom{n}{1}\binom{n}{n-1}$ ways to sample 1 object from the first set and $n-1$ objects from the second set, $\binom{n}{2}\binom{n}{n-2}$ ways to sample 2 objects from the first set and $n-2$ objects from the second set. The pattern is clear

$$\sum_{k=0}^{n}\binom{n}{k}\binom{n}{n-k} = \sum_{k=0}^{n}\binom{n}{k}^2$$

1.2.2 problem 18

Consider the right hand side of the equation. Since a committe chair can only be selected from the first group, there are n ways to choose them. Then, for each choice of a committee chair, there are $\binom{2n-1}{n-1}$ ways to choose the remaining members. Hence, the total number of committees is $n\binom{2n-1}{n-1}$.

Now consider the left side of the equation. Suppose we pick k people from the first group and $n-k$ people from the second group, then there are k ways to assign a chair from the members of the first group we have picked.

k can range from 1 to n giving us a total of $\sum_{k=1}^{n} k\binom{n}{k}\binom{n}{n-k} = \sum_{k=1}^{n} k\binom{n}{k}^2$ possible committees.

Since, both sides of the equation count the same thing, they are equal.

1.2.3 problem 21

(a) Case 1: If Tony is in a group by himself, then we have to break the remaining n people into $k-1$ groups. This can be done in

$$\left\{ \begin{matrix} n \\ k-1 \end{matrix} \right\}$$

ways.

Case 2: If Tony is not in a group by himself, then we first break up the remaining n people into k groups. Then, Tony can join any of them. The number of possible groups then is

$$k \left\{ \begin{matrix} n \\ k \end{matrix} \right\}$$

Case 1 and 2 together count the number of ways to break up $n+1$ people into k non empty groups, which is precisely what the left side of the equation counts.

(b) Say Tony wants to have m in his group. That is to say he does not want $n-m$ people. These $n-m$ people must then be broken into k groups.

The number of people Tony wants to join his group can range from 0 to $n-k$. The reason for the upper bound is that at least k people are required to make up the remaining k groups.

Taking the sum over the number of people in Tony's group we get

$$\sum_{j=0}^{n-k} \binom{n}{j} \left\{ \begin{matrix} n-j \\ k \end{matrix} \right\}$$

Now, instead of taking the sum over the number of people Tony wants in his group, we can equivalently take the sum over the number of people Tony does not want in his group. Hence,

$$\sum_{j=0}^{n-k} \binom{n}{j} \left\{ \begin{matrix} n-j \\ k \end{matrix} \right\} = \sum_{i=n}^{k} \binom{n}{i} \left\{ \begin{matrix} i \\ k \end{matrix} \right\}$$

Since the sum counts all possible ways to group $n+1$ people into $k+1$ groups, we have

$$\sum_{i=n}^{k} \binom{n}{i} \left\{ \begin{matrix} i \\ k \end{matrix} \right\} = \left\{ \begin{matrix} n+1 \\ k+1 \end{matrix} \right\}$$

as desired.

1.2.4 problem 22

(a) Let us count the number of games in a round-robin tournament with $n+1$ participants in two ways.

Method 1: Since every player plays against all other players exactly once, the problem reduces to finding the number of ways to pair up $n+1$ people. There are $\binom{n+1}{2}$ ways to do so.

Method 2: The first player participates in n games. The second one also participates in n games, but we have already counted the game against the first player, so we only care about $n-1$ games. The third player also participates in n games, but we have already counted the games against the first and second players, so we only care about $n-2$ games.

In general, player i will participate in $n+1-i$ games that we care about. Taking the sum over i we get

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

Since both methods count the same thing, they are equal.

An Alternative Approach The RHS expression counts number of ways to pair two people from a group of $n+1$ people of different ages. Let's say the eldest person in the subgroup of two is also the eldest person in the whole group. Then, we would have n people to choose from as the second person of the sub group. If subgroup's eldest is the second-eldest in the whole group, we'd have $n-1$ people to choose from, and so on all the way to 1. By adding these cases, we get the

LHS expression which covers all possibilities of pairing two people from group of $n+1$, and hence is equivalent to RHS.

(b) LHS: If n is chosen first, then the subsequent 3 numbers can be any of $0, 1, \ldots, n-1$. These 3 numbers are chosen with replacement resulting in n^3 possibilities. Summing over possible values of n we get $1^3 + 2^3 + \cdots + n^3$ total number of possibilities.

RHS: We can count the number of permutations of the 3 numbers chosen with replacement from a different perspective. The 3 numbers can either all be distinct, or all be the same, or differ in exactly 1 value.

Case 1: All 3 numbers are distinct.

Selecting 4 (don't forget the very first, largest selected number) distinct numbers can be done in $\binom{n+1}{4}$ ways. The 3 smaller numbers are free to permute amongst themselves. This gives us a total of $6\binom{n+1}{4}$ possibilities.

Case 2: All 3 numbers are the same.

In this case, we have to select 2 digits. The smaller digit will be sampled 3 times and there are no ways to permute identical numbers, so the number of possiblities is $\binom{n+1}{2}$.

Case 3: Two of the 3 numbers are distinct.

In this case, we have to select 3 digits in total. One of the smaller 2 digits will be sampled twice, giving us 3 permutations. Since, there are 2 choices for which digit gets sampled twice, we get a total of 6 permutations. The total number of possibilities then is $6\binom{n+1}{3}$.

Adding up the number of possibilities in each of the cases we get a total of

$$6\binom{n+1}{4} + 6\binom{n+1}{3} + \binom{n+1}{2}$$

possibilities.

Since the LHS and the RHS count the same set, they are equal.

1.3 Naive Definition Of Probability

1.3.1 problem 23

We are interested in the case of 3 consecutive floors. There are 7 equally likely possibilities

$$(2,3,4),(3,4,5),(4,5,6),(5,6,7),(6,7,8),(7,8,9),(8,9,10).$$

For each of this possibilities, there are 3 ways for 1 person to choose button, 2 for second and 1 for third (3! in total by multiplication rule).

So number of favorable combinations is

$$7 * 3!$$

Generally each person have 9 floors to choose from so for 3 people there are 9^3 combinations by multiplication rule.

Hence, the probability that the buttons for 3 consecutive floors are pressed is

$$\frac{7 * 3!}{9^3}$$

1.3.2 problem 26

(a) The problem is isomorphic (having same structure) to the birthday problem. When sampling with replacement, each person corresponds to a date in the birthday problem, and the size of sample corresponds to the number of people in birthday problem. Hence, taking a random sample of 1000 from a population of a million corresponds to asking a thousand people their birth date where there are a total of a million dates. Number of ways to take such a sample is K^{1000} where K is size of population. Similarly, number of ways to take sample without replacement corresponds to number of ways of having no birthday match in that situation: $K(K-1)\dots(K-1000+1)$

(b)

$$P(A) = 1 - P(A^c) = 1 - \frac{K(K-1)\dots(K-1000+1)}{K^{1000}}$$

where $K = 1000000$.

1.3.3 problem 27

For each of the k names, we sample a memory location from 1 to n with equal probability, with replacement. This is exactly the setup of the birthday problem. Hence, the probability that at least one memory location has more than 1 value is

$$P(A) = 1 - P(A^c) = 1 - \frac{n(n-1)\dots(n-k+1)}{n^k}$$

Also, $P(A) = 1$ if $n < k$.

1.3.4 problem 30

Suppose the word consists of 7 letters. Once we choose the first letter, the seventh one has to be the same. Once we choose the second letter, the sixth one has to be the same. In general, we are free to choose 4 letters. Hence, the probability that a 7 letter word is a palindrome is

$$\frac{26^4}{26^7} = \frac{1}{26^3}$$

If the word consists of 8 letters, then there are 26^8 possible words, but for a palindrome, the number of letters we are free to choose is still 4. Hence, the probability is

$$\frac{26^4}{26^8} = \frac{1}{26^4}$$

1.3.5 problem 32

Call the two black cards B_1, B_2 and the two red cards R_1, R_2. Since every configuration of the 4 cards is equally likely, each outcome has a probability of $\frac{1}{24}$ of occurance.

Case 1: $j = 0$.

If both guesses are incorrect, then both of them are black cards. There are two choices for the configuration of the black cards and for each, there are two choices for the configuration of the red cards for a total of 4 possibilities.

$$P(j = 0) = \frac{4}{24} = \frac{1}{6}$$

Case 2: $j = 4$

Notice that to guess all the cards correctly, we only need to guess correctly the two red cards, which, by symmetry, is as likely as guessing both of them wrong.

Hence,

$$P(j = 4) = P(j = 0) = \frac{1}{6}$$

Case 3: $j = 2$

One of the guesses is red the other is black. Like before, there are two choices for the red and two choices for the black cards. This undercounts the possibilities by a factor of 2, since we can switch the places of the red and the black cards. Hence,

$$P(j = 2) = \frac{2}{6} + \frac{2}{6} = \frac{2}{3}$$

Notice that getting both right, none right and one right are all the possible outcomes. Hence,

$$P(j = 1) = P(j = 3) = 0$$

1.3.6 problem 35

We can generate a random hand of 13 cards with the desired property by the following process:

1. Pick a suite to sample 4 cards from
2. Sample 3 cards for each one of the other suites

There are 4 suites and $\binom{13}{4}$ ways to sample 4 cards for any of one of them.

By the multiplication rule, there are $\binom{13}{3}^3$ ways to sample 3 cards of every one of the remaining 3 suits.

By the multiplication rule, the total number of possibilities is $4\binom{13}{4}\binom{13}{3}^3$.

The unconstrained number of 13-card hands is $\binom{52}{13}$.

Since each hand is equally likely, by the naive definition of probability, the desired likelihood is

$$\frac{4\binom{13}{4}\binom{13}{3}^3}{\binom{52}{13}}$$

1.3.7 problem 36

We can think of the problem as sampling with replacement where order matters.

There are 6^{30} possible sequences of outcomes. We are interested in the cases where each face of the die is rolled exactly 5 times. Since each sequence is equally likely, we can use the naive definition of probability.

There are $\binom{30}{5}$ ways to select the dice that fall on a 1. Then, $\binom{25}{5}$ ways to select the dice falling on a 2, $\binom{20}{5}$ falling on a 3, $\binom{15}{5}$ falling on a 4, $\binom{10}{5}$ falling on a 5 and finally, $\binom{5}{5}$ falling on a 6.

Thus, the desired probability is

$$\frac{\binom{30}{5}\binom{25}{5}\binom{20}{5}\binom{15}{5}\binom{10}{5}\binom{5}{5}}{6^{30}}$$

Alternatively, imagining the sample space to be a 30 digit long sequence of $1, 2 \ldots 6$, we want the cases in which each of $1, 2 \ldots 6$ numbers appear exactly five times. There are $\frac{30!}{(5!)^6}$ ways to arrange such a sequence. Hence, the probability is

$$\frac{30!}{(5!)^6 6^{30}}$$

1.3.8 problem 37

(a) Ignore all the cards except J, Q, K, A. There are 16 of those, 4 of which are aces. Each card has an equal chance of being first in the list, so the answer is $\frac{1}{4}$.

Source: https://math.stackexchange.com/a/3726869/649082

(b) Ignore all the cards except J, Q, K, A. There are 4 choices for a king, 4 choices for a queen and 4 choices for a jack with 3! permutations of the cards. Then, there are 4 choices for an ace. The remaining 12 cards can be permuted in 12! ways, so the answer is $\frac{4^3 \times 3! \times 4 \times 12!}{16!}$.

1.3.9 problem 38

(a) There are 12 choices of seats for Tyron and Cersei so that they sit next to each other (11 cases, where they take $i-1$ and i positions and 1 case, where they take 1 and 12th position, because table is round). Tyron can sit to the left or to the right of Cersei. The remaining 10 people can be ordered in 10! ways, so the answer is

$$\frac{24 \times 10!}{12!} = \frac{2}{11}$$

(b) There are $\binom{12}{2}$ choices of seats to be assigned to Tyron and Cersei, but only 12 choices where they sit next to each other. Since every assignment of seats is equally likely the answer is

$$\frac{12}{\binom{12}{2}} = \frac{2}{11}$$

1.3.10 problem 39

There are a total of $\binom{2N}{K}$ possible committees of K people. There are $\binom{N}{j}$ ways to select j couples for the committee. $K-2j$ people need to be selected from the remaining $N-j$ couples such that only one person is selected from a couple. First, we select $K-2j$ couples from the remaining $N-j$ couples. Then, for each of the selected couples, there are 2 choices for committee membership.

$$\frac{\binom{N}{j}\binom{N-j}{K-2j}2^{K-2j}}{\binom{2N}{K}}$$

1.3.11 problem 40

(a) Counting strictly increasing sequences of k numbers amounts to counting the number of ways to select k elements out of the n, since for any such selection, there is exactly one increasing ordering. Thus, the answer is

$$\frac{\binom{n}{k}}{n^k}$$

(b) The problem can be thought of sampling with replacement where order doesn't matter, since there is only one non decreasing ordering of a given sequence of k numbers. Thus, the answer is

$$\frac{\binom{n-1+k}{k}}{n^k}$$

1.3.12 problem 41

We can treat this problem as sampling numbers 1 to n with replacement with each number being equally likely. There are n^n possible sequences. To count the number of sequences with exactly one of the numbers missing, we first select the missing number. There are n ways to do this. The rest of the numbers have to be sampled at least once with one number being sampled exactly twice. There are $n - 1$ choice to select the number that will be sampled twice. Finally, we have n sampled numbers which can be ordered in any of $\frac{n!}{2}$ ways, since one of the numbers is repeated. Thus, the answer is

$$\frac{n(n-1)\frac{n!}{2}}{n^n} = \frac{n(n-1)n!}{2n^n}$$

1.4 Axioms Of Probability

1.4.1 problem 43

(a) Inequality can be demonstrated using the first property of probabilities,

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

and the first axiom of probabilities,

$$P(S) = 1.$$

$$P(A) + P(B) - P(A \cap B) \leq 1 \implies P(A) + P(B) - 1 \leq P(A \cap B).$$

Strict equality holds if and only if $A \cup B = S$ where S is the sample space.

(b) Since $A \cap B \subseteq A \cup B$, $P(A \cap B) \leq P(A \cup B)$ by the second property of probabilites.

Strict equality holds if and only if $A = B$.

(c) Inequality follows directly from the first property of probabilities with strict equality if and only if $P(A \cap B) = 0$.

1.4.2 problem 44

Since $B = (B - A) \cup A$, $P(B) = P(A) + P(B - A)$ by the second axiom of probability. Rearranging terms,

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

1.4.3 problem 45

$B \triangle A = (A \cup B) - (A \cap B)$. By problem 44,

$$\begin{aligned} P(B \triangle A) &= P(A \cup B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) - P(A \cap B) \\ &= P(A) + P(B) - 2P(A \cap B) \end{aligned}$$

1.4.4 problem 46

$B_k = C_k - C_{k+1}$. Since $C_{k+1} \subseteq C_k$, $P(B_k) = P(C_k) - P(C_{k+1})$.

1.4.5 problem 47

(a) Consider the experiment of flipping a fair coin twice. The sample space S is $\{HH, HT, TH, TT\}$. Let A be the event that the first flip lands heads and B be the event that the second flip lands heads. $P(A \cap B) = \frac{1}{4}$ since $A \cap B$ corresponds to the outcome HH.

On the other hand, A corresponds to the outcomes $\{HH, HT\}$ and B corresponds to the outcomes $\{HH, TH\}$. Thus, $P(A) = P(B) = \frac{1}{2}$.

Since $P(A \cap B) = P(A)P(B)$, A and B are independent events.

(b) A_1 and B_1 should intersect such that the ratio of the area of $A_1 \cap B_1$ to the area of A_1 equals the ratio of the area of B_1 to the area of R.

As a simple, extreme case, if $A_1 = B_1$, then A and B are dependent, since the condition above is violated.

(c)

$$\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
&= P(A) + P(B) - P(A)P(B) \\
&= P(A)(1 - P(B)) + P(B) \\
&= P(A)P(B^c) + P(B) \\
&= P(A)P(B^c) + 1 - P(B^c) \\
&= 1 + P(B^c)(P(A) - 1) \\
&= 1 - P(B^c)P(A^c)
\end{aligned}$$

1.5 Inclusion Exclusion

1.5.1 problem 49

Let A_i be the event that i is never rolled for $1 \leq i \leq 6$. The event of interested then is $\bigcup\limits_{i=1}^{6} A_i$.

By inclusion-exclusion, $P(\bigcup\limits_{i=1}^{6} A_i) = \sum_{i=1}^{6} P(A_i) - \sum_{1 \leq i < j \leq 6} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq 6} P(A_i \cap A_j \cap A_k) - \cdots - P(\bigcap\limits_{i=1}^{6} A_i)$.

Now,
$P(A_i) = \frac{5^n}{6^n} = (\frac{5}{6})^n$
$P(A_i \cap A_j) = \frac{4^n}{6^n} = (\frac{4}{6})^n$
$P(A_i \cap A_j \cap A_k) = \frac{3^n}{6^n} = (\frac{3}{6})^n$
$P(A_i \cap A_j \cap A_k \cap A_w) = \frac{2^n}{6^n} = (\frac{2}{6})^n$
$P(A_i \cap A_j \cap A_k \cap A_w \cap A_z) = \frac{1^n}{6^n} = (\frac{1}{6})^n$
$P(\bigcap\limits_{i=1}^{6} A_i) = 0$

Thus, $P(\bigcup\limits_{i=1}^{6} A_i) = 6(\frac{5}{6})^n - \binom{6}{2}(\frac{4}{6})^n + \binom{6}{3}(\frac{3}{6})^n - \binom{6}{4}(\frac{2}{6})^n + \binom{6}{5}(\frac{1}{6})^n$

1.5.2 problem 52

Let A_i be the event that the i-th student takes the same seat on both days. The desired probability then is $1 - P(\bigcup\limits_{i=1}^{20} A_i)$. By inclusion exclusion principle,

$P(\bigcup_{i=1}^{20} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{21} P(A_1 \cap \cdots \cap A_{20})$,
where $P(A_i) = \frac{19!}{20!}$, $P(A_i \cap A_j) = \frac{18!}{20!}$ and so on by naive definition of probability.
Hence,

$$\begin{aligned} P(\bigcup_{i=1}^{20} A_i) &= \sum_{i=1}^{20} \frac{1}{20} - \sum_{1\le i<j\le 20} \frac{1}{20*19} + \sum_{1\le i<j<k\le 20} \frac{1}{20*19*18} - \cdots + \frac{1}{20!} \\ &= 1 - \binom{20}{2}\frac{1}{20*19} + \binom{20}{3}\frac{1}{20*19*18} - \cdots + \frac{1}{20!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{1}{20!} \\ &\approx 1 - e^{-1} \end{aligned}$$

1.5.3 problem 53

(a) $62^8 - 36^8$

(b) $62^8 - 36^8 - 36^8 + 10^8$

(c) $62^8 - 36^8 - 36^8 - 10^8 + 2(62^8 - 36^8 - 52^8 + 26^8) + 62^8 - 36^8 - 36^8 + 10^8$

1.5.4 problem 55

(a) $\frac{\binom{15}{3}\binom{22}{2}}{\binom{37}{5}}$

(b) $\frac{\binom{37}{5}-\binom{27}{5}-\binom{25}{5}-\binom{22}{5}+\binom{15}{5}+\binom{10}{5}+\binom{12}{5}}{\binom{37}{5}}$

1.6 Mixed Practice

1.6.1 problem 56

(a) $>$

(b) $<$

(c) =

We are interested in two outcomes of the samme sample space. This is, $S = \{(a_1, a_2, a_3) : a_i \in \{1, 2, 3, ..., 365\}\}$ The first outcome is $(1, 1, 1)$, and the second outcome is $(1, 2, 3)$. The answer follows, since every outcome of the sample space is equally likely.

(d) <

If the first toss is T, Martin can never win, since as soon as H is seen on any subsequent toss, the game stops, and Gale is awarded the win.

If the first toss is H, then if the second toss is also H, Martin wins. Otherwise, if the second toss is T, Gale wins, since as soon as a subsequent toss shows H, Gale is awarded a win.

Thus, Martin loses $\frac{3}{4}$ of the time.

1.6.2 problem 57

The desired event can be expressed as $\bigcup\limits_{i=1}^{10^{22}} A_i$, where A_i is the event that the i-th molecule in my breath is shared with Caesar. We can compute the desired probability using inclusion exclusion.

Since every molecule in the universe is equally likely to be shared with Caesar, and we assume our breath samples molecules with replacement, $P(\bigcap\limits_{i=1}^{n} A_i) = (\frac{1}{10^{22}})^n$.

Thus,

$$\begin{aligned}
P(\bigcup_{i=1}^{10^{22}} A_i) &= \sum_{i=1}^{10^{22}} (-1)^{i+1} \left(\frac{1}{10^{22}}\right)^i \\
&= \left(1 - \frac{1}{10^{22}}\right)^{10^{22}} \\
&\approx e^{-1}
\end{aligned}$$

1.6.3 problem 58

Explanation: https://math.stackexchange.com/questions/1936525/inclusion-exclusion-problem

(a) Let A be the event that at least 9 widgets need to be tested.

$$P(A) = 1 - P(A^c) = 1 - \frac{\binom{8}{3}3!9!}{12!}$$

(b) Similar to part a,

$$P(A) = 1 - P(A^c) = 1 - \frac{\binom{9}{3}3!9!}{12!}$$

1.6.4 problem 59

(a) $\binom{15+9}{9}$

(b) $\binom{5+9}{9}$

(c) Each of 15 bars can be given to any of 10 children, so by orderd sampling with replcement formula we have 10^{15} combinations

(d) To count amount of suitable combinations, we can subtract amount of combination, where at least one child doesn't get any bars (is example of inclusion-exclusion usage case) from total amount of combinations. $10^{15} - \sum_{i=1}^{9} (-1)^{i+1} \binom{10}{i} (10-i)^{15}$

1.6.5 problem 60

(a) n^n

(b) $\binom{2n-1}{n-1}$

(c) The least likely bootstrap sample is one where $a_1 = a_2 = \cdots = a_n$. Such a sample occurs with probability $\frac{1}{n^n}$. The most likely bootstrap sample is one where all the terms are different. Such a sample occurs with probability $\frac{n!}{n^n}$. Thus, the ratio of the probabilities is $n!$

1.6.6 problem 62

(a) $1 - k!e_k(\overrightarrow{p})$

(b) Consider the extreme case where $p_1 = 1$ and $p_i = 0$ for $i \neq 1$. Then, the probability that there is at least one birthday match is 1. In general, if $p_i > \frac{1}{365}$ for a particular i, then a birthday match is more likely, since that particular day is more likely to be sampled multiple times. Thus, it makes intuitive sense that the probability of at least one birthday match is minimized when $p_i = \frac{1}{365}$.

(c) First, consider $e_k(x_1, ..., x_n)$. We can break up this sum into the sum of three disjoint cases.

  (a) Sum of terms that contain both x_1 and x_2. This sum is given by $x_1 x_2 e_{k-2}(x_3, ..., x_n)$
  (b) Sum of terms that contain either x_1 or x_2 but not both. This sum is given by $(x_1 + x_2)\, e_{k-1}(x_3, ..., x_n)$
  (c) Sum of terms that don't contain either x_1 or x_2. This sum is give by $e_k(x_3, ..., x_n)$

Thus,

$$e_k(x_1, ..., x_n) = x_1 x_2 e_{k-2}(x_3, ..., x_n) + (x_1 + x_2) e_{k-1}(x_3, ..., x_n) + e_k(x_3, ..., x_n) \blacksquare$$

Next, compare $e_k(\overrightarrow{p})$ and $e_k(\overrightarrow{r})$. Expanding the elementary symmetric polynomials, it is easy to see that the only difference between the two are the terms that contain either the first, the second or both terms from $\overrightarrow{p}$ and $\overrightarrow{r}$ respectively.

Notice that because $r_1 = r_2 = \frac{p_1 + p_2}{2}$, the sum of the terms with only r_1 and only r_2 but not both is exactly equal to $(p_1 + p_2) e_{k-1}(x_3, ..., x_n)$. Thus, the only difference between $e_k(\overrightarrow{p})$ and $e_k(\overrightarrow{r})$ are the terms $p_1 p_2 e_{k-2}(x_3, ..., x_n)$ and $r_1 r_2 e_{k-2}(x_3, ..., x_n)$.

By the arithmetic geometric mean inequailty, $r_1 r_2 e_{k-2}(x_3, ..., x_n) \geq p_1 p_2 e_{k-2}(x_3, ..., x_n)$. Hence, $1 - k! e_k(\overrightarrow{p}) \geq 1 - k! e_k(\overrightarrow{r})$.

In other words, given birthday probabilities $\overrightarrow{p}$, we can potentially reduce the probability of having at least one birthday match by taking any two birthday probabilities and replacing them with their average. For a minimal probability of at least one birthday match then, all values p_i in $\overrightarrow{p}$ must be equal, so that averaging any p_i and p_j does not change anything.

Chapter 2

Conditional Probability

2.1 Conditioning On Evidence

2.1.1 problem 3

Let S be event that a man in the US is a smoker and C be event man has cancer. From problem conditions:

$$P(S) = 0.216$$
$$P(C|S) = 23P(C|S^c)$$
$$P(C|S^c) = \frac{1}{23}P(C|S)$$

Lets use Bayes' theorem

$$\begin{aligned}
P(S|C) &= \frac{P(S)P(C|S)}{P(C)} \\
&= \frac{P(S)P(C|S)}{P(S)P(C|S) + P(S^c)P(C|S^c)} \\
&= \frac{P(S) \cdot 23P(C|S^c)}{P(S) \cdot 23P(C|S^c) + P(S^c)P(C|S^c)} \\
&= \frac{P(S)}{23P(S) + P(S^c)} \\
&= \frac{23 \cdot 0.216}{23 \cdot 0.216 + 0.784} \\
&\approx 0.864
\end{aligned}$$

2.1.2 problem 4

(a)

$$\begin{aligned}
P(K|R) &= \frac{P(K)P(R|K)}{P(R)} \\
&= \frac{P(K)P(R|K)}{P(K)P(R|K) + P(K^c)P(R|K^c)} \\
&= \frac{p}{p + (1-p)\frac{1}{n}}
\end{aligned}$$

(b) Since $p + (1-p)\frac{1}{n} \leq 1$, $P(K|R) \geq p$ with strict equality only when $p = 1$. This result makes sense, since if Fred gets the answer right, it is more likely that he knew the answer.

2.1.3 problem 5

By symmetry, all 50 of the remaining cards are equally likely. Thus, the probability that the third card is an ace is $\frac{3}{50}$.

We can reach the same answer using the definition of conditional probability. 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. Then,

$$P(C|A,B) = \frac{P(C,A,B)}{P(A,B)} = \frac{\frac{3*49!}{52!}}{\frac{50!}{52!}} = \frac{3}{50}$$

2.1.4 problem 6

Let H be the event that 7 tosses of a coin land Heads. Let A be the event that a randomly selected coin is double-headed.

$$P(A|H) = \frac{P(A)P(H|A)}{P(A)P(H|A) + P(A^c)P(H|A^c)} = \frac{\frac{1}{100}}{\frac{1}{100} + \frac{99}{100} * \left(\frac{1}{2}\right)^7}$$

2.1.5 problem 7

(a)

$$\begin{aligned} P(D|H) &= \frac{P(D)P(H|D)}{P(H)} \\ &= \frac{P(D)P(H|D)}{P(D)P(H|D) + P(D^c)P(H|D^c)} \\ &= \frac{\frac{1}{2}(\frac{1}{100} + \frac{99}{100}\left(\frac{1}{2}\right)^7)}{\frac{1}{2}(\frac{1}{100} + \frac{99}{100}\left(\frac{1}{2}\right)^7) + \frac{1}{2}\left(\frac{1}{2}\right)^7} \\ &= 0.69 \end{aligned}$$

(b) Let C be the event that the chosen coin is double-headed.

$$\begin{aligned} P(C|H) &= P(D|H)P(C|D,H) + P(D^c|H)P(C|D^c,H) \\ &= 0.69 * 0.56 + 0 \\ &= 0.39 \end{aligned}$$

2.1.6 problem 8

Let A_1 be the event that the screen is produced by company A, B_1 be the event that the screen is produced by company B, and C_1 be the event that the screen is produced by company C. Let D be the event that the screen is defective.

$$\begin{aligned}
P(A_1|D) &= \frac{P(A_1)P(D|A_1)}{P(A_1)P(D|A_1) + P(A_1^c)P(D|A_1^c)} \\
&= \frac{P(A_1)P(D|A_1)}{P(A_1)P(D|A_1) + P(A_1^c)(P(B_1|A_1^c)P(D|B_1, A_1^c) + P(C_1|A_1^c)P(D|C_1, A_1^c))} \\
&= \frac{0.5 * 0.01}{0.5 * 0.01 + 0.5 * (0.6 * 0.02 + 0.4 * 0.03)} \\
&= 0.29
\end{aligned}$$

$\blacksquare$

2.1.7 problem 9

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

(b) If B is implied by both A_1 and A_2, knowing that B occured does not tip the probability of occurence in favor of either A_1 or A_2.

For example, let A_1 be the event that the card in my hand is the Ace of Spades. Let A_2 be the event that the card in my hand is the Ace of Hearts. Let B be the event that there are 3 aces left in the deck.

B is implied by both A_1 and A_2, and $P(A_1) = P(A_2)$. Knowing that B occured does not give one any information on whether they are holding the Ace of Spades or the Ace of Hearts, since B would have occured in both cases. Thus, $P(A_1|B) = P(A_2|B)$.

2.1.8 problem 10

(a)

$$\begin{aligned}
P(A_3|A_1) &= P(A_2|A_1)P(A_3|A_2, A_1) + P(A_2^c|A_1)P(A_3|A_2^c, A_1) \\
&= 0.8 * 0.8 + 0.2 * 0.3 = 0.7
\end{aligned}$$

(b)

$$\begin{aligned}
P(A_3|A_1^c) &= P(A_2|A_1^c)P(A_3|A_2, A_1^c) + P(A_2^c|A_1^c)P(A_3|A_2^c, A_1^c) \\
&= 0.3 * 0.8 + 0.7 * 0.3 = 0.45
\end{aligned}$$

$$\begin{aligned}P(A_3) &= P(A_1)P(A_3|A_1) + P(A_1^c)P(A_3|A_1^c)\\ &= 0.75 * 0.7 + 0.25 * 0.45 = 0.64\end{aligned}$$

2.1.9 problem 11

Using the odds form of Baye's Theorem,

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

$$\frac{0.6}{0.4} = \frac{P(A)}{P(A^c)}\frac{0.7}{0.3}$$

$$\frac{P(A)}{P(A^c)} = 0.643$$

$$P(A) = 0.39$$

2.1.10 problem 12

(a) Let A_i be the event that Alice sends bit i. Let B_j be the event that Bob recieves bit j.

$$\begin{aligned}P(A_1|B_1) &= \frac{P(A_1)P(B_1|A_1)}{P(A_1)P(B_1|A_1) + P(A_0)P(B_1|A_0)}\\ &= \frac{0.5 * 0.9}{0.5 * 0.9 + 0.5 * 0.05}\\ &= 0.95\end{aligned}$$

(b) Let $B_{j,k,l}$ be the event that Bob recieves bit tuple j, k, l.

$$\begin{aligned}P(A_1|B_{110}) &= \frac{P(A_1)P(B_{110}|A_1)}{P(A_1)P(B_{110}|A_1) + P(A_0)P(B_{110}|A_0)}\\ &= \frac{0.5 * 0.9^2 * 0.1}{0.5 * 0.9^2 * 0.1 + 0.5 * 0.05^2 * 0.95}\\ &= 0.97\end{aligned}$$

2.1.11 problem 13

(a) Let B be the event that the test done by company B is successfull. Let A be the event that the test done by company A is successfull. Let D be the event that a random person has the disease.

$$\begin{aligned} P(B) &= P(D)P(B|D) + P(D^c)P(B|D^c) \\ &= 0.01 * 0 + 0.99 * 1 \\ &= 0.99 \end{aligned}$$

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

Thus, $P(B) > P(A)$.

(b) Since the disease is so rare, most people don't have it. Company B diagnoses them correctly every time. However, in the rare cases when a person has the disease, company B fails to diagnose them correctly. Company A however shows a very good probability of an accurate diagnoses for afflicted patients.

(c) If the test conducted by company A has equal specifity and sensitivity, then it's accuracy surpasses that of company B's test if the specifity and the sensitivity are larger than 0.99. If company A manages to achieve a specifity of 1, then any positive sensitivity will result in a more accurate test. If company A achieves a sensitivity of 1, it still requires a specificity larger than 0.98, since positive cases are so rare.

2.1.12 problem 14

(a) Intuitively, $P(A|B) > P(A|B^c)$, since Peter will be in a rush to install his alarm if he knows that his house will be burglarized before the end of next year.

(b) Intuitively $P(B|A^c) > P(B|A)$, since Peter is more likely to be robbed if he doesn't have an alarm by the end of the year.

(c) See https://math.stackexchange.com/a/3761508/649082.

(d) An explanation might be that in part a, we assume Peter to be driven to not let burglars rob him, but in part b we assume the burglars to not necessarily be as driven, since we assume that if the burglers know that Peter will install an alaram before the end of the next year they might not rob him. If the burglers are driven, they might actually be more inclined to rob Peter sooner, before he actually installs the alarm.

2.1.13 problem 15

Given the inequailities and the fact that $P(A \cap B) = P(A) + P(B) - P(A \cup B)$, to maximize $P(A \cap B)$ we maximize the smallest of the three expressions. Namely, $P(A)$. Thus, we would like to know that event A occured.

2.1.14 problem 16

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

Given $P(A|B) \leq P(A)$, if $P(A|B^c) < P(A)$, then the right hand side of the equation above is strictly less than the left hand side, and we have a contradiction.

We can intuitively think of this problem as asking "How likely is X to be elected as president?" and hearing "It depends" in response. The implication is that there exists some latent event (major states vote against X) that reduces the chances of X getting elected, and if we know that the former does not occure, the chances of X getting elected improve.

2.1.15 problem 17

(a) $P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B)+P(B^c)P(A|B^c)} = 1 \implies P(B^c)P(A|B^c) = 0$.

Since $P(B^c) \neq 0$ by assumption, $P(A|B^c) = 0 \implies P(A^c|B^c) = 1$.

(b) Let A and B be independent events. Then, $P(B|A) \approx 1 \implies P(B) \approx 1$. Thus, $P(B^c) \approx 0$, and so the term $P(A|B^c)$ in the denominator in part a may be large, implying $P(A^c|B^c) \approx 0$.

For example, consider a deck of 52 cards, where all but one of the cards are the Queen of Spades. Let A be the event that the first turned card is a Queen of Spades, and let B be the event that the second turned card is a Queen of Spades, where sampling is done with replacement. Then, $P(A) = P(B) \approx 1$. Then, by independence, $P(A|B^c) \approx 1 \implies P(A^c|B^c) \approx 0$.

2.1.16 problem 18

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

Note that $P(A^c) = P(A^c \cap B) + P(A^c \cap B^c)$. Hence, $P(B \cap A^c) = P(A^c) - P(A^c \cap B^c) = 0 - P(A^c \cap B^c) = -P(A^c \cap B^c)$.

Since probabilities are nonnegative, this implies $P(B \cap A^c) = P(A^c \cap B^c) = 0$.

Thus, $P(A|B) = 1$.

2.1.17 problem 19

See https://math.stackexchange.com/q/3292400/649082

2.1.18 problem 20

(a) Since the second card is equally likely to be any of the remaining 3 cards, the probability that both cards are queens is $\frac{1}{3}$.

(b) Our sample space now consists of all order pairs of the two queens and the two jacks, where at least one card is a queen. Since all the outcomes are equally likely, the answer is $\frac{2}{10} = \frac{1}{5}$.

(c) Now, the sample space consists of all order pairs of the two queens and the two jacks, where one of the cards is the Queen of Hearts. Thus, the answer is $\frac{2}{6} = \frac{1}{3}$.

2.1.19 problem 21

(a) The sample space is $(H, H, H), (H, H, T), (H, T, H), (T, H, H)$. Since each outcome is equally likely, the answer is $\frac{1}{4}$.

(b) Since the last throw is independent of the first two, the probability that all three throws landed heads given two of them landed heads equals the probability that the third throw landed heads, which is $\frac{1}{2}$.

2.1.20 problem 27

Let G be the event that the suspect is guilty. Let T be the event that one of the criminals has blood type 1 and the other has blood type 2.

Thus,

$$P(G|T) = \frac{P(G)P(T|G)}{P(G)P(T|G) + P(G^c)P(T|G^c)} = \frac{pp_2}{pp_2 + (1-p)2p_1p_2} = \frac{p}{p + 2p_1(1-p)}$$

For $P(G|T)$ to be larger than p, p_1 has to be smaller than $\frac{1}{2}$. This result makes sense, since if $p_1 = \frac{1}{2}$, then half of the population has blood type 1, and finding it at the crime scene gives us no information as to whether the suspect is guilty.

2.1.21 problem 28

(a) $\frac{P(D|T)}{P(D^c|T)} = \frac{P(D)}{P(D^c)} \frac{P(T|D)}{P(T^c|D^c)}$.

(b) Suppose our population consists of 10000 people, and only one percent of them is afflicted with the disease. So, 100 people have the disease and 9900 people don't. Suppose the specificity and sensitivity of our test are 95 percent. Then, out of the 100 people who have the disease, 95 test positive and 5 test negative, and out of the 9900 people who do not have the disease, 9405 test negative and 495 test positive.

Thus, $P(D|T) = \frac{95}{95+495}$.

Here, we can see why specificity matters more than sensitivity. Since, the disease is rare, most people do not have it. Since specificity is measured as a percentage of the population that doesn't have the disease, small changes in specificity equate to much larger changes in the number of people than in the case of sensitivity.

2.1.22 problem 29

Let G_i be the event that the i-th child is a girl. Let C_i be the event that the i-th child has property C.

$P(G_1 \cap G_2 | (G_1 \cap C_1) \cup (G_2 \cap C_2)) = \frac{0.25(2p-p^2)}{0.5p+0.5p-0.25p^2} = \frac{0.5(2-p)}{2-0.5p} = \frac{2-p}{4-p}$.

This result confirms the idea that the more rare characteristic C is, the closer we get to specifying which child we mean when we say that at least one of the children has C.

2.2 Independence and Conditional Independence

2.2.1 problem 33

(a) $\frac{1}{2^{|C|}}$

(b) For each element of C we have four options for whether this element is in A and/or B and each option has equal probability of occurring. An element $x \in C$ is in $A \subseteq B$ if and only if ($x \in B$ and $x \in A$) or ($x \in B$ and $x \notin A$) or ($x \notin B$ and $\notin A$), i.e., 3 times out of 4. Thus, the probability that $x \in A \subseteq B$ is 3/4 and $P(A \subseteq B) = (3/4)^{100}$.

Another way to see this is by using the naive definition of probability. The sample space consists of 100 binary pairs where 1 in the 1st slot of the i-th pair indicates that the i-th element of C is in A and 1 in the 2nd slot indicates that the element is in B. Hence, $|S| = 4^{100}$. The number of elements in the set X of outcomes corresponding to $A \subseteq B$ can be counted as

$$|X| = \sum_{i=0}^{100} \binom{100}{i} 2^i = 3^{100}.$$

The binomial coefficient accounts for the number of i-element subsets B of C and 2^i is the number of all subsets A of B. This gives $P(A \subseteq B) = |X|/|S| = (3/4)^{100}$.

(c) Let p be a randomly selected person from C sampled without replacement.

$P(p \in A \cup p \in B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$.

$P(A \cup B = C) = (P(p \in A \cup p \in B))^{|C|} = \left(\frac{3}{4}\right)^{|C|}$.

2.2.2 problem 34

(a) A and B are not independent, since knowing that A occured makes G^c more likely, which in turn makes B makes more likely.

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

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

2.2.3 problem 36

(a) Since any applicant who is good at baseball is accepted to the college, the proportion of admitted students good at baseball is higher than the proportion of applicants good at baseball, because applicants include people who aren't good at either math or baseball.

(b) Let S denote the sample space. Then,

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

2.2.4 problem 37

See https://math.stackexchange.com/a/3789043/649082

2.2.5 problem 38

Let S be the event that an email is spam. Let $L = W_1^c, ..., W_{22}^c, W_{23}, W_{24}^c, ..., W_{64}^c, W_{64}, W_{65}, W_{66}^c, ..., W$

Let $q = \prod_j (1-p_j)$ where $1 \leq j \leq 100 : j \notin 23, 64, 65$.

Let $x = \prod_j (1-r_j)$ where $1 \leq j \leq 100 : j \notin 23, 64, 65$.

$$P(S|L) = \frac{P(S)P(L|S)}{P(L)} = \frac{pp_{23}p_{64}p_{65}q}{pp_{23}p_{64}p_{65}q + (1-p)r_{23}r_{64}r_{65}x}.$$

2.3 Monty Hall

2.3.1 problem 41

Let G_i be the event that the i-th door contains a goat, and let D_i be the event that Monty opens door i.

$$P(G_1|D_2, G_2) = \frac{P(G_1)P(D_2, G_2|G_1)}{P(D_2, G_2)}.$$

$$\begin{aligned}
P(G_1)P(D_2, G_2|G_1) &= \frac{2}{3}\left(P(G_2|G_1)P(D_2|G_1, G_2)\right) \\
&= \frac{2}{3}\left(\frac{1}{2}(p + (1-p)\frac{1}{2})\right) \\
&= \frac{2}{3}\left(\frac{1}{2}p + \frac{1}{4} - \frac{1}{4}p\right) \\
&= \frac{1}{6}(p+1).
\end{aligned}$$

Thus,

$$\begin{aligned}
P(G_1|D_2, G_2) &= \frac{\frac{1}{6}(p+1)}{\frac{1}{6}(p+1) + \frac{1}{3} \times \frac{1}{2}} \\
&= \frac{p+1}{p+2}.
\end{aligned}$$

Note that when $p = 1$, the result matches that of the basic Monty Hall problem.

2.3.2 problem 42

Let G_i be the event that the i-th door contains a goat, and let D_i be the event that Monty opens door i.

Let S be the event of success under the specified strategy.

(a)

$$\begin{aligned}
P(S) &= P(G_1)P(S|G_1) + P(G_1^c)P(S|G_1^c) \\
&= \frac{2}{3}p + 0 \\
&= \frac{2}{3}p.
\end{aligned}$$

Note that when $p = 1$, the problem reduces to the basic Monty Hall problem, and we get the correct solution $\frac{2}{3}$. In the case when $p =$

0, Monty never gives the contestant a chance to switch their initial, incorrect choice to the correct one, resulting in a definite failure under the specified strategy.

(b)

$$\begin{aligned}
P(G_1|D_2) &= \frac{P(G_1)P(D_2|G_1)}{P(D_2)} \\
&= \frac{P(G_1)P(D_2|G_1)}{P(G_1)P(D_2|G_1) + P(G_1^c)P(D_2|G_1^c)} \\
&= \frac{\frac{2}{6}p}{\frac{2}{6}p + \frac{1}{6}} \\
&= \frac{2p}{2p+1}.
\end{aligned}$$

Note that if $p = 1$, the problem reduces to the basic Monty Hall problem, and the solution matches that of the basic, conditional Monty Hall problem. If $p = 0$ on the other hand, then the reason Monty has opened a door is because the contestant's initial guess (Door 1) is correct. By choosing the strategy to switch, the contestant always loses.

2.3.3 problem 43

Let C_i be the event that Door i contains the car. Let D_i be the event that Monty opens Door i. Let O_i be the event that Door i contains the computer, and let G_i be the event that Door i contains the goat.

(a)

$$
\begin{aligned}
P(C_3|D_2, G_2) &= \frac{P(C_3)P(D_2, G_2|C_3)}{P(D_2, G_2)} \\
&= \frac{P(C_3)P(D_2, G_2|C_3)}{P(C_3)P(D_2, G_2|C_3) + P(C_3^c)P(D_2, G_2|C_3^c)} \\
&= \frac{\frac{1}{3} * \frac{1}{2}}{\frac{1}{3} * \frac{1}{2} + \frac{2}{3}P(G_2|C_3^c)P(D_2|G_2, C_3^c)} \\
&= \frac{\frac{1}{3} * \frac{1}{2}}{\frac{1}{3} * \frac{1}{2} + \frac{2}{3} * \frac{1}{4}} \\
&= \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{6}} \\
&= \frac{1}{2}.
\end{aligned}
$$

(b)

$$
\begin{aligned}
P(C_3|D_2, O_2) &= \frac{P(C_3)P(D_2, O_2|C_3)}{P(C_3)P(D_2, O_2|C_3) + P(C_3^c)P(D_2, O_2|C_3^c)} \\
&= \frac{P(C_3)P(D_2, O_2|C_3)}{P(C_3)P(D_2, O_2|C_3) + P(C_3^c)P(D_2, O_2|C_3^c)} \\
&= \frac{\frac{1}{3}P(O_2|C_3)P(D_2|O_2, C_3)}{\frac{1}{3}P(O_2|C_3)P(D_2|O_2, C_3) + P(C_3^c)P(D_2, O_2|C_3^c)} \\
&= \frac{\frac{1}{3} * \frac{1}{2} * p}{\frac{1}{3} * \frac{1}{2} * p + \frac{2}{3} * \frac{1}{4} * q} \\
&= \frac{\frac{1}{6}p}{\frac{1}{6}p + \frac{1}{6}q} \\
&= \frac{p}{p + (1 - p)} \\
&= p.
\end{aligned}
$$

2.3.4 problem 44

Let G_i be the event that the i-th door contains a goat, and let D_i be the event that Monty opens door i. Let S be the event that the contestant is successful under his strategy.

(a) There are two scenarios which result in the contestant selecting door 3 and Monty opening door 2. Either the car is behind door 3 and Monty randomly opens door 2, or doors 3 and 2 contain goats, and Monty opens door 2. Only the latter scenario results in a win for the contestant.

Thus,

$$P(S|D_2, G_2) = \frac{(p_1 + p_2)\frac{p_1}{p_1+p_2}}{p_3\frac{1}{2} + (p_1 + p_2)\frac{p_1}{p_1+p_2}} = \frac{p_1}{p_1 + \frac{1}{2}p_3}.$$

(b) We can slightly modify the scenario in part a where doors 3 and 2 contain goats by multiplying the probability of the scenario by $\frac{1}{2}$ to accomodate the chance that Monty might open the door with the car behind it.

$$P(S|D_2, G_2) = \frac{\frac{1}{2}(p_1 + p_2)\frac{p_1}{p_1+p_2}}{p_3\frac{1}{2} + \frac{1}{2}(p_1 + p_2)\frac{p_1}{p_1+p_2}p} = \frac{\frac{1}{2}p_1}{\frac{1}{2}p_1 + \frac{1}{2}p_3} = \frac{p_1}{p_1 + p_3}.$$

(c)

$$P(S|D_2, G_2) = \frac{p_3}{p_3 + \frac{1}{2}p_1}.$$

(d)

$$P(S|D_2, G_2) = \frac{p_3}{p_3 + p_1}.$$

2.3.5 problem 45

(a) Since the prizes are independent for each door, and since the strategy is switch doors every time, what is behind Door 1 is irrelevant.

Possible outcomes for doors 2 and 3 are Goat and Car with probability $2pq$, in which case the contestant wins, Car and Car with probability p^2, in which case the contestant wins again, and Goat and Goat with probability q^2, in which case the contestant loses.

Thus,

$$P(S) = \frac{p^2 + 2pq}{p^2 + 2pq + q^2} = \frac{p^2 + 2pq}{(p + q)^2} = p^2 + 2pq.$$

(b) There are two scenarios in which Monty opens Door 2. Either Door 3 contains a Car and Door 2 contains a Goat, which happens with probability pq, or both doors contain Goats and Monty randomly chooses to open Door 2, which happens with probability $\frac{1}{2}q^2$. Contestant wins in the first case and loses in the second case.

Thus,

$$P(S|D_2, G_2) = \frac{pq}{pq + \frac{1}{2}q^2}.$$

2.3.6 problem 46

Let S be the event of successfully getting the Car under the specified strategy. Let C_i be the event that Door i contains the Car. Let A be the event that Monty reveals the Apple, and let A_i be the event that Door i contains the Apple.

(a)

$$\begin{aligned}
P(S) &= P(S \cap C_1) + P(S \cap C_2) + P(S \cap C_3) + P(S \cap C_4) \\
&= P(C_1)P(S|C_1) + P(C_2)P(S|C_2) + P(C_3)P(S|C_3) + P(C_4)P(S|C_4) \\
&= \frac{1}{4} * 0 + 3 * \frac{1}{4} * (p+q)\frac{1}{2} \\
&= 3 * \frac{1}{4} * \frac{1}{2} \\
&= \frac{3}{8}
\end{aligned}$$

$\blacksquare$

(b)

$$\begin{aligned}
P(A) &= P(A \cap G_1) + P(A \cap A_1) + P(A \cap B_1) + P(A \cap C_1) \\
&= P(G_1)P(A|G_1) + P(A_1)P(A|A_1) + P(B_1)P(A|B_1) + P(C_1)P(A|C_1) \\
&= \frac{1}{4}p + 0 + \frac{1}{4}q + \frac{1}{4}q \\
&= \frac{1}{4}(1+q)
\end{aligned}$$

$\blacksquare$

(c)

$$P(S|A) = \frac{P(S \cap A)}{P(A)} = \frac{\frac{1}{4} * p * \frac{1}{2} + \frac{1}{4} * q * \frac{1}{2}}{\frac{1}{4}(1+q)} = \frac{\frac{1}{8}}{\frac{1}{4}(1+q)}$$

2.3.7 problem 47

(a) Contestant wins under the "stay-stay" strategy if and only if the Car is behind Door 1.

$$P(S) = \frac{1}{4}$$

(b) If the Car is not behind Door 1, Monty opens one of the two doors revealing a Goat. Contestant stays. Then, Monty opens the other door with a Goat behind it. Finally, contestant switches to the Door concealing the Car.

$$P(S) = P(C_1)P(S|C_1) + P(C_1^c)P(S|C_1^c)$$

$$P(S) = 0 + \frac{3}{4} * 1 = \frac{3}{4}$$

(c) Under the "switch-stay" strategy, if the Car is behind Door 1 the contestant loses. Given that the Car is not behind Door 1, Monty opens one of the Doors containing a Goat. The contestant will win if they switch to the Door containing the Car and will lose if they switch to the Door containing the last remaining Goat.

Thus,

$$P(S) = P(C_1)P(S|C_1) + P(C_1^c)P(S|C_1^c) = 0 + \frac{3}{4} * \frac{1}{2} = \frac{3}{8}$$

(d) Under the "switch-switch" strategy, if the car is behind Door 1, then Monty opens a door with a Goat behind it. The contestant switches to a door with a Goat behind it. Monty then opens the last door containing a Goat, at which point the contestant switches back to the door containing the Car.

If Door 1 contains a Goat, Monty opens another Door containing a Goat and presents the contestant with a choice. If the contestant switches to the remaining door containing a Goat, then Monty is forced to open Door 1, revealing the final Goat. The contestant switches to the one remaining Door which contains the Car. If, on the other hand, the contestant switches to the door containing the Car, then on the subsequent switch they lose the game.

Thus,

$$P(S) = \frac{1}{4} * 1 + \frac{3}{4} * \frac{1}{2} = \frac{5}{8}$$

(e) "Stay-Switch" is the best strategy.

2.4 First-step Analysis and Gambler's Ruin

2.4.1 problem 49

(a)

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

(b) By strong induction,

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

for $n \leq 2$.

Suppose the statement holds for all $n \leq k-1$. Let S_i be the event that the i-th trial is a success.

$$\begin{aligned}
P(A_k) &= p_k P(A_{k-1}^c|S_k) + q_k P(A_{k-1}|S_k^c) \\
&= p_k\left(1 - \left(\frac{1}{2} + 2^{k-2}b_1b_2...b_{k-1}\right)\right) + q_k\left(\frac{1}{2} + 2^{k-2}b_1b_2...b_{k-1}\right) \\
&= p_k\left(\frac{1}{2} - 2^{k-2}b_1b_2...b_{k-1}\right) + q_k\left(\frac{1}{2} + 2^{k-2}b_1b_2...b_{k-1}\right) \\
&= \frac{1}{2} + (q_k - p_k)2^{k-2}b_1b_2...b_{k-1} \\
&= \frac{1}{2} + 2b_k2^{k-2}b_1b_2...b_{k-1} \\
&= \frac{1}{2} + 2^{k-1}b_1b_2...b_{k-1}b_k
\end{aligned}$$

(c) if $p_i = \frac{1}{2}$ for some i, then $b_i = 0$ and $P(A_n) = \frac{1}{2}$.

if $p_i = 0$ for all i, then $b_i = \frac{1}{2}$ for all i. Hence, the term $2^{k-1}b_1b_2...b_{k-1}b_k$ equals $\frac{1}{2}$. Thus, $P(A_n) = 1$. This makes sense since the number of successes will be 0, which is an even number.

if $p_i = 1$ for all i, then $b_i = -\frac{1}{2}$ for all i. Hence, the term $2^{k-1}b_1b_2...b_{k-1}b_k$ will either equal to $\frac{1}{2}$ or $-\frac{1}{2}$ depending on the parity of the number of trials. Thus, $P(A_n)$ is either 0 or 1 depending on the parity of the number of trials.

This makes sense since, if every trial is a success, the number of successes will be even if the number of trials is even. The number of successes will be odd otherwise.

2.4.2 problem 52

The problem is equivalent to betting 1 increments and having A start with ki dollars, while B starts with $k(N-i)$ dollars.

Thus, $p < \frac{1}{2}$,

$$p_i = \frac{1 - \left(\frac{q}{p}\right)^{ki}}{1 - \left(\frac{q}{p}\right)^{kN}}.$$

Note that,

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

This result makes sense, since $p < \frac{1}{2}$ implies that A should lose a game with high degree of certainty over the long run.

2.4.3 problem 53

See https://math.stackexchange.com/a/2706032/649082

2.4.4 problem 54

(a) $p_k = pp_{k-1} + qp_{k+1}$ with boundary condition $p_0 = 1$.

(b) Let A_j be the event that the drunk reaches k before reaching $-j$. Then, $A_j \subseteq A_{j+1}$ since to reach $-(j+1)$ the drunk needs to pass $-j$. Note that $\bigcup_{j=1}^{\infty} A_j$ is equivalent to the event that the drunk ever reaches k, since the complement of this event, namely the event that the drunk reaches $-j$ before reaching k for all j implies that the drunk never has the time to reach k.

By assumption, $P(\bigcup_{j=1}^{\infty} A_j) = \lim_{n\to+\infty} P(A_n)$. $P(A_n)$ can be found as a result of a gambler's ruin problem.

If $p = \frac{1}{2}$,

$$P(A_n) = \frac{n}{n+k} \to 1.$$

If $p > \frac{1}{2}$,

$$P(A_n) = \frac{1-\left(\frac{q}{p}\right)^{n}}{1-\left(\frac{q}{p}\right)^{n+k}} \to 1.$$

If $p < \frac{1}{2}$,

$$P(A_n) = \frac{1-\left(\frac{q}{p}\right)^{n}}{1-\left(\frac{q}{p}\right)^{n+k}} \to \left(\frac{p}{q}\right)^{k}.$$

2.5 Simpson's Paradox

2.5.1 problem 57

(a) Suppose C_1 contains 7 green gummi bears and 8 red ones, M_1 contains 1 green gummi bear and 2 red gummi bears, C_2 contains 5 green gummi bears and no red gummi bears, M_2 contains 12 green gummi bears and 5 red gummi bears.

The proportion of green gummi bears in C_1 is $\frac{7}{15}$, which is larger than that of M_1, which is $\frac{1}{3}$. The proportion of green gummi bears in C_2 is $\frac{5}{5}$, which is larger than that of M_2, which is $\frac{12}{17}$. However, the proportion of green gummi bears in C_1+C_2 is $\frac{12}{20}$, which is less than that of M_1+M_2, which is $\frac{13}{20}$.

(b) We can imagine that it is much more difficult to get a green gummi bear out of a jar with subscript 1 than it is out of a jar with subscript 2. C jars have a lower overall success rate, because most of their green gummi bears are in C_1, which is harder to sample from compared to the jars with subscript 2.

Let A be the event that a sampled gummi bear is green. Let B be the event that the jar being sampled from is an M jar. Let C be the event that the jar being sampled from has subscript 1.

Then, by Simpson's Paradox, $P(A|B,C) < P(A|B^c,C)$, $P(A|B,C^c) < P(A|B^c,C^c)$, however, $P(A|B) > P(A|B^c)$.

2.5.2 problem 58

(a) If A and B are independent, then

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

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

Thus, Simpson's Paradox does not hold.

(b) If A and C are independent, then $P(A|B,C) < P(A|B^c,C) \implies P(A|B) < P(A|B^c)$. Thus, Simpson's Paradox does not hold.

(c) If B and C are independent, then

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

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

Since $P(A|B,C) > P(A|B^c,C)$ and $P(A|B,C^c) > P(A|B^c,C^c)$, $P(A|B) >$ $P(A|B^c)$, so Simpson's Paradox does not hold.

2.6 Mixed Problems

2.6.1 problem 60

Let D be the event that a person has the disease. Let T be the event that a person tests positive for the disease.

(a)

$$\begin{aligned}
P(D|T) &= \frac{P(D)P(T|D)}{P(T)} \\
&= \frac{p(P(A|D)P(T|D,A) + P(B|D)P(T|D,B))}{P(A)P(T|A) + P(B)P(T|B)} \\
&= \frac{p\left(\frac{1}{2}a_1 + \frac{1}{2}b_1\right)}{\frac{1}{2}\left(P(D|A)P(T|D,A) + P(D^c|A)P(T|D^c,A)\right) + \frac{1}{2}\left(P(D|B)P(T|D,B) + P(D^c|B)\right.} \\
&= \frac{\frac{1}{2}p(a_1 + b_1)}{\frac{1}{2}(pa_1 + (1-p)(1-a_2)) + \frac{1}{2}(pb_1 + (1-p)(1-b_2))}
\end{aligned}$$

(b)

$$\begin{aligned}
P(A|T) &= \frac{P(A)P(T|A)}{P(T)} \\
&= \frac{\frac{1}{2}(pa_1 + (1-p)(1-a_2))}{P(A)P(T|A) + P(B)P(T|B)} \\
&= \frac{\frac{1}{2}(pa_1 + (1-p)(1-a_2))}{\frac{1}{2}(pa_1 + (1-p)(1-a_2)) + \frac{1}{2}(pb_1 + (1-p)(1-b_2))}
\end{aligned}$$

2.6.2 problem 61

(a)

$$
\begin{aligned}
P(D|\bigcap_{i=1}^{n} T_i) &= \frac{P(D)P(\bigcap_{i=1}^{n} T_i|D)}{P(\bigcap_{i=1}^{n} T_i)} \\
&= \frac{p\prod_{i=1}^{n} a}{p\prod_{i=1}^{n} a + q\prod_{i=1}^{n} b} \\
&= \frac{pa^n}{pa^n + qb^n}
\end{aligned}
$$

(b)

$$
\begin{aligned}
P(D|\bigcap_{i=1}^{n} T_i) &= \frac{P(D)P(\bigcap_{i=1}^{n} T_i|D)}{P(\bigcap_{i=1}^{n} T_i)} \\
&= \frac{p(P(G)P(\bigcap_{i=1}^{n} T_i|D,G) + P(G^c)P(\bigcap_{i=1}^{n} T_i|D,G^c))}{P(\bigcap_{i=1}^{n} T_i)} \\
&= \frac{p(\frac{1}{2} + \frac{1}{2}a_0^n)}{P(G)P(\bigcap_{i=1}^{n} T_i|G) + P(G^c)P(\bigcap_{i=1}^{n} T_i|G^c)} \\
&= \frac{p(\frac{1}{2} + \frac{1}{2}a_0^n)}{P(G)(P(D|G)P(\bigcap_{i=1}^{n} T_i|D,G) + P(D^c|G)P(\bigcap_{i=1}^{n} T_i|D^c,G)) + P(G^c)(P(D|G^c)P(} \\
&= \frac{p(\frac{1}{2} + \frac{1}{2}a_0^n)}{\frac{1}{2} + \frac{1}{2}(pa_0^n + (1-p)b_0^n)} \\
&= \frac{p(1 + a_0^n)}{1 + pa_0^n + (1-p)b_0^n}
\end{aligned}
$$

2.6.3 problem 62

Let D be the event that the mother has the disease. Let C_i be the event that the i-th child has the disease.

(a)

$$\begin{aligned}
P(C_1^c \cap C_2^c) &= P(D)P(C_1^c \cap C_2^c|D) + P(D^c)P(C_1^c \cap C_2^c|D^c) \\
&= \frac{1}{3} * \frac{1}{4} + \frac{2}{3} \\
&= \frac{9}{12}
\end{aligned}$$

(b) The two events are not independent. If the elder child has the disease, the mother has the disease, which means the younger child has probability $\frac{1}{2}$ of having the disease. Unconditionally, the younger child has probability $\frac{1}{6}$ of having the disease.

(c)

$$\begin{aligned}
P(D|C_1^c \cap C_2^c) &= \frac{P(D)P(C_1^c \cap C_2^c|D)}{P(C_1^c \cap C_2^c)} \\
&= \frac{\frac{1}{3} * \frac{1}{4}}{\frac{1}{3} * \frac{1}{4} + \frac{2}{3}} \\
&= \frac{1}{9}
\end{aligned}$$

2.6.4 problem 63

This problem is similar to the variations on example 2.2.5 (Two Children) in the textbook.

It is true that conditioned on specific two of the three coins matching, the probability of the third coin matching is $\frac{1}{2}$, but the way the problem statement is phrased, *at least two* of the coins match. According to the Two Children problem, the result is no longer $\frac{1}{2}$. In fact, the probability of all the coins matching given at least two match is $\frac{1}{4}$.

2.6.5 problem 64

Let R_i, G_i, and B_i be the events that the i-th drawn ball is red, green or blue respectively. Let A be the event that a green ball is drawn before a blue ball.

(a) Note that if a red ball is drawn, it is placed back, as if the experiment never happened. Draws continue until a green or a blue ball is drawn. The red balls are irrelevant in the experiment. Thus, the problem reduces to removing all the red balls, and finding the probability of the first, randomly drawn ball being green.

$$\begin{aligned} P(A) &= P(R_1)P(A|R_1) + P(R_1^c)P(A|R_1^c) \\ &= rP(A) + (g+b)\frac{g}{g+b} \\ &= rP(A) + g \end{aligned}$$

Thus,

$$P(A) = \frac{g}{1-r} = \frac{g}{g+b}.$$

(b) We are interested in draws in which the first ball is green. Each completed sequence of $g+b+r$ draws is equally likely. Since the red balls are once again irrelevant, we focus on the $g+b$ draws of green or blue balls.

Thus,

$$P(A) = \frac{\binom{g+b-1}{g-1}}{\binom{g+b}{g}} = \frac{g}{g+b}.$$

(c) Let $A_{i,j}$ be the event that type i occurs before type j. Generalizing part a, we get

$$P(A_{i,j}) = \frac{p_i}{p_i + p_j}.$$

2.6.6 problem 65

(a) All $(n+1)!$ permutations of the balls are equally likely, so the probability that we draw the defective ball is $\frac{1}{n+1}$ irrespective of when we choose to draw.

(b) Consider the extreme case of the defective ball being super massive ($v >> nw$). Then, it is more likely that a person draws the defective ball rather than a non defective ball, so we want to draw last. On the other hand, if v is much smaller than nw, then, at any stage of the experiment, drawing the defective ball is less likely than not, but after each draw of a non defective ball, the probability of it being drawn increases since there are less balls left in the urn. Thus, we want to be one of the first ones to draw.

So the answer depends on the relationship of w and v.

2.6.7 problem 66

Let $S_{i,k}$ be the event that sum after i rolls of the die is k. Let l denote the roll after which $k \geq 100$. Let X_i be the event that the die lands on i.

$$P(S_{l,100}) = \sum_{i=94}^{99} P(S_{l-1,i})P(X_{100-i}|S_{l-1,i}) = \frac{1}{6}\sum_{i=94}^{99} P(S_{l-1,i})$$

$$P(S_{l,101}) = \sum_{i=95}^{99} P(S_{l-1,i})P(X_{101-i}|S_{l-1,i}) = \frac{1}{6}\sum_{i=95}^{99} P(S_{l-1,i})$$

$$P(S_{l,102}) = \sum_{i=96}^{99} P(S_{l-1,i})P(X_{102-i}|S_{l-1,i}) = \frac{1}{6}\sum_{i=96}^{99} P(S_{l-1,i})$$

$$P(S_{l,103}) = \sum_{i=97}^{99} P(S_{l-1,i})P(X_{103-i}|S_{l-1,i}) = \frac{1}{6}\sum_{i=97}^{99} P(S_{l-1,i})$$

$$P(S_{l,104}) = \sum_{i=98}^{99} P(S_{l-1,i})P(X_{104-i}|S_{l-1,i}) = \frac{1}{6}\sum_{i=98}^{99} P(S_{l-1,i})$$

$$P(S_{l,105}) = \sum_{i=99}^{99} P(S_{l-1,i})P(X_{105-i}|S_{l-1,i}) = \frac{1}{6}\sum_{i=99}^{99} P(S_{l-1,i})$$

Thus, $S_{l,100}$ is the most likely.

2.6.8 problem 67

(a) Unconditionally, each of the $c+g+j$ donuts is equally likely to be the last one. Thus, the probability that the last donut is a chocolate donut is $\frac{c}{c+g+j}$.

(b) We are interested in the event that the last donut is chocolate and the last donut that is either glazed or jelly is jelly. The probability that the last donut is chocolate is $\frac{c}{c+g+j}$. Since any ordering of glazed and jelly donuts is equally likely, the probability that the last one is a jelly donut is $\frac{j}{g+j}$. Thus, the probability of the desired event is $\frac{c}{c+g+j} * \frac{j}{g+j}$.

2.6.9 problem 68

(a)

$$OR = \frac{P(D|C)}{P(D^c|C)} * \frac{P(D^c|C^c)}{P(D|C^c)}$$

Since the disease is rare among both exposed and not exposed groups, $P(D^c|C) \approx 1$ and $P(D^c|C^c) \approx 1$. Thus,

$$OR \approx \frac{P(D|C)}{P(D|C^c)} = RR$$

(b)

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

(c) Since $P(C,D)$ also equals $P(D)P(C|D)$, reversing the roles of C and D in part b gives the result.

2.6.10 problem 69

(a)

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

(b) The worst choice for p is $\frac{1}{2}$, because then the fraction of "yes" responses is $\frac{1}{2}$ irrespective of the fraction of drug users. In other words, the number of "yes" responses tells us nothing.

(c) We can extend the result from part a.

A drug user says "yes" either if they get a "Have you used drugs" slip, or if they get a "I was born in winter" slip and they are, in fact, born in winter.

A person who has not used drugs says "yes" only in the case that they get a "I was born in winter" slip and they were, in fact, born in winter.

$$y = d(p + \frac{1}{4}(1-p)) + (1-d)(1-p)\frac{1}{4}$$

Thus,

$$d = \frac{4y + p - 1}{2p}$$

2.6.11 problem 70

Let F be the event that the coin is fair, and let H_i be the even that the i-th toss lands Heads.

(a) Both Fred and his friend are correct. Fred is correct in that the probability of there being no Heads in the entire sequence is very small. For example, there are $\binom{92}{45}$ sequences with 45 Heads and 47 Tails, but only 1 sequence of all Heads.

On the other hand, Fred's friend is correct in his assessment that any particular sequence has the same likelihood of occurance as any other sequence.

(b)

$$P(F|H_{1\leq i\leq 92}) = \frac{P(F)P(H_{1\leq i\leq 92}|F)}{P(F)P(H_{1\leq i\leq 92}|F) + P(F^c)P(H_{1\leq i\leq 92}|F^c)} = \frac{p\left(\frac{1}{2}\right)^{92}}{p\left(\frac{1}{2}\right)^{92} + (1-p)}$$

(c) For $P(F|H_{1\leq i\leq 92})$ to be larger than $\frac{1}{2}$, p must be greater than $\frac{2^{92}}{2^{92}+1}$, which is approximately equal to 1, where as for $P(F|H_{1\leq i\leq 92})$ to be less than $\frac{1}{20}$, p must be less than $\frac{2^{92}}{2^{92}+19}$, which is also approximately equal to 1. In other words, unless we know for a fact that the coin is fair, 92 Heads in a row will convince us otherwise.

2.6.12 problem 71

(a) To have j toy types after sampling i toys, we either have $j-1$ toy types after sampling $i-1$ toys, and the i-th toy is of a previously unseen type, or, we have j toy types after sampling $i-1$ toys, and the i-th toy has an already seen type.

Thus,

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

(b) Note that $p_{1,0} = 0, p_{1,1} = 1$ and $p_{i,j} = 0$ for $j > i$. Using strong induction, a proof of the recursion in part a follows.

2.6.13 problem 72

(a)

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

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

(b)

$$p_{n+1} = (a - b)a_{n+1} + b$$
$$p_{n+1} = (a - b)((a + b - 1)a_n + 1 - b) + b$$
$$p_{n+1} = (a - b)\left((a + b - 1)\frac{p_n - b}{a - b} + 1 - b\right) + b$$
$$p_{n+1} = (a + b - 1)p_n + a + b - 2ab$$

(c) Let $p = \lim_{n\to\infty} p_n$. Taking the limit of both sides of the result of part b, we get

$$p = (a + b - 1)p + a + b - 2ab$$
$$p = \frac{a + b - 2ab}{2 - (a + b)}$$

2.6.14 problem 74

a. See the first paragraph of part d.

b.

$$P(B|C_j) = P(B \wedge C_j)/P(C_j) = \frac{\binom{48}{j-1} * (j-1)! * 4 * 3 * (52-j-1)!}{\binom{48}{j-1} * (j-1)! * 4 * (52-j)!} = \frac{3}{52-j}$$

The first equality comes from Bayes' theorem. The $\binom{48}{i-1} * (i-1)!$ terms come from the ways to have the first j-1 cards be non-aces. 4*3 refers to combinations of 2 adjacent aces in the numerator. The ending factorials in the numerator and denominator come from ordering the rest of the cards.

c. We have

$$P(C_j) = \frac{\binom{48}{j-1} * (j-1)! * 4 * (52-j)!}{52!}$$

With the LOTP, part b, and the power of R, we can compute

$$P(B) = \sum_{i=1}^{49} (P(B|C_j) * P(C_j))$$

d. Argument by symmetry: Consider the events "the first card after the first ace is an ace" and "the last card after the first ace is an ace". The second event is equivalent to the last card in the deck being an ace. In addition, the two events must have the same probability, as every card drawn after the first ace is equally likely to be an ace. Therefore, the probability of the first event is 1/13.

For a proof using conditional probability:

Consider the ace of hearts and the ace of spades. The probability that the ace of hearts is the first ace to appear followed immediately by the ace of spades (call this event A) is the probability that they appear adjacent to each other in that order (call this event B) and that those two aces appear before the other two aces (call this event C).

We have

$$P(B \wedge C) = P(C|B) * P(B)$$

Now, to compute $P(C|B)$ consider that if the aces of hearts and of spades appear adjacent to each other in that order, we can consider them as a card "glued together". There are 3!=6 possible orderings of the glued together card and the other 2 aces - in 2 of them, the glued together card is first. So $P(C|B) = 1/3$.

We also have $P(B) = \frac{51*(50!)}{52!} = 1/52$, as there are 51 sets of two adjacent spaces the two aces could be, and the rest of the cards can be ordered in 50! ways.

Then, we now find $P(B \wedge C) = (1/3) * (1/52)$.

Now, instead of the aces of hearts and spades specifically, consider there are 12 possible pairs of aces that can be adjacent to each other. Then the total probability that the card after the first ace is another ace is

$$12 * (1/3) * (1/52) = 1/13$$

Chapter 3

Random Variables and Their Distributions

3.1 PMFs and CDFs

3.1.1 problem 1

If the k-th person's arrival results in the first birthday match, the first $k-1$ people have $365 * 364 * \cdots * (365-k+2)$ choices of birthday assignments such that no two people have the same birthday. The k-th person has $k-1$ choices of birthdays, since their birthday must match that of one of the first $k-1$ people.

Thus,

$$P(X=k) = \frac{365 * 364 * \cdots * (365-k+2)}{365^{k-1}} \frac{k-1}{365}$$

3.1.2 problem 2

(a) Since the trials are independent, the probability that the first $k-1$ trials fail is $(\frac{1}{2})^{k-1}$, and the probability that the k-th trial is successful is $\frac{1}{2}$. Thus, for $k \geq 1$,

$$P(X = k) = (\frac{1}{2})^{k-1} * \frac{1}{2}.$$

(b) This problem reduces to part a once a trial is performed. Whatever it's outcome, we label it failure and proceed to perform more trials until the opposite outcome is observed. Thus, for $k \geq 2$,

$$P(X = k) = (\frac{1}{2})^{k-2} * \frac{1}{2}.$$

3.1.3 problem 3

$P(Y \leq k) = P(X \leq \frac{k-\mu}{\sigma}) = F(\frac{k-\mu}{\sigma})$.

3.1.4 problem 4

To show that $F(x)$ is a CDF, we need to show that F is increasing, right-continuous, and converges to 0 and 1 in the limits.

The first condition is true since $\lfloor x \rfloor$ is increasing.

Since $\lim_{x \to a^+} F(x) = F(a)$ when $a \in \mathbb{N}$ by the definition of $F(x)$, the second condition is satisfied.

$\lim_{x \to \infty} F(x) = 1$ by the definition of $F(x)$, and also, by definition, $\lim_{x \to -\infty} F(x) = 0$. Thus, the third condition is satisfied, and $F(x)$ is a CDF.

The PMF F corresponds to is

$$P(X = k) = \frac{1}{n}$$

for $1 \leq k \leq n$ and 0 everywhere else.

3.1.5 problem 5

(a) $p(n)$ is clearly non-negative. Also,

$$\sum_{n=0}^{\infty} p(n) = \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{1}{2} * \frac{1}{1 - \frac{1}{2}} = 1.$$

Thus, $p(n)$ is a valid PMF.

(b)

$$F(x) = \sum_{n=0}^{\lfloor x \rfloor} p(n) = \frac{1}{2}\sum_{n=0}^{\lfloor x \rfloor} \frac{1}{2^n} = \frac{1}{2} * \frac{1 - \frac{1}{2^{\lfloor x \rfloor + 1}}}{1 - \frac{1}{2}} = 1 - \frac{1}{2^{\lfloor x \rfloor + 1}}$$

for $x \geq 0$ and 0 for $x < 0$.

3.1.6 problem 7

To find the probability mass function (PMF) of X, we need to determine the probabilities of Bob reaching each level from 1 to 7.

Level 1 is the highest level reached by Bob if he fails to pass level 1, which happens with probability $1 - p_1$. Thus, $P(X = 1) = 1 - p_1$.

For $2 \leq j \leq 6$, $P(X = j)$ is the probability of reaching level j but not reaching level $j + 1$. This can be calculated as

$P(X = j) = p_1 \cdots p_{j-1}(1 - p_j)$.

Bob reaches level 7 if he passes each level from 1 to 6:

$P(X = 7) = p_1\, p_2\, p_3\, p_4\, p_5\, p_6$.

3.1.7 problem 8

$P(X = k) = \frac{\binom{k-1}{4}}{\binom{100}{5}}$ for $k \geq 5$.

$P(X = k) = 0$ for $k < 5$.

3.1.8 problem 9

(a) $F(x) = pF_1(x) + (1 - p)F_2(x)$.

Let $x_1 < x_2$. Then

$$F(x_1) = pF_1(x_1) + (1 - p)F_2(x_1) < pF_1(x_2) + (1 - p)F_2(x_2) = F(x_2).$$

Since $F(x)$ is a weighted sum of right continuous functions, it is itself a right continuous function.

$$\lim_{x\to\infty} F(x) = p \lim_{x\to\infty} F_1(x) + (1 - p) \lim_{x\to\infty} F_2(x) = p + 1 - p = 1.$$

Similarly,

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

(b) Let X be an r.v. created as described. Let H be the event that coin lands heads, and T be the event that the coin lands tails.

Then, $F(X=k) = P(H)F_1(k) + P(T)F_2(k) = pF_1(k) + (1-p)F_2(k)$.

Note that this is the same CDF as in part a.

3.1.9 problem 10

(a) Let $P(n) = \frac{k}{n}$ for $n \in \mathbb{N}$. By principles of probability, $\sum_{n\in\mathbb{N}} P(n)$ must equal 1.

$$\sum_{n\in\mathbb{N}} P(n) = k \sum_{n\in\mathbb{N}} \frac{1}{n}.$$

The sum on the right side of the equality is a divergent, harmonic series. Hence. the aforementioned principle of probably is violated. Contradiction.

(b) $\sum_{n\in\mathbb{N}} \frac{1}{n^2} = \frac{\pi^2}{6}$. Thus, letting k equal $\frac{6}{\pi^2}$, the principle of probability is satisfied.

3.1.10 problem 12

(a) https://drive.google.com/file/d/1vAAxLU7hvihAHOEcHx8Nc-9xapGlzc-I/view?usp=sharing

(b) Let $I \subset X$ be the subset of the support where $P_1(x) < P_2(x)$. Then

$$\sum_{x\in X} P_1(x) = \sum_{x\in I} P_1(x) + \sum_{x\in X\setminus I} P_1(x) < \sum_{x\in I} P_2(x) + \sum_{x\in X\setminus I} P_2(x) = 1.$$

Thus, having such a property in PMFs is impossible.

3.1.11 problem 13

$$P(X=a) = \sum_{z\in Z} P(Z=z)P(X=a|Z=z) = \sum_{z\in Z} P(Z=z)P(Y=a|Z=z) = P(Y=a).$$

3.1.12 problem 14

(a)

$$1 - P(X = 0) = 1 - e^{-\lambda}$$

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = (1 - e^{-\lambda}) - e^{-\lambda}\lambda$$

(b)

$$P(X = k|X > 0) = \frac{P(X = k)}{P(X > 0)} = \begin{cases} 0 & \text{if } k = 0 \\ \frac{\lambda^k}{(e^{\lambda}-1)k!} & \text{if } k \geq 1 \end{cases}$$

3.2 Named Distributions

3.2.1 problem 15

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & , \text{ if } x < 1 \\ \frac{\lfloor x \rfloor}{n} & , \text{ if } 1 \leq x \leq n \\ 1 & , \text{ if } x > n \end{cases}$$

where $\lfloor x \rfloor$ equals the largest integer that is less than or equal to x .

3.2.2 problem 16

$$P(X = k|X \in B) = \frac{\frac{1}{|C|}}{\frac{|B|}{|C|}} = \frac{1}{|B|}$$

3.2.3 problem 17

$$P(X \leq 100) = \sum_{i=0}^{100} \binom{110}{i} (0.9)^i (0.1)^{110-i}$$

3.2.4 problem 19

The pmf of the number of games ending in a draw is $P(X = k) = \binom{n}{k}(0.6)^k(0.4)^{n-k}$ for $0 \leq k \leq n$.

Let X be the number of games that end in draws. The number of players whose games end in draws is $Y = 2X$.

3.2.5 problem 20

(a) $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $0 \leq k \leq 3$.

(b) To use the complement of the desired deven,

$$P(X > 0) = 1 - P(X = 0) = 1 - (1-p)^3 = 1 - (-p^3 + 3p^2 - 3p + 1) = p^3 - 3p^2 + 3p.$$

To prove the same by Inclusion-Exclusion,

$$P(X > 0) = \sum_{i=1}^{3} P(I_{X_i} = 1) - 3p^2 + P(\cap_{i=1}^{3} I_{X_i} = 1) = 3p - 3p^2 + p^3.$$

(c) Since p^2 and p^3 go to 0 asymptotically faster than p, when p is small, $3p - 3p^2 + p^3 \approx 3p$.

3.2.6 problem 22

(a) Let C_i be the event that i-th type of coin is chosen. Let H_k be the event that k out of the n flips land heads.

$$P(X=k) = P(C_1)P(H_k|C_1) + P(C_2)P(H_k|C_2) = \frac{1}{2}\binom{n}{k} p_1^k (1-p_1)^{n-k} + \frac{1}{2}\binom{n}{k} p_2^k (1-p_2)^{n-k}$$

(b) if $p_1 = p_2$, then X is Binomial n, k.

(c) If $p_1 \neq p_2$, then the Bernoulli trials are not independent. If, for instance, p_1 is small and p_2 is large, and after the first million flips we see two heads, this increases the likelihood that we are using the coin with probability p_1 of landing heads, which in turn tells us that subsequent flips are unlikely to be land heads.

3.2.7 problem 23

Let I_i be the indicator of the i-th person voting for Kodos. Then, $P(I_i = 1) = p_1 p_2 p_3$. Since the voters make their decisions independently, we have

n independent Bernoulli trials, which is precisely the story for a Binomial distribution.

Thus,

$$P(X=k)=\binom{n}{k}(p_1p_2p_3)^k(1-p_1p_2p_3)^{n-k}$$

3.2.8 problem 24

(a) Since tosses are independent, we expect information about two of the tosses to not provide any information about the remaining tosses. In other words, we expect the required probability to be

$$\binom{8}{k}(0.5)^k(0.5)^{8-k}=\binom{8}{k}(0.5)^8$$

for $0 \leq k \leq 8$.

To prove this, let X be the number of Heads out of the 10 tosses, and let $X_{1,2}$ be the number of Heads out of the first two tosses.

$$\begin{aligned}
P(X=k|X_{1,2}=2) &= \frac{P(X=k\cap X_{1,2}=2)}{P(X_{1,2}=2)} \\
&= \frac{(0.5)^2\binom{8}{k-2}(0.5)^{k-2}(0.5)^{8-k+2}}{(0.5)^2} \\
&= \binom{8}{k-2}(0.5)^{k-2}(0.5)^{8-k+2} \\
&= \binom{8}{k-2}(0.5)^8
\end{aligned}$$

for $2 \leq k \leq 10$, which is equivalent to $\binom{8}{k}(0.5)^8$ for $0 \leq k \leq 8$.

(b) Let $X_{\geq 2}$ be the event that at least two tosses land Heads.

$$\begin{aligned}
P(X = k | X_{\geq 2}) &= \frac{P(X = k \cap X_{\geq 2})}{X_{\geq 2}} \\
&= \frac{\binom{10}{k}(0.5)^k(0.5)^{10-k}}{1 - (0.5^{10} + 10 * 0.5^{10})}
\end{aligned}$$

for $2 \leq k \leq 10$.

To see that this answer makes sense, notice that if we over all values of k from 2 to 10, we get exactly the denominator, which means the said sum equals to 1.

3.2.9 problem 26

If $X \sim HGeom(w, b, n)$, then $n - X \sim HGeom(b, w, n)$.

If X counts the number of items sampled from the set of w items in a sample of size n, then $n - X$ counts the number of items from the set of b items in the same sample.

To see this, notice that

$$P(n - X = k) = P(X = n - k) = \frac{\binom{w}{n-k}\binom{b}{k}}{\binom{w+b}{n}}$$

3.2.10 problem 27

X is not Binomial becuase the outcome of a card is not independent of the previous cards' outcomes. For instance, if the first $n - 1$ cards match, then the probability of the last card matching is 1.

The Hypergeometric story requires sampling from two finite sets, but the matching cards isn't a set of predetermined size, so the story doesn't fit.

$$P(X = k) = \frac{\binom{n}{k}!(n-k)}{n!}$$

where $!(n - k)$ is a subfactorial.

3.2.11 problem 30

(a) The distribution is hypergeometric. We select a sample of t employees and count the number of women in the sample.

$$P(X = k) = \frac{\binom{n}{k}\binom{m}{t-k}}{\binom{n+m}{t}}$$

(b) Decisions to be promoted or not are independent from employee to employee. Thus, we are dealing with Binomial distributions.

Let X be the number of women who are promoted. Then, $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$. The number of women who are not promoted is $Y = n - X$ and so is also Binomial.

Distribution of the number of employees who are promoted is also Binomial, since each employee is equally likely to be promoted and promotions are independent of each other.

(c) Once the total number of promotions is fixed, they are no longer independent. For instance, if the first t people are promoted, the probability of the $t+1$-st person being promoted is 0.

The story fits that of the hypergeometric distribution. t promoted employees are picked and we count the number of women among them.

$$P(X = k|T = t) = \frac{\binom{n}{k}p^k(1-p)^{n-k}\binom{m}{t-k}p^{t-k}(1-p)^{m-t+k}}{\binom{n+m}{t}p^t(1-p)^{n+m-t}} = \frac{\binom{n}{k}\binom{m}{t-k}}{\binom{n+m}{t}}$$

3.2.12 problem 31

(a) Note that the distribution is not Binomial, since the guesses are not independent of each other. If, for instance, the woman guesses the first three cups to be milk-first, and she is correct, then the probability of her guessing milk-first on subsequent guesses is 0, since it is known in advance that there are only 3 milk-first cups.

Hypergeometric story fits. Let X_i be the probability that the lady guesses exactly i milk-first cups correctly.

$$P(X_i) = \frac{\binom{3}{i}\binom{3}{3-i}}{\binom{6}{3}}$$

Thus, $P(X_2) + P(X_3) = \frac{10}{\binom{6}{3}} = \frac{1}{2}$

(b) Let M be the event that the cup is milk first, and let T be the event that the lady claims the cup is milk first. Then,

$$\frac{P(M|T)}{P(M^c|T)} = \frac{P(M)}{P(M^c)}\frac{p_1}{1-p_2} = \frac{p_1}{1-p_2}$$

3.2.13 problem 32

(a) The problem fits the story of Hypergeometric distributions.

$$P(X = k) = \frac{\binom{s}{k}\binom{100-s}{10-k}}{\binom{100}{10}}$$

for $0 \leq k \leq s$.

(b)
```
> x = 75
> y_interval <- sum(dhyper(7:10, x, 100-x, 10))
> y_interval
[1] 0.7853844
```

3.2.14 problem 33

(a) The probability of a typo being caught is $p_1 + p_2 - p_1p_2$. Then,

$$P(X = k) = \binom{n}{k}(p_1 + p_2 - p_1p_2)^k(1 - (p_1 + p_2 - p_1p_2))^{n-k}$$

(b) When we know the total number of caught typos in advance, the typos caught by the first proofreader are no longer independent. For example, if we know that first proofreader has caught the first t typos, and the total number of caught typos is t, then the probability of the first proofreader catching subsequent typos is 0, since the total number of caught typos was t.

Thus, we employ a Hypergeometric distribution. Since $p_1 = p_2$, all $\binom{2n}{t}$ t-tuples of caught typos are equally likely. Hence,

$$P(X_1 = k | X_1 + X_2 = t) = \frac{\binom{n}{k}\binom{n}{t-k}}{\binom{2n}{t}}$$

3.2.15 problem 34

(a) Let Y be the number of Statistics students in the sample of size m.

$$P(Y = k) = \sum_{i=k}^{n} P(X = i)P(Y = k|X = i) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} \frac{\binom{i}{k}\binom{n-i}{m-k}}{\binom{n}{m}}$$

(b) Consider a student in a random sample of size m. Independently of other students, the student has probability p of being a statistics major. Then, the probability of k students in the sample being statistics majors is $\binom{m}{k} p^k (1-p)^{m-k}$. Thus, $Y \sim Binom(m, p)$.

3.2.16 problem 36

(a)

$$P(X = \frac{n}{2}) = \binom{n}{\frac{n}{2}} (\frac{1}{2})^n$$

(b) Using Sterling's formula

$$\binom{n}{\frac{n}{2}} = \frac{\sqrt{2\pi n}(\frac{n}{e})^n}{\sqrt{2\pi \frac{n}{2}}(\frac{\frac{n}{2}}{e})^{\frac{n}{2}} \sqrt{2\pi \frac{n}{2}}(\frac{\frac{n}{2}}{e})^{\frac{n}{2}}} = \frac{\sqrt{2}2^n}{\sqrt{\pi n}}$$

Thus,

$$P(X = \frac{n}{2}) = \sqrt{\frac{2}{\pi n}} 2^n \frac{1}{2^n} = \frac{1}{\sqrt{\frac{\pi n}{2}}}$$

3.3 Independence of r.v.s

3.3.1 problem 38

(a) Let $Y = X + 1$. Then, X and Y are clearly dependent, and $P(X < Y) = 1$.

(b) Let X be the value of a toss of a six sided die, with values 1 to 6. Let Y be the value of a toss of a six sided die, with values 7 to 12. Tosses of the two die are independent, but $P(X < Y) = 1$.

3.3.2 problem 39

Let X have a discrete uniform distribution over values $1, 2, 3, ...10$. Let $Y = 11 - X$. Then Y is also discrete uniform over the same sample space, but $P(X = Y) = 0$.

If X and Y are independent, then $P(X = Y) = \sum_{i \in S} P(X = i)P(Y = i) > 0$.

3.3.3 problem 40

(a) Suppose, toward a contradiction, that X and Y do not have the same PMF. Then there is at least one k in the support of X such that $P(X = k)$ and $P(Y = k)$ are not equal.
Note that if $P(X = Y) = 1$, then $P(X = Y|X = k) = P(X = Y|Y = k) = P(X = k|Y = k) = P(Y = k|X = k) = 1$, as an event with probability 1 will still have probability 1 conditioned on any non-zero event.
Using the above and examining Bayes' theorem, we have $P(X = k|Y = k) = P(Y = k|X = k) * P(X = k)/P(Y = k)$, which simplifies to $1 = P(X = k)/P(Y = k)$ as the conditional probabilities equal 1 as previously shown. However, this equality is impossible if $P(X = k) = / = P(Y = k)$. This contradicts the assumption that $P(X = Y) = 1$ - therefore, X and Y must have the same PMF if they are always equal.

(b) Let X, Y be r.v.s with probability 1 of equalling 1, and probability 0 of equalling any other value.
Then for $x = y = 1$ $P(X = x \wedge Y = y) = 1 = P(X = x)P(Y = y)$,

and for all other possible pairs of values x, y, $P(X = x \wedge Y = y) = 0 = P(X = x)P(Y = y)$. Therefore, X, Y can be independent in this extreme case.

3.3.4 problem 41

Let X be the event that Tom woke up at 8 in the morning. Let Y be the event that Tom has blue eyes. Let Z be the event that Tom made it to his 7 a.m. class.

Clearly Tom's eye color is independent of the time he woke up and whether he made it to his early morning class or not. However, if Tom woke up at 8, then he definitely did not make it to his 7 am class.

3.3.5 problem 43

(a) Let $X \equiv a \pmod{b}$ and $Y \equiv X + 1 \pmod{b}$. Then, $\lim_{b\to\infty} P(X < Y) = 1$.

For finite random variables X and Y, the case of $P(X < Y) \geq 1$ is not possible, since then Y can never achieve the smallest value of X, contradicting the assumption that X and Y have the same distribution.

(b) If X and Y are independent random variables with the same distribution, then $P(X < Y) \leq \frac{1}{2}$

3.3.6 problem 44

(a) $P(X \bigoplus Y) \sim \text{Bern}(\frac{p}{2})$

(b) If $p \neq \frac{1}{2}$, $X \bigoplus Y$ and Y are not independent. Imagine that $X = 0$ is extremely unlikely. Then, knowing that $Y = 0$ makes it very likely that $X \bigoplus Y = 1$. If $p = \frac{1}{2}$, then $X \bigoplus Y$ and Y are independent.

$X \bigoplus Y$ and X are independent, since knowledge of X still keeps the probability of $Y = 1$ at $\frac{1}{2}$

(c) Let the largest element in J be m.

$$\begin{aligned} P(Y_J = 1) &= P(X_m = 1)P(Y_{J\setminus\{m\}} = 0) + P(X_m = 0)P(Y_{J\setminus\{m\}} = 1) \\ &= \frac{1}{2}(P(Y_{J\setminus\{m\}} = 0) + P(Y_{J\setminus\{m\}} = 1)) \\ &= \frac{1}{2} \end{aligned}$$

Thus, $Y_J \sim Bern(\frac{1}{2})$

To prove pairwise independence: Let J, J' be two arbitrary subsets of $\{1...n\}$. We want to show that $P(Y_J = a \wedge Y_{J'} = b) = P(Y_J = a)P(Y_J = b) = 1/4$ for $a, b \in \{0, 1\}$, with the second equality coming from our knowledge that $Y_J \sim Bern(1/2)$ for all J.

First, let us note that for all J, J' that are disjoint, $Y_J, Y_{J'}$ are independent - this follows from the independence of the X_i.

Now, suppose J, J' are not disjoint. Let $A = J \cap J'$, let $B = J \setminus A$, and let $C = J' \setminus B$. By definition, A, B, C are disjoint.

Now, we have

$$P(Y_J = a \wedge Y_{J'} = b) = P(Y_J = a, Y_{J'} = b | Y_A = 1)P(Y_A = 1) + P(Y_J = a, Y_{J'} = b | Y_A = 0)P(Y_A$$

using the LOTP. Continuing, we have

$$= P(Y_B = 1-a, Y_C = 1-b | Y_A = 1)P(Y_A = 1) + P(Y_B = a, Y_C = b | Y_A = 0)P(Y_A = 1)$$

by noting that if $x \oplus 1 = y$, we must have $y = 1 - x$ and if $x \oplus 1 = y$, we have $y = x$. Continuing, we get

$$= P(Y_B = 1-a)P(Y_C = 1-b)P(Y_A = 1) + P(Y_B = a)P(Y_C = b)P(Y_A = 0)$$

We can remove the conditioning since A, B, C are disjoint, and therefore Y_B, Y_C, Y_A are all independent r.v.s. Finally, we realize that since all Y are $Bern(1/2)$, this results in

$$= (1/2)^3 + (1/2)^3 = 1/4 = P(Y_J = a)P(Y_{J'} = b)$$

as desired - $Y_J, Y_{J'}$ are independent for any pair J, J'.

To prove that the Y_J are not all independent, consider the subsets $S = \{1\}, S' = \{2\}, S''\{1,2\}$. It is clear that if $Y_S = 1$ and $Y_{S'} = 1$, then $Y_{S''} = Y_S \oplus Y_{S'} = 0$. However, this implies that

$$P(\bigcap_{J \subseteq \{1..n\}} Y_J = 1) = 0$$

i.e. it is impossible for all Y to simultaneously equal 1. However, we know that

$$\prod_{J \subseteq \{1..n\}} P(Y_J = 1) = (1/2)^{2n-1} \neq 0$$

Thus, the Y_J are not independent.

3.4 Mixed Practice

3.4.1 problem 46

If a failure is seen on the first trial, then there are 0 successes and 1 failure, so it is clearly possible that there are more than twice as many failures as successes.

(a) If we think of the Bernoulli trial success as a win for player A, and the Bernoulli trial failure as a loss for player A, then have more than twice as many failures as successes is analogous to A losing the Gambler's Ruin starting with 1 dollar. For instance, if A wins the first gamble, then A has 3 dollars, and B needs $2 * 1 + 1$ gamble wins for A to lose the entire game.

Thus, we need to find p_1.

(b) $p_k = \frac{1}{2}p_{k+2} + \frac{1}{2}p_{k-1}$ with conditions $p_0 = 1$ and $\lim_{k\to\infty} p_k = 0$

The characteristic equation is $\frac{1}{2}t^3 - t + \frac{1}{2} = 0$ with roots 1 and $\frac{-1\pm\sqrt{5}}{2}$. Thus,

$$p_k = c_1 + c_2(\frac{-1+\sqrt{5}}{2})^k + c_3(\frac{-1-\sqrt{5}}{2})^k$$

Using the hint that $\lim_{k\to\infty} p_k = 0$, c_1 and c_3 must be 0. Thus,

$$p_k = c_2(\frac{-1+\sqrt{5}}{2})^k$$

Using $p_0 = 0$, we get that $c_2 = 1$. Thus,

$$p_k = (\frac{-1+\sqrt{5}}{2})^k$$

(c)

$$p_1 = \frac{-1+\sqrt{5}}{2}$$

3.4.2 problem 47

(a) Consider the simple case of $m < \frac{n}{2}$. Then, the trays don't have enough pages to print n copies. Desired probability is 0.

On the other hand, if $m \geq n$, then desired probability is 1, since each tray individually has enough pages.

Now, consider the more interesting case that $\frac{n}{2} \leq m < n$. Associate n pages being taken from the trays with n independent Bernoulli trials. Sample from the first tray on success, and sample from the second tray on failure. Thus, the assignment of trays can be modeled as a Binomial random variable, $X \sim \text{Bin}(n, p)$. As long as not too few pages are sampled from the first tray, the remaining pages can be sampled from the second tray. What is too few? $n - m - 1$ is too few, because $n - m - 1 + m < n$.

Hence,

$$P = \begin{cases} 0 & m < \frac{n}{2} \\ pbinom(m, n, p) - pbinom(n - m - 1, n, p) & \frac{n}{2} \leq m < n \\ 1 & m \geq n \end{cases}$$

(b) Typing out the hinted program in the R language, we get that the smallest number of papers in each tray needed to have 95 percent confidence that there will be enough papers to make 100 copies is 60.

Chapter 4

Expectation

4.1 Expectations And Variances

4.1.1 problem 1

Let N be the number of ameobas in the pond after a minutes.

$$E(N) = \frac{1}{3}(0 + 2 + 1) = 1$$

$$\text{Var}(N) = E(N^2) - (E(N))^2 = \frac{1}{3}(0 + 4 + 1) - 1 = \frac{2}{3}$$

4.1.2 problem 2

Let N be the number of days in a randomly chosen year.

$$E(N) = \frac{3}{4}365 + \frac{1}{4}366 = 365.25$$

$$\text{Var}(N) = E(N^2) - (E(N))^2 = \frac{3}{4}365^2 + \frac{1}{4}366^2 - 365.25^2 = 0.1875$$

4.1.3 problem 3

(a) Let D be the value of the die roll.

$$E(D) = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

(b) Let T_4 be the total sum of the four die rolls, and let D_i be the value of the i-th roll. Note that $T_4 = D_1 + D_2 + D_3 + D_4$. Then, by linearity of expectation,

$$E(T_4) = 4E(D_i) = 4 * 3.5 = 12.2$$

4.1.4 problem 4

Let's start defining some convenient r.v.s for this problem:

- D_i: value of the i-th roll.
- w_1: winning if one keeps playing after the first roll.
- w_2: winning if one keeps playing after the second roll.

The optimal strategy is to stop if the value of the last roll is greater than the expected winning if one keeps playing. In other words, keep rolling if doing so brings winnings that are, on average, greater than the last roll:

1. If $D_1 > E(w_1)$, STOP after 1 roll.
2. Else if $D_2 > E(w_2)$, STOP after 2 rolls.

Since the rolls are independent, we can calculate the expectations of w_1 and w_2 in reverse order.

The winning w_2 is equal to the value of the third roll:

$$E(w_2) = E(D_3) = \sum_{x=1}^{6} x\, P(D_3 = x) = \frac{1}{6} * (1+2+3+4+5+6)$$

$$E(w_2) = 3.50 \text{ dollars}$$

This reveals the second part of the optimal strategy: stop after 2 rolls if $D_2 \geq 4$.

The PMF of w_1 is given by

$$\begin{aligned} P(w_1 = x) = P(D_2 < 4, D_3 = x) &= \frac{3}{6} * \frac{1}{6} \\ &= \frac{3}{36} \text{ for } x = 1, 2, 3 \end{aligned}$$

$$\begin{aligned} P(w_1 = x) = P(D_2 = x \cup D_2 < 4, D_3 = x) &= \frac{1}{6} + \frac{3}{6} * \frac{1}{6} \\ &= \frac{9}{36} \text{ for } x = 4, 5, 6 \end{aligned}$$

Using those probabilities to calculate the expected winning w_1 from the definition of expectation:

$$E(w_1) = \sum_{x=1}^{6} x\, P(w_1 = x) = 4.25 \text{ dollars}$$

Now we can fully describe the optimal strategy, which maximizes the expected winnings:

1. If the value of the first roll is ≥ 5, STOP after 1 roll.

2. Else if the value of the second roll is ≥ 4, STOP after 2 rolls.

Finally, let's calculate the expected winning W^* of the optimal strategy. The PMF of W^* is calculated below

$$\begin{aligned} P(W^* = x) = P(D_1 < 5, D_2 < 4, D_3 = x) &= \frac{4}{6} * \frac{3}{6} * \frac{1}{6} \\ &= \frac{12}{6^3} \text{ for } x = 1, 2, 3 \end{aligned}$$

$$\begin{aligned}
P(W^* = 4) &= P(D_1 < 5, D_2 = 4 \cup D_1 < 5, D_2 < 4, D_3 = 4) \\
&= \frac{4}{6} * \frac{1}{6} + \frac{4}{6} * \frac{3}{6} * \frac{1}{6} \\
&= \frac{36}{6^3}
\end{aligned}$$

$$\begin{aligned}
P(W^* = x) &= P(D_1 = x \cup D_1 < 5, D_2 = x \cup D_1 < 5, D_2 < 4, D_3 = x) \\
&= \frac{1}{6} + \frac{4}{6} * \frac{1}{6} + \frac{4}{6} * \frac{3}{6} * \frac{1}{6} \\
&= \frac{72}{6^3} \text{ for } x = 5, 6
\end{aligned}$$

Plugging these probabilities into the definition of expectation:

$$E(W^*) = \sum_{x=1}^{6} x\, P(W^* = x) = 4.67 \text{ dollars}$$

4.1.5 problem 5

Let $X \sim \text{DUniform}(n)$.

$$E(X) = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{2}(n+1)$$

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{1}{n} \sum_{i=1}^{n} i^2 - (\frac{1}{2}(n+1))^2 = \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2$$

4.1.6 problem 6

Let N be the number of games played. Then the probability than $N = i$ is the probability of exactly 3 wins in the first $i-1$ games, and the last game being a win. $P(N = i) = 2\binom{i-1}{3}(\frac{1}{2})^3(\frac{1}{2})^{i-1-3}\frac{1}{2} = 2\binom{i-1}{3}(\frac{1}{2})^i$. Note that the factor of 2 in $P(N = i)$ is to account for either of the two players winning after i games.

Then,

$$\text{E}(N) = 2\sum_{i=4}^{7} i \binom{i-1}{3} (\frac{1}{2})^i \approx 5.81$$

$$\text{Var}(N) = \text{E}(N^2) - (\text{E}(N))^2 \approx 1.06$$

4.1.7 problem 7

(a) Let R be the birthrank of the chosen child. Then,

$P(R=3) = \frac{20}{100}\frac{1}{3} = \frac{4}{60}$

$P(R=2) = \frac{50}{100}\frac{1}{2} + \frac{20}{100}\frac{1}{3} = \frac{19}{60}$

$P(R=1) = \frac{30}{100} + \frac{50}{100}\frac{1}{2} + \frac{20}{100}\frac{1}{3} = \frac{37}{60}$

$\mathrm{E}(R) = 1\frac{37}{60} + 2\frac{19}{60} + 3\frac{4}{60} = \frac{29}{20}$

$\mathrm{Var}(R) = \mathrm{E}(R^2) - (\mathrm{E(R)})^2 = \frac{149}{60} - \frac{841}{400} \approx 0.38$

(b) $\mathrm{E}(R) = 1\frac{100}{190} + 2\frac{70}{190} + 3\frac{20}{190} = \frac{30}{19}$

$\mathrm{Var}(R) = \frac{56}{19} - (\frac{30}{19})^2 \approx 0.45$

4.1.8 problem 8

(a) Let C_i be the population of the i-th city, such that the first four cities are in the Northern region, the next three cities are in the Eastern region, the next two cities are in the Southern region, and the last city is in the Western region.

Let C be the population of a randomly chosen city.

Then $\mathrm{E}(C) = \frac{1}{10}\sum_{i=1}^{10} C_i = 2\text{million}$.

(b) $\mathrm{Var}(C) = \mathrm{E}(C^2) - (\mathrm{E}(C))^2$. $\mathrm{E}(C^2)$ can not be computed without the knowledge of population sizes of individual cities.

(c) $\mathrm{Var}(C) = \frac{1}{4}(\frac{1}{4}3\text{million} + \frac{1}{3}4\text{million} + \frac{1}{2}5\text{million} + 8\text{million}) \approx 3\text{million}$

(d) Since regions with smaller population have more cities, if a city is randomly selected, it is more likely that the city belongs to a low population region. On the other hand, if a region is selected uniformly at random first, then a randomly selected city is as likely to belong to a region with a large population as it is to belong to a region with a smaller population.

4.1.9 problem 9

Let X be the amount of money Fred walks away with.

(a) $\mathrm{E}(X) = 16000$. There is no variance under this scenario, since Fred's take home amout is fixed.

(b) $\mathrm{E}(X) = \frac{1}{2}1000 + \frac{1}{2}\frac{3}{4}32000 + \frac{1}{2}\frac{1}{4}64000 = 20500$. $\mathrm{Var}X = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \approx 4.76 * 10^8$.

(c) $\mathrm{E}(X) = \frac{3}{4}1000 + \frac{1}{4}\frac{1}{2}32000 + \frac{1}{4}\frac{1}{2}64000 = 12750$. $\mathrm{Var}X = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \approx 4.78 * 10^8$.

Option b has a higher expected win than option c, but it also has a higher variance.

4.1.10 problem 10

The probability that the game lasts n rounds is $1/2^n$.

Thus, if the winnings for n rounds is n, we must compute $\sum_{i=1}^{\infty}(i/2^i)$.

We know that $\sum_{i=1}^{\infty}(x^i) = \frac{x}{1-x}$. Deriving both sides with respect to x gives $\sum_{i=1}^{\infty}(ix^{i-1}) = \frac{1}{(1-x)^2}$. Multiplying by both sides gives $\sum_{i=1}^{\infty}(ix^i) = \frac{x}{(1-x)^2}$. Plugging in $x = 1/2$ gives the answer 2.

For the second part of the problem we need to find $\sum_{i=1}^{\infty}(i^2/2^i)$.

We know $\sum_{i=1}^{\infty}(ix^i) = \frac{x}{(1-x)^2}$. Deriving both sides with respect to x again, using the quotient rule, gives $\sum_{i=1}^{\infty}(i^2x^{i-1}) = \frac{1+x}{(1-x)^3}$. Multiplying both sides by x gives $\sum_{i=1}^{\infty}(i^2x^i) = \frac{x+x^2}{(1-x)^3}$. Plugging in $x = 1/2$ gives the answer of 6.

4.1.11 problem 11

Note that $31 = 2^4 + 2^3 + 2^2 + 2^1 + 1$. Thus, Martin can play at most 5 rounds. For every possible win, Martin makes 1 dollar. If the game reaches the fifth round, it is also possible that Martin loses and walks away with nothing.

Let X be Martin's winnings.

Then,

$$\mathrm{E}(X) = \sum_{i=1}^{5} (\frac{1}{2^i} 1) + (\frac{1}{2^5} 0) \approx 0.97$$

4.1.12 problem 12

Since $P(X = k) = P(X = -k)$, $\sum_{i=1}^{n} (iP(X = i) + (-i)P(X = -i)) = 0$. Hence, $\mathrm{E}(X) = 0$.

4.1.13 problem 14

$\mathrm{E}(X) = c \sum_{i=1}^{\infty} p^k = c(\frac{1}{1-p} - 1) = -\frac{1}{\log(1-p)} \frac{p}{1-p}$

$\mathrm{E}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = c \sum_{i=1}^{\infty} ip^i - (-\frac{1}{\log(1-p)} \frac{p}{1-p})^2 = -\frac{1}{\log(1-p)} \frac{p}{(p-1)^2} - (-\frac{1}{\log(1-p)} \frac{p}{1-p})^2$

4.1.14 problem 15

(a) Let X be the earnings by player B. Suppose B guesses a number j with probability b_j. Then,

$$\mathrm{E}(X) = \sum_{j=1}^{100} j p_j b_j$$

To maximize $\mathrm{E}(X)$ then, B should set $b_j = 1$ for the j for which jp_j is maximal. Since p_j are known, this quantity is known.

(b) Suppose player $P(A = k) = \frac{c_A}{k}$, and $P(B = k) = b_k$. Then,

$$\mathrm{E}(X) = \sum_{k=1}^{100} (k \frac{c_A}{k} b_k) = c_A$$

.

Thus, irrespective of what strategy B adopts, their expected earnings are the same, so B has no incentive to change strategies. Similar argument can be made for A.

(c) part b answers this part as well.

4.1.15 problem 16

(a) From the student's perspective, the average class size is $\mathrm{E}(X) = \frac{200}{360}100 + \frac{160}{360}10 = 60$. From the dean's perspective, the average class size is $\mathrm{E}(X) = \frac{16}{18}10 + \frac{2}{18}100 = 20$. The discrepancy comes from the fact that when surveying the dean, there are only two data points with a large number of students. However, when surveying students, there are two hundred data points with a large number of students. In a sense, the student's perspective overcounts the classes.

(b) Let C be a set of n classes with c_i students for $1 \leq i \leq n$. The dean's view of average class size then is $\mathrm{E}(X) = \sum_{i=1}^{n} \frac{c_i}{n}$. The students' view of average class size is $\mathrm{E}(X) = \sum_{i=1}^{n}(c_i \frac{c_i}{\sum_{i=1}^{n} c_i})$. In the dean's perspective, all c_i are equally weighted - $\frac{1}{n}$. However, in the students' perspective, weights scale with the size of the class. Thus, the students' perspective will always be larger than the dean's, unless all classes have the same number of students.

4.1.16 problem 17

(a) The expected number of children in a randomly selected family during a particular era is $\mathrm{E}(X) = \sum_{k=0}^{\infty} k \frac{n_k}{\sum_{k=0}^{\infty} n_k} = \frac{m_1}{m_0}$.

(b) The expected number of children in the family of a randomly selected child is $\mathrm{E}(X) = \sum_{k=0}^{\infty} k \frac{k n_k}{\sum_{k=0}^{\infty} k n_k} = \frac{m_2}{m_1}$.

(c) answer in part b is larger than the answer in part a. Since the average in part a is taken over randomly selected families, families with fewer children are weighted the same as families with more children. The average in part b, on the other hand, is taken over individual children, skewing the weights in favor of families with more children.

4.2 Named Distributions

4.2.1 problem 20

(a) This is not possible, since Y has a non-zero probability of being a number larger than 100, where as X is capped at 100.

(b) Let X be the number of contestants who enter a tournament, and let Y be the number of contestants who pass the first round. Clearly, $P(X \geq Y) = 1$.

(c) This is not possible, because if X always produces values smaller or equal to the values produced by Y, then $\mathrm{E}(X) \leq \mathrm{E}(Y)$. However, $\mathrm{E}(X) = 90$, and $\mathrm{E}(Y) = 50$.

4.2.2 problem 24

One way to think about the problem is that the event $X < r$ counts all sequences of n independent Bernoulli trials, where the number of failures is larger than $n - r$. If we extend the number of trials indefinitely, this implies that more than $n-r$ failures occured before the r-th success, because otherwise, we'd have $X \geq r$. The probability of this event is $P(Y > n - r)$.

Implication in the reverse direction can be shown analogously.

4.2.3 problem 26

(a) Let Z represent the number of flips until both Nick and Penny flip Heads. Then is $Z \sim \mathrm{FS}(p_1 p_2)$, since Nick's and Penny's flips are independent. $\mathrm{E}(Z) = \frac{1}{p_1 p_2}$

(b) The logic is analogous to part a, but success probability is $p_1+p_2-p_1p_2$.

(c)

$$P(X_1 = X_2) = \sum_{k=1}^{\infty}(((1-p)^2)^{k-1}p^2) = \frac{p}{2-p}$$

(d) By symmetry,

$$P(X_1 < X_2) = \frac{1-(\frac{p}{2-p})}{2} = \frac{1-p}{2-p}$$

4.2.4 problem 28

Let I_k be the indicator variable for the k-th location, so that $I_k = 1$ if k-th location has a treasure and $I_k = 0$ otherwise.

Let X be number of locations William checks to get t treasures, and $X_j \sim \text{HGeom}(t-1, n-1-(t-1), j)$ be the number of treasures found within j checked locations.

By symmetry. $P(I_k = 1) = \frac{t}{n}$. Then,

$$P(X = k) = P(I_k = 1)P(X_{k-1} = t-1) = \frac{t}{n}\frac{\binom{t-1}{t-1}\binom{n-1-t+1}{k-1-t+1}}{\binom{n-1}{k-1}} = \frac{t}{n}\frac{\binom{n-t}{k-t}}{\binom{n-1}{k-1}}$$

$$\text{E}(X) = \sum_{k=t}^{n} kP(I_k = 1)P(X_{k-1} = t-1) = \sum_{k=t}^{n}(k\frac{t}{n}\frac{\binom{n-t}{k-t}}{\binom{n-1}{k-1}}) = \frac{(n+1)t}{t+1}$$

4.2.5 problem 29

Random variable $f(X)$ takes values that are the probabilities of a random value taken by X. Since $X \sim \text{Geom}(p)$, $f(X) \in \{(1-p)^k p | k \in \mathbb{Z}_{\geq 0}\}$, and each value $(1-p)^k p$ of $f(X)$ occurs with probability $(1-p)^k p$. Thus,

$$\text{E}(X) = \sum_{k=0}^{\infty}((1-p)^k p)^2 = -\frac{p}{p-2}$$

for $|p-1|^2 < 1$.

4.2.6 problem 30

(a)

$$\begin{aligned}
\text{E}(Xg(X)) &= \sum_{x=0}^{\infty} xg(x)\frac{\text{e}^{-\lambda}(\lambda)^x}{x!} \\
&= \sum_{x=1}^{\infty} xg(x)\frac{\text{e}^{-\lambda}(\lambda)^x}{x!} \\
&= \lambda\sum_{x=1}^{\infty} g(x)\frac{\text{e}^{-\lambda}(\lambda)^{x-1}}{(x-1)!} \\
&= \lambda\sum_{x=0}^{\infty} g(x+1)\frac{\text{e}^{-\lambda}(\lambda)^x}{(x)!} = \lambda\text{E}(g(X+1))
\end{aligned}$$

(b)

$$\begin{aligned}
\mathrm{E}(X^3) &= \mathrm{E}(XX^2)\\
&= \lambda\mathrm{E}((X+1)^2)\\
&= \lambda(\mathrm{E}(X^2)+\mathrm{E}(2X)+1)\\
&= \lambda(\lambda\mathrm{E}(X+1)+2\lambda+1) = \lambda(\lambda(\lambda+1)+2\lambda+1)\\
&= \lambda(\lambda^2+3\lambda+1)
\end{aligned}$$

4.2.7 problem 31

(a)

$$P(X) = \begin{cases} p+(1-p)\mathrm{Poiss}(X=k) & k=0\\ (1-p)\mathrm{Poiss}(X=k) & k>0 \end{cases}$$

(b) First, notice that $(1-I)Y \in \{0,1,2,\dots\}$. $(1-I)Y = 0$ if $I=1$, or $Y=0$. Thus $P((1-I)Y=0) = p+(1-p)P(Y=0)$. For any other value k of $(1-I)Y$, it is achieved if $I=0$ and $Y=k$. Thus, $P((1-I)Y=k) = (1-p)P(Y=k)$.

(c) $\mathrm{E}(X) = (1-p)\sum_{k=1}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = (1-p)\mathrm{E}(\mathrm{Poss}(\lambda)) = (1-p)\lambda$.
$\mathrm{E}(X) = \mathrm{E}((1-I)Y) = \mathrm{E}(1-I)\mathrm{E}(Y) = (1-p)\lambda$.

(d) $\mathrm{Var}(X) = \mathrm{E}(X^2)-(\mathrm{E}(X))^2$. $\mathrm{E}(X^2) = (1-p)e^{-\lambda}\sum_{k=1}^{\infty} k^2\frac{\lambda^k}{k!} = (1-p)\lambda(1+\lambda)$. Thus, $\mathrm{Var}(X) = (1-p)\lambda(1+\lambda)-((1-p)\lambda)^2 = (1-p)\lambda(1+p\lambda)$.

4.2.8 problem 33

Suppose $w=r=1$. The white ball is equally likely to be any of the $w+b$ balls. Also, note that the event *k-th drawn ball is the white ball* is equivalent to the event *$k-1$ black balls are drawn until the white ball is drawn.* Thus, for $X \sim \mathrm{NHGeom}(1, n-1, 1)$, $P(X=k) = P(k+1\text{-th drawn ball is white}) = \frac{1}{n}$ for $0 \le k \le n-1$.

$$P(X=k) = \frac{\binom{1+k-1}{1-1}\binom{1+n-1-1-k}{1-1}}{\binom{1+n-1}{1}} = \frac{1}{n}$$

4.3 Indicator r.v.s

4.3.1 problem 38

Let I_j be the indicator random variable for j-th person drawing the slip of paper containing their name.

Let $X = \sum_{j=1}^{n} I_j$ be the number of people who draw their name. Then, by linearity of expectation, $\mathrm{E}(X) = \mathrm{E}(\sum_{j=1}^{n} I_j) = \sum_{j=1}^{n} \mathrm{E}(I_j) = \sum_{j=1}^{n} \frac{1}{n} = 1$.

4.3.2 problem 39

Let $I_{j,1}$ and $I_{j,2}$ be the indicator random variables for the j-th person being sampled by the first and second researchers respectively.

$P(I_{j,1} = 1) = \frac{\binom{N-1}{m-1}}{\binom{N}{m}}$. $P(I_{j,2} = 1) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}}$. Since sampling is done independently by the two researchers, $P(I_{j,1} = 1, I_{j,2} = 1) = \frac{\binom{N-1}{m-1}\binom{N-1}{n-1}}{\binom{N}{m}\binom{N}{n}}$.

Let $X = \sum_{j=1}^{n}(I_{j,1}I_{j,2})$ be the number of people sampled by both researchers. Then,

$$\mathrm{E}(X) = \mathrm{E}(\sum_{j=1}^{n}(I_{j,1}I_{j,2})) = \sum_{j=1}^{n} \mathrm{E}(I_{j,1}I_{j,2}) = \sum_{j=1}^{n} \frac{\binom{N-1}{m-1}\binom{N-1}{n-1}}{\binom{N}{m}\binom{N}{n}} = n\frac{\binom{N-1}{m-1}\binom{N-1}{n-1}}{\binom{N}{m}\binom{N}{n}}$$

4.3.3 problem 40

Let I_j be the indicator random variable for HTH pattern starting on the j-th toss. Since the tosses are independent, $P(I_j = 1) = \frac{1}{8}$ for $1 \leq j \leq n-2$.

Let $X = \sum_{j=1}^{n-2} I_j$ be the number of HTH patterns in n independent coin tosses. Then,

$$\mathrm{E}(X) = \mathrm{E}(\sum_{j=1}^{n-2} I_j) = \sum_{j=1}^{n-2} \mathrm{E}(I_j) = \sum_{j=1}^{n-2} \frac{1}{8} = \frac{n-2}{8}$$

4.3.4 problem 41

Let I_j be the indicator variable for j-th card being red. Let $R_j = I_j I_{j+1}$ be the indicator variable for the j-th and $j+1$-st cards being red. Let $X = \sum_{j=1}^{51} R_j$

be the number of consecutive red pairs in a well shuffled deck of 52 cards. Then,

$$\mathrm{E}(X) = \mathrm{E}(\sum_{j=1}^{51} R_j) = \sum_{j=1}^{51} \mathrm{E}(R_j) = \sum_{j=1}^{51} \frac{\binom{26}{2}}{\binom{52}{2}} = 51\frac{\binom{26}{2}}{\binom{52}{2}}$$

4.3.5 problem 42

Let I_j be the indicator variable for the j-th toy being of a new type. The number of toy types after collecting t toys is $X = \sum_{j=1}^{t} I_j$. $P(I_j = 1) = (\frac{n-1}{n})^{j-1}$. Thus,

$$\mathrm{E}(X) = \mathrm{E}(\sum_{j=1}^{t} I_j) = \sum_{j=1}^{t} \mathrm{E}(I_j) = \sum_{j=1}^{t} (\frac{n-1}{n})^{j-1} = n - n(\frac{n-1}{n})^t$$

4.3.6 problem 43

(a) This problem is a special case of problem 42 with $t = k$ and $n-1$ floors. Thus, the expected number of stops is $(n-1) - (n-1)(\frac{n-2}{n-1})^k$.

(b) Let I_j be the indicator variable for the j-th floor being selected for $2 \leq j \leq n$. Then, the number of stops is $X = \sum_{j=2}^{n} I_j$. Thus,

$$\mathrm{E}(X) = \mathrm{E}(\sum_{j=2}^{n} I_j) = \sum_{j=2}^{n} \mathrm{E}(I_j) = \sum_{j=2}^{n} (1 - (1 - p_j)^k)$$

4.3.7 problem 45

Notice that

$$I(A_1 \cap A_2 \cdots \cap A_n) \geq \sum_{i=1}^{n} I(A_i) - n + 1$$

because the left-hand side is either 0, or 1, so the question reduces to whether the left-hand side is ever 0, while the right-hand side is 1. Notice that this is not possible, becuase if the left-hand side is 0, then $A_j = 0$ for some j. Thus, $\sum_{i=1}^{n} I(A_i) < n \implies$ R.H.S. < 1.

Then,

$$I(\bigcap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} I(A_i) - n + 1 \implies \mathrm{E}(I(\bigcap_{i=1}^{n} A_i)) \geq \mathrm{E}(\sum_{i=1}^{n} I(A_i) - n + 1) \implies P(\bigcap_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - n +$$

4.3.8 problem 46

Let $X \sim \text{NHGeom}(4, 48, 1)$ be the number of non-aces before the first ace. Then, $\text{E}(X) = \frac{rb}{w+1} = \frac{1*48}{5} = 9.6$.

Let $Y \sim \text{NHGeom}(4, 48, 2)$ be the number of non-aces before the second ace is drawn. Then, $\text{E}(X) = \frac{rb}{w+1} = \frac{2*48}{5} = 19.2$

Let $Z = Y - X$. Notice that Z represents the number of non-aces between the first and the second ace. $\text{E}(Z) = \text{E}(Y) - \text{E}(X) = 19.2 - 9.6 = 9.6$.

4.3.9 problem 47

(a) Let $X = \sum_{i=1}^{52} I_i$ be the number of cards that are called correctly. $\text{E}(X) = \sum_{i=1}^{52} P(I_i = 1) = 52\frac{1}{52} = 1$.

(b) Source: https://math.stackexchange.com/a/1078747/649082

Let $X = \sum_{i=1}^{52} I_i$ be the number of cards that are called correctly. $\text{E}(X) = \sum_{i=1}^{52} P(I_i = 1)$. To find $P(I_i = 1)$, consider the first i cards, with the i-th card correctly guessed. Let k be the number of correctly guessed cards within the i cards. For instance, for $i = 5$, $k = 2$, Y representing a correctly guessed card and N representing an incorrectly guessed card, one possible sequence of i draws is $NYNNY$.

$$P(NYNNY) = \frac{51}{52}\frac{1}{51} \times \frac{50}{51}\frac{49}{50}\frac{1}{49} = \frac{1}{52} \times \frac{1}{51}$$

Notice that the second N in the sequence has probability $\frac{50}{51}$, because the second card is guessed correctly. The only piece of information we have is that the third card is not the card that was correctly guessed, leaving a total of 51 possibilities. Generalizing, the probability of a string of length i with k Ys is $\frac{(52-k)!}{52!}$. There are $\binom{i-1}{k-1}$ strings of length i with k Ys that end in a Y, and since $1 \leq k \leq i$,

$$P(I_i = 1) = \sum_{k=1}^{i} \binom{i-1}{k-1} \frac{(52-k)!}{52!}$$

Thus,

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{i=1}^{52}\sum_{k=1}^{i}\binom{i-1}{k-1}\frac{(52-k)!}{52!} \\
&= \sum_{k=1}^{52}\sum_{i=k}^{52}\binom{i-1}{k-1}\frac{(52-k)!}{52!} \\
&= \sum_{k=1}^{52}(\frac{(52-k)!}{52!}\sum_{i=k}^{52}\binom{i-1}{k-1}) \\
&= \sum_{k=1}^{52}(\frac{(52-k)!}{52!}\binom{52}{k}) \\
&= \sum_{k=1}^{52}\frac{1}{k!}
\end{aligned}$$

Note that $e^x = \sum_{i=0}^{\infty}\frac{x^i}{i!} \implies e^1 \approx 1 + \mathrm{E}(X) + 10^{-15}$. Thus,

$$\mathrm{E}(X) \approx e - 1$$

.

(c) Since at any given time, we know all the cards remaining in the deck, the probability of the i-th card being the card guessed correctly is $\frac{1}{52-i+1}$. Thus, $\mathrm{E}(X) = \sum_{i=1}^{52}\mathrm{E}(I_i) = \sum_{i=1}^{52}P(I_i = 1) = \sum_{i=1}^{52}\frac{1}{52-i+1} = \sum_{i=0}^{51}\frac{1}{52-i} \approx 4.54$.

4.3.10 problem 49

Let I_j be the indicator variable for the j-th prize being selected. The value recieved from the j-th prize is jI_j. Then, the total value X is $\sum_{j=1}^{n}(jI_j)$. $\mathrm{E}(X) = \sum_{j=1}^{n}jP(I_j = 1) = \sum_{j=1}^{n}j\frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \sum_{j=1}^{n}j\frac{k}{n} = \frac{k}{n}\frac{n(n+1)}{2} = \frac{k(n+1)}{2}$.

4.3.11 problem 50

Let C_1 be a random chord that spans a minor arch of length x on a circle of radius r. To generate a chord C_2, with endponts A and B, such that C_2 intersects C_1, either A is on the minor arch and B is on the major arch, or A is on the major arch and B is on the minor arch.

Let I_x be the indicator variable for C_2 intersecting C_1 when the minor arch generated by C_1 has length x. Then, $P(I_x = 1) = 2\frac{x}{2\pi * r} * \frac{2\pi * r - x}{2\pi * r}$.

Since the length of the minor arch x generated by C_1 can span from 0 to $2\pi * r$, we integrate $P(I_x = 1)$.

$$\frac{1}{2\pi * r}\int_0^{2\pi * r} 2\frac{x}{2\pi * r} * \frac{2\pi * r - x}{2\pi * r}\,\mathrm{d}x = \frac{1}{3}.$$

4.3.12 problem 52

Let I_j be the indicator variable for the j-toss landing on an outcome different from the previous toss for $2 \leq j \leq n$. Then, the total number of such tosses is $X = \sum_{j=2}^{n} I_j$. The total number of runs is $Y = X + 1$. Since $\mathrm{E}(X) = \sum_{j=2}^{n} P(I_j = 1) = \sum_{j=2}^{n} \frac{1}{2} = \frac{n-1}{2}$, $\mathrm{E}(Y) = \frac{n-1}{2} + 1 = \frac{n+1}{2}$.

4.3.13 problem 53

Let I_j be the indicator variable for tosses j and $j + 1$ landing heads for $1 \leq j \leq 3$. Then, the expected number of such pairs is $\mathrm{E}(X) = \sum_{j=1}^{3} P(I_j = 1) = 3p^2$. $\mathrm{Var}(X) = \mathrm{E}(X^2) - 9p^4$.

$\mathrm{E}(X^2) = \mathrm{E}((\sum_{j=1}^{3} I_j)^2) = \mathrm{E}((I_1+I_2)^2+2(I_1+I_2)I_3+I_3^2) = \mathrm{E}(I_1^2+2I_1I_2+I_2^2 + 2I_1I3 + 2I_2I3 + I_3^2)$.

Note that $I_j^2 = I_j$.

Note that $E(I_1I_2) = E(I_2I_3) = p^3$ as these require 3 consecutive heads to equal 1, but $E(I_1I_3) = p^4$ as this requires 4 consecutive heads to equal 1. Thus, $\mathrm{E}(X^2) = (p^2 + 2p^3 + p^2 + 2p^4 + 2p^3 + p^2) = 4p^3 + 3p^2 + 2p^4$.

Thus,

$$\mathrm{Var}(X) = 4p^3 + 3p^2 - 7p^4$$

4.3.14 problem 54

(a) Since $P(W_j = y_k)$ for $1 \leq k \leq N$ is $\frac{\binom{N-1}{n-1}}{\binom{N}{n}}\frac{1}{n} = \frac{n}{N}\frac{1}{n} = \frac{1}{N}$,

$$\mathrm{E}(W_j) = \frac{1}{N}\sum_{k=1}^{N} y_k.$$

Thus,

$$\begin{aligned}
\mathrm{E}(\overline{W}) &= \frac{1}{n}\sum_{j=1}^{n}(\frac{1}{N}\sum_{k=1}^{N}y_k) \\
&= \frac{1}{N}\sum_{k=1}^{N}y_k \\
&= \overline{y}
\end{aligned}$$

(b) Since

$$\overline{W} = \frac{1}{n}\sum_{j=1}^{N}(I_j y_j)$$

where I_j is the indicator variable for the j-th person being in the sample. Then,

$$\begin{aligned}
\mathrm{E}(\overline{W}) &= \frac{1}{n}\sum_{j=1}^{N}\frac{n}{N}y_j \\
&= \frac{1}{N}\sum_{k=1}^{N}y_k \\
&= \overline{y}
\end{aligned}$$

4.3.15 problem 56

(a) Let I_j be the indicator variable for shots j to $j+6$ being successful. The total number of succesful, consecutive, 7 shots is $X = \sum_{i=1}^{n-6} I_j$. Then,

$$\mathrm{E}(X) = \sum_{i=1}^{n-6}\mathrm{E}(I_j) = \sum_{i=1}^{n-6}P(I_j = 1) = \sum_{i=1}^{n-6}p^7 = (n-6)p^7$$

(b) Thinking of each block of 7 shots as a single trial with probability p^7 of success, let $Y \sim \mathrm{Geom}(p^7)$ be the number of failed 7-block shots taken until the first succesful 7-shot block. Then,

$$\mathrm{E}(X) = 7(1+\mathrm{E}(Y)) = 7 + \frac{7-7p^7}{p^7} = \frac{7p^7+7-7p^7}{p^7} = \frac{7}{p^7}$$

Note that it is possible that a consecutive sequence of 7 shots could happen "between" blocks - for example, this way of solving the problem does not consider the scenario where shots 2 to 8 are made. Therefore, the above calculation is a "worst case scenario" that assumes the consecutive 7 made shots must always happen in the last possible block - the actual number of blocks (and therefore shots) taken to make 7 consecutive shots is strictly less than or equal to the above calculated expectation.

4.3.16 problem 59

(a) WLOG, let $m_1 > m$ be the second median of X. Then, by the definition of medians, $P(X \leq m) \geq \frac{1}{2}$ and $P(X \geq m_1) \geq \frac{1}{2}$. Then, $P(X \in (m, m_1)) = 0$. If $m_1 > m + 1$, then there exists an $m_2 \in (m, m_1)$, such that $P(X = m_2) = 0$. This implies that $m_2 = 1$, since that is the only value of X with probability 0. However, then $m < 1$, which precludes m from being a median. Thus, m_1 must be $1 + m$. Since we know 23 to be a median of X, we need to check whether 22 or 24 are medians of X. Computation via the CDF of X shows that niether 22, nor 24 are medians. Hence, 23 is the only median of X.

(b) Let I_j be the indicator variable for the event $X \geq j$. Notice that the event $X = k$ (the first occurance of a birthday match happens when there are k people) implies that $I_j = 1$ for $j \leq k$ and vice versa. Thus,

$$X = \sum_{j=1}^{366} I_j$$

.

Then,

$$\mathrm{E}(X) = \sum_{j=1}^{366} P(I_j = 1) = 1 + 1 + \sum_{j=3}^{366} P(I_j = 1) = 2 + \sum_{j=3}^{366} p_j$$

(c) $2 + 22.61659 = 24.61659$

(d) $\mathrm{E}(X^2) = \mathrm{E}(I_1^2 + \cdots + I_{366}^2 + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1}(I_i I_j))$. Note that $I_i^2 = I_i$ and $I_i I_j = I_j$ for $i < j$. Thus,

$$\begin{aligned}
\mathrm{E}(X^2) &= \mathrm{E}(I_1 + \cdots + I_{366} + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1} I_j) \\
&= 2 + \sum_{j=3}^{366} p_j + 2\sum_{j=2}^{366}((j-1)\mathrm{E}(I_j)) \\
&= 2 + \sum_{j=3}^{366} p_j + 2\sum_{j=2}^{366}((j-1)p_j) \\
&\approx 754.61659
\end{aligned}$$

$$\mathrm{Var}(X) \approx 754.61659 - (\mathrm{E}(X))^2 \approx 754.61659 - 605.98 = 148.63659.$$

4.3.17 problem 60

(a) By the story of the problem, $X \sim \mathrm{NHGeom}(n, N-n, m)$. Then, $Y = X + m$.

(b) According to part a, $\mathrm{E}(Y) = \mathrm{E}(X) + m = \frac{m(N-n)}{n+1} + m$. The implied indicator variables are the same as in the proof of the expectation of Negative Hypergeometric random variables.

(c) The problem can be modeled with a Hypergeometric random variable $Z \sim \mathrm{HGeom}(n, N-n, \mathrm{E}(Y))$. Then, $\mathrm{E}(Z) = \mathrm{E}(Y)\frac{n}{N} = (\frac{m(N-n)}{n+1} + m)\frac{n}{N} = m \times \frac{N+1}{n+1} \times \frac{n}{N}$. Since $\frac{N+1}{n+1} \times \frac{n}{N} < 1 \implies (n+1)N(n-N) < 0 \implies \frac{N-n}{(n+1)N} > 0 \implies n < N$ for positive n and N, $\mathrm{E}(Z) < m$.

4.4 LOTUS

4.4.1 problem 62

$\mathrm{E}(2^X) = \sum_{k=0}^{\infty} 2^k P(X=k) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(2\lambda)^k}{k!} = e^{-\lambda}e^{2\lambda} = e^{\lambda}$.

4.4.2 problem 63

$\mathrm{E}(2^X) = \sum_{k=0}^{\infty} 2^k(1-p)^k p = p\sum_{k=0}^{\infty}(2-2p)^k = \frac{p}{2p-1}$ when $2 - 2p < 1 \implies p > \frac{1}{2}$.

$E(2^{-X}) = \sum_{k=0}^{\infty} 2^{-k}(1-p)^k p = p \sum_{k=0}^{\infty} (\frac{1-p}{2})^k = \frac{p}{\frac{1+p}{2}} = \frac{2p}{1+p}$ when $\frac{1-p}{2} < 1$ which is always true.

4.5 Poisson approximation

4.5.1 problem 69

There are $\binom{1000}{2}$ pairs of sampled individuals - each pair has a $\frac{1}{10^6}$ chance of being the same person. Therefore, we can estimate the "rate of occurrence" of a pair being the same person as $\binom{1000}{2} * \frac{1}{10^6} = \frac{1000*999}{2*10^6} \approx \frac{10^6}{2*10^6} = 1/2$. Therefore, the number of pairs in the sample that are the same person can be approximated by Pois(1/2).

Then the probability that there is at least one pair in the sample that are the same person is $1 - e^{-0.5} = 0.393$. This can be verified as a close approximation in R - the probability that every individual in the sample is unique is the last value resulting from the command cumprod(1-(0:999)/1000000), which is .6067. 1 minus this value gives .3933, the actual probability some two sampled individuals are the same person, which is very close to our Poisson approximation.

4.5.2 problem 71

Let I_j be the indicator random variable for pair j having the aforementioned property. $P(I_j = 1) = \frac{1}{365^2}$, under the assumption that the probability of being born on a particular day is $\frac{1}{365}$. Note that since we don't know anything about the age of the kids, we are assuming their mothers are also equally likely to be born on any of the 365 days.

Then, the expected number of pairs with the aforementioned property is $E(X) = \binom{90}{2}\frac{1}{365^2} \approx 0.03$.

Let $Z \sim$ Poiss(0.03) model the distribution of pairs with the desired property. Then, probability that there is at least one such pair is $1 - P(Z = 0) = 1 - e^{-0.03} \approx 1 - (1 - 0.03) = 0.03 = \frac{3}{100}$.

4.5.3 problem 72

(a) Suppose the population consists of n people (excluding me). Let I_j be the indicator variable for the j-th person having the same birthday as me. Then, the expected number of people with the same birthday as me is $\mathrm{E}(X) = \sum_{i=1}^{n} P(I_j = 1) = \sum_{i=1}^{n} \frac{1}{365} = \frac{n}{365}$.

Let $Z \sim \mathrm{Poss}(\frac{n}{365})$ model the distribution of the number of people in the population with the same birthday as me. Then, the probability that there is at least one person with the same birthday as me is $1 - P(Z = 0) = 1 - e^{-\frac{n}{365}}$.

$$1 - e^{-\frac{n}{365}} >= 0.5 \implies \frac{n}{365} > -\ln(0.5) \implies n > 252$$

(b) By similar logic to part a, $\mathrm{E}(X) = \frac{\binom{n}{2}}{365 \times 24}$. $1 - P(Z = 0) = 1 - e^{-\frac{\binom{n}{2}}{365 \times 24}}$.

$$1 - e^{-\frac{\binom{n}{2}}{365 \times 24}} >= 0.5 \implies \frac{\binom{n}{2}}{365 \times 24} > -\ln(0.5) \implies n > 110$$

(c) Since Poisson approximation is completely determined by the expectation of the underlying random variable, we need to increase the population size so that the expectation of the number of pairs with the desired property is the same as the expectation of the number of pairs with the same birthday when population size is 23. Since, $\mathrm{E}(X) = \frac{1}{24}\mathrm{E}(Y)$, where Y is the number of pairs of people that share a birthday, the population needs to be increased to have 24 times more pairs.

$$\binom{n}{2} = 24 * \binom{23}{2} \implies n \approx 110$$

.

(d) Let X be the number of triplets with the same birthday. Let I_j be the indicator random variable for triplet j having the same birthday. Then, $\mathrm{E}(X) = \binom{100}{3}(\frac{1}{365})^2 \approx 1.21$. Then, X can be approximated with $Z \sim \mathrm{Poiss}(1.21)$. $P(\text{at least one triplet with the same birthday}) \approx 1 - P(Z = 0) = 1 - e^{-1.21} \approx 0.7$.

Another way to approximate the desired probability is to let I_j be the indicator variable that there is a triplet born on day j. $P(I_j = 1) =$

$1-((\frac{364}{365})^{100}+100\frac{1}{365}(\frac{364}{365})^{99}+\binom{100}{2}\frac{1}{365^2}(\frac{364}{365})^{98})\approx 0.003$. Then, the expected number of days for which there is a triplet born on that day is approximately equal to $365*0.003=1.095$.

Then, the probability that there is at least one triplet born on the same day can be approximated using $Z\sim\text{Poiss}(1.095)$ - the number of days for which there is a triplet born on that day. The desired probability is $1-P(Z=0)=1-e^{-1.095}\approx 0.66$.

Thus, the second method is a closer approximation for the desired probability.

4.5.4 problem 73

(a) Let X be the number of people that play the same opponent in both rounds. Let I_j be the indicator variable that person j plays against the same opponent twice. $P(I_j=1)=\frac{1}{99}$. Then, $\text{E}(X)=\sum_{j=1}^{100}P(I_j=1)=100/99$.

(b) There is a strong dependence between trials. For instance, if we know that the first 50 players played the same opponent twice, then all of the players played the same opponents twice. Moreover, knowing each of the I_i gives us perfect information about one other I - they are strongly pairwise dependent.

(c) Consider the 50 pairs that played each other in round one. Let I_j be the indicator variable for pair j playing each other again in the second round. $P(I_j=1)=\frac{1}{99}$. Then, the expected number of pairs that play the same opponent twice is $\text{E}(Z)=\frac{50}{99}\approx\frac{1}{2}$.

We can approximate the number of pairs that play against one another in both rounds with $Z\sim\text{Poiss}(\frac{1}{2})$. Note that $X=2Z$. $P(X=0)\approx P(Z=0)=e^{-\frac{1}{2}}\approx 0.6$.

$P(X=2)\approx P(Z=1)=\frac{(\frac{1}{2})^1e^{-\frac{1}{2}}}{1!}\approx 0.32$.

Note that the approximation in part C is more accurate - the independence of the same pairs playing against each other is much stronger than the independence of individuals who play the same opponent. Knowing that the players in Game 1 of round 2 played against each other in round 1 gives us very little information about whether players in any other games also played against each other. Whereas, knowing that Player 1 in round 2 plays against the same player (say, player 71) guarantees that we know that that player 71 also plays against the same player.

4.6 Mixed practice

4.6.1 problem 79

(a) Let $X \sim FS(\frac{1}{m})$ be the number of guesses made by the hacker. Then, $\mathrm{E}(X) = m$.

(b) Suppose $w_1, w_2, w_3, \ldots, w_m$ is the sequence of passwords sampled by the hacker. Since, every permutation of the m words is equally likely, the probability that the correct password is w_i is $\frac{(m-1)!}{m!} = \frac{1}{m}$. Then $\mathrm{E}(X) = \frac{1}{m}\sum_{i=1}^{m} i = \frac{1}{m}\frac{m(m+1)}{2} = \frac{m+1}{2}$.

(c) Both m and $\frac{m+1}{2}$ are positively sloped lines, intersecting at $m = 1$. For $m = 2$, $m > \frac{m+1}{2}$. Thus, $m > \frac{m+1}{2}$ for all $m > 1$. This makes intuitive sense since when the hacker samples passwords without replacement, the number of possible passwords reduces.

(d) With replacement, $P(X = k) = (\frac{m-1}{m})^{k-1}\frac{1}{m}$ for $1 \leq k < n$ and $P(X = n) = (\frac{m-1}{m})^{n-1}\frac{1}{m} + (\frac{m-1}{m})^{n}$.

In the case of sampling without replacement, since all orderings of the passwords sampled by the hacker are equally likely, $P(\text{hacker samples k passwords}) = \frac{1}{m}$ for $1 \leq k < n$, and $P(\text{hacker samples n passwords}) = \frac{1}{m} + \frac{m-n}{m}$.

4.6.2 problem 80

(a) $X \sim \mathrm{FS}(\frac{20-m+1}{20}) \implies \mathrm{E}(X) = \frac{20}{20-m+1}$

(b) $\mathrm{E}(\sqrt{X}) = \sum_{x \in X} \sqrt{x} P(X = x) = \sum_{i=1}^{20} \sqrt{i} (\frac{m-1}{20})^{i-1} \frac{20-m+1}{20}$

4.6.3 problem 86

(a) $P(X = x, Y = y, Z = z) = \frac{\binom{n_A}{x}\binom{n_B}{y}\binom{n_C}{z}}{\binom{n}{m}}$

(b) Let I_j be the indicator variable for person j in the sample being a member of party A. Then, $X = \sum_{i=1}^{m} I_i \implies \mathrm{E}(X) = m\frac{n_A}{n}$ by symmetry.

(c) Let's find $E(X^2)$. If we square the expression for the sum of X's constituent indicator r.v.s, we get

$E(X^2) = \sum_{i=1}^{m} E(I_i^2) + 2 * \sum_{i<j, 1\leq j\leq m} E(I_i I_j)$

Since $I_i^2 = I_i$, we have $\sum_{i=1}^{m} E(I_i^2) = \frac{m*n_A}{n}$

Additionally, for any pair i, j, the r.v. $I_i I_j$ equals 1 only when some pair of samples are both members of party A, which occurs with probability $\frac{n_A(n_A-1)}{n(n-1)}$. There are $\binom{m}{2}$ pairs i, j. Therefore, the expression $2 * \sum_{i<j, 1\leq j\leq m} E(I_i I_j)$ evaluates to $\frac{n_A m(m-1)(n_A-1)}{n(n-1)}$.

Finally, we have $E(X^2)$, so now we can write

$$Var(X) = E(X^2) - EX^2 = \frac{m * n_A}{n} + \frac{n_A m(m-1)(n_A-1)}{n(n-1)} - \frac{m^2 n_A^2}{n^2}$$

When $m = 1$, $\mathrm{Var}(X) = \frac{n_A}{n} - (\frac{n_A}{n})^2 = \frac{n_A}{n}(1 - \frac{n_A}{n}) = \frac{n_A}{n} \times \frac{n_B + n_C}{n}$.
When $m = n$, Var(X) = 0. This makes sense, as if the sample is the entire population, we always get the same number of members of party A in our sample (all of them), so there is no variation.

4.6.4 problem 87

(a) Let I_j be the indicator random variable for j person in the sample being a democrat. Let X be the total number of democrats in the sample. Then, $\mathrm{E}(X) = \sum_{j=1}^{c} I_j = c\frac{d}{100}$

(b) Let I_j be the indicator random variable for state j being represented by at least one person in the sample. Then $P(I_j = 1) = 1 - \frac{\binom{98}{c}}{\binom{100}{c}}$. Then, the expected number of states represented in the sample is $\mathrm{E}(X) = 50(1 - \frac{\binom{98}{c}}{\binom{100}{c}})$.

(c) Similarly to part b, $\mathrm{E}(X) = 50\frac{\binom{98}{c-2}}{\binom{100}{c}}$.

(d) $P(X = k) = \frac{\binom{50}{k}\binom{50}{20-k}}{\binom{100}{20}}$ for $0 \leq k \leq 20$.

Letting I_j be the indicator variable for person j in the sample being a junior senator of a state, $\mathrm{E}(X) = 20\frac{50}{100} = 10$.

(e) Similar to part b, $\mathrm{E}(X) = 50\frac{\binom{98}{18}}{\binom{100}{20}}$.

4.6.5 problem 88

(a) $X \sim \mathrm{Geom}(\frac{g}{g+b}) \implies \mathrm{E}(X) = \frac{b}{g}$

(b) The answer should be equal to $\frac{b}{g}$, as since X is geometrically distributed.

4.6.6 problem 89

(a) Since $\mathrm{E}(N_C) = 115p_C$, $\mathrm{Var}(N_C) = \sum_{k=0}^{115} k^2 p_C^k - (115p_C)^2$.

(b) Let I_j be the indicator random variable that CATCAT starts at position j. Then, the expected number of CATCAT is $\mathrm{E}(X) = 110(p_C p_A p_T)^2$.

(c) In a sequence of length 6, the desired options are CATxxx, xxxCAT. Thus, $P(\text{at least one CAT}) = 2(p_C p_A p_T(1 - p_C p_A p_T)) + (p_C p_A p_T)^2$.

4.6.7 problem 90

(a) Let I_j be the indicator variable that j person in Bob's sample is also sampled by Alice. Then, $P(I_j = 1) = \frac{1}{10}$. Then, the expected number of people in Bob's and Alice's samples is 2.

(b) $|A \cup B| = 100 + 20 - |A \cap B| \implies \mathrm{E}(|A \cup B|) = 100 + 20 - \mathrm{E}(|A \cap B|) = 100 + 20 - 2 = 118$.

(c) Let I_j be the indicator random variable for couple j being in Bob's sample. Then $P(I_j = 1) = \frac{\binom{998}{18}}{\binom{1000}{20}}$. Thus, the expected number of couples in Bob's sample is $\mathrm{E}(X) = 500\frac{\binom{998}{18}}{\binom{1000}{20}} \approx 0.2$.

4.6.8 problem 91

(a) If $F = G$, Then, X_j is equally likely to be in any of the $m+n$ positions in the ordered list.

$$\mathrm{E}(R) = \sum_{j=1}^{m} \mathrm{E}(R_j) = \sum_{j=1}^{m} \frac{(m+n)(m+n+1)}{2} \frac{1}{m+n} = m\frac{m+n+1}{2}.$$

(b) $R_j = (\sum_{k=1}^{n} I_{Y_k} + \sum_{k \neq j} I_{X_k} + 1)$ where I_{Y_k} are the indicator random variables for X_j being larger than Y_k and I_{X_k} are the indicator random variables for X_j being larger than X_k. Note that $E(I_{Y_k}) = p$ for all k since the Ys are iid, and $E(I_{X_k}) = 1/2$ - X_j and X_k are iid and never equal, so they are equally likely to be bigger or smaller than the other. Then $\mathrm{E}(R_j) = np + (m-1)/2 + 1$.Thus, $\mathrm{E}(R) = m(np + (m-1)/2 + 1)$.

4.6.9 problem 92

(a) Let S be the sum of the ranks of the dishes we eat during both phases. $S = (m-k+1)X + \sum_{j=1}^{k-1} R_j$, where R_j is the rank of dish j, excluding the highest ranked dish, from the exploration phase. Since $\mathrm{E}(R_j) = \frac{(X-1)X}{2} \times \frac{\binom{X-2}{k-2}}{(k-1)\binom{X-1}{k-1}} = \frac{(X-1)X}{2} \times \frac{1}{X-1} = \frac{X}{2}$, $\mathrm{E}(S) = (m-k+1)\mathrm{E}(X) + (k-1)\frac{\mathrm{E}(X)}{2} = (m-k)\mathrm{E}(X) + (k+1)\frac{\mathrm{E}(X)}{2}$.

(b) $P(X = x) = \frac{\binom{x-1}{k-1}}{\binom{n}{k}}$.

(c)

$$
\begin{aligned}
\mathrm{E}(X) &= \frac{1}{\binom{n}{k}} \sum_{i=k}^{n} i \binom{i-1}{k-1} \\
&= \frac{1}{\binom{n}{k}} \sum_{i=k}^{n} k \binom{i}{k} \\
&= \frac{k}{\binom{n}{k}} \sum_{i=k}^{n} \binom{i}{k} \\
&= \frac{k}{\binom{n}{k}} \binom{n+1}{k+1} \\
&= \frac{k(n+1)}{k+1}
\end{aligned}
$$

(d) Plugging $\frac{k(n+1)}{k+1}$ into the result of part b and derivating $(m-k)\frac{k(n+1)}{k+1} + k\frac{n+1}{2}$ with respect to k provides an extremum of $k = \sqrt{2(m+1)} - 1$.

Chapter 5

Continuous Random Variables

5.1 PDFs and CDFs

5.1.1 problem 1

The PDF is

$$f(x) = xe^{-x^2/2}$$

$$\begin{aligned} P(x \leq a) &= \int_0^a f(x)dx && (5.1) \\ &= \int_0^a xe^{-x^2/2}dx && (5.2) \\ &= 1 - e^{-a^2/2} && (5.3) \end{aligned}$$

(a)

$$\begin{aligned} P(1 < X < 3) &= P(X \leq 3) - P(X \leq 1) && (5.4) \\ &= e^{-1/2} - e^{-9/2} && (5.5) \end{aligned}$$

(b) For first quantile q_1

$$P(X \leq q_1) = \frac{1}{4}$$
$$1 - e^{-q_1^2/2} = \frac{1}{4}$$
$$q_1 = 0.54$$

For second quantile q_2

$$P(X \leq q_2) = \frac{2}{4}$$
$$1 - e^{-q_2^2/2} = \frac{2}{4}$$
$$q_1 = 0.83$$

For third quantile q_3

$$P(X \leq q_3) = \frac{1}{4}$$
$$1 - e^{-q_3^2/2} = \frac{3}{4}$$
$$q_3 = 1.17$$

5.1.2 problem 2

Take Unif$(0, \frac{1}{2})$,

$$f(x) = \{2, x \in (0, \frac{1}{2})\}$$
$$\int_C f(x)dx = 1$$

where C is the complete space. Say, $f(x) > 1$ in a given domain X
$f(x) > 0$, so $\int_D f(x)dx \leq \int_C f(x)dx$
Here $\int_D 1dx < \int_D f(x)dx$
We can say, $|D| < 1$

5.1.3 problem 3

(a) The new PDF is,

$$g(x) = 2F(x)f(x)$$

$g(x) \geq 0$ in the same range as f

$$\begin{aligned}
\int_{-\infty}^{\infty} g(x)dx &= \int_{-\infty}^{\infty} 2F(x)f(x)dx && (5.6)\\
&= \int_{-\infty}^{\infty} d(F^2(x)) && (5.7)\\
&= \lim_{x\to\infty} F^2(x) - \lim_{x\to-\infty} F^2(x) && (5.8)\\
&= 1 - 0 && (5.9)
\end{aligned}$$

(b) The new PDF is,

$$g(X) = \frac{1}{2}(f(x) + f(-x))$$

$$\begin{aligned}
\int_{-\infty}^{\infty} g(x)dx &= \frac{1}{2}\int_{-\infty}^{\infty} f(x)dx + \frac{1}{2}\int_{-\infty}^{\infty} f(x)dx && (5.10)\\
&= 1 && (5.11)
\end{aligned}$$

5.1.4 problem 5

a. We have $A = \pi R^2$, so $E(A) = \pi E(R^2)$. We have $E(R^2) = \int_0^1 x^2 * 1\, dx = 1/3$, since the PDF of R is always 1. Then $E(A) = \pi/3$.

We have $Var(A) = E(A^2) - E(A)^2 = \pi^2 E(R^4) - \pi^2/9$ using linearity. $E(R^4) = \int_0^1 x^4 * 1\, dx = 1/5$, so $Var(A) = \pi^2/5 - \pi^2/9 = 4\pi^2/45$

b. CDF: $P(A < k) = P(\pi R^2 < k) = P(R < \sqrt{k/\pi}) = \sqrt{k/\pi}$ for $0 < k < \pi$ using the CDF of Unif(0,1). The CDF of A is 0 for $k < 0$ and $k > \pi$.

PDF: $\frac{d}{dk}(\sqrt{k/\pi}) = \frac{1}{2\sqrt{k\pi}}$ for $0 < k < \pi$ and 0 elsewhere.

5.1.5 problem 6

a. For $X \sim Unif(0,1)$, $F(k) = k$ for $0 < k < 1$, $E(X) = 1/2$, $Var(X) = 1/12$, $STD(X) = 1/2\sqrt{3}$.
Then the probability X is within one standard deviation of its mean is $P(\frac{\sqrt{3}-1}{2\sqrt{3}} < X < \frac{\sqrt{3}+1}{2\sqrt{3}}) = F(\frac{\sqrt{3}+1}{2\sqrt{3}}) - F(\frac{\sqrt{3}-1}{2\sqrt{3}}) = \frac{1}{\sqrt{3}}$.

The probability that X is within two standard deviations of its mean is 1, as the mean plus two standard deviations $1/2 + 1/\sqrt{3}$ exceeds 1 and the mean minus two standard deviations is less than 0 - since X always takes values between 0 and 1, X is always within 2 standard deviations of its mean. Similarly, it is always within 3 standard deviations of the mean.

b. We have $E(X) = 1$ and $Var(X) = 1$. $F(k) = 1 - e^{-k}$. Also note that $P(X < 0) = 0$ for an exponential distribution.

1 standard deviation: $P(0 < X < 2) = F(2) - F(0) = F(2) = 1 - e^{-2}$
2 standard deviation: $P(-1 < X < 3) = P(0 < X < 3) = F(3) - F(0) = F(3) = 1 - e^{-3}$
3 standard deviation: $P(-2 < X < 4) = P(0 < X < 4) = F(4) - F(0) = F(4) = 1 - e^{-4}$

c. If $Y \sim Expo(1/2)$, then $Y = 2X$ where $X \sim Expo(1)$, $E(Y) = 2$, $Var(Y) = 4$, $STD(Y) = 4$. In general, we note that if $Y \sim Expo(\lambda)$ then $Y = X/\lambda$ and $E(Y) = 1/\lambda$ and $STD(Y) = 1/\lambda$.

Then we can realize the following pattern: the probability that Y is n standard deviations away from its mean is $P(-(n-1)/\lambda < Y < (n+1)\lambda) = P(0 < Y < (n+1)/\lambda) = F((n+1)/\lambda) = 1 - e^{\lambda(n+1)/\lambda} = 1 - e^{n+1}$

5.1.6 problem 7

a. F(x) is continuous at its given endpoints: $F(1) = \frac{2}{\pi} arcsin(1) = \frac{2}{\pi}\frac{\pi}{2} = 1$ and $F(0) = 0$, and F(x) is differentiable between 0 and 1:

b. $F'(x) = f(x) = \frac{2}{\pi}\frac{d}{dx}(arcsin(\sqrt{x})) = \frac{2}{\pi}(\frac{1}{\sqrt{1-x}})(\frac{1}{\sqrt{x}})$

This is a valid PDF despite the discontinuities at 0 and 1 as the integral of $f(x)$ from 0 to 1 converges, and $f(x)$ is always positive - the same reason why $\frac{1}{\sqrt{x}}$ has a discontinuity at $x = 0$ but can be integrated from 0 to any positive real number.

5.1.7 problem 8

(a) Let F be the CDF of the Beta distribution with parameters $a = 3$, $b = 2$. Due to the properties of the CDF, F must be 0 for $x \leq 0$ and 1 for $x \geq 1$. For $0 < x < 1$:

$$\begin{aligned} F(x) &= \int_0^x f(t)dt = \int_0^x 12t^2(1-t)dt = 12\int_0^x t^2 dt - 12\int_0^x t^3 dt \\ &= 12\left(\frac{x^3}{3} - \frac{x^4}{4}\right) \\ &= 4x^3 - 3x^4 \end{aligned}$$

Then:

$$F(x) = \begin{cases} 0 & \text{, for } x \leq 0 \\ x^3(4-3x) & \text{, for } 0 < x < 1 \\ 1 & \text{, for } x \geq 1 \end{cases}$$

(b)

$$\begin{aligned} P(0 < X < 1/2) &= F(1/2) - F(0) = \left(\frac{1}{2}\right)^3\left(4 - \frac{3}{2}\right) - 0 \\ &= \frac{5}{16} \end{aligned}$$

(c) The mean of X can be calculated by the definition of expectation:

$$\begin{aligned} E(X) &= \int_0^1 xf(x)dx = \int_0^1 12x^3(1-x)dx \\ &= 12\int_0^1 x^3 dx - 12\int_0^1 x^4 dx = 12\left(\frac{1}{4} - \frac{1}{5}\right) \\ &= \frac{3}{5} \end{aligned}$$

By the definition of variance:

$$\text{Var}(X) = E(X^2) - (EX)^2$$

Let's calculate the second moment of X by LOTUS:

$$\begin{aligned}
E(X^2) &= \int_0^1 x^2 f(x)dx = \int_0^1 12x^4(1-x)dx \\
&= 12\int_0^1 x^4 dx - 12\int_0^1 x^5 dx = 12\left(\frac{1}{5} - \frac{1}{6}\right) \\
&= \frac{2}{5}
\end{aligned}$$

Substituting this value into the definition of variance:

$$\text{Var}(X) = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25}$$

5.2 Mixed Practice

5.2.1 problem 50

By symmetry, and by X and Y being independent and identically distributed, the chance that X should be smaller than Y should not be different than the chance that Y should be smaller than X. Furthermore, from independence, joint pdf, $g(x, y) = f_X(x)f_Y(y)$

$$\begin{aligned}
P(X < Y) &= \int_{y=-\infty}^{\infty}\int_{t=-\infty}^{y} f_X(x)f_Y(y)dtdy && (5.12)\\
&= \int_{y=-\infty}^{\infty} F(y)f(y)dy && (5.13)\\
&= \frac{1}{2} && (5.14)
\end{aligned}$$

When X and Y are not independent say X = Y + 1, (assume the existence of such X and Y), then, $P(X < Y) = 0$ and $P(Y < X) = 1$ When X and Y are not identically distributed, say $X \sim \text{Unif}(0, 1)$ and $Y \sim \text{Unif}(-1, 0)$ then, $P(X < Y) = 1$ and $P(Y < X) = 0$

5.2.2 problem 51

(a) We know that, $X^2 \leq X$ with probability 1. So $\mathbb{E}[X^2] \leq \mathbb{E}X$

$$V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{5.15}$$
$$\leq \mu - \mu^2 \tag{5.16}$$
$$\leq \frac{1}{4}\text{taking the minimum of the above quadratic} \tag{5.17}$$

(b) I have to show that $V(X) = 1/4$ leads to a unique distribution. From (a), $V(X) \leq \mu - \mu^2 \leq 1/4$ implies that, $\mu = 1/2$ Now

$$\mathbb{E}[(X - 1/2)^2] = 1/4$$

But

$$0 \geq (X - 1/2)^2 \leq 1/4$$

with probability 1. To get $\mathbb{E}[(X - 1/2)^2] = 1/4$, we need $(X - 1/2)^2 = 1/4$ with probability 1. So,

$$X = \begin{cases} 0 & \text{with prob p} \\ 1 & \text{with prob 1 - p} \end{cases} \tag{5.18}$$

Using $\mu = 1/2$ gives us, $p = 1/2$.

5.2.3 problem 52

$$\mathbb{E}[X] = \int_0^\infty x f(x) dx \tag{5.19}$$
$$= \int_0^\infty x^2 e^{-\frac{x^2}{2}} dx \tag{5.20}$$
$$= \frac{1}{2} \int_{-\infty}^\infty x^2 e^{-\frac{x^2}{2}} dx \tag{5.21}$$
$$= \frac{1}{2} \tag{5.22}$$

$$\mathbb{E}[X^2] = \int_0^\infty x^2 f(x) dx \tag{5.23}$$
$$= \int_0^\infty x^3 e^{\frac{x^2}{2}} dx \tag{5.24}$$
$$= \int_0^\infty 2ue^u du = 2 \tag{5.25}$$
$$\tag{5.26}$$

5.2.4 problem 56

$$\mathbb{E}[Z^2\Phi(z)] = \int_{-\infty}^{\infty} x^2 \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx \tag{5.27}$$

Now

$$\int_{-\infty}^{\infty} f(z)dz \int_{-\infty}^{\infty} f(z)dz$$

if $\int_{-\infty}^{\infty} f(z)dz$ is meant to be $\lim_{a\to\infty} \int_{-a}^{a} f(z)dz$ So

$$\mathbb{E}[z^2\Phi(z)] = \int x^2\Phi(x)e^{-\frac{x^2}{2}} dx \tag{5.28}$$

$$= \int x^2\Phi(-x)e^{-\frac{x^2}{2}} dx \tag{5.29}$$

$$= \int x^2(1-\Phi(x))e^{-\frac{x^2}{2}} dx \tag{5.30}$$

So

$$E[\Phi(z)z^2] = \int x^2\Phi(x)e^(- \frac{x^2}{2})dx \tag{5.31}$$

$$= \frac{1}{2}\int x^2 e^{-\frac{x^2}{2}} \tag{5.32}$$

$$= \frac{1}{2} \tag{5.33}$$

(b)

$$P(\Phi(z) \leq \frac{2}{3}) = P(z \leq \Phi^{-1}(\frac{2}{3})) = \Phi(\Phi^{-1}(\frac{2}{3}))$$

(c)

$$\frac{1}{3!}$$

5.2.5 problem 57

(a)

$$\Phi(W)(t) = P(\Phi(z)^2 \leq t) \tag{5.34}$$

$$= P(\Phi(z) \leq \sqrt{t}) \tag{5.35}$$

$$= P(z \leq \Phi^{-1}(\sqrt{t})) \tag{5.36}$$

$$= \Phi(\Phi^{-1}(\sqrt{t})) \tag{5.37}$$

$$= \sqrt{t} \tag{5.38}$$

(b) $\mathbb{E}[W] = \int_0^1 w^3 f_W(w)dw$ $\mathbb{E}[W] = \int_0^1 \Phi(z)^6\phi(z)dz$

(c) $P(X + 2Y < 2Z + 3) = P(X + 2Y - 2Z < 3)$ here, $X + 2Y - 2Z \sim N(0, \sqrt{1^2 + 2^2 + 2^2})$ So, $P(X + 2Y < 2Z + 3) = \Phi(1)$

5.2.6 problem 58

(a)

$$\mathbb{E}[Y] = \frac{1}{2} \cdot 0 + \int_0^\infty xf(x)dx \tag{5.39}$$

$$= \int_0^\infty \frac{1}{\sqrt{2\pi}} xe^{-\frac{x^2}{2}} dx \tag{5.40}$$

(b) N is the first success distribution with $p = \frac{1}{2}$ $\mathbb{E}[N] = 2$

(c)

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \Phi(y) & y \geq 0 \end{cases} \tag{5.41}$$

5.2.7 problem 59

(a) Length biased sampling

$$L_1 + L_2 + L_3 = 2\pi$$

$$\mathbb{E}[L_1] = \mathbb{E}[L_2] = \mathbb{E}[L_3] = \frac{2\pi}{3}$$

But our point is more likely to be a part of the longest arc. If there was a $\frac{1}{3}$ chance of the point being in any one of the three points then $\mathbb{E}[L] = \frac{2\pi}{3}$

(b)

$$\theta_1 = \text{Unif}(0, 2\pi)$$

$$\theta_2 = \text{Unif}(0, 2\pi)$$

$$\theta_3 = \text{Unif}(0, 2\pi)$$

$$L_1 = \min(\theta_1, \theta_2, \theta_3)$$

CDF,

$$F(y) = 1 - P(\min(\theta_1, \theta_2, \theta_3) > y) \tag{5.42}$$

$$= 1 - \frac{2\pi - y^3}{2\pi} \tag{5.43}$$

PDF,

$$f(y) = \frac{d}{dy}F(y) \tag{5.44}$$

$$= \frac{3}{2\pi}(1 - \frac{y}{2\pi})^2 \tag{5.45}$$

(c)

$$\mathbb{E}[L] = 2\mathbb{E}[L_1] \tag{5.46}$$

$$= 2\int_0^{2\pi} y\frac{3}{2\pi}(1 - \frac{y}{2\pi})^2 dy \tag{5.47}$$

$$= \frac{3}{\pi}\int_0^{2\pi} y + \frac{y^3}{4\pi^2} - \frac{y^2}{\pi}dy \tag{5.48}$$

$$= \frac{3}{\pi}[\frac{4\pi^2}{2} + \frac{1}{4\pi^2}\frac{16\pi^4}{4} - \frac{1}{\pi}\frac{8\pi^3}{3}] \tag{5.49}$$

$$= \pi \tag{5.50}$$

5.2.8 problem 61

(a)

$$I_k = \begin{cases} 1 & \text{if k arrives when fun} \\ 0 & \text{if k arrives when not fun} \end{cases} \tag{5.51}$$

Now $I = I_1 + \cdots + I_n$

$$\mathbb{E}[I] = n\mathbb{E}[I_1] \tag{5.52}$$

$$= \frac{n}{3} \tag{5.53}$$

as $P(I_k = 1) = \frac{1}{3}$
(b)

$$P(I_1 I_2) = P(I_1)P(I_2)$$

Given that Jaime and Robert are guests 1 and 2

$$P(I_1 I_2) = P(\text{both 1 and 2 arrive when fun})$$

Out of the possible 4! orderigns of Tyrion, Cersei, 1, and 2 for both 1 and 2 to arrive when fun, the following orderings are possible
Tyrion 1 2 Cersei Tyrion 2 1 Cersei Cersei 1 2 Tyrion Cersei 2 1 Tyrion

So $P(I_1 I_2) = \frac{1}{4!}4 = \frac{1}{6}$

$$P(I_1 I_2) \neq P(I_1)P(I_2)$$

(c)

We already know the answer. Conditioning on the event that 1 arrives when it's fun, the chances of 2 arriving when it's fun are higher than the unconditional probability of 2 arriving when it's fun. When we have information of 1 arriving when it's fun, we know that there's someone arriving betweenTyrion and Cersei and this forces the conditional sample space to have a skew towards having Tyrion and Cersei further apart than if we have no information about 1. This skewing of the the conditional sample space increasing chances that 2 arrives at a time when it's fun.

However, the events ARE conditionally independent. If we know the length of the interval of time between Cersei and Tyrion's arrivals L, then the probability of any other guest arriving at a fun time just becomes L. It is no longer true that any order of Tyrion, Cersei, and some specific guest is equally likely - for example, if we know Tyrion is the first guest and Cersei is the last guest, it is obvious that the only possible ordering is Tyrion Guest Cersei. Moreover, since the arrival times of other guests are independent of each other, knowing that Jaime arrives at a fun time no longer makes it more likely that Robert will arrive at a fun time - Robert still has to arrive in the interval L, as opposed to Jaime's arrival making the expected amount of fun time larger. Therefore, the probability of both Jaime and Robert arriving when it is fun, given that the amount of time between Cersei and Tyrion's arrival is L, is L^2.

5.2.9 problem 62

(a)

$$I_k = \begin{cases} 1 & \text{if k sets a low or high record} \\ 0 & \text{if k doesn't set a low or high record} \end{cases} \tag{5.54}$$

$$P(I_1) = 1$$
$$P(I_2) = 1$$
$$P(I_3) = \frac{2}{3}$$

$$P(I_4) = \frac{2}{4}$$

and so on.

Now $I = I_1 + \cdots + I_n$
$\mathbb{E}[I] = 1 + 1 + \frac{2}{3} \cdots \frac{2}{100}$

(b)

$$I_k = \begin{cases} 1 & \text{if k sets a low followed by a high} \\ 0 & \text{otherwise} \end{cases} \tag{5.55}$$

$$P(I_k) = \frac{1}{k}\frac{1}{k+1}$$

Now $I = I_1 + \cdots + I_n$
$\mathbb{E}[I] = \frac{1}{1 \cdot 2} + \cdots + \frac{1}{100 \cdot 101}$
$\mathbb{E}[I] = 1 - \frac{1}{101}$

(c)

$$P(N > n) = P(\text{all of 2 to n + 1 fall short of 1}) \tag{5.56}$$
$$= \frac{1}{n+1} \tag{5.57}$$

$$P(N = n) = P(N > n-1) - P(N > n) \tag{5.58}$$
$$= \frac{1}{n(n+1)} \tag{5.59}$$

(d)

$$\mathbb{E}[N] = \sum_{i=1}^{\infty} iP(N = i) \tag{5.60}$$
$$= \sum_{1}^{\infty} \frac{1}{i+1} \tag{5.61}$$

is unbounded.

5.3 Exponential

5.3.1 problem 37

a. We need to find the value of t such that $F(t) = 1/2$ - this will indicate that there is a 1/2 chance that the particle has decayed before time t.

$1 - e^{-\lambda t} = 1/2$ implies $ln(2) = \lambda t$, so $t = ln(2)/\lambda$

b. We need to compute $P(t < T < t+\epsilon | T > t) = P(t < T < t+\epsilon)/P(T > t)$. This is $\frac{(1-e^{-\lambda(t+\epsilon)})-(1-e^{-\lambda t})}{e^{-\lambda t}} = 1 - e^{-\lambda\epsilon}$. Using the approximation given in the hint and the assumption that ϵ is small enough that $\epsilon\lambda \approx 0$, this is about $1 - (1 - \epsilon\lambda) = \epsilon\lambda$.

c. $P(L > t) = P(T_1 > t)P(T_2 > t)...P(T_n > t) = e^{-n\lambda t}$, so $L \sim Expo(n\lambda)$. Therefore, if $X \sim Expo(1)$, we have $L = X/n\lambda$. Then since $E(X) = 1$ and $Var(X) = 1$, we can get $E(L) = 1/(n\lambda)$ and $Var(L) = 1/(n^2\lambda^2)$

d. M must be equal to the sum of $D_1 + D_2 + D_3 + ... + D_n$, where D_i is the amount of time between the $i-1$th and ith decay event. We observe that D_i must then be the minimum of $n - i + 1$ $Expo(\lambda)$ variables - for example, D_1 is the first particle to decay out of n particles, D_2 is the first particle to decay out of the remaining n-1 particles, etc. Since $Expo$ is memoryless, D_{i+1} is independent of D_i as the amount of time it takes for the next particle to decay is not affected by the amount of time it took the previous particle to decay. Therefore, $D_i \sim Expo((n - i + 1)\lambda)$.

Then $E(M) = E(D_1) + E(D_2) + ... + E(D_n) = \frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i}$

Now we calculate the variance of M, recalling that the variance of the sum of independent r.v.s is equal to the sum of variances
$\text{Var}(M) = \text{Var}(D_1 + D_2 + \cdots + D_n)$
$\text{Var}(M) = \text{Var}(D_1) + \text{Var}(D_2) + \cdots + \text{Var}(D_n)$
Since $\text{Var}(D_i) = 1/[(n - i + 1)^2\lambda^2]$,

$$\text{Var}(M) = \sum_{i=1}^{n}\frac{1}{(n-i+1)^2\lambda^2} = \frac{1}{\lambda^2}\sum_{i=1}^{n}\frac{1}{i^2}$$

5.3.2 problem 39

Let O_1, O_2, O_3 be the offers received, all distributed as $Expo(1/12000)$. We want to find $E(max(O_1, O_2, O_3))$. Imagine ordering the offers as lowest, middle, and highest price. Let D_1 be the lowest price, let D_2 be how much more the middle price is than the lowest price, and D_3 be how much more the highest price is than the middle price. Then D_1 is the minimum of 3 $Expo(1/12000)$ variables (since it is just the minimum of the 3 offers), D_2 is the minimum of 2 $Expo(1/12000)$ variables (by the memoryless property, given that all other offers are at least the lowest offer, the likelihood that any offer is at least k more than the lowest is just the probability the offer was at least k in the first place), and similarly, D_3 is the minimum of 1 $Expo(1/12000)$ variable. In addition, we realize that $D_1 + D_2 + D_3$ must equal the highest offer.

Then $E(D_1 + D_2 + D_3) = \frac{1}{3\lambda} + \frac{1}{2\lambda} + \frac{1}{\lambda} = 4000 + 6000 + 12000 = 22000$.

5.3.3 problem 45

Let N be the number of emails received within the first 0.1 hours. Then
$P(T > 0.1) = 1 - P(T < 0.1) = 1 - P(N \geq 3) = 1-(1-P(N = 0)-P(N = 1)-P(N = 2)) = P(N = 0)+P(N = 1)+P(N = 2)$

That is, to find the probability that it takes longer than 0.1 hours for 3 emails to arrive, we find the probability it takes less than 0.1 hours for 3 emails to arrive. To do this, we realize that this is equivalent to at least 3 emails arriving in the first 0.1 hours. And to find that probability, we realize we can find the probabilities of exactly 0, 1, or 2 emails arriving in the first 0.1 hours.

Next, we note that $N \sim Pois(0.1 * \lambda) = Pois(2)$, per the definition of a Poisson process. Then we get

$P(T > 0.1) = P(N = 0)+P(N = 1)+P(N = 2) = e^{-2}+2e^{-2}+4e^{-2}/2 = 5e^{-2} \approx 0.68$ $\blacksquare$

5.4 Normal

5.4.1 problem 20

(a)

$$\begin{aligned}\Phi(z) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{0} e^{-x^2/2}dx + \int_{0}^{z} e^{-x^2/2}dx\right)\end{aligned}$$

We know from the Normal PDF that $\int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}$. Since the integrand is an even function, the integral from $-\infty$ to 0 is half this value

$$\Phi(z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\int_{0}^{z} e^{-x^2/2}dx$$

By applying the variable substitution $u = x/\sqrt{2}$ in the integral, we obtain $du = dx/\sqrt{2}$. The lower and upper limits of integration become 0 and $z/\sqrt{2}$, respectively.

$$\begin{aligned}\Phi(z) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\int_{0}^{z/\sqrt{2}} e^{-u^2}\sqrt{2}\,du \\ &= \frac{1}{2} + \frac{1}{2}\frac{2}{\sqrt{\pi}}\int_{0}^{z/\sqrt{2}} e^{-u^2}du \\ \Phi(z) &= \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\end{aligned}$$

(b)

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_{0}^{z} e^{-x^2}dx$$

The integrand is an even function, so the integrals from 0 to z and from $-z$ to 0 are the same

$$\begin{aligned}\operatorname{erf}(z) &= \frac{2}{\sqrt{\pi}}\int_{-z}^{0} e^{-x^2}dx \\ &= -\frac{2}{\sqrt{\pi}}\int_{0}^{-z} e^{-x^2}dx \\ \operatorname{erf}(-z) &= -\operatorname{erf}(z)\end{aligned}$$

5.4.2 problem 26

a. Let $T_w \sim N(w, \sigma^2)$ be the time it takes Walter to arrive, $T_c \sim N(c, 4\sigma^2)$ be the time it takes Carl to arrive. We have $-T_w \sim N(-c, \sigma^2)$ since flipping the sign of the rv flips the sign of the expectation but does not change the variance. Then $T_c - T_w \sim N(c - w, 5\sigma^2)$ per the important fact given in the problem. If $Z \sim N(0, 1)$, then $T_c - T_w = \sqrt{5}\sigma Z + c - w$.

For Carl to arrive first, we require $T_c - T_w < 0$ (the time it takes Carl is less than the time it takes Walter). Let us find this probability:

$$P(T_c - T_w < 0) = P(Z < \frac{w-c}{\sigma\sqrt{5}}) = \Phi(\frac{w-c}{\sigma\sqrt{5}})$$

b. If Carl has a greater than 1/2 chance of arriving first, then $\Phi(\frac{w-c}{\sigma\sqrt{5}}) > 1/2$. Since Φ is an increasing function and equals 1/2 when its input is 0, this implies we need $\frac{w-c}{\sigma\sqrt{5}} > 0$, which in turn implies $c < w$. So, as long as Carl's car lets him be faster on average than Walter's walking, Carl has a better than 1/2 chance of arriving first.

c. To make it to the meeting at time, either individual needs to make sure the amount of time they take to arrive is less than $w + 10$.

$$P(T_c < w + 10) = P(2\sigma Z + c < w + 10) = \Phi(\frac{w + 10 - c}{2\sigma})$$

$$P(T_w < w + 10) = P(\sigma Z + w < w + 10) = \Phi(\frac{10}{\sigma})$$

Since Φ is an increasing function, if we want Carl to have a greater chance than Walter to make it on time, then we require $\frac{w+10-c}{2\sigma} > \frac{10}{\sigma}$. This then implies that we need $w > c + 10$.

5.4.3 problem 28

Starting from the fact given in the problem

$$P(|Y - \mu| < 1.96\sigma) \approx 0.95$$

$$P(-1.96\sigma < Y - \mu < 1.96\sigma) \approx 0.95$$

$$P(Y - 1.96\sigma < \mu < Y + 1.96\sigma) \approx 0.95$$

$$P(\mu \in (Y - 1.96\sigma, Y + 1.96\sigma)) \approx 0.95$$

Therefore, the random interval that contains μ about 95% of the time is $(Y - 1.96\sigma, Y + 1.96\sigma)$.

5.4.4 problem 35

Let $g(Z) = max(Z - c, 0)$. We have $g(Z) = 0$ for $Z < c$ and $g(Z) = Z - c$ for $Z > c$.

Then

$$E(g(Z)) = \int_{-\infty}^{\infty} g(k)\varphi(k)dk = \int_{c}^{\infty} (k - c)\varphi(k)dk$$

since the expression inside the integral is 0 for $k < c$. Next we have

$$\int_{c}^{\infty} (k-c)\varphi(k)dk = \int_{c}^{\infty} \frac{ke^{-k^2/2}}{\sqrt{2\pi}}dk - \int_{c}^{\infty} c\varphi(k)dk = \frac{e^{-c^2/2}}{\sqrt{2\pi}} - c\int_{-\infty}^{-c} \varphi(k) = \varphi(c) - c\Phi(c)$$

with the first equality following from splitting the integral via subtraction and expanding out φ for the left integral, the second equality comes from observing that the antiderivative of $k\varphi(k)$ is $-\varphi(k)$ and that we can change the limits of the right integral due to the symmetry of tail areas of the curve of φ, and the last equality comes from applying the definitions of the PDF and CDF of the standard normal distribution.

5.5 Uniform and Universality

5.5.1 problem 13

Recall from problem 12 that the CDF of Y, the length of the longer piece, is $F(k) = 2k - 1$.

a. Let us find the CDF of X/Y.

$$P(X/Y < k) = P(\frac{1-Y}{Y} < k) = P(Y > \frac{1}{k+1}) = 1 - (\frac{2}{k+1} - 1) = \frac{2k}{k+1}$$

To find the PDF, we derive the CDF using the quotient rule:

$$\frac{d}{dk}(\frac{2k}{k+1}) = 2(k+1)^{-2}$$

b. Note that X/Y is minimized at when X is 0 and Y is 1 and maximized at 1 when X and Y are both 1/2. So, to find $E(X/Y)$, we must find $\int_0^1 2k(k+1)^{-2}dk$. This can be done with integration by parts, factoring out the constant 2, with $u = k$, $dv = (k+1)^{-2}dk$, $du = 1$, and $v = -(k+1)^{-1}$:

$$\int_0^1 2k(k+1)^{-2}dk = 2((\frac{-k}{k+1})|_0^1 - \int_0^1 -(k+1)^{-1}dk) = 2(-\frac{1}{2}+ln(2)) = 2ln(2)-1$$

c. Following similar steps as in part A, the CDF of Y/X is $P(\frac{Y}{1-Y} < k) = P(Y < \frac{k}{k+1}) = \frac{k-1}{k+1}$. Then the PDF using the quotient rule is $2(k+1)^{-2}$, the same as the PDF for X/Y!

Then, we need to evaluate the same integral as in part b to find $E(Y/X)$, but now with the limits set from 1 to infinity, since Y/X is minimized when X=Y and maximized when Y=1 and X=0:

$$\int_1^\infty 2k(k+1)^{-2}dk = 2((\frac{-k}{k+1})|_1^\infty - \int_1^\infty -(k+1)^{-1}dk) = 2((-1/2)+ln(\infty)-ln(2)) = \infty$$

5.5.2 problem 14

Let F be the CDF and f the PDF of X.

The maximum is less than or equal to some x if and only if all U_i are less than or equal to x.

$$P(X \leq x) = P(U_1 \leq x, U_2 \leq x, \ldots, U_n \leq x) \text{ , for } 0 < x < 1$$

Since the r.v.s U_i are independent, the events $\{U_i \leq x\}$ are independent as well.

$$P(X \leq x) = P(U_1 \leq x)P(U_2 \leq x) \dots P(U_n \leq x)$$

The CDF of the standard uniform is equal to x for $0 < x < 1$. Thus:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{, for } x \leq 0 \\ x^n & \text{, for } 0 < x < 1 \\ 1 & \text{, for } x \geq 1 \end{cases}$$

The PDF is equal to the derivative of the CDF

$$f(x) = F'(x) = nx^{n-1} \text{ , for } 0 < x < 1$$

By the definition of expectation for continuous r.v.s

$$E(X) = \int_0^1 xf(x)dx = \int_0^1 nx^n dx = \frac{n}{n+1}$$

Chapter 6

Moments

6.1 Means, Medians, Modes, Moments

6.1.1 problem 1

The median is $(a+b)/2$, the solution to $P(X < k) = \frac{k-a}{b-a} = 1/2$.

The mode is every real number between a and b, since the PDF is a constant.

6.1.2 problem 2

Let k be the median, therefore $P(X \geq k) \geq \frac{1}{2}$ and $P(X \leq k) \geq \frac{1}{2}$, therefore $P(X \leq k) \leq \frac{1}{2}$.
Hence,

$$\begin{aligned} P(X \leq k) &= \frac{1}{2} \\ F(k) &= \frac{1}{2} \\ k &= \frac{\ln(2)}{\lambda} \end{aligned}$$

where F is the CDF of X.

Let c be the mode, f be the PDF of X, since $f(x)$ is decreasing for all $x \geq 0$ ($f'(x) = -\lambda^2 e^{-\lambda x}$) or is 0 otherwise, $f(x)$ has its maxima when $x = 0$, hence, the mode is 0.