

# Probability Blitzstein - Solutions Manual

Mingruifu Lin, Mina Lin

September 2025



# Contents

<b>1</b>	<b>Probability and Counting</b>	<b>5</b>
1.1	Counting . . . . .	5
1.2	Story Proofs . . . . .	10
1.3	Naive Definition of Probability . . . . .	13
1.4	Axioms of Probability . . . . .	16
<b>2</b>	<b>Conditional Probability</b>	<b>19</b>
2.1	Conditioning on Evidence . . . . .	19
2.2	Independence . . . . .	22
<b>3</b>	<b>Random Variables and Distributions</b>	<b>27</b>
3.1	PMFs and CDFs . . . . .	27
3.2	Named Distributions . . . . .	31
<b>4</b>	<b>Expectation</b>	<b>33</b>
4.1	Expectations and Variances . . . . .	33
4.2	Named Distributions . . . . .	37
4.3	Indicator Random Variables . . . . .	37
4.4	LOTUS . . . . .	39
4.5	Poisson Approximation . . . . .	40
<b>5</b>	<b>Continuous Random Variables</b>	<b>43</b>
5.1	PDFs and CDFs . . . . .	43



# Chapter 1

# Probability and Counting

## 1.1 Counting

### Exercise 1

11 places to put the 4 "S". Remains 7 place to put the 4 "I". Remains 3 place to the 2 "P". Remains 1 place to put the 1 "M". Combination is used because the order of identical letters does not matter.

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

### Exercise 2

(a) 8 choices for the first digit, 10 choices for the remaining 6 digits.

$$8 \cdot 10^6$$

(b) From the total count, we subtract numbers that start with 911. A number that begins with 911 has 4 remaining digits, so a total of  $10^4$  numbers begin with 911.

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

### Exercise 3

(a) This is simply the amount of permutations of 10 restaurants over 7 days. The first day, he has 10 choices. The last day, only 4 choices remain.

$$10 \cdot 9 \cdot \dots \cdot 4 = \frac{10!}{3!}$$

(b) The only thing that matters is that he does not dine at the same restaurant 2 days in a row. The first day, he has 10 choices. The remaining days, he cannot repeat yesterday's restaurant, so he has 9 choices.

$$10 \cdot 9^6$$

**Exercise 4**

- (a) There are  $\binom{n}{2}$  games played in total. Each game has 2 outcomes.

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

- (b) We already found it in (a).

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

**Exercise 5**

- (a) For each game played between 2 players, 1 of them is eliminated. Thus, each round, half of the players are eliminated. Hence, for  $2^n$  players, there are  $n$  rounds.

- (b) We are simply adding powers of 2.

$$\sum_{i=0}^n 2^i = 2^n - 1$$

- (c) There remains 1 champion at the end. Thus, from the original  $2^n$  players,  $2^n - 1$  players must be eliminated. Hence,  $2^n - 1$  games are played, since each game eliminates exactly 1 player.

**Exercise 6****Approach 1**

The first player can be paired with any of the 19 other players, the third player can be paired with any of the 17 remaining players, and so on. Also, each game has 2 configurations: white or black.

$$\begin{aligned} & (2 \cdot 19) \cdot (2 \cdot 17) \cdot \dots \cdot (2 \cdot 1) \\ &= (\frac{20}{10} \cdot 19) \cdot (\frac{18}{9} \cdot 17) \cdot \dots \cdot (\frac{2}{1} \cdot 1) \\ &= \frac{20 \cdot 19}{10} \cdot \frac{18 \cdot 17}{9} \cdot \dots \cdot \frac{2 \cdot 1}{1} \\ &= \frac{20!}{10!} \end{aligned}$$

**Approach 2**

Align all 20 players in a particular order, in a single row. There are  $20!$  ways to align them.

Group them in pairs: the first in the row plays with the second, the third plays with the fourth, and so on. There are 10 pairs in total, which means there are  $10!$  ways to shuffle these pairs.

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

### Exercise 7

(a) A total of 7 games are played. There are  $\binom{7}{3}$  ways to put the 3 wins among the 7 games. There are  $\binom{4}{2}$  ways to put the 2 draws among the remaining 4 games. There are  $\binom{2}{2}$  ways to put the 2 losses among the remaining 2 games.

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

(b) B's win/draw/loss are uniquely determined by A's win/draw/loss. So we only consider A. To enumerate all cases, notice that if A's wins are not enough, we must fill those missing points with draws. Thus, A's wins uniquely determine their draws. We count just like in (a).

Case 1: A has 4 wins and 3 losses.

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

Case 2: A has 3 wins and 2 draws and 2 losses.

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

Case 3: A has 2 wins and 4 draws and 1 loss.

$$\binom{7}{2} \binom{5}{4}$$

Case 4: A has 1 win and 6 draws.

$$\binom{7}{1}$$

All that's left is to add them up.

(c) We go back to the cases. Notice that A must not end on a loss, which would imply that less than 7 games were necessary to accumulate 4 points.

Case 1: Among the first 6 games, we must place 3 wins. The last win is on the 7th game:  $\binom{6}{3}$ .

Case 2: Either A draws or wins on the 7th game. If he draws, he places the 3 wins among the 6 games, and he places the 1 draw among the remaining 3 games:  $\binom{6}{3} \binom{3}{1}$ . If he wins, he places the 2 wins among the 6 games, and he places the 2 draws among the remaining 4 games:  $\binom{6}{2} \binom{4}{2}$ .

Case 3: Similar to Case 2. If he draws the 7th:  $\binom{6}{2} \binom{4}{3}$ . If he wins the 7th:  $\binom{6}{1} \binom{5}{4}$ .

Case 4: If he draws the 7th:  $\binom{6}{1} \binom{5}{5}$ . If he wins the 7th:  $\binom{6}{6}$ .

All that's left is to add them up.

**Exercise 8**

(a) Choose 2 from the original 12 people. Choose 5 from the remaining 10. Then, choose 5 from the remaining 5. Since the teams are unlabelled, there are 2 ways to permute the groups of 5, i.e. they are counted twice.

$$\frac{\binom{12}{2} \binom{10}{5}}{2!}$$

(b) Similar to (a), but this time, there are -not 2- but 3 groups of same size.

$$\frac{\binom{12}{4} \binom{8}{4}}{3!}$$

**Exercise 9**

(a) There are a total of 110 steps to the right, and 111 steps upwards. We simply need to choose the order in which to perform the steps. Among these 221 total actions, we must choose which ones are steps to the right. (Using steps upwards gives the same answer.)

$$\binom{221}{110}$$

(b) We simply multiply the answer in (a) by the amount of ways to go from  $(110, 111)$  to  $(210, 211)$ . There are  $\binom{200}{100}$  ways. Thus, the total is

$$\binom{221}{110} \binom{200}{100}$$

**Exercise 10**

(a) If he chooses 1 stat courses, he chooses 6 of the 15 non-stat courses. If he chooses 2 stat courses, he chooses 5 of the 15 non-stat courses. And so on. He stops at 5 stat courses.

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

**Approach 2**

(a) Finding all combinations with at least 1 statistics course is equivalent to the complement of having 0 statistics course. The total amount of combinations can be written as  $\binom{20}{7}$ . The amount of combinations without statistics courses is  $\binom{15}{7}$ .

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

(b) Some configurations are repeated if we use that formula.

For example, we can label the classes from 1 to 20. In one configuration, he can choose stat class 1 in  $\binom{5}{1}$ , and stat class 2 in  $\binom{19}{6}$ , with no other stat classes. But in another configuration, he can choose stat class 2 in  $\binom{5}{1}$ , and stat class 1 in  $\binom{19}{6}$ , with no other stat classes, but with the same non-stat class as the previous configuration. These two configurations hold effectively the same classes, but are counted separately as if they were different.

### Exercise 11

- (a) The function need not be injective, so each of the  $n$  domain elements can choose any of the  $m$  codomain elements:

$$m^n$$

- (b) Now, it's injective. This is simply permutation.  $m$  choices for the first domain element,  $m - 1$  choices for the second, and so on.

$$\frac{m!}{(m - n)!}$$

This only works if  $m \geq n$ , i.e. the codomain is larger than the domain.

### Exercise 12

- (a) Each player receives 13 cards among the 52.

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

- (b) First player chooses 13 cards among 52. Second player chooses 13 among remaining 39. And so on. Last player has no choice.

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

- (c) A card can only be included in a one hand. Once 13 cards are picked, they are not replaced in the pack.

### Exercise 13

Think of each card type as a box where you throw pebbles. There are 52 boxes, and you have 10 pebbles. Notice that each box has a capacity of 10, so we are safe from pebble overflow.

If you recall Bose-Einstein, we have a total of 53 bars and 10 pebbles. A bar must occupy the far left, and a bar must occupy the far right. So 51 bars remain. Thus, there are  $51 + 10 = 61$  seats to place the remaining bars and pebbles. We can focus on placing the 10 pebbles among the 61 seats, since the bars fill out the remaining seats.

$$\binom{61}{10}$$

**Exercise 14**

There are  $\binom{8}{8} + \binom{8}{7} + \cdots + \binom{8}{1} + \binom{8}{0} = 2^8$  topping combinations. There are 4 pizza sizes.

Consider the case when the sizes are different. If the sizes are different, then there are  $\binom{4}{2}$  size combinations, each with  $2^8 \cdot 2^8$  topping combinations.

$$\binom{4}{2} \cdot 2^{16}$$

If the size are the same, then there are 4 size combinations, each with  $2^8 \cdot 2^8$  topping combinations. In total, we have

$$\binom{4}{2} \cdot 2^{16} + 4 \cdot 2^{16}$$

**1.2 Story Proofs****Exercise 15**

Suppose we want to count the number of subsets of  $n$  elements.

Method 1: There are  $2^n$  subsets, because each element is either in, or not in, the current subset.

Method 2: We could also decide to count subsets of increasing size. We start with subset of size 0, then size 1, and so on.

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n}$$

**Exercise 16**

(a) Algebraically:

$$\begin{aligned} \binom{n}{k} + \binom{n}{k-1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\ &= \frac{n!(n-k+1)}{k!(n-k+1)!} + \frac{n! \cdot k}{k!(n-k+1)!} \\ &= \frac{n!(n+1)}{k!(n-k+1)!} \\ &= \frac{(n+1)!}{k!((n+1)-k)!} \\ &= \binom{n+1}{k} \end{aligned}$$

(b) An organization has 1 president and  $n$  other members. Let's say we want to choose  $k$  members from the  $n+1$  members (which include the president).

Method 1: This is just  $\binom{n+1}{k}$ .

Method 2: We either choose the president or we don't. If we choose the president, then there are  $n$  members left, from which we choose  $k - 1$  members to complete the group. This is  $\binom{n}{k-1}$ . If we don't choose the president, then we choose  $k$  members from the  $n$  members. This is  $\binom{n}{k}$ . The total is

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

### Exercise 17

Suppose we want to choose  $n$  objects from a set of  $2n$  objects.

Method 1: This is  $\binom{2n}{n}$ .

Method 2: Split the  $2n$  objects into two disjoint groups of  $n$  objects. We wish to choose  $n$  objects in total.

We can decide to choose 0 objects from the 1st group, and  $n$  objects from the 2nd group. We can choose 1 object from the 1st group, and  $n - 1$  objects from the 2nd group. And so on.

But equivalently, we can say which of the objects from the 2nd group are NOT chosen. So we choose 0 object from the 1st group, and we choose 0 object in the second group to NOT be chosen. Then we choose 1 object from the 1st group, and we choose 1 object in the second group to NOT be chosen (the remaining  $n - 1$  objects are chosen). Recall that each group has  $n$  objects. This translates to:

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

### Exercise 18

We want to form a committee of  $n$  people chosen from 2 separate groups of  $n$  people ( $2n$  people total). The president (of the committee) must have been a member of the 1st group.

Method 1: Choose the president from amongst the 1st group. There are  $n$  members to choose from. Then, choose the  $n - 1$  remaining committee seats from the remaining  $2n - 1$  people. This is

$$n \binom{2n-1}{n-1}$$

Method 2: Similar to Exercise 17. We select  $k$  members from the 1st group, then we say which of the  $n$  members from the 2nd group are NOT chosen. For the latter, we select  $k$  such members that are NOT chosen, because we want the total committee size to be  $k + (n - k) = n$ . Then, among the chosen members from the first group (there are  $k$  of them), we choose one to be the president. Notice that the sum starts at 1 (to make sure there is a president):

$$\sum_{k=1}^n k \binom{n}{k} \binom{n}{k} = \sum_{k=1}^n k \binom{n}{k}^2$$

**Exercise 19**

From a set of  $n + 3$  objects, with  $n \geq 2$ , we want to choose 5 objects.

Method 1:

$$\binom{n+3}{5}$$

Method 2: Select the  $(k + 1)$ th object as your third object. This splits your set into left and right. From the  $k$  left objects, choose 2 objects. Then from the  $(n + 3) - (k + 1) = (n - k + 2)$  right objects, choose 2 objects. Thus, for each  $k$ , we have

$$\binom{k}{2} \binom{n-k+2}{2}$$

What's left is to sum up based on different chosen 3rd object.

**Exercise 20**

(a) From  $n + 1$  sticks, I wish to choose  $k + 1$  sticks.

Method 1: This is

$$\binom{n+1}{k+1}$$

Method 2: Order the sticks in a row. I decide to include the 1st stick, so I only need to choose the remaining  $k$  sticks from the remaining  $n$  sticks. I can also decide to exclude the 1st stick, but always include the 2nd stick, so I only need to choose the remaining  $k$  sticks from the remaining  $n - 1$  sticks (which exclude the 1st stick). Then I can decide to exclude both the 1st and 2nd stick, and always include the 3rd stick, so again I need to choose  $k$  sticks from the  $n - 2$  remaining sticks. And so on.

$$\binom{n}{k} + \binom{n-1}{k} + \cdots + \binom{k+1}{k} + \binom{k}{k}$$

(b) We loop from 30 to 50 gummies. For each, we apply Bose-Einstein.

$$\sum_{i=30}^{50} \binom{i+4}{4}$$

Applying part (a), since we start at  $i = 30$ , we truncate to get

$$\binom{55}{5} - \binom{34}{5}$$

**Exercise 21**

(a) In a class of  $n + 1$  students, I want to form  $k$  groups. If I'm in a group by myself, then there are  $n$  students left to form the remaining  $k - 1$  groups:

$\{ \binom{n}{k-1} \}$ . If I'm not in a group by myself, then I can decide to join any of the  $k$  groups already formed:  $k \{ \binom{n}{k} \}$ . Hence,

$$\binom{n+1}{k} = \binom{n}{k-1} + k \binom{n}{k}$$

(b) With  $n+1$  students, I want to form  $k+1$  groups. Suppose I am the first one to form my group. From the  $n$  students (other than me), I choose  $j$  students that are NOT in my group. Since there remains  $k$  groups to be formed, and each group is at least 1 student, then I need to choose at least  $k$  students NOT in my group. In any case, I leave the  $j$  remaining students to form their own  $k$  groups. Hence, I have

$$\binom{n+1}{k+1} = \sum_{j=k}^n \binom{n}{j} \binom{j}{k}$$

### Exercise 22

(a) Round robin is given as a hint already, but let's consider a different, visual approach: We need to choose 2 items from  $n+1$  items. Order the items in a row. We can fix our left finger on item 1, and glide our right finger on the subsequent  $n$  items. We can also fix our left finger on item 2, and glide our right finger on the subsequent items  $n-1$  (we don't go back to item 1, or else we would be recounting). We can repeat for item 3, gliding our right finger over the remaining  $n-2$  items. And so on. Thus,

$$\binom{n+1}{2} = n + (n-1) + \cdots + 2 + 1$$

(b)

## 1.3 Naive Definition of Probability

### Exercise 23

There are  $9^3$  combinations total. There are 7 places to put the consecutive triple. There are  $3!$  permutations among the triple. Thus, the probability is

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

### Exercise 24

There are  $6!$  birth orders. If the eldest are girls, there are  $3!$  ways to permute them among the eldest, and  $3!$  ways to permute the boys among the youngest. The probability is

$$\frac{3!3!}{6!}$$

Equivalently, there are  $\binom{6}{3}$  arrangements of boys and girls. Only one of them, BBBGGG, is valid. Hence, we have

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

### Exercise 25

The only way to have at most 1 robbery per district to have exactly 1 robbery per district. So we take the total, minus this, all over the total. The probability is

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

### Exercise 26

(a) This is a year with 1 million days, and you have a crowd of 1000 people. You want to know how many share the same birthday.

(b) To get the probability of resampling the same resident, we instead calculate the complement probability, i.e. probability to get different residents in all samples. Using this method, we get

$$\frac{(10^6)^{1000} - \frac{(10^6)!}{(10^6-1000)!}}{(10^6)^{1000}}$$

### Exercise 27

$$\frac{n^k - \frac{n!}{(n-k)!}}{n^k}$$

### Exercise 28

$$\frac{10^3 - 10 \cdot 9 \cdot 8}{10^3}$$

### Exercise 29

(a)  $\frac{4 \cdot 6}{2} = 12$ , so 21, which is closer to 12, is more likely.

(b) For the 2-letter word, whatever the first letter, the second letter must be the same:  $\frac{1}{26}$ . The same goes for the 3-letter word, with the relation between the first and third letter.

### Exercise 30

For  $n = 7$ , we have  $(\frac{1}{26})^3$ . For  $n = 8$ , we have  $(\frac{1}{26})^4$ .

**Exercise 31**

The total ways to recapture  $m$  random elk is  $\binom{N}{m}$ . The ways to recapture exactly  $k$  tagged elk is to choose  $k$  tagged elk from among the  $n$  tagged elk, and to choose  $m - k$  non-tagged elk from the  $N - n$  non-tagged elks:  $\binom{n}{k} \binom{N-n}{m-k}$ . The probability is thus:

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

Note: This will be further discussed in the hypergeometric distribution section, later in the book.

**Exercise 32**

$j = 3$  is impossible: once you guess 3 colors correctly, the 4th is implicitly guessed correct. Likewise,  $j = 1$  is impossible: suppose you guessed one red correctly, so you must choose from the remaining 3 cards such that they are all wrong, but since there are 2 black cards left, you can only select one of them as red, leaving the other black as implicitly guessed correctly.  $j = 0$  is the same as  $j = 4$ , which have probability  $\frac{1}{\binom{4}{2}}$ . For  $j = 2$ , we take their complement, which is  $1 - 2\frac{1}{\binom{4}{2}}$ .

**Exercise 33**

- (a) This experiment is equivalent to drawing two balls at once. By symmetry, they have equal probability to be green.
- (b) The sample space contains all pairs of balls. The probability that the first ball is green is  $\frac{g}{r+g}$ . The probability that the second ball is green is

$$\begin{aligned} & \frac{g}{r+g} \frac{g-1}{r+g-1} + \frac{r}{r+g} \frac{g}{r-1+g} \\ &= \frac{g(g-1) + rg}{(r+g)(r+g-1)} \\ &= \frac{g(g-1+r)}{(r+g)(r+g-1)} = \frac{g(r+g-1)}{(r+g)(r+g-1)} \\ &= \frac{g}{r+g} \end{aligned}$$

- (c) From  $r + g$  balls, to get same color balls, we either choose 2 red balls or 2 green balls:

$$\binom{r}{2} + \binom{g}{2}$$

To get different color:

$$\binom{r}{1} \binom{g}{1}$$

We equate them and replace  $g = 16 - r$ :

$$\begin{aligned} \binom{r}{2} + \binom{16-r}{2} &= \binom{r}{1} \binom{16-r}{1} \\ \frac{r(r-1)}{2} + \frac{(16-r)(16-r-1)}{2} &= r(16-r) \\ 2r^2 - 32r + 120 &= 0 \\ 2(r-10)(r-6) &= 0 \end{aligned}$$

Either  $r = 10$  and  $g = 6$ , or  $r = 6$  and  $g = 10$ .

### Exercise 34

$$\frac{4\binom{13}{5}}{\binom{52}{5}}$$

### Exercise 35

$$\frac{44\binom{13}{2}\binom{4}{2}^2}{\binom{52}{5}}$$

### Exercise 36

This is the same problem as shuffling letters of a word, with possibly repeated letters.

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

## 1.4 Axioms of Probability

### Exercise 43

(i)

$$P(A) = P((A \cap B) \cup (A \cap B^c))$$

Since  $B$  and  $B^c$  are disjoint, then their subsets  $A \cap B$  and  $A \cap B^c$  are disjoint as well. Hence their probability are additive.

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

Using  $P(A \cap B^c) \leq P(B^c)$ , and the fact that  $P(B^c) = 1 - P(B)$ , we have

$$P(A \cap B) + P(A \cap B^c) \leq P(A \cap B) + P(B^c) = P(A \cap B) + 1 - P(B)$$

Rearranging, we get

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

(ii) Since  $A \cap B \subseteq A \subseteq A \cup B$ , we get

$$P(A \cap B) \leq P(A) \leq P(A \cup B)$$

(iii) Using  $A \cup B = (A \cap B) \cup (A^c \cap B) \cup (A \cap B^c) = B \cup (A \cap B^c)$ , with  $B$  and  $A \cap B^c$  disjoint, we get

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

### Exercise 44

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

Since  $A \subseteq B$ , then  $A$  is disjoint from  $B^c$ .

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

### Exercise 45

We use the fact:

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

Now, we derive

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

### Exercise 46

The event that at least  $k$  events occur is either:

1. There are exactly  $k$  events that occur.
2. There are more than  $k$  events that occur, i.e.  $k + 1$  events or more.

Hence,

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

**Exercise 47**

- (a) I'll not give examples. I'll just provide a new perspective:  $B$  is independent of  $A$  if

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

where  $U$  is the entire sample space (universe). In other words, the proportion of  $A$  inside  $B$  is equivalent to the proportion that  $A$  occupies in the universe originally. In other words, if we condition  $A$  on  $B$ , there is no change to the new probability of  $A$ . Notice that the relation is symmetric.

- (b) You can easily translate the above logic into areas.

(c)

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

**Exercise 48**

Unclear exercise.

# Chapter 2

## Conditional Probability

### 2.1 Conditioning on Evidence

#### Exercise 1

$$\begin{aligned} P(S|F) &= \frac{P(F|S)P(S)}{P(F)} \\ &= \frac{P(F|S)P(S)}{P(F|S)P(S) + P(F|S^c)P(S^c)} \\ &= \frac{10\% \cdot 80\%}{10\% \cdot 80\% + 1\% \cdot 20\%} \\ &= \frac{0.08}{0.08 + 0.002} \\ &\approx 100\% \end{aligned}$$

#### Exercise 2

$$\begin{aligned} P(I|B) &= \frac{P(B|I)P(I)}{P(B)} \\ P(I|B) &= \frac{P(B|I)P(I)}{P(B|I)P(I) + P(B|I^c)P(I^c)} \\ &= \frac{\frac{1}{2} \frac{1}{3}}{\frac{1}{2} \frac{1}{3} + \frac{1}{4} \frac{2}{3}} \\ &= \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{6}} \\ &= \frac{1}{2} \end{aligned}$$

**Exercise 3**

Let's translate the first sentence.

$$P(C|S) = 23P(C|S^c)$$

Now, we isolate  $P(C)$ .

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

Now, we solve:

$$\begin{aligned} P(S|C) &= \frac{P(C|S)P(S)}{P(C)} \\ &= \frac{P(C|S)P(S)}{P(C|S)P(S) + \frac{P(C|S)}{23}P(S^c)} \\ &= \frac{P(S)}{P(S) + \frac{1}{23}P(S^c)} \\ &= \frac{21.6\%}{21.6\% + \frac{1}{23} \cdot 21.6\%} \\ &\approx 100\% \end{aligned}$$

**Exercise 4**

(a)

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

(b) The denominator is smaller than 0. Thus,  $p$  divided by the denominator is greater than  $p$ . It equals  $p$  when

$$p + \frac{1-p}{n} = 1$$

Clearly,  $n = 1$  is a solution. This is because, now, getting the correct answer does not tell us more information about what he knows.

**Exercise 5**

Skip

**Exercise 6**

Let  $C$  be the event that we chose the double-headed coin. Let  $H$  be the event that we got 7 heads in a row.

$$\begin{aligned} P(C|H) &= \frac{P(H|C)P(C)}{P(H)} \\ &= \frac{P(H|C)P(C)}{P(H|C)P(C) + P(H|C^c)P(C^c)} \\ &= \frac{1 \cdot \frac{1}{100}}{1 \cdot \frac{1}{100} + \frac{1}{2^7} \frac{99}{100}} \\ &\approx 56\% \end{aligned}$$

**Exercise 7**

(a)  $D$  is the event that one of the coins is double-headed. It does not mean that we have chosen that coin. Preliminary calculations:

$$P(H|D) = \frac{1}{2^7} \frac{99}{100} + 1 \cdot \frac{1}{100} = \frac{227}{12800}$$

As you see, by conditioning on  $D$ , we only know that there is a double headed coin. 99% of the time, we choose a fair coin. 1% of the time, we choose the double-headed coin.

$$P(H) = P(H|D)P(D) + P(H|D^c)P(D^c) = \frac{227}{12800} \frac{1}{2} + \left(\frac{1}{2^7}\right) \frac{1}{2} = \frac{327}{25600}$$

With the preliminary calculations done, we solve using Bayes:

$$\begin{aligned} P(D|H) &= \frac{P(H|D)P(D)}{P(H)} \\ &= \frac{\frac{227}{12800} \frac{1}{2}}{\frac{327}{25600}} \\ &\approx 69\% \end{aligned}$$

It makes sense that it is higher than Exercise 6, since we only require that the double-headed coin exists. We do not need to choose it.

(b) Now, we have to choose it. Let  $C$  be the event that we chose the double-headed coin.

$$P(C|H) = \frac{P(H|C)P(C)}{P(H)}$$

$$= \frac{1 \cdot \frac{1}{100} \frac{1}{2}}{\frac{327}{25600}} \\ \approx 39\%$$

It makes sense that it is smaller than Exercise 6, because the double-headed coin now only appears half of the time.

### Exercise 8

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)} \\ = \frac{0.01 \cdot 0.5}{0.01 \cdot 0.5 + 0.02 \cdot 0.3 + 0.03 \cdot 0.2} \\ \approx 29\%$$

### Exercise 9

(a) The assumptions are the following:

$$P(A_1) = P(A_2)$$

$$P(B|A_1) = 1$$

$$P(B|A_2) = 1$$

Now, using simple algebra:

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

(b) Let's say that there are many varieties of berry. Two of the varieties are poisonous and they appear equally often in nature. Suppose that ingesting any of the two types of poisonous berry always leads to death. If someone dies, there is no way to use that information to get a better guess as to which berry they ingested.

## 2.2 Independence

### Exercise 30

(a) A cannot be the youngest, leaving C more often with the choice of the youngest.

(b) There are  $6!$  ways to arrange the children. If A is older than C, then only 3 arrangements remain. 2 of these arrangements satisfy the criteria. Hence, the probability is

$$\frac{2}{3}$$

**Exercise 31**

Two events are independent if  $P(A \cap B) = P(A)P(B)$ . If  $A = B$ , then we have

$$P(A \cap A) = P(A)P(A)$$

$$P(A) = P(A)^2$$

Then either  $P(A) = 0$  or  $P(A) = 1$ .

**Exercise 32**

- (a) Dice B always returns 3. We only need to count the times when dice A returns 4.

$$P(A > B) = \frac{4}{6} = \frac{2}{3}$$

Similarly, for dice C.

$$P(B > C) = \frac{4}{6} = \frac{2}{3}$$

Between dice C and D, we decompose the problem. We look at when dice C is 6, which happens  $\frac{2}{6}$  of the time, then look at when dice C is 2, which happens  $\frac{4}{6}$  of the time. When dice C is 2, dice D must be 1, which happens  $\frac{3}{6}$  of the time.

$$P(C > D) = \frac{2}{6} \cdot 1 + \frac{4}{6} \cdot \frac{3}{6} = \frac{2}{3}$$

We decompose the problem for dice A and D as well.

$$P(D > A) = \frac{3}{6} \cdot 1 + \frac{3}{6} \cdot \frac{2}{6} = \frac{2}{3}$$

- (b)  $A > B$  is independent of  $B > C$ . This is because B is the only shared information between the two events, but B is always 3, so knowing that  $A > B$  does not tell us additional information about B.

However,  $B > C$  is not independent from  $C > D$ . This is because knowing that  $B > C$  eliminates some crucial possibilities of C. For example, if  $B > C$ , then C loses the 6's which were originally guaranteed to beat D. With this conditioning, C only beats D half of the time, instead of  $\frac{2}{3}$ .

**Exercise 33**

- (a) The question asks us to find the probability that the selected group D are all friends of Alice. For each person in D, there is a  $\frac{1}{2}$  chance that they are friends with Alice. For a group of size  $n$ , since each friendship is independent, the probability that every single one of them is Alice's friend is

$$P(D = A) = \left(\frac{1}{2}\right)^n$$

(b) The question asks us to find the probability that every friend of Alice is a friend of Bob. Consider a single person. If that person is not a friend of Alice, we don't care. This happens  $\frac{1}{2}$  of the time. If that person IS a friend of Alice, then it must be a friend of Bob. This happens  $\frac{1}{2}$  of the time, and among those times,  $\frac{1}{2}$  of the time the person is also a friend of Bob. Thus, the probability that any friend of Alice is also a friend of Bob is

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

Every single one of the 100 people must satisfy this criteria independently:

$$\left(\frac{3}{4}\right)^{100}$$

(c) This question is identical to (b). Taking the universe to be C, we have

$$P(A \cup B = C) = P(A^c \subseteq B)$$

In other words, A already covers a portion of people, and B must cover what is left out by A, which is  $A^c$ , by being a superset of what is left out. Since the chance of being friend with Alice is equal to the chance of being non-friend with Alice, then

$$\begin{aligned} P(A^c \subseteq B) &= P(A \subseteq B) \\ &= \left(\frac{3}{4}\right)^{100} \end{aligned}$$

### Exercise 34

(a) They are dependent. This is because knowing that an accident A occurred makes it more likely that it's a bad driver, which makes it more likely that he gets into a second accident B.

When they say that A and B are conditionally independent given G, it makes sense. This is because, now, a first accident A does not tell us more information about whether it's a good or bad driver (we already know it). Equivalently, the probability of having a second accident B is already determined by the knowledge that he is a good or bad driver.

(b)

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

(c)

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

Since  $A$  and  $B$  are conditionally independent given  $G$  (i.e.  $P(B|A) = P(B)$  given  $G$ ), we have

$$P(B|G, A^c) = P(B|G)$$

$$P(B|G^c, A^c) = P(B|G^c)$$

Hence, we end up with

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

### Exercise 35

(a)

$$\begin{aligned} P(W) &= P(W|B)P(B) + P(W|I)P(I) + P(W|M)P(M) \\ &= 90\% \frac{1}{3} + 50\% \frac{1}{3} + 30\% \frac{1}{3} \\ &= \frac{17}{30} \end{aligned}$$

(b) Let  $H$  be the event that I won a game already.  $P(H) = \frac{17}{30}$ . I agree that the letters are confusing here. Sometimes, probabilities with  $H$  are equivalent to probabilities with  $W$ .

$$\begin{aligned} &P(W|H) \\ &= P(W|B, H)P(B|H) + P(W|I, H)P(I|H) + P(W|M, H)P(M|H) \\ &= P(W|B)P(B|H) + P(W|I)P(I|H) + P(W|M)P(M|H) \\ &= P(W|B) \frac{P(H|B)P(B)}{P(H)} + P(W|I) \frac{P(H|I)P(I)}{P(H)} + P(W|M) \frac{P(H|M)P(M)}{P(H)} \\ &= 90\% \frac{90\% \frac{1}{3}}{\frac{17}{30}} + 50\% \frac{50\% \frac{1}{3}}{\frac{17}{30}} + 30\% \frac{30\% \frac{1}{3}}{\frac{17}{30}} \\ &\approx 68\% \end{aligned}$$

(c) If we assume that outcomes are independent, we are ignoring the information we can glean from winning the first game, which can give us hints to the level of the player. So this is false.

On the other hand, once we know the skill level of the opponent, then there is only one probability left, because there is no more weighing of probabilities.

### Exercise 36

(a) For a random person, there are 4 combinations of being good/bad at math and good/bad at baseball. If we condition on students that are admitted, then there are 3 combinations: good at both, good at only math, good at only baseball. Here, 2 of the 3 combinations are good at baseball. If we condition on math, then there are 2 combinations left: good at both, good at math. Here, only 1 of the 2 combinations is good at baseball.

Intuitively, the general proportion of students good at baseball is higher than if we focused only on the subset of students good at math, because we ignore the subset of students who are purely good at baseball.

(b) Given

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

$$C = A \cup B$$

we wish to show

$$P(A|B, C) \neq P(A, C)$$

in particular, we want

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

We derive the left:

$$\begin{aligned} P(A|B, C) &= P(A|B \cap (A \cup B)) \\ &= P(A|B) \end{aligned}$$

The right:

$$\begin{aligned} P(A|C) &= P(A|A \cup B) \\ &= \frac{P(A \cup B|A)P(A)}{P(B)} \\ &= (P(A|A) + P(B|A) - P(A \cap B|A)) \frac{P(A)}{P(B)} \\ &= \frac{P(A)}{P(B)} + P(A|B) - P(A \cap B|A) \frac{P(A)}{P(B)} \end{aligned}$$

Hence, the right expression differs from the left by this quantity

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

We need to show that this quantity is greater than zero. This is trivial:

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

$$1 > P(A \cap B|A)$$

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

which is evident.

## Chapter 3

# Random Variables and Distributions

### 3.1 PMFs and CDFs

#### Exercise 1

Each previous person from 1 to  $k - 1$  must have different birthdays. The  $k$ -th person must choose one of the  $k - 1$  birthdays to copy.

$$\begin{aligned} P(X = k) &= \underbrace{\left( \frac{365}{365} \frac{364}{365} \frac{363}{365} \cdots \frac{365 - (k - 1) + 1}{365} \right)}_{k-1 \text{ terms}} \frac{k - 1}{365} \\ &= \frac{\frac{365!}{(365 - (k - 1))!} (k - 1)}{365^k} \end{aligned}$$

#### Exercise 2

(a)  $k - 1$  failures, 1 success:

$$P(X = k) = \left(\frac{1}{2}\right)^{k-1} \frac{1}{2} = \left(\frac{1}{2}\right)^k$$

(b) If we start with success, then we must continue with success. As soon as we hit failure, the experiment stops. Likewise, if we start with failure, we must continue with failure. Thus, the 1st choice does not matter; what matters is that the following 2 to  $k - 1$  results ( $k - 2$  terms in total) are equal to the 1st result, and the  $k$ -th result is the opposite.

$$P(X = k) = \left(\frac{1}{2}\right)^{k-2} \frac{1}{2} = \left(\frac{1}{2}\right)^{k-1}$$

**Exercise 3**

$$\begin{aligned} G(Y = \mu + \sigma k) &= F(X = k) \\ \Rightarrow G(Y = k) &= F\left(X = \frac{k - \mu}{\sigma}\right) \end{aligned}$$

**Exercise 4**

This is the CDF of the uniform distribution for elements 1 to  $n$ :

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

Let's verify its CDF:

$$F(x) = P(X \leq x) = P(X \leq \lfloor x \rfloor) = \sum_{k=1}^{\lfloor x \rfloor} P(X = k) = \sum_{k=1}^{\lfloor x \rfloor} \frac{1}{n} = \frac{\lfloor x \rfloor}{n}$$

**Exercise 5**

(a) Clearly, it's non-negative. We need only to show that it sums to 1:

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{k+1} = \frac{1}{2} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = \frac{1}{2} \frac{1}{1 - \frac{1}{2}} = \frac{1}{2} \frac{1}{\frac{1}{2}} = 1$$

(b) We could use calculus to get the sum of geometric series. But here,  $p = \frac{1}{2}$ , so each iteration, half of the remaining area gets covered. Hence, the remaining area is halved every time.

$$F(x) = 1 - \left(\frac{1}{2}\right)^x$$

**Exercise 6**

If  $j+1 > j$ , then  $\frac{j+1}{j} > 1$ . The log of numbers greater than 1 is positive, hence the first property is satisfied.

To verify the second property, we simply use log product rule.

$$\sum_{j=1}^9 \log_{10}\left(\frac{j+1}{j}\right) = \log_{10}\left(\frac{2}{1} \frac{3}{2} \cdots \frac{10}{9}\right) = \log_{10}\left(\frac{10}{1}\right) = 1$$

**Exercise 7**

For  $X = k \in \{1, 2, 3, 4, 5, 6\}$

$$P(X = k) = \left(\prod_{j=1}^{k-1} p_j\right) (1 - p_k)$$

For  $X = 7$ , we remove the factor  $1 - p_k$ .

**Exercise 8**

If the most valuable is  $k$ , there is only one choice for the biggest of the 5 prizes (namely,  $k$ ), then all 4 other prizes must be smaller than  $k$ .

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

**Exercise 9**

- (a) Eventually,  $F_1(x) = 1$  and  $F_2(x) = 1$ . Hence,  $F(x) = p + (1 - p) = 1$  eventually. They also start at zero. So the first property is satisfied.

Also, since  $F_1$  and  $F_2$  are monotonically increasing, then any scalar multiple of them is increasing. The addition of increasing and non-negative function is increasing.

- (b) This is an application of part (a).

**Exercise 10**

- (a) The probability of  $n$  is proportional to  $\frac{1}{n}$ :

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

The sum must converge to 1:

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

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

This series diverges, thus there exist no such distribution.

- (b) This time, we have

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

$$= C \frac{\pi^2}{6}$$

So we can choose

$$C = \frac{6}{\pi^2}$$

**Exercise 11**

For CDF, we have

$$F(x) = P(X \leq x)$$

For the other one, we have

$$G(x) = P(X < x)$$

When  $x$  reaches an element  $n$  of the support,  $F$  adds the probability of  $n$  immediately, but  $G$  stays the same. When  $x$  nudges past  $n$ , then  $F$  stays the same, but  $G$  increase by the probability of  $n$ . Hence, the difference is that the staircase is left-continuous for  $G$ , whereas it is right-continuous for  $F$ .

Let  $S$  denote the support. To get  $G$  from  $F$ :

$$G(x) = \begin{cases} F(x) & x \notin S \\ F(x) - P(x) & x \in S \end{cases}$$

**Exercise 12**

(a)

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

$$P(Y = 0) = 0$$

where  $P(X = k)$  can be any distribution that is non-zero for every  $k \in \mathbb{N}$ . The CDF of  $Y$  will always be delayed by 1.

(b) It is impossible. The probabilities have to sum to 1. If the probability of  $Y$  sum to 1, then the probabilities of  $X$  sum to less than 1, since the inequality is strict for some.

**Exercise 13**

For all  $a$ , with  $z_i$  which are obviously disjoint:

$$\begin{aligned} P(X = a) &= P(X = a \mid Z = z_1) + P(X = a \mid Z = z_2) + \dots \\ &= P(Y = a \mid Z = z_1) + P(Y = a \mid Z = z_2) + \dots \\ &= P(Y = a) \end{aligned}$$

Since it works for all  $a$  (their domain), their probability distribution is the same.

**Exercise 14**

(a) For the first:

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

$$\begin{aligned}
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \left( \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - \frac{\lambda^0}{0!} \right) \\
&= e^{-\lambda}(e^\lambda - 1) \\
&= 1 - e^{-\lambda}
\end{aligned}$$

For the second:

$$\begin{aligned}
P(X \geq 2) &= \sum_{k=2}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} \\
&= e^{-\lambda} \left( \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - \frac{\lambda^0}{0!} - \frac{\lambda^1}{1!} \right) \\
&= e^{-\lambda}(e^\lambda - 1 - \lambda) \\
&= 1 - e^{-\lambda} - \lambda e^{-\lambda}
\end{aligned}$$

(b) I don't know.

## 3.2 Named Distributions

### Exercise 15

$$F(x) = \frac{\lfloor x \rfloor}{n}$$

### Exercise 16

$$\begin{aligned}
P(X = x \mid X \in B) &= \frac{P(X = x, X \in B)}{P(X \in B)} \\
&= \frac{P(X = x)}{P(X \in B)} = \frac{\frac{1}{|X|}}{\frac{|B|}{|X|}} = \frac{1}{|B|}
\end{aligned}$$

### Exercise 17

Let  $X$  be the number of people who show up.  $X \sim \text{Bin}(110, 0.9)$ . We want  $P(X \leq 100)$ :

$$\begin{aligned}
P(X \leq 100) &= 1 - P(X > 100) = 1 - \sum_{k=101}^{110} P(X = k) \\
&= 1 - \sum_{k=101}^{110} \binom{110}{k} 0.9^k (1 - 0.9)^{110-k}
\end{aligned}$$

The calculator tells us:

$$\approx 67\%$$



# Chapter 4

## Expectation

### 4.1 Expectations and Variances

#### Exercise 1

$$\mathbb{E}X = \sum_{k=0}^2 p(k) \cdot k = \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 2 = 1$$

$$\sigma^2 = \mathbb{E}[(X - \mathbb{E}X)^2] = \sum_{k=0}^2 p(k) \cdot (k - 1)^2 = \frac{2}{3}$$

or using the other derived formula:

$$\sigma^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \left( \sum_{k=0}^2 p(k) \cdot k^2 \right) - 1^2 = \frac{5}{3} - 1 = \frac{2}{3}$$

#### Exercise 2

$$\mathbb{E}X = \frac{3}{4} \cdot 365 + \frac{1}{4} \cdot 366 = 365.25$$

$$\sigma^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \left( \frac{3}{4} \cdot 365^2 + \frac{1}{4} \cdot 366^2 \right) - 365.25^2 = 0.1875$$

#### Exercise 3

(a)

$$\mathbb{E}X = \frac{1}{6} \cdot 1 + \dots + \frac{1}{6} \cdot 6 = 3.5$$

(b)

$$\mathbb{E}Y = 4 \cdot \mathbb{E}X = 4 \cdot 3.5 = 14$$

**Exercise 4**

We always roll when the expected value of the future is higher than the current value. We can use a recursive approach.

Suppose I'm at my third roll. The fourth dice expectation is 3.5. So if I roll lower than this, then I reroll. Else, I keep the value. The expected value is

$$\mathbb{E}X_3 = \frac{1}{6} \cdot 6 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 4 + \frac{3}{6} \cdot 3.5 = 4.25$$

Suppose I'm at my second roll. The third dice expectation is 4.25. So if I roll lower than this, then I reroll. The expected value is

$$\mathbb{E}X_2 = \frac{1}{6} \cdot 6 + \frac{1}{6} \cdot 5 + \frac{4}{6} \cdot 4.25 = 4.67$$

Suppose I'm at my first roll. The second dice expectation is 4.67. The expected value is

$$\mathbb{E}X_1 = \frac{1}{6} \cdot 6 + \frac{1}{6} \cdot 5 + \frac{4}{6} \cdot 4.67 = 4.95$$

Hence, the expectation value of this experiment is around 4.95.

**Exercise 5**

$$\begin{aligned}\mathbb{E}X &= \sum_{k=1}^n \frac{k}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2} \\ \sigma^2 &= \sum_{k=1}^n \frac{k^2}{n} - (\mathbb{E}X)^2 = \frac{\frac{n(n+1)(2n+1)}{6}}{n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2 - 1}{12}\end{aligned}$$

**Exercise 6**

Obviously,  $P(X < 4) = 0$ . For others:

$$P(X = 4) = 2 \cdot \left(\frac{1}{2}\right)^4$$

$$P(X = 5) = 2 \cdot \binom{4}{1} \left(\frac{1}{2}\right)^5$$

$$P(X = 6) = 2 \cdot \binom{5}{2} \left(\frac{1}{2}\right)^6$$

$$P(X = 7) = 2 \cdot \binom{6}{3} \left(\frac{1}{2}\right)^7$$

The mean is 5.8125. The variance is 1.02734375.

**Exercise 7**

(a)

$$P(X = k)$$

$$= P(X = k \mid F = 3)P(F = 3) +$$

$$P(X = k \mid F = 2)P(F = 2) +$$

$$P(X = k \mid F = 1)P(F = 1)$$

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

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

$$P(X = 3) = \frac{1}{3} \frac{20}{100} = \frac{1}{15}$$

The mean is 1.45. The variance is approximately 0.380833.

(b) Total children:  $30 + 50 \cdot 2 + 20 \cdot 3 = 190$ . The difference with the previous question is that we are no longer choosing a family. Instead, we are choosing the child, then looking at the family in which they belong. Families with more children are more likely to have their child chosen, even if that particular type of family is less common.

$$P(X = k)$$

$$= P(X = k \mid C \in F3)P(C \in F3) +$$

$$P(X = k \mid C \in F2)P(C \in F2) +$$

$$P(X = k \mid C \in F1)P(C \in F1)$$

$$P(X = 1) = \frac{1}{3} \frac{60}{190} + \frac{1}{2} \frac{100}{190} + \frac{1}{1} \frac{30}{190} \approx 0.5263$$

$$P(X = 2) = \frac{1}{3} \frac{60}{190} + \frac{1}{2} \frac{100}{190} \approx 0.3684$$

$$P(X = 3) = \frac{1}{3} \frac{60}{190} \approx 0.1053$$

The mean is approximately 1.579. The variance is approximately 0.4544.

**Exercise 8**

- (a) There are 20 million people. There are 10 cities. The average population per city is

$$\frac{20 \text{ million}}{10} = 2 \text{ million}$$

- (b) Only the population per region is given, so it is impossible to know the distribution within each region's cities.

- (c) The averages of each region are:

$$N_{avg} = \frac{3 \text{ million}}{4} = 750000$$

$$E_{avg} = \frac{4 \text{ million}}{3} = 1333333.33\dots$$

$$S_{avg} = \frac{5 \text{ million}}{2} = 2500000$$

$$W_{avg} = \frac{8 \text{ million}}{1} = 8000000$$

Choosing a region uniformly means  $\frac{1}{4}$  for each.

$$\mathbb{E}X = N_{avg}P(N) + E_{avg}P(E) + S_{avg}P(S) + W_{avg}P(W) = 3145833.33\dots$$

- (d) Now, the regions with less cities are given the same weight as other regions with more cities. Thus, the cities are not chosen uniformly anymore. Region D, containing only one city of 8 million, is now weighing  $\frac{1}{4}$ , whereas before it was only  $\frac{1}{10}$  of the cities.

**Exercise 9**

Option A: He walks away with 16000\$. This one obviously has the lowest variance.

Option B:  $\frac{1}{2}$  chance to leave with 1000\$.  $\frac{1}{2}$  chance to move on to the next question, which itself has  $\frac{3}{4}$  chance to get 32000\$ and  $\frac{1}{4}$  chance to get 64000\$. The expected value is  $\frac{1}{2} \cdot 1000 + \frac{1}{2} (\frac{3}{4} \cdot 32000 + \frac{1}{4} \cdot 64000) = 20500$ .

Option C:  $\frac{3}{4}$  chance to leave with 1000\$.  $\frac{1}{4}$  chance to move on to the next question, which itself has  $\frac{1}{2}$  chance to get 32000\$ and  $\frac{1}{2}$  chance to get 64000\$. The expected value is  $\frac{3}{4} \cdot 1000 + \frac{1}{4} (\frac{1}{2} \cdot 32000 + \frac{1}{2} \cdot 64000) = 12750$ .

**Exercise 13**

Yes. Let

$$P(X = 1) = 1$$

Let

$$P(Y = 2) = 0.99$$

$$P(Y = -999999999999) = 0.01$$

The expected value of  $X$  is 1. The expected value of  $Y$  is  $0.99 \cdot 2 + 0.01 \cdot -999999999999 = -99999999998$ . Obviously, the expected value of  $X$  is greater than the expected value of  $Y$  by at least a factor of 100. But  $Y$  is greater than  $X$  at 99 percent of the time.

## 4.2 Named Distributions

### Exercise 18

$$P(X = k) = 2 \cdot \left(\frac{1}{2}\right)^{k-1} \frac{1}{2} = \left(\frac{1}{2}\right)^{k-1}$$

$$\mathbb{E}X = \sum_{k=1}^{\infty} k \cdot P(X = k) = \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^{k-1}$$

$$\sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^{k-1} = \sum_{k=0}^{\infty} k \left(\frac{1}{2}\right)^{k-1} = \frac{1}{(1 - \frac{1}{2})^2} = 4$$

### Exercise 20

No, for all three questions. Simply because, under each of the four distributions, each integer (from 0 to 100) has a non-zero probability of being reached. Hence, we cannot have  $X$  always greater than  $Y$ , because there is a non-zero chance that  $Y$  lands on 100, and  $X$  lands on 0.

## 4.3 Indicator Random Variables

### Exercise 34

We can sum up the probability of each box being empty, through the usage of indicator variables. For the  $i$ -th box to be empty, all the  $k$  balls must be placed in the  $n - 1$  other boxes.  $I(empty_i) = 1$  indicates the emptiness of box  $i$ . The expectation of the indicator is equal to its probability of being 1, which is

$$I(empty_i) = \left(\frac{n-1}{n}\right)^k$$

The expected number of empty boxes is the sum of these individual expectations:

$$\mathbb{E} = \sum_{i=1}^n \left(\frac{n-1}{n}\right)^k$$

**Exercise 35**

There are  $\binom{50}{2}$  pairs of people. The probability that a pair shares the same birthday is  $\frac{1}{365}$ . The expectation of the corresponding indicator is thus

$$I = \frac{1}{365}$$

The expected number of shared birthdays is the sum of the indicators for all the pairs

$$\mathbb{E} = \binom{50}{2} \frac{1}{365}$$

**Exercise 36**

Each pair of indicator is independent, since knowing that two people have the same birthday does not give us any more information about the fourth person having the same birthday as the first or second or third. However, if we have more than a pair, then they are not necessarily independent. For instance, we can make a chain  $I_{12}$  and  $I_{23}$ , which will give us information about  $I_{13}$ .

**Exercise 37**

The average number of bags for the first three students is simply the sum of all their bags divided by 3. The probability that a bag lands among the three is  $\frac{3}{20}$ , which is also the expectation of the indicator variable that a particular bags lands among the three. The expected total bag among the three is the sum over the 20 bags

$$20 \cdot \frac{3}{20} = 3$$

The average is thus

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

The expected number of students who get at least one bag is the sum of the expectation of the indicators that each student gets at least a bag. The probability of getting at least one bag is the complement of getting no bag.

$$I = 1 - P(0) = 1 - \left(\frac{19}{20}\right)^{20}$$

Taking the sum, we have that the expected number of students with at least one bag is

$$20 \cdot \left(1 - \left(\frac{19}{20}\right)^{20}\right) \approx 12.83$$

## 4.4 LOTUS

### Exercise 61

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

For  $|\lambda| < 1$ , we have

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

### Exercise 62

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

### Exercise 64

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \sum_{k=1}^{\infty} e^{tk} P(X = k) \\ &= \sum_{k=1}^{\infty} e^{tk} (1-p)^{k-1} p \\ &= pe^t \sum_{k=1}^{\infty} e^{t(k-1)} (1-p)^{k-1} \\ &= pe^t \sum_{k=0}^{\infty} (e^t(1-p))^k\end{aligned}$$

For  $|e^t(1-p)| < 1$ , we have

$$f(t) = \frac{pe^t}{1 - e^t(1-p)}$$

**Exercise 65**

(a)

$$\begin{aligned}
\mathbb{E}(Y^2) &= \sum_{k=1}^{\infty} k^2 P(Y = k) = \sum_{k=0}^{\infty} k^2 P(Y = k) \\
&= \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda} \lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= e^{-\lambda} (\lambda^2 e^\lambda) \\
&= \lambda^2
\end{aligned}$$

(b)

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{Y}\right) &= \sum_{k=1}^{\infty} \frac{1}{k} P(Y = k) \\
&= \sum_{k=1}^{\infty} \frac{1}{k} \frac{e^{-\lambda} \lambda^k}{k!} \\
&= e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{1}{k} \frac{\lambda^{k-1}}{k!} \\
&= e^{-\lambda} \lambda \left(\frac{e^\lambda}{\lambda^2}\right) \\
&= \frac{1}{\lambda}
\end{aligned}$$

**4.5 Poisson Approximation****Exercise 67**

(a) Let class 1 be fixed. There are  $100!$  permutations of class 2. We want to count the portion of permutations of class 2 where none of the students sit at the same place as class 1. Student 1 has 99 choices of seats in class 2 which are not his original seat. Call his choice "seat i".

Consider "student i" who originally sat in "seat i" in class 1. There are two possibilities to consider for "student i" in class 2. If "student i" chooses student 1's seat in class 1 for his class 2, then we are left with finding the permutations of the remaining 98 students in class 2 (since they effectively swapped places).

In any other case, "seat i" is taken by student 1, so our space reduces to seats

$$1, 2, \dots, i-1, i+1, \dots, n-1, n$$

(i.e. excluding "seat i"). Call this space  $N - 1$ . In this space, we want to prevent "student i" from going to seat 1 (since that's our previous case). So we put "student i" at seat 1 and pretend he actually had seat 1 in class 1. This effectively prompts him to choose a seat other than seat 1, but also prevents him from choosing "seat i" because our space now doesn't have "seat i" anymore! This effectively reduces the problem to permuting the student in the space  $N - 1$ , where "student i" takes student 1's place. Hence,

$$D(n) = (n - 1)(D(n - 1) + D(n - 2))$$

Using induction, we can prove that it's equal to the formula

$$n! \sum_{k=0}^n \frac{(-1)^k}{k!}$$

Base case works:

$$D(1) = 0 \quad \text{obviously}$$

$$n! \sum_{k=0}^1 \frac{(-1)^k}{k!} = 1(1 - 1) = 0$$

$$D(2) = 1 \quad \text{obviously}$$

$$n! \sum_{k=0}^2 \frac{(-1)^k}{k!} = 2(1 - 1 + \frac{1}{2}) = 1$$

Now, suppose

$$D(n) = n! \sum_{k=0}^n \frac{(-1)^k}{k!}$$

$$D(n - 1) = (n - 1)! \sum_{k=0}^{n-1} \frac{(-1)^k}{k!}$$

Then

$$\begin{aligned} D(n + 1) &= n \left( n! \sum_{k=0}^n \frac{(-1)^k}{k!} + (n - 1)! \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} \right) \\ &= n \cdot n! \sum_{k=0}^n \frac{(-1)^k}{k!} + n! \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} \\ &= n \cdot n! \sum_{k=0}^n \frac{(-1)^k}{k!} + n! \left( \sum_{k=0}^n \frac{(-1)^k}{k!} - \frac{(-1)^n}{n!} \right) \\ &= (n + 1) \left( n! \sum_{k=0}^n \frac{(-1)^k}{k!} \right) - (-1)^n \end{aligned}$$

$$\begin{aligned}
&= (n+1)! \left( \sum_{k=0}^{n+1} \frac{(-1)^k}{k!} - \frac{(-1)^{n+1}}{(n+1)!} \right) - (-1)^n \\
&= (n+1)! \left( \sum_{k=0}^{n+1} \frac{(-1)^k}{k!} \right) - (-1)^{n+1} - (-1)^n
\end{aligned}$$

One of the  $n$  or  $n+1$  is odd, and the other is even, so the  $(-1)$  cancel out.  
Hence

$$D(n+1) = (n+1)! \sum_{k=0}^{n+1} \frac{(-1)^k}{k!}$$

Now that we have proved the formula, we can simply plug in 100:

$$D(100) = 100! \sum_{k=0}^{100} \frac{(-1)^k}{k!}$$

The probability is

$$P = \frac{D(100)}{100!} = \sum_{k=0}^{100} \frac{(-1)^k}{k!}$$

(b) This is the beginning of the taylor series of  $e^x$ . For  $n$  very large, the sum approaches

$$e^{-1} \approx 0.3679$$

## Chapter 5

# Continuous Random Variables

### 5.1 PDFs and CDFs

#### Exercise 1

(a) The CDF of Rayleigh is

$$F(x) = 1 - e^{-\frac{x^2}{2}}$$

Hence

$$P(1 < X < 3) = F(3) - F(1) = e^{-\frac{1^2}{2}} - e^{-\frac{3^2}{2}} \approx 0.5954$$

(b) We want  $F(x) = \frac{1}{4}, \frac{2}{4}, \frac{3}{4}$ . The inverse function (percentile function) is

$$G(y) = \sqrt{-2 \log(1 - y)}$$

Plugging in the values, we get

$$G\left(\frac{1}{4}\right) = 0.4999$$

$$G\left(\frac{2}{4}\right) = 0.7759$$

$$G\left(\frac{3}{4}\right) = 1.0973$$

#### Exercise 2

(a) Uniform random number generator between 0 and 0.5. It has PDF:

$$P(x) = 2$$

for all  $x$ .

(b) If  $P(x) > 1$  for all  $x$ , then its integral over the interval  $[a, b]$  is

$$\int_a^b P(x)dx > \int_a^b 1dx = b - a$$

Since  $P$  is a probability distribution, we have

$$1 \geq \int_a^b P(x)dx$$

So

$$1 > b - a$$

### Exercise 3

(a) First,  $F$  and  $f$  are non-negative, so  $g$  is non-negative. Next, we integrate

$$\begin{aligned} & \int_{-\infty}^{\infty} 2F(x)f(x)dx \\ &= 2 \int_{-\infty}^{\infty} F(x)F'(x)dx \\ &= 2 \left( [F(t)F(t)]_{t=-\infty}^{t=\infty} - \int_{-\infty}^{\infty} F'(x)F(x)dx \right) \\ &= 2 \left( 1 - \int_{-\infty}^{\infty} f(x)F(x)dx \right) \end{aligned}$$

Isolating the integral, we get

$$\int_{-\infty}^{\infty} f(x)F(x)dx = \frac{1}{2}$$

So,

$$\int_{-\infty}^{\infty} 2f(x)F(x)dx = 1$$

(b)  $f$  is positive, so  $h$  is positive. Next, we integrate

$$\begin{aligned} & \int_{-\infty}^{\infty} \left( \frac{1}{2}f(-x) + \frac{1}{2}f(x) \right) dx \\ &= \frac{1}{2} \int_{-\infty}^{\infty} f(-x)dx + \frac{1}{2} \int_{-\infty}^{\infty} f(x)dx \\ &= \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 \\ &= 1 \end{aligned}$$