Introduction to Probability by Blitzstein: Exercise Solutions

Contents

Chapter 1

Probability and counting

1.1 Counting

Exercise 1
How many ways are there to permute the letters in the word MISSISSIPPI?

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

Exercise 2
Solve the following

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

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

Solution.

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

2. **The solution is** $10^7 - 10^4 = 9990000$

Exercise 4
A round-robin tournament is being held with n tennis players; this means that every player will play against every other player exactly once.

1. How many possible outcomes are there for the tournament (the outcome lists out who won and who lost for each game)?

2. How many games are played in total?

Solution.

1. **1) There are** $2^{\frac{n(n-1)}{2}}$ **possible outcomes.**

2. **There are** $[n(n-1)]/2$ **games being played.**

Exercise 5
A knock-out tournament is being held with 2^n tennis players. This means that for each round, the winners move on to the next round and the losers are eliminated, until only one person remains. For example, if initially there are $2^4 = 16$ players, then there are 8 games in the first round, then the 8 winners move on to round 2, then the 4 winners move on to round 3, then the 2 winners move on to round 4, the winner of which is declared the winner of the tournament. (There are various systems

for determining who plays whom within a round, but these do not matter for this problem.)

1. How many rounds are there?
2. Count how many games in total are played, by adding up the numbers of games played in each round.
3. Count how many games in total are played, this time by directly thinking about it without doing almost any calculation.

Hint: How many players need to be eliminated?

Solution.

1. **n**
2. $2^n/2 + 2^n/4 + \cdots + 2^n/2^n$ **or according to the geometric series** $1(2^n - 1)/(2 - 1) = 2^n - 1$
3. **Since in each match you eliminate 1 person you need to do** $2^n - 1$

Exercise 6
There are 20 people at a chess club on a certain day. They each find opponents and start playing. How many possibilities are there for how they are matched up, assuming that in each game it does matter who has the white pieces (in a chess game, one player has the white pieces and the other player has the black pieces)?

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

Exercise 10
To fulfill the requirements for a certain degree, a student can choose to take any 7 out of a list of 20 courses, with the constraint that at least 1 of the 7 courses must be a statistics course. Suppose that 5 of the 20 courses are statistics courses.
(a)How many choices are there for which 7 courses to take?
(b)Explain intuitively why the answer to (a) is not $\binom{5}{1} \times \binom{19}{6}$

Solution.
(a) $\binom{5}{1} \times \binom{15}{6} + \binom{5}{2} \times \binom{15}{5} + \binom{5}{3} \times \binom{15}{4} + \binom{5}{3} \times \binom{15}{4} + \binom{5}{4} \times \binom{15}{3} + \binom{5}{5} \times \binom{15}{2}$
(b) Because this means from the 5 statistics courses chose only 1 and then form the remaining 19 courses left chose 6. The second binomial coefficient might or might not add the remaining possibilities of statistic courses.

Exercise 14
You are ordering two pizzas. A pizza can be small, medium, large, or extra large, with any combination of 8 possible toppings (getting no toppings is allowed, as is getting all 8). How many possibilities are there for your two pizzas?

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

1.2 Story proofs

Exercise 15
Give a story proof that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

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

1.3 Naive definition of probability

Exercise 24
A certain family has 6 children, consisting of 3 boys and 3 girls. Assuming that all birth orders are equally likely, what is the probability that the 3 eldest children are the 3 girls?

Solution.
In three methods(just to test my understanding):
As individual probabilities,

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

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

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

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

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

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

Exercise 25
A city with 6 districts has 6 robberies in a particular week. Assume the robberies are located randomly, with all possibilities for which robbery occurred where equally likely. What is the probability that some district had more than 1 robbery?

Solution.
The probability is

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

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

Exercise 26
A survey is being conducted in a city with 1 million residents. It would be far too expensive to survey all of the residents, so a random sample of size 1000 is chosen (in practice, there are many challenges with sampling, such as obtaining a complete list of everyone in the city, and dealing with people who refuse to participate). The survey is conducted by choosing people one at a time, with replacement

and with equal probabilities.
(a) Explain how sampling with vs. without replacement here relates to the birthday problem.
(b) Find the probability that at least one person will get chosen more than once.

Solution.
(a)In this problem we use the same process as in the birthday problem. A sampling with replacement in the denominator. In the birthday problem the sampling with replacement corresponds to the ways in which we can choose birth dates. And in the denominator we use sampling without replacement, but in this case we use the binomial coefficient instead of n permutation k.
(b)We will get it using the negation of the statement

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

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

Exercise 28
A college has 10 time slots for its courses, and blithely assigns courses to completely random time slots, independently. The college offers exactly 3 statistics courses. What is the probability that 2 or more of the statistics courses are in the same time slot?

Solution.

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

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

Exercise 29
For each part, decide whether the blank should be filled in with $=, <, or >$, and give a clear explanation.
(a) (probability that the total after rolling 4 fair dice is 21)>(probability that the total after rolling 4 fair dice is 22)
(b) (probability that a random 2-letter word is a palindrome)= (probability that a random 3-letter word is a palindrome)

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

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

they can each be written in

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

ways respectively and therefore the probability of getting 21 is

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

and 22 can be written as

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

and these can be written in

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

ways respectively and therefore the probability of getting 22 is

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

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

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

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

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

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

Exercise 30
With definitions as in the previous problem, find the probability that a random n-letter word is a palindrome for n = 7 and for n = 8.

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

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

and

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

are the solutions.

Exercise 31
Elk dwell in a certain forest. There are N elk, of which a simple random sample of size n are captured and tagged ("simple random sample" means that all $\binom{N}{n}$ sets of n elk are equally likely). The captured elk are returned to the population, and then a new sample is drawn, this time with size m. This is an important method that is widely used in ecology, known as capture-recapture. What is the probability that exactly k of the m elk in the new sample were previously tagged? (Assume that an elk that was captured before doesn't become more or less likely to be captured again.)

Solution.

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

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

Exercise 34
A random 5-card poker hand is dealt from a standard deck of cards. Find the probability of each of the following possibilities (in terms of binomial coefficients).
(a) A flush (all 5 cards being of the same suit; do not count a royal flush, which is a flush with an ace, king, queen, jack, and 10).

(b) Two pair (e.g., two 3's, two 7's, and an ace).

Solution. (a)

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

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

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

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

Exercise 39
An organization with 2n people consists of n married couples. A committee of size k is selected, with all possibilities equally likely. Find the probability that there are exactly j married couples within the committee.

Solution.

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

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

Exercise 41
Each of n balls is independently placed into one of n boxes, with all boxes equally likely. What is the probability that exactly one box is empty?

Solution.
Let's do the contrary to this problem so we need to search for $1-$ probability of each ball landing in a different box. Therefore the result is

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

Exercise 42
A norepeatword is a sequence of at least one (and possibly all) of the usual 26 letters a,b,c,...,z, with repetitions not allowed. For example, "course" is a norepeatword, but "statistics" is not. Order matters, e.g., "course" is not the same as "source". A norepeatword is chosen randomly, with all norepeatwords equally likely. Show that the probability that it uses all 26 letters is very close to 1/e.

Solution.

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

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

For the $1/e$ part

$$\frac{26!}{\sum_{i=1}^{26}\binom{26}{i}\cdot i!} = \frac{26!}{\sum_{i=1}^{26}\frac{26!}{\cancel{i!}\cdot(26-i)!}\cdot\cancel{i!}} =$$

$$\frac{\cancel{26!}}{\cancel{26!}\cdot\sum_{i=1}^{26}\frac{1}{(26-i)!}} =$$

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

1.4 Axioms of probability

Exercise 44
Let A and B be events. The difference $B - A$ is defined to be the set of all elements of B that are not in A. Show that if $A \subseteq B$, then

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

directly using the axioms of probability.

Solution.

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

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

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

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

□

Exercise 46
Let $A_1, A_2, ..., A_n$ be events. Let Bk be the event exactly k of the A_i occur, and C_k be the event that at least k of the A_i occur, for $0 \leq k \leq n$. Find a simple expression for $P(B_k)$ in terms of $P(C_k)$ and $P(C_{k+1})$.

Solution.
The expression would be:

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

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

Exercise 47
Events A and B are independent if $P(A \cap B) = P(A)P(B)$ (independence is explored in detail in the next chapter).
(a) Give an example of independent events A and B in a finite sample space S (with neither equal to $\emptyset$ or S), and illustrate it with a Pebble World diagram.
(b) Consider the experiment of picking a random point in the rectangle

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

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

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

Solution.
(a) S is 2 dice throws (dice-1,dice-2). A is the event of getting 2 in dice-1. B is the event of getting 4 in dice-2.
(b) They are independent if they do not share a value of x or y at any point; They are dependent if they do share values. (c) We know from the properties of probability that:

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

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

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

We also know that:

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

so:

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

1.5 Inclusion-exclusion

Exercise 50
A card player is dealt a 13-card hand from a well-shuffled, standard deck of cards. What is the probability that the hand is void in at least one suit ("void in a suit" means having no cards of that suit)?

Solution.
We will do it using the union:

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

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

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

Exercise 53
Fred needs to choose a password for a certain website. Assume that he will choose an 8-character password, and that the legal characters are the lowercase letters a, b, c, ..., z, the uppercase letters A, B, C, ..., Z, and the numbers 0, 1, ..., 9.
(a) How many possibilities are there if he is required to have at least one lowercase letter in his password?
(b) How many possibilities are there if he is required to have at least one lowercase letter and at least one uppercase letter in his password?
(c) How many possibilities are there if he is required to have at least one lowercase letter, at least one uppercase letter, and at least one number in his password?

Solution.
(a)P(At least one letter) =
Incorrect version:

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

Correct version:

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

(b) We will get the number of possibilities of this event using the contrary 1(in this case 62^8) - the possibilities of the password not having any letter.

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

Explanation: The sum of the lowercase plus the digits is 36 same with uppercase and digits in both this cases we might have the case in which everything in digits so as we are subtracting it two time we need to add it once (that's where 10^8 comes from).
(c) In the same way we did the last one using inclusion exclusion and the 1 - trick.

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

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

Exercise 55
A club consists of 10 seniors, 12 juniors, and 15 sophomores. An organizing committee of size 5 is chosen randomly (with all subsets of size 5 equally likely).
(a) Find the probability that there are exactly 3 sophomores in the committee.
(b) Find the probability that the committee has at least one representative from each of the senior, junior, and sophomore classes.

Solution.
(a)

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

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

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

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

1.6 Mixed practice

Exercise 58
A widget inspector inspects 12 widgets and finds that exactly 3 are defective. Unfortunately, the widgets then get all mixed up and the inspector has to find the 3 defective widgets again by testing widgets one by one.
(a) Find the probability that the inspector will now have to test at least 9 widgets.

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

Solution.
(a)

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

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

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

Explanation: Same as above

Chapter 2

Conditional probability

2.1 Conditioning on evidence

2.2 Independence and conditional independence

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

1. (a) What is your probability of winning the first game?
2. (b) Congratulations: you won the first game! Given this information, what is the prob- ability that you will also win the second game (assume that, given the skill level of your opponent, the outcomes of the games are independent)?
3. (c) Explain the distinction between assuming that the outcomes of the games are in-dependent and assuming that they are conditionally independent given the opponent's skill level. Which of these assumptions seems more reasonable, and why?

Solution.

1.

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

$$= (.3)(1/3) + (.5)(1/3) + (.9)(1/3) = 0.566667 = \frac{17}{30}$$

2.

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

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

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

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

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

Therefore:

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

Other solution

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

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

Intermediate step:

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

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

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

Continue:

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

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

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

Exercise 36
Hi again :)

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

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

2. (b) Show that if A and B are independent and $C = A \cup B$, then A and B are conditionally dependent given C (as long as $P(A \cap B) > 0$ and $P(A \cup B) < 1$), with $P(A|B, C) < P(A|C)$. This phenomenon is known as Berkson's paradox, especially in the context of admissions to a school, hospital, etc.

Solution.

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

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

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

And it is important to note that:

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

and:

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

Then:

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

□

Exercise 37
Two different diseases cause a certain weird symptom; anyone who has either or both of these diseases will experience the symptom. Let D_1 be the event of having the first disease, D_2 be the event of having the second disease, and W be the event of having the weird symptom. Suppose that D_1 and D_2 are independent with $P(D_j) = p_j$, and that a person with neither of these diseases will have the weird symptom with probability w_0. Let $q_j = 1 - p_j$,and assume that $0 < p_j < 1$.

1. (a) Find $P(W)$.
2. (b) Find $P(D_1|W)$,$P(D_2|W)$, and $P(D_1, D_2|W)$.
3. (c) Determine algebraically whether or not D_1 and D_2 are conditionally independent given W.
4. (d) Suppose for this part only that $w_0 = 0$. Give a clear, convincing intuitive explanation in words of whether D_1 and D_2 are conditionally independent given W.

Solution.

1.

$$P(W) = 1 - P(D_1^C, D_2^C, w_0^C)$$
$$= 1 - (P(D_1^C) \cdot P(D_2^C) \cdot P(W_0^C))$$
$$= 1 - (q_1 \cdot q_2 \cdot (1 - w_0))$$
$$= 1 - ((q_1q_2) - (q_1q_2 \cdot w_0))$$
$$= 1 - (q_1q_2) + (q_1q_2w_0)$$

Other version, let's say that the event of not having both symptoms is T

$$P(W) = P(D_1) + P(D_2) - P(D_1 \cap D_2) + P(W \cap T)$$
$$= P(D_1) + P(D_2) - P(D_1 \cap D_2) + P(W|T)P(T)$$
$$= p_1 + p_2 - (p_1p_2) + P(W|T)P(T)$$
$$= p_1 + p_2 - (p_1p_2) + P(T)w_0$$

Notice this is the same as above, since:

$$P(W) = 1 - P(D_1^C, D_2^C, w_0^C)$$

$$
\begin{aligned}
&= 1 - (q_1 q_2) + (q_1 q_2 w_0) \\
&= 1 - ((1 - p_1)(1 - p_2)) + ((1 - p_1)(1 - p_2)(w_0)) \\
&= 1 - (1 - p_2 - p_1 + p_1 p_2) + (1 - p_2 - p_1 + p_1 p_2) w_0 \\
&= 1 - 1 + p_2 + p_1 - p_1 p_2 + (1 - p_2 - p_1 + p_1 p_2) w_0 \\
&= p_2 + p_1 - p_1 p_2 + (1 - p_2 - p_1 + p_1 p_2) w_0 \\
&= p_1 + p_2 - (p_1 p_2) + (1 - p_2 - p_1 + p_1 p_2) w_0 \\
&= p_1 + p_2 - (p_1 p_2) + (1 - p_2 - p_1 + p_1 p_2) w_0 \\
&= p_1 + p_2 - (p_1 p_2) + P(T) w_0
\end{aligned}
$$

2.

$$
\begin{aligned}
P(D_1|W) &= \frac{P(W|D_1)P(D_1)}{P(W)} \\
&= \frac{1 \cdot p_1}{1 - (q_1 q_2) + (q_1 q_2 w_0)} \\
&= \frac{p_1}{1 - (q_1 q_2) + (q_1 q_2 w_0)} \\
P(D_2|W) &= \frac{P(W|D_2)P(D_2)}{P(W)} \\
&= \frac{1 \cdot p_2}{1 - (q_1 q_2) + (q_1 q_2 w_0)} \\
&= \frac{1 \cdot p_2}{1 - (q_1 q_2) + (q_1 q_2 w_0)} \\
P(D_1, D_2|W) &= \frac{P(W|D_1, D_2)P(D_1, D_2)}{P(W)} \\
&= \frac{1 \cdot p_1 p_2}{1 - (q_1 q_2) + (q_1 q_2 w_0)} \\
&= \frac{p_1 p_2}{1 - (q_1 q_2) + (q_1 q_2 w_0)}
\end{aligned}
$$

3. We know from Definition 2.5.7 that: "Events A and B are said to be conditionally independent given E if $P(A \cap B|E) = P(A|E)P(B|E)$" and we have:

$$
P(D_1 \cap D_2|W) = \frac{P(W|D_1 \cap D_2)P(D_1 \cap D_2)}{P(W)}
$$

$$
= \frac{P(W|D_1 \cap D_2)P(D_1)P(D_2)}{P(W)}
$$

Since you automatically have the weird symptoms once you have both diseases

$$
= \frac{P(W)P(D_1)P(D_2)}{P(W)}
$$

$$
= P(D_1)P(D_2) \neq P(D_1|W)P(D_2|W)
$$

Therefore it is not the case that they are conditionally independent

4. Taking into account that $w_0 = 0$ means that a person without neither of the diseases will not have the weird symptom then if you have the weird symptom you must either have D1 or D2. If you don't have D2 but you have the weird symptom you have to have D1 and the same situation with D1, and if you have D2 and the weird symptom there is a chance that you have D1 same with D2.

Exercise 38
We want to design a spam filter for email. As described in Exercise 1, a major strategy is to find phrases that are much more likely to appear in a spam email than in a non- spam email. In that exercise, we only consider one such phrase: "free money". More realistically, suppose that we have created a list of 100 words or phrases that are much more likely to be used in spam than in non-spam. Let Wj be the event that an email contains the jth word or phrase on the list. Let

$$p = P(spam), p_j = P(W_j|spam), r_j = P(W_j|notspam),$$

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

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

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

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

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

Solution. We will use the following notation:

$$1 - p_j = q_j$$

$$1 - r_j = s_j$$

$$1 - P(spam) = 1 - p = q = P(notSpam)$$

$$E = (W_1^c, ..., W_{22}^c, W_{23}, W_{24}^c, ..., W_{63}^c, W_{64}, W_{65}, W_{66}^c, ..., W_{100}^c)$$

$$x = (q_1 \cdot \ldots q_{22} \cdot p_{23} \cdot q_{24} \cdot \cdots \cdot q_{63} \cdot p_{64} \cdot p_{64} \cdot p_{65} \cdot q_{66} \cdot \ldots q_{100} = P(E|spam)$$

$$y = (s_1 \cdot \ldots s_{22} \cdot r_{23} \cdot s_{24} \cdot \cdots \cdot s_{63} \cdot r_{64} \cdot r_{64} \cdot s_{65} \cdot s_{66} \cdot \ldots s_{100}) = P(E|notSpam)$$

Now for the actual problem, using Bayes rule we have:

$$P(spam|E) = \frac{P(E|spam)P(spam)}{P(E)}$$

$$= \frac{P(E|spam)P(spam)}{P(E|spam)P(spam) + P(E|notSpam)P(notSpam)}$$

Substituting:

$$= \frac{x(p)}{x(p) + y(q)}$$

2.3 Monty Hall

Exercise 39
Hi :)

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

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

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

Solution.

1. Let C_i be the event of the car being in the ith door ($1 \leq i \leq 7$) and assume without loss of generality door 1 and we switch, then:

$$P(getCar) = \sum_{i=1}^{7} P(getCar|C_i) \cdot P(C_i)$$

$$= (0 \cdot \frac{1}{7}) + (\frac{1}{3} \cdot \frac{1}{7}) + (\frac{1}{3} \cdot \frac{1}{7}) + (\frac{1}{3} \cdot \frac{1}{7}) + (\frac{1}{3} \cdot \frac{1}{7}) + (\frac{1}{3} \cdot \frac{1}{7}) + (\frac{1}{3} \cdot \frac{1}{7}) = \frac{2}{7}$$

   Explanation: We know Monty is going to choose 3 doors so $P(getCar|C_i)$ when $2 \leq i$ is going to be affected we are assuming that the Car is in this ith door but we have to take into account that when we switch we are going to fall in one of the 3 remaining doors so we get the car 1/3 of the time. Now you also have to remember that this will be applied to the 6 other doors, why? Because Monty will open 3 doors from the 6 doors with goats.

   Now for the non switch case we have:

$$P(getCar) = \sum_{i=1}^{7} P(getCar|C_i) \cdot P(C_i)$$

$$= (1 \cdot \frac{1}{7}) + (0 \cdot \frac{1}{7}) + (0 \cdot \frac{1}{7}) + (0 \cdot \frac{1}{7}) + (0 \cdot \frac{1}{7}) + (0 \cdot \frac{1}{7}) + (0 \cdot \frac{1}{7}) = \frac{1}{7}$$

   Therefore switching is better than not switching.

2. *Same notation as above.*

   Assuming without loss of generality that we choose door 1 and we have a switching strategy:

$$P(getCar) = \sum_{i=1}^{n} P(getCar|C_i) \cdot P(C_i)$$

$$= \sum_{i=1}^{n} P(getCar|C_i) \cdot \frac{1}{n} = (0 \cdot \frac{1}{n}) + \sum_{i=2}^{n} (\frac{1}{n-m-1} \cdot \frac{1}{n}) = n-1 \cdot \frac{1}{n-m-1} \cdot \frac{1}{n} = \frac{n-1}{n(n-m-1)}$$

   On the other hand assuming we do not switch:

$$P(getCar) = \sum_{i=1}^{n} P(getCar|C_i) \cdot P(C_i)$$

$$= \sum_{i=1}^{n} P(getCar|C_i) \cdot \frac{1}{n} = (1 \cdot \frac{1}{n}) + \sum_{x=1}^{n-1}(0 \cdot \frac{1}{n}) = \frac{1}{n}$$

Exercise 43

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

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

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

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

Solution.

1. Assume we chose door 1 and not switch

   If you don't switch the probability for getting the car is 1/2 because we know that we did not land on the Goat (because Monty would not reveal it otherwise) therefore there are two possibilities either we landed on the Goat or we landed on the Car which is the only favorable option.

   If you switch since we know that we did not land on the Goat (because Monty would not reveal it otherwise) we have a 1/2 chance of switching to the car.

2. Assumptions: we will choose door 1. Notation: Goat Revealed = GR, Computer Revealed = PR, Car Revealed = CR, cC = door 1 is a Car, cG = door 1 is a Goat

$$P(cC|PR) = \frac{P(PR|cC)P(cC)}{P(PR)} = \frac{q \cdot \frac{1}{3}}{P(PR)} = \frac{q}{3 \cdot P(PR)}$$

$$P(cG|PR) = \frac{P(PR|c)P(cG)}{P(PR)} = \frac{p \cdot \frac{1}{3}}{P(PR)} = \frac{p}{3 \cdot P(PR)}$$

   Since we cannot select the computer we know that:

$$P(cC|PR) + P(cG|PR) = 1$$

   Therefore:

$$\frac{q}{3 \cdot P(PR)} + \frac{p}{3 \cdot P(PR)} = 1$$

$$(p-1) + p = 3 \cdot P(PR)$$

$$P(PR) = \frac{1}{3}$$

So:

$$q + p = 1$$

in other words:

$$(1 - p) + p = 1$$

In conclusion we should switch if $p > \frac{1}{2}$ and stay otherwise.

Exercise 45
Monty Hall is trying out a new version of his game. In this version, instead of there always being 1 car and 2 goats, the prizes behind the doors are generated independently, with each door having probability p of having a car and $q = 1 - p$ of having a goat. In detail: There are three doors, behind each of which there is one prize: either a car or a goat. For each door, there is probability p that there is a car behind it and $q = 1 - p$ that there is a goat, independent of the other doors.
The contestant chooses a door. Monty, who knows the contents of each door, then opens one of the two remaining doors. In choosing which door to open, Monty will always reveal a goat if possible. If both of the remaining doors have the same kind of prize, Monty chooses randomly (with equal probabilities). After opening a door, Monty offers the contestant the option of switching to the other unopened door.
The contestant decides in advance to use the following strategy: first choose door 1. Then, after Monty opens a door, switch to the other unopened door.

1. (a) Find the unconditional probability that the contestant will get a car.
2. (b) Monty now opens door 2, revealing a goat. Given this information, find the conditional probability that the contestant will get a car.

Solution.

1. So there are 4 situations in which we can be either the two doors where we can land are Goat (G) Goat (G)= first, Goat (G) Car (C) = second, Car (C) Goat (G) = third,or Car (C) Car (C) = fourth. Therefore:

$$P(getCar) = P(getCar|first)P(first)+$$
$$P(getCar|second)P(second)+$$
$$P(getCar|third)P(third)+$$
$$P(getCar|fourth)P(fourth)+$$
$$= (0 \cdot q^2) + (1 \cdot pq) + (1 \cdot pq) + (1 \cdot p^2)$$
$$= p^2 + 2pq$$

Notice that the probability of $P(getCar|third)$ and $P(getCar|second)$ is 1 because Monty has revealed the Goat so the only one remaining is the car.

2. This means we are in one of three situations:

   (a) Goat (G) Goat (G) = first
   (b) Goat (G) Car (C) = second
   (c) Car (C) Goat (G) = third

Therefore the probability of getting the car is 2/3 since there are only two favourable positions in this case.

WRONG APPROACH: We will use the following notation C_i for a car being in position i and C_i^- for a goat being in that position.

$$P(getCar) = P(getCar|C_1)P(C_1) + P(getCar|C_1^-)P(C_1^-) +$$
$$P(getCar|C_2)P(C_2) + P(getCar|C_2^-)P(C_2^-) +$$
$$P(getCar|C_3)P(C_3) + P(getCar|C_3^-)P(C_3^-)$$
$$= (0 \cdot p) + 0 \cdot q + (1/2 \cdot p) + 0 \cdot q + (1/2 \cdot p) + 0 \cdot q$$
$$= \frac{p}{2} + \frac{p}{2} = p$$

This solution is wrong. As to why it is wrong, I still have to think about that...

2.4 First-step analysis and gambler's ruin

Exercise 48
A fair die is rolled repeatedly, and a running total is kept (which is, at each time, the total of all the rolls up until that time). Let p_n be the probability that the running total is ever exactly n (assume the die will always be rolled enough times so that the running total will eventually exceed n, but it may or may not ever equal n).

1. (a) Write down a recursive equation for p_n (relating p_n to earlier terms p_k in a simple way). Your equation should be true for all positive integers n, so give a definition of p_0 and p_k for $k < 0$ so that the recursive equation is true for small values of n.
2. (b) Find p_7.
3. (c) Give an intuitive explanation for the fact that $p_n \to 1/3.5 = 2/7$ as $n \to \infty$.

Solution.

1. Define $p_0 = 1$ and $p_k = 0$ for $k < 0$ then:

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

2.

$$p_7 = \frac{1}{6}(p_1 + p_2 + p_3 + p_4 + p_5 + p_6)$$
$$= \frac{1}{6}(\frac{1}{6} + \frac{1}{6}(1 + \frac{1}{6}) + \frac{1}{6}(1 + \frac{1}{6})^2 + \frac{1}{6}(1 + \frac{1}{6})^3 + \frac{1}{6}(1 + \frac{1}{6})^4 + \frac{1}{6}(1 + \frac{1}{6})^5 \approx 0.2536$$

3. The total amount we can get from all the possible sides is 21 but since we have 6 sides we get an expected value of $21/6 = 7/2 = 3.5$. To get some intuition we will think of the following imagine that we have a die in which all of the sides are 1 then our p_n from this problem would be 1 because well we will eventually get there, now imagine the same case but with 2 then we have 1/2 chance of getting the number as we may skip it by one with each throw. Now do it with 3 we get 1/3. In this problem since it is a normal die we know that on average we get a value of 3.5 so we have a $p_n = 1/3.5$ as $n \to \infty$. But we can also think of 1/3.5 as we will get 2 numbers every seven throws.

Exercise 49
A sequence of $n \geq 1$ independent trials is performed, where each trial ends in "success" or "failure" (but not both). Let p_i be the probability of success in the ith trial, $q_i = 1 - p_i$, and $b_i = q_i - 1/2$, for $i = 1, 2, \ldots, n$. Let An be the event that the number of successful trials is even.

1. (a) Show that for $n = 2$, $P(A_2) = 1/2 + 2b_1b_2$.
2. (b) Show by induction that
$$P(A_n) = 1/2 + 2^{n-1}b_1b_2 \dots b_n.$$
(This result is very useful in cryptography. Also, note that it implies that if n coins are flipped, then the probability of an even number of Heads is 1/2 if and only if at least one of the coins is fair.) Hint: Group some trials into a supertrial.
3. (c) Check directly that the result of (b) is true in the following simple cases: $p_i = 1/2$ for some i; $p_i = 0$ for all i; $p_i = 1$ for all i.

Solution.

1. We will first show what the answer means:
$$P(A_2) = 1/2 + 2b_1b_2 = 1/2 + 2(q_i - 1/2)(q_i - 1/2)$$
$$= 1 - 2q_i + 2q_i^2$$
$$= 1 + 2(1 - p_i)(1 - p_i) - 2(1 - p_i)$$
$$1 - 2p_i + 2p_i^2$$
Then we solve the problem with the event of having a success as S and the event of a failure as F:
$$P(A_2) = P(A_2|SS)P(SS) + P(A_2|FF)P(FF)$$
$$= 1 \cdot p_i^2 + 1 \cdot (1 - p_i)(1 - p_i)$$
$$= 1 - 2p_i + 2p_i^2$$
TADAH!
2. *Proof.* By induction,
Base Case: $i = 1$, the only way in which we can get an even number of success in one game is if we fail it.
$$P(A_1) = 1 - p_i = 1/2 + 2^{1-1}b_1 = 1^2 + 1 \cdot (1 - p_i - 1/2) = 1 - p_i$$
Induction Step: Suppose that for $1 \geq i \leq n$
$$P(A_n) = 1/2 + 2^{n-1}b_1b_2 \dots b_n.$$
then, the only way we can get A_{n+1} is by having a success after the event A_n^- which is an odd number of success or a failure after an even number of sucess. Notice:
$$P(A_n^-) = 1 - (1/2 + 2^{n-1}b_1b_2 \dots b_n) = 1/2 - 2^{n-1}b_1b_2 \dots b_n$$
Therefore:
$$P(A_{n+1}) = P(S|A_n^-)P(A_n^-) + P(F|A_n)P(A_n)$$
$$= (p_i)(1/2 - 2^{n-1}b_1b_2 \dots b_n) + (1 - p_i)(1/2 + 2^{n-1}b_1b_2 \dots b_n)$$
$$= (p_i)(1/2 - 2^{n-1}b_1b_2 \dots b_n) + (1/2 + 2^{n-1}b_1b_2 \dots b_n) + (-p_i)(1/2 + 2^{n-1}b_1b_2 \dots b_n)$$
$$= 1/2 - 2(p_i2^{n-1}b_1b_2 \dots b_n) + 2^{n-1}b_1b_2 \dots b_n$$
$$= 1/2 - p_i2^nb_1b_2 \dots b_n + 2^{n-1}b_1b_2 \dots b_n$$
$$= 1/2 - p_i2^nb_1b_2 \dots b_n + 2^{n-1}b_1b_2 \dots b_n$$
$$= 1/2 + 2^{n-1}b_1b_2 \dots b_n - p_i2^nb_1b_2 \dots b_n + 0$$
$$= 1/2 + 2^{n-1}b_1b_2 \dots b_n - p_i2^nb_1b_2 \dots b_n + (2^{n-1}b_1b_2 \dots b_n - 2^{n-1}b_1b_2 \dots b_n)$$

$$= 1/2 + 2(2^{n-1}b_1b_2 \dots b_n) - p_i 2^n b_1b_2 \dots b_n - 2^{n-1}b_1b_2 \dots b_n$$
$$= 1/2 + 2^n b_1b_2 \dots b_n - p_i 2^n b_1b_2 \dots b_n - 2^{n-1}b_1b_2 \dots b_n$$
$$= 1/2 + 2^n b_1b_2 \dots b_n(1 - p_i - 1/2)$$
$$P(A_{n+1}) = 1/2 + 2^n b_1b_2 \dots b_n b_{n+1}.$$

□

3. If $p_i = 0$ for all i this would mean that we would start with 0 success and it will remain that way forever so the probability of having $P(A_n) = 1$ checking the result

$$P(A_n) = 1/2 + 2^{n-1}(1 - p - 1/2)(1 - p - 1/2) \dots (1 - p - 1/2)$$
$$= 1/2 + 2^{n-1}(1/2)(1/2) \dots (1/2)$$
$$= 1/2 + 2^{n-1} \cdot 2^{-n} = 1/2 + 2^{n-n-1} = 1/2 + 1/2 = 1$$

If $p_i = \frac{1}{2}$ for some i. No matter if we had an even or odd amount of success before the first i where $p_i = 1/2$, getting or not getting an event amount of success would end up being $P(A_n) = 1/2$ as this sole entry already change the running amount and even if after it we start getting just sucess we would still have to take into account this 1/2 initial chance. Verification:

$$P(A_n) = 1/2 + 2^{n-1}(1 - p - 1/2)(1 - p - 1/2) \dots (1 - 1/2 - 1/2) \dots (1 - p - 1/2)$$
$$= 1/2 + 2^{n-1}(1 - p - 1/2)(1 - p - 1/2) \dots (o) \dots (1 - p - 1/2) = 1/2$$

If $p_i = 0$ for all i this would mean we always have success so the only thing that changes our probability of even sucessful trials is how many trials we do if we do an even number of trials then $P(A_n) = 1$ if we do an odd number of trials then $P(A_n) = 0$. To verify the result we will do it by cases where a is some positive integer:

Case 1: $n = 2a$ then $n - 1 = 2a - 1$ so

$$P(A_n) = 1/2 + 2^{n-1}(1 - 1 - 1/2)(1 - 1 - 1/2) \dots (1 - 1 - 1/2)$$
$$= 1/2 + 2^{n-1}(-1/2)(-1/2) \dots (-1/2)$$
$$= 1/2 + (-1)(-1/2) = \frac{1}{2} + \frac{1}{2} = 1$$

Case 2: $n = 2a + 1$ then $n - 1 = 2a$ so

$$P(A_n) = 1/2 + 2^{n-1}(1 - 1 - 1/2)(1 - 1 - 1/2) \dots (1 - 1 - 1/2)$$
$$= 1/2 + 2^{n-1}(-1/2)(-1/2) \dots (-1/2)$$
$$= 1/2 + (1)(-1/2) = \frac{1}{2} - \frac{1}{2} = 0$$

Exercise 50
Calvin and Hobbes play a match consisting of a series of games, where Calvin has probability p of winning each game (independently). They play with a "win by two" rule: the first player to win two games more than his opponent wins the match. Find the probability that Calvin wins the match (in terms of p), in two different ways:

1. (a) by conditioning, using the law of total probability.
2. (b) by interpreting the problem as a gambler's ruin problem.

Solution. Hi

1. still
2. not

Exercise 51
A gambler repeatedly plays a game where in each round, he wins a dollar with probability 1/3 and loses a dollar with probability 2/3. His strategy is "quit when he is ahead by \$2". Suppose that he starts with a million dollars. Show that the probability that he'll ever be ahead by \$2 is less than 1/4.

Solution. We will use the gambler's ruin problem solution formula. You can think of this game as a two person game where the other person is the game or the house however you want to call it. Then:

$$p = \frac{1}{3}$$

$$q = \frac{2}{3}$$

$$N = 1000002$$

$$i = 1000000$$

Since $p \neq \frac{1}{2}$ we have:

$$\begin{aligned} p_i &= \frac{1-(\frac{q}{p})^i}{1-(\frac{q}{p})^{i+2}} \\ &= \frac{1-2^i}{1-2^{i+2}} \\ &= \frac{1-2^i}{1-2^i\cdot 4} \\ &< \frac{1-2^i}{4-2^i\cdot 4} = \frac{1}{4} \end{aligned}$$

Exercise 52
As in the gambler's ruin problem, two gamblers, A and B, make a series of bets, until one of the gamblers goes bankrupt. Let A start out with i dollars and B start out with $N-i$ dollars, and let p be the probability of A winning a bet, with $0 < p < 1$. Each bet is for $\frac{1}{k}$ dollars, with k a positive integer, e.g., $k = 1$ is the original gambler's ruin problem and $k = 20$ means they're betting nickels. Find the probability that A wins the game, and determine what happens to this as $k \to \infty$.

Solution. We will use the gambler's ruin problem solution formula. But modified with $20i$ as starting wealth and $20N$ as the money to win. Then: If $p \neq \frac{1}{2}$ then

$$p_i = \frac{1-(\frac{q}{p})^{20i}}{1-(\frac{q}{p})^{20N}}$$

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

$$p_i = \frac{20i}{20N} = \frac{i}{N}$$

If $k \to \infty$ then if $p \neq \frac{1}{2}$ then

$$lim_{k\to\infty} p_{ki} = \frac{1 - (\frac{q}{p})^{ki}}{1 - (\frac{q}{p})^{kN}} = 1$$

We can use L'Hôpitals rule here and derivate numerator and denominator in respect to p

$$\frac{-ki \cdot (\frac{q}{p})^{ki-1}}{-kN \cdot (\frac{q}{p})^{kN-1}}$$

$$= (\frac{i}{N}) \cdot ((\frac{q}{p})^{ki-1} / (\frac{q}{p})^{kN-1})$$

$$= (\frac{i}{N}) \cdot ((\frac{q}{p})^{ki-1} (\frac{q}{p})^{-(kN-1)})$$

$$= (\frac{i}{N}) \cdot (\frac{q}{p})^{ki-1-kN+1}$$

$$= (\frac{i}{N}) \cdot (\frac{q}{p})^{k(i-N)}$$

Depending on the size of i and N and p we would get different results. For instance if $i = N$ then $p_i = \frac{i}{N}$. If $p = \frac{1}{2}$ then

$$lim_{k\to\infty} p_{ki} = \frac{ki}{kN} = \frac{i}{N}$$

Exercise 53
There are 100 equally spaced points around a circle. At 99 of the points, there are sheep, and at 1 point, there is a wolf. At each time step, the wolf randomly moves either clockwise or counterclockwise by 1 point. If there is a sheep at that point, he eats it. The sheep don't move. What is the probability that the sheep who is initially opposite the wolf is the last one remaining?

Solution.

2.5 Simpson's paradox

Exercise 59
The book Red State, Blue State, Rich State, Poor State by Andrew Gelman [12] discusses the following election phenomenon: within any U.S. state, a wealthy voter is more likely to vote for a Republican than a poor voter, yet the wealthier states tend to favor Democratic candidates!

1. (a) Assume for simplicity that there are only 2 states (called Red and Blue), each of which has 100 people, and that each person is either rich or poor, and either a Democrat or a Republican. Make up numbers consistent with the above, showing how this phenomenon is possible, by giving a $2x2$ table for each state (listing how many people in each state are rich Democrats, etc.). So within each state, a rich voter is more likely to vote for a Republican than a poor voter, but the percentage of Democrats is higher in the state with the higher percentage of rich people than in the state with the lower percentage of rich people.

2. (b) In the setup of (a) (not necessarily with the numbers you made up there), let D be the event that a randomly chosen person is a Democrat (with all 200 people equally likely), and B be the event that the person lives in the Blue State. Suppose that 10 people move from the

Blue State to the Red State. Write P_{old} and P_{new} for probabilities before and after they move. Assume that people do not change parties, so we have $P_{new}(D) = P_{old}(D)$. Is it possible that both $P_{new}(D|B) > P_{old}(D|B)$ and $P_{new}(D|B_c) > P_{old}(D|B_c)$ are true? If so, explain how it is possible and why it does not contradict the law of total probability $P(D) = P(D|B)P(B) + P(D|B_c)P(B_c)$; if not, show that it is impossible.

Solution.

1. The tables are as follows:

Blue State

-	Rich	Poor
Dem	2	75
Rep	15	8

Red State

-	Rich	Poor
Dem	1	20
Rep	10	79

2. Suppose 5 democrats and 5 republicants move then,

$$0.82 \approx \frac{74}{90} = P_{new}(D|B) > P_{old}(D|B) = \frac{79}{100} = 0.79$$

and

$$0.2363 \approx \frac{26}{110} = P_{new}(D|B_c) > P_{old}(D|B_c) = \frac{21}{100} = 0.21$$

and it doesn't break the laws of probability since $P(B)$ will be equal to $\frac{90}{200}$ and $P(B_c) = \frac{110}{200}$

$$\frac{79}{100} \cdot \frac{1}{2} + \frac{21}{100} \cdot \frac{1}{2} = \frac{1}{2}$$
$$\frac{1}{2} = \frac{74}{90} \cdot \frac{90}{200} + \frac{26}{110} \cdot \frac{110}{200}$$

2.6 Mixed practice

Exercise 60

A patient is being given a blood test for the disease conditionitis. Let p be the prior probability that the patient has conditionitis. The blood sample is sent to one of two labs for analysis, lab A or lab B. The choice of which lab to use is made randomly, independent of the patient's disease status, with probability 1/2 for each lab. For lab A, the probability of someone testing positive given that they do have the disease is a_1, and the probability of someone testing negative given that they do not have the disease is a_2. The corresponding probabilities for lab B are b_1 and b_2.

1. Find the probability that the patient has the disease, given that they tested positive.

2. Find the probability that the patient's blood sample was analyzed by lab A, given that the patient tested positive.

Solution. We will use the following notation A for testing in lab A, B for lab B. We will also denote C as being the event of having conditionitis and C^- as not having conditionitis. The $P(C^-) = 1 - p = q$. The problem statement says:

$$P(+|C \cap A) = a_1$$
$$P(+|C \cap B) = b_1$$

$$P(+|C^- \cap A) = a_2$$
$$P(+|C^- \cap B) = b_2$$
$$P(C) = p$$
$$P(A) = 1/2$$
$$P(B) = 1/2$$

1. We need the following information:

$$\begin{aligned}
P(+|C) &= P(+|C \cap A)P(A|C) + P(+|C \cap B)P(B|C) \\
&= a_1 \cdot \frac{P(A \cap C)}{P(C)} + b_1 \cdot \frac{P(B \cap C)}{P(C)} \\
&= a_1 \cdot \frac{P(A)P(C)}{P(C)} + b_1 \cdot \frac{P(B)P(C)}{P(C)} \\
&= \frac{1}{2}(a_1 + b_1) \\
P(+|C^-) &= P(+|C^- \cap A)P(A|C^-) + P(+|C^- \cap B)P(B|C^-) \\
&= a_2 \cdot \frac{P(A \cap C^-)}{P(C^-)} + b_2 \cdot \frac{P(B \cap C^-)}{P(C^-)} \\
&= a_2 \cdot \frac{P(A)P(C^-)}{P(C^-)} + b_2 \cdot \frac{P(B)P(C^-)}{P(C^-)} \\
&= \frac{q}{2}(a_2 + b_2) \\
P(+) &= P(+|C)P(C) + P(+|C^-)P(C^-) \\
&= \frac{1}{2}(a_1 + b_1) \cdot p + \frac{q}{2}(a_2 + b_2) \cdot q \\
&= \frac{p}{2}(a_1 + b_1) + \frac{q}{2}(a_2 + b_2)
\end{aligned}$$

Therefore what we are looking for is:

$$\begin{aligned}
P(C|+) &= \frac{P(+|C)P(C)}{P(+)} \\
&= \frac{\frac{1}{2}(a_1 + b_1) \cdot p}{\frac{p}{2}(a_1 + b_1) + \frac{q}{2}(a_2 + b_2)} \\
&= \frac{\frac{p}{2}(a_1 + b_1)}{\frac{1}{2} \cdot (p(a_1 + b_1) + q(a_2 + b_2))}
\end{aligned}$$

2. We first need:

$$\begin{aligned}
P(+|A) &= P(+|A \cap C)P(C|+) + P(+|A \cap C^-)P(C^-|A) \\
&= a_1 \frac{P(C \cap A)}{P(P(A))} + b_1 \frac{P(C^- \cap A)}{P(P(A))} \\
&= a_1 \frac{P(C)P(A)}{P(A)} + b_1 \frac{P(C^-)P(A)}{P(A)} \\
& a_1 p + b_1 q
\end{aligned}$$

Therefore what we are looking for is:

$$\begin{aligned}
P(A|+) &= \frac{P(+|A)P(A)}{P(+)} \\
&= \frac{(a_1 p + b_1 q) \cdot \frac{1}{2}}{\frac{p}{2}(a_1 + b_1) + \frac{q}{2}(a_2 + b_2)}
\end{aligned}$$

Exercise 61
Fred decides to take a series of n tests, to diagnose whether he has a certain disease (any individual test is not perfectly reliable, so he hopes to reduce his uncertainty by taking multiple tests). Let D be the event that he has the disease, $p = P(D)$ be the prior probability that he has the disease, and $q = 1 - p$. Let T_j be the event that he tests positive on the jth test.

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

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

Solution.

1. We are looking for

$$P(D|T_1 \cap \dots \cap T_j \cap \dots T_n) = \frac{P(T_1 \cap \dots \cap T_j \cap \dots T_n|D)P(D)}{P(T_1 \cap \dots \cap T_j \cap \dots T_n)}$$

Since we know that the test results are conditionally independent given Fred's disease status, we know:

$$P(T_1 \cap \dots \cap T_j \cap \dots T_n|D) = P(T_1|D) \dots P(T_j|D) \dots P(T_n|D) = \prod_{j=1}^{n} P(T_j|D) = \prod_{k=1}^{n} a = a^n$$

Notice also that since the statement says "conditionally independent given Fred's disease status" this also includes D_c so:

$$P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c) = P(T_1|D_c) \dots P(T_j|D_c) \dots P(T_n|D_c) = \prod_{j=1}^{n} P(T_j|D_c) = \prod_{k=1}^{n} b = b^n$$

Therefore:

$$P(D|T_1 \cap \dots \cap T_j \cap \dots T_n) = \frac{a^n \cdot p}{P(T_1 \cap \dots \cap T_j \cap \dots T_n)}$$

by the law of total probability we have:

$$P(D|T_1 \cap \dots \cap T_j \cap \dots T_n) = \frac{a^n \cdot p}{P(T_1 \cap \dots \cap T_j \cap \dots T_n|D)P(D) + P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c)P(D_c)}$$

$$= P(D|T_1 \cap \dots \cap T_j \cap \dots T_n) = \frac{a^n \cdot p}{a^n \cdot p + b^n \cdot q}$$

2. We first need the following knowledge:

$$P(T_1 \cap \dots \cap T_j \cap \dots T_n|D) = P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G)P(G|D) + P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G_c)P(G_c|D)$$

, but in the part where G_c we can simply because the test results are conditionallly independent given his disease status.

$$P(T_1 \cap \dots \cap T_j \cap \dots T_n|D) = P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G)P(G|D) + \prod_{j=i}^{n} P(T_j|D, G_c)P(G_c|D)$$

therefore

$$P(T_1 \cap \dots \cap T_j \cap \dots T_n|D) = P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G)P(G|D) + \prod_{j=i}^{n} a_0 \cdot P(G_c|D)$$

and given that disease status is independent from having the gene and a_0 is independent from j

$$P(T_1 \cap \dots \cap T_j \cap \dots T_n|D) = P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G) \cdot \frac{1}{2} + a_0^n \cdot \frac{1}{2}$$

In the same way

$$P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c) = P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c \cap G) \cdot \frac{1}{2} + P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c \cap G_c) \cdot \frac{1}{2}$$

$$= P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c \cap G) \cdot \frac{1}{2} + \prod_{j=1}^{n} P(T_j|D_c, G_c) \cdot \frac{1}{2}$$

$$= P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c \cap G) \cdot \frac{1}{2} + \prod_{j=1}^{n} b_0 \cdot \frac{1}{2}$$

Since b_0 is independent of j

$$= P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c \cap G) \cdot \frac{1}{2} + b_0^n \cdot \frac{1}{2}$$

Now for the actual problem we are looking for

$$P(D|T_1 \cap \dots \cap T_j \cap \dots T_n) = \frac{P(T_1 \cap \dots \cap T_j \cap \dots T_n|D)P(D)}{P(T_1 \cap \dots \cap T_j \cap \dots T_n)}$$

$$= \frac{P(T_1 \cap \dots \cap T_j \cap \dots T_n|D) \cdot p}{P(T_1 \cap \dots \cap T_j \cap \dots T_n)}$$

$$= \frac{[P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G) \cdot \frac{1}{2} + a_0^n \cdot \frac{1}{2}] \cdot p}{P(T_1 \cap \dots \cap T_j \cap \dots T_n)}$$

From the law of total probability we have

$$= \frac{[P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G) \cdot \frac{1}{2} + a_0^n \cdot \frac{1}{2}] \cdot p}{P(T_1 \cap \dots \cap T_j \cap \dots T_n|D)P(D) + P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c)P(D_c)}$$

$$= \frac{[P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G) \cdot \frac{1}{2} + a_0^n \cdot \frac{1}{2}] \cdot p}{[P(T_1 \cap \dots \cap T_j \cap \dots T_n|D \cap G) \cdot \frac{1}{2} + a_0^n \cdot \frac{1}{2}] \cdot p + [P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c \cap G) \cdot \frac{1}{2} + b_0^n \cdot \frac{1}{2}] \cdot q}$$

Now for the final part let's remember that the statement says "some people have a certain gene that makes them always test positive. Let G be the event that Fred has the gene" therefore $P(T_1 \cap \dots \cap T_j \cap \dots T_n|D, G) = 1 = P(T_1 \cap \dots \cap T_j \cap \dots T_n|D_c, G)$ so

$$= \frac{[\frac{1}{2} + a_0^n \cdot \frac{1}{2}] \cdot p}{[\frac{1}{2} + a_0^n \cdot \frac{1}{2}] \cdot p + [\frac{1}{2} + b_0^n \cdot \frac{1}{2}] \cdot q}$$

Exercise 63
Three fair coins are tossed at the same time. Explain what is wrong with the following argument: "there is a 50% chance that the three coins all landed the same way, since obviously it is possible to find two coins that match, and then the other coin has a 50% chance of matching those two".

Solution. This solution is written as if it said "given that we have two coins that have the same side the probability of getting a third coin with the same side is 50%". While this statement is correct (since these statements are independent), it does not answer the question about what's the probability of getting three coins to land on the same side.

Exercise 64
An urn contains red, green, and blue balls. Let r, g, b be the proportions of red, green, blue balls, respectively (r + g + b = 1).

1. (a) Balls are drawn randomly with replacement. Find the probability that the first time a green ball is drawn is before the first time a blue ball is drawn. Hint: Explain how this relates to finding the probability that a draw is green, given that it is either green or blue.
2. (b) Balls are drawn randomly without replacement. Find the probability that the first time a green ball is drawn is before the first time a blue ball is drawn. Is the answer the same or different than the answer in (a)? Hint: Imagine the balls all lined up, in the order in which they will be drawn. Note that where the red balls are standing in this line is irrelevant.
3. (c) Generalize the result from (a) to the following setting. Independent trials are performed, and the outcome of each trial is classified as being exactly one of type 1, type 2, . . . , or type n, with probabilities $p_1, p_2, ..., p_n$, respectively. Find the probability that the first trial to result in type i comes before the first trial to result in type j, for $i \neq j$.

Solution.

1.
2.
3.

Exercise 66
A fair die is rolled repeatedly, until the running total is at least 100 (at which point the rolling stops). Find the most likely value of the final running total (i.e., the value of the running total at the first time when it is at least 100).
Hint: Consider the possibilities for what the running total is just before the last roll.

Solution. We have 6 numbers that could reach at least 100 in the next throw those numbers are $94 \ldots 99$. In the next throw we can reach the numbers in the following way

1. 95 - 1 time
2. 96 - 2 times
3. 97 - 3 times
4. 98 - 4 times
5. 99 - 5 times
6. 100 - 6 times
7. 101 - 5 times

8. 102 - 4 times
9. 103 - 3 times
10. 104 - 2 times
11. 105 - 1 time

We can therefore see that the most likely value for the final running total is 100.

Chapter 3

Random variables and their distributions

3.1 PMFs and CDFs

Exercise 1
People are arriving at a party one at a time. While waiting for more people to arrive they entertain themselves by comparing their birthdays. Let X be the number of people needed to obtain a birthday match, i.e., before person X arrives no two people have the same birthday, but when person X arrives there is a match. Find the PMF of X.

Solution.

$$P(X = k) = 1 - \frac{365 \cdot 364 \cdot 363 \dots (365 - k + 1)}{365^k}$$

Exercise 3
Let X be an r.v. with CDF F, and $Y = \mu + \sigma X$, where μ and σ are real numbers with $\sigma > 0$. (Then Y is called a location-scale transformation of X; we will encounter this concept many times in Chapter 5 and beyond.) Find the CDF of Y , in terms of F .

Solution.

$$P(Y \leq y) = P(\mu + \sigma X \leq y) = P(\sigma X \leq y - \mu) = P(X \leq \frac{y - \mu}{\sigma})$$

Exercise 5
(a) Show that $p(n) = (\frac{1}{2})^{n+1}$ for $n = 0, 1, 2, \dots$ is a valid PMF for a discrete r.v.
(b) Find the CDF of a random variable with the PMF from (a).

Solution. We will call this r.v X
(a) We assume n is the support of X we also see that n has the property of being infinite countable.
Since $(\frac{1}{2})^{n+1} > 0$ for all n we see that $p(n) > 0$
We will now prove that this PMF as a series tends to one. We can use the ratio test to see that it converges absolutely.

$$L = \lim_{n\to\infty} \frac{|a_{n+1}|}{|an|} = \frac{(1/2^{n+2})}{(1/2^{n+1})} = \frac{1}{2}$$

Since $L < 1$ the series converges absolutely. Where does it converge? If we define the partial sum as

$$S_n = \frac{1}{2} + \frac{1}{4} + \dots + \frac{1}{2^{n-1}} + \frac{1}{2^n}$$

then we can see that as n approaches infinity

$$1 - S_n = 1 - \frac{1}{2^n} = 1$$

Therefore the series converges at 1 and we have the two requirements to make a valid PMF.
(b)

$$P(X \leq n) = \sum_{i=0}^{n} P(X = i)$$

Exercise 9
Let F1 and F2 be CDFs, $0 < p < 1$, and $F(x) = pF_1(x) + (1-p)F_2(x)$ for all x.
(a) Show directly that F has the properties of a valid CDF (see Theorem 3.6.3). The distribution defined by F is called a mixture of the distributions defined by F_1 and F_2.
(b) Consider creating an r.v. in the following way. Flip a coin with probability p of Heads. If the coin lands Heads, generate an r.v. according to F_1; if the coin lands Tails, generate an r.v. according to F_2. Show that the r.v. obtained in this way has CDF F.

Solution.
(a) Suppose $x_1 \leq x_2$ then $F_1(x_1) \leq F_1(x_2)$ and $F_2(x_1) \leq F_2(x_2)$. We may also notice that p is positive and bigger than 1 and so is $(1-p)$ therefore $pF_1(x_1) \leq pF_1(x_2)$ and $(1-p)F_2(x_1) \leq (1-p)F_2(x_2)$. Furthermore by adding them together we see that

$$F(x_1) = pF_1(x_1) + (1-p)F_2(x_1) \leq pF_1(x_2) + (1-p)F_2(x_2) = F(x_2)$$

We have the first property.
For the second property notice that $F_1(a) = \lim_{x \to a^+} F_1(x)$ and $F_2(a) = \lim_{x \to a^+} F_2(x)$. Therefore

$$\lim_{x \to a^+} F(x) = pF_1(a) + (1-p)F_2(a)$$

and also

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

so

$$F(a) = \lim_{x \to a^+} F(x)$$

.
For the third condition since $\lim_{x \to -\infty} F_1(x) = 0$ and $\lim_{x \to -\infty} F_2(x) = 0$ then

$$\lim_{x \to -\infty} F(x) = p \cdot 0 + (1-p) \cdot 0 = 0$$

similarly since $\lim_{x \to \infty} F_1(x) = 1$ and $\lim_{x \to \infty} F_2(x) = 1$ then

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

(b) Since F is a CDF we have $F(x) = P(X \leq x)$. We define H to be the even of getting Heads and naturally H^c the even of getting Tails. Then using the law of total probability

$$P(X \leq x) = P(H)P(X \leq x|H) + P(H^c)P(X \leq x|H^c)$$

then

$$P(X \leq x) = pF_1(x) + (1-p)F_2(x)$$

Exercise 10
(a) Is there a discrete distribution with support $1, 2, 3, \ldots$, such that the value of the PMF at n is proportional to $\frac{1}{n}$?
Hint: See the math appendix for a review of some facts about series.
(b) Is there a discrete distribution with support $1, 2, 3, \ldots$, such that the value of the PMF at n is proportional to $\frac{1}{n^2}$?

Solution.
(a)No, we know the sum of the PMF should be in total 1 and we also know the value at n is proportional to n so set a to be a positive real number then

$$\sum_{n=1}^{\infty} = \frac{a}{n}$$

since $a \geq 1$ then since

$$\sum_{n=1}^{\infty} = \frac{1}{n}$$

is divergent then so is

$$\sum_{n=1}^{\infty} = \frac{a}{n}$$

therefore it is not possible to have a PMF like this. There is no way changing the constant a can fix this.
(b)Yes, from the P-test we know that

$$\sum_{n=1}^{\infty} = \frac{1}{n^2}$$

is convergent. Suppose it converges to C a real number then set our constant a to be $\frac{1}{C}$ then

$$\sum_{n=1}^{\infty} = \frac{a}{n^2} = \frac{\frac{1}{C}}{n^2} = 1$$

. So there is a valid PMF.

3.2 Named distributions

Exercise 15
Find the CDF of an r.v. $X \sim DUnif(1, 2, \ldots, n)$.

Solution. Since we know $C = \{1, 2, \ldots, n\}$ then we know that the PMF of X is

$$P(X = x) = \frac{1}{|C|} = \frac{1}{n}$$

therefore we have the following CDF

$$P(X \leq x) = x \cdot \frac{1}{n}$$

It is important to note that if $\lim_{x \to -\infty}$ our CDF is defined to go to 0 and in the same way it should go to 1 when it approaches infinity, so it is valid.

Exercise 16
Let $X \sim DUnif(C)$, and B be a nonempty subset of C. Find the conditional distribution of X, given that X is in B.

Solution.
$$P(X = k|X \in B) = \frac{P(X = k \cap X \in B)}{P(X \in B)}$$

By cases:
If $k \in B$ then
$$= \frac{P(X = k)}{P(X \in B)}$$
$$= \frac{\frac{1}{|C|}}{\frac{|B|}{|C|}} = \frac{1}{|B|}$$

The Uniform distribution tells us how X is distributed in a set C so if we are asking how is X distributed over B which is in C then the probability of this should be
$$\frac{|B|}{|C|}$$

If $k \notin B$ then $P(X = k \cap X \in B) = 0$ so
$$= \frac{0}{P(X \in B)} = 0$$

Finally since we have that if $x \in B$ then $P(X = k|X \in B) = \frac{1}{|B|}$ and 0 otherwise this is by definition $X \sim Dunif(B)$.

Exercise 17
An airline overbooks a flight, selling more tickets for the flight than there are seats on the plane (figuring that it's likely that some people won't show up). The plane has 100 seats, and 110 people have booked the flight. Each person will show up for the flight with probability 0.9, independently. Find the probability that there will be enough seats for everyone who shows up for the flight.

Solution. So we are basically testing how many of the 110 people will show up. So we have a binomial r.v
$$X \sim Bin(110, 0.9)$$

But we want to know the probability that there will be enough seats for everyone who shows up for the flight. Which by using R we know is
$$P(X \leq 100) = 0.1245667$$

```
>dbinom(100,110,0.9)
[1] 0.1245667
```

I would personally choose an air line with better odds.

Exercise 20
Suppose that a lottery ticket has probability p of being a winning ticket, independently of other tickets. A gambler buys 3 tickets, hoping this will triple the chance of having at least one winning ticket.
(a) What is the distribution of how many of the 3 tickets are winning tickets?
(b) Show that the probability that at least 1 of the 3 tickets is winning is $3p - 3p^2 + p^3$, in two different ways: by using inclusion-exclusion, and by taking the complement of the desired event and then using the PMF of a certain named distribution.
(c) Show that the gambler's chances of having at least one winning ticket do not quite triple (compared with buying only one ticket), but that they do approximately triple if p is small.

Solution.
(a) $X \sim Bin(3, p)$
(b) Using inclusion-exclusion we know $(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$. In this case we get

$$P(Ticket1 \cup Ticket2 \cup Ticket3) =$$

$$p + p + p - p^2 - p^2 - p^2 + p^3 =$$
$$3p - 3p^2 + p^3$$

On the other hand we know we can also get this result if we take the complement which is 1 minus the event of not having any winning ticket. We will use the r.v from (a)

$$1 - P(X = 0) =$$

$$1 - (1(p)^0(1 - p)^3) =$$
$$1 - (1 - 3p(1 - p) - p^3) =$$
$$1 - (1 - 3p + 3p^2 - p^3) =$$
$$3p - 3p^2 + p^3$$

(c) Since the value of p is between 0 and 1 we have

$$3p^2 \geq p^3$$

$$3p + 3p^2 \geq 3p + p^3$$
$$3p \geq 3p - 3p^2 + p^3$$

So you don't actually triple your chances, which is quite sad. But if p is very very small then

$$(-3p^2 + p^3) \to 0$$

so $3p \geq 3p - (0) = 3p$

Exercise 36
There are n voters in an upcoming election in a certain country, where n is a large, even number. There are two candidates: Candidate A (from the Unite Party) and Candidate B (from the Untie Party). Let X be the number of people who vote for Candidate A. Suppose that each voter chooses randomly whom to vote for, independently and with equal probabilities.
(a) Find an exact expression for the probability of a tie in the election (so the candidates end up with the same number of votes).
(b) Use Stirling's approximation, which approximates the factorial function as

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

to find a simple approximation to the probability of a tie. Your answer should be of the form $1/\sqrt{cn}$, with c a constant (which you should specify).

Solution. (a) We will use r.v $X \sim Bin(n,p)$ and we will calculate for $\frac{n}{2}$ half the votes to any party since there are only two parties this is general enough.

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

$$= \binom{n}{\frac{n}{2}} p^{n/2}(1-p)^{n/2}$$

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

Since n is even we can do the following

$$= \frac{n\cdot(n-1)\ldots(\frac{n}{2})\cdot(\frac{n}{2}-1)\cdot\dots\cdot 1}{(\frac{n}{2}!)(\frac{n}{2}!)} p^{n/2}(1-p)^{n/2}$$

$$= \frac{n\cdot(n-1)\ldots(\frac{n}{2}!)}{(\frac{n}{2}!)(\frac{n}{2}!)} p^{n/2}(1-p)^{n/2}$$

$$= \frac{n\cdot(n-1)\ldots\cancel{(\frac{n}{2}!)}}{\cancel{(\frac{n}{2}!)}(\frac{n}{2}!)} p^{n/2}(1-p)^{n/2}$$

$$= \frac{n\cdot(n-1)\ldots(\frac{n}{2}+1)}{(\frac{n}{2}!)} p^{n/2}(1-p)^{n/2}$$

Since we have equal probabilities we have $p=\frac{1}{2}$ and so

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

(b) Using the Stirling function we have

$$\frac{\sqrt{2\pi n}(\frac{n}{e})^n}{2\pi\frac{n}{2}(\frac{n}{2e})^{2n/2}} p^{n/2}(1-p)^{n/2}$$

$$= \frac{\sqrt{2\pi n}(\frac{n}{e})^n}{2\pi\frac{n}{2}(\frac{n}{2e})^{n}} p^{n/2}(1-p)^{n/2}$$

$$= \frac{\sqrt{2\pi n}(\frac{n}{e})^n}{2\pi\frac{n}{2}(\frac{1}{2})^n(\frac{n}{e})^{n}} p^{n/2}(1-p)^{n/2}$$

$$= \frac{\sqrt{2\pi n}}{2\pi\frac{n}{2}(\frac{1}{2})^n} p^{n/2}(1-p)^{n/2}$$

$$= \frac{\sqrt{2\pi n}}{2\pi\frac{n}{2}(\frac{1}{2})^n} p^{n/2}(1-p)^{n/2}$$

$$= \frac{\sqrt{2\pi n}}{2\pi n\frac{1}{2}(\frac{1}{2})^n} p^{n/2}(1-p)^{n/2}$$

$$= \frac{1}{\sqrt{2\pi n}\frac{1}{2}(\frac{1}{2})^n} p^{n/2}(1-p)^{n/2}$$

$$= \frac{1}{\sqrt{2\pi n}(\frac{1}{2})^{n+1}} p^{n/2}(1-p)^{n/2}$$

Since we have equal probabilities we have $p = \frac{1}{2}$ and so

$$= \frac{1}{\sqrt{2\pi n}(\frac{1}{2})^{n+1}} \frac{1}{2}^{n/2} \frac{1}{2}^{n/2}$$

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

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

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

$$= \frac{1}{\sqrt{2\pi\frac{1}{4}n}}$$

$$= \frac{1}{\sqrt{\pi\frac{1}{2}n}}$$

So $c = \pi\frac{1}{2}$

Exercise 37
A message is sent over a noisy channel. The message is a sequence $x_1, x_2, \ldots, x_n of n bits (x_i \in \{0,1\})$. Since the channel is noisy, there is a chance that any bit might be corrupted, resulting in an error (a 0 becomes a 1 or vice versa). Assume that the error events are independent. Let p be the probability that an individual bit has an error ($0 < p < 1/2$). Let $y_1, y_2, \ldots, y_n$ be the received message (so $y_i = x_i$ if there is no error in that bit, but $y_i = 1 - x_i$ if there is an error there).
To help detect errors, the nth bit is reserved for a parity check: x_n is defined to be 0 if $x_1+x_2+\cdots+x_{n-1}$ is even, and 1 if $x_1+x_2+\cdots+x_{n-1}$ is odd. When the message is received, the recipient checks whether y_n has the same parity as $y_1 + y_2 + \cdots + y_{n-1}$. If the parity is wrong, the recipient knows that at least one error occurred; otherwise, the recipient assumes that there were no errors.
(a) For $n = 5, p = 0.1$, what is the probability that the received message has errors which go undetected?
(b) For general n and p, write down an expression (as a sum) for the probability that the received message has errors which go undetected.
(c) Give a simplified expression, not involving a sum of a large number of terms, for the probability that the received message has errors which go undetected.
Hint for (c): Letting

$$a = \sum_{k even, k \geq 0} \binom{n}{k} p^k (1-p)^{n-k}$$

$$b = \sum_{k odd, k \geq 1} \binom{n}{k} p^k (1-p)^{n-k}$$

the binomial theorem makes it possible to find simple expressions for $a+b$ and $a-b$, which then makes it possible to obtain a and b.

Solution.
(a) It is important to note that we are counting just the errors that go undetected so if in the first 4 digits we have a pair number of errors it will go undetected. But the thing is that the last digit is also sent so there is also a probability of it getting kapput, and if we have an odd error in the 4 first but then the last one in the last digit the other errors will go undetected. Therefore

$$\binom{5}{2} p^2 (1-p)^3 + \binom{5}{4} p^4 (1-p)$$

when p=0.1
we have 0.07335
(b)Where t is even

$$\sum_{t \geq 2}^{n} \binom{n}{t} p^t (1-p)^{n-t}$$

(c)

$$a + b = \sum_{t=0}^{n} \binom{n}{t} p^t (1-p)^{n-t} = 1$$

This is the CDF of a binomial distribution so we know the result is 1.

$$a - b = \sum_{t=0}^{n} \binom{n}{t} (-p)^t (1-p)^{n-t} = (-p + 1 - p)^n = (1-2p)^n$$

By the binomial theorem.
Now

$$2a = 1 + (1-2p)^n$$

so

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

$$b = 1 - a$$

so

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

Since the probability of errors going undetected plus the probability of no errors (the k=0 case) is

$$\sum_{keven, k \geq 0} \binom{n}{k} p^k (1-p)^{n-k} = \frac{1 + (1-2p)^n}{2}$$

by subtracting the probability of no errors which is $(1-p)^n$ we have

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

3.3 Independence of r.v.s

Exercise 38
(a) Give an example of dependent r.v.s X and Y such that $P(X < Y) = 1$.
(b) Give an example of independent r.v.s X and Y such that $P(X < Y) = 1$.

Solution. (a) Suppose you are throwing 2 coins at the same time and Y is equal to 1 if you have heads and 0 otherwise and X is 1 if you have 2 heads and 0 otherwise. Then they are dependent since knowing that Y happened gives you a higher chance that X will happen and also

$$P(X < Y) = 1$$

since we have Bern(1/2) and Binomial(2,1/2) which say $0.25 < 0.5$
(b) Suppose you are throwing 2 coins and then another 2 coins and Y is equal to 1 if you have 1 head in the first 2 throws and 0 otherwise and X is 1 if you have 2 heads in the second 2 throws and 0 otherwise. Then they are independent since they are independent sets of two throws. And

$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1),$$

but

$$P(X < Y) = 1$$

since $0.25 < 0.5$ as above.

Exercise 39
Give an example of two discrete random variables X and Y on the same sample space such that X and Y have the same distribution, with support

$$\{1, 2, \ldots, 10\},$$

but the event $X = Y$ never occurs. If X and Y are independent, is it still possible to construct such an example?

Solution. Suppose $P(X = k) = 0.1$ for all $k \in 1, \ldots, 10$ Now suppose $Y = 15 - X$ and Y has the same distribution then

$$P(X = Y) = P(X = 15 - X) = P(2X = 15) = P(X = \frac{15}{2}) = 0$$

because 7.5 is not in the support it is equal to 0.
Now for the other part suppose X and Y are independent and $P(X = Y) = 0$. Now for all t,

$$0 = P(X = Y) = \sum_{t=1}^{10} P(X = t, Y = t) = \sum_{t=1}^{10} P(X = t)P(Y = t)$$

but since X and Y have he same distribution

$$= \sum_{t=1}^{10} P(X = t)^2 = 0$$

But since $P(X = t) \geq 0$ for all $t \in 1, \ldots, 10$ there is no way $\sum_{t=1}^{10} P(X = t)^2 = 0$ so we have a contradiction.

Exercise 40
Suppose X and Y are discrete r.v.s such that $P(X = Y) = 1$. This means that X and Y always take on the same value.
(a) Do X and Y have the same PMF?
(b) Is it possible for X and Y to be independent?

Solution.
(a) If they always take the same value they have to have the same distribution. For instance if both translate events s from the sample space as $P(s \in \Omega \mid X(s) = Y(s)) = 1$. Then no matter the s we have $X(s) = x$ for some $x \in \mathbb{R}$.

$$X(s) = x = Y(s)$$

Since X has a certain distribution of how it translates events from the sample space to $\mathbb{R}$, then Y must have that same distribution since X and Y always take the same value of the same event.
(b) Suppose $P(X = Y) = 1$ and X and Y are independent. Then, for all x_i in the support.

$$1 = P(X = Y) = \sum_{x_i} P(X = x, Y = x)$$

by independence

$$1 = \sum_{x_i} P(X = x)P(Y = x)$$

because they have the same distribution

$$1 = \sum_{x_i} P(X = x)^2$$

So the only way in which they can be independent is if we have two independent events with only one support and $P(X = Y) = 1$ and $P(X = x) = 1$ for that sole support. Otherwise it is not possible since $\sum_{x_i} P(X = x_i) = 1$ and $\sum P(X = x_i)^2 = 1$. Since we are going to prove it for more than one support, we may suppose we have supports a such that

$$0 < P(X = a) < 1$$

it follows that

$$P(X = a)^2 < P(X = a)$$

therefore

$$\sum_{x_i} P(X = x_i)^2 < \sum_{x_i} P(X = x_i)$$

since we know

$$\sum_{x_i} P(X = x_i) = 1$$

then

$$\sum_{x_i} P(X = x_i)^2 < 1$$

Which contradicts the definition of PMF.

Exercise 41
If X,Y,Z are r.v.s such that X and Y are independent and Y and Z are independent, does it follow that X and Z are independent?
Hint: Think about simple and extreme examples.

Solution.
No, for a counter example suppose we throw a coin and lets have X which tells us if the coin landed on heads and Z that tells us if the coin landed on tails. Now suppose you throw another coin and Y is the indicator of this coin landing on heads, then X is independent from Y and Y is independent from Z but X and Z are dependent. If you know the first coing landed on heads you know if it landed on tails.

Exercise 42
Let X be a random day of the week, coded so that Monday is 1, Tuesday is 2, etc. (so X takes values $1, 2, \ldots, 7$, with equal probabilities). Let Y be the next day after X (again represented as an integer between 1 and 7). Do X and Y have the same distribution? What is $P(X < Y)$?

Solution.
They have the same distribution as you are not really changing anything in the experiment. $Y = X + 1$ but modulus 7 since it wraps around if you are on sunday. So for some $k \in \{1, \ldots, 7\}$

$$P(Y = k) = P(X + 1 = k) = P(X = k - 1)$$

The $k - 1$ is modulus 7 so it is still a number say $t = k - 1(\%7)$ from 1 to 7. So we have

$$P(Y = k) = P(X = t)$$

so they indeed have the same distribution.
For the second part well the only situation in which $P(X < Y)$ is when $X = 7$ so $Y = 1$ which only happens 1 out of 7 times, since we have a uniform distribution. Therefore $P(X < Y) = 1 - P(X \geq Y) = 1 - \frac{1}{7} = \frac{6}{7}$

3.4 Mixed Practice

Exercise 46
Independent Bernoulli trials are performed, with success probability 1/2 for each trial. An important question that often comes up in such settings is how many trials to perform. Many controversies have arisen in statistics over the issue of how to analyze data coming from an experiment where the number of trials can depend on the data collected so far.
For example, if we can follow the rule “keep performing trials until there are more than twice as many failures as successes, and then stop”, then naively looking at the ratio of failures to successes (if and when the process stops) will give more than 2 : 1 rather than the true theoretical 1 : 1 ratio; this could be a very misleading result! However, it might never happen that there are more than twice as many failures as successes; in this problem, you will find the probability of that happening.
(a) Two gamblers, A and B, make a series of bets, where each has probability 1/2 of winning a bet, but A gets \$2 for each win and loses \$1 for each loss (a very favorable game for A!). Assume that the gamblers are allowed to borrow money, so they can and do gamble forever. Let p_k be the probability that A, starting with \$k, will ever reach \$0, for each $k \geq 0$. Explain how this story relates to the original problem, and how the original problem can be solved if we can find p_k.
(b) Find p_k.
Hint: As in the gambler’s ruin, set up and solve a difference equation for p_k. We have $p_k \to 0$ as $k \to \infty$(you don’t need to prove this, but it should make sense since the game is so favorable to A, which will result in A’s fortune going to ∞; a formal proof, not required here, could be done using the law of large numbers, an important theorem from Chapter 10). The solution can be written neatly in terms of the golden ratio.
(c) Find the probability of ever having more than twice as many failures as successes with independent $Bern(1/2)$ trials, as originally desired.

Solution.
(a) This is the same problem since this is a repetition of bernoulli trials with $p = 1/2$ and we also need a ratio of 2 : 1 for A to ever lose the game. For instance if A has won 100 games consecutively they should have \$200 so they would need to lose $2 \cdot \$200$ to have no money. If we manage to get p_k we can determine what’s the probability of losing the game aka achieving 2 : 1 after each trial. It is important to do it like this because the world is unfortunately not perfect and you will most definetely not win 100 throws consecutively.
(b)
$p_k = P(W|A \text{ starts at } k, \text{ wins round } 1) \cdot p + P(W|A \text{ starts at } k, \text{ loses round } 1) \cdot q$
$= P(W|A \text{ starts at } k+2) \cdot p + P(W|A \text{ starts at } k-1) \cdot q$

$$= p_{k+2} \cdot p + p_{k-1} \cdot q.$$

$$= p_{k+2} \cdot \frac{1}{2} + p_{k-1} \cdot \frac{1}{2}$$

set $p_i = x_i$ then

$$x^k = x^{k+2} \cdot \frac{1}{2} + x^{k-1} \cdot \frac{1}{2}$$

we divided by x_{k-1}

$$0 = x^3 \cdot \frac{1}{2} + x^0 \cdot \frac{1}{2} - x$$

$$0 = \frac{1}{2} \cdot x^3 - x + \frac{1}{2}$$

$$0 = x^3 - 2x + 1$$

$$0 = (x-1)(x^2+x-1)$$

solutions

$$x = 1$$
$$x = \frac{\sqrt{5}-1}{2}$$
$$x = \frac{-\sqrt{5}-1}{2}$$

We have 3 distinct roots so

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

and we know

$$p_k \to 0$$

as $k \to \infty$
this can only happen when

$$a = c = 0$$

and

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

we also know

$$p_k = 1$$

when $k = 0$ so

$$b = 1$$

therefore

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

(c) Since

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

for whichever k successes. Then the probability of ever having more than twice as many failures is that p_k but for all cases possible so.

$$p = \sum_{k=1}^{n} p_k$$

We start at 1 since you need to have at least one trial.

Chapter 4

Expectation

4.1 Expectations and variances

Exercise 3
(a) A fair die is rolled. Find the expected value of the roll.
(b) Four fair dice are rolled. Find the expected total of the rolls.

Solution.
(a) Let's call X the r.v for the value of the roll, then

$$E(X) = \sum_{x=1}^{6} xP(X = x)$$

$$= 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6}$$

$$= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5$$

(b) Let's call the events X_1, X_2, X_3, X_4 and $Y = X_1 + X_2 + X_3 + X_4$ then by the linearity of expectation

$$E(Y) = E(X_1) + E(X_2) + E(X_3) + E(X_4) = 14$$

Exercise 5
Find the mean and variance of a Discrete Uniform r.v. on $1, 2, \ldots, n$.
Hint: See the math appendix for some useful facts about sums.

Solution. I am supposing Blitz and Hwang mean that we have the set $1, 2, \ldots, n = C$. So we have an r.v $X \sim DUnif(C)$ then the mean should be

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

I used the Gauss sum to simplify this.
For the variance it would be

$$Var(X) = E(X^2) - (EX)^2$$

First we will calculate $E(X^2)$

$$E(X^2) = \sum_{x=1}^{n} x^2 P(X = x) = \frac{1}{n} \cdot \sum_{x=1}^{n} x^2$$

$$= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

$$= \frac{2n^2 + n + 2n + 1}{6} = \frac{2n^2 + 3n + 1}{6}$$

and since $(E(X))^2 = \frac{n^2+2n+1}{4}$ then

$$Var(X) = \frac{2n^2 + 3n + 1}{6} - \frac{n^2 + 2n + 1}{4}$$

$$= \frac{4n^2 + 6n + 2}{12} - \frac{3n^2 + 6n + 3}{12}$$

$$= \frac{n^2 - 1}{12}$$

Exercise 9
Consider the following simplified scenario based on Who Wants to Be a Millionaire?, a game show in which the contestant answers multiple-choice questions that have 4 choices per question. The contestant (Fred) has answered 9 questions correctly already, and is now being shown the 10th question. He has no idea what the right answers are to the 10th or 11th questions are. He has one "lifeline" available, which he can apply on any question, and which narrows the number of choices from 4 down to 2. Fred has the following options available.
(a) Walk away with \$16,000.
(b) Apply his lifeline to the 10th question, and then answer it. If he gets it wrong, he will leave with \$1,000. If he gets it right, he moves on to the 11th question. He then leaves with \$32,000 if he gets the 11th question wrong, and \$64,000 if he gets the 11th question right.
(c) Same as the previous option, except not using his lifeline on the 10th question, and instead applying it to the 11th question (if he gets the 10th question right).
Find the expected value of each of these options. Which option has the highest expected value? Which option has the lowest variance?

Solution. I'm supposing X will be the r.v that counts the amount of money Fred will win.
(a) Given that fred will walk away and this was decided before hadn

$$E(X) = 16000 \cdot 1 = 16000$$

Now we will do the Variance

$$Var(X) = E(X^2) - (E(X))^2$$

We will calculate $E(X^2)$

$$E(X^2) = 16000^2 \cdot 1 = 16000^2$$

so

$$Var(X) = 16000^2 - (16000)^2 = 0$$

We didn't really need to calculate it since we already know from a theorem that if you have an event X=x with probability 1 then you have a variance of 0.
(b) We will calculate the E(X), we have to remember that we are calculating the outcomes and their probability not the events

$$E(X) = 1000 \cdot \frac{1}{2} + 32000 \cdot \frac{3}{8} + 64000 \cdot \frac{1}{8}$$

$$= 500 + 12000 + 8000 = 20500$$

Now we will do the Variance

$$Var(X) = E(X^2) - (E(X))^2$$

We will calculate $E(X^2)$

$$E(X^2) = 1000^2 + 32000^2 \cdot \frac{3}{8} + 64000^2 \cdot \frac{1}{8}$$

$$= 500000 + 384000000 + 512000000 = 896500000$$

so

$$Var(X) = 896500000 - (20500)^2 = 476250000$$
$$\sqrt{Var(X)} = 21823.15284$$

(c) We will do it similarly

$$E(X) = E(X_{10Wrong}) + E(X_{10Right})$$

Remember Getting the 10th question right just means you go to the 11th.

$$= 1000 \cdot \frac{3}{4} + 32000 \cdot \frac{1}{8} + 64000 \cdot \frac{1}{8}$$
$$= 750 + 4000 + 8000 = 12750$$

Now we will do the Variance

$$Var(X) = E(X^2) - (E(X))^2$$

We will calculate $E(X^2)$

$$E(X^2) = 1000^2 \cdot \frac{3}{4} + 32000^2 \cdot \frac{1}{8} + 64000^2 \cdot \frac{1}{8}$$
$$= 750000 + 128000000 + 512000000 = 640750000$$

so

$$Var(X) = 896500000 - (12750)^2 = 478187500$$
$$\sqrt{Var(X)} = 21867.49871$$

Therefore the option with the best expectation for Fred is to use his lifeline on the 10th question which is option (b) and the worst option is (c)
The option with the biggest variance is (c) and the option with the smalles variance is (a).

THE NEXT SOLUTION IS WRONG, BUT I LEAVE IT HERE TO KNOW I MADE THIS MISTAKE
(b) We will calculate the E(X) in 10th given that he uses the lifeline and 11th separetely and by linearity we will add them together and get the actual one after

$$E(X)$$
$$= E(X_{10}|Get10thWrong) + E(X_{11}|Get11thWrong \cap Get10thRight) + E(X_{11}|Get11thRight \cap Get10thRight)$$

Since getting 11th wrong is independent of getting 10th right

$$= 1000 \cdot \frac{1}{2} + 32000 \cdot \frac{\frac{3}{4} \cdot \frac{1}{2}}{\frac{1}{2}} + 64000 \cdot \frac{\frac{1}{4} \cdot \frac{1}{2}}{\frac{1}{2}}$$
$$= 500 + 32000 \cdot \frac{\frac{3}{8}}{\frac{1}{2}} + 64000 \cdot \frac{\frac{1}{8}}{\frac{1}{2}}$$
$$= 500 + 32000 \cdot \frac{3}{4} + 64000 \cdot \frac{1}{4} = 40500$$

(C) We will calculate the E(X) in 10th and 11th separetely and by linearity we will add them together and get the actual one after. Remember than in this one Fred uses his lifeline on the 11th question.

$$E(X)$$
$$= E(X_{10}|Get10thWrong) + E(X_{11}|Get11thWrong \cap Get10thRight) + E(X_{11}|Get11thRight \cap Get10thRight)$$

Since getting 11th wrong is independent of getting 10th right

$$= 1000 \cdot \frac{3}{4} + 32000 \cdot \frac{\frac{1}{2} \cdot \frac{1}{4}}{\frac{1}{4}} + 64000 \cdot \frac{\frac{1}{2} \cdot \frac{1}{4}}{\frac{1}{4}}$$
$$= 750 + 32000 \cdot \frac{\frac{1}{8}}{\frac{1}{4}} + 64000 \cdot \frac{\frac{1}{8}}{\frac{1}{4}}$$
$$= 7500 + 32000 \cdot \frac{1}{2} + 64000 \cdot \frac{1}{2} = 48750$$

Exercise 11
Martin has just heard about the following exciting gambling strategy: bet \$1 that a fair coin will land Heads. If it does, stop. If it lands Tails, double the bet for the next toss, now betting \$2 on Heads. If it does, stop. Otherwise, double the bet for the next toss to \$4. Continue in this way, doubling the bet each time and then stopping right after winning a bet. Assume that each individual bet is fair, i.e., has an expected net winnings of 0. The idea is that

$$1 + 2 + 2^2 + 2^3 + \cdots + 2^n = 2^{n+1} - 1,$$

so the gambler will be \$1 ahead after winning a bet, and then can walk away with a profit.
Martin decides to try out this strategy. However, he only has \$31, so he may end upwalking away bankrupt rather than continuing to double his bet. On average, how much money will Martin win?

Solution. Ok so assuming each individual bet is fair we have identicall probabilities for winning $p = \frac{1}{2}$. Martin has the following expectation. MY GUESS BEFORE DOING THE EXERCISE IS THAT MARTINS STRATEGY WONT WORK. We will first do it taking into account that martin can do it for $n = 31$

$$E(X) = 1 \cdot \frac{1}{2} + -(1) \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} + -(2) \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{2} + -(2^2) \cdot \frac{1}{2} + \cdots + 2^{31} \cdot \frac{1}{2} + -(2^{31}) \cdot \frac{1}{2}$$

We can rearrange this equation and use the formula above

$$= 2^{32} - 1 \cdot \frac{1}{2} + -(2^{32} - 1) \cdot \frac{1}{2}$$

$$= \frac{1}{2}(2^{32} - 1 + -(2^{32} - 1))$$

$$= \frac{1}{2}(2^{32} - 1 - 2^{32} + 1)$$

$$= \frac{1}{2}(0)$$

$$= 0$$

The result for a limited number of times, say n would be the same. The result is like this because each individual bet remains fair. We might be changing the amount that we bet, but that doesn't change the game.

Exercise 15
Player A chooses a random integer between 1 and 100, with probability p_j of choosing j (for $j = 1, 2, \ldots, 100$). Player B guesses the number that player A picked, and receives from player A that amount in dollars if the guess is correct (and 0 otherwise).
(a) Suppose for this part that player B knows the values of p_j. What is player B's optimal strategy (to maximize expected earnings)?
(b) Show that if both players choose their numbers so that the probability of picking j is proportional to $1/j$, then neither player has an incentive to change strategies, assuming the opponent's strategy is fixed. (In game theory terminology, this says that we have found a Nash equilibrium.)
(c) Find the expected earnings of player B when following the strategy from (b). Express your answer both as a sum of simple terms and as a numerical approximation. Does the value depend on what strategy player A uses?

Solution.
(a) We want a way of choosing them that strikes the balance between a big number and a big probability so the best option since we already know the probability of all the numbers is to do

$$New_j = j \cdot p_j$$

for all j from 1 to 100 and then choosing the biggest number. This way we get the "real" value of each number, because even if 100 is very valuable, if it is going to happen with $p_{100} = 1/100$ then a 10 with probability p
(b) If both Player A and Player B follow the $1/j$ strategy the expectation for A should be

$$E(X_A) = \sum_{i,j} a_{i,j} P(A = i, B = j)$$

But since if $i \neq j$ we have $a_{i,j} = 0$ we will only count for $i = j$. Therefore, since if $i = j$ means $a_{i,j} = -j$ and since A and B choose their numbers independently

$$E(X_A) = \sum_{j=1}^{100} -j \cdot P(A = i, B = j) = \sum_{j=1}^{100} -j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})2$$

We are taking the square because they have the same strategy and $i = j$
for B it should be

$$E(X_B) = \sum_{i,j} a_{i,j} P(A = i, B = j)$$

But since if $i \neq j$ we have $a_{i,j} = 0$ we will only count for $i = j$. Therefore, since if $i = j$ means $a_{i,j} = j$ and since A and B choose their numbers independently

$$E(X_B) = \sum_{j=1}^{100} j \cdot P(A = i, B = j) = \sum_{j=1}^{100} j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})2$$

We are taking the square because they have the same strategy and $i = j$
We can see that the game is fair since adding both of their expectations we get 0, no one wins more than the other on average.

$$E(X) = E(X_A) + E(X_B) = \sum_{j=1}^{100} -j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2 + \sum_{j=1}^{100} j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2 = 0$$

But if A follows her strategy and B has j such that $p_j \neq 1/j$ then

$$E(X_A) = \sum_{j=1}^{100} -j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2$$

but

$$E(X_B) = \sum_{j=1}^{100} j \cdot (p(j))(\frac{1/j}{\sum_{n=1}^{100} 1/n})$$

$$\sum_{j=1}^{100} j \cdot (p(j))(\frac{1/j}{\sum_{n=1}^{100} 1/n})$$

$$= 1 \cdot p(1) \cdot \frac{\frac{1}{1}}{\sum_{n=1}^{100} 1/n} + 2 \cdot p(2) \cdot \frac{\frac{1}{2}}{\sum_{n=1}^{100} 1/n} + \ldots 99 \cdot p(99) \cdot \frac{\frac{1}{99}}{\sum_{n=1}^{100} 1/n} + 100 \cdot p(100) \cdot \frac{\frac{1}{100}}{\sum_{n=1}^{100} 1/n}$$

$$= \frac{1}{\sum_{n=1}^{100} \frac{1}{n}} (\sum_{j=1}^{100} j \cdot p(j) \cdot \frac{1}{j})$$

$$= \frac{1}{\sum_{n=1}^{100} \frac{1}{n}} (\sum_{j=1}^{100} p(j))$$

Since the probability should add up to 1 we get

$$= \frac{1}{\sum_{n=1}^{100} \frac{1}{n}}$$

so

$$E(X) = E(X_A) + E(X_B) = \sum_{j=1}^{100} -j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2 + \frac{1}{\sum_{n=1}^{100} \frac{1}{n}}$$

$$= \sum_{j=2}^{100} -j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2 + -1 \cdot (\frac{1}{\sum_{n=1}^{100} \frac{1}{n}}) + \frac{1}{\sum_{n=1}^{100} \frac{1}{n}}$$

$$= \sum_{j=2}^{100} -j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2 + 0$$

$$= \sum_{j=2}^{100} -j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2$$

As you may notice we will get a negative expectation that is lower than the 0 expectation we were getting before. Therefore, there is no incentive for B to change, he will a lower expectation. And in the same way we can prove that if B keeps its strategy but A changes it the game has no 0 expectation. (Doing it again would be kind of long)
(c)If B follows the strategy from part (b) then

$$E(X_B) = \sum_{j=1}^{100} j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2 = \sum_{j=1}^{100} j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2$$

$$= \sum_{j=1}^{100} j \cdot (\frac{1/j}{\sum_{n=1}^{100} 1/n})^2$$

```
(define (harmonic n) (harmonic-iter 0 1 n ))
(define (harmonic-iter a b count) (if (= count 0)
      a
      (harmonic-iter (+ a (/ 1 b)) (+ b 1) (- count 1))))
(define (Bstrategy n) (strat-iter 0 1 n ))
(define (strat-iter a b count) (if (= count 0)
      a
      (strat-iter (+ a (expt (/ (/ 1 b) (harmonic 100)) 2)) (+ b 1) (- count 1))))
> (harmonic 1)
1
> (harmonic 2)
1 1/2
> (harmonic 3)
1 5/6
> (harmonic 4)
2 1/12
> (harmonic 100)
5((522561233577855727 3)/(278881500918849908 6))
> (Bstrategy 100)
5050(15737709842901365080319781470470830294083004936076335246430028067364535951389801)/(259014313420
71092547751420114800805382909763373437683196479484957617374842467762 5)
```

By my very nice code in Scheme

$$E(X_B) \approx 5050 \cdot \frac{1573770984290136508031978147047083029408300493607633524643002806736453595138}{2590143134207109254775142011480080538290976337343768319647948495761737484 2467}$$

Some numbers wouldn't fit in the page so I removed them.
For the other part of the question **"Does the value depend on what strategy player A uses?"** yes it does because

$$E(X_B) = \sum_{j=1}^{100} j \cdot P(A = i, B = j)$$

for

$$i = j$$

you may notice that the Probability is

$$P(A = i, B = j)$$

. So for instance if A choosed only small numbers then the expectation would change even if B has the same strategy as always. We are supposed to match the numbers and since both people choose a number then changing the strategy changes the expectation.

Exercise 16
The dean of Blotchville University boasts that the average class size there is 20. But the reality experienced by the majority of students there is quite different: they find themselves in huge courses, held in huge lecture halls, with hardly enough seats or Haribo gummi bears for everyone. The purpose of this problem is to shed light on the situation. For simplicity, suppose that every student at Blotchville University takes only one course per semester.
(a) Suppose that there are 16 seminar courses, which have 10 students each, and 2 large lecture courses, which have 100 students each. Find the dean's-eye-view average class size (the simple average of the class sizes) and the student's-eye-view average class size (the average class size experienced by students, as it would be reflected by surveying students and asking them how big their classes are). Explain the discrepancy intuitively.
(b) Give a short proof that for any set of class sizes (not just those given above), the dean's-eye-view average class size will be strictly less than the student's-eye-view average class size, unless all classes have exactly the same size.
Hint: Relate this to the fact that variances are nonnegative.

Solution.
(a) The mean for the deans eye view is

$$\mu = \frac{16 \cdot 10 + 2 \cdot 100}{18} = \frac{160 + 200}{18} = \frac{360}{18} = 20$$

The student's-eye-view would be

$$E(X) = \sum^{16} (10 \cdot \frac{10}{360}) + \sum^{2} (100 \cdot \frac{100}{360})$$

$$= 16(0.27777777777) + 55.5555555556 = 4.44444444444 + 55.5555555556 = 60$$

(b) The deans average is the number of students n over the number of classes c

$$\frac{n}{c}$$

If for each i class we have n_i students then the student's-eye-view is

$$\sum_{i=1}^{c} n_i \cdot \frac{n_i}{n} = \sum_{i=1}^{c} \frac{n_i^2}{n}$$

We already know $c < n$ it wouldn't make any sense to have a class for each individual student, so we are assuming they have common sense. And since we know that we also know

$$\sum_{i=1}^{c} \frac{n_i^2}{n} < \sum_{i=1}^{c} \frac{n_i^2}{c}$$

and also

$$\frac{n}{n} < \frac{n}{c}$$

If we set a r.v X for the number of students in a class.

$$Var(X) = E(X - EX)^2$$

and since we know

$$0 \leq E(X - EX)^2$$

$$0 \leq \sum_i (n_i - \sum_j \frac{n_j^2}{n})^2 \frac{1}{n}$$

$$0 \leq \sum_i (n_i^2 - 2n_i \sum_j \frac{n_j^2}{n} + (\sum_j \frac{n_j^2}{n})^2)$$

$$0 \leq \sum_i (n_i^2 - \frac{2n_i}{n} \sum_j n_j^2 + (\sum_j \frac{n_j^2}{n})^2)$$

since

$$(\sum_j \frac{n_j^2}{n})^2 = (\frac{n_1^2}{n} + \frac{n_2^2}{n} + \frac{n_3^2}{n} + \cdots + \frac{n_c^2}{n})^2 = (\frac{1}{n} \cdot (n_1^2 + n_2^2 + n_3^2 + \cdots + n_c^2))^2 = \frac{1}{n^2}(n_1^2 + n_2^2 + n_3^2 + \cdots + n_c^2)^2$$

$$= \frac{1}{n^2}(\sum_j n_j^2)^2$$

we have

$$0 \leq \sum_i (n_i^2 - \frac{2n_i}{n} \sum_j n_j^2 + \frac{1}{n^2} \sum_j n_j^2)$$

to make our lives easier for the moment being we will set $A = \sum_j n_j^2$ so

$$0 \leq \sum_i (n_i^2 - \frac{2n_i}{n} A + \frac{1}{n^2} A^2)$$

$$0 \leq \sum_i (n_i^2 - \frac{2A}{n} n_i + \frac{A^2}{n^2})$$

$$0 \leq \sum_i n_i^2 - \frac{2A}{n} \sum_i n_i + \sum_i \frac{A^2}{n^2}$$

We know $\sum_i n_i = n$ **the sum of all classes gives the total number of students (no student is in two classes) I GUESS THIS IS KIND OF AN ASSUMPTION THAT SEEMS PLAUSIBLE FROM THE SETUP AND PREVIOUS RESULTS**

$$0 \leq \sum_i n_i^2 - \frac{2A}{n} \cdot n + \sum_i \frac{A^2}{n^2}$$

Remember that we are summing i or j a total c times therefore.

$$0 \leq \sum_i n_i^2 - \frac{2A}{n} \cdot n + c \cdot \frac{A^2}{n^2}$$

$$0 \leq \sum_i n_i^2 - 2A + c \cdot \frac{A^2}{n^2}$$

Notice $A = \sum_j n_j^2 = \sum_i n_i^2$ since the number of classes remain the same, therefore

$$0 \leq \sum_i n_i^2 - 2\sum_j n_j^2 + c \cdot \frac{A^2}{n^2}$$

$$0 \leq -\sum_j n_j^2 + c \cdot \frac{A^2}{n^2}$$

$$0 \leq -A + c \cdot \frac{A^2}{n^2}$$

multiplying by n

$$0 \leq -nA + c \cdot \frac{A^2}{n}$$

dividing by A

$$0 \leq -n + c \cdot \frac{A}{n}$$

$$-c \cdot \frac{A}{n} \leq -n$$

dividing by $-c$

$$\frac{A}{n} \geq \frac{n}{c}$$

substituting A

$$\sum_j \frac{n_j^2}{n} \geq \frac{n}{c}$$

We now notice that if all the classes have the same size $n = c \cdot n_j$ therefore

$$\sum_j \frac{n_j^2}{n} = \frac{\sum_j n_j^2}{n} = \frac{c \cdot n_j^2}{c \cdot n_j} = n_j$$

$$\frac{c \cdot n_j}{c} = n_j$$

Therefore

$$\sum_j \frac{n_j^2}{n} = \frac{n}{c}$$

when the class size is equal. Otherwise we shall have an strict inequality.

$$\sum_j \frac{n_j^2}{n} > \frac{n}{c}$$

Exercise 17
The sociologist Elizabeth Wrigley-Field posed the following puzzle [29]:

> American fertility fluctuated dramatically in the decades surrounding the Second World War. Parents created the smallest families during the Great Depression, and the largest families during the postwar Baby Boom. Yet children born during the Great Depression came from larger families than those born during the Baby Boom. How can this be?

(a) For a particular era, let n_k be the number of American families with exactly k children, for each $k \geq 0$. (Assume for simplicity that American history has cleanly been separated into eras, where each era has a well-defined set of families, and each family has a well-defined set of children; we are ignoring the fact that a particular family's size may change over time, that children grow up, etc.) For each

$j \geq 0$, let

$$m_j = \sum_{k=0}^{\infty} k^j n_k$$

For a family selected randomly in that era (with all families equally likely), find the expected number of children in the family. Express your answer only in terms of the m_j 's.
(b) For a child selected randomly in that era (with all children equally likely), find the expected number of children in the child's family, only in terms of the m_j's.
(c) Give an intuitive explanation in words for which of the answers to (a) and (b) is larger, or whether they are equal. Explain how this relates to the Wrigley-Field puzzle.

Solution.
(a)

$$E(X) = \frac{1}{\sum_{k=0}^{n} n_k} \cdot m_1 = \frac{m_1}{m_0}$$

(b)

$$E(X) = \frac{k}{m_1} \cdot m_1 = \frac{m_2}{m_1}$$

for k from 0 to infinity.
(c)The answer in b should be larger since choosing a child from a big family is more likely than choosing one from a smaller family. We are basically giving more weight to big families. But choosing any one family is as likely as choosing any other.
Now for the other part if in the Great depression we have the smallest families so that means we have possibly lots of families that are small. As for the Baby Boom we have less families with a large number of children. So now when we take them into account there are many more children, so long as n_k is big enough, that makes it so that the formula from (b) is bigger for the Great Depression than the Baby Boom.

4.2 Named distributions

Exercise 26
Nick and Penny are independently performing independent Bernoulli trials. For concreteness, assume that Nick is flipping a nickel with probability p_1 of Heads and Penny is flipping a penny with probability p_2 of Heads. Let $X_1, X_2, \ldots$ be Nick's results and $Y_1, Y_2, \ldots$ be Penny's results, with $X_i \sim Bern(p_1)$ and $Y_j \sim Bern(p_2)$.
(a) Find the distribution and expected value of the first time at which they are simultaneously successful, i.e., the smallest n such that $X_n = Y_n = 1$. Hint: Define a new sequence of Bernoulli trials and use the story of the Geometric.
(b) Find the expected time until at least one has a success (including the success). Hint: Define a new sequence of Bernoulli trials and use the story of the Geometric.
(c) For $p_1 = p_2$, find the probability that their first successes are simultaneous, and use this to find the probability that Nick's first success precedes Penny's.

Solution.
(a) So the probability of them being simultaneaoulsy succesful is $p_1 \cdot p_2$ because the nickel throw is independent. Therefore we define a new experiment in which they throw two different coins, bernoulli trials, multiple times and we have a $p = p_1 \cdot p_2$ probability of success and $q = 1 - p = 1 - p_1 \cdot p_2$. And using the story of the Geometric we can find the distribution of the first time were they are both succesful simultaneously. It is

$$X \sim Geom(p)$$

$$P(X = k) = q^k p$$

$$P(X = k) = (1 - p_1 \cdot p_2)^k (p_1 \cdot p_2)$$

The expectation should be

$$E(X) = \frac{1 - p_1 p_2}{p_1 p_2}$$

which is the expected value of the first time at which they are simultaneously successful, i.e., the smallest n such that $X_n = Y_n = 1$.
(b) We define a new Bernoulli Trial where the succes is having at least one success that is the sum of the independent events of Penny having one success or Nick having one success. Therefore the probability of success is

$$p = p_1 \cdot q_2 + p_2 \cdot q_1 + p_1 \cdot p_2$$

we will now define a geometric distribution using this

$$X \sim Geom(p)$$

and then the first sucess distribution to get the actual distribution

$$Y = X + 1 \sim FS(p)$$

and we get its expectation

$$E(Y) = E(X + 1) = \frac{q}{p} + 1 = \frac{1}{p} = \frac{1}{p_1 \cdot q_2 + p_2 \cdot q_1 + p_1 \cdot p_2}$$

which is the expected time until at least one has a success (including the success).
(c) We know that we have 2 random variables X, Ythat follow a $Geom(p_1)$ or $Geom(p_2)$, doesn't matter as $p_1 = p_2$, distribution.
We now need to get

$$P(X = Y)$$

$$\sum_{k=0}^{\infty} P(X = k|Y = k)P(Y = k)$$

and since X and Y are independent

$$\sum_{k=0}^{\infty} P(X = k)P(Y = k)$$

$$\sum_{k=0}^{\infty} (q^k p_1)^2$$

$$p_1^2 \sum_{k=0}^{\infty} q^{2k}$$

$$p_1^2 \frac{1}{1 - q^2}$$

$$\frac{p_1^2}{1 - q^2}$$

and since $q = 1 - p_1$

$$\frac{p_1^2}{1 - (1 - p_1)^2}$$

and

$$P(X < Y)$$

(We are setting X for Nick and Y for Penny) we know $1 = P(X < Y) + P(X = Y) + P(Y < X)$ and $P(X < Y) = P(Y < X)$ so

$$1 = P(X < Y) + P(X = Y) + P(Y < X)$$

$$1 = 2P(X < Y) + P(X = Y)$$

$$-2P(X < Y) = P(X = Y) - 1$$
$$2P(X < Y) = 1 - P(X = Y)$$
$$P(X < Y) = \frac{1 - P(X = Y)}{2}$$

Exercise 27
Let X and Y be Pois(λ) r.v.s,and $T = X + Y$.Suppose that X and Y are not independent, and in fact $X = Y$. Prove or disprove the claim that $T \sim$Pois(2λ) in this scenario

Solution. We will compare the PMF of $T \sim Pois(2\lambda)$ with $P(T = \frac{k}{2})$ which is the result of $T = X + Y$ where $X = Y$. The former gives

$$P(T = k) = \frac{e^{-2\lambda}2\lambda^k}{k!}$$

The latter gives

$$P(T = \frac{k}{2}) = \frac{e^{-\lambda}\lambda^{k/2}}{\frac{k}{2}!}$$

we see that for any k which is an odd positive integer we will have 0 by definition, because there cannot be any fraction in the k it doesn't make sense.
But this does not happen in $T \sim Pois(2\lambda)$ so they are not the same.

Exercise 30
(a) Use LOTUS to show that for $X \sim$ Pois(λ) and any function g,

$$E(Xg(X)) = \lambda E(g(X + 1))$$

,assuming that both sides exist. This is called the Stein-Chen identity for the Poisson.
(b) Find the third moment E(X_3) for $X \sim Pois(\lambda)$ by using the identity from (a) and a bit of algebra to reduce the calculation to the fact that X has mean λ and variance λ.

Solution.
(a)

$$E(Xg(X)) = \sum k = 0^\infty kg(k)P(X = k) = \sum k = 0^\infty kg(k)\frac{\lambda^k e^{-\lambda}}{k!} = \lambda \sum_{k=0}^{\infty} g(k)\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!}$$

on the other side

$$\lambda E(g(X + 1)) = \lambda \sum_{k=0}^{\infty} g(k)P(X + 1 = k) = \lambda \sum_{k=0}^{\infty} g(k)P(X = K - 1) = \lambda \sum_{k=0}^{\infty} g(k)\frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!}$$

(b)

$$E(X \cdot X^2) = \lambda E((X + 1)^2)$$

by (a)

$$= \lambda E((X + 1)^2) = \lambda E(X^2 + 2X + 1)$$
$$= \lambda(E(X^2) + E(2X) + E(1))$$

by linearity of Expectation but we know $E(X) = \lambda$ and $\lambda+(EX)^2 = E(X^2)$ from the fact that $Var(X) = \lambda$ so

$$= \lambda(\lambda + (EX)2 + 2E(X) + 1)$$
$$= \lambda(\lambda + \lambda^2 + 2\lambda + 1)$$
$$= \lambda^3 + 3\lambda^2 + \lambda$$

Exercise 31
In many problems about modeling count data, it is found that values of zero in the data are far more common than can be explained well using a Poisson model (we can make $P(X = 0)$ large for $X \sim$Pois(λ) by making λ small, but that also constrains the mean and variance of X to be small since both are λ). The Zero-Inflated Poisson distribution is a modification of the Poisson to address this issue, making it easier to handle frequent zero values gracefully.
A Zero-Inflated Poisson r.v. X with parameters p and λ can be generated as follows. First flip a coin with probability of p of Heads. Given that the coin lands Heads, $X = 0$. Given that the coin lands Tails, X is distributed Pois(λ). Note that if $X = 0$ occurs, there are two possible explanations: the coin could have landed Heads (in which case the zero is called a structural zero), or the coin could have landed Tails but the Poisson r.v. turned out to be zero anyway. For example, if X is the number of chicken sandwiches consumed by a random person in a week, then $X = 0$ for vegetarians (this is a structural zero), but a chicken-eater could still have $X = 0$ occur by chance (since they might happen not to eat any chicken sandwiches that week).
(a) Find the PMF of a Zero-Inflated Poisson r.v. X.
(b) Explain why X has the same distribution as $(1 - I)Y$, where $I \sim$ Bern(p) is independent of $Y \sim$ Pois(λ).
(c) Find the mean of X in two different ways: directly using the PMF of X, and using the representation from (b). For the latter, you can use the fact (which we prove in Chapter 7) that if r.v.s Z and W are independent, then $E(ZW) = E(Z)E(W)$.
(d) Find the variance of X.

Solution.
(a) We set an event H where the coin lands heads and H^c if the coin lands tails with $P(H) = p$

$$P(X = 0) = P(X = 0|H)P(H) + P(X = 0|H^c)P(H^c) = 1 \cdot p + \frac{e^{-\lambda}\lambda^0}{0!} \cdot (1 - p)$$

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

as for $k \neq 0$

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}(1 - p)$$

so the PMF for all k is

$$P(X = k) = P(k = 0) + P(k \neq 0) = p + e^{-\lambda} \cdot (1 - p) + \frac{e^{-\lambda}\lambda^k}{k!}(1 - p)$$

let's see if it is valid, it is already positive so we test if it is equal to 1 when added

$$\sum_{k=0}^{\infty} P(X = k) = p + e^{-\lambda} \cdot (1 - p) + \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}(1 - p)$$

$$= p + e^{-\lambda} \cdot (1 - p) + \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}(1 - p)$$

$$= p + (1 - p)(e^{-\lambda} + \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}) = 1$$

(b) That is because we are doing a Bernoulli to know if it landed heads or tails and we are also doing a poisson distribution if it landed tails.
(c) Method 1:

$$\sum_{k=0}^{\infty} kP(X = k) = 0(p + e^{-\lambda} \cdot (1 - p)) + \sum_{k=1}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!}(1 - p)$$

$$= \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!}(1-p)$$
$$= \lambda(1-p)$$

Method 2: We use the fact that if r.v.s Z and W are independent, then $E(ZW) = E(Z)E(W)$.

$$E((1-I)Y) = E(1-I)E(Y)$$

we already know $E(Y) = \lambda$ as for $E(1-I)$ we will calculate it

$$E(1-I) = E(1) - E(I) = 1-p$$

so

$$E((1-I)Y) = E(1-I)E(Y) = (1-p)\lambda = \lambda - \lambda p$$

(d)

$$Var(X) = E(X^2) - (EX)^2$$

we will first calculate $E(X^2)$

$$E(X^2) = 0^2(p + e^{-\lambda} \cdot (1-p)) + \sum_{k=1}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!}(1-p)$$
$$= \sum_{k=1}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!}(1-p)$$
$$= (1-p)e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!}$$
$$= (1-p)e^{-\lambda}e^{\lambda}\lambda(1+\lambda)$$
$$= (1-p)\lambda(1+\lambda)$$
$$Var(X) = E(X^2) - (EX)^2 = (1-p)\lambda(1+\lambda) - (\lambda(1-p))^2$$
$$= \lambda + \lambda^2 - \lambda p - \lambda^2 p - (\lambda^2 - 2\lambda p + \lambda^2 p^2)$$
$$= \lambda + \lambda^2 - \lambda p - \lambda^2 p - \lambda^2 + 2\lambda p - \lambda^2 p^2$$
$$= \lambda - \lambda^2 p + \lambda p - \lambda^2 p^2$$
$$= \lambda(1+p) - \lambda^2(1+p^2)$$

4.3 Indicator r.v.s

Exercise 38
Each of $n \geq 2$ people puts their name on a slip of paper (no two have the same name). The slips of paper are shuffled in a hat, and then each person draws one (uniformly at random at each stage, without replacement). Find the average number of people who draw their own names.

Solution. Setting X to be the number of matches A_i to be the event of person i having a match we have $X = A_1 + A_2 + \cdots + A_n$ so

$$E(X) = E(I_{A_1}) + E(I_{A_2}) + \cdots + E(I_{A_n}) = P(A_1) + P(A_2) \cdots + P(A_n)$$

by definition. Then we know that the probability of each indiviual match is $\frac{1}{n}$ so

$$= n \cdot \frac{1}{n} = 1$$

the number of expecter matches is 1

Exercise 45
Show that for any events $A_1, \ldots, A_n$,

$$P(A_1 \cap A_2 \cdots \cap A_n) \geq \sum_{j=1}^{n} P(A_j) - n + 1$$

. Hint: First prove a similar-looking statement about indicator r.v.s, by interpreting what the events $I(A_1 \cap A_2 \cdots \cap A_n) = 1$ and $I(A_1 \cap A_2 \cdots \cap A_n) = 0$ mean.

Exercise 47
You are being tested for psychic powers. Suppose that you do not have psychic powers. A standard deck of cards is shuffled, and the cards are dealt face down one by one. Just after each card is dealt, you name any card (as your prediction). Let X be the number of cards you predict correctly. (See Diaconis [5] for much more about the statistics of testing for psychic powers.)
(a) Suppose that you get no feedback about your predictions. Show that no matter what strategy you follow, the expected value of X stays the same; find this value. (On the other hand, the variance may be very different for different strategies. For example, saying "Ace of Spades" every time gives variance 0.) Hint: Indicator r.v.s.
(b) Now suppose that you get partial feedback: after each prediction, you are told immediately whether or not it is right (but without the card being revealed). Suppose you use the following strategy: keep saying a specific card's name (e.g., "Ace of Spades") until you hear that you are correct. Then keep saying a different card's name (e.g., "Two of Spades") until you hear that you are correct (if ever). Continue in this way, naming the same card over and over again until you are correct and then switching to a new card, until the deck runs out. Find the expected value of X, and show that it is very close to $e - 1$. Hint: Indicator r.v.s.
(c) Now suppose that you get complete feedback: just after each prediction, the card is revealed. Call a strategy "stupid" if it allows, e.g., saying "Ace of Spades" as a guess after the Ace of Spades has already been revealed. Show that any non-stupid strategy gives the same expected value for X; find this value. Hint: Indicator r.v.s.

Solution. (a) We have X which is the addition of the individuals matches A_i for $1 \leq i \leq 52$. So

$$E(X) = E(I_{A_1}) + E(I_{A_2}) + \cdots + E(I_{A_{52}}) = P(A_1) + P(A_2) \cdots + P(A_{52})$$

by definition. Then we know that the probability of each indiviual match is $\frac{1}{52}$, no matter the strategy as we are considering individual matches, so

$$= 52 \cdot \frac{1}{52} = 1$$

the number of expected matches is 1
(b) We have X which is the addition of the individuals matches, $P(A_1)$ indicates having one match and so on, A_i for $1 \leq i \leq 52$. So

$$E(X) = E(I_{A_1}) + E(I_{A_2}) + \cdots + E(I_{A_{52}}) = P(A_1) + P(A_2) \cdots + P(A_{52})$$

by definition. We notice that there are 52! ways in which we can have 1 match and 51! in which we can have 2 and so on. therefore

$$P(A_1) + P(A_2) \cdots + P(A_{52}) = \frac{1}{52!} + \frac{1}{51!} + \ldots \frac{1}{2!} + 1$$

we notice this is the same as the taylor series for e minus 1.
(c) Since knowing that you are wrong won't actually change anything while it comes to the individual probability of getting one or two right. The Expectation is the same as in (b) which is $e - 1$.

Exercise 58
Job candidates $C_1, C_2, \ldots$ are interviewed one by one, and the interviewer compares them and keeps an updated list of rankings (if n candidates have been interviewed so far, this is a list of the n candidates, from best to worst). Assume that there is no limit on the number of candidates available, that for any n the candidates $C_1, C_2, \ldots, C_n$are equally likely to arrive in any order, and that there are no ties in the rankings given by the interview.
Let X be the index of the first candidate to come along who ranks as better than the very first candidate C_1 (so C_X is better than C_1, but the candidates after 1 but prior to X (if any) are worse than C_1. For example, if C_2 and C_3 are worse than C_1 but C_4 is better than C_1, then $X = 4$. All 4! orderings of the first 4 candidates are equally likely, so it could have happened that the first candidate was the best out of the first 4 candidates, in which case $X > 4$.
What is $E(X)$ (which is a measure of how long, on average, the interviewer needs to wait to find someone better than the very first candidate)?
Hint: Find $P(X > n)$ by interpreting what $X > n$ says about how C_1 compares with other candidates, and then apply the result of Theorem 4.4.8.

Solution. So we will first find $P(X > n)$ which basically means what's the probability that the firt candidate is the best among the first n. We see that they can come in any random orther the n candidates so the best candidate can only be placed in n ways therefore.

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

Now, we see from Theorem 4.4.8 that the expectation via survival function is

$$E(X) = \sum_{n=1}^{\infty} P(X > n) = \sum_{n=1}^{\infty} \frac{1}{n}$$

We know this is the harmonic series which diverges if the terms are infinite. So the time is inifinite if we are assuming we have infinite candidates, which kind of makes sense if there is always someone better. If there is a limited number of candidates then it is different, the expectation converges and we have a limited amount of time.

Exercise 59
People are arriving at a party one at a time. While waiting for more people to arrive they entertain themselves by comparing their birthdays. Let X be the number of people needed to obtain a birthday match, i.e., before person X arrives there are no two people with the same birthday, but when person X arrives there is a match.
Assume for this problem that there are 365 days in a year, all equally likely. By the result of the birthday problem from Chapter 1, for 23 people there is a 50.7% chance of a birthday match (and for 22 people there is a less than 50% chance). But this has to do with the median of X (defined below); we also want to know the mean of X, and in this problem we will find it, and see how it compares with 23.
(a) A median of a random variable Y is a value m for which $P(Y \leq m) \geq 1/2$ and $P(Y \geq m) \geq 1/2$ (this is also called a median of the distribution of Y ; note that the notion is completely determined by the CDF of Y). Every distribution has a median, but for some distributions it is not unique. Show that 23 is the unique median of X.
(b) Show that $X = I_1 + I_2 + \cdots + I_{366}$, where I_j is the indicator r.v. for the event $X \geq j$. Then find E(X)in terms of pj's defined by $p_1 = p_2 = 1$ and for $3 \leq j \leq 366$,

$$p_j = (1 - \frac{1}{365})(1 - \frac{2}{365}) \ldots (1 - \frac{j-2}{365})$$

(c) Compute $E(X)$ numerically. In R, the pithy command cumprod(1-(0:364)/365) produces the vector $(p_2, \ldots, p_{366})$.
(d) Find the variance of X, both in terms of the pj's and numerically. Hint: What is I_i^2, and what is

I_iI_j for $i < j$? Use this to simplify the expansion

$$X^2 = I_1^2 + \cdots + I_{366}^2 + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1} I_iI_j$$

Solution. (a)Let's think about it this way. We have no matches till we get to person k. So how can we arrange the people before the person k so that we get no matches. The first person has 365 available spaces the second one 364 and so on, we divide this numer by 365 since they are the total number of spaces (days). From this we have that

$$P(X = k) = (\prod_{n=0}^{k-2} \frac{365-n}{365}) \cdot \frac{k-1}{365}$$

The multiplication at the end is because the last person, person k has to be in one of the k-1 days of the previous people.
We can simplify that into

$$= \frac{365!}{(365-(k-3))!365^{k-1}} \cdot \frac{k-1}{365}$$

Now we get the CDF

$$P(X \leq m) = \sum_{k=0}^{m} (\frac{365!}{(365-(k-3))!365^{k-1}} \cdot \frac{k-1}{365})$$

$$= \sum_{k=0}^{m} (\frac{365!}{(365-(k-3))!365^{k}} \cdot k - 1)$$

and we have to find the m for which we have

$$P(X \leq m) \geq \frac{1}{2}$$

For this we see that we need to know when

$$\sum_{k=0}^{m} (\frac{365!}{(365-(k-3))!365^{k}} \cdot k - 1) \geq \frac{1}{2}$$

We also find that since you cannot divide a person, aka this is a discrete distribution.

$$P(X < m) = \sum_{k=0}^{m-1} (\frac{365!}{(365-(k-3))!365^{k}} \cdot k - 1)$$

Therefore we can get

$$P(X \geq m) = 1 - P(X < m) = 1 - \sum_{k=0}^{m-1} (\frac{365!}{(365-(k-3))!365^{k}} \cdot k - 1) \geq \frac{1}{2}$$

Well apparently it is 23, i'll later paste the code here.

```
# This one is P(X  <=23)
total3 = 1
for (k in 0:23) {
  term <-(365 - k)/365
  total3 <- total3 * term
}
psecond2=1-total3
print(psecond2)
print(psecond2>=1/2)
```

```
# P(X>=23)=P(X>22)
total4 = 1
for (k in 0:21) {
  term <-(365 - k)/365
  total4 <- total4 * term
}
print(total4)
print(total4>=1/2)

[1] 0.5383443
[1] TRUE
[1] 0.5243047
[1] TRUE
```

(b) We can think of

$$X = I_1 + I_2 + \cdots + I_{366}$$

as follows, X indicates the number of people needed to obtain a birthday match so each I_j in which $X \geq j$ would basically indicate that in the past $j-1$ we haven't gotten a match and then in the j person we got a match. We also do till I_{366} as that is the smallest number for which we have probaility 1 that there is a match for certain, by pigeon-hole principle.
Now we will do

$$E(X) = E(I_1) + E(I_2) + \cdots + E(I_366)$$

by the fundamental bridge we have

$$= P(I_1) + P(I_2) + \cdots + P(I_{366})$$

$$= 1 + 1 + p_3 + \cdots + p_{366}$$

$$= 1 + 1 + \frac{364}{365} + \frac{364}{365}\frac{363}{365} + \dots$$

$$= 1 + 1 + \sum_{k=3}^{366} \frac{365!}{365^{j-1} \cdot (365 - (j-1))!}$$

(c)

```
actual=0
for (k in 3:366) {
  totalF=1
    for (j in 1:(k-2)) {
    tempterm <-(365 - j)/365
    totalF <- totalF * tempterm
    }
  actual=actual+totalF
}
print(actual+2)
[1] 24.61659
```

(d) We have

$$X^2 = I_1^2 + \cdots + I_{366}^2 + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1} I_i I_j$$

and since we know $E(I^2) = E(I)$ by LOTUS

$$E(X^2) = E(I_1) + \cdots + E(I_{366}) + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1} E(I_i I_j)$$

Since $j > i$ for the $E(I_i I_j)$ we have$P(X \geq j)$ since both of the events are not separate and the probability of $X \geq j > i$ is the same as the probability of $X \geq j$ therefore

$$= E(I_1) + \cdots + E(I_{366}) + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1} P(X \geq j)$$

Since we already got the first part in part (c) we write

$$= 24.61659 + 2\sum_{j=2}^{366}\sum_{i=1}^{j-1} P(X \geq j)$$

$$= 24.61659 + 2\sum_{j=2}^{366}[(j-1) \cdot P(X \geq j)]$$

We will do the calculation with R now

```
    actual=0
for (k in 3:366) {
  totalF=1
    for (j in 1:(k-2)) {
    tempterm <-(365 - j)/365
    totalF <- totalF * tempterm
    }
  totalF=totalF*(k-1)
  actual=actual+totalF
}
actual=(actual+1) #We add one because we need to start from 2
actual=(actual*2) #This because the sum is mulitplied by 2
actual=actual+24.61659 #This because of the formula
print(actual)
[1] 754.6166
```

Therefore

$$= 24.61659 + 730 = 754.6166$$

And now we can do the Variance

$$Var(X) = E(X^2) - E(X)^2 = 754.6166 - (24.61659)^2 = 754.6166 - 605.9765 = 148.6401$$

SUCCESSSSSSSSS

Exercise 60

Elk dwell in a certain forest. There are N elk, of which a simple random sample of size n is captured and tagged (so all $\binom{N}{n}$ sets of n elk are equally likely). The captured elk n are returned to the population, and then a new sample is drawn. This is an important method that is widely used in ecology, known as capture-recapture. If the new sample is also a simple random sample, with some fixed size, then the number of tagged elk in the new sample is Hypergeometric.

For this problem, assume that instead of having a fixed sample size, elk are sampled one by one without replacement until m tagged elk have been recaptured, where m is specified in advance (of course, assume that $1 \leq m \leq n \leq N$). An advantage of this sampling method is that it can be used to avoid ending up with a very small number of tagged elk (maybe even zero), which would be problematic in many applications of capture-recapture. A disadvantage is not knowing how large the

sample will be.
(a) Find the PMFs of the number of untagged elk in the new sample (call this X) and of the total number of elk in the new sample (call this Y).
(b) Find the expected sample size EY using symmetry, linearity, and indicator r.v.s.
(c) Suppose that m,n,N are such that EY is an integer. If the sampling is done with a fixed sample size equal to EY rather than sampling until exactly m tagged elk are obtained, find the expected number of tagged elk in the sample. Is it less than m, equal to m, or greater than m (for $n < N$)?

Solution.
(a)

$$P(Y = k) = \binom{k-1}{m-1} \frac{(N-n)^{(k-m)} n^{(m)}}{N^{(k)}}$$

Since $X = Y - m$ then

$$P(X = k - m)$$

(b) For this part we will set an indicator r.v at each step that indicates if we got a tagged elk or not as such

$$Y = I_1 + I_2 + \cdots + I_m$$

so

$$E(Y) = E(I_1) + E(I_2) + \cdots + E(Y_m)$$

The probability of sucess at each step is $\frac{n}{N}$ therefore

$$E(Y) = \frac{n}{N} + \frac{n}{N} + \cdots + \frac{n}{N} = \frac{m \cdot n}{N}$$

For this part that we have done remember that we are not doing conditional probability. The probability of a tagged elk being on the first one and the last one is technically the same, it can be affected by what we have drawn before but we do not take into account that. And since we can find the last of the tagged elks at the end of the entire forest we have to do this till N.
(c) Similarly as above with $A = \frac{m \cdot n}{N}$

$$E(Z) = E(I_1) + E(I_2) + \cdots + E(Y_A)$$

at each step we have probability $\frac{n}{N}$ as above and therefore

$$E(Z) = \frac{n}{N} + \frac{n}{N} + \cdots + \frac{n}{N} = A \cdot \frac{n}{N} = m \cdot (\frac{n}{N})^2$$

Since we have that $n < N$ we see that

$$(\frac{n}{N})^2$$

is smaller than 1 and therefore when we multiply

$$(\frac{n}{N})^2$$

by m we get something smaller than m therefore $E(Z) < m$
It is kind of interesting that the average of the process, sampling until we get m tagged elks, will give us less tagged elks than what we need m if we render it as our limit of sampling.

4.4 LOTUS

Exercise 61
For $X \sim Pois(\lambda)$, find $E(X!)$ (the average factorial of X), if it is finite.

Solution.

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

$$= e^{-\lambda} \sum_{k=0}^{\infty} \lambda^k$$

using the geometric sereies we simplify to

$$= e^{-\lambda} \cdot \frac{1}{1-\lambda}$$

Note that this only applies if $|\lambda| < 1$

$$= \frac{e^{-\lambda}}{1-\lambda}$$

Exercise 62
For $X \sim Pois(\lambda)$, find $E(2^X)$, if it is finite.

Solution.

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

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!}$$

Apparently the sum is the same as $e^{2\lambda}$, it is a Taylor series. Here there is an explanation for that

$$= e^{-\lambda} e^{2\lambda} = e^{\lambda}$$

I accidentaly made a typo and did $E(2X)$ instead of $E(2^X)$. I will leave this solution here still.

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

$$= 2e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!}$$

$$= 2\lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

$$= 2\lambda e^{-\lambda} e^{\lambda}$$

$$= 2\lambda$$

Exercise 63
For $X \sim Geom(p)$, find $E(2X)$ (if it is finite) and $E(2-X)$ (if it is finite). For each, make sure to clearly state what the values of p are for which it is finite.

Solution.

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

The sum was simplified using the explanation from page 160
Or we can also just see that by the laws of expectation we just multiply $E(X)$ by 2 and get $E(2X)$

$$\begin{aligned}
E(2-X) &= \sum_{k=0}^{\infty}(2-k)q^k p = \sum_{k=0}^{\infty} 2q^k p - \sum_{k=0}^{\infty} kq^k p \\
&= 2p\sum_{k=0}^{\infty} q^k - \sum_{k=0}^{\infty} kq^k p \\
&= 2p\frac{1}{1-q} - p\sum_{k=0}^{\infty} kq^k \\
&= \frac{2p}{1-q} - \frac{q}{p}
\end{aligned}$$

Exercise 66
Let X be a $Pois(\lambda)$ random variable, where λ is fixed but unknown. Let $\theta = e^{-3\lambda}$, and suppose that we are interested in estimating θ based on the data. Since X is what we observe, our estimator is a function of X, call it $g(X)$. The bias of the estimator $g(X)$ is defined to be $E(g(X)) - \theta$, i.e., how far off the estimate is on average; the estimator is unbiased if its bias is 0.
(a) For estimating λ, the r.v. X itself is an unbiased estimator. Compute the bias of the estimator $T = e^{-3X}$. Is it unbiased for estimating θ?
(b) Show that $g(X) = (-2)^X$ is an unbiased estimator for θ. (In fact, it turns out to be the only unbiased estimator for θ.)
(c) Explain intuitively why g(X) is a silly choice for estimating θ, despite (b), and show how to improve it by finding an estimator h(X) for θ that is always at least as good as g(X) and sometimes strictly better than $g(X)$. That is,

$$|h(X) - \theta| \leq |g(X) - \theta|,$$

with the inequality sometimes strict.

Solution.
(a)

$$\begin{aligned}
E(T) - \theta &= E(e^{-3X}) - \theta \\
&= \sum_{k=0}^{\infty} e^{-3k}\frac{e^{-\lambda}\lambda^k}{k!} - e^{-3\lambda} \\
&= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^{-3}\lambda)^k}{k!} - e^{-3\lambda}
\end{aligned}$$

We use taylor series to simplify to

$$= e^{-\lambda}e^{e^{-3\lambda}} - e^{-3\lambda}$$

We can clearly see this is not 0 so it is a biased estimator
(b)

$$\begin{aligned}
E(g(X)) - \theta &= E((-2)^X) - \theta \\
&= \sum_{k=0}^{\infty}(-2)^k\frac{e^{-\lambda}\lambda^k}{k!} - e^{-3\lambda}
\end{aligned}$$

$$= e^{-\lambda}\sum_{k=0}^{\infty}(-2)^k\frac{\lambda^k}{k!} - e^{-3\lambda}$$

$$= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(-2\lambda)^k}{k!} - e^{-3\lambda}$$

using Taylor Series

$$= e^{-\lambda}e^{-2\lambda} - e^{-3\lambda}$$

$$e^{-3\lambda} - e^{-3\lambda} = 0$$

Since

$$E(g(X)) - \theta = 0$$

$g(X)$ is an unbiased estimator
(c) The problem is that $g(X) = (-2)^X$ can give negative values of enormous magnitude and $\theta = e^{-3\lambda}$ is entirely positive. It just so happen that $g(X)$ is on average following θ. I'm guessing g(X) as a function goes supper positive and super negative and in the middle when we add both and average both sides of the function we have θ.
Using a graphic calculator I realized that modifying the original estimator that we have we get the same graph up to a certain point

$$|(-2)^{-x}|$$

4.5 Poisson approximation

Exercise 68
A group of n people play "Secret Santa" as follows: each puts their name on a slip of paper in a hat, picks a name randomly from the hat (without replacement), and then buys a gift for that person. Unfortunately, they overlook the possibility of drawing one's own name, so some may have to buy gifts for themselves (on the bright side, some may like self-selected gifts better). Assume $n \geq 2$.
(a) Find the expected value of the number X of people who pick their own names.
(b) Find the expected number of pairs of people, A and B, such that A picks B's name and B picks A's name (where $A \neq B$ and order doesn't matter).
(c) What is the approximate distribution of X if n is large (specify the parameter value or values)? What does $P(X = 0)$ converge to as $n \to \infty$?

Solution.
(a) We can see that X is a sum of indicators of each person from 1 to n getting their own gift, therefore.

$$E(X) = E(I_1) + E(I_2) + \cdots + E(I_n) = \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = \frac{n}{n} = 1$$

(b) We will call this Z and we see that Z can be expressed as the sum of each of all the exisitng pairs of being in the same pair.

$$E(X_{xy}) = \sum_{xy} E(I_{xy}) = \binom{n}{2}E(I_{xy})$$

For this probability we can think about it this way x and y can be literally anyone and can be anywhere in the list of people that are selected, but they cannot be in their original position. And that's how we will select each match

$$E(X_{xy}) = \binom{n}{2}\frac{(n-2)!}{n!} = \frac{n!}{2!\cdot(n-2)!}\cdot\frac{(n-2)!}{n!} = \frac{1}{2!} = \frac{1}{2}$$

(c) As we saw in (a) we have $E(X) = 1$ so we can estimate X by Pois(1) when n is large. And from there by deifniton we can see that

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

Exercise 70
Ten million people enter a certain lottery. For each person, the chance of winning is one in ten million, independently.
(a) Find a simple, good approximation for the PMF of the number of people who win the lottery.
(b) Congratulations! You won the lottery. However, there may be other winners. Assume now that the number of winners other than you is $W \sim Pois(1)$, and that if there is more than one winner, then the prize is awarded to one randomly chosen winner. Given this information, find the probability that you win the prize (simplify).

Solution.
(a)

$$P(X = k) = (\frac{1}{10^7})^k$$

And the approximation is $X \sim Pois(1)$

$$P(X = k) = \frac{e^{-1}}{k!} = \frac{1}{e \cdot k!}$$

We put $\lambda = 1$ because the expected number of people winning is

$$E(X) = \frac{10^7}{10^7}$$

. We see this with $X = I_1 + I_2 + \cdots + I_{10^7}$ where each person has a probability of sucess of 1 in 10 million.
(b)Let W be the event of me winning the lottery and let O be the even of there being others then

$$P(W) = \sum_{k=0}^{\infty} \frac{1}{k+1} \cdot \frac{1}{e \cdot k!} = \frac{1}{e} \sum_{k=0}^{\infty} \frac{1}{k+1} \cdot \frac{1}{k!} = \frac{1}{e} \sum_{k=0}^{\infty} \frac{1}{(k+1)!}$$

We notice that the sum is similar to the Taylor Series for e but we notice that it is the same but skipping the first term which is 1, so it is the Taylor series for e-1, therefore

$$P(W) = \frac{e-1}{e} = 1 - \frac{1}{e}$$

4.6 *Existence

Exercise 76
A hundred students have taken an exam consisting of 8 problems, and for each problem at least 65 of the students got the right answer. Show that there exist two students who collectively got everything right, in the sense that for each problem, at least one of the two got it right.

Solution. We will set X to be the number of people who got everything right we see that $X = I_1 + I_2 + \cdot + I_8$ where $I_i = 1$ if everyone got the question right. We then see

$$E(X) = E(I_1) + E(I_2) + \cdot + E(I_8) = \frac{8}{65} \approx 0.12$$

By the Good score principle and given that the number of people expected can only be an integer, then we now that there are at least two people who got everything right.

4.7 Mixed Practice

Exercise 79
A hacker is trying to break into a password-protected website by randomly trying to guess the password. Let m be the number of possible passwords.
(a) Suppose for this part that the hacker makes random guesses (with equal probability), with replacement. Find the average number of guesses it will take until the hacker guesses the correct password

(including the successful guess).
(b) Now suppose that the hacker guesses randomly, without replacement. Find the average number of guesses it will take until the hacker guesses the correct password (including the successful guess). Hint: Use symmetry.
(c) Show that the answer to (a) is greater than the answer to (b) (except in the degenerate case $m = 1$), and explain why this makes sense intuitively.
(d) Now suppose that the website locks out any user after n incorrect password attempts, so the hacker can guess at most n times. Find the PMF of the number of guesses that the hacker makes, both for the case of sampling with replacement and for the case of sampling without replacement.

Solution.
(a) Let X be the number of tries until the hackers success including the succes, we have that $X \sim FS(\frac{1}{m})$.Therefore by definition the expected number of tries is

$$E(X) = \frac{1}{\frac{1}{m}} = m$$

So if the number of possible passwords is very very large, the hacker will probably die before they can guess the password.
(b) Let X be the number of tries until the hackers succeds without replacement we can do this as a sum of getting the correct password at each try, like

$$X = I_1 + I_2 + \cdots + I_m$$

Now we can do

$$\begin{aligned} E(X) &= E(I_1) + E(I_2) + \cdots + E(I_m) \\ &= \frac{1}{m} + \frac{1}{m-1} + \cdots + \frac{1}{m-(m-1)} \\ &= \sum_{k=1}^{m} \frac{1}{k} \approx log(m) + 0.577 \end{aligned}$$

This approximation is according to the annex of the book and works for m large.
(c) We can see that (a) grows linearly and (b) grows logarithmically, therefore (a) is bigger than (b). This makes sense because in one we are eliminating wrong options while in the other one we could still choose a singular bad option multiple times in a row if we are unlucky enough. But we see that for $m = 1$ both (a) and (b) are equal to 1. This makes sense because in both cases if there is only one posibility then it only takes one try.
(d) With replacement the answer is

$$(1-\frac{1}{m})^n + (1-\frac{1}{m})^{n-1} \cdot \frac{1}{m}$$

Without replacement we have

$$(1-\frac{1}{m}) \cdot (1-\frac{1}{m-1}) \ldots (1-\frac{1}{m-n+1}) + (1-\frac{1}{m}) \cdot (1-\frac{1}{m-1}) \ldots (1-\frac{1}{m-n+2})(\frac{1}{m-n+1})$$

Exercise 81
A group of 360 people is going to be split into 120 teams of 3 (where the order of teams and the order within a team don't matter).
(a) How many ways are there to do this?
(b) The group consists of 180 married couples. A random split into teams of 3 is chosen, with all possible splits equally likely. Find the expected number of teams containing married couples.

Solution.
(a) We are once again using what we practiced in Chapter 1. We will over count and then eliminate the over counting. Let's arrange all of the persons in a line and see in how many ways we can arrange them which is 360! and then since we are overcounting we need to eliminate the orders for which we are over counting. We don't care about the order in which people are lined up in each time so we don't care for $(3!)^{120}$ and we also don't care for the way in which we order the teams 120!. Therefore

$$\frac{360!}{(3!)^{120} \cdot 120!}$$

(b) Let set X to be the number of teams containing married couples we know there are 120 teams so we can say that

$$X = I_1 + I_2 + \cdots + I_{120}$$

where each indicator indicates if we have a couple in that team for the 120 teams then

$$E(X) = E(I_1) + E(I_2) + \cdots + E(I_{120})$$

then

$$E(X) = \sum_{i=1}^{120} E(I_i) = 120E(I_i)$$

by simmetry and linearity then for the probability we know that to put a couple in a team of 3 we have 180 possible couples and then 358 people available for the third person, but in general we have 360 choose 3 ways of choosing a team of 3

so

$$E(X) = 120 \cdot \frac{180 \cdot 358}{\binom{360}{3}} = \frac{7732800}{7711320} = 1.00278551532 \approx 1$$

Exercise 91
The Wilcoxon rank sum test is a widely used procedure for assessing whether two groups of observations come from the same distribution. Let group 1 consist of i.i.d. $X_1, \dots, X_m$ with CDF F and group 2 consist of i.i.d. $Y_1, \dots, Y_n$ with CDF G, with all of these r.v.s independent. Assume that the probability of 2 of the observations being equal is 0 (this will be true if the distributions are continuous).
After the $m + n$ observations are obtained, they are listed in increasing order, and each is assigned a rank between 1 and $m+n$: the smallest has rank 1, the second smallest has rank 2, etc. Let Rj be the rank of X_j among all the observations for $1 \leq j \leq m$, and let $R = \sum_{j=1}^{m} R_j$ be the sum of the ranks for group 1.
Intuitively, the Wilcoxon rank sum test is based on the idea that a very large value of R is evidence that observations from group 1 are usually larger than observations from group 2 (and vice versa if R is very small). But how large is "very large" and how small is "very small"? Answering this precisely requires studying the distribution of the test statistic R.
(a) The null hypothesis in this setting is that F = G. Show that if the null hypothesis is true, then $E(R) = m(m + n + 1)/2$.
(b) The power of a test is an important measure of how good the test is about saying to reject the null hypothesis if the null hypothesis is false. To study the power of the Wilcoxon rank sum test, we need to study the distribution of R in general. So for this part, we do not assume $F = G$. Let $p = P(X_1 > Y_1)$. Find $E(R)$ in terms of m, n, p. Hint: Write R_j in terms of indicator r.v.s for X_j being greater than various other r.v.s.

Solution.
(a) We assume $F = G$ and we are looking for

$$E(R) = E(\sum_{j=1}^{m} R_j) = E(R_1) + E(R_2) + \cdots + E(R_m) = \sum_{j=1}^{m} E(R_j) = mE(R_1)$$

Here we use linearity and simmetry, since the expectation of the rank of 1 should not be different from the expectation of the rank of any other one from 1 to $m + n$.
Since both F and G have the same distribution it is technically random which one out of them will be bigger than the other as the distribution is the same so the probability of getting any one rank is $\frac{1}{n+m}$ but the thing is that for instance $E(R_j)$ is the expectation of the rank of X_j and we know that the rank can be $1, 2, \ldots, n + m$ therefore

$$E(R) = E(\sum_{j=1}^{m} R_j) = E(R_1) + E(R_2) + \cdots + E(R_m) = \sum_{j=1}^{m} E(R_j) = mE(R_1) = m \sum_{k=1}^{n+m} kP(R_1 = k)$$

$$= m \cdot \frac{(n+m)(n+m+1)}{2} \cdot \frac{1}{n+m} = \frac{m(n+m+1)}{2}$$

(b)

$$E(R) = E(\sum_{j=1}^{m} R_j) = E(R_1) + E(R_2) + \cdots + E(R_m) = \sum_{j=1}^{m} E(R_j) = mE(R_1)$$

As the expectation of R1 means the expectation for the rank of X_1 among all the observations. So we set an indicator of another random variable being bigger than X_1, because that's technically how we get the rank. We know that for $m + 1$ till $m + n$ it will be $p = P(X_1 > Y_1)$ as we don't have to worry for which Y it is exactly as all come from the same distribution. Now, for 1 to m what's the probability of X_k for any $k \in \{1, \ldots, m\}$ to be bigger than X_1?

$$mE(R_1) = m((m-1)P(I) + \sum_{m+1}^{n+m} P(X_1 > Y_1))$$

The indicator is for X_k being bigger than X_1 for the m variables, but we put $m - 1$ because if you get the biggest value at first on the X_1 the probability is 0 so we eliminate that one.

$$E(R) = mE(R_1) = m[(m-1)\frac{(m-1)!}{m^2} + np]$$

The $\frac{(m-1)!}{m^2}$ comes from the fact that in the indicator r.v we are looking for the probability that for a distribution if I choose a set draw from that distribution say 1 like in this case that the next draws will be bigger than the first draw. We can look for the probability easily thinking of a dice.

Exercise 92
The legendary Caltech physicist Richard Feynman and two editors of The Feynman Lectures on Physics (Michael Gottlieb and Ralph Leighton) posed the following problem about how to decide what to order at a restaurant. You plan to eat m meals at a certain restaurant, where you have never eaten before. Each time, you will order one dish. The restaurant has n dishes on the menu, with $n \geq m$. Assume that if you had tried all the dishes, you would have a definite ranking of them from 1 (your least favorite) to n (your favorite). If you knew which your favorite was, you would be happy to order it always (you never get tired of it).
Before you've eaten at the restaurant, this ranking is completely unknown to you. After you've tried some dishes, you can rank those dishes amongst themselves, but don't know how they compare with the dishes you haven't yet tried. There is thus an exploration exploitation tradeoff : should you try new dishes, or should you order your favorite among the dishes you have tried before?
A natural strategy is to have two phases in your series of visits to the restaurant: an exploration phase, where you try different dishes each time, and an exploitation phase, where you always order the best dish you obtained in the exploration phase. Let k be the length of the exploration phase (so $m - k$ is the length of the exploitation phase). Your goal is to maximize the expected sum of the ranks of the dishes you eat there (the rank of a dish is the "true" rank from 1 to n that you would give that dish if you could try all the dishes). Show that the optimal choice is

$$k = \sqrt{2(m+1)} - 1$$

or this rounded up or down to an integer if needed. Do this in the following steps:
(a) Let X be the rank of the best dish that you find in the exploration phase. Find the expected sum of the ranks of all the dishes you eat (including both phases), in terms of k, n, and E(X).
(b) Find the PMF of X, as a simple expression in terms of binomial coefficients.
(c) Show that

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

Hint: Use Example 1.5.2 (about the team captain) and Exercise 20 from Chapter 1 (about the hockey stick identity).
(d) Use calculus to find the optimal value of k.

Solution.
(a) Let's set Y to be the r.v for the ranks of both sections explotation and exploration.

$$E(Y) = k \cdot \frac{n(n+1)}{2} \cdot \frac{1}{n} + (m-k)[E(X) \cdot \frac{1}{n}] = \frac{k(n+1)}{2} + (m-k)[E(X) \cdot \frac{1}{n}]$$

For the first phase I'm using $\frac{n(n+1)}{2}$ because that's the expected rank for each individual dish even if we are eating just k dishes (that's why it is multiplied by k) and then $\frac{1}{n}$ because that's the possibility of each individual dish having that rank. For the second part it is nearly the same think but we are just doing $m-k$ dishes because well that's the length of the exploitation phase. I don't know why the instruction says to only do it without m. In that case we would just go n times which wouldn't change this formula a lot just change the m for the n.
(b)

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

So we chose z and set it and after that we have the numbers behind z to choose from $k-1$ and this is divided by how many ways we have to choose k plates of n total plates.
(c)

$$E(X) = \sum_{z=1}^{n-k+1} zP(X=z) = \sum_{z=1}^{n-k+1} z\frac{\binom{n-z}{k-1}}{\binom{n}{k}} = \frac{1}{\binom{n}{k}} \sum_{z=1}^{n-k+1} z\binom{n-z}{k-1}$$

The sums ends at $n-k+1$ because if we take a sample of k dishes in the worst case scenario we will have the n-k+1 best dish. We won't have the very last one. If you think about it if we have 10 dishes and we choose a sample of 3 there is no way that your best dish will be the number 10 because you grabbed 3 so you will have to have 10 11 and 12 for your best dish to be 10, therefore your best dish is (in this scenario) 8, because the worst case scenario is 8 9 10.
We will set a variable called $o = n - z$ then the sum will be

$$= \frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} (n-o)\binom{o}{k-1}$$

$$= \frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} n\binom{o}{k-1} - \frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} o\binom{o}{k-1}$$

$$= \frac{n}{\binom{n}{k}} \sum_{o=k-1}^{n-1} \binom{o}{k-1} - \frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} o\binom{o}{k-1}$$

using Exercise 20 from chapter 1 on the left sum we have

$$= \frac{n}{\binom{n}{k}}\binom{n}{k} - \frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} o\binom{o}{k-1}$$

$$= n - \frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} o\binom{o}{k-1}$$

Since we know $o\binom{o}{k-1} = (o+1)\binom{o}{k-1} - \binom{o}{k-1}$

$$= n - (\frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} (o+1)\binom{o}{k-1} - \frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} \binom{o}{k-1})$$

using Exercise 20 from chapter 1 we can again reduce the right sum

$$= n - (\frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} (o+1)\binom{o}{k-1} - \frac{\binom{n}{k}}{\binom{n}{k}})$$

and now using the Team captain identity we change the sum to

$$= n - (\frac{1}{\binom{n}{k}} \sum_{o=k-1}^{n-1} k\binom{o+1}{k} - \frac{\binom{n}{k}}{\binom{n}{k}})$$

we simplify to make it prettier

$$= n - (\frac{k}{\binom{n}{k}} \sum_{o=k-1}^{n-1} \binom{o+1}{k} - 1)$$

usinge Exercise 20 from chapter 1 we simpligy the sum to

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

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

now we do this manually using the coefficient

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

$$= n - \frac{k \cdot k!(n-k)!}{n!} \cdot \frac{(n+1)!}{(k+1)!(n+1-(k+1))!} + 1$$

$$= n - \frac{k \cdot k!(n-k)!}{n!} \cdot \frac{(n+1)!}{(k+1)!(n-k)!} + 1$$

$$= n - \frac{k \cdot k!(n-k)! \cdot (n+1)}{(k+1)!(n-k)!} + 1$$

$$= n - \frac{k \cdot k! \cdot (n+1)}{(k+1)!} + 1$$

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

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

$$= \frac{nk + n - nk - k + k + 1}{(k+1)}$$

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

$$= \frac{n+1}{(k+1)}$$

(d)

$$\frac{k(n+1)}{2} + (m-k)[\frac{k(n+1)}{k+1} \cdot \frac{1}{n}]$$

we basically derive this and put it to 0

BAD TRIAL– I LEAVE THIS HERE JUST TO SHOW MY MISTAKE– by using Example 1.5.2

$$= n\binom{n}{k} - \sum_{o=k-1}^{n-1} \frac{k\binom{n}{k}}{n} o$$

$$= n\binom{n}{k} - \frac{k\binom{n}{k}}{n} \sum_{o=k-1}^{n-1} o$$

So how do we do $\sum_{o=k-1}^{n-1} o$ we can do the sum from 1 to $k-1$ and also the sum from 1 to $n-1$ and subtract them $\sum_{o=k-1}^{n-1} o = \frac{n(n-1)}{2} - \frac{k(k-1)}{2}$

$$n\binom{n}{k} - \frac{k\binom{n}{k}}{n}[\frac{n(n-1)}{2} - \frac{k(k-1)}{2}]$$

$$\frac{n \cdot n!}{k!(n-k)!} - \frac{k \cdot n!}{n \cdot k!(n-k!)}[\frac{n(n-1)}{2} - \frac{k(k-1)}{2}]$$

$$\frac{n \cdot n!}{k!(n-k)!} - \frac{k \cdot n! \cdot n \cdot (n-1)}{n \cdot k!(n-k!) \cdot 2} - \frac{k \cdot n! \cdot k \cdot (k-1)}{n \cdot k!(n-k!) \cdot 2}$$

$$\frac{n \cdot n!}{k!(n-k)!} - \frac{k \cdot n! \cdot n \cdot (n-1) - [k \cdot n! \cdot k \cdot (k-1)]}{n \cdot k!(n-k!) \cdot 2}$$

$$\frac{n \cdot n \cdot n! \cdot 2}{n \cdot k!(n-k)! \cdot 2} - \frac{k \cdot n! \cdot n \cdot (n-1) - [k \cdot n! \cdot k \cdot (k-1)]}{n \cdot k!(n-k!) \cdot 2}$$

$$\frac{n^2 \cdot n! \cdot 2}{n \cdot k!(n-k)! \cdot 2} - \frac{k \cdot n! \cdot n^2 - k \cdot n! \cdot n - [k^3 \cdot n! - k^2 \cdot n!]}{n \cdot k!(n-k!) \cdot 2}$$

$$\frac{n^2 \cdot n! \cdot 2}{n \cdot k!(n-k)! \cdot 2} - \frac{k \cdot n! \cdot n^2 - k \cdot n! \cdot n - k^3 \cdot n! + k^2 \cdot n!}{n \cdot k!(n-k!) \cdot 2}$$

$$\frac{n^2 \cdot n! \cdot 2 - [k \cdot n! \cdot n^2 - k \cdot n! \cdot n - k^3 \cdot n! + k^2 \cdot n!]}{n \cdot k!(n-k!) \cdot 2}$$

$$\frac{n^2 \cdot n! \cdot 2 - k \cdot n! \cdot n^2 + k \cdot n! \cdot n + k^3 \cdot n! - k^2 \cdot n!}{n \cdot k!(n-k!) \cdot 2}$$

$$\frac{n!(n^2 \cdot 2 - k \cdot n^2 + k \cdot n + k^3 - k^2)}{n \cdot k!(n-k!) \cdot 2}$$

UNFORTUNATELY I COULDN'T GET TO THE ANSWER THIS WAY