STA 473 - Probability

Derek L. Sonderegger

2017-04-24

Contents

Preface

This is a set of questions to be used in an *Inquiry Based Learning* class for an undergraduate level course in probability.

Acknowledgements

Many people have helped. I should thank them.

Chapter 1

Introduction to Probability

1.1 History of Probability

1.2 Interpretations of Probability

1.3 Experiments and Events

1.4 Review of Set Theory (D&S 1.4)

1. Create a sample space $\mathcal{S}$ where
   a) Where the the number of outcomes is finite.
   b) Define events (subsets of $\mathcal{S}$) that do not have a 1-to-1 correspondence with the outcomes.
2. Create a sample space $\mathcal{S}$ where
   a) Where the the number of outcomes is countably infinite.
   b) Define a finite number of events (subsets of $\mathcal{S}$) that do not have a 1-to-1 correspondence with the outcomes and that the union of all your events is $\mathcal{S}$.
   c) Define an infinite number of events (subsets of $\mathcal{S}$) that do not have a 1-to-1 correspondence with the outcomes and that the union of all your events is $\mathcal{S}$.
3. Create a sample space $\mathcal{S}$ where
   a) Where the the number of outcomes is uncountably infinite.
   b) Define a finite number of events (subsets of $\mathcal{S}$) that do not have a 1-to-1 correspondence with the outcomes and that the union of all your events is $\mathcal{S}$.
   c) Define an countably infinite number of events (subsets of $\mathcal{S}$) that do not have a 1-to-1 correspondence with the outcomes and that the union of all your events is $\mathcal{S}$.

It is time to define the set of events more carefully. The take-home idea is that if you add an event, say A, you also add some other events related to A. The rules are summarized below.

- $\mathcal{S}$ is an event. This is to say *something* will happen.
- If A is an event, then A^c is also an event.
- If A_i is a countable sequence of events, then $\cup_{i=1}^{\infty} A_i$ is also an event.

4. Prove that $\emptyset$ is is an event.

5. Prove two of the conclusions of theorem 1.4.4. I would expect a proof of $A \cup A^c = \mathcal{S}$ to look something like, *"Let e be an arbitrary event in $\mathcal{S}$. Due to the nature of complements, either $e \in A$ or $e \in A^c$.*

Therefore $e in A \cup A^c$ *but because* e *was an arbitrary element of* $\mathcal{S}$ *then* $\mathcal{S} \subset A \cup A^c$*. However because* $\mathcal{S}$ *is the set of all possible events, then* $A \cup A^c \subset \mathcal{S}$ *and thus* $A \cup A^c = \mathcal{S}$*."*

6. Chapter problem 1.4.1. Suppose $A \subset B$. Show that $B^c \subset A^c$. *Do this in a similar fashion as problem 5.*
7. Chapter problem 1.4.2. *Show this by Venn diagrams.*
8. Chpater problem 1.4.3. Prove DeMorgan's Laws. *Prove this via Venn diagrams.*
9. Chapter problem 1.4.6.
10. Chapter problem 1.4.7
11. Chapter problem 1.4.13
12. Chapter problem 1.4.14

1.5 Definition of Probability (D&S 1.5)

Axiom 1 For every event A, the probabilitity of the event, denoted $Pr(A)$ has the property $Pr(A) \geq 0$

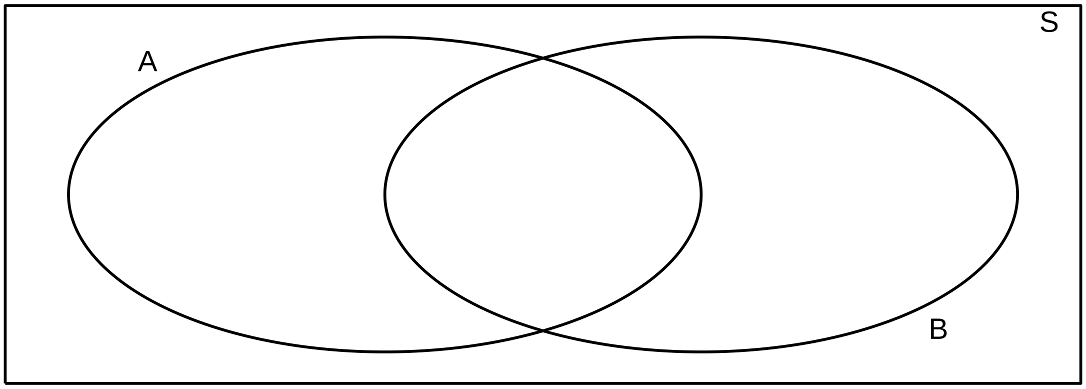
Axiom 2 If an event is sure to occur, then the event has probability 1. That is, $Pr(\mathcal{S}) = 1$.

Axiom 3 For every finite or countably infinite sequence of events A_1, A_2, $\dots$ where $A_i \cap A_j = \emptyset \;\; \forall \, i, j$ (that is the sequence A_i is pairwise disjoint), then

$$Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr\left(A_i\right)$$

1. Consider drawing a single card from a well shuffled deck of playing cards (4 suits, each with 13 cards Ace, two, ..., Queen, King). Consider the events H, S, C, D, which are drawing a Heart, Spade, Club, and Diamond. Explain why H and S are disjoint but H^c and S^c are not.
2. Prove $Pr(\emptyset) = 0$
3. Argue that *Axiom 3* should have been "For every countably infinite sequence of events A_i" because you can pad any finite sequence with an infinite sequence of empty sets.
4. Prove $Pr(A^c) = 1 - Pr(A)$

Often we will draw Venn diagrams where the area of the event is its probability.



Many of the probability calculations can be most easily understood using a Venn diagram along with the algebraic proof.

5. Prove if $A \subset B$ then $Pr(A) \leq Pr(B)$. *Show this formally and via Venn diagrams*

6. Prove that $Pr(A) = Pr(A \cap B) + Pr(A \cap B^c)$ *Show this formally and via Venn diagrams*

7. Prove that $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$ *Show this formally and via Venn diagrams. Notice our Axiom 3 addresses the case where A and B are disjoint.*

8. Consider events A and B where $Pr(A) = 1/3$ and $Pr(B) = 1/2$. Determine the value of $Pr(A \cap B^c)$ when

   a) A and B are disjoint
   b) $A \subset B$
   c) $Pr(A \cap B) = 1/8$

9. Suppose that Adam has a probability of failing an exam of $Pr(A) = 0.5$ while Bob only has a probability of failing the exam of $Pr(B) = 0.2$. Suppose the probability of both students failing is 0.1

   a) What is the probability that at least one of these two students will fail?
   b) What is the probability that neither student will fail?
   c) What is the probability that exactly one student will fail?

10. A point (x, y) is to be selected from the unit square $\mathcal{S}$ $(0 \le x \le 1,\ 0 \le y \le 1)$. Suppose that the probability that the point is selected from a specific region is equal to the area of the region. Find the probabiliy the point selected is from each of the following regions:

    a) (x, y) such that $(x - 1/2)^2 + (y - 1/2)^2 \ge 1/4$
    b) (x, y) such that $1/2 \le x + y \le 3/2$
    c) (x, y) such that $y \le 1 - x^2$
    d) (x, y) such that $x = y$

11. Bonferroni's Inequality. Let $A_1, A_2, \ldots$ be an arbitrary infinite squence of events. Define the seqence of events $B_1, B_2, \ldots$ as

$$B_1 = A_1$$

$$B_2 = A_1^c \cap A_2$$

$$B_3 = A_1^c \cap A_2^c \cap A_3$$

$$B_4 = A_1^c \cap A_2^c \cap A_3^c \cap A_4$$

    a) Prove that $B_i \subset A_i$, $B_i \cap B_j = \emptyset$ for $i \ne j$, and that $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i$.
    b) Prove that

$$Pr\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n Pr(B_i)$$

    c) Prove that

$$Pr\left(\bigcup_{i=1}^n A_i\right) \le \sum_{i=1}^n Pr(A_i)$$

    d) Using the previous result (b), prove that for sets $D_1, D_2, \ldots, D_n$ that

$$Pr\left(\bigcap_{i=1}^n D_i\right) \ge 1 - \sum_{i=1}^n Pr(D_i^c)$$

1.6 Finite Sample Spaces (D&S 1.6)

When dealing with sample spaces with only a finite number of outcomes (say n outputcomes s_i), it is often convenient to define each outcome as an event.

Let s_i be outcomes in the sample space $\mathcal{S}$. Let each of these outcomes have probability p_i.

For the axioms of probability to hold then:

$$p_i \geq 0$$

$$\sum_{i=1}^{n} p_i = 1$$

1. When fair dice, we assume that each side has equal probability of being rolled. For rolling a 6-sided die, what is the probability of rolling an even number?

2. When rolling two (or more) differently colored dice, we assume that the die do not affect the outcome of the other and that every pair of is equally likely. Alternatively you can think of rolling 1 die and then the other. So for rolling two six sided dice, there are 36 possible rolls, and notice, for example, $(2, 3)$ is a different roll that $(3, 2)$. What is the probability that the sum of the two rolls is even?

3. If a fair coin is flipped three times...

   a) What are the possible outcomes (enumerate these)?
   b) Explain why it is reasonable that each outcome is equally probable?
   c) What is the probability that all three faces will be the same?

1.7 Ordered Counting (D&S 1.7)

Often we situations where it is reasonable to beleve that each outcome of an event is equally likely and therefore we can figure out the probability if we knew how many events there were. E.g. there are 36 different outcomes for rolling two fair 6-sided dice, so each outcome has a 1/36 probability.

1. Prove/argue/justify that if the outcome of experiment is composed of 2 parts, where the first part has m outcomes $x_1, \ldots, x_m$ and the second part has n outcomes, $y_1, \ldots, y_n$ then there is a total of mn outcomes (x_i, y_j). This is often called the Multiplication Rule for Counting.

2. I own 3 pair of pants that are "work appropriate." I also own 6 different shirts and 5 pairs of shoes that are "work appropriate." How many different outfits are possible?

3. How many way can the numbers $1, 2, 3, 4,$ and 5 be arranged?

4. **Ordered Sampling without Replacement** For distinct objects $1, 2, \ldots, n$ prove that there are are $P_{n,k} = \frac{n!}{(n-k)!}$ arrangements of k elements (where the order is important, i.e. $1, 2, 3$ is distinct from $2, 1, 3$) and $n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot (2) \cdot (1)$ and by definition $0! = 1$. We call $P_{n,k}$ the number of *permutations* of k elements taken from a set of n distinct objects. *Hint: First consider base cases of $k = 1$ and then $k = 2$ and that the formula is appropriate. Then, to complete the induction arguement, show that if we have a $P_{n,k}$ permutations of k objects, then increasing to $k + 1$ elements simply results in $(n-k) \cdot P_{n,k}$ arrangements due to the Multiplation Rule of counting.*

5. From $n = 17$ students, one student will get a candy bar, another will get a soda, and a third will receive some gummi bears. How many different ways could the treats be distributed to the students?

6. Suppose we are going to randomly select 3 elements from the digits $0, 1, 2, \ldots, 9$ but we will select these *with replacement* (so we could get the 022). How many outcomes are there?

7. **Ordered Sampling with Replacement** Suppose that we have n distinct objects labeled $1, 2, \ldots, n$ and we are going to sample k of these objects *with replacement.* Justify/derive a formula for the number of outcomes.

8. Consider the sequence of numbers $0000, 0001, 0002, \ldots, 9998, 9999$. How many of these numbers are composed of 4 different digits?

Often times I am interested in calculating the probability of a particular event and we can often do it in the following manner:

- First count the number of equally likely outcomes there are.

- Count the number of outcomes where the event of interest occures.
- Then calculate

$$Pr\,(\text{Event}) = \frac{\text{Number of outcomes where event happens}}{\text{Total number of equally likely outcomes}}$$

9. As I work in the evenings, I often listen to music. Suppose that I have a playlist of $n = 300$ songs and I listen to them on shuffle where the software always selects from the list with equal probability when selecting which song to play next. If I listen to $k = 10$ songs, what is the probability that at least one of the songs will be duplicated? What about if I listen to $k = 30$ songs?

10. If 14 balls are randomly thrown into 25 boxes such that there is equal chance for a ball to land in any box, what is the probability that no box recieves more than one ball?

1.8 Combinations (D&S 1.8)

Often we want to count the number of arrangements of k elements selected without replacement from n distinct objects but where the order doesn't matter. Another way of saying this is that we want to count the number of sets of size k taken from n distinct objects.

1. **Unordered Sampling without Replacement**For a set of k elements, prove that there are $k!$ permutations of those elements. Using this information, argue that the number of distinct sets of k objects taken from n elements is (which the book denotes as $C_{n,k}$ and many others denote $\binom{n}{k}$) is

$$C_{n,k} = \binom{n}{k} = \frac{P_{n,k}}{k!} = \frac{n!}{k!(n-k)!}$$

2. I have 3 identical cans of soda that I will distribute randomly to 17 students. I will select (with equal probabilities per student) 3 students. How many ways could I choose 3 students?

3. Suppose I have a character string composed of only 0s and 1s. The character string is 20 characters long and 8 of them are 0s. How many different strings are there?

4. **Unordered Sampling with Replacement** Suppose that I have $n = 7$ boxes into which I will randomly throw $k = 3$ balls.

```
n=7 boxes
+----+-----+-----+-----+-----+-----+-----+
| 1  |  2  |  3  |  4  |  5  |  6  |  7  |
+----+-----+-----+-----+-----+-----+-----+
```

Now suppose that we throw, at random, $k = 3$ balls into the boxes. We might end up with one ball in box 3 and two in box 6.

```
 n=7 boxes
+----+-----+-----+-----+-----+-----+----+
|    |     | 0   |     |     | 00  |    |
+----+-----+-----+-----+-----+-----+----+
```

a) Argue that throwing k balls randomly into n boxes is equivalent to selecting a set of k elements from n distinct objects with replacement. *Hint show that every set chosen with resampling can be represented via boxes/balls and that every boxes/balls combination represents a possible set of k elements from n distinct objects with replacement.*

b) The guts of the boxes/balls diagram is the arrangement of box partitions and balls because the outer box walls don't matter because the balls get into a box.

```
|    |      | 0   |     |     | 00   |     |
```

which we can clean up a bit by remembering that the other walls have to be there and we'll represent the balls with 0 and the box partitions with a 1. |110111001|. This reduces the problem into how many binary strings can I produce with $n-1$ 1s and k 0s. How many are there?

5. Suppose that I will distribute my three soda cans to 17 students by drawing names out of a hat, but replacing the student's name after it is draw. How many different outcomes could occur?

6. Suppose I draw 2 cards from a standard deck of 52 cards, what is the probability that I draw two cards of the same suit?

7. Ten teams are playing in a tournament. In the first round there, there will be five games played. How many possible arrangements are there? What is the probability that the Ashville Avalanch plays the Boston Behemoths (these are two of the ten teams playing)?

8. Suppose that I flip a fair coin 10 times. What is the probability I observe 3 heads?

1.9 Multinomial Coefficients (D&S 1.9)

1. From the Math/Stat department faculty, a committee of 5 members is to be selected. There are 8 Math, 4 Statistics, and 4 Math Ed Professors. What is the probability that committee is composed of 3 Math, 1 Stats, and 1 Math Ed professor.

2. Suppose that we are creating a string of beads from 9 red, 7 blue, and 10 yellow beads. How many different arrangements can be made?

Chapter 2

Conditional Probability

2.1 Defining Conditional Probability (D&S 2.1)

1. Out of $n = 17$ students in a class, I will chose one at random student to give gummi bears to. In this class there are 12 men and 5 women. Student A is very interested in her probability of being selected. Denote the event A as being that the student gets the gummi bears. Denote event W as a woman is selected.
   a) What is the probability that student A is selected?
   b) What is the probability that a woman is selected?
   c) Suppose I restricted my selection to *only* the women? What is the probability that student A is selected. Can you write this probability as a function of your answers in parts (a) and (b)?
2. Consider the case where we take the sum of 2 six-sided dice. Define A as the event that the sum is greater than or equal to 9, and B being the event that the sum is greater than or equal to 6.
   a) Create a table of the possible outcomes. Notice that each outcome is equally likely. What is the probability that A occurs? *Use correct notation and leave it as a fraction* $\frac{???}{36}$
   b) Suppose that you are told that event B has occured. How many equally likely outcomes are there and what is the probability that A occurs? This will be denoted as $Pr(A|B)$. *(Leave this as a fraction).*
   c) Notice that you the numerator in answer in part (a) is the same as the numerator as you had in part (b), but the denominators are different. What number do you need to multiple your part (a) answer by to get your part (b) answer. How does this relate to $P(B)$?
3. It seems that my children (Casey and Elise) get sick with annoying frequency. Suppose the probability that my son Casey gets sick is $Pr(C) = 0.05$ and furthermore the probability that **both** children get sick is $Pr(E \cap C) = 0.03$ If Casey is sick right now, what is the probability that Elise is also sick?

When we talk about events A and B, we defined them as possible outcomes, but it isn't until we define probability of the events $Pr(A)$ $Pr(B)$ that we care about the sample space $\mathcal{S}$ of all possible events. What we are now trying to do is to claim that some addition knowledge allows us to refine the sample space to some smaller subset of $\mathcal{S}$, perhaps $B \subset \mathcal{S}$. So for events in B, we now need to re-scale all the probabilities to reflect that we now know that event B did happen.

If we previously know that $Pr(A \cap B) = 0.3$ and $Pr(B) = 0.6$, then half of the probability associated with B is overlapping with A. So if we just restrict ourselves to cases where B occurs, then the probability that A will occur is 1/2.

Notice that our notation $Pr(A|B)$ is addressing the refinement of the sample space, but we've just defined our notation this way. We have not, and will not ever define $A|B$ because event A is event A, regardless of the sample space we use to figure out its probability.

We formally define the conditional probability of A given B as

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} \quad \text{Assuming } Pr(B) \neq 0$$

If $Pr(B) = 0$ then the conditional probablity is undefined. Notice that this can be happily re-arranged to

$$Pr(A \cap B) = Pr(A|B)Pr(B)$$

I can also notice that I could condition on either event A or event B, so we could also have

$$Pr(A \cap B) = Pr(B|A)Pr(A)$$

4. If $B \subset A$ and $Pr(B) > 0$, what is $Pr(A|B)$
5. If A and B are disjoint and $Pr(B) > 0$, what is $Pr(A|B)$?
6. Suppose that events $A_1, A_2, \dots, A_n$ are events such that $Pr\left(\bigcap_{i=1}^{n} A_i\right) > 0$. Show that

$$Pr\left(\bigcap_{i=1}^{n} A_i\right) = Pr\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right) Pr\left(A_{n-1} \middle| \bigcap_{i=1}^{n-2} A_i\right) \dots Pr\left(A_3 \middle| A_2 \cap A_1\right) Pr\left(A_2 \middle| A_1\right) Pr\left(A_1\right)$$

7. For events A, B, and D such that $Pr(D) > 0$ show that:
   a) $Pr(A^C|D) = 1 - Pr(A|D)$.
   b) $Pr(A \cup B \Big| D) = Pr(A|D) + Pr(B|D) - Pr(A \cap B|D)$

Suppose that we have K events B_k such that the $B_1, B_2, \dots, B_K$ are disjoint and $\bigcup B_k = \mathcal{S}$. Then we call the events $B_1, \dots, B_K$ a *partition* of the sample space $\mathcal{S}$. A partition of $\mathcal{S}$ is often useful for caculating probabilities due to the disjoint nature of the B_k elements.

8. **Law of Total Probability** Prove that for for a partition $B_1, \dots, B_K$ of $\mathcal{S}$, that

$$Pr(A) = \sum_{k=1}^{K} Pr(A \cap B_k) = \sum_{k=1}^{K} Pr(A \Big| B_k) Pr(B_k)$$

Note: There is a conditional version of the Law of Total probability, which is proved in an analogous fashion:

$$Pr(A|C) = \sum_{k=1}^{K} Pr(A \cap B_k \big| C) = \sum_{k=1}^{K} Pr(A \big| B_k \cap C) Pr(B_k \big| C)$$

9. A child's bookshelf contains three shelves. On the shelves are $n_1 = 10$,$n_2 = 20$ and $n_3 = 30$ books. Within each set of books, there are some number of Dr Suess books $m_1 = 5$, $m_2 = 4$, $m_3 = 2$. The child will select a shelf at random (equal probability) and then from the shelf will select a book at random. What is the probability the child selects a Dr Suess book?

10. A camera with a motion detector was mounted facing a forest trail. 50% of the pictures were taken during the daytime, 15% were taken during twilight hours (dawn and dusk), and 35% were taken during the night. Of the pictures taken during the daytime, 80% were of hikers and 20% were of wild animals. Of the pictures taken at twilight 30% were of hikers and 70% were of wild animals. Finally, of the pictures taken during the night, 100% were wild animals.
    a) What is the probability that a randomly selected photo is of a hiker and was taken at twilight?
    b) What is the probability a photo was taken at night given that is of a wild animal?

11. I have kept track of the probabilities of how many cats will sit with me and/or my wife on the couch. Below is a table of probabilities.

	0 Cats	1 Cat	2 Cats	3 Cats
0 People	0.08	0.08	0.03	0.01
1 Person	0.1	0.25	0.125	0.025
2 People	0.03	0.09	0.12	0.06

a) What is the probability that one human is sitting on the couch?
b) What is the probability that at least two cats are sitting on the couch?
c) Given that there are two cats sitting on the couch, what is the probability that there are two humans also on the couch?

12. My cat Kaylee occasionally likes to sit on people's laps while they are seated at the table. My wife is strongly opposed to this and will scold the cat when she catches her in the act. Suppose that that Kaylee will select my lap 60% of the time and the remaining 40% of the time she jumps into my wife's lap. If Kaylee jumps into my wife's lap, there is a 100% chance of being scolded, while if she jumps into mine, there is only a 20% chance of being scolded. Given that Kaylee was just scolded for being in a lap, what is the probability she was in my wife's lap?

13. There are two brands of Mac & Cheese that my daughter will eat. When I go shopping I will pick from the two brands with a 70% probability of choosing the brand that I bought the previous time. The first time I went shopping, I chose from the two brands with equal probability. What is the probability that I chose brand A on the first and second trips, and brand B on the third and fourth trips?

2.2 Independence

Definition: Two events, A and B are independent if $Pr(A \cap B) = Pr(A)Pr(B)$.

Definition: Events $A_1, A_2, \ldots, A_K$ are pairwise independent if A_i and A_j are independent for any i, j.

Definition: Events $A_1, A_2, \ldots, A_K$ are mutually independent if for all subsets I of $1, 2, \ldots, K$, $Pr(\bigcap_{i \in I} A_i) = \prod_{i \in I} Pr(A_i)$

1. Show that if $Pr(A) > 0$ and $Pr(B) > 0$, then A and B are independent if and only if $Pr(A|B) = Pr(A)$ and $Pr(B|A) = Pr(B)$

2. Give an example of three events that are pairwise independent but not mutually independent.

Convention If I say that a set of events are “independent”“, then we intend to say”mutually independent“” but are being lazy.

3. Show that if A and B are indendent, then A and B^c are also independent.

4. Suppose that we flip a fair coin three times. Denote H_i as the event that I flip a head on the ith flip.

   a) Find $Pr(H_1 \cap H_2 \cap H_3)$
   b) Find $Pr(H_1 \cap H_2^c \cap H_3)$
   c) Find $Pr(H_1^c \cap H_2 \cap H_3)$
   d) How many ways can we have 2 heads?
   e) What is the probability of 2 heads?

5. I will roll a 20-sided die three times. Define the event H_i as the event that I roll a 17 or greater on the ith roll.

   a) Find $Pr(H_1 \cap H_2 \cap H_3)$
   b) Find $Pr(H_1 \cap H_2^c \cap H_3)$
   c) Find $Pr(H_1^c \cap H_2 \cap H_3)$
   d) How many ways can we have exactly 2 H events happen?
   e) What is the probability of exactly 2 H events happening?

6. I will roll my 20-sided die until I roll a 20.

   a) What is the probability that I roll a 20 on my first roll?
   b) What is the probability that the first 20 I roll is on the 5th roll?

7. A family has two children. It is known that at least one is a boy. What is the probability that the family has two boys, given that at least is one a boy? Assume that genders are equally likely and that genders of siblings are independent.

2.3 Bayes' Theorem

The goal of Bayes' Theorem is to reverse the order of conditioning. Suppose we are interested in two events A and B. We might be given some information about $P(A|B)$ but we want to know about $P(B|A)$.

1. **Bayes' Theorem** Prove that for two events A and B such that $Pr(A) > 0$ then

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

2. **Bayes' Theorem (general case)** Again consider an event A such that $Pr(A) > 0$. For an arbitrary partition of $\mathcal{S}$, say B_k $k = 1, \dots, K$, prove that

$$Pr(B_k|A) = \frac{Pr(A|B_k)Pr(B_k)}{\sum_{i=1}^{K} Pr(A|B_k)Pr(B_k)}$$

3. A softball team has two pitchers, Jeff and Bob. Of the two Jeff is the better pitcher and wins 80% of the games he pitches for but can only play in 30% of the games. Bob pitches in the rest of the games, but only wins 40% of his games.

   a) What is the probability that the softball team wins a game?
   b) Given that the team won, what is the probability that Jeff pitched?

4. One card is selected at random from a standard deck of 52 playing cards. It is inserted into a second standard deck and the second deck is then well shuffled.

   a) A card is drawn at random from the second deck. What is the probability it is an ace?
   b) Given that an ace was drawn from the second deck, what is the probability that an ace was transfered from the first deck?

5. My three cats love licking up the milk out of my cereal bowl if I leave it unattended. If unattended, there is a 30% chance that Beau will clean the bowl, a 50% chance that Tess will, and 20% chance that Kaylee will. Unfortunately the milk makes the cats nauseous and if a cat gets milk there is a good chance the cat will puke. In particular the probability that Beau will puke given he has had milk is 30%, for Tess it is 60%, and for Kaylee it is 40%. My daughter recently left a cereal bowl out and a cat finished the milk.

   a) What is the probability that a cat has puked as a result.
   b) Given that a cat has puked in response, what is the probability it was Kaylee?

6. I have two decks of cards. The first deck has 40 red cards and 10 black. The second deck has 25 red and 25 black. I select a deck at random, and then draw two cards. Given that I've selected two red cards, what is the probability that I initially chose the first deck?

7. An inexpensive and convenient enzyme immunoassay screening tests for HIV in a human. If the person is actually HIV negative then the test returns negative with a probability of 0.985. If the person is HIV positive, the test returns a positive result with probability 0.9997. HIV is a major epidemic in Sub-Saharan Africa with approximately 5% of the adult population having HIV. Major aid organizations

want to help identify people with HIV for treatment and will use this cheap and convenient test in their efforts. Suppose that an adult in Sub-Sahran Africa is selected and tested and the test result is that the person has HIV. What is the probability that the person actual has HIV?

Chapter 3

Random Variables and Distributions

3.1 Defining Random Variables and Discrete Distributions

A random variable is a *function* that takes outcomes in the sample space $\mathcal{S}$ and maps them to numeric values in $\mathbb{R}$. Often times we abbreviate random variable as RV.

The idea is that random events such as flipping Heads, a medical test showing the patient has a disease, Chris Froome winning the Tour de France, or rolling a Leaning Jowler in *Pass the Pigs* are all random events but to do math on them, we need to turn them into numbers.

In cases where the sample space $\mathcal{S}$ is already numeric, the random variable can just be the identity, but in other cases, we might have to be more careful. For example, if my experiment is flipping a fair coin $n = 4$ times, I could define the random variable $X =$ number of heads and X could take on any of the values $x \in \{0, 1, 2, 3, 4\}$. I could similarly define

$$Y = \begin{cases} 0 & \text{if number of heads } < 2 \\ 1 & \text{if number of heads } > 2 \end{cases}$$

A RV function doesn't have to be one-to-one and it doesn't have to map to the entire set of real values.

Because events in the sample space $\mathcal{S}$ have an associated probability, then it is natural to define an event, say B to be all the outcomes $s \in \mathcal{S}$ such that $X(s) = x$ and then define $Pr(X = x) = Pr(B)$.

Notation: We will refer to the random variable using the capital letters, (e.g. X, Y, Z, W) and the possible values they take on using lower case letters. With this notation, the RV X could take on values $x \in \{0, 1, 2, 3, 4\}$

1. We consider flipping a fair coin $n = 4$ times.
    a) What is the set of outcomes? As usual, we will define an event for each outcome.
    b) What is the probability of each outcome?
    c) For the RV X defined above, what outcomes define the event B such that $s \in B \implies X(s) = 2$?
    d) What is $Pr(B)$? Therefore what is $Pr(X = 2)$?
    e) For the RV Y defined above, what outcomes define the event A such that $s \in A \implies Y(s) = 1$?
    f) What is $Pr(A)$? Therefore what is $Pr(Y = 1)$?

For each value that X or Y could take on, we could figure out the probability of the associated event. We define the random variables *distribution* as a description of what values the RV can take on and $Pr(X \in C)$, for any interval $C = \{c : a \le c \le b\}$ for any $a < b$. This is actually a very awkward definition and we will examine more convenient ways to specify these same probabilities.

Discrete random variables are RVs that can only take on a finite or countably infinite set of values.

Continuous random variables are RVs that can take on an unaccountably infinite set of values.

We now define the *probability function* of a discrete RV X as

$$f(x) = Pr(X = x)$$

and the closure of the set $\{x : \text{ such that } f(x) > 0\}$ is referred to as the *support of X*. Notice that this function is defined for all $x \in \mathbb{R}$, but for only a countable number of cases is $f(x) > 0$.

2. Suppose that RV X can take on the values $\{x_1, x_2, \ldots, x_K\}$. Prove that

$$\sum_{k=1}^{K} f(x_k) = 1$$

3. **Bernoulli Distribution**. Suppose that the random variable W takes on the values 0 and 1 with the probabilities $Pr(W = 1) = p$ and $Pr(W = 0) = 1-p$. Then we say that W has a Bernoulli distribution with probability of success p, which I might write as $W \sim Bernoulli(p)$. Show that for any interval $C = \{c : a \leq c \leq b\}$ in $\mathbb{R}$, you can find $Pr(W \in C)$.

4. **Uniform Distribution on Integers**. Suppose that we have integers a and b such that $a < b$. Suppose that the RV X is equally likely to be any of the consecutive integers $a, \ldots, b$. What is $f(x)$? *(Make sure your definition applies to any $x \in \mathbb{R}$)*

5. **Binomial Distribution** Suppose that we have an experiment that consists of n independent Bernoulli(p) trials. We are interested in the distribution of $X = \#$ of successful trials. That is $X \sim Binomial(n, p)$.

   a) For any integer $x \in \{0, 1, \ldots, n\}$, what is $Pr(X = x)$?
   b) Define $f(x)$ for all values of $x \in \mathbb{R}$.

6. Give two examples of random variables which have Bernoulli($p = 1/2$) distributions. These two RVs should not the same RVs, but they have the same distribution. That is to say, RVs have distributions, but distributions are not RVs.

7. Suppose that two fair six-sided dice are rolled and the RV of interest is the absolute value of the difference between the dice. Give the probability distribution along with an illuminating graph.

8. Suppose that a box contains 7 red balls and 3 green balls. If five balls are selected at random, without replacement, determine the probability function of X where X is the number of red balls selected.

9. Suppose that a random variable X has a discrete distribution with the following probability function:

$$f(x) = \begin{cases} \frac{c}{2^x} & \text{for } x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

   Find the value for c that forces the requirement that

$$\sum_{x=0}^{\infty} f(x) = 1$$

   Hint: This is a particular power series. Go to any Calculus book (or internet) for an appropriate result. Make sure you introduce the result and show why the result applies to $f(x)$ in your solution.

For the binomial distribution (and many distributions we will consider this semester), it would be nice to not have to calculate various probabilities by hand. Most mathematical software packages include some way to calculate these probabilities.

System	Documentation Link or site link
Matlab	https://www.mathworks.com/help/stats/working-with-probability-distributions.html

System	Documentation Link or site link
Mathematica	http: //reference.wolfram.com/language/howto/WorkWithStatisticalDistributions.html
R	https://dereksonderegger.github.io/570L/3-statistical-tables.html
Web App	https://ismay.shinyapps.io/ProbApp/

10. Each time I ride home or to work, there is a $p = 0.1$ probability that I will get stopped by a train. Let X be the number of times I'm stopped by the train in the next 10 trips. Assume that the probability I'm stopped on trip i is independent of all other trips j.
    a) What is the distribution of X? Remember, to specify the distribution by name, you must specify the name and the value of all parameters.
    b) What is $Pr(X = 6)$?
    c) What is $Pr(X < 6)$?
    d) What is $Pr(X \geq 6)$?

3.2 Continuous Distributions

We wish to define something similar to the probability function for continuous RVs. However, because there are an uncountable number of values that the RV could take on, we have to be careful and define probability on intervals of the form $[a, b]$. We define the *probability density function* (usually denoted pdf) as the function $f(x)$ such that

$$Pr(a \leq X \leq b) = \int_a^b f(x)dx$$

We further define the support of the distribution as the closure of the set $x : f(x) > 0$. This is the second time we've defined the support as the *closure* of the set. In the discrete case, it didn't really matter, but here it does. If we define the pdf on the set $[0, 1]$ versus $(0, 1)$, we want the support to contain the end points of the interval, that is the support is the closure of $(0, 1)$ which is $[0, 1]$.

1. Show that, for any $a \in \mathbb{R}$, $Pr(X = a) = 0$ is consistent with this definition of $f(x)$. *Hint: What happens as $b - a$ gets small? Take the limit.*

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

Further notice that $f(x) \geq 0$ for all x because otherwise that would imply that there are events with negative probabilities.

3. **Continuous Uniform**. Suppose that the random variable X will be a randomly chosen real value from the interval $[a, b]$. Suppose that for any sub interval $d = [d_1, d_2]$ where $a \leq d_1 \leq d_2 \leq b$, that $Pr(X \in d)$ is proportional to the length of the interval d. What is the pdf of the distribution of X?

It is unfortunate that your book doesn't introduce indicator functions at this point in time. We can define the following:

$$I(\text{logical test}) = \begin{cases} 1 & \text{if logical test is true} \\ 0 & \text{if logical test is false} \end{cases}$$

4. Suppose that the pdf of a random variable X is

$$f(x) = \frac{4}{3}\left(1 - x^3\right)\ I(0 < x < 1)$$

    a) Create a graph of the pdf.
    b) Find $Pr\left(X \leq \frac{1}{2}\right)$?
    c) Find $Pr\left(\frac{1}{4} < X < \frac{3}{4}\right)$
    d) Find $Pr\left(X > \frac{3}{4}\right)$

5. Suppose the random variable X has pdf

$$f(x) = c \cdot e^{-5x} I(x \geq 0) = \begin{cases} c \cdot e^{-5x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

   a) What is the value of c that makes this a valid pdf? That is, what value of c makes this function integrate to 1?
   b) Graph this function. What is $f(0)$? Many students dislike that $f(0) > 1$ is greater than one. Why isn't this a problem?
   c) What is the probability $Pr(X \leq 0.5)$?

6. Suppose that the pdf of the random variable X is

$$f(x) = c \cdot x \;\; I(0 \leq x \leq 3)$$

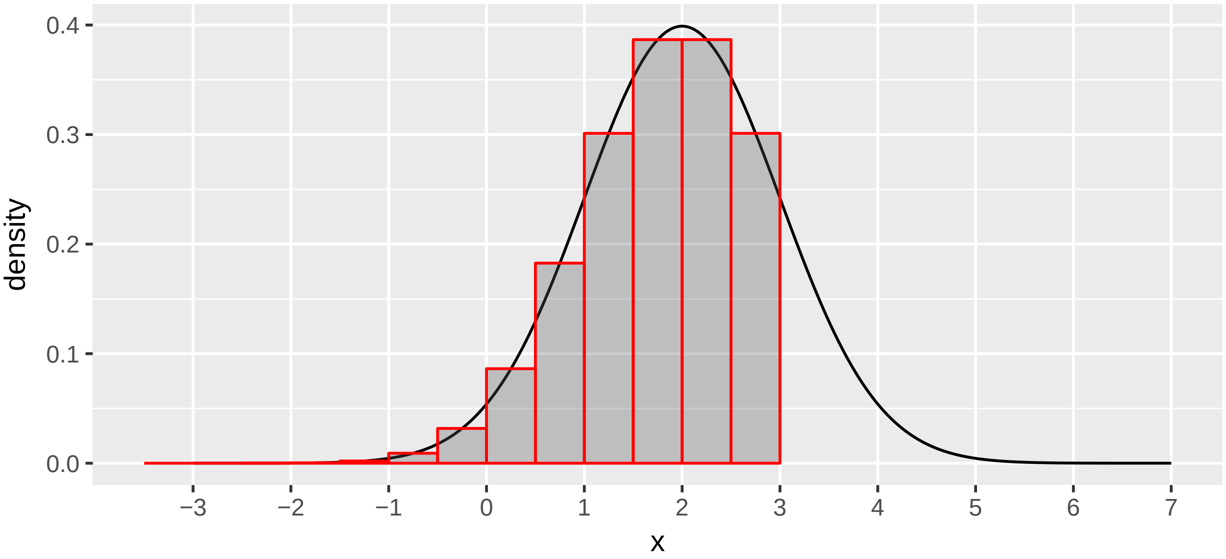
   a) What value of c makes this a valid pdf?
   b) Find a value of t such that $Pr(X \leq t) = 0.25)$
   c) Find a value of t such that $Pr(X > t) = 0.5)$

7. Given the same random variable X described in problem 3.2.5, consider the random variable Y which is the simply the nearest integer to X. What is the pf of Y?

We now have defined both the probability function (pf) for discrete random variables and probability density function (pdf) for continuous random variables. We now try to make a physics analogy to describe the difference between the two. In both cases we want to think about probability as physical mass.

Discrete Case In this case, all the probability mass is concentrated at precise points. So we have little nuggets of probability mass, at discrete points. If I want to know, say $Pr(X \leq 3)$ then we have to sum the probability across all the discrete locations such that $X \leq 3$.

Continuous Case In this case, the probability mass is spread out across $\mathbb{R}$ and the concentration of mass is not uniform, some spots have more concentrated mass. In this case, we don't have a definite amount of mass at a particular point, but we do have a description of how dense the mass is at any point. In this case if we want to know $Pr(X \leq 3)$ then we have to break up $\mathbb{R}$ into a bunch of intervals and then combine the length of the interval along with information about the average density of each interval.



Each bar has some probability mass (mass is the length of the interval times the average density along the interval) and then we just sum up the bars. If we take the limit as we make the bars narrower, then we end up with

$$Pr(X \leq 3) = \int_{-\infty}^{3} f(x)\, dx$$

You might be a little freaked out because typically you would think about mass being the *area* or *volume* times the density, but we need to start in the 1-dimension case before we address the 2-dimension and 3-dimension cases.

3.3 Cumulative Distribution Function

It is somewhat annoying to mathematically describe distributions in two different ways, depending on if the random variable is discrete or continuous.

$$\begin{aligned} Pr\left(X=x\right) &= f(x) \quad \text{if X is discrete RV} \\ Pr\left(a \le X \le b\right) &= \int_a^b f(x)\,dx \quad \text{if X is continuous RV} \end{aligned}$$

While the probability function and probability density function are both useful functions, it is mathematically convenient to have a mathematical description of the distribution that has the same interpretation regardless of if the distribution is discrete or continuous.

Definition: For a random variable X (notice we don't specify if it is continuous or discrete) the **Cumulative Distribution Function** (CDF) is defined as

$$F(x) = Pr(X \le x)$$

Notice that this is defined for all x in $\mathbb{R}$.

1. Suppose that random variable X has a Uniform distribution on the integers $1, 2, 3, 4, 5$. These are the only values that X can take on, and

   $$f(x) = \frac{1}{5} \cdot I\Big(x \in \{1, 2, \dots, 5\}\Big)$$

   Draw a graph of $F(x)$ and make sure the graph demonstrates:
   a) That $F(x)$ is defined for all $x \in \mathbb{R}$.
   b) That $F(x)$ is a step function.
   c) That $F(x)$ is continuous from the right. *That is, for $\epsilon > 0$, $\lim_{\epsilon \to 0} F(x+\epsilon) = F(x)$.*
   d) That $\lim_{x \to -\infty} F(x) = 0$.
   e) That $\lim_{x \to \infty} F(x) = 1$.

Define $B_x = \{s \text{ such that } X(s) \le x\}$. Then we have for $x_1 < x_2$ that $B_{x_1} \subset B_{x_2}$ and therefore:

$$\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} Pr\left(B_x\right) = Pr(\emptyset) = 0$$

$$\lim_{x \to \infty} F(x) = \lim_{x \to \infty} Pr\left(B_x\right) = Pr(\mathcal{S}) = 1$$

2. Show that $F(x)$ must be non-decreasing. *Notice this allows for $F(x)$ to be a flat function, but cannot decrease.*

3. Suppose that the r.v. X has a Binomial($n = 5$, $p = 0.8$) distribution. Sketch the CDF.

4. **Geometric Distribution** Suppose that we flip a biased coin that has probability of heads as $p \in [0, 1]$. Let the r.v. X be the number of coin flips until the first head is observed.

   a) What values could X take? Mathematically, we say, what is the support of X?
   b) Is X a continuous or discrete random variable?
   c) Find the probability function, $f(x)$.
   d) Show that cumulative distribution function is $F(x) = 1 - (1-p)^x$. *Hint: Geometric Series!*

For continuous random variables it is relatively easy to go back and forth from the cdf to the pdf (assuming the integration and differentiation isn't too hard).

$$F(x) = \int_{\infty}^{x} f(u)\,du$$

$$f(x) = \frac{d}{dx}\,F(x)$$

5. **Exponential Distribution** (Warning! There are two ways to parameterize the Exponential Distribution. Before you look anything up, make sure it is using the same parameterization you are.) Suppose that we have a continuous random variable X with pdf

$$f(x) = \beta e^{-\beta x} \ \cdot I(x > 0)$$

   a) Find the cdf function $F(x)$.
   b) For $\beta = 2$ sketch the pdf and cdf.
   c) On the pdf and cdf graphs, represent $Pr(X < 1)$. *In the pdf it will be some shaded area, in the cdf it is something else.*

6. Suppose that the cdf of a random variable X is as follows:

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \frac{1}{9}x^2 & \text{for } 0 \le x \le 3 \\ 1 & \text{for } x > 3 \end{cases}$$

   a) Find the pdf function $f(x)$.
   b) Sketch the pdf and cdf.
   c) On the pdf and cdf graphs, represent $Pr(X \le 2)$.

7. Suppose that a point in the xy-plane is chosen at random from the interior of the unit circle, which is the circle centered at $(0, 0)$ with radius 1. Notice the probability that the chosen point belongs to a given region is proportional to the area of the region. Let the random variable Z represent the distance of the point to the origin.
   a) Find and sketch the cdf of Z.
   b) Find and sketch the pdf of Z.
   c) On the pdf and cdf graphs, represent $Pr(Z \le 0.5)$.

We think of the cdf as a function that takes some value x and produces a probability. In the case where $F(x)$ is monotonically increasing, we could define $F^{-1}(p)$ which takes a probability and tells us what value of x produces $F(x) = p$.

The *quantile function* of a distribution generalizes the inverse function to work similarly for non-decreasing functions by defining

$$F^{-1}(p) = \min(x) \text{ such that } F(x) \ge p$$

8. Suppose that a point in the xy-plane is chosen at random from the interior of the unit circle, which is the circle centered at $\{0, 0\}$ with radius 1. Notice the probability that the chosen point belongs to a given region is proportional to the area of the region. Let the random variable Z represent the distance of the point to the origin.
   a) What is the median of the distribution? That is, find the value z such that $Pr(Z \le z) = 0.5$.
   b) Again sketch the pdf and cdf of this distribution and represent the median on both graphs along with pertinent information showing that indeed the value you found is the median.
   c) What is the 75th percentile? Follow your previous steps in part (a) and (b).

9. **Binomial Distribution** Again we consider the binomial distribution but now with parameters $n = 6$ and probability of success $p = 0.4$.
   a) Find and graph the pdf and cdf of this distribution.
   b) What is the median of this distribution?
   c) What is the 75th percentile? What is the 80th percentile?

3.4 Bivariate Distributions

We now consider the case of having two random variables. We can think of selecting an individual from a population and measuring two (or more!) variables. For example, we might select a NAU student and measure both their height and weight and we want to understand how those two measurements vary. It should be clear that there is a positive correlation (taller people also tend to weigh more) and we want to establish the mathematical framework to address questions such as this.

In general we will consider the distribution of 2 or more random variables and we will call this the *joint distribution*. In the bivariate case, we will consider the joint distribution of X and Y.

3.4.1 Bivariate Discrete

We first consider the case where both variables are discrete. In this case, the bivariate distribution can be defined by simply defining the probabilities $f(x,y) = Pr(X = x, Y = y)$. Your book likes to emphasize the notation of an (X,Y) pair and writes these as $Pr\Big((X,Y) = (x,y)\Big)$, but I dislike so many parentheses.

1. Consider the experiment of rolling two six-sided fair dice. Define the discrete random variable X as the number of ones we roll and Y as the number of sixes.
   a) Fill in the table of $f(x,y)$ values.

$f(x,y)$	$Y=0$	$Y=1$	$Y=2$
$X=0$			
$X=1$			
$X=2$			

   b) Find $Pr(X \geq 1 \text{ and } Y \geq 1)$
2. Suppose that X and Y have a discrete joint distribution for which the joint p.f. is defined as follows:

$$f(x,y) = \begin{cases} c|x+y| & \text{for } x \in \{-2,-1,0,1,2\} \\ & \text{and } y \in \{-2,-1,0,1,2\} \\ 0 & \text{otherwise.} \end{cases}$$

   a) Find the value of the constant c.
   b) Find $Pr(X = 0 \text{ and } Y = -2)$.
   c) Find $Pr(X = 1)$
   d) Find $Pr(|XY| \leq 1)$

3.4.2 Bivariate Continuous

Two random variables X and Y have a bivariate continuous distribution if there exists a non-negative function $f(x,y)$ such that for every subset C of the xy-plane

$$Pr\Big[(X,Y) \in C\Big] = \iint_C f(x,y)$$

Unsurprisingly, the requirements for a function $f(x,y)$ to be a joint pdf are:

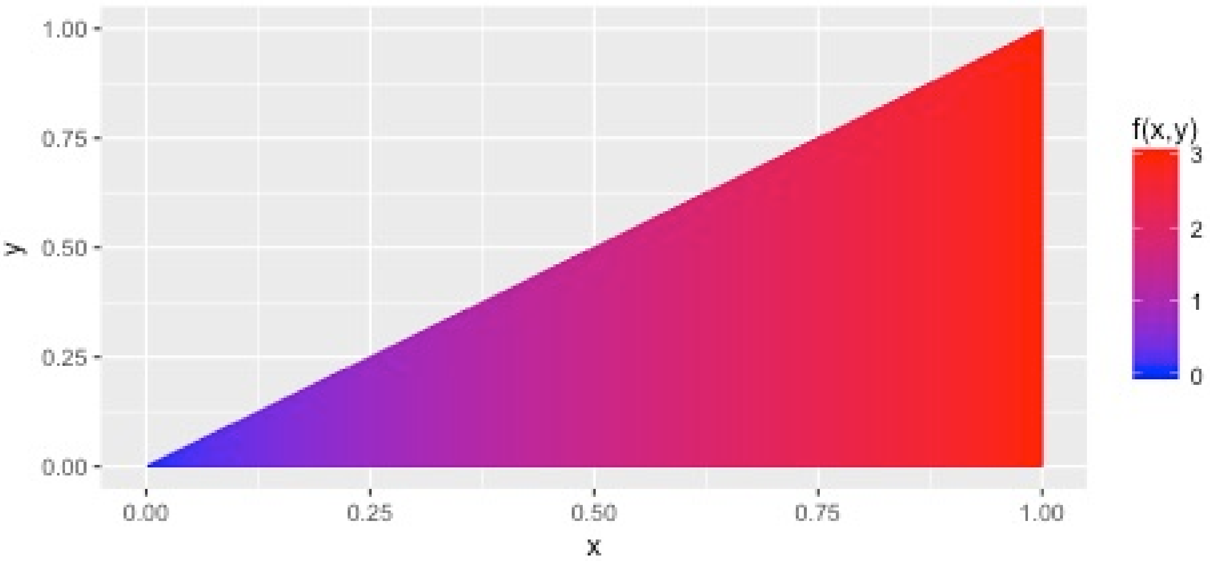
$$f(x,y) \geq 0 \text{ for all } (x,y) \text{ in } \mathbb{R}^2$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

Consider the case where we have the joint pdf

$$f(x,y) = 3x \ \cdot I(0 \le y \le x \le 1)$$

We could visualize this in 3-D but I find it easier to visualize the xy-plane and then let the height of the density function be represented by color.



We might now ask question such as "What is the probability that $X \le 0.5$?" or "What is the probability that $Y > X^2$? In both cases, we just need to integrate the density across the full area of interest.

$$\begin{aligned}
Pr(X \le 0.5) &= \int_0^{0.5} \int_0^x f(x,y)\, dy\, dx \\
&= \int_0^{0.5} \int_0^x 3x\, dy\, dx \\
&= \int_0^{0.5} 3xy|_{y=0}^{x}\, dx \\
&= \int_0^{0.5} 3x^2\, dx \\
&= x^3|_{x=0}^{0.5} \\
&= 0.5^3 \\
&= \frac{1}{8}
\end{aligned}$$

Notice that you could ask Wolfram Alpha for this by using the following:

```
Integrate[ Integrate[ 3*x, {y,0,x}], {x,0,0.5} ] ]
```

Similarly we could ask

$$\begin{aligned}
Pr(Y > X^2) &= \int_0^1 \int_{x^2}^x f(x,y)\,dy\,dx \\
&= \int_0^1 \int_{x^2}^x 3x\,dy\,dx \\
&= \int_0^1 3xy|_{y=x^2}^{x}\,dx \\
&= \int_0^1 3x^2 - 3x^3\,dx \\
&= x^3 - \frac{3}{4}x^4|_{x=0}^{1} \\
&= \frac{1}{4}
\end{aligned}$$

and verify this via Wolfram…

```
Integrate[ Integrate[ 3*x, {y,x^2,x}], {x,0,1} ] ]
```

The joint CDF of the bivariate distribution is defined as we would expect it:

$$F(x,y) = Pr(X \le x, \text{ and } Y \le y)$$

Just as in the univariate case, $F(x,y)$ is non-decreasing x for every fixed value of y and as well as non-decreasing in y for every fixed value of x.

Claims: Suppose that X and Y have joint cdf $F(x,y)$. Define the cdf of X as $F_1(x)$ and the cdf of Y as $F_2(y)$. Then

$$\begin{aligned}
F_1(x) &= \lim_{y\to\infty} F(x,y) \\
F_2(y) &= \lim_{x\to\infty} F(x,y) \\
F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dv\,du \\
f(x,y) &= \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial^2 F(x,y)}{\partial y \partial x}
\end{aligned}$$

assuming the integrals and derivatives exist.

3. Suppose the random variables X and Y have joint pdf

$$f(x,y) = cx^2y \cdot I(x^2 \le y \le 1)$$

   a) Draw the support of this distribution by drawing the parabola $y = x^2$ on the xy-plane and shade in the area for which $f(x,y) > 0$. Denote this region as D
   b) Integrate $f(x,y)$ over its support. That is, find $\iint_D f(x,y)\,dx\,dy$.
   c) What is the value of c such that $\iint f(x,y)\,dx\,dy = 1$?
   d) Find $Pr(X > 0 \text{ and } Y > X)$ by shading in the area of the xy-plane corresponding to this event and then integrating $f(x,y)$ over this region.

4. Suppose that we have random variables X and Y that measure the lifespan of two components in a electronic system. Their joint pdf is

$$f(x,y) = \frac{1}{8}xe^{-(x+y)/2} \cdot I(x > 0,\, y > 0)$$

   a) Graph the support of this function along with contour lines or shading that indicates where the density is high and where is is near zero.

b) Find $Pr(X > 1 \text{ and } Y > 1)$

5. Suppose that X and Y have a continuous joint distribution for which the joint pdf is defined as follows:

$$f(x, y) = cy^2 \ \cdot I(0 \le x \le 2, \ \ 0 \le y \le 1)$$

a) Graph the support of this pdf along with contour lines or shading that indicates the relative value of the density.
b) Determine the value of the constant c.
c) Find $Pr(X + Y \ge 2)$
d) Find $Pr(X \le 1)$

6. Suppose that X and Y have a continuous joint distribution for which the joint pdf is defined as follows:

$$f(x, y) = c(x^2 + y \ \cdot I(0 \le y \le 1 - x^2)$$

a) Graph the support of this pdf along with contour lines or shading that indicates the relative value of the density.
b) Determine the value of the constant c.
c) Find $Pr(X \le 1/2)$
d) Find $Pr(Y \le X + 1)$

3.5 Marginal Distributions

We might be given a joint distribution via its cdf $F(x, y)$ or pf/pdf $f(x, y)$. Often we want to be able to ignore one or the other random variables and just consider X or Y. So we want to understand how to take the joint distribution and obtain the *marginal* distributions of X and Y. We call them marginal because we are taking the xy-plane (or table) and pushing all the probability density (or actual probability) to the X or Y margins.

3.5.1 Discrete Case

Example Suppose we have the discrete pf

$$f(x, y) = \frac{3 - x - y}{8} \ \cdot I(x \in \{0, 1\}, \ y \in \{0, 1\}$$

. Then the table of values for $f(x, y)$ is

$f(x, y)$	$x = 0$	$x = 1$
$y = 0$	$\frac{3}{8}$	$\frac{2}{8}$
$y = 1$	$\frac{2}{8}$	$\frac{1}{8}$

We can easily calculate that $Pr(X = 0)$ by simply summing the column of $x = 0$ probabilities. In general we could write

$$f_1(x) = \sum_{\text{all } y} f(x, y)$$

We can do the same for Y and calculate

$$f_2(y) = \sum_{\text{all } x} f(x, y)$$

In the table, the just means summing across the rows and columns:

$f(x,y)$	$x=0$	$x=1$	**Total**
$y=0$	$\frac{3}{8}$	$\frac{2}{8}$	$\mathbf{\frac{5}{8}}$
$y=1$	$\frac{2}{8}$	$\frac{1}{8}$	$\mathbf{\frac{3}{8}}$
Total	$\mathbf{\frac{5}{8}}$	$\mathbf{\frac{3}{8}}$	

Similarly we could see that

$$\begin{aligned} F_1(x) &= Pr(X \leq x) \\ &= Pr(X \leq x, Y = \text{ anything }) \\ &= \sum_{u \leq x} \sum_{\text{all } y} f(u,y) \end{aligned}$$

3.5.2 Continuous Case

In the continuous case, we have a similar relationship, but we want to be careful and define things first using the cdf.

Again we define the marginal cdf of X as

$$\begin{aligned} F_1(x) &= Pr(X \leq x) \\ &= Pr(X \leq x, Y = \text{anything}) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u,y)\, dy\, du \end{aligned}$$

But recall how we initially defined the pdf of the distribution. It was whatever function you had to integrate such that you happily got the cdf. So therefore by that definition

$$F_1(x) = \int_{-\infty}^{x} \underbrace{\int_{-\infty}^{\infty} f(u,y)\; dy}_{f_1(u)} \;\; du$$

And we can now see that

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

and similarly

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

1. Recall the joint pdf
$$f(x,y) = \frac{21}{4} x^2 y \;\cdot I(x^2 \leq y \leq 1)$$
   a) Graph the joint pdf $f(x,y)$ by graphing the support of the distribution on the xy-plane and
   b) Find and graph the marginal pdf of X, $f_1(x)$.
   c) Find and graph the marginal pdf of Y, $f_2(y)$.
2. Suppose we have the continuous joint distribution
$$f(x,y) = \frac{3}{2} y^2 \;\cdot I(0 \leq x \leq 2, 0 \leq y \leq 1)$$
   a) Graph the joint pdf $f(x,y)$ by graphing the support of the distribution on the xy-plane and
   b) Find and graph the marginal pdf of X, $f_1(x)$.
   c) Find and graph the marginal pdf of Y, $f_2(y)$.

3. **Indentical marginal distrubutions don't imply identical joint distributions.** Create two discrete joint distributions that are different but yet have the same marginal distributions. Feel free to define the joint distributions by two tables of values instead of working out some functional form for $f(x, y)$.

3.5.3 Independence

Recall that we had defined that events A and B are independent if $Pr(A \cap B) = Pr(A)Pr(B)$. Equivalently we can define that random variables X and Y are independent if $Pr(X \leq x, Y \leq y) = Pr(X \leq x)Pr(Y \leq y)$ for all x, y. This is equivalent to

$$F(x, y) = F_1(x)F_2(y)$$

In the discrete case, it is clear that

$$Pr(X \leq x, Y \leq y) = Pr(X \leq x)Pr(Y \leq y) \text{ for all } x, y$$

is equivalent to

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y) \text{ for all } x, y$$

and therefore we could use an equivalent criteria for independence that

$$f(x, y) = f_1(x)f_2(y)$$

In the continuous case, we see that if $F(x, y) = F_1(x)F_2(y)$ then

$$\begin{aligned} f(x, y) &= \frac{\partial}{\partial x \partial y} F(x, y) \\ &= \frac{\partial}{\partial x \partial y} F_1(x)F_2(y) \\ &= \frac{\partial}{\partial x} F_1(x) \frac{\partial}{\partial y} F_2(y) \\ &= f_1(x)f_2(y) \end{aligned}$$

There is nothing in this sequence of equalities that can't be done in reverse so we see this is another equivalent criteria for independence.

The critical aspect of this definition of independence is that the joint cdf or pf/pdf factors into two functions $h_1(x)$ and $h_2(y)$. Typically we think of these as the marginal distributions, but they might be off by some constant value and that is acceptable. That is to say, X and Y are independent if and only if

$$f(x, y) = h_1(x)h_2(y)$$

where $f_1(x) = c_1h_1(x)$ and $f_2(y) = c_2h_2(y)$ for some c_1 and c_2. We note this just so that we can show two variables are independent without bothering to figure out how the integrating constant gets divided up between the two marginal distributions.

Example Suppose that

$$f(x, y) = \frac{3}{16} xy^2 \cdot I(0 \leq x \leq 2,\ 0 \leq y \leq 2)$$

Then it is clear that we could factor $f(x, y)$ into

$$f(x, y) = \underbrace{\frac{3}{16} x \cdot I(0 \leq x \leq 2)}_{h_1(x)} \ \underbrace{y^2 \cdot I(0 \leq y \leq 2)}_{h_2(y)}$$

but I know that $h_2(y)$ isn't exactly the marginal distribution of Y because it doesn't integrate to 1. I'm too lazy to figure out how I need to break the 3/16 multiplier into two pieces and distribute it to make $h_1(x)$ and $h_2(y)$ into the marginals, but I know that I could do it if I had to.

Notice that the indicator function has to be split up and appropriately grouped into the $h_1(x)$ and $h_2(y)$ terms. If that cannot happen, then the variables cannot be independent.

Finally notice that if X and Y are independent, then for any functions $g_1(\cdot)$ and $g_2(\cdot)$ we have that there is $C_1 = \{x : g_1(x) \le u\}$ and $C_2 = \{y : g_2(y) \le v\}$ and therefore

$$\begin{aligned} Pr\Big(g_1(X) \le u, g_2(Y) \le v\Big) &= Pr(X \in C_1, Y \in C_2) \\ &= Pr(X \in C_1)\ Pr(Y \in C_2) \\ &= Pr(g_1(X) \le v)\ Pr(g_2(Y) \le u) \end{aligned}$$

and therefore separate functions of independent random variables are also independent.

4. Suppose we have the continuous joint distribution

$$f(x, y) = \frac{3}{2}y^2\ \cdot I(0 \le x \le 2, 0 \le y \le 1)$$

   a) Are X and Y independent?
   b) Are the events $\{X < 1\}$ and $\{Y \ge 1\}$ independent?
5. Suppose that the joint pdf of X and Y is

$$f(x, y) = \frac{15}{4}x^2\ \cdot I(0 \le y \le 1 - x^2)$$

   a) Sketch the region of support of the joint distribution.
   b) Determine the marginal pdfs of X and Y.
   c) Are X and Y independent? Justify your answer.
6. Suppose that the amount of sugar (in grams) I add to my tea each day has a pdf

$$f(x) = \frac{3}{8}x^2\ \cdot I(0 \le x \le 2)$$

   Let X and Y be the amount added on two successive days and suppose that X and Y are independent.
   a) Find the joint pdf of X and Y.
   b) Find $Pr(X < Y)$.
   c) Find $Pr(X + Y \le 2)$.
7. Suppose that a point (X, Y) is chosen randomly from the unit disc

$$S = \{(x, y) : x^2 + y^2 \le 1\}$$

   . *(In general, when we say something like "a point is chosen randomly", we mean that $f(x, y) = c$ for some constant c for all values of (x, y) in the support. In other words, "at random" means "follows a uniform distribution on the support.")*
   a) Determine the joint and marginal pdfs of X and Y.
   b) Are X and Y independent?
8. Suppose that Alice and Barbara plan to meet at Macy's Coffee shop. Each person will arrive randomly between 5 pm and 6 pm. Each person will arrive and wait for the other for at most 15 minutes. What is the probability that they will meet up? *Hint: Draw the region of support and sub-region where they meet up. Because the density is constant, you could skip the integration and just do some geometry, but be careful!*

3.6 Conditional Distributions

The last distribution we wish to derive from the joint distribution is that *conditional distribution.* Recall for events A and B we had

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} \quad \text{if } Pr(B) > 0$$

This definition immediately leads to the definition of the condition pf in the discrete case:

$$g_1(x|y) = Pr(X = x|Y = y) = \frac{f(x, y)}{f_2(y)} \quad \text{if } f_2(y) > 0$$

and

$$g_2(y|x) = Pr(Y = y|X = x) = \frac{f(x, y)}{f_1(x)} \quad \text{if } f_1(x) > 0$$

Example Suppose we have the discrete pf

$$f(x, y) = \frac{3 - x - y}{8} \cdot I(x \in \{0, 1\},\ y \in \{0, 1\}$$

The table of values for $f(x, y)$ is

$f(x, y)$	$x = 0$	$x = 1$
$y = 0$	$\frac{3}{8}$	$\frac{2}{8}$
$y = 1$	$\frac{2}{8}$	$\frac{1}{8}$

We might be interested in

$$Pr(Y = 1|X = 0) = g_2(1|0) = \frac{f(0, 1)}{f_1(0)} = \frac{2/8}{5/8} = \frac{2}{5}$$

For the continuous case... we'll just *cheat and define* it as what we want:

$$g_1(x|y) = \frac{f(x, y)}{f_2(y)} \quad \text{if } f_2(y) > 0$$

and

$$g_2(y|x) = \frac{f(x, y)}{f_1(x)} \quad \text{if } f_1(x) > 0$$

In graduate level probability courses, this gets a bit more treatment and using measure theory we could appropriately derive this, but that is too much for now.

In the continuous case, we should at least check that $g_1(x|y)$ is a valid pdf, though.

1. For continuous random variables X and Y and y such that $f_2(y) > 0$, show that $g_1(x|y)$ is a valid pdf by showing that:

   a) $g_1(x|y) \geq 0$ for all x.
   b) $\int_{-\infty}^{\infty} g_1(x|y)\, dx = 1$

2. The population distribution of blood donors in the United States based on race/ethnicity and blood type as reported by the American Red Cross is given here:

$f(x,y)$	O	A	B	AB
White	36%	32.2%	8.8%	3.2%
Black	7%	2.9%	2.5%	0.5%
Asian	1.7%	1.2%	1%	0.3%
Other	1.5%	0.8%	0.3%	0.1%

a) What is the probability that a randomly selected donor will be Asian and have Type O blood?
b) What is the probability that a randomly selected donor is white?
c) What is the probability that a randomly selected donor has Type A blood?
d) What is the probability that a donor will have Type A blood given that the donor is white?

3. Suppose that continuous random variables X and Y have joint pdf

$$f(x,y) = e^{-x} \;\; \cdot I(0 \le y \le x)$$

This might be an appropriate model for the number of minutes it takes my daughter to put on her socks and shoes. Here Y is the amount of time it takes to put on her socks, and $X - Y$ is the takes to put the shoes over the socks, so the whole process takes X minutes.

a) Derive the marginal distribution for the total time to put on her shoes and socks.
b) Derive the conditional distribution for how long it took to put on her socks given the total amount of time taken. *Notice this should be a function of the total amount of time taken!*
c) Derive the conditional distribution of how long the total process took given the amount of time it took to put on her socks.
d) What is the probability the process takes more that 5 minutes given that she took 2 minutes to put on her socks?

4. Suppose we have a joint pdf of

$$f(x,y) = c(x+y^2) \;\; \cdot I(0 \le x \le 1,\ 0 \le y \le 1)$$

a) Determine the value of c such that this is a valid joint pdf.
b) Find the conditional pdf $g_1(x|y)$
c) Find $Pr(X \le 1/2 \mid Y = 1/2)$

5. **Law of Total Probability** Prove that

a) If X and Y are discrete RVs then

$$f_1(x) = \sum_y g_1(x|y) f_2(y)$$

b) If X and Y are continuous RVs then

$$f_1(x) = \int_{-\infty}^{\infty} g_1(x|y) f_2(y)\, dy$$

6. **Bayes Theorem** Prove that

a) If X and Y are discrete RVs then

$$g_1(y|x) = \frac{g_1(x|y) f_2(y)}{f_1(x)} = \frac{g_1(x|y) f_2(y)}{\sum_y g_1(x|y) f_2(y)}$$

b) If X and Y are continuous RVs then

$$g_1(y|x) = \frac{g_1(x|y) f_2(y)}{f_1(x)} = \frac{g_1(x|y) f_2(y)}{\int_{-\infty}^{\infty} g_1(x|y) f_2(y)\, dy}$$

Notice that it was arbitrary which variable we conditioned on and we could similarly define $f_2(x)$ and $g_2(y|x)$. Furthermore, all the formulas in this section work if one variable is continuous and the other is discrete with just the appropriate switch between summation and integration.

7. We are interested in modeling the rate of hits on a web page server. Let X be the number of page requests in a minute and R is the *rate* of page requests. Suppose that we model

$$f_2(r) = e^{-r} \ \cdot I(r > 0)$$

and

$$g_1(x|r) = \frac{(2r)^x}{x!} e^{-2r} \ \cdot I\Big(x \in \{0, 1, 2, \dots\}\Big)$$

   a) Find the marginal pf of X.
   b) Find the conditional pdf of R given X, $g_2(r|x)$.
   c) Consider the cases where $X = 1$ and $X = 2$. For what values of r is $g_2(r|1)$ larger than $g_2(r|2)$?

8. **Beta-Binomial** Suppose we are considering a model for seedling success of a tree. We will randomly select an adult tree to harvest seeds from and then plant $n = 40$ seeds and observe how many of them germinate. We will model the number of germinating seeds as $X \sim Binomial(n = 40, \pi)$ where π is the probability a randomly selected seed will germinate. However because seed fitness depends on the parent tree, we allow $\pi \sim Beta(\alpha = 2, \beta = 2)$. The beta distribution has the pdf:

$$f_2(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1} \ \cdot I(0 < \pi < 1)$$

where $\alpha > 0, \beta > 0$ and $\Gamma(\cdot)$ is the gamma function (see the appendix). Recall that the $Binomial(n, \pi)$ distribution has pf

$$g_1(x|\pi) = \binom{n}{x} \pi^x (1-\pi)^{n-x} \ \cdot I(x \in \{0, 1, \dots, n\})$$

For this problem, we have $n = 40$ and $\alpha = \beta = 2$ and we are interested in random variables X and Π.

   a) Graph the marginal distribution of Π.
   b) Find the marginal distribution of X. *Hint: Distributions must integrate/sum to one, and therefore distributions lists are a convenient set of integration/sumation identities. In particular from our definition of the Beta Distribution we have:*

$$\int_0^1 \pi^{a-1}(1-\pi)^{b-1} \, d\pi = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

   c) Find the conditional pdf of π given X, $g_2(\pi|x)$.
   d) Consider the cases where $X = 20$ and $X = 30$. For what values of π is $g_2(\pi|x = 20)$ larger than $g_2(\pi|x = 30)$?

3.7 Functions of Random Variables

Often we have a distribution for the random variable X but actually want to know aobut the random variable $Y = r(X)$.

Example Suppose that we know the distribution of X, the number of customers per hour at a resturant. However, what we really want to know about is $Y = r(X) = 60/X$ which is the typical number of minutes between customers.

Example Suppose that I know the distribution of X which is the number of docter appointments necessary for treatment of some issue. But what I really care about is $Y = 200 + 100 * X$ which is the amount of money total that these appointments will cost.

3.7.1 CDF Method

Of the two methods we'll investigate in this section, the CDF method is by far the most general, but can be somewhat annoying to implement.

The idea is that

$$G(y) = Pr(Y \leq y) = Pr(r(X) \leq y) = \begin{cases} \sum_{x:r(x)\leq y} f(x) \\ \\ \int_{x:r(x)\leq y} f(x) \end{cases}$$

and the only annoying part is figuring out what set of x values forces the $r(x) \leq y$ inequality to hold.

Example Suppose that X, the number of customers per hour, follows a distribution with CDF

$$F(x) = 1 - e^{-6x} \quad \cdot I(x > 0)$$

We are actually interested in $Y = r(X) = 60/X$. Then

$$\begin{aligned} G(y) &= Pr(Y \leq y) \\ &= Pr(r(X) \leq y) \\ \\ &= Pr\left(\frac{60}{X} \leq y\right) \\ \\ &= Pr\left(X \geq \frac{60}{y}\right) \\ &= 1 - F\left(\frac{60}{y}\right) \\ &= \left[1 - \left(1 - e^{-6(\frac{60}{y})}\right)\right] \quad \cdot I(0 < y) \\ &= e^{\frac{-360}{y}} \quad \cdot I(0 < y) \end{aligned}$$

Example Suppose that the random variable X is the amount of time waiting on a child to eat her breakfast before leaving the house and has pdf

$$f(x) = \left(\frac{x}{50}\right) \quad \cdot I(0 \leq x \leq 10)$$

but what I really want to know about is the proportion of breatkfast time remaining $Y = r(X) = 1 - \frac{X}{10}$. Furthermore, suppose that I want the pdf of Y and not the CDF. So my plan is:

$$f(x) \overset{Integrate}{\Longrightarrow} F(x) = Pr(X \leq x) \overset{CDF\ Method}{\Longrightarrow} Pr(Y \leq y) = G(y) \overset{Differentiate}{\Longrightarrow} g(y)$$

Step 1: Find $F(x)$. We will do this by integrating $f(x)$.

$$F(x) = \int_{-\infty}^{x} f(u)\, du = \int_{0}^{x} \frac{u}{50}\, du = \frac{u^2}{100}$$

Step 2: CDF method to find $G(y)$.

$$\begin{aligned} G(y) &= Pr(Y \leq y) \\ &= Pr\left(1 - \frac{X}{10} \leq y\right) \\ &= Pr\left(X \geq 10(1-y)\right) \\ &= 1 - F\Big(10(1-y)\Big) \\ &= 1 - \frac{\Big(10(1-y)\Big)^2}{100} \\ &= 1 - (1-y)^2 \cdot I(y > 0) \end{aligned}$$

Step 3: Differentiate $G(y)$ to obtain $g(y)$

$$g(y) = \frac{d}{dy}G(y) = \frac{d}{dy}\left[1 - (1-y)^2\right] = 2(1-y) \ \cdot I(0 \leq y \leq 1)$$

3.7.2 pdf Method

In some cases it is possible to go straight from $f(x)$ to $g(y)$. In particular if $r(x)$ is a one-to-one and differentiable. If $r(x)$ is one-to-one, then the inverse function $r^{-1}(y)$ exists.

Theorem *Suppose X is a continuous random variable with pdf $f(x)$ such that $Pr(a \leq X \leq b) = 1$ for some (possibly infinite) boundaries a,b. If the transformation $r(x)$ is a one-to-one and differentiable and has range (α, β) for $x \in (a, b)$, then*

$$g(y) = f\Big(r^{-1}(y)\Big) \cdot \left|\frac{d}{dy}r^{-1}(y)\right|$$

where $r^{-1}(y)$ is the inverse function of $r(x)$.

Proof Suppose that $r(x)$ is an increasing function. Then its inverse function is also an increasing function and

$$G(y) = Pr[Y \leq y] = Pr[r(X) \leq y] = Pr[X \leq r^{-1}(y)] = F[r^{-1}(y)]$$

Next we can take the derivative and see

$$g(y) = \frac{d}{dy}G(y) = \frac{d}{dy}F[r^{-1}(y)] = f[r^{-1}(y)]\frac{d}{dy}r^{-1}(y)$$

Because r^{-1} is an increasing function, its derivative is positive, $\frac{d}{dy}r^{-1}(y) = |\frac{d}{dy}r^{-1}(y)|$.

$$g(y) = f[r^{-1}(y)]\left|\frac{d}{dy}r^{-1}(y)\right|$$

Now suppose that $r(x)$ is a decreasing function. The its inverse is also decreasing and we have

$$G(y) = Pr[Y \leq y] = Pr[r(X) \leq y] = Pr[X \geq r^{-1}(y)] = 1 - F[r^{-1}(y)]$$

Again we take the derivative and see

$$g(y) = \frac{d}{dy}G(y) = -f[r^{-1}(y)]\frac{d}{dy}r^{-1}(y)$$

But because $r^{-1}(y)$ is decreasing, its derivative is negative and the negative terms would cancel. So we could write

$$g(y) = f[r^{-1}(y)]\left|\frac{d}{dy}r^{-1}(y)\right|$$

and cover both the increasing and deceasing $r(x)$ cases.

Example Suppose the rate of growth of some bacteria is a random variable X with pdf

$$f(x) = 3(1-x)^2 \ \cdot I(0 \le x \le 1)$$

but what we are interested in is the amount of bacteria at time t which is a function of the initial amount of bacteria $\nu > 0$ and time $t > 0$. Suppose that ν and t are known values, we are interested in the pdf of

$$Y = \nu e^{Xt}$$

Step 1: Is this function a one-to-one increasing or decreasing function on the support of X? Yes it is so we can use the pdf method.

Step 2: Find the inverse function.

$$\begin{aligned} Y &= \nu e^{Xt} \\ \log\left(\frac{Y}{\nu}\right) &= Xt \\ X &= \frac{1}{t}\log\left(\frac{Y}{\nu}\right) \end{aligned}$$

Step 3: Find the deriviative with respect to the new variable.

$$\frac{d}{dy}\left[\frac{1}{t}\log\left(\frac{Y}{\nu}\right)\right] = \frac{1}{tY}$$

Step 4: Stick these into the formula

$$g(y) = f\left(\frac{1}{t}\log\left(\frac{Y}{\nu}\right)\right)\left|\frac{1}{tY}\right| = 3\left(1-\frac{1}{t}\log\left(\frac{Y}{\nu}\right)\right)\frac{1}{tY}\cdot I(\nu \le y \le \nu e^t)$$

Problems

1. Suppose that random variable X has pdf

$$f(x) = 3x^2 \ \cdot I(0 < x < 1)$$

Further suppose that $Y = 1 - X^2$.

a) Find the pdf of Y using the CDF method.
b) Find the pdf of Y using the pdf method.

2. Suppose that random variable X has pdf

$$f(x) = 1 \cdot I(0 \le x \le 1)$$

Further suppose that $Y = X(1-X)$. Find the pdf of Y.

a) Graph the function $y = x(1-x)$ on the support of x. Is this a one-to-one function?
b) What is the support of the random variable Y?
c) On your graph of $y = x(1-x)$, pick a value for y and find the region(s) of x such that $x(1-x) \le y$. *Hint: quadratic equation!*
d) Find the CDF of Y by integrating over the appropriate region(s).

e) Find the pdf of Y by differentiating.

3. Suppose that random variable X has pdf

$$f(x) = e^{-x} \ \cdot I(0 < x)$$

Further suppose that $Y = \left|\sqrt{X}\right|$. Find the pdf of Y.

4. Suppose that X has a uniform distribution on (a, b). Let $c > 0$. Show that $Y = cX + d$ has a uniform distribution on $(ca + d, cb + d)$.

3.8 Multivariate Transformations

Often we have a bivariate or multivariate distribution and we wish to consider some transformation.

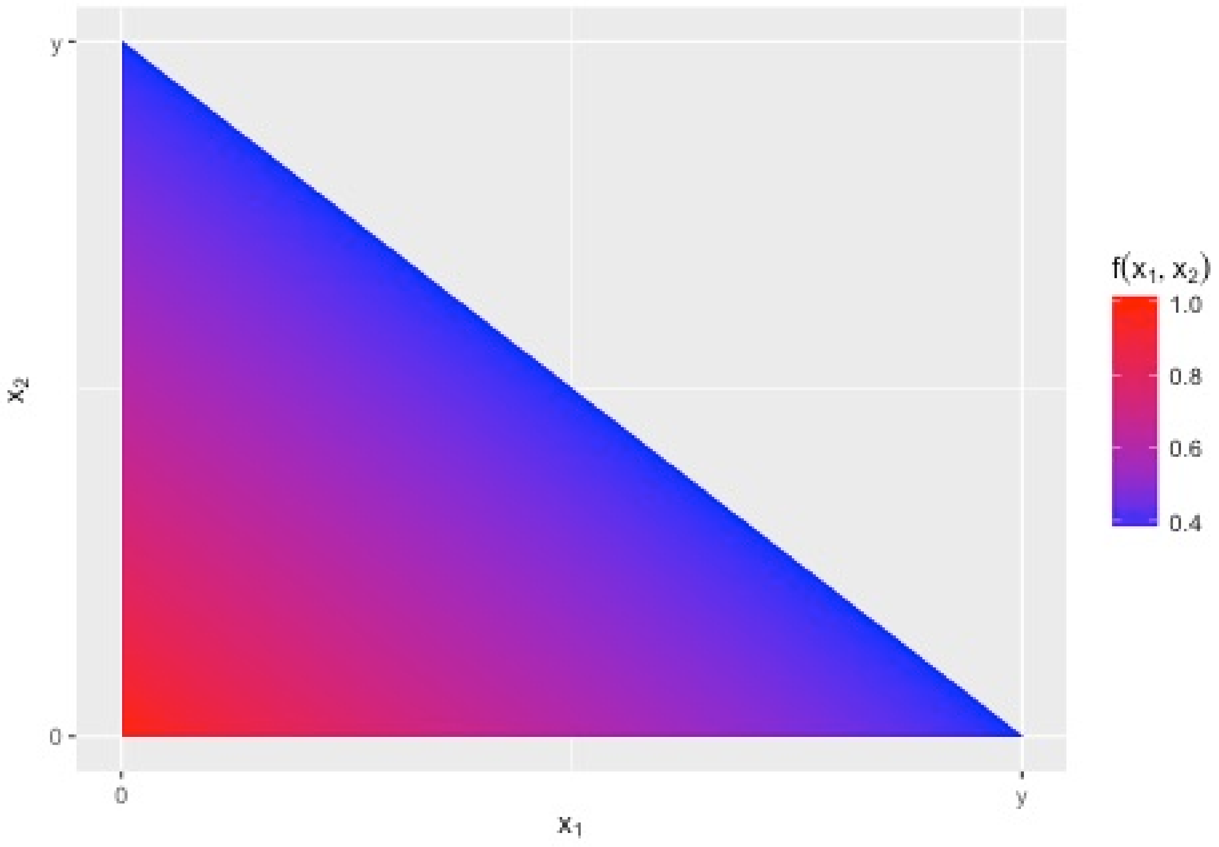
CDF Method Our most widely applicable method for dealing with transformations is to consider how the tranformation can affect the CDF.

Example Suppose that we have X_1 and X_2 that are independent with joint pdf

$$\beta^2 e^{-\beta(x_1+x_2)} \ \cdot I(x_1 > 0, x_2 > 0)$$

where $\beta > 0$. What we really care about is the sum of the two variables $Y = X_1 + X_2$. Then we could figure out the cdf of Y by

$$G(Y) = Pr(Y \leq y) = Pr(X_1 + X_2 \leq y)$$



$$\begin{aligned}
G(Y) &= Pr(X_1 + X_2 \leq y) \\
&= \int_0^y \int_0^{y-x_1} f(x_1, x_2)\, dx_2\, dx_1 \\
&= \int_0^y \int_0^{y-x_1} \beta^2 e^{-\beta(x_1+x_2)}\, dx_2\, dx_1 \\
&= \vdots \\
&= 1 - e^{-\beta y} - \beta y e^{-\beta y}
\end{aligned}$$

We can now use this CDF to obtain the pdf by differentiating

$$\begin{aligned}
g(y) = \frac{d}{dy} G(y) = \frac{d}{dy}\left[1 - e^{-\beta y} - \beta y e^{-\beta y}\right] \\
= 0 + \beta e^{-\beta y} - \beta e^{-\beta y} + \beta^2 y e^{-\beta y} \\
= \beta^2 y e^{-\beta y} \ \cdot I(0 < y)
\end{aligned}$$

Convolutions We are often interested in the sum of two independent random variables. Suppose that X_1 has density function $f_1(x_1)$ and X_2 has density function $f_2(x_2)$ and X_1 is independent of X_2. Define $Y = X_1 + X_2$. We say that the distribution of Y is the *convolution* of the distributions of X_1 and X_2.

1. Suppose that X_1 and X_2 have support along the entire real line. Show that the density function of $Y = X_1 + X_2$ is
$$g(y) = \int_{-\infty}^{\infty} f_1(y-u) f_2(u)\, du = \int_{-\infty}^{\infty} f_1(z) f_2(y-z)\, dz$$
   a) Define the set A_y which is the set of (x_1, x_2) values such that $x_1 + x_2 \leq y$.
   b) Define $G(y)$ by setting up the integration across the set A_y.
   c) Recognize that in the set up of $G(y)$ we have indirectly figured out what $g(y)$ is.

Example Suppose we have independent random variables X_1 and X_2 with uniform distributions on $(0,1)$ and we are interested in $Y = X_1 + X_2$. Because X_1 and X_2 are constrained to $(0,1)$ then $Y \in (0,2)$. Because $f_1(x_1) = I(0 < x_1 < 1)$ and $f_2(x_2) = I(0 < x_2 < 1)$ then

$$\begin{aligned}
g(y) &= \int_{-\infty}^{\infty} f_1(y-u) f_2(u)\, du \\
&= \int_{-\infty}^{\infty} I(0 < y - u < 1) I(0 < u < 1)\, du
\end{aligned}$$

We'll look at this in cases, $y \in (0,1)$ and $y \in (1,2)$. In the first case where $y \in (0,1)$, then $u \in (0,y)$. In the second case where $y \in (1,2)$ then $u \in (y-1, 1)$.

In the first case we have

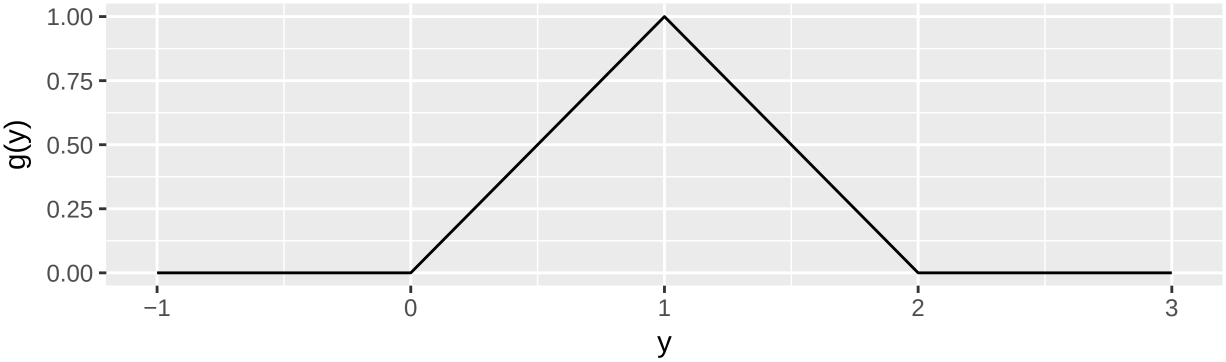
$$g(y) = \int_0^y 1\, du = y \qquad \text{if } y \in (0,1)$$

and in the second case

$$g(y) = \int_{y-1}^1 1\, du = 2 - y \qquad \text{if } y \in (1,2)$$

so all together

$$g(y) = \begin{cases} 0 & \text{if } \ y \leq 0 \\ y & \text{if } \ 0 < y \leq 1 \\ 2 - y & \text{if } \ 1 < y < 2 \\ 0 & \text{if } \ 2 \leq y \end{cases}$$



2. Suppose that we have independent random variables X_1 and X_2 with uniform distributions on a, b where $a < b$.
$$f_1(x_1) = \frac{1}{b-a} \ \cdot I(a < x_1 < b)$$
$$f_2(x_2) = \frac{1}{b-a} \ \cdot I(a < x_2 < b)$$
Find the pdf of $Y = X_1 + X_2$.

3. Suppose that we have independent random variables X_1 and X_2 with pdfs
$$f_1(x_1) = \beta e^{-\beta x_1} \ \cdot I(x_1 > 0)$$
$$f_2(x_2) = \beta e^{-\beta x_2} \ \cdot I(x_2 > 0)$$
where $\beta > 0$. Show that the pdf of $Y = X_1 + X_2$ is
$$g(y) = \beta^2 y e^{-\beta y} \ \cdot I(y > 0)$$
Hint: Don't ignore the indicator functions.

4. **Distribution of the Maximum of a sample.** Suppose that $X_1, X_2, \dots, X_n$ are independent random variables, and all have a common distribution with CDF $F(x)$ and pdf $f(x)$. Let Y_n be the maximum value of these $X_1, X_2, \dots, X_n$ values. *For convenience, let X_i be the ith of these.*

   a) If for some y we have that $Y_n \le y$, what inequality must hold for each X_i?
   b) Write $Pr(Y_n \le y)$ as some probability statement involving all $X_1, X_2, \dots, X_n$ independent variables.
   c) Write down the formula for $G_n(y)$.
   d) Write down the formula for $g_n(y)$.

5. **Distribution of the Minimum of a sample.** Suppose that $X_1, X_2, \dots, X_n$ are independent continuous random variables, and all have a common distribution with CDF $F(x)$ and pdf $f(x)$. Let Y_1 be the minimum value of these X_i values.

   a) If for some y we have that $Y_1 > y$, what inequality must hold for each X_i?
   b) Write $1 - Pr(Y_1 > y)$ as some probability statement involving all $X_1, X_2, \dots, X_n$ independent variables.
   c) Write down the formula for $G_1(y)$.
   d) Write down the formula for $g_1(y)$.

6. Suppose that we have independent random variables X_1 and X_2 with uniform distributions on a, b where $a < b$.
$$f_1(x_1) = \frac{1}{b-a} \ \cdot I(a \le x_1 \le b)$$
$$f_2(x_2) = \frac{1}{b-a} \ \cdot I(a \le x_2 \le b)$$
Find the pdf of the maximum of X_1 and X_2.

7. Suppose that we have independent random variables X_1 and X_2 with pdfs

$$f_1(x_1) = \beta e^{-\beta x_1} \cdot I(x_1 > 0)$$

$$f_1(x_2) = \beta e^{-\beta x_2} \cdot I(x_2 > 0)$$

where $\beta > 0$. Find the pdf of the minimum of X_1 and X_2.

pdf method Just as in the univariate case, we might ask if there is a way to just go straight from the joint pdf to the pdf of the variable of interest. The answer is yes provided, we convert the distribution of $X_1, X_2, \dots, X_n$ into $Y_1, Y_2, \dots, Y_n$ via transformations $r_1(), r_2(), \dots, r_n()$ where

$$\begin{aligned} y_1 &= r_1(x_1, x_2, \dots, x_n) \\ y_2 &= r_2(x_1, x_2, \dots, x_n) \\ &\vdots \\ y_n &= r_n(x_1, x_2, \dots, x_n) \end{aligned}$$

and these functions are invertable so that there exists

$$\begin{aligned} x_1 &= s_1(y_1, y_2, \dots, y_n) \\ x_2 &= s_2(y_1, y_2, \dots, y_n) \\ &\vdots \\ x_n &= s_n(y_1, y_2, \dots, y_n) \end{aligned}$$

For notational compactness, denote $s_1 = s_1(y_1, y_2, \dots, y_n)$. Then

$$g(y_1, y_2, \dots, y_n) = f(s_1, s_2, \dots, s_n) \cdot \left|J\right|$$

where $|J|$ is the absolute value of the determinant

$$J = \det \begin{bmatrix} \frac{\partial s_1}{\partial y_1} & \cdots & \frac{\partial s_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_n}{\partial y_1} & \cdots & \frac{\partial s_n}{\partial y_n} \end{bmatrix}$$

Example
Suppose that we have independent random variables X_1 and X_2 with pdfs

$$f_1(x_1) = \beta e^{-\beta x_1} \cdot I(x_1 > 0)$$

$$f_2(x_2) = \beta e^{-\beta x_2} \cdot I(x_2 > 0)$$

where $\beta > 0$. We will find the joint pdf of $Y_1 = \frac{X_1}{X_1 + X_2}$ and $Y_2 = X_1 + X_2$.

In this case X_1 and X_2 are positive values, so $Y_1 \in (0, 1)$ and $Y_2 > 0$. Doing the backsolving for X_1 and X_2 we have

$$x_1 = y_1 y_2$$

$$x_2 = y_2(1 - y_1)$$

and therefore

$$J = \det \begin{bmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{bmatrix} = y_2(1 - y_1) + y_1 y_2 = y_2$$

Therefore

$$\begin{aligned}
g(y_1, y_2) &= f\Big(y_1 y_2,\ y_2(1-y_1)\Big) \cdot \big|J\big| \\
&= f_1(y_1 y_2)\, f_2(y_2(1-y_1))|y_2| \\
&= \beta e^{-\beta y_1 y_2} \beta e^{-\beta y_2(1-y_2)} \\
&= \beta^2 e^{-\beta y_2}\ I(y_2 > 0)\ I(0 < y_1 < 1)
\end{aligned}$$

Notice that this pdf is easily broken into the marginals of Y_1 and Y_2 and we see that we have confirmed our calculation for the pdf of Y_2 and conveniently discovered that Y_1 has a $Uniform(0, 1)$ distribution.

8. Suppose we have random variables X_1 and X_2 with joint distribution

$$f(x_1, x_2) = 2(1 - x_1)\ I(0 \le x_1 \le 1) I(0 \le x_2 \le 1)$$

Find the pdf of $Y_2 = X_1 X_2$. *Because the pdf method of transformation requires having a Y_1 variable, we will consider a second random variable that is convenient, $Y_1 = X_1$.*

9. Suppose that X_1 and X_2 have a continuous joint distribution with pdf

$$f(x_1, x_2) = 8 x_1 x_2\ \cdot I(0 < x_1 < x_2 < 1)$$

Define the random variables $Y_1 = X_1/X_2$ and $Y_2 = X_2$

a) Find the joint pdf of Y_1 and Y_2.
b) Show that Y_1 and Y_2 are independent.

Chapter 4

Expectations

4.1 Expectation of a RV

The sample mean is a useful measure of centrality of a set of data and we would like a similar quantity for a distribution.

Suppose we have a sample from a distribution that can take on integer values in the range of 1 to 5. For example suppose we have the data $\{1,1,2,3,3,3,4,5\}$. Then the sample mean is

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{1}{8}(1+1+2+3+3+3+4+5) \\ &= \sum_{i=1}^{5} \hat{p}_i \; i = \left(\frac{2}{8}\cdot 1\right) + \left(\frac{1}{8}\cdot 2\right) + \left(\frac{3}{8}\cdot 3\right) + \left(\frac{1}{8}\cdot 4\right) + \left(\frac{1}{8}\cdot 5\right)\end{aligned}$$

where $\hat{p}_i$ values are the observed proportions for each possible value. If we have a really large sample then $\hat{p}_i \approx Pr(X = i)$ and it is natural to define the Expected Value as

$$E(X) = \begin{cases} \sum x\, f(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x\, f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

We need to be careful to avoid the $\infty - \infty$ case and note that if

$$\sum_{\text{Negative } x} x\, f(x) = -\infty \qquad \text{and} \qquad \sum_{\text{Positive } x} x\, f(x) = \infty$$

or

$$\int_{-\infty}^{0} x\, f(x)\, dx = -\infty \qquad \text{and} \qquad \int_{0}^{\infty} x\, f(x)\, dx = \infty$$

then the resulting expectation could be written as $-\infty + \infty$ and that quantity *does not exist.*

Example Suppose that the lifetime, X, of an appliance has a pdf

$$f(x) = 2e^{-2x} \cdot I(x > 0)$$

Then the expectation of X is

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_{0}^{\infty} x\, 2e^{-2x}\, dx = 2\int_{0}^{\infty} x\, e^{-2x}\, dx$$

To finish solving this integral, we need to do integration by parts letting

$$\begin{aligned} u &= 2x \qquad & dv &= e^{-2x}\,dx \\ du &= 2\,dx \qquad & v &= -\frac{1}{2}e^{-2x} \end{aligned}$$

and therefore

$$\begin{aligned} E(X) &= -xe^{-2x}\Big|_0^\infty + \int_0^\infty e^{-2x}\,dx \\ &= 0 + -\frac{1}{2}e^{-2x}\Big|_0^\infty \\ &= \frac{1}{2} \end{aligned}$$

Expectations of functions of a RV Suppose we have a random variable X and some function $Y = r(X)$, then I might want to know the expectation of the random variable Y. We could just derive the pdf of Y and calculate its expectation, but that is just a bunch of integration and differentiation that cancels out because (*in the continuous case and assuming* $r(x)$ *is invertable*)

$$\begin{aligned} E(Y) &= \int y\,g(y)\,dy \\ &= \int r(x)\,f\big(r(x)\big)\left|\frac{dx}{dy}\right|\,dy \\ &= \int r(x)\,f(x)\,dx \end{aligned}$$

We could prove that if the expectation of $Y = r(X)$ exists, then for any function $r(x)$ is equal to

$$E(Y) = E[r(X)] = \begin{cases} \sum r(x)\,f(x) \\ \int r(x)\,f(x)\,dx \end{cases}$$

Example Suppose that we have a random variable with pdf

$$f(x) = 3x^2\,I(0 < x < 1)$$

then

$$E(X) = \int_0^1 x\,f(x)\,dx = \int_0^1 x\,3x^2\,dx = \int_0^1 3x^3\,dx = \frac{3}{4}x^4\Big|_0^1 = \frac{3}{4}$$

and if we consider $Y = X^2$ then

$$E(Y) = E(X^2) = \int_0^1 x^2\,f(x)\,dx = \int_0^1 3x^4 = \frac{3}{5}x^5\Big|_0^1 = \frac{3}{5}$$

Notice that $E(X^2) \ne \Bigg(E(X)\Bigg)^2$. For general $r(X)$, we typically see that $E(r(X)) \ne r\Bigg(E(X)\Bigg)$. However for linear functions, it is true.

1. Suppose that X has a Uniform(a, b) distribution. Find the expected value of X.
2. Suppose that X has a Uniform(a, b) distribution. Find the expected value of $Y = \sqrt{X}$.
3. Suppose that X has a Uniform(a, b) distribution. Find the expected value of $Y = 1/X$.

4. A 1-meter stick is broken at a random spot along the stick. Find the expected value of the length of the longer piece.

Expectations of functions of several variables Suppose that a multivariate distribution with joint pdf $f(x_1, x_2, \dots, x_n)$ and we define $Y = r(X_1, X_2, \dots, X_n)$ then

$$E(Y) = \iint \cdots \int r(x_1, x_2, \dots, x_n)\, f(x_1, x_2, \dots, x_n)\, dx_1 dx_2 \dots dx_n$$

Example Suppose that we have a bivariate distribution

$$f(x_1, x_2) = 8x_1x_2\ \cdot I(0 < x_1 < x_2 < 1)$$

and we wish to know the expectation of $Y = X_1 + X_2$. Then

$$\begin{aligned} E(X_1 + X_2) &= \int_0^1 \int_0^{x_2} (x_1 + x_2)\, 8x_1x_x\ dx_1dx_2 \\ &= \vdots \\ &= \frac{4}{3} \end{aligned}$$

4.2 Properties of Expectations

1. Prove that for finite constants a and b and continuous random variable X, we have

$$E(aX + b) = aE(X) + b$$

2. Show that if continuous random variable X has support on the interval (a, b) where $a < b$, then $a < E(X) < b$. This is true in the discrete case as well.

3. Show that if X_1 and X_2 are (possibly not independent!) continuous random variables with joint pdf $f(x_1, x_2)$ that $E(X_1 + X_2) = E(X_1) + E(X_2)$. By induction this result will hold for the sum of a finite number number random variables. Notice the proof for the discrete case is similar with simply replacing integrals with summations.

4. Suppose that three random variables X_1, X_2, X_3 are sampled from a distribution that has mean $E(X_i) = 5$. Find the expectation of $2X_1 - 3X_2 - X_3 - 5$. *When we say $X_1, X_2, \dots$ are sampled from a distribution, we actually mean that $X_1, X_2, \dots$ are independent and each have marginal distribution as given. So when you heard "We sampled from population ..." in your Introduction to Statistics course, they actually were telling you that the observations are independent.*

5. Suppose that three random variables X_1, X_2, X_3 are sampled from a uniform distribution on $(0, 1)$. What is the expectation of $(X_1 - 2X_2 + X_3)^2$.

6. *Bernoulli Expectation* Suppose that the random variable X_i can take on values 0 or 1 and the probability it takes on 1 is $f(1) = p$. What is the expected value of X_i?

7. *Binomial Expectation* Suppose that we have n identically distributed Bernoulli random variables, each of which having probability of success $f_i(1) = p$. Letting $Y = \sum_1^n X_i$, what is the expected value of Y?

8. *Hypergeometric Expectation* Suppose that we have a bag with N balls, of which M are red and $N - M$ are blue. We will draw n balls out (without replacement) and we are interested in the total number of red balls drawn. Let X_i be 1 if the ith draw was a red ball.

   a) First consider the case where we draw $n = 1$ ball. What is the probability that we draw a red ball on the first draw and therefore what is $E(X_1)$?

b) Next consider the case where we draw $n = 2$ balls. What is the probability that we draw a red ball on the second draw and therefore what is $E(X_2)$?

c) The same pattern holds for the rest of $X_3, X_4, \dots X_n$. Given that, what is the expected value of $Y = \sum_1^n X_i$?

9. Suppose that continuous random variables $X_1, X_2, \dots, X_n$ are independent, each having a marginal pdf $f_i(x_i)$. Show that

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i)$$

Notice that this result requires that the variables are independent, whereas the result in 4.2.3 did not require independence.

10. A gambler will play a game where he is equally likely to win or lose on a given play. When the gambler wins, her fortune is doubled, but when she loses, it is cut in half. Given that the gambler started the game with a fortune of c, what is the expected fortune after n plays?

It can be shown that for non-negative, continuous random variables

$$E(X) = \int_0^\infty (1 - F(x))\, dx$$

and for non-negative discrete random variables

$$E(X) = \sum_{x=1}^\infty Pr(X \ge x)$$

The proof in the discrete case is a reordering of

$$E(X) = \sum_{x=0}^\infty x\, f(x) = \sum_x^\infty x\, Pr(X = x)$$

to summing one copy of $Pr(X = 1)$ and two copies of $Pr(X = 2)$ and three copies of $Pr(X = 3)$ and so on.

$$\begin{array}{ccccccc}
\hline
E(X) & = & 0 & +Pr(X=1) & +Pr(X=2) & +Pr(X=3) & \dots \\
 & & & & +Pr(X=2) & +Pr(X=3) & \dots \\
 & & & & & +Pr(X=3) & \dots \\
 & & & & & & \ddots \\
\hline
\end{array}$$

Notice that the first row of this sum is $Pr(X \ge 1)$ and the second is $Pr(X \ge 2)$ and that establishes the result.

11. *Geometric Expectation* Suppose that each time a person plays a game, they have a probability p of winning. Let the random variable X be the number of games played until the person wins. We have previously shown that

$$f(x) = (1-p)^{x-1} p \;\; I(x \in \{1, 2, \dots\})$$

$$Pr(X \le x) = 1 - (1-p)^x$$

for $x \in \{1, 2, \dots\}$ What is expected number of times a player must play until they win?

12. *Gamma Expectation* Suppose that we have a random variable X with a $Gamma(\alpha, \beta)$ distribution and therefore

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta} \;\cdot I(x > 0)$$

Show that

$$E(X)=\frac{\alpha}{\beta}$$

4.3 Variance

Although the mean of a distribution is quite useful, it is not the only measure of interest. A secondary measure of interest is a measure of *spread* of the distribution. Just as the sample variance is interpreted as the "typical squared distance to the mean" we will define the distribution variance as the "expected squared distance to the mean".

For notational convenience, let $\mu = E(X)$ and define

$$Var(X) = E\big[(X-\mu)^2\big]$$

Because expectations don't necessarily exist, we'll say that $Var(X)$ does not exist if $E(X)$ does not exist or if $E[(X-\mu)^2]$ does not exist. Notice that the Variance is non-negative because of the square.

Finally, we will define the standard deviation of X as the positive square-root of the variance. That is $StdDev(X) = \left|\sqrt{Var(X)}\right|$.

Notationally all of this is a bit cumbersome and we'll use

$$E(X)=\mu_X \qquad Var(X)=\sigma_X^2 \qquad StdDev(X)=\sigma_X$$

If we have only a single random variable in a situation, we will suppress the subscript.

1) Suppose that RV X has $Var(X)$ that exists, then for constants a and b, show that the RV

$$Y = aX + b$$

has variance

$$Var(Y) = a^2 Var(X)$$

Notice that shifting the entire distribution of X by some constant b does not affect the *spread* of the shifted distribution.

Next we consider the sum of independent random variables $X_1 + X_2$.

$$\begin{aligned}
Var(X_1+X_2) &= E\left[\left[(X_1+X_2)-E(X_1+X_2)\right]^2\right]\\
&= E\left[\left[X_1+X_2-\mu_1-\mu_2\right]^2\right]\\
&= E\left[\left[(X_1-\mu_1)+(X_2-\mu_2)\right]^2\right]\\
&= E\left[(X_1-\mu_1)^2+2(X_1-\mu_1)(X_2-\mu_2)+(X_2-\mu_2)^2\right]\\
&= E\big[(X_1-\mu_1)^2\big]+E\big[2(X_1-\mu_1)(X_2-\mu_2)\big]+E\big[(X_2-\mu_2)^2\big]\\
&= Var(X_1) \qquad\quad +2E\big[(X_1-\mu_1)(X_2-\mu_2)\big]+Var(X_2)
\end{aligned}$$

Because X_1 and X_2 are independent then

$$E\big[(X_1-\mu_1)(X_2-\mu_2)\big] = E[(X_1-\mu_1)]E[(X_2-\mu_2)] = (\mu_1-\mu_1)(\mu_2-\mu_2) = 0$$

Repeating this argument for n independent random variables, we therefore have

$$Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Notice that this result *requires* independence so that the cross-product terms are zero!

2) Show that $Var(X) = E(X^2) - \mu^2$. This formula is far more convenient to use, generally.

3) Suppose that X has a uniform distribution on the interval $[0, 1]$. Compute the variance of X.

4) Suppose that Y has a uniform distribution on the interval $[a, b]$ where $a < b$. Compute the variance of Y.

5) Suppose that X has expectation μ and variance σ^2. Show that

$$E\Big[X(X-1)\Big] = \mu(\mu - 1) + \sigma^2$$

6) Suppose that X has a $Gamma(\alpha, \beta)$ distribution. Find that the variance of X is $\frac{\alpha}{\beta^2}$.

7) Suppose that X has a $Bernoulli(p)$ distribution, that is X takes on values 1 or 0 with probabilities $Pr(X = 1) = p$ and $Pr(X = 0) = 1 - p$. Find the variance of X.

8) Suppose that Y has a $Binomial(n, p)$ distribution. That is that

$$Y = \sum_{i=1}^{n} X_i$$

where X_i are independent $Bernoulli(p)$ random variables. Show that

$$Var(Y) = np(1-p)$$

4.4 Moments and Moment Generating Functions

Definition 4.1. Just as the $E(X)$ defines the center of a distribution and $E[(X-\mu)^2] = E(X^2) - \mu^2$ defines the variance, the quantities

$$M_k = E\big(X^k\big) \qquad \text{where } k \in \{1, 2, 3, \dots\}$$

are what we call the kth moment of the distribution. These moments define other attributes of the distribution, but sometimes it is useful to define a similar quantity called the kth *central moment*

$$m_k = E\big((X-\mu)^k\big) \qquad \text{where } k \in \{1, 2, 3, \dots\}$$

These two quantities can define several aspects of the distribution. For example, $M_1 = E(X)$ is the distribution mean, while $M_2 = E(X^2)$ is related to the variance. Other attributes are related to higher moments (e.g. the distribution skew is related to m_3).

Somewhat obnoxiously, the book defines M_k to exist if and only if $E\Big(|X|^k\Big) < \infty$. *(This is obnoxious because we used a different criteria to say if $E(X)$ existed in section 4.1 and the definition for the kth moment is a more strict requirement.)*

Theorem 4.1. *For positive integers $j < k$, if M_k exists, then M_j must also exist.*

Proof.

$$\begin{aligned}
E\Big(|X|^j\Big) &= \int_{-\infty}^{\infty} |x|^j\, f(x)\, dx \\
&= \int_{|x|\leq 1} |x|^j\, f(x)\, dx + \int_{|x|>1} |x|^j\, f(x)\, dx \\
&\leq \int_{|x|\leq 1} 1\, f(x)\, dx + \int_{|x|>1} |x|^k\, f(x)\, dx \\
&\leq 1 + M_k \\
&< \infty \text{ by assumption}
\end{aligned}$$

□

Definition 4.2. Let X be a random variable. For each real number t, define

$$\psi(t) = E\left(e^{tX}\right)$$

as the *Moment Generating Function of X, which we denote mgf of X.*

Theorem 4.2. *Let X be a random variable whose mgf $\psi(t)$ is finite for some neighborhood about $t = 0$. Then for positive integer k, the kth momement is the kth derivative of $\psi(t)$ evaluated at $t = 0$. That is*

$$M_k = E\big(X^k\big) = \psi^{(k)}(0)$$

Proof. A full proof is quite technical, but it revolves around showing that it is permissible to interchange the order of integration/summation and differentiation in this case and that therefore:

$$\begin{aligned}
\psi^{(n)}(0) &= \frac{d^n}{dt^n} E(e^{tX})\Bigg|_{t=0} \\
&= E\left[\Big(\frac{d^n}{dt^n} e^{tX}\Big)\Bigg|_{t=0}\right] \\
&= E\Big[X^n e^{tX}|_{t=0}\Big] \\
&= E\big[X^n\big]
\end{aligned}$$

□

Theorem 4.3. *If the mgfs of two random variables are finite and identical for all values of t in a neighborhood of $t = 0$, then the probability distributions of the two variables are the same.*

Remark. This theorem allows us to compare the mgf of some variable of interest to the set of known mgfs and claim that because the mgfs match, then the variable of interest must follow the matching distribution. This is often a very easy way to show that a variable has a particular distribution and is the reason that we have introduced moment generating functions.

1. Suppose that the random variable X has an $Exponential(\beta)$ distribution which is a special case of the Gamma distribution. $Exponential(\beta) = Gamma(1, \beta)$ distribution. The table of distributions in your book shows that the $Var(X) = \frac{1}{\beta^2}$. Derive the mgf of the Exponential distribution and use it to derive both the expectation and variance.

2. **Mgf of a linear transformation** Suppose that we have a random variable X with mgf $\psi_1(t)$. Show that $Y = aX + b$ for real constants a and b has the mgf

$$\psi_2(t) = e^{bt}\psi_1(at)$$

3. Suppose that $X \sim Exponential(\beta)$. Show that $Y = 2X$ has the mgf of an $Exponential(\beta/2)$ distribution and therefore $Y \sim Exponential(\beta/2)$.

4. Derive the moment generating function of the $Gamma(\alpha, \beta)$ distribution.

5. Derive the moment generating function of $Bernoulli(p)$ distribution.

6. Suppose that independent random variables $X_1, X_2, \dots, X_n$ each have mgfs $\psi_1(t), \psi_2(t), \dots, \psi_n(t)$. Show that the mgf of $Y = \sum X_i$ is

$$\psi_Y(t) = \prod_{i=1}^{n} \psi_i(t)$$

7. Suppose that $X_1, X_2, \dots, X_n$ are independent $Bernoulli(p)$ random variables. Derive the mgf of $Y = \sum X_i$. That is, derive the mgf of the Binomial distribution.

8. Suppose that random variable X has support on the non-negative integers $0, 1, 2, \dots$. The probability function of X is

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} \cdot I(x \in \{0, 1, 2, \dots\})$$

   The random variable X has a $Poisson(\lambda)$ distribution. Derive the mgf of this distribution. *Hint: what is the summation constant in the pf?*

9. Suppose that $X_1, X_2, \dots, X_n$ are independent random variables each with distribution $Poisson(\lambda)$. Derive the moment generating function of the distribution of $Y = \sum X_i$ and state its distribution.

10. Suppose that $X_1, X_2, \dots, X_n$ are independent random variables each with distribution $Exponential(\beta) = Gamma(1, \beta)$. Derive the moment generating function of the distribution of $Y = \sum X_i$ and state its distribution.

4.5 Mean vs Median

We will skip this section.

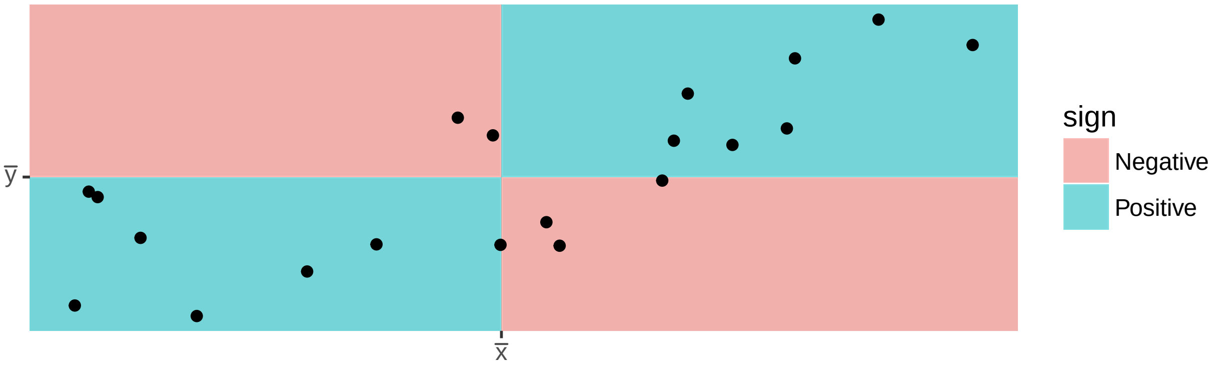
4.6 Covariance and Correlation

In introductory statistics classes, we often define the *correlation coefficient* which describes if there is a positive or negative linear relationship between two variables.

The sample correlation coefficient is defined as

$$r = \frac{\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}{n-1}$$

The heart of this formula is the sign of each of the $(x_i - \bar{x})(y_i - \bar{y})$ terms. If the x-value is big (greater than $\bar{x}$) and the y-value is large (greater than $\bar{y}$), then after multiplication, the result is positive. Likewise if the x-value is small and the y-value is small, both standardized values are negative and therefore after multiplication the result is positive. If a large x-value is paired with a small y-value, then the first value is positive, but the second is negative and so the multiplication result is negative.



We will define a similar quantity for two random variables.

$$Cov(X,Y) = E\left[\big(X - E(X)\big)\big(Y - E(Y)\big)\right]$$

1. Prove that for random variables X and Y such that $Var(X) < \infty$ and $Var(Y) < \infty$,

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Remark. This computational formula for the covariance is the same as the computational formula for $Var(X)$ is we let $Y = X$.

$$Cov(X,X) = E(XX) - E(X)E(X) = E\left(X^2\right) - [E(X)]^2$$

In fact, we can consider $Var(X)$ as a special case of $Cov(X,Y)$.

2. *Cauchy-Schwartz Inequality.* Let X and Y be random variables, each with finite variances. Show that $[E(XY)]^2 \le E(X^2)E(Y^2)$. *Hint: Observe that* $E\left[(tX - Y)^2\right] \ge 0$ *for any real value* t *and therefore*

$$t^2E(X^2) - 2tE(XY) + E(Y) > 0$$

This a polynomial of degree 2 *in* t *but has no roots. Combine this fact with the quadradic equation to achieve the desired result.*

Definition 4.3. We can now define the correlation between random variables X and Y such that they have finite variances as

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

3. Use the Cauchy-Schwartz Inequality to show that

$$-1 \le \rho(X,Y) \le 1$$

4. Prove that if X and Y are independent random variables, each with finite variances, then $Cov(X,Y) = 0$.

5. Show that the preceding statement is not an “if and only if” statement by considering the following case. Let be uniformly X distributed on the integers $-1, 0,$ and 1. Let $Y = X^2$. Show that X and Y are not independent but that $Cov(X,Y) = 0$.

6. Prove that for random variables X and Y and constants a, b, c and d that

$$Cov(aX + b, cY + d) = ac\,Cov(X,Y)$$

7. Prove that for random variables X and Y

$$Var(X + Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$$

8. Prove that for random variables X and Y and constants a, b, and c that

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X, Y)$$

9. Suppose that X and Y have a continuous joint distribution with pdf

$$f(x, y) = \frac{1}{3}(x + y)\; I(0 < x < 1) I(0 < y < 2)$$

Determine the value of $Var(2X - 3Y + 8)$

10. Suppose that $X_1, X_2, \dots, X_n$ are independent and identically distributed random variables, each with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Show that the sample mean $\bar{X} = \frac{1}{n}\sum X_i$ has expectation $E(\bar{X}) = \mu$ and variance $Var(\bar{X}) = \frac{\sigma^2}{n}$.

4.7 Conditional Expectation

In many real-world scenarios, the phenomena of interest is best modeled in a "multilevel" fashion. For example, suppose that we are interested in understanding radon levels within homes and we have radon levels from thousands of houses across the US. Because houses within a county are more similar to each other than houses that are states apart, it makes sense to model the data using a multilevel approach that models each county and then the houses within the county. In this section, we will develop certain mathematical tools that will be useful for these situations.

Example Recall the Beta-Binomial hierarchical relationship where we have some observation X that is the number of successes out of n independent Bernoulli(P) trials, and the probability of success P was also a random variable but with a Beta(α,β) distribution.

$$X|P \sim Binomial(n, P) \qquad \text{where} \qquad P \sim Beta(\alpha, \beta)$$

Recall the probability and probability density functions were

$$\begin{aligned} g(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} \quad \cdot I(x \in 0, 1, \dots, n) \\ f(p) &= \frac{\Gamma(\alpha, \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \quad \cdot I(0 < p < 1) \end{aligned}$$

If we regard $P = p = 0.6$ as a known quantity, then

$$E(X \mid P = 0.6) = np = n(0.60)$$

However, if we don't know the value of P and continue to consider it a random variable, then

$$E(X|P) = nP$$

is just a function of the random variable P and therefore $E(X|P)$ also a random variable. We could ask questions like what is the probability density function of $E(X|P)$ and what is the expectation of $E(X|P)$.
Definition 4.4. If Y has a continuous conditional distribution given $X = x$, define the conditional expectation as

$$E(Y|x) = \int_{-\infty}^{\infty} y\, g_2(y|x)\, dy.$$

If Y has a discrete conditional distribution given $X = x$, replace the integral with the summation over all values of y.

Similarly we can define the conditional expectation of some function $h(y)$ of Y as

$$E(h(Y)|x) = \int_{-\infty}^{\infty} h(y)\, g_2(y|x)\, dy.$$

1. Suppose that X and Y are continuous random variables where Y has a finite expectation. Prove that

   $$E\left[E(Y|X)\right] = E(Y)$$

   Hint: Start with the definition of $E(Y)$ using the double integral and joint distribution and then split the joint distribution into the product of the conditional and marginal. Finally recognize the inner integral as $E(Y|X = x)$.

2. Suppose that $X|P \sim Binomial(n, P)$ and $P \sim Beta(\alpha = 4, \beta = 6)$. Find $E(X)$. *Hint: Though we haven't proven it (yet!), the expectation of a Beta random variable is $\alpha/(\alpha + \beta)$.*

3. Suppose that an unknown number of individuals (denoted as N) will be independently randomly chosen for "additional screening" by the TSA. Suppose that we allow $N \sim Poisson(\lambda)$ individuals are selected and that the probability that a selected person is female is 0.4. Let X denote the number of females (out of N individuals additionally screened) and so $X|N \sim Binomial(N, p = 0.4)$. Find $E(X)$. *Hint: Though we haven't proven it yet, the variance of a Poisson random variable is also λ.*

Definition 4.5. The conditional variance of $Y|X = x$ is defined similarly to the unconditional case

$$Var(Y|x) = E\left[\Big(Y - E(Y|x)\Big)^2\right] = E\left[\Big(Y|x\Big)^2\right] - \Big[E(Y|x)\Big]^2$$

Theorem 4.4. *If X and Y are random variables which have finite expectation and variances, then*

$$Var(Y) = E\Big[Var(Y|X)\Big] + Var\Big[E(Y|X)\Big]$$

Proof. We start by noticing

$$Var(Y|X) = E\Big(Y^2|X\Big) - \Big[E(Y|X)\Big]^2$$

and therefore

$$E\left[Var(Y|X)\right] = E\left[E\Big(Y^2|X\Big) - \Big[E(Y|X)\Big]^2\right]$$

Furthermore by the definition of variance

$$Var\left[E(Y|X)\right] = E\left[\Big(E(Y|X)\Big)^2\right] - \left[E\Big(E(Y|X)\Big)\right]^2$$

Finally

$$\begin{aligned}
Var(Y) &= E\left(Y^2\right) - [E(Y)]^2 \\
&= E\left[E\Big(Y^2|X\Big)\right] - \left[E\Big(E(Y|X)\Big)\right]^2 \\
&= E\left[E\Big(Y^2|X\Big)\right] - E\left[\Big(E(Y|X)\Big)^2\right] + E\left[\Big(E(Y|X)\Big)^2\right] - \left[E\Big(E(Y|X)\Big)\right]^2 \\
&= E\left[E\Big(Y^2|X\Big) - \Big(E(Y|X)\Big)^2\right] + E\left[\Big(E(Y|X)\Big)^2\right] - \left[E\Big(E(Y|X)\Big)\right]^2 \\
&= E\left[Var(Y|X)\right] + Var\left[E(Y|X)\right]
\end{aligned}$$

□

4. Suppose that an unknown number of individuals (denoted as N) will be independently randomly chosen for "additional screening" by the TSA. Suppose that we allow $N \sim Poisson(\lambda)$ individuals are selected and that the probability that a selected person is female is 0.4. Let X denote the number of females (out of N individuals additionally screened) and so $X|N \sim Binomial(N, p = 0.4)$. Find $Var(X)$.

5. The number of defects per yard in a certain fabric, Y, was known to have a Poisson distribution with parameter λ. The parameter λ was assumed to be a random variable with a Exponential(1) distribution. Find $E(Y)$ and $Var(Y)$. Notice that $E(\lambda) = 1$ and $1 = Var(Y|\lambda = 1) < Var(Y)$.

6. Suppose we have random variables X and Y with joint pdf

$$f(x, y) = (x + y) \cdot I(0 \leq x \leq 1)I(0 \leq y \leq 1)$$

Find $E(Y|X)$ and $Var(Y|X)$. *Feel free to set up all the necessary integrals and then use software to calculate them.*

Chapter 5

Common Distributions

5.1 Bernoulli and Binomial

Definition 5.1. Define a Bernoulli random variable as a variable that takes on the value 0 with probability $1-p$ and the value 1 with probability p where $0 \leq p \leq 1$. We will write this as $X \sim Bernoulli(p)$.

The pf of a Bernoulli random variable is written as

$$f(x) = p^x (1-p)^{1-x} \quad \cdot I(x \in \{0,1\})$$

1. Show that if $X \sim Bernoulli(p)$ then
   a) The expectation of X is $E(X) = p$
   b) The variance of X is $Var(X) = p(1-p)$
   c) The moment generating function of X is $\psi(t) = pe^t + (1-p)$

Definition 5.2. Define a Binomial random variable as the sum of n independent and identically distributed Bernoulli(p) random variables. We will write $X \sim Bin(n,p)$

The pf of a Binomial random variable is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \cdot I(x \in \{0,1,2,\dots,n\})$$

2. Show that if $X \sim Bin(n,p)$ then
   a) The expectation of X is $E(X) = np$
   b) The variance of X is $Var(X) = np(1-p)$
   c) The moment generating function of X is $\left(pe^t + (1-p)\right)^n$

5.2 Hypergeometric

We next consider a distribution that is the sum of *dependent* Bernoulli random variables. Consider the scenario where there are A balls that are *amber* and B balls that are *blue*. The balls are thoroughly mixed and we will select n of those *without replacement*. Of interest is Y, the number of amber balls drawn. It is helpful to think of randomly arranging all $A+B$ balls into some order and then selecting the first n balls. Define $X_i = 1$ if the ith ball is amber and $X_i = 0$ if the ith ball is blue. Finally we note that X_i is *not* independent of X_j and that

$$Y = \sum_{i=1}^{n} X_i$$

1. Derive the pf of $Y \sim Hypergeometric(A,B,n)$.

a) How many ways are there to draw, without replacement, n balls from $A + B$ when order doesn't matter?
b) How many ways are there to draw x amber balls and $n - x$ blue balls (assuming $x \leq A$ and $n - x \leq B$)?
c) Give the pf of a Hypergeometric(A,B,n) distribution.

2. Notice that absent any information about the other X_j balls, $X_i \sim Bernoulli\left(\frac{A}{A+B}\right)$. Use this information to derive the expectation of Y.

3. Because X_i is negatively correlated with X_j, we can't easily derive the variance of Y. Instead we will be *obnoxiously* clever.

a) Let $n = A + B$ so that we are selecting all the balls. Argue that $Var(Y) = 0$.
b) Because of the symmetry of the random assignments of balls, then $Cov(X_i, X_j)$ is the same for all $i \neq j$. *There isn't anything thing to do here, but this part of our arguement is critical.*
c) We know that for any set of random variables

$$Var(Y) = \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$$

Use this information, along with parts (a) and (b), to solve for the $Cov(X_i, X_j)$. *The hard part here is figuring out how many covariance terms there are.*
d) Finally, show that

$$Var(Y) = \frac{nAB}{(A+B)^2} \cdot \frac{A+B-n}{A+B-1}$$

4. Suppose that we have $Y \sim Hypergeometric(A, B, n)$ and separately we have $W \sim Bernoulli\left(n, \frac{A}{A+B}\right)$. Show that for $n > 1$ that $Var(W) > Var(Y)$.

5. My daughter recently mixed 20 M&Ms and 30 Skittles in a bowl and left them for me to find.

a) What is the probability that I select **only** M&Ms when I select 6 pieces of candy?
b) What is the expected number of M&Ms in the next 6 pieces (from the 50)?

5.3 Poisson

The Poisson distribution is used to model the number of events that happen in some unit time. The critical idea is that the number of events that occur in any two disjoint time periods are independent, regardless of the length of the period. By splitting the time unit into *many* sub-intervals, each of which that could only have 0 or 1 event and considering those sub-intervals as independent Bernoulli RVs, it is possible to justify the following probability function

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \cdot I(x \in \{0, 1, 2, \dots\})$$

where the parameter λ represents the mean number of events that should happen per time interval of some specific size.

1. Suppose that $X \sim Poisson(\lambda)$. Show that
a) This is a valid pf by showing that $f(x) \geq 0$ for all x and that $\sum_{x=0}^{\infty} f(x) = 1$. *Hint, look at the series expansion of* e^{λ}.
b) The expectation of X is $E(X) = \lambda$.
c) The variance of X is $Var(X) = \lambda$. *Hint, derive* $E[X(X-1)]$ *and use that to figure out* $E[X^2]$.
d) The moment generating function of X is

$$\psi(t) = e^{\lambda(e^t - 1)}$$

2. Show that if $X_1, \dots, X_n$ are independent and identically distributed $Poisson(\lambda)$ random variables, which we denote as

$$X_i \overset{iid}{\sim} Poisson(\lambda)$$

then

$$Y = \sum(X_i) \sim Poisson(n\lambda).$$

5.4 Geometric and Negative Binomial

We will first define the *geometric* distribution. Here we consider another version of multiple Bernoulli random variables. This time, we consider an experiment where we repeatedly sample from a $Bernoulli(p)$ distribution, where p is the probability the draw was a success, and each draw is independent of all previous draws. We are interested in *"the number of failures before the first success."*

1. We first consider $Y \sim Geometric(p)$.

   a) What is the probability that there are no failures? That is, what is the probability that the first success occurs on the first draw.

   $$f(0) = Pr(Y = 0) = \ ?$$

   b) What is the probability that there is one failure followed by a success? What is the probability that there are y failures before the first success?
   c) Use this to derive the pf of a $Geometric(p)$ distribution.

2. Show that the Moment Generating Function for the $Geometric(p)$ distribution is

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

   Hint, the Geometric distribution is named as such because the Geometric Series result is necessary to show this.

3. Utilize the mfg to derive the expected value and variance of a $Geometric(p)$ distribution.

The Negative Binomial distribution extends the idea of "number of failures until the first success" to the number of failures until the rth success.

4. The pf of the $Y \sim Negative\ Binomial(r, p)$ distribution can be derived with the following:
   a) What is the probability of observing exactly y failures in a row, followed by r successes?
   b) How many ways are there to distribute the r successes among the $r + y$ draws, keeping in mind that the final draw must be a success?
   c) Use the above ideas to derive the pf.
5. A second way of thinking about the negative binomial distribution is as the sum of r independent Geometric random variables. Utilize this interpretation to derive the expectation, variance, and moment generating function of a $Negative\ Binomial(r, p)$ distribution.

Some books parameterize the geometric (and negative binomial) distributions as the total number of draws before the first (or nth) success. Whenever you are looking up properties of this distribution, make sure it is defined how you want it. For example, the wikipedia page for the geometric (and negative binomial) have the information for both definitions.

5.5 Uniform

5.6 Exponential and Gamma

5.7 Beta

5.8 Normal

5.9 Bivariate Normal

Appendix A

Useful functions

A.1 Gamma Function

We want to generalize the factorial function which was

$$n! = n(n-1)(n-2)\dots(2)(1)$$

for $n \in \{0, 1, 2, \dots\}$ and we define $0! = 1$ for notational convenience because that is what we need in many formulas. Notice that $1! = 1$, $2! = 2$ and $3! = 3 * 2!$.

The problem is that this is only defined for the whole numbers (natural numbers and zero). We want to define a function that allows us to expand this function to the positive real numbers.

Consider the funtion:

$$\Gamma(x) = \int_0^\infty z^{x-1} e^{-z}\, dz$$

First we show that $\Gamma(x) = (x-1)\Gamma(x-1)$ utilizing integration by parts.

$$\begin{aligned} u &= z^{x-1} & dv &= e^{-z}\, dz \\ du &= (x-1)z^{x-2}\, dx & v &= -e^{-z} \end{aligned}$$

and therefore

$$\begin{aligned} \Gamma(x) &= \int_0^\infty z^{x-1} e^{-z}\, dz \\ &= \int_0^\infty u\; dv \\ &= uv|_0^\infty - \int_0^\infty v\, du \\ &= -z^{x-1}e^{-z}|_{z=0}^\infty + \int_0^\infty (x-1)z^{x-2}e^{-z}\, dz \\ &= (x-1)\int_0^\infty z^{x-2}e^{-z}\, dz \\ &= (x-1)\Gamma(x-1) \end{aligned}$$

Next we notice that $\Gamma(1) = 1 = 0!$ via the following integration

$$\Gamma(1) = \int_0^\infty z^{1-1} e^{-z}\, dz = \int_0^\infty e^{-z}\, dz = -e^{-z}|_0^\infty = 1$$

Using these bits, we can see that

$$\Gamma(2) = 1 \cdot \Gamma(1) = 1 = 1!$$

$$\Gamma(3) = 2\Gamma(2) = 2 = 2!$$

$$\Gamma(4) = 3\Gamma(3) = 3 \cdot 2! = 3!$$

and for positive integers n,

$$\Gamma(n) = (n-1)!$$

which can be re-arranged to see that

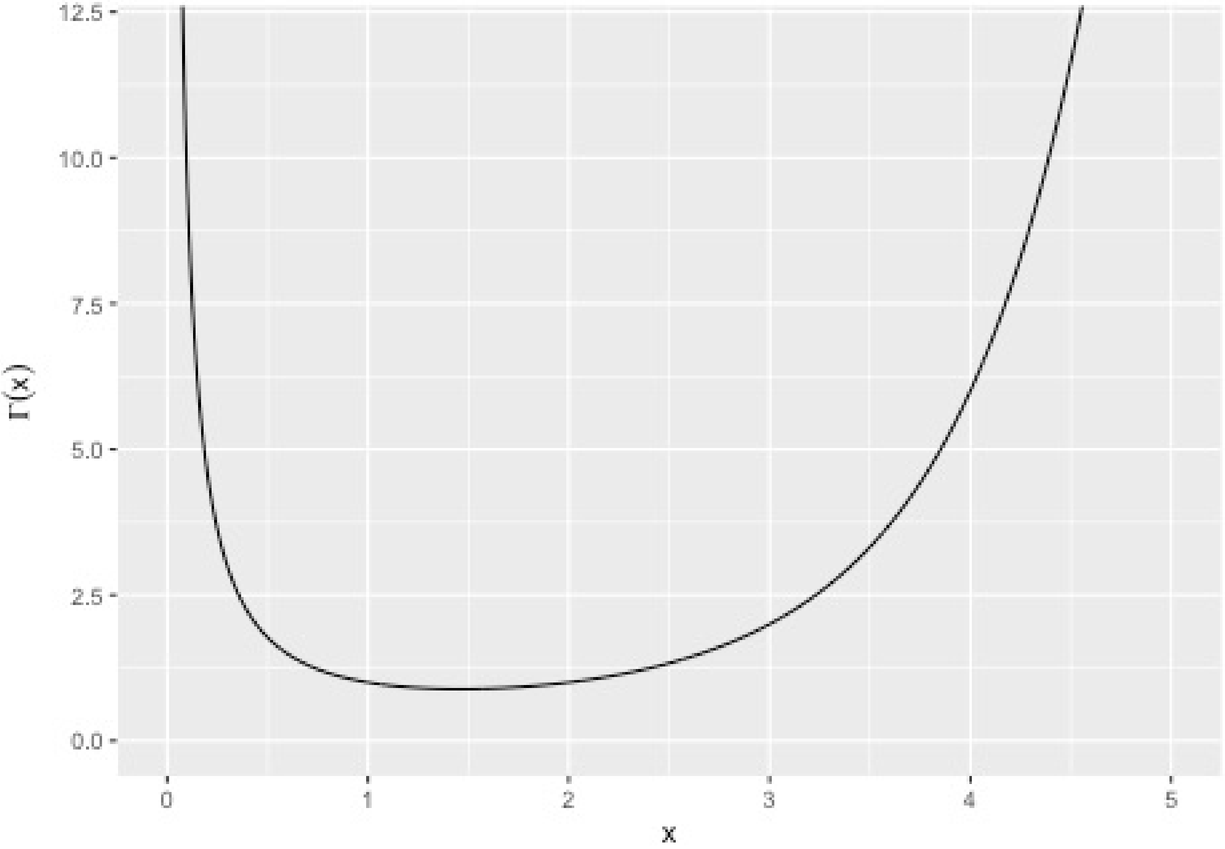
$$n! = n\Gamma(n)$$

We conclude by noting that

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

(This result can be obtained by looking at the square of the integral and then doing a trig substitution.)

A graph of this on the postive real numbers is given below:



Finally we note that there is a reasonable approximation (known as Stirling's approximation) to the function for large values of x as

$$\Gamma(x+1) \approx \sqrt{2\pi x}\left(\frac{x}{e}\right)^x$$

A.2 Useful Series Results

A.2.1 e^x

In many calculus books, the following is shown:

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

A.2.2 Geometric Series

For $0 < \alpha < 1$, it is known that

$$\sum_{x=0}^{\infty} \alpha^x = \frac{1}{1-\alpha}$$

Bibliography