Probability and Statistics Short Course

Brad Jacobs

October 11, 2017

Table of contents

Introduction

Counting (Combinatorics)

Probability Basics

Conditional Probability

Independence

Bayes' Formula

Random Variables

Common Distributions

Introduction

What is Probability?
What is Statistics?
What is the difference?

Probability refers to the study of patterns in a random process.

When we solve problems in probability we assume that all basic features of the random process are **known**, and our goal is to discover other, deeper features.

For example:

If I have a coin which is **known** to land on heads exactly half of the time, it is a problem in **probability** to determine how often the coin will never land on heads over ten consecutive flips.

Statistics refers to the study of random process where some basic features of the random process are **unknown**, and our goal is to **infer from observations** basic, hidden features of the random process.

For example:

It is a problem in **statistics** to determine, when presented with a coin which has landed tails ten consecutive times, whether one should continue to believe it fair.

In this course:

Day 1 (Tuesday): Basics of Probability.
Day 2 (Wednesday): Basics of Statistics.

Counting (Combinatorics)

The basic problem solving skill you need to solve problems in probability is **counting** (no, really).

For example:

- How many ways are there to arrange four letters of the alphabet?
- How many ways are there to arrange four *different* letters of the alphabet.
- How many ways are there to arrange 25 math books on a bookshelf.
- How many five card hands are full houses.
- How many five card hands have three of a kind.
- How many five card hands have three of a kind and are not also full houses or four of a kind.

Basic Counting Principle

If a task can be accomplished as a series steps, then the number of outcomes of the task is the **product** of the number of outcomes of each individual step.

How many ways are there to arrange four letters of the alphabet?

Think: How can we accomplish this task as a step by step process.

How many ways are there to arrange four letters of the alphabet?

Think: How can we accomplish this task as a step by step process.

Pick the first letter, write it down.
$\Rightarrow$ Pick the second letter, write it down.
$\Rightarrow$ Pick the third letter, write it down.
$\Rightarrow$ Pick the fourth letter, write it down.

$$26 \times 26 \times 26 \times 26 = 456976$$

How would this change if we could **not** re-use a letter?

The previous example is a common situation: we are pulling from a pool of objects, and we **cannot** re-use an object once selected.

This is called **selection without replacement**.

The number of **ordered** selections **of k objects** without replacement **from a population of n objects** is called the **number of permutations of k objects taken from n**.

$$P(n,k) = \underbrace{n \times (n-1) \times (n-2) \times \ldots \times (n-k+1)}_{\text{k total factors}}$$

You have 25 math and stats books on a bookshelf. How many ways are there to arrange these books in any order?

$$25 \times 24 \times 23 \times \ldots \times 2 \times 1 = 15511210043330985984000000$$

What if we have a procedure in which the order of choices does not matter?

How many 5 card hands are possible when drawing from a standard 52 card deck?

Notice that the **order in which we draw cards is not important here**.

Think: To choose an ordered list of five cards I can first chose the five cards I want to use and then choose a way to order them.

$$\begin{aligned}&\#\text{ of ordered hands}\\&\qquad = \#\text{ of unordered hands} \times \#\text{ of ways to order five cards}\end{aligned}$$

$$52 \times 51 \times 50 \times 49 \times 48 = \# \text{ of unordered hands} \times 5 \times 4 \times 3 \times 2 \times 1$$

$$\# \text{ of unordered hands} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1}$$

The number of unordered selections **of k objects** without replacement **from a population of n objects** is called the **number of combinations of k objects taken from n**.

$$C(n, k) = \frac{P(n, k)}{P(k, k)}$$

Another common notation:

$$\binom{n}{k} = \frac{P(n, k)}{P(k, k)}$$

$$\begin{aligned}&\#\text{ of unordered collections something}\\&\qquad=\frac{\#\text{ of ordered sequences of the thing}}{\#\text{ of ways to order a single sequence}}\end{aligned}$$

A **full house** is a hand of five cards that has a **pair** of the same value, and a **three of a kind** of the same value. How many hands of five cards are full houses?

A **full house** is a hand of five cards that has a **pair** of the same value, and a **three of a kind** of the same value. How many hands of five cards are full houses?

Choose a value for the pair
$\Rightarrow$ Choose two cards of that value
$\Rightarrow$ Chose a value for the three of a kind
$\Rightarrow$ Choose three cards of that value

A **full house** is a hand of five cards that has a **pair** of the same value, and a **three of a kind** of the same value. How many hands of five cards are full houses?

Choose a value for the pair
$\Rightarrow$ Choose two cards of that value
$\Rightarrow$ Chose a value for the three of a kind
$\Rightarrow$ Choose three cards of that value

$$13 \times C(4,2) \times 12 \times C(4,3) = 13 \times 6 \times 12 \times 4 = 3744$$

How many hands are there that contain a three-of-kind that are not full houses and also not four of a kind?

Probability Basics

An **outcome** is a single thing that can happen.

An **event** is a collection of things that can happen, usually given by a short description.

For example:

When thinking about poker hands, an **outcome** is a single hand (any unordered collection of five cards).

The collection of all full-houses, three of a kinds, etc are **events**.

The **probability** of an event is:

$$P(\text{event}) = \frac{\#\text{ of ways event can happen}}{\text{total }\#\text{ of things that can happen}}$$

So to compute basic probabilities, we use our knowledge from counting.

$$P(\text{full-house}) = \frac{\#\text{ of full houses}}{\text{total }\#\text{ of hands}}$$

$$\# \text{ of full houses} = 13 \times 6 \times 12 \times 4 = 3744$$

$$\text{total } \# \text{ of hands} = \binom{52}{5} = 2598960$$

$$P(\text{full-house}) = \frac{3744}{2598960} = 0.0014$$

When rolling two fair, six-sided dice, what is the probability that they sum to 7?

$$P(\text{Die One} + \text{Die Two} = 7) = ?$$

When rolling two fair, six-sided dice, what is the probability that they sum to 7?

$$
\begin{aligned}
P(\text{Die One} + \text{Die Two} = 7) &= \\
&= \frac{6}{36} \\
&= 0.1667
\end{aligned}
$$

What is the probability of drawing a hand containing a three of a kind **that is not a full house or four of a kind**?

Conditional Probability

Suppose we know that one event C has already happened or will happen (the condition), and we want to know the probability of different event B.

Then the **conditional probability of A given B** is defined by

$$P(A \mid B) = \frac{\text{\# of ways A and B \textbf{both} happen}}{\text{\# of ways B can happen}}$$

When rolling two fair, six-sided dice, what is the conditional probability that Die One shows a 2, given that the sum of Die One and Die Two is less than or equal to 5?

$$P(\text{Die One} = 2 \mid \text{Die One} + \text{Die Two} \leq 5) =$$
$$= \frac{\#\text{ of ways of Die One is 2 while total for both is 5 or less}}{\#\text{ of ways of Die One and Die Two total 5 or less}}$$

#of ways of Die One is 2 while total for both is 5 or less $= 3$

# of ways of Die One and Die Two totaling 5 or less $= 10$

$$
\begin{aligned}
P(\text{Die One} = 2 \mid \text{Die One} + \text{Die Two} \leq 5) = \\
&= \frac{3}{10} \\
&= 0.3
\end{aligned}
$$

There is another way to look at conditional probabilities

$$
\begin{aligned}
&P(A \mid B) \\
&\quad = \frac{\#\text{ of ways A and B } \textbf{both} \text{ happen}}{\#\text{ of ways B can happen}} \\
&\quad = \frac{\frac{\#\text{ of ways A and B } \textbf{both} \text{ happen}}{\text{Total }\#\text{ of things that can happen}}}{\frac{\#\text{ of ways B can happen}}{\text{Total }\#\text{ of things that can happen}}} \\
&\quad = \frac{P(A \text{ and } B)}{P(B)}
\end{aligned}
$$

$$
\begin{aligned}
P(\text{Die One} = 2 \mid \text{Die One} + \text{Die Two} \leq 5) = & \\
&= \frac{3/36}{10/36} \\
&= \frac{0.0833}{0.2778} \\
&= 0.3
\end{aligned}
$$

What is the conditional probability that you draw a full house, given that you know you have at least a three of a kind

$$
\begin{aligned}
P(\text{Drawing a straight flush} \mid \text{Drawing a straight}) &= \\
&= \frac{\#\text{ of ways of drawing straight flush}}{\#\text{ of ways of drawing a straight}} \\
&= \frac{10 \times 4}{10 \times 4^5} \\
&= \frac{40}{10240} \\
&= 0.0039
\end{aligned}
$$

Important: Any hand straight flush **automatically** contains a straight, and we are including a royal flush as a type of straight flush!

What is the conditional probability that you draw a full house, given that you have drawn a four of a kind.
What is the conditional probability that you draw a four of a kind, given that you know you already have a pair?

Independence

Two events A and B are called **independent** when

$$P(A \mid B) = P(A)$$

This means that **knowledge that B has or will occur does not change our knowledge about whether A will occur**.

Remember that the definition of conditional probability is

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Combining this with our definition of independence (so, below, A and B are independent

$$P(A) = P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Or, rearranging things

$$P(A \text{ and } B) = P(A)P(B)$$

This equation is sometimes used as the **definition** of independence, though I think it is less convincing.

Common applications of independence generally go like this:

- We deduce from context that some event bears no influence on another.
- We conclude that the two events are independent.
- We use either of the two equations defining independence in our calculations.

You roll a six sided die six times, what is the probability that you roll **all the possible numbers, in decreasing order**?

Let's call the values of the rolls R_1, R_2, R_3, R_4, R_5 and R_6. Then we are looking for

$$P(R_1 = 6 \text{ and } R_2 = 5 \text{ and } R_3 = 4 \text{ and } R_4 = 3 \text{ and } R_5 = 2 \text{ and } R_6 = 1)$$

Since each individual roll of the die does not influence the others, all size rolls are independent. This means we can break up and multiply

$$P(R_1 = 6) \times P(R_2 = 5) \times P(R_3 = 4) \times P(R_4 = 3) \times P(R_5 = 2) \times P(R_6 = 1)$$

Since each individual roll of the die has **six possible outcomes** and **only one of them is the number we are looking for** we get

$$P(\cdots) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{46656}$$

Suppose you have a bucket with 5 red and 5 yellow balls in it, which you draw in sequence, without replacement. Are the events "You draw a red ball first" and "You draw a yellow ball second" independent?

Suppose that it rains with probability 80% each day in seattle, and a winter month has four work weeks. What is the probability that there is a work week in which it rains **every day**? What assumptions do you have to make to solve this problem? Are they reasonable?

Bayes' Formula

Remember our definition of conditional probability:

$$\begin{aligned} P(A \text{ and } B) &= P(A \mid B)P(B) \\ P(B \text{ and } A) &= P(B \mid A)P(A) \end{aligned}$$

Setting these equal to one another leads us to **Bayes' Formula**

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

This is probably the most important simple formula in both probability and statistics

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Let's study a classic thought experiment: the disease screening problem.

Suppose we have developed a test for a certain disease:

- Only 1 % of people have the disease.
- If a person has the disease, the test will be positive 99.9 % of the time.
- If a person does not have the disease, the test will be negative 98 % of the time.

You get tested for the disease, and the test is positive. **What is the probability that you actually have the disease**?

We are after the following conditional probability

$$P(\text{Have Disease} \mid \text{Test is Positive})$$

And we **know**

$$P(\text{Have Disease}) = 0.01$$

$$P(\text{Test is Positive} \mid \text{Have Disease}) = 0.999$$

$$P(\text{Test is Positive} \mid \text{Do Not Have Disease}) = 0.02$$

Bayes formula says:

$$P(\text{Have Disease} \mid \text{Test is Positive}) \\ = \frac{P(\text{Test is Positive} \mid \text{Have Disease})P(\text{Have Disease})}{P(\text{Test is Positive})}$$

We know all the things appearing in this formula except $P(\text{Test is Positive})$.

We can calculate the last piece by breaking things down

$$\begin{aligned} P(\text{Test is Positive}) \\ &= P(\text{Test is Positive and Have Disease}) \\ &\quad + P(\text{Test is Positive and Don't Have Disease}) \end{aligned}$$

$$
\begin{aligned}
&P(\text{Test is Positive and Have Disease}) \\
&\quad = P(\text{Test is Positive} \mid \text{Have Disease})P(\text{Have Disease}) \\
&\quad = 0.999 \times 0.01
\end{aligned}
$$

$$
\begin{aligned}
&P(\text{Test is Positive and Don't Have Disease})\\
&\quad= P(\text{Test is Positive} \mid \text{Don't Have Disease})P(\text{Don't Have Disease})\\
&\quad= 0.02 \times 0.99
\end{aligned}
$$

We have all the pieces, so let's put them together

$$
\begin{aligned}
P(\text{Have Disease} \mid \text{Test is Positive}) = & \\
&= \frac{P(\text{Test is Positive} \mid \text{Have Disease})P(\text{Have Disease})}{P(\text{Test is Positive})} \\
&= \frac{0.999 \times 0.01}{0.999 \times 0.01 + 0.02 \times 0.99} \\
&= 0.34
\end{aligned}
$$

The probability we have the disease is only 34%, **even knowing we have received a positive test**.

This kind of result is unintuitive to almost all humans, a mental bias called the **base rate fallacy**.

Pretty much **everyone's** intuition says that it should be much more likely that the person does have the disease after a test comes back positive. Pretty much **everyone** undervalues the prior information that

$$P(\text{Have Disease}) = 0.01$$

I takes a lot of evidence to take an unlikely situation and make it likely.

There is some terminology that is often used to understand these relationships:

- $P(\text{Have Disease})$ is called the **prior probability**. It is what we know **before collecting evidence/data**.
- $P(\text{Test is Positive} \mid \text{Have Disease})$ is called the **likelihood**. It is the **strength of the evidence/data we collected**.
- $P(\text{Have Disease} \mid \text{Test is Positive})$ is called the **posterior**. It is what we know, **after collecting evidence/data**.

These ideas form the basis of **Bayesian statistics**.

Suppose you now get a second test, which **also** comes out **positive**. What is the posterior probability that you actually have the disease?

Suppose you get a second test, which comes back **negative**. What is the posterior probability that you actually have the disease.

Random Variables

A random variable X is an object that can be used to generate numbers, in a way that valid probabilistic statements about the generated numbers can be made.

For example:

$$P(X > 0) = 0.25$$

$$P(-1 < X < 1) = 0.25$$

$$P(X < -1) = 0.02$$

$$P(X > 1 \mid X > 0) = 0.5$$

Are all probabilistic statements about an unknown random variable X.

Random variables can be used to model real life events or measurements when there is some variation in their outcomes that we cannot account for

- The number of heads seen in ten flips of a quarter.
- The number of heads seen in ten flips of a dime.
- The number of Buses that arrive late to a particular bus stop in a single day.
- The number of times my cat asks for food between 5 and 6 pm (when she is always fed) in a given day.
- The temperature on a mid-summers day in San Francisco.
- The rainfall in a mid-winters day in Seattle.

We naturally have a feeling that there is something **the same** about these two situations

- The number of heads seen in ten flips of a quarter.
- The number of heads seen in ten flips of a dime.

What is it?

We expect that the probabilities

$$P(\text{We get 5 heads in 10 flips of a quarter})$$

$$P(\text{We get 5 heads in 10 flips of a dime})$$

are **equal**.

So are

$$P(\text{We get 2 heads in 10 flips of a quarter})$$

$$P(\text{We get 2 heads in 10 flips of a dime})$$

and so on...

The **sameness** that we sense between

- The number of heads seen in ten flips of a quarter.
- The number of heads seen in ten flips of a dime.

is that the probabilities of all events like

$$P(\text{We get } N \text{ heads in 10 flips of a quarter})$$

$$P(\text{We get } N \text{ heads in 10 flips of a dime})$$

are all equal.

We summarize this by saying that the random variables

- The number of heads seen in ten flips of a quarter.
- The number of heads seen in ten flips of a dime.

have the same distribution.

The **distribution** of a random variable is the pattern of all probabilities we assign to all outcomes of the random variable.

So two random variables have the same distribution if **they assign the same probabilities to all of the possible outcomes**.

In this case, we use the shorthand **equally distributed**.

Common Distributions

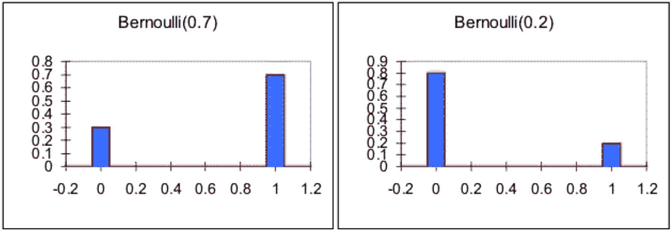
Some distributions are so common, they have been named and entered our shared statistical conciousness.

Bernoulli Distribution

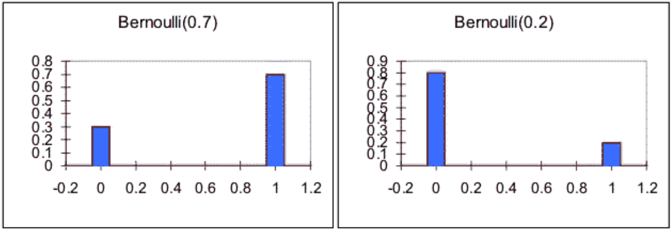
A single coin flip is an example of a Bernoulli Distributed random variable.

The **Bernoulli distribution** describes any random event with only two possible outcomes.

We can draw a picture of a distribution by letting the height of bars represent the probabilities of certain outcomes occurring:

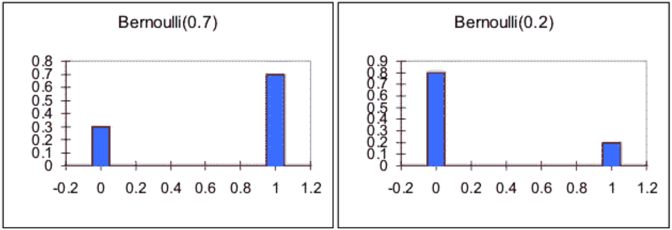


This picture is called the **probability mass function** of the distribution.



Notice how the first picture represents the occurrence of a **common event** and the second a rare event.

Changing the probability that the event occurs changes the shape of the probability mass function. This is called **varying a parameter**.



Draw pictures of the probability mass functions of the following Bernoulli distributions:

- Flipping a fair coin.
- Rolling a six on a six sided die.
- Rolling a twenty on a twenty sided die.
- Rolling greater than a ten on a twenty sided die.

Are any of these equally distributed?

Binomial Distribution

The two familiar random variables

- The number of heads seen in ten flips of a quarter.
- The number of heads seen in ten flips of a dime.

Have a **binomial distribution**.

$$
\begin{aligned}
&P(\text{We get 2 heads in 10 flips of a quarter}) \\
&= \binom{10}{2} \times \left(\frac{1}{2}\right)^{10} \\
&= 0.044
\end{aligned}
$$

The **binomial distribution** describes the number of events that happen in a fixed number of **attempts** when the events **individually happen with the same probability**.

When we are flipping a fair coin, the heads happen with probability $\frac{1}{2}$, and

$$P(\text{We get k heads in n flips of a coin})$$
$$= \binom{n}{k} \times \left(\frac{1}{2}\right)^n$$

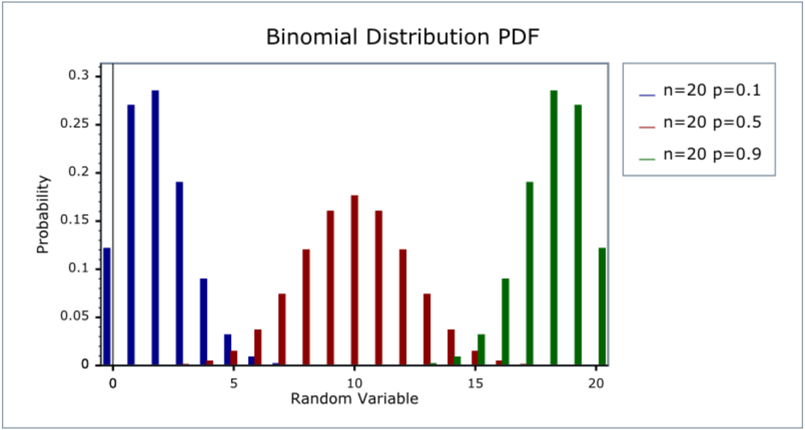
If the coin in **unfair**, so that the probability of an individual head is p, then

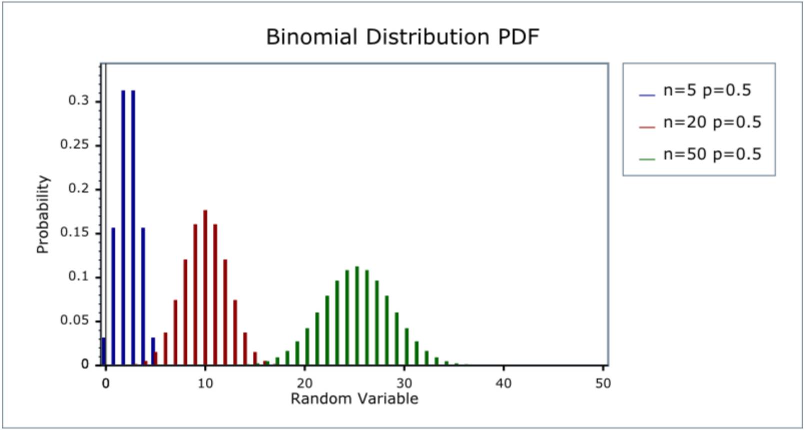
$$P(\text{We get k heads in n flips of an unfair coin})$$
$$= \binom{n}{k} \times p^k \times (1-p)^{n-k}$$

The binomial distribution has **two parameters**:

- The number of attempts, usually called n.
- The probability the event occurs in a single attempt, usually called p.

Changing either n or p changes the shape of the binomial probability mass function.




Binomial Distribution PDF
n=5 p=0.5
n=20 p=0.5
n=50 p=0.5
Probability
0.3
0.25
0.2
0.15
0.1
0.05
0
0
10
20
30
40
50
Random Variable

A critical hit is a roll of 20 on a 20 sided die. In a session of Dungeons and Dragons, you roll the die 20 times. What is the probability that you roll **at least two** critical hits.

A saving throw is a roll of at least 15 on a 20 sided die. In a session of Dungeons and Dragons you roll the die to attempt a saving throw 10 times. What is the probability you fail all of your saving throws?

Poisson Distribution

- The number of buses that arrive late to a particular bus stop in a single day.
- The number of times my cat asks for food between 5 and 6 pm (when she is always fed) in a given day.

We see a similarity: they are both about the number of times an event happens **in a given span of time (or space)**.

If we assume that the buses arrive at a fixed rate (but possibly unknown), and the cat meows at a fixed rate, then these are both examples of the **Poisson Distribution**.

$$P(\text{Cat meows k times in one hour}) = e^{-\lambda}\frac{\lambda^k}{k!}$$

The λ above is the **rate the event occurs**.

Suppose we observe the cat meow 5 times in ten minutes. What is the probability that the cat will not meow at all in the next five minutes?

The rate the cat meows is:

$$\lambda = \frac{5 \text{ meows}}{10 \text{ minutes}} = 5 \frac{\text{meows}}{10 \text{ minutes}}$$

So using the Poisson equation

$$P(\text{Cat meows zero times in ten minutes}) = e^{-5}\frac{5^0}{0!} = 0.007$$

What is the probability the cat meows zero times in the next hour?

$$\lambda = \frac{5 \text{ meows}}{10 \text{ minutes}} = 30 \frac{\text{meows}}{60 \text{ minutes}} = 30 \frac{\text{meows}}{\text{hour}}$$

$$P(\text{Cat meows zero times in the next hour}) = e^{-30} \frac{5^0}{0!} = 9.3 \times 10^{-14}$$

...it's basically impossible.

In the same cat problem setup, what is the probability the cat meows at least two times in the next ten minutes?

Overview

The general strategy for solving problems using probability distributions like is:

- Use the information in the statement of the problem to determine a likely distribution for the quantity of interest.
- Use the information in the problem to determine the values to use for the parameters of this distribution.
- Use the probability function of the distribution to compute the needed probability.

You may have to break the problem down into multiple steps to succeed. Practice and you'll start to see common patterns!

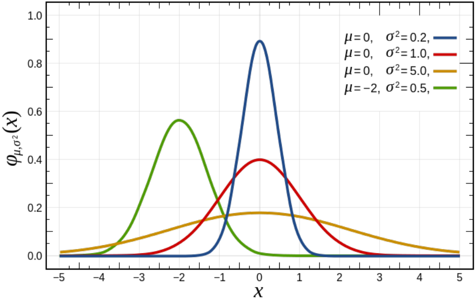
The distributions we discussed are **discrete**: the outcomes can only be individualized numbers (0, 1, 2, 3, ...).

There are also **continuous distributions**.

Normal Distribution

This distribution usually shows up due to the **central limit theorem**.

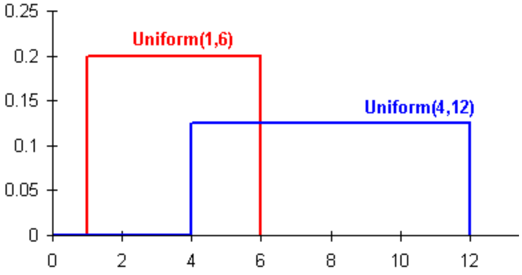
Parameters: The mean μ, and the standard deviation σ.



Uniform Distribution

This distribution shows up when when a random event can take any value in a range, each result being equally likely.

Parameters: The minimum a and the maximum b.



Exponential Distribution

This distribution describes the **time you have to wait before observing an event** when the events happen at a fixed rate.

Parameters: The rate α.