My name is Leon Hostetler. I am currently a student at Florida State University majoring in physics as well as applied and computational mathematics. Feel free to download, print, and use these class notes. If you find them useful, consider buying me a coffee.

All of my class notes can be found at `www.leonhostetler.com/classnotes`

Please bear in mind that these notes will contain errors. If you find one, please email me at leonhostetler@gmail.com with the name of the class notes, the page on which the error is found, and the nature of the error. If you include your name, I will probably list your name on the thank you page if I decide to compile and sell my notes.

Last Updated: August 25, 2016

Chapter 1

Probability and Statistics

1.1 Counting

1.1.1 Multiplication Principle

The **rule of product** or the **multiplication principle** is the basic rule of counting. It states that if there are m ways of doing one thing and n ways of doing another thing, there are mn ways of doing both things. In other words, if one experiment has m possible outcomes and for each of those, there are n possible outcomes of a second experiment, then the total number of outcomes is mn (assuming the second experiment is independent of the first). This principle can be generalized to more than two experiments. For example, if there are n_1 possible outcomes for experiment one, for each of those, there are n_2 possible outcomes of experiment two, and for each of those there are n_3 possible outcomes of experiment three, and so on, then there are a total of $n_1 \cdot n_2 \cdot n_3 \cdots$ possible outcomes.

Example:

How many different 6-place license plates are possible if the first and last place must be letters and the others must be numbers.

There are 26 possible letters and 10 possible numbers, so the total possibilities are

$$26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 26 = 6,760,000.$$

1.1.2 Permutations

A **permutation** is an *ordered arrangement*. For example, when permuting *abc*, the results *abc* and *bac* are distinct. With permutations, there are two possibilities: permutation with repetition and permutation with no repetition. In the first case, the order matters, but we are allowed to repeat elements. To calculate the permutations with repetition, just use the rule of product. If we just say "permutation" we assume permutation without repetition, which simply means that if you choose an item for first place, then that item is removed from consideration for subsequent places.

If there are n objects, then there are

$$n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1 = n!,$$

ways of permuting them without repetition.

Example:

How many different 6-place license plates are possible if the first and last place must be letters and the others must be numbers, and none of the letters or numbers can appear more than once?

There are 26 possible letters and 10 possible numbers, so the total possibilities are

$$26 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 25 = 3,276,000.$$

The number of permutations of n distinct objects taken r at a time is called "n permute r" and is given by

$${}_nP_r = \frac{n!}{(n-r)!}.$$

Think of it this way, how many ways can you permute the numbers 123456789 if you take them 3 at a time? You will end up with a list of three digit numbers. For the first place, there are 9 possibilities, for the second place there are 8 possibilities remaining, and for the third place there are 7 possibilities, so by the multiplication principle, there are $9 \cdot 8 \cdot 7$ possibilities. This is the same as $\frac{9!}{6!} = \frac{9!}{(9-3)!}$.

Example:

Give the permutations of 1234 taken two digits at a time.

We know that ${}_4P_2 = 12$, so we should have a total of 12 items. Listing them, we have $\{12, 13, 14, 21, 23, 24, 31, 32, 34, 41, 42, 43\}$.

Example:

A president and vice president are to be chosen from a committee of 20 people. How many different choices of those officers are possible?

For the first choice, there are 20 possibilities, leaving 19 possibilities for the second choice. You can also think of this as permuting 20 distinct objects taken 2 at a time or 20 permute 2.

$$20 \cdot 19 =_{20} P_2 = 380.$$

Example:

Consider the example above. If Alex is a member of the committee, how many ways can the two officers be chosen if Alex will not serve as vice president?

Split the problem into the two possible paths. If he is chosen as the president (i.e. there is one option for the first slot), then there are 19 possibilities remaining for the vice presidency (i.e. 19 options for the second slot).

$$1 \cdot 19 =_{19} P_1 = 19$$

If he is not chosen as the president, then there were 19 possibilities for the presidency and 18 possibilities for the vice presidency (since Alex is not being counted as a possibility for the vice presidency.

$$19 \cdot 18 =_{19} P_2 = 342.$$

The total number of possibilities is the sum of the two paths or 361.

If some of the objects being permuted are indistinguishable from each other, then out of the permutations of all the objects, we must remove those that look the same. If there are n total objects and n_1 are the same as each other, n_2 are the same as each other and so on, then the total number of different permutations is given by

$$\frac{n!}{n_1!n_2!\cdots n_r!}.$$

Example:

How many distinct arrangements are there of the letters in MISSISSIPPI?

There are a total of 11 letters, so there are 11! different permutations if the S's, P's, and I's are distinguishable from each other, say, they're different colors. Since they're not distinguishable however, we must remove those repeated cases from our total set. There are 4 S's, 2 P's and 4 I's, so the total number of different permutations is

$$\frac{11!}{4!2!4!} = 34,650.$$

1.1.3 Combinations

The **combinations** of n distinct objects taken r at a time is

$${}_nC_r = \frac{n!}{(n-r)!r!} = \binom{n}{r}.$$

With combinations, order doesn't matter. The number of permutations for a given set is larger than the number of combinations for the same set. With permutations, *abc* is counted separately from *bac*, but with combinations, *abc* is the same as *bac* so the pair only counts for one combination.

Note that

$${}_nC_r = \frac{{}_nP_r}{r!}.$$

To keep from confusing permutations and combinations, think of permutations as lists (order matters) and combinations as groups (order does not matter).

Example:

How many different ways can a committee of two be chosen from a committee of 20?

This problem is different from the president and vice president example given above because the officers are no longer being distinguished from each other. The number of possibilities are given by

$${}_{20}C_2 = 190.$$

Tip:

The key to learning how to do counting problems is to practice (or at least try) a lot of them, and then check a solution manual.

To divide n distinct objects into groups $n_1, n_2, \ldots, n_r$ where $n_1 + n_2 + \ldots + n_r = n$, use the formula

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}.$$

Notice that this is the same as permuting n objects when n_1 are the same, n_2 are the same, and so on. Notice that if you're choosing 3 from 10, for example, that is the same as dividing 10 distinct objects into groups of size 3 (the chosen ones) and size 7 (the ones not chosen), so

$$\binom{10}{3} = \binom{10}{3,7}.$$

This formula becomes more useful when considering more than two groups. For example, if you have to divide 20 distinct objects into groups of size 10, 7, and 3, then

$$\binom{20}{10,7,3} = \frac{20!}{10!7!3!} = 22170720,$$

which can also be thought of as choosing 10 from 20, and then 7 from the remaining 10

$$\binom{20}{10}\binom{10}{7} = 22170720,$$

or choosing 10 from 20, and then 3 from the remaining 10

$$\binom{20}{10}\binom{10}{3} = 22170720,$$

or even choosing 3 from 20 and 7 from the remaining 17

$$\binom{20}{3}\binom{17}{7} = 22170720.$$

Tip:

You can brute force counting problems by enumerating all of them, but this is rarely the desired or most efficient manner. The key is to learning tricks, noting patterns, and using reasoning. For example, the easy way of counting the arrangements of 123456 in which 1 comes before 2, is simply to note that 1 comes before 2 in exactly half of all the possible arrangements.

1.1.4 Set Notation and Venn Diagrams

The **sample space**, S, of a probability experiment is the set of all possible outcomes. Each possible outcome is called an **element** or **sample point**. Typically, we use the counting methods detailed in the previous sections to enumerate the sample space, but sometimes, it's easier to use a tree diagram.

There can be infinite sample spaces, which are described by some kind of rule, for example

$$\{x|x^2 + y^2 = 1\},$$

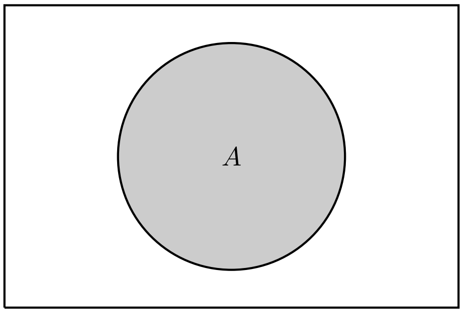
is the sample space containing all the points on the unit circle.

An **event** is a subset of the sample space. We are typically interested in the probability of an event.

The **null set** or the **empty set** contains no elements and is denoted ϕ.

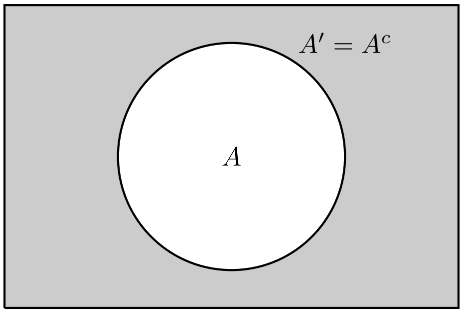
We can depict the sample space as a Venn diagram as shown here.

Since an event is a subset of the sample size, we can depict an event A as in the Venn diagram below.



The **complement** of a set is all the elements of the set that are not in a specified subset. For example, the complement of event A is denoted below by shading in all the points that are not in A. The complement is denoted by a prime symbol or a superscript c as in

$$A' = A^c.$$



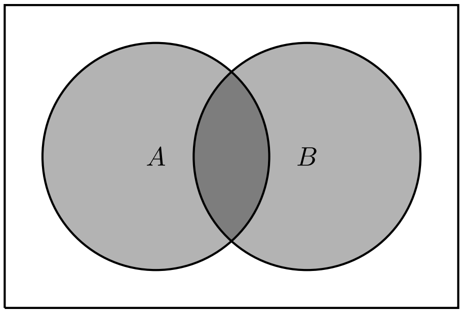
The **union** of two sets is all the elements that are in either of the two sets. The union of sets A and B is denoted

$$A \cup B.$$

It is depicted in the Venn diagram below as all the gray shaded area. The **intersection** of two sets is the elements that are in both sets. The intersection of sets A and B is denoted

$$A \cap B.$$

Note, the intersection of A and B may also be denoted simply AB. It is depicted in the Venn diagram below as only the *darker* gray area.



Set notation involves chaining together these symbols to denote a specific set of elements. For example, the set of events in A, but not in B, could be represented by $A - B$, but in set notation, it would be represented by $A \cap B^c$. In general, to subtract, you use the intersection of the complement.

Two sets A and B are **mutually exclusive** or **disjoint** if $A \cap B = \phi$, that is, if they have no elements in common. To present this with a Venn diagram, you would draw two circles that do not overlap.

Tip:

When doing Venn diagram word problems, it is often easier to work backwards. First you draw the diagram of overlapping circles, then you find the value corresponding to the intersection of all the circles, and you work outward from there.

1.2 Probability

Probability is a set function that assigns to each event A in the sample space S a number $P(A)$ called the "probability of A" such that the following properties are satisfied:

1. $0 \leq P(A) \leq 1$
   This simply states that probabilities are numbers between 0 and 1.
2. $P(S) = 1$
   This just states that the outcome of an experiment will in a point in the sample space.
3. If A_i are disjoint events, then

$$P(A_i \cup A_2 \cdots) = \sum_{i=1} P(A_i).$$

   This just means, for example, if events A_1 and A_2 are mutually exclusive, then the probability of at least one of them occuring is just the sum of their individual probabilities. Think of it in terms of a Venn diagram.

If an experiment has N equally likely outcomes, and if exactly n of them correspond to event E, then

$$P(E) = \frac{n}{N}.$$

This simply means that for an event E

$$P(E) = \frac{\text{favorable outcomes}}{\text{total possible outcomes}}.$$

Example:

With a five card poker hand, what is the probability of having 2 aces and 3 jacks?

Since the order of the cards doesn't matter, the total sample size is $S =_{52} C_5 = \binom{52}{5}$. There are four aces and four jacks. The number of ways to choose two aces from four is $\binom{4}{2}$ and the number of ways to choose three jacks from four is $\binom{4}{3}$, so the number of ways to choose two aces from four *and* three jacks from four is $E = \binom{4}{2}\binom{4}{3}$. The probability is

$$P(E) = \frac{E}{S} = \frac{\binom{4}{2}\binom{4}{3}}{\binom{52}{5}} = 0.00092\%.$$

One proposition is that

$$P(E^c) = 1 - P(E).$$

This simply states that the probability of something not happening is 1 minus the probability of it happening. This is a trivial consequence of the fact that the probability of something happening plus the probability of it not happening must be 1.

Another proposition is that if E is a subset of F, then

$$P(E) \leq P(F).$$

Just think of the Venn diagram.

If A and B are two events, then

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

In other words, the probability of A or B happening is the probability of A happening plus the probability of B happening minus the probability that A and B both happen. This is an important proposition and it makes sense when visualized as a Venn diagram. The intersection of A and B must be subtracted because otherwise, that contribution would be counted twice (since the circles overlap).

This also means that if the events are disjoint, then the proposition simplifies to

$$P(A \cup B) = P(A) + P(B).$$

By solving the earlier proposition for $P(A \cap B)$, we find that

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

In other words, the probability of A and B happening is the probability of A happening plus the probability of B happening minus the probability that A or B.

We can use a similar formula to count outcomes. If $N(A)$ means the number of outcomes in event A, then a Venn diagram shows us that

$$N(A \cup B) = N(A) + N(B) - N(A \cap B).$$

We can use these propositions to prove that $P(\phi) = 0$. Since an event A and ϕ are disjoint, we have that $A \cap \phi = \phi$ and $A \cup \phi = A$, so

$$\begin{aligned} P(A \cup \phi) &= P(A) + P(\phi) \\ &= P(A), \end{aligned}$$

and so $P(\phi) = 0$.

Example:

What is the probability of getting a sum of 6 or 10 when a pair of dice are thrown?

The problem is asking for either case, so we need to think of a union. If we let A be the event in which we get a sum of 6 and B be the event in which we get a sum of 10, then $P(A \cup B) = P(A) + P(B)$. The two events are disjoint since you can't get both 6 and 10 on the same toss. The total possible outcomes of a throw is $6 \times 6 = 36$. We can enumerate the outcomes favorable to A by considering that if the first dice gives 1, the second dice must be 5, and so on. The favorable outcomes are (1,5), (2,4), (3,3), (4,2), and (5,1), so $P(A) = \frac{5}{36}$. Similarly, the outcomes favorable to B are (4,6), (5,5), and (6,4), so $P(B) = \frac{3}{36}$. The solution is then

$$P(A \cup B) = \frac{5}{36} + \frac{3}{36} = \frac{2}{9}.$$

For three events A, B, and C, the union of all three is

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

Example:

Suppose you flip a coin five times. Let A be the event in which you get an odd number of heads and B be the event in which you get an even number of heads.

1. What is $A \cup B$.
   This is simply the sample size S since the number of heads must be either even or odd.
2. What is $P(A \cup B)$?
   Since every outcome is in $A \cup B$, $P(A \cup B) = P(S) = 1$.
3. What is $A \cap B$?
   This is the null set ϕ since it is impossible to have both an odd number of heads and an even number of heads.
4. What is $P(A \cap B)$?
   This is just $P(A \cap B) = P(\phi) = 0$.
5. What is $P(A)$?
   We know that A and B are disjoint, so $1 = P(A) + P(B)$. The probability of there being exactly 1 head is the same as the probability of their being exactly four tails. The probability of exactly 2 heads is the same as that of exactly 3 tails. In other words, every point in A maps to a point in B and vice versa, so

$$P(A) = P(B) = \frac{1}{2}.$$

Tip:

One check to see if your solution makes sense is to look at the limiting behavior. For example, if you are choosing a number of objects from a larger group of objects, see what your solution says when you select all the objects. Does it still make sense? Does the probability go to 1 when it should? Does it go to 0 when it should?

1.2.1 Table Method

One common type of probability problem is to calculate the probability of selecting a subset with a specific makeup from a set with a specific makeup. For example, given a group of 14 children—3 five-year-olds, 4 six-year-olds, 4 seven-year-olds, and 3 eight-year-olds, if you chose four of them at random, what is the probability of ending up with exactly 2 six-year-olds and 2 seven-year-olds?

The solution is calculated as 4 choose 2 (the six-year-olds) times 4 choose 2 (the seven-year-olds) divided by 14 choose 4 (the entire groups).

$$\frac{\binom{3}{0}\binom{4}{2}\binom{4}{2}\binom{3}{0}}{\binom{14}{4}} = \frac{\binom{4}{2}\binom{4}{2}}{\binom{14}{4}}$$

This calculation can be depicted nicely with the table:

Age:	5	6	7	8	Total
Old Group	3	4	4	3	14
New Group	0	2	2	0	4

Notice that the probability of choosing the 'New Group' from the 'Old Group' is the product of the four binomials $\binom{3}{0}\binom{4}{2}\binom{4}{2}\binom{3}{0}$ where the top numbers come from the old group and the bottom numbers come from the new group, all divided by $\binom{14}{4}$, which comes from the group totals.

You draw 8 numbers from 1 to 40 when playing a lottery. What is the probability that the 8 numbers you choose will match the 8 drawn by the lottery commission? We can think of this problem in terms of the group problem where the old group (the numbers chosen by the lottery commission) is composed of the 8 winning numbers and the 32 non-winning numbers.

			Total
Old Group	8	32	40
New Group	8	0	8

The answer is

$$\frac{\binom{8}{8}\binom{32}{0}}{\binom{40}{8}} = \frac{1}{\binom{40}{8}}.$$

What is the probability that exactly six of your numbers match? This can be illustrated with the table:

			Total
Old Group	8	32	40
New Group	6	2	8

and the answer is

$$\frac{\binom{8}{6}\binom{32}{2}}{\binom{40}{8}}.$$

What is the probability that at least six of your numbers match? Here we have to be careful, since the table method only works when exactly six of the numbers match. However, by making three different tables, we can calculate the probability that exactly six match, exactly 7 match, and exactly 8 match and then add up the three probabilities to get the total.

If you draw five cards from a standard deck, what is the probability that you end up with at least one card from each of the four suits?

We can think of this as choosing 1 from each of the four suits, but what about the fifth card? It could be any of the four suits. If we think of the favorable outcomes, it will be either $\{HHDSC\}, \{HDDSC\}, \{HDSSC\}$ or $\{HDSCC\}$ and each of these favorable outcomes will have equal probability. Therefore, we can calculate the probability of one occuring and just multiply by four. Here's the table for calculating the case where you get one from each suit and two hearts:

Suit:	H	D	S	C	Total
Old Group	13	13	13	13	52
New Group	2	1	1	1	5

The solution, then, is

$$4 \times \frac{\binom{13}{2}\binom{13}{1}\binom{13}{1}\binom{13}{1}}{\binom{52}{5}}.$$

We can represent the table method symbolically as the probability of choosing the set with the specific makeup $\{a, b, c, \ldots, z\}$ from the set with the specific makeup $\{A, B, C, \ldots, Z\}$ is

$$\frac{\binom{A}{a}\binom{B}{b}\binom{C}{c}\cdots\binom{Z}{z}}{\binom{A+B+C+\cdots+Z}{a+b+c+\cdots+z}}.$$

Here are some general tips for counting and calculating probabilities:

- Always double-check that your solution is the complete one by explicitly noting the meaning of your solution. Does it match what the question is asking? It's easy to miss something that is integral but subtle.
- Make sure you understand the physical meaning of every statement by saying it in words.
- Always think of how you can represent a problem graphically or geometrically. Think of things in terms of Venn diagrams and probability trees.
- Be careful when using trees that you really are dealing with a partition, which requires that the events are disjoint *and* fill the sample space.
- Solve the problems on your own (if you can), but check the solution manual for alternate, and perhaps better, ways of solving them.
- Being able to do well at calculating probabilities requires that you practice a lot of different problems.
- Try a simple version of the problem to gain insight into its mechanics.
- Try the opposite question. What is the probability that it does not ...
- Try to rephrase the question in terms of equivalent probabilities. For example, given that a person went to the store, the probability that she purchased exactly one item is equivalent to the probability that she purchased at least one item minus the probability that she purchased at least two items.
- When asked to calculate the probability that a set of things occur together in an arrangement, think of

the set as being glued together–they can be treated as a single unit.

- Be careful you don't mistake a purely counting problem for a probability problem. If the problem asks for the number of possible arrangements, don't make the mistake of giving the probability of getting that arrangement.
- If it's easy to enumerate all the possible outcomes (e.g. two dice), do so.

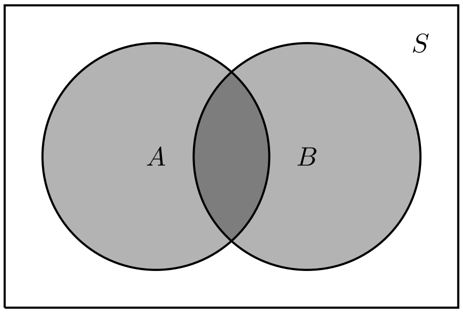
1.3 Conditional Probability

Conditional probability helps us if we have partial information about an experiment. Given event A and its probability of occuring and event B and its probability of occuring, what is the probability that B occurs given that A has occured. One example of a conditional probability is the probability that the second card chosen from a shuffled deck is an ace given that the first one was a ten.

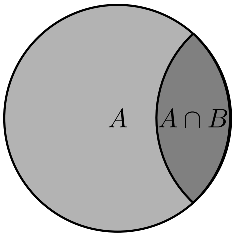
A conditional probability is denoted $P(B|A)$ and read as "the probability of B given A", and it is calculated as

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

We can think of this in terms of a Venn Diagram. Imagine that we have two events A and B as shown in the diagram below.



If event A has occured, then we are concerned only with the restricted sample size of A. In terms of the Venn diagram, we are now concerned only with the circle A. Then the probability of B occuring given that A has occured is essentially the area $A \cap B$ divided by A. This demonstrates that we can think of conditional probability as a shrinking of the sample space.



For the probability of A given B and C, we can write

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

Rearranging the formula for conditional probability by multiplying both sides by $P(A)$ gives us

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

This form is often useful when we want to find $P(B \cap A)$. An extension of this is known as the **multiplication rule**.

Example:

If A is the event in which a black king is the first card selected from a deck of shuffled cards, and event B is the event that the second card selected is a jack or a queen, what is the probability that the first card is a black king *and* the second card is a jack or queen.

The probability that the first card is a black king is $P(A) = \frac{2}{52}$ since there are two black kings. The probability that the second card is a jack or queen is $\frac{8}{52}$ since there are four jacks and four queens. The probability that the second card is a jack or a queen given that the first card was a black king is $P(B|A) = \frac{8}{51}$ since there are only 51 cards remaining to choose from. This allows us to calculate $P(A \cap B)$ as

$$P(A \cap B) = P(B|A)P(A) = \frac{8}{51} \cdot \frac{2}{52} = 0.0060.$$

	Nonsmoker	Moderate Smoker	Heavy Smoker	Total
Lung Cancer	8	22	15	45
No Lung Cancer	30	18	7	55
Total	38	40	22	100

One way of presenting statistical data is with a **contingency table** like the one shown above with visitors to a cancer clinic. It shows that out of a 100 visitors, 38 are nonsmokers, 40 are moderate smokers, and 22 are heavy smokers. It also shows that 45 of the visitors have lung cancer and 55 do not. We can use a contingency table to calculate conditional probabilities such as, given that a visitor is a heavy smoker, what is the probability of them having lung cancer. We could also ask, given a visitor has lung cancer, what is the probability that they are heavy smokers.

What is the probability that a visitor to the clinic is a heavy smoker given that he has lung cancer? We could let A be the event in which a visitor is a heavy smoker and B be the event in which the visitor has lung cancer and then calculate $P(A|B)$ using the formula. It's faster with a contingency table to recall that a conditional probability is a shrinking of the sample size. Since we are conditioning on lung cancer, we can ignore the row containing the numbers for no lung cancer. Using only the row for lung cancer, we can calculate the probability as the number of heavy smokers (15) divided by the total number of people in that row (45). So the probability is $15/45 = 1/3$.

What is the probability that a visitor is a nonsmoker *and* has cancer? If A is the event that a visitor is a nonsmoker and B is the event that a visitor has cancer, then we are asking for $P(A \cap B)$. Looking at the table, we see that there are 8 people that are both nonsmokers and have cancer. This is out of 100 people, though, so the probability is $8/100 = 2/25$.

One proposition we can make is that

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

For example, if A is the event that you're in severe pain and B is the event in which you stub your toe, then given that you have stubbed your toe, you will either be in severe pain or not. If there is a 60% probability that you will be in severe pain given that you've stubbed your toe, then there is necessarily a 40% probability that you will not be in severe pain even though you stubbed your toe. A similar relationship does not exist between $P(A|B)$ and $P(A|B^c)$. For example, a person with a chronic illness may have a 60% probability of being in severe pain whether or not they stub their toe.

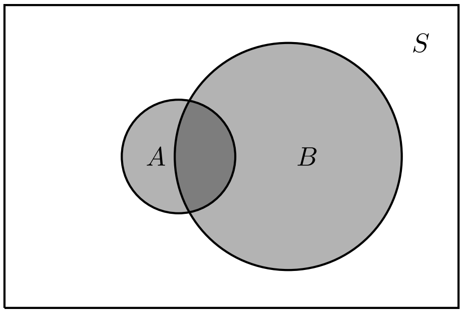
In general the probability of A given B is not equal to the probability of B given A. That is,

$$P(A|B) \neq P(B|A).$$

In fact, $P(A|B) = P(B|A)$ only if $P(A) = P(B)$. However, **Bayes' theorem** relates the two conditional probabilities as

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

We can understand this in terms of an example. Consider the sample S of people that drink coffee. Let A be the event that a coffee drinker adds sugar to their coffee and B be the event that they add cream to their coffee, then $A \cap B$ is the event in which a a coffee drinker adds both to their coffee. If $P(A) = 0.1$, $P(B) = 0.3$ and $P(A \cap B) = 0.05$. The Venn diagram for this situation might be as shown below.



Calculating the conditional probabilities, we find that

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{0.05}{0.1} = 0.5$$
$$P(A|B) = \frac{P(B \cap A)}{P(B)} = \frac{0.05}{0.3} = 0.17$$

Notice that $P(B|A) \neq P(A|B)$. We can 'see' this in the Venn diagram by noting that the intersection of A and B takes up a much larger portion of A than it does B. We can also use this example to verify Bayes' theorem

$$0.5 = \frac{0.3}{0.1}0.17.$$

If we want to use Bayes' theorem when $P(A \cap B)$ is known but $P(A)$ is unknown, we can make the substitution $P(A) = P(A \cap B) + P(A \cap B^c)$ to get

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

Tip:

There are often multiple ways of approaching the same problem, and sometimes, one approach is a lot easier than the other. For example, you might wonder what is the probability that the read marble is in the green bag, but you could also wonder, given marbles and bags, how many placement options does each marble have if I distribute them into the bags.

Mistakenly equating $P(B|A)$ and $P(A|B)$ causes reasoning errors such as the *base rate fallacy*. For example, if the police are using breathalyzers that always detect very drunk drivers, but falsely declare that sober drivers are drunk in 2% of the cases, and if 1 out of every thousand drivers is drink, then given a specific traffic stop in which a breathalyzer detected drunkenness, what is the probability that the driver is actually drunk?

If we let A be the event in which the breathalyzer declares that a driver is drunk and B be the event in which the driver is actually drunk then, $P(B)$, the probability that a driver is drunk is $1/1000$, $P(A)$, the probability that the breathalyzers says "drunk" is $20.98/1000$ (1 person in a thousand plus 2% of 999 people out of a thousand), and $P(A|B)$, the probability that the breathalyzer says "drunk" given that the driver is drunk is 100%.

Bayes' theorem tells us that the probability that the driver is drunk given that the breathalyzer says "drunk" is

$$P(B|A) = \frac{0.001}{0.02098}(1) = 0.0477 = 4.77\%.$$

This probability is much lower than what most people would intuitively predict, but it can be easily understood in terms of more concrete numbers. Out of 1000 drivers, 1 of them is drunk. Out of 1000 people, the breathalyzer claims that about 21 of them are drunk (the one actually drunk person plus 2% of the remaining people. Clearly, the probability that a driver is drunk given that the breathalyzer claims he is drunk is only 1/21.

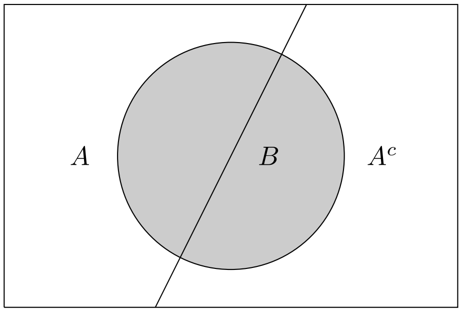
Tip:

As in the breathalyzer example, reforming questions in terms of concrete numbers (e.g. numbers of people instead of probabilities and percents) can make things a lot easier to understand. It can also serve as a check for problems that require Bayes' theorem.

If the sample space is partitioned into two pieces A and A^c, then an event B in the sample space is partitioned into the portion that is an A and the portion that is in A^c. In other words, we can write

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c) \\ &= P(B|A)P(A) + P(B|A^c)P(A^c). \end{aligned}$$

This situation can be understood with the help of the Venn diagram shown here:



So when the sample space is split into two partitions, Bayes' theorem can be written as

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

Bayes' theorem can be extended to any number of partitions named $A_1, \ldots, A_k$. Any two partitions cannot overlap. Since the partitions are disjoint and they cover the entire sample space,

$$\begin{aligned} A_i \cap A_i &= \phi, \quad \text{for } i \neq j \\ A_1 \cup A_2 \cup \cdots \cup A_k &= S. \end{aligned}$$

Taking the probability of both sides of each of the above equations gives us

$$\begin{aligned} P(A_i \cap A_i) &= 0, \quad \text{for } i \neq j \\ P(A_1) + P(A_2) + \cdots + P(A_k) &= 1. \end{aligned}$$

Then Bayes' theorem tells us that

$$\begin{aligned} P(A_j|B) &= \frac{P(A_j \cap B)}{P(B)} \\ &= \frac{P(A_j \cap B)}{\sum_{i=1}^{k}(A_i \cap B)} \\ &= \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}. \end{aligned}$$

A **probability tree** can be very helpful in organizing and understanding the probabilities for both independent and dependent events, especially when the sample space is partitioned.

The events in a partitioned system are mutually exclusive and collectively exhaustive.

At each node, the probabilities for the branches sum to 1. For the probability tree drawn below

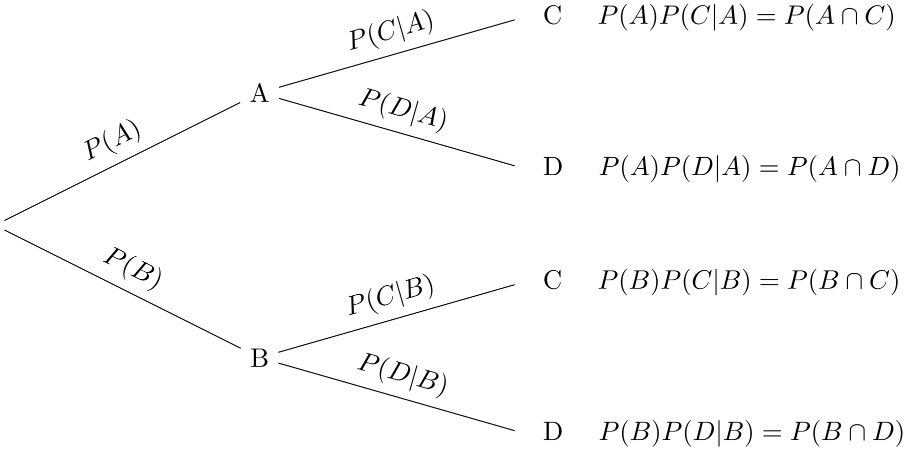
$$\begin{aligned} P(A) + P(B) &= 1 \\ P(C|A) + P(D|A) &= 1 \\ P(C|B) + P(D|B) &= 1. \end{aligned}$$

We also know that

$$\begin{aligned} P(C) &= P(A \cap C) + P(B \cap C) \\ &= P(A)P(C|A) + P(B)P(C|B) \\ P(D) &= P(A \cap D) + P(B \cap D) \\ &= P(A)P(D|A) + P(B)P(D|B). \end{aligned}$$

Tip:

It's sometimes hard to remember if you should multiply a pair of probabilities or add them. A good rule is that probabilities are multiplied if the experiment is layered and they are added if the sample space is partitioned. This is made much clearer with the use of a probability tree. You multiply if you're going from left to right and you add if you're going from top to bottom.



Two events A and B are **independent** if and only if any one of the following statements are true

- $P(B|A) = P(B)$
- $P(A|B) = P(A)$
- $P(A \cap B) = P(A)P(B)$

Recall that two events A and B are **disjoint** if and only if any one of the following statements are true

- $A \cap B = \phi$
- $P(A \cap B) = 0$

Independent and disjoint are not the same thing. Notice that $P(A \cap B) = 0$ for disjoint events but $P(A \cap B) = P(A)P(B)$ for independent events. Disjoint probabilities are mutually exclusive—if one happens, the other cannot happen. Independent probabilities have no impact on each other.

Example:

Two factories, A and B produce phones. Of the ones produced by A, 5% turn out to be defective as do 2% of those produced by B.

1. Your friend purchased a phone with equal likelihood from A or B. What is the probability that the phone is defective?
2. Given a defective phone purchased from either factory with equal likelihood, what is the probability that it came from A?
3. If you purchase a pair of phones from either A or B with equal likelihood and the first one you open is defective, what is the conditional probability that the second one is also defective?

A good way to start with these questions is the probability tree shown below. We let D denote a defective phone and D^c denote a phone that is not defective.

For the first question, we know that a defective phone can either be purchased at A or B. The total probability of getting a defective phone is the probability of going to A and buying a defective phone $P(A \cap D)$ or going to B and buying a defective phone $P(B \cap D)$. So the probability is

$$\begin{aligned} P(D) &= P(A \cap D) + P(B \cap D) \\ &= P(D|A)P(A) + P(D|B)P(B) \\ &= (0.05)(0.5) + (0.02)(0.5) \\ &= 0.035 \end{aligned}$$

This result makes sense as it is the average of the two probabilities for getting a defective phone given A and given B.

For the second question, we are being asked to find $P(A|D)$. This is just a matter of using the probability tree and applying Bayes' theorem.

$$\begin{aligned} P(A|D) &= \frac{P(A)}{P(D)} P(D|A) \\ &= \frac{0.5}{0.035} 0.05 \\ &= 0.71 \end{aligned}$$

The third question is a little more difficult. If we call D_1 the event that the first phone is defective and D_2 the event that the second phone is defective, then we're asked to find $P(D_2|D_1)$. Expanding the conditional probability gives us

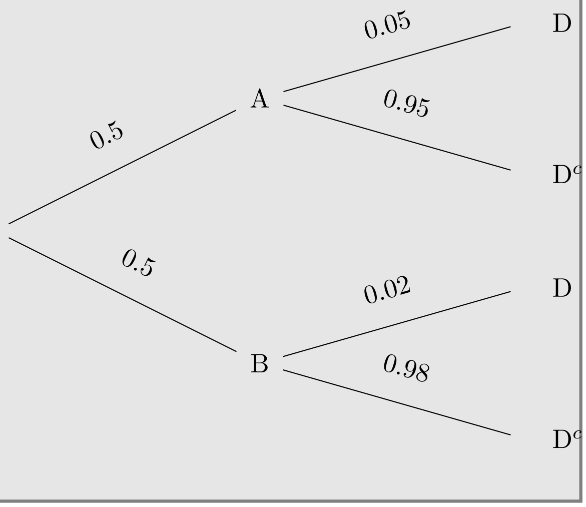
$$P(D_2|D_1) = \frac{P(D_2 \cap D_1)}{P(D_1)}.$$

Some of the defective phones come from factory A and the rest come from factory B. That is to say, the sample space of defective phones is *partioned* in two, and we can expand the right side of the equation above to

$$\frac{P(D_2 \cap D_1|A)P(A) + P(D_2 \cap D_1|B)P(B)}{P(D_1|A)P(A) + P(D_1|B)P(B)}.$$

We can find each of these probabilities using the tree diagram. For example, $P(D_2 \cap D_1|A)$ is that probability that both phones are defective given that they were purchased at factory A. The factory A condition puts us on the line A—D, which has a probability of 0.05. So the probability for both phones to be defective is (0.05)(0.05). The total probability is given by

$$\begin{aligned} P(D_2|D_1) &= \frac{(0.05^2)(0.5) + (0.02)^2(0.5)}{(0.05)(0.5) + (0.02)(0.5)} \\ &= 0.041. \end{aligned}$$



1.4 Discrete Random Variables

A **random variable** is typically a function that assigns a real number to each element in the sample space. For example, if you're interested in the sum of a pair of dice where X is their sum, then X is a random variable.

- Each random variable has a range of values that it can take on the real number line. Our example X, can take on the integer values from 2 (e.g. $1+1$) to 12 (e.g. $6+6$) since it is the sum of the values of two dice. Notice that the possible values of a random variable are generally not equivalent to the values of the sample space. In our sample space the values go from 1 to 6, but X goes from 2 to 12.
- Each random variable has a probability assigned to each point in its range. In our example, $P(X = 12) = \frac{1}{36}$ since that is the probability that throwing two dice will yield a sum of 12.
- These probabilities form a **probability distribution** of the random variable.

There are different kinds of random variables.

- **Categorial random variables:** These are non-numerical random variables. For example, the blood type of a random person or their state of residence may be categorical random variables.
- **Numerical random variables:** We will be focusing on this kind.
  - **Discrete random variables:** This kind of random variable has a countable range. We can write down all the possible values of the random variable and calculate each probability.
  - **Continuous random variables:** The range of a continuous random variable is a part of the real numbers. The possible values of the random variable are not countable. For example, the length of randomly selected phone call is a continuous random variable since it can take on any real-values from 0 to some large number of seconds.

For a random variable X, the probability that X has the value x, represented by $P(X = x)$ is typically denoted $p(x)$ called the **probability mass function** or sometimes "probability distribution" or just "probability function". The analogous function for continuous random variables is the "probability density function".

$$p(x) = P(X = x)$$

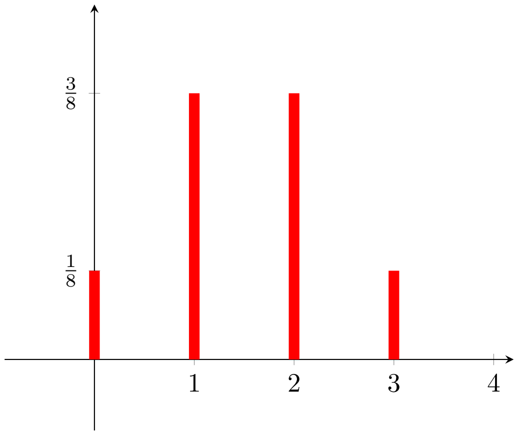
For discrete random variables, we enumerate the range of the random variable and their individual probabilities in a table.

For example, if you toss a coin three times in a row, and let the random variable X be the number of throws that resulted in heads, then the possible outcomes are {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}. The variable X can take on the value 0 (for no heads), 1, 2, or 3 (all heads). There are 8 possibilities in the sample space, but only one in which the number of heads is zero, so $P(X = 0) = \frac{1}{8}$, and so on. We enumerate the possible values x of X and their probabilities in the table

x:	0	1	2	3
S:	TTT	HTT THT TTH	THH HTH HHT	HHH
$p(x)$:	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Keep in mind that $p(x)$ takes in a number and $P(X = x)$ takes in an event.

A graph of the probability mass function looks like this:



When asked to give the probability mass function for a discrete random variable, your answer should be a table like the one shown above. The middle row in the table doesn't need to be in the answer–it's just displayed above to show how the elements of the sample space are distributed.

Notice that the probabilities in the bottom row of the table above sum to 1, and this should always happen. In other words,

$$\sum_x p(x) = 1.$$

When doing a probability mass function, always sum the probabilities (as a double-check) ensuring that they sum to 1. Since $P(X = x)$ is a probability, the values of $p(x)$ will always be greater than or equal to zero.

The **cumulative distribution function** computes the probability that the observed value of a random variable X is less than or equal to some real number x. The cumulative distribution function $F(x)$ abbreviated CDF is defined as

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t).$$

Some properties of the CDF are

- It stays between 0 and 1
$$0 \leq F(x) \leq 1.$$
- It is always increasing.
- It is defined on all $\mathbb{R}$. So x (the input of $F(x)$) can be any real number rather than just a discrete value.
- It is right-continuous.

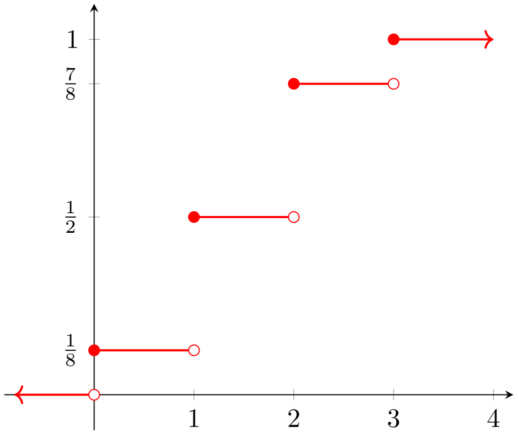
So the CDF is a piecewise function for discrete random variables.

We can easily determine $p(x)$ (i.e. the table from earlier), by looking at a plot of $F(x)$. This plot gives the CDF for the three coin toss problem given above. In the graph below, we can tell by looking at the horizontal axis, that X can take on the values 0, 1, 2, and 3. We can also see that $p(0) = P(X = 0) = \frac{1}{8}$ since the CDF jumps from 0 to $\frac{1}{8}$ at that point. The next jump, occurring at $x = 1$ is from $\frac{1}{8}$ to $\frac{1}{2}$, so we know that $p(1) = P(X = 1) = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$.

Notice that, $p(1) = F(1) - F(0)$, and in general,

$$p(x_2) = F(x_2) - F(x_1),$$

where x_1 and x_2 are consecutive values taken on by X.



The CDF is defined over all real numbers. Notice in the example above that $F(-100) = F(-0.0001) = 0$, that $F(100) = F(3) = 1$, and that $F(1.5) = F(1) = \frac{1}{2}$.

Notice in the above example, that $P(0.1 < X < 2.5) = \frac{3}{4}$ is the probability that X has a value between 0.1 and 2.5. In general, we can calculate this as

$$\begin{aligned} P(x_1 < X < x_2) &= P(X < x_2) - P(X < x_1) \\ &= F(x_2) - F(x_1). \end{aligned}$$

Given the cumulative distribution function ($F(x)$) as a piecewise-defined function, to find the probability mass function ($p(x)$), start by graphing the CDF from the piecewise-defined function. From that, you can deduce the graph of the PMF and the table.

The **expectation**, also called the "mean" or the "expected value" is the value you expect to get in a statistical experiment. The expectation of a discrete random variable X with a probability mass distribution $p(x)$ is calculated as

$$\mu = E(X) = \sum_x xp(x).$$

In other words, you just multiply each value of x by its probability and sum them up. This is easy if you have the table relating x and $p(x)$. You just multiply the corresponding values from the top row and the bottom row and add them all up. In our example with the three dice, the expection is

$$\mu = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5.$$

For continuous random variables, the expectation is calculated using an integral rather than a sum.

The expectation value is very helpful when talking about games in which you can win or lose money. If the random variable X takes on the possible gains/losses that you can make in a single game, then $p(x)$ gives the probabilities associated with each of those values. The expectation is then the average that you would expect to make in a single game. More accurately, μn gives you the dollar amount that you should expect to have gained or lost after playing n games.

Suppose you have a new random variable $g(X)$, which depends on X, and that X is a discrete random variable with probability mass function $p(x)$. Then $g(X)$ is a **function of a random variable**. For example, if you were interested in X^2 rather than just X, then $g(X) = X^2$. The expectation of $g(X)$ is

$$\mu_{g(X)} = E\left[g(X)\right] = \sum_x g(x)p(x).$$

The average of the possible values of a random variable's squared distance from its expected value is termed its **variance** and is denoted $\text{Var}(X)$ or σ^2. There's nothing special about variance—it just quantifies how much the possible values of a random variable are spread out. The variance of a random variable X with probability mass function $p(x)$ and expected value $\mu = E(X)$ is always positive and it is calculated as

$$\begin{aligned}\mathrm{Var}(X) &= \sum_x (x-\mu)^2 p(x) \\ &= \sum_x (x^2 - 2x\mu + \mu^2) p(x) \\ &= \sum_x (x^2 p(x) - 2x\mu p(x) + \mu^2 p(x)) \\ &= \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\ &= E(X^2) - 2\mu(\mu) + \mu^2(1) \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2 .\end{aligned}$$

Notice that the variance is always positive.

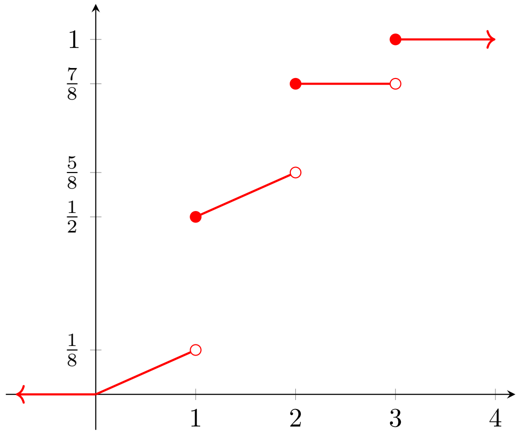
Some properties of the expectation and variance are

- $E(aX+b) = aE(X)+b$
- $\mathrm{Var}(X) = E(X^2) - (E(X))^2$
- $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

The **standard deviation** of X is the positive square root of the variance of X

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Consider cumulative distribution function shown in the graph below.



Notice that this is not the cumulative distribution function of a discrete random variables because in $[0,2]$, there are two continuous segments. Calculating probabilities from this kind of graph is a little different from graphs of purely discrete functions.

Notice that $P(X=3) = F(3) - F(2) = \frac{1}{8}$ is calculated as usual since those segments are discrete. However, $P(X=2)$ is calculated as $F(2) - F(\to 2) = \frac{1}{4}$. That is, it is the value of $F(x)$ at $x=2$ minus the value of $F(x)$ as it approaches 2 from the left. Similarly, $P(X=1) = F(1) - F(\to 1) = \frac{3}{8}$. Notice that these are still just the values of the gaps.

Beware... the probability of a specific value in a continuous segmet is zero. For example, $P(X=1.5)=0$ because there is no vertical gap (i.e. sudden jump) there. Another result of the continuous segments is that $P(X=1)+P(X=2)+P(X=3) \neq 1$ which is different from the purely discrete case.

To calculate the probability for a range, split it into multiple expressions. For example,

$$\begin{aligned}P(0.5 < X < 1.5) &= P(0.5 < X \leq 1.5) - P(X=1.5) \\ &= P(0.5 < X \leq 1.5) \\ &= P(X \leq 1.5) - P(X \leq 0.5) \\ &= F(1.5) - F(0.5).\end{aligned}$$

If you're given the piecewise function $F(x)$, you can now evaluate the above exactly. Otherwise, you can estimate the value from the given graph, which looks like $F(1.5) - F(0.5) = \frac{9}{16} - \frac{1}{16} = \frac{1}{2}$.

1.4.1 Transformations of Random Variables

A transformation is when you take a discrete random variable, such as X, with a given probability mass function, then you create a new random variable, such as Y, which is a function of X.

The probability mass function of X looks like:

x:	x_1	x_2	$\cdots$	x_k
$p(x)$:	p_1	p_2	$\cdots$	p_k

For a linear transformation such as $g(X) = Y = a + bX$, the transformation is one-to-one, and the probabilities stay the same. The probability mass function for Y looks like

y:	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_k)$
$p(y)$:	p_1	p_2	$\cdots$	p_k

or for $g(X) = Y = a + bX$, it is

y:	$a+bx_1$	$a+bx_2$	$\cdots$	$a+bx_k$
$p(y)$:	p_1	p_2	$\cdots$	p_k

For a nonlinear transformation, such as $g(x) = Y = X^2$, the transformation is not necessarily one-to-one, so the bottom row (i.e. the probabilities) won't necessarily stay the same. However, the same procedure is followed in that the probability mass function

x:	x_1	x_2	$\cdots$	x_k
$p(x)$:	p_1	p_2	$\cdots$	p_k

transforms to

y:	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_k)$
$p(y)$:	p_1	p_2	$\cdots$	p_k

However, in the end, if the $g(x_i)$ are not all unique, that is there are multiple columns with the same value in the top row, then these get combined into unique columns by adding the probabilities in the second row. For example, if the probability mass function is

x:	-1	1	10
$p(x)$:	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

and the transformation is $g(x) = Y = X^2$, then the transformed probability mass function

y:	1	1	100
$p(y)$:	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

must be adjusted by combining the first two columns since the entries in the top row are the same. The final transformation is

y:	1	100
$p(y)$:	$\frac{5}{6}$	$\frac{1}{6}$

1.5 Continuous Random Variables

A random variable that can take on values from a continuous interval, it is a **continuous random variable**. For example, if X is the random variable defined as the time it takes for a fast food order to be completed, then X is a continuous random variable since the time can take on real number values. For continuous random variables, the possible outcomes are not countable.

For a continuous random variable X, $f(x)$ is its **probability density function** if

- $f(x) \geq 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- $P(a < X < b) = \int_a^b f(x)\,dx$

The probability density function is the continuous analogue to the probability mass function for discrete random variables. For the continuous case, the total area under the curve must be one. Notice that $f(x)$ can be greater than 1 since it is a probability density function rather than a probability.

Example:

If X is a continuous random variable with the probability density function $f(x)$, what is the value of the constant k?

$$f(x) = \begin{cases} k(6x - 3x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Recall that $f(x) \geq 0$ for all x. Factoring the equation gives us $3kx(2-x)$, so it is a downward opening quadratic which is nonnegative between $x = 0$ and $x = 2$. Recall also that the integral from negative infinity to positive infinity must yield exactly 1. Since $f(x) = 0$ everywhere except in the interval $[0, 2]$, we can integrate over this interval and ignore the rest.

$$\begin{aligned} \int_0^2 (6x - 3x^2)\,dx &= k(3x^2 - x^3)\Big|_0^2 \\ &= 4k. \end{aligned}$$

So in order for this to be a valid probability density function, $k = \frac{1}{4}$.

To check if a given function is a valid probability density function

1. Check that the function is greater than or equal to zero everywhere. If you're given a piecewise defined function, compare each function with its given interval to make sure that the function is greater than or equal to zero in that interval.
2. Check that the total area under the curve is one, by integrating each function over its given interval and summing them.

Example:

Given the probability density function

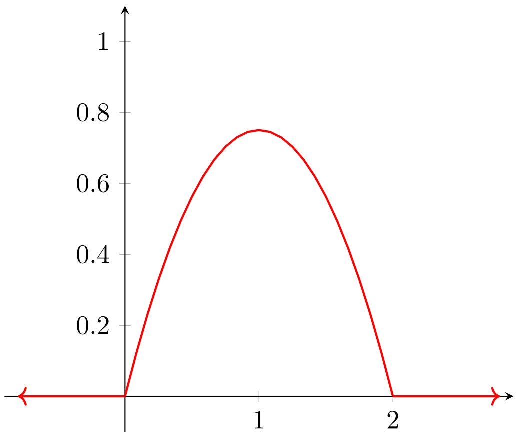
$$f(x) = \begin{cases} \frac{1}{4}(6x - 3x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

for a continuous random variable X, what is $P(X > 1)$ the probability that X is greater than 1?

To find this, we simply integrate the probability density function above $x = 1$, that is, from $[1, \infty]$. However, since $f(x) = 0$ for $[2, \infty]$, we only have to integrate from $[1, 2]$.

$$\int_1^2 \frac{1}{4}(6x - 3x^2)\,dx = \frac{1}{2}.$$

The graph of the probability density function in the above example looks like this:



The **cumulative distribution function** for continuous random variables is very similar to the one for discrete random variables except that we integrate instead of sum. If $f(x)$ is the probability density function for a continuous random variable X, then the cumulative distribution function is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x')\,dx'.$$

It follows that

$$P(a < X < b) = \int_a^b f(x)\,dx = F(b) - F(a).$$

Notice that

$$\begin{aligned} P(X > a) &= \int_a^{\infty} f(x)\,dx = 1 - F(a) \\ P(X < b) &= \int_{-\infty}^{b} f(x)\,dx = F(b) \\ P(X = c) &= \int_c^c f(x)\,dx = 0. \end{aligned}$$

So the probability of X being a specific value is 0 for continuous random variables. This means that, unlike with discrete random variables, we don't have to be as careful with out inequality signs because $P(a \leq X \leq b) = P(a < X < b)$.

Example:

Find the cumulative distribution function if the probability density function is

$$f(x) = \begin{cases} \frac{1}{4}(6x - 3x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

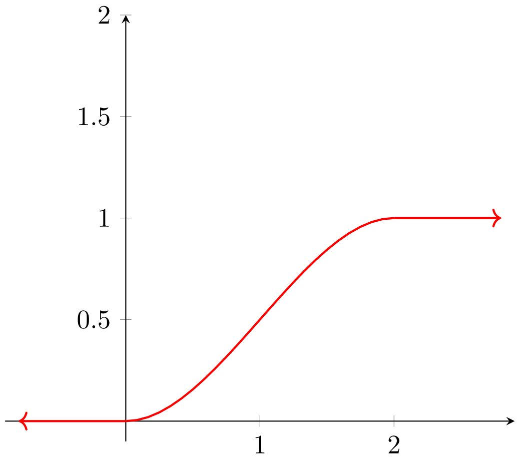
We know that on the interval $[-\infty, 0]$, $F(x) = 0$ since it has not accumulated any area under the curve. We also know that on the interval $[2, \infty]$, $F(x) = 1$, since it has accumulated 1 unit of area under the curve. We need to integrate the given function to find the function for $F(x)$ in the interval $[0, 2]$. Integrating, we find that

$$\int_0^x \frac{1}{4}(6x' - 3x'^2)\, dx' = \frac{1}{4}(3x^2 - x^3),$$

so our cumulative distribution function is

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{4}(3x^2 - x^3) & 0 < x < 2 \\ 1 & x \geq 2 \end{cases}$$

The graph of the cumulative distribution function of the example above is

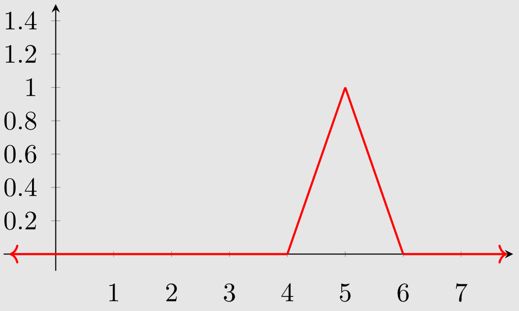


Example:

Determine the cumulative distribution function given the probability density function

$$f(x) = \begin{cases} 1 - |x-5| & 4 < x < 6 \\ 0 & \text{otherwise} \end{cases}$$

The graph of $f(x)$ looks like



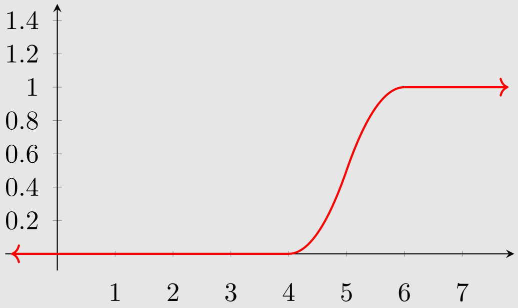
To find the cumulative distribution function, we first get rid of the absolute value by splitting the function into two functions. We know that $1 - |x-5|$ can be split into $1-(5-x)$ and $1-(x-5)$, so

$$f(x) = \begin{cases} x-4 & 4 < x < 5 \\ 6-x & 5 < x < 6 \\ 0 & \text{otherwise} \end{cases}$$

The cumulative distribution function is then broken into four intervals. For $[-\infty, 4]$, we know that $F(x) = 0$, and for $[6, \infty]$, we know that $F(x) = 1$. For the first nonzero interval, $[4, 5]$, we know that $F(x) = \int_4^x (x-4)\, dx$. For the second interval, it is $\int_5^x (6-x)\, dx$ $\underline{\text{plus}}$ the previous area under the curve, which is $\int_4^5 (x-4)\, dx$. Performing the integrations, we have that

$$F(x) = \begin{cases} 0 & x \leq 4 \\ \frac{1}{2}x^2 - 4x + 8 & 4 < x \leq 5 \\ -\frac{1}{2}x^2 + 6x - 17 & 5 < x < 6 \\ 1 & x \geq 6 \end{cases}$$

The graph of $F(x)$ looks like



After finding the cumulative distribution function, there are several things you should do to double-check your work:

1. Differentiate $F(x)$. You should get $f(x)$.
2. $F(x)$ must be continuous. Evaluate $F(x)$ on both sides of every cusp in $f(x)$. In the example above, $f(x)$ has cusps at $x = 4, 5, 6$. We check the first cusp by evaluating the first and second lines at $x = 4$. They both give $F(4) = 0$ as it should. Then we check the second cusp by evaluating the second and third lines of $F(x)$ at $x = 5$. They both give $F(5) = \frac{1}{2}$ as expected. Finally, we check the third cusp by evaluating the third and fourth lines of $F(x)$ at $x = 6$. They both give $F(6) = 1$ as expected. Therefore, $F(x)$ is continuous everywhere. We can also confirm just by graphing $F(x)$.
3. $F(x)$ must be a monotonically increasing function bounded at the bottom by $F(x) = 0$ and at the top by $F(x) = 1$. This can be confirmed by graphing it.

In general, if the probability density function $f(x)$ is given as the piecewise function

$$f(x) = \begin{cases} f_1(x) & x < a \\ f_2(x) & a < x < b \\ f_3(x) & x > b \end{cases},$$

then the cumulative distribution function will have the form

$$F(x) = \begin{cases} \int_{-\infty}^x f_1(x')\, dx' & x < a \\ \int_{-\infty}^a f_1(x')\, dx' + \int_a^x f_2(x')\, dx' & a < x < b \\ \int_{-\infty}^a f_1(x')\, dx' + \int_a^b f_2(x')\, dx' + \int_b^\infty f_3(x')\, dx' & x > b \end{cases}.$$

Typically, however, we'll be given a simple probability density function of the form

$$f(x) = \begin{cases} f_1(x) & a < x < b \\ 0 & \text{otherwise} \end{cases},$$

in which case the cumulative distribution function has the form

$$F(x) = \begin{cases} 0 & x \leq a \\ \int_{-\infty}^x f_1(x')\, dx' & a < x < b \\ 1 & x \geq b \end{cases}.$$

Recall that $P(X = c) = 0$ for continuous random variables, so where exactly we place the equality part of the inequalities in the cumulative distribution function doesn't matter provided that we define $F(x)$ for all real numbers.

For a continuous random variable X with a probability density function $f(x)$, the **expectation** of X is

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

and for a function $g(X)$ of a random variable X, the expectation is

$$\mu_{g(X)} = E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx.$$

The **variance** can be expressed in several different ways including

$$\begin{aligned}\mathrm{Var}(X) &= E((x-\mu)^2) \\ &= \int_{-\infty}^{\infty} (x - E[X])^2 f(x)\, dx \\ &= E(X^2) - [E(X)]^2 \\ &= \int_{-\infty}^{\infty} x^2 f(x)\, dx - [E(X)]^2 .\end{aligned}$$

Keep in mind the following properties of expectation and variance which hold true for both discrete and continuous random variables:

$$\begin{aligned}E(aX+b) &= aE(X) + b \\ \mathrm{Var}(X) &= E(X^2) - [E(X)]^2 \\ \mathrm{Var}(aX+b) &= a^2 \mathrm{Var}(X).\end{aligned}$$

Example:

Calculate the expectation and variance of X if

$$f(x) = \begin{cases} \frac{1}{4}(6x - 3x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}.$$

The expectation is

$$\begin{aligned}\int_{-\infty}^{\infty} x f(x)\, dx &= \frac{1}{4}\int_0^2 x(6x - 3x^2)\, dx \\ &= \frac{1}{4}\left(2x^3 - \frac{3}{4}x^4\right)\Big|_0^2 \\ &= 1.\end{aligned}$$

The variance is

$$\begin{aligned}\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - E[X])^2 f(x)\, dx \\ &= \frac{1}{4}\int_0^2 (x-1)^2(6x - 3x^2)\, dx \\ &= \frac{1}{5}.\end{aligned}$$

We could also have calculated the variance as

$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{1}{4}\int_0^2 x^2(6x - 3x^2)\, dx - (1)^2 \\ &= \frac{1}{5}.\end{aligned}$$

Usually, the second method is easier if we already have the value of $E(X)$ because we don't have to expand $(x - E(X))^2$ in the integrand.

Example:

Find the expectation and variance of $g(X) = 3X + 2$ if

$$f(x) = \begin{cases} \frac{1}{4}(6x - 3x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}.$$

We could calculate this using the definition

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

but it's a much easier if we use the property

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

From the previous example, we know that $E(X) = 1$, and since $a = 3$ and $b = 2$, we have that

$$E(3X + 2) = 5.$$

Similarly, we can find the variance of $g(X)$ using the property

$$\text{Var}(aX + b) = a^2 \text{Var}(X),$$

to get

$$\text{Var}(3X + 2) = \frac{9}{5}.$$

Notice that in the example above, we could also have calculated the variance of $g(X)$ using the property

$$\text{Var}(X) = E(X^2) - [E(X)]^2 .$$

In our case

$$\begin{aligned} \text{Var}(X) &= E\left([3X+2]^2\right) - [E(3X+2)]^2 \\ &= E\left(9X^2 + 12X + 4\right) - [E(3X+2)]^2 \\ &= 9E(X)^2 + 12E(X) + 4 - [3E(X) + 2]^2 \\ &= 9\left(\frac{6}{5}\right) + 12(1) + 4 - [3(1) + 2]^2 \\ &= \frac{9}{5}. \end{aligned}$$

By using the expectation operator wisely, we can almost always avoid doing the tedious integration.

1.6 Discrete Probability Distributions

1.6.1 Uniform Distribution

If the random variable takes on any of its possible values with equal probability, then it is a **uniform distribution**. For example, if the random variable X is the value of a dice throw, then the random variable can take on the values from 1 to 6, and it does so with equal probabilities. That is the probability that $X = 1$ is the same as the probability that $X = 2$, and so on.

If X is the uniform discrete random variable, which takes on the values $x_1, x_2, \ldots$, then the probability is calculated as

$$p(x; k) = P(X = x) = \frac{1}{k}, \qquad x = x_2, x_2, \ldots.$$

The expectation of X is

$$\mu = E(X) = \frac{1}{k}\sum_{i=1}^{k} x_i,$$

and the variance is

$$\sigma^2 = \mathrm{Var}(X) = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2.$$

For example, if X is the value of a dice throw, then

$$\begin{aligned} P(X = x) &= \frac{1}{6}, \qquad x = 1,2,3,4,5,6 \\ E(X) &= \frac{1}{6}(1+2+3+4+5+6) = 3.5 \\ \mathrm{Var}(X) &= \frac{1}{6}\sum_{x=1}^{6}(x - 3.5)^2 = 2.92. \end{aligned}$$

For uniform distributions, the possible values of the random variable do not have to be equally spaced–they only have to occur with equal probability. For example, if the two on a dice was scraped off and replaced with a 7, then the possible values would be $1, 3, 4, 5, 6, 7$. The possible values are not equally spaced, but they still have equal probability, so it is still a uniform distribution.

1.6.2 Bernoulli Distribution

A **bernoulli trial** is an experiment that has two possible outcomes. For example, a single flip of a coin has two possible outcomes–heads or tails, and so a coin flip is a bernoulli trial.

If X is a discrete random variable associated with a single bernoulli trial, then the probability mass function for the bernoulli distribution is

$$P(X = x) = p^x(1-p)^{1-x}, \qquad x = 0, 1.$$

Typically, we let $x = 1$ mean "success" and $x = 0$ mean failure. Notice that there are only two possibilities: $P(X = 0) = 1 - p$ and $P(X = 1) = p$. The expectation for a bernoulli distribution is

$$E(X) = p,$$

and the variance is

$$\mathrm{Var}(X) = p(1-p).$$

For example, for a single coin flip, we have a bernoulli distribution with $p = 0.5$. We say $X \sim \mathrm{Ber}(0.5)$ to indicate that X follows a bernoulli distribution and $p = 0.5$. For a coin flip, if we let heads be 1 and tails 0, then the probability of getting heads is $P(X = 1) = 0.5$ and the probability of getting tails is $P(X = 0) = 0.5$. The expected value is just $E(X) = 0.5$, and the variance is $\mathrm{Var}(X) = 0.25$.

1.6.3 Binomial Distribution

A **binomial trial** is an experiment that is repeated a fixed number of n times, where each trial is independent of the others, each trial has only two possible outcomes (i.e. success or failure), and the probability of a certain outcome (e.g. success) is p, the same for each trial.

The number of successes X in a fixed number n of bernoulli trials is called a **binomial random variable**. The associated probability distribution is called a **binomial distribution** and denoted $b(x; n, p)$ since it depends on the two fixed parameters n and p.

For example, the number of heads X obtained after $n = 10$ coin flips follows a binomial distribution, where the probability of getting heads in each trial is $p = 0.5$.

The probability mass function for a binomial random variable X is

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots n,$$

the expected value is

$$E(X) = np,$$

and the variance is

$$\mathrm{Var}(X) = np(1-p).$$

A binomial trial is effectively repeating a bernoulli trial n times. Keep in mind that n is the fixed number of trials–it can't be modified during the binomial experiment. Notice that a bernoulli distribution is the special case of a binomial distribution with $n = 1$.

Continuing our example, the probability of getting exactly 3 heads in 10 coin flips is

$$P(X = 3) = \binom{10}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{10-3} = 11.7\%.$$

The expect value (i.e. the expected number of heads is $E(X) = 10 \cdot 0.05 = 5$, and the variance is $\mathrm{Var}(X) = 2.5$.

Example:

Twenty of a certain component are manufactured and need to be tested. If each is good with a probability of 0.8, what is the probability that at least 15 of the components are good? What is the probability that between 5 and 15 of them are good? What is the probability that exactly 10 of them are good.

If we let X be the number of good components, then X follows a binomial distribution with $n = 20$ and $p = 0.8$.

The probability that at least 15 of the components are good can be calculated as

$$\begin{aligned} P(X \geq 15) &= P(X = 15) + \cdots + P(X = 20) \\ &= \sum_{x=15}^{20} \binom{20}{x} (0.8)^x (1 - 0.8)^{20-x} \\ &= 80.4\%. \end{aligned}$$

The probability that between 5 and 15 (including 5 and 15) of the components are good is

$$\begin{aligned} P(5 \leq X \leq 15) &= \sum_{x=5}^{15} \binom{20}{x} (0.8)^x (1 - 0.8)^{20-x} \\ &= 37.0\%. \end{aligned}$$

The probability that exactly 10 of them are good is

$$P(X = 10) = \binom{20}{10} (0.8)^{10} (1 - 0.8)^{20-10} = 0.20\%.$$

1.6.4 Poisson Distribution

A random variable X that is the number of outcomes occurring in a given continuous interval, is a **poisson random variable**. The given interval can be an interval of time such as an hour or a year or even a region or area such as a county or a city. Some examples of poisson random variables include

- The number of radioactive decay events in a minute
- The number of accidents per month in a city
- The number of calls received in an hour
- The number of deer in a given square mile
- The number of typos in a book

In order to be a poisson process, the number of outcomes occurring in nonoverlapping regions must be independent, the probability of exactly one outcome in an interval of length h is approximately λh, and the probability of two or more outcomes in a sufficiently short interval is essentially zero. A poisson process has paramter λ, which is the average number of outcomes in unit interval.

The probability distribution for a poisson random variable X is given by

$$P(X = x) = p(x; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

and the expectation and variance are

$$E(X) = \mathrm{Var}(X) = \lambda t.$$

If the number of customers walking into a store in a given hour is a poisson random variable, then λ is the average rate of customers per hour. If the average rate of customers per hour is 20, then the expected number of customers in 8 hours is simply $\lambda t = 20 \cdot 8 = 160$.

Example:

A geiger counter pointed at radioactive sample is detecting an average of 3 particles per second. What is the probability that no particles are detected in a given second? What is the probability that 20 particles are detected in a given second? What is the probability that exactly 3 particles are detected in a given second? What is the probability that more than 5 particles will be detected in a given second.

The number of decay events in a second is a poisson random variable with $t = 1$ second and $\lambda = 3$.

In the first case, we have

$$P(X = 0) = \frac{e^{-3} (3)^0}{0!} = 4.98\%.$$

In the second case, we have

$$P(X = 20) = \frac{e^{-3} (3)^{20}}{20!} = 0.00000000714\%.$$

In the third case, we have

$$P(X = 3) = \frac{e^{-3} (3)^3}{3!} = 22.4\%.$$

In the last case, we have

$$\begin{aligned} P(X > 5) &= 1 - P(X \leq 5) \\ &= 1 - \sum_{x=0}^{5} \frac{e^{-3} (3)^x}{x!} \\ &= 8.39\% \end{aligned}$$

The last case shows a trick, $P(X > x) = 1 - P(X \leq x)$ that we always need for poisson random variables. The resulting summation can easily be calculated on a calculator.

A poisson probability distribution can never be symmetric. The lower bound on x is zero since it doesn't make sense to count a negative number of outcomes in an interval, but it does not have an upper bound. Theoretically, there is a nonzero probability that any large number of events can occur in the specified interval.

1.6.5 Geometric Distribution

For repeated, independent trials, with a probability of success of p and a probability of failure of $1-p$, then if we let X be the number of the trial in which the first success occurs, we say that X follows a **geometric distribution** where

$$P(X=x)=g(x;p)=p(1-p)^{x-1}, \qquad x=1,2,3,\ldots.$$

The probability $P(X=x)$ for a geometric random variable, means the probability that the first success occurs on the xth trial. The expectation is

$$E(X)=\frac{1}{p},$$

and the variance is

$$\mathrm{Var}(x)=\frac{1-p}{p^2}.$$

We can understand the probability mass function for the geometric distribution fairly easily. The probability that the first trial is the first success $P(X=1)=p$ is simply the probability of any given trial being a success. The probability that the second trial is the first success $P(X=2)=p(p-1)$ is the probability that the first trial is a failure $1-p$ *and* the second trial is a success p. Similarly the probability that the xth trial is the first sucess $P(X=x)=p(1-p)^{x-1}$ is the probability that all $x-1$ of the previous trials failed with probability of $1-p$ for each of them and that the xth trial succeeded with probability p.

Example:

When repeatedly tossing a dice, what is the probability that the first '5' appears on the 20th toss?

If we let X be the number of the trial in which the first '5' appears, then X follows a geometric distribution and

$$P(X=20)=\frac{1}{6}\left(1-\frac{1}{6}\right)^{20-1}=0.0052.$$

If makes sense that the probability is very low because we would expect to see a '5' come up long before the 20th throw.

Notice that we could also have thought of this problem as failing the first 19 tosses and succeeding on the 20th toss and the probability is calculated as

$$\frac{5}{6}\cdot\frac{5}{6}\cdots\frac{5}{6}\cdot\frac{1}{6}=\left(\frac{5}{6}\right)^{19}\frac{1}{6}.$$

In the dice example above, the expected number of tosses before a 5 is encountered is $E(X)=6$. What is the most likely number of tosses before you encounter the first '5'? This is not the same as the expected value. If we plug in different values for x, we calculate

$$\begin{aligned} P(X=1) &= 0.167 \\ P(X=2) &= 0.139 \\ P(X=3) &= 0.116 \\ P(X=4) &= 0.096. \end{aligned}$$

It is always the case for geometric distributions that the largest probability is associated with the first trial. At first this seems strange, but keep in mind that for a geometric distribution, we are interested in the probability of some trial yielding the *first* success. Subsequent trials are less likely to yield the first success because not only does it require a success on that trial, it also requires a failure on every previous trial.

An interesting property of geometric distributions is the **memoryless property**. The memoryless property asserts that if a given event has not occurred by time t, then the probability that it occurs in additional time T is simply the probability that it occurs in time T. That is, the probability that the event occurs in time T is independent of how much time t has passed previously without it occurring.

For example, if the life of a car (given in terms of its mileage) follows a geometric distribution, then the probability that it dies in the next 10,000 miles is independent of the number of miles it has already survived.

1.6.6 Negative Binomial Distribution

A **negative binomial experiment** is one in which a trial is repeated until a fixed number k of successes occur. X is a negative binomial random variable if it is the number of the trial on which the kth success occurs. In a binomial experiment, we are interested in the number X of successes that occur in a fixed number of n trials, whereas in a negative binomial experiment, we are interested in the number of the trial X in which a fixed number k of successes has occurred. The binomial and negative binomial distributions really have the same parameters, but the random variable is switched. Notice that the geometric distribution is a special case of the negative binomial experiment with $k = 1$.

For a negative binomial distribution

$$P(X = x) = nb(x; k, p) = \binom{x-1}{k-1} p^k (1-p)^{x-k},$$

where $x = k, k+1, k+2, \ldots$. The probability $P(X = x)$ for a negative binomial experiment is understood of as the probability that the kth success occurs on the xth trial. The expectation is

$$E(X) = \frac{k}{p},$$

and the variance is

$$\text{Var}(X) = \frac{k(1-p)}{p^2}.$$

In order for the kth trial to be the kth success, $P(X = k) = p^k$, is the probability that the first k trials were all successes, each with probability p. The probability that the x trial is the kth success, $P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$, is the probability that there are $k-1$ successes in the first $x-1$ trials and exactly one success on the xth trial. This means we can break the problem into two regular binomial experiments $\binom{x-1}{k-1} p^{k-1} (1-p)^{x-k}$ and $\binom{1}{1} p^1 (1-p)^0$ and then multiply them together to get the final probability $\binom{x-1}{k-1} p^k (1-p)^{x-k}$.

1.6.7 Hypergeometric Distribution

If a subset n is randomly selected without replacement from N items where k of the N items are labelled as successes and $N - k$ of the N items are labelled as failures, then if we let X be the number of successes in our n selected items, we say that X follows a hypergeometric distribution.

If X is a hypergeometric random variable, then

$$P(X = x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}.$$

The expectation is

$$E(X) = \frac{nk}{N},$$

and the variance is

$$\text{Var}(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right).$$

Once you have defined it as a hypergeometric random variable, then the expectation is easy to calculate. It is simply the ratio of defective items per total items, $\frac{k}{N}$ times the number of items selected, n.

Example:

If you have a box of 10 radios containing 2 defective radios and 8 good radios, what is the probability all 4 of the radios you select will be good? If X is the number of good radios in our selection of 4, then X follows a hypergeometric distribution with $n = 4$, $N = 10$, and $k = 8$. Then

$$P(X = 4) = \frac{\binom{8}{4}\binom{10-8}{4-4}}{\binom{10}{4}} = 0.333.$$

1.7 Continuous Probability Distributions

1.7.1 Continuous Uniform Distribution

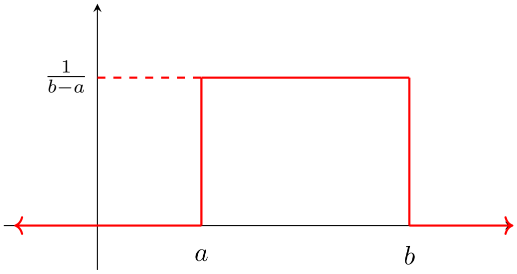
A continuous random variable that is uniformly distributed on the interval $[a,b]$ has the probability density function

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}.$$

The expectation and variance are

$$E(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{2}.$$

The graph of the probability density function looks like:



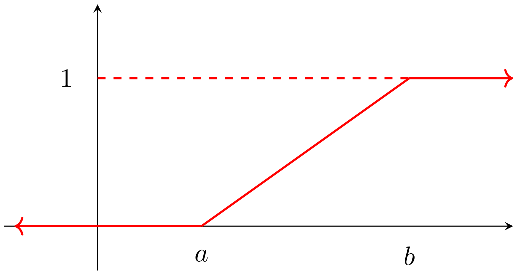
The cumulative distribution function is then

$$F(x) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x \ge b \end{cases}.$$

since

$$\int_{-\infty}^{x} \frac{1}{b-a}\, dx' = \frac{1}{b-a}\int_{a}^{x} dx' = \frac{x-a}{b-a}.$$

The graph of the cumulative distribution function looks like:



Example:

If a real number between 1 and 3 is selected at random, what is the probability that it is between 1.5 and 2?

If the random variable X is the value of the number selected, then X follows a continuous uniform distribution, and its cumulative distribution function is

$$F(x) = \begin{cases} 0, & x \le 1 \\ \frac{x-1}{2}, & 1 < x < 3 \\ 1, & x \ge 3 \end{cases}.$$

Then

$$\begin{aligned} P(1.5 \le X \le 2) &= F(2) - F(1.5) \\ &= \frac{1}{2} - \frac{1}{4} \\ &= \frac{1}{4}. \end{aligned}$$

1.7.2 Normal Distribution

For a normally distributed random variable X with mean μ and variance σ^2, the probability density function is

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty.$$

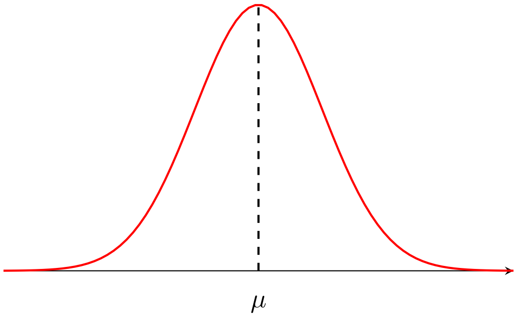
The expectation is

$$E(X) = \mu,$$

and the variance is

$$\mathrm{Var}(X) = \sigma^2.$$

A normal distribution with mean 0 and variance 1 is called the **standard normal distribution**. The standard normal distribution is simply the special case where it is centered on the y-axis.



A normal distribution has the properties

- It is a bell-shaped curve

- Its mode occurs at $x = \mu$. That is, it is centered on its mean value. The mode is the value that occurs most often, so it corresponds to the peak of the bell curve.
- It is symmetric about $x = \mu$
- It is always positive
- The total area under the curve is 1

The mean μ is often called the **location parameter** since it determines where along the x-axis, the bell curve is located, and σ^2 is often called the **scale parameter** since it determines the shape of the curve–whether it is tall and skinny or short and fat. If you increase σ, the bell curve flattens out.

Like the other distributions, probabilities of a normally distributed random variable are calculated as areas under the density curve. For example, $P(x_1 < X < x_2)$ is calculated as the area under the bell curve from x_1 to x_2.

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx.$$

However, the normal density function is hard to integrate, so we a table of values is typically used instead.

The table gives the values for the *standard* normal distribution, so before we can use it, we have to convert whatever our normal distribution is into the standard normal distribution with $\mu = 0$ and $\sigma^2 = 1$. We have to be able to transform any normal distribution into the standard normal distribution. Given a normal distribution $X \sim N(\mu, \sigma^2)$, we can transform it into the standard normal distribution $Z \sim N(0, 1)$ using the transformation

$$Z = \frac{X - \mu}{\sigma}, \qquad X = \sigma Z + \mu.$$

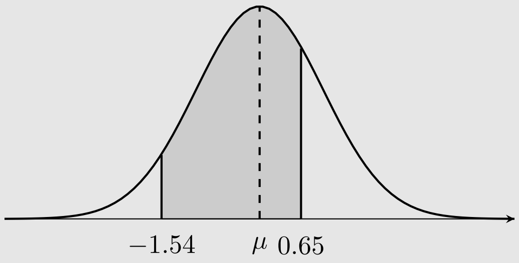
We reserve Z for the standard normal random variable.

Remember the following useful formulas for calculating probabilities using the table:

$$\begin{aligned}
P(Z < a) &= \Phi(a) \\
P(Z > a) &= 1 - \Phi(a) \\
P(a < Z < b) &= \Phi(b) - \Phi(a) \\
\Phi(-a) &= 1 - \Phi(a) \\
P(Z > -a) &= 1 - \Phi(-a) = \Phi(a) \\
P(-a < Z < b) &= \Phi(b) - \Phi(-a) \\
&= \Phi(b) - 1 + \Phi(a).
\end{aligned}$$

Example:

Calculate $P(-1.54 < Z < 0.65)$.



We start by expressing it in terms of $\Phi(z)$.

$$\begin{aligned}
P(-1.54 < Z < 0.65) &= \Phi(0.65) - \Phi(-1.54) \\
&= \Phi(0.65) - (1 - \Phi(1.54)) \\
&= \Phi(0.65) + \Phi(1.54) - 1.
\end{aligned}$$

We can now look up the two values in the table. We find that

$$\begin{aligned}
P(-1.54 < Z < 0.65) &= \Phi(0.65) + \Phi(1.54) - 1 \\
&= 0.74215 + 0.93822 - 1 \\
&= 0.68037.
\end{aligned}$$

Example:

If X is a normally distributed random variable with mean 3 and variance 16, what is the probability that X is between 3 and 5? We want to find $P(3 < X < 5)$, but first we have to transform it to the standard normal distribution.

$$\begin{aligned}
P(3 < X < 5) &= P\left(\frac{3-3}{4} < Z < \frac{5-3}{4}\right) \\
&= P(0 < Z < 0.5) \\
&= \Phi(0.5) - \Phi(0) \\
&= 0.69146 - 0.50000 \\
&= 0.19146.
\end{aligned}$$

1.7.3 Binomial Approximation

For sufficiently large n a binomial distribution approaches a normal distribution, so we can use a normal distribution to approximate a binomial distribution. Recall that if the random variable X follows a binomial distribution, then $b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$. The normal approximation for the binomial distribution is given by

$$P(X \leq x) \approx P\left(Z \leq \frac{x - np}{\sqrt{np(1-p)}}\right),$$

and

$$P(x_1 \leq X \leq x_2) \approx P\left(\frac{x_1 - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{x_2 - np}{\sqrt{np(1-p)}}\right).$$

When np and $np(1-p)$ are greater than 5, then it makes sense to use the normal approximation for a binomial distribution.

Example:

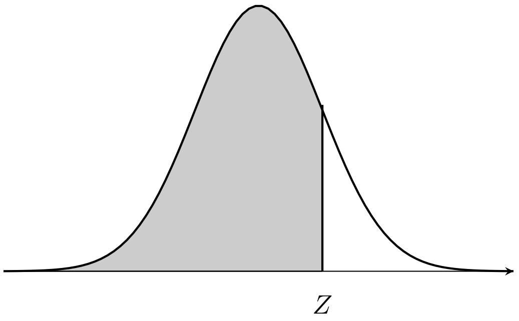
If X is a random variable that follows a binomial distribution with $n = 100$ and $p = 0.4$, find the normal approximation. Find $P(X < 30)$. Notice that X is strictly less than 30, but we need it to be less than or equal to use the formula given above, so we write $P(X < 30) = P(X \leq 29)$.

Plugging $n = 100$ and $p = 0.4$ into the formula given above gives us

$$\begin{aligned} P(X \leq 29) &\approx P\left(Z \leq \frac{29 - (100)(0.4)}{\sqrt{(100)(0.4)(1-0.4)}}\right) \\ &\approx P(Z \leq -2.25) \end{aligned}$$

Now it's a standard normal distribution problem, and we can use the table.

$$\begin{aligned} P(Z \leq -2.25) &= \Phi(-2.25) \\ &= 1 - \Phi(2.25) \\ &= 1 - 0.98778 \\ &= 0.01222. \end{aligned}$$



Values in the table are the values of $\Phi(z)$, which is the area under the standard normal distribution to the left of Z. The function $\Phi(z)$ is the cumulative distribution function for the standard normal distribution.

To find $\Phi(z)$ locate z by adding the values of the left column and the top row. The value of $\Phi(z)$ is given at the intersection of the row and the column. For example, to find $\Phi(0.12)$ find 0.1 in the far left column and 0.02 in the top row (since $0.12 = 0.1 + 0.02$), then $\Phi(0.12) = 0.54776$ is found at the intersection of that row and column.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976

1.8 Summary

When solving problems on an exam, one of the first things we often have to do is identify the random variable as a discrete random variable or a continuous random variable. Look at the CDF. The CDF for a discrete random variable contains only constants, whereas the CDF for a continuous random variable often contains a function of x. Keywords to note include the probability mass function or PMF, which is in the form of a table, and is used only with discrete random variables. The PDF, or probability density function, is a piecewise function $f(x)$ and is used only with continuous random variables.

1.8.1 Combinatorial Analysis

Know the basic rule of counting, and how it is involved in more complex problems.

1.8.2 Conditional Probability

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

Remember that $A \cap B$ means *and* and $A \cup B$ means *or*.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \qquad (1.1)$$

1.8.3 Independence

If A and B are independent events, then

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

We can plug this into Eq. (1.1) if the events are independent. If A and B are independent, we also have that

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

1.8.4 Discrete Random Variables

Need to be able to calculate

- PMF, probability mass function
- CDF, cumulative distribution function. Need to know how to go back and forth between $p(x)$ (the PMF) and $F(x)$ the (CDF).
- $E(X)$, expectation
- $\mathrm{Var}(X)$, variance

Transformation of a Random Variable

Need to be able to transform from $p(x)$ to $p(y)$ given $Y = g(x)$.

Binomial Distribution

$$\begin{array}{rcl} X & \sim & Bin(n,p) \\ P(X=k) & = & \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n \\ E(X) & = & np \\ \mathrm{Var}(X) & = & np(1-p). \end{array}$$

Remember that k starts at 0.

Poisson Distribution

$$\begin{array}{rcl} X & \sim & Poi(\lambda t) \\ P(X=k) & = & e^{-\lambda t}\frac{(\lambda t)^h}{k!}, \quad k = 0, 1, \ldots, \infty \\ E(X) & = & \lambda t \\ \mathrm{Var}(X) & = & \lambda t. \end{array}$$

1.8.5 Continuous Random Variable

First of all, check that a given function really is a probability mass function by checking that it is always nonnegative and that the integral from negative infinity to infinity is exactly 1. Need to be able to solve for a constant by using this integral.

Uniform Distribution

For both discrete and continuous random variables, the PDF of a uniform distribution is a horizontal line.

Normal Distribution

$$\begin{array}{rcl} X & \sim & N(\mu, \sigma^2) \\ z & = & \frac{X-\mu}{\sigma} \sim N(0,1) \end{array}$$

Once you have a *standard* normal distribution in Z, use the normal table to find the probabilities.

$$\begin{array}{rcl} P(a < Z < b) & = & \Phi(b) - \Phi(a) \\ \Phi(-a) & = & 1 - \Phi(a). \end{array}$$

Binomial Approximation

If $X \sim Bin(n,p)$, then for large n,

$$Z = \frac{X - np}{\sqrt{np(1-p)}}.$$

Index

Bayes' theorem, 11
Bernoullli Distribution, 24
Binomial approximation, 29
Binomial Distribution, 24

Combinations, 4
Complement, 5
Conditional probability, 10
Continuous random variables, 19
Cumulative distribution function, 16, 19

Discrete random variables, 15
Disjoint events, 13

Empty set, 5
Event, 5
Expectation, 16, 22

Geometric Distribution, 26

Hypergeometric Distribution, 27

Independent events, 13

Multiplication principle, 3
Multiplication rule, 10

Negative Binomial Distribution, 26
Normal distribution, 28
Null set, 5

Permutations, 3
Poisson Distribution, 25
Probability density function, 19
Probability distribution, 15
Probability tree, 12

Random variables, 15, 19
Rule of product, 3

Sample space, 5
Set notation, 5
Standard deviation, 17
Standard normal distribution, 28

Tree diagram, 12

Uniform Distribution, 24
Uniform distribution, 28
Union, 5

Variance, 17, 22
Venn diagram, 5