**Continuous vs Discrete**

Gaussian distributions are a type of continuous probability distribution. So before getting into the details about Gaussian distributions, let's review the differences between a continuous and discrete distribution.

As a reminder, a continuous probability distribution is associated with a continuous variable like height, weight, distance, velocity, angle, etc.

You've been working with a uniform continuous probability distribution like the following example.

A screenshot of a cell phone

Description automatically generated

On the other hand, a discrete distribution is associated with variables that can only take on certain values like coin flips, dice rolls or location on a grid. This is an example of a discrete distribution.

A screenshot of a cell phone

Description automatically generated

**Continuous Distributions**

The uniform continuous probability distribution is just one of many continuous probability distributions that exist. Take a look at this Wikipedia page, [**List of Continuous Distributions**](https://en.wikipedia.org/wiki/List_of_probability_distributions#Continuous_distributions), just to get a sense for how many there are.

What do all of these distributions have in common? They represent the probability of events occurring for **continuous variables**. Different applications will use different distributions.

For example, when modeling probabilities of a spinning wheel, you used a uniform distribution. When modeling uncertainty in a sensor measurement, you'll use a Gaussian distribution.

**Gaussian Distribution**

In order to understand uncertainty in self-driving cars, you should at least have a basic knowledge of the Gaussian distribution. The uncertainty in a sensor measurement or the location of a pedestrian, for example, is oftentimes modeled with a Gaussian distribution.

This next part of the lesson will give a broad overview of the Gaussian distribution and where the distribution comes from.

However, this is not a complete statistics course. This will be a brief overview of this particular distribution and assumes you're familiar with terms like mean, standard deviation, population, and sample where:

* **population** refers to the entire set of all data points. Like if you were measuring people's weights, then the population would be all people in the world.
* **sample** refers to a part of the population. In the weights example, you might take a random sample of the population since it would be practically impossible to measure the weights of all humans.
* **mean** is the average value, which in this case would be the average weight of all humans.
* **standard deviation** measures the spread in the data. Does the data tend to hover around the mean, or is the data more spread out?

If you aren't familiar with these terms or need a refresher, here is a resource that you might find helpful: [**sample vs population**](http://stattrek.com/sampling/populations-and-samples.aspx).

**The Gaussian Distribution**

So let's go straight to looking at a Gaussian distribution, and then we'll talk about what it is and how it relates to self-driving cars.

A screenshot of a cell phone

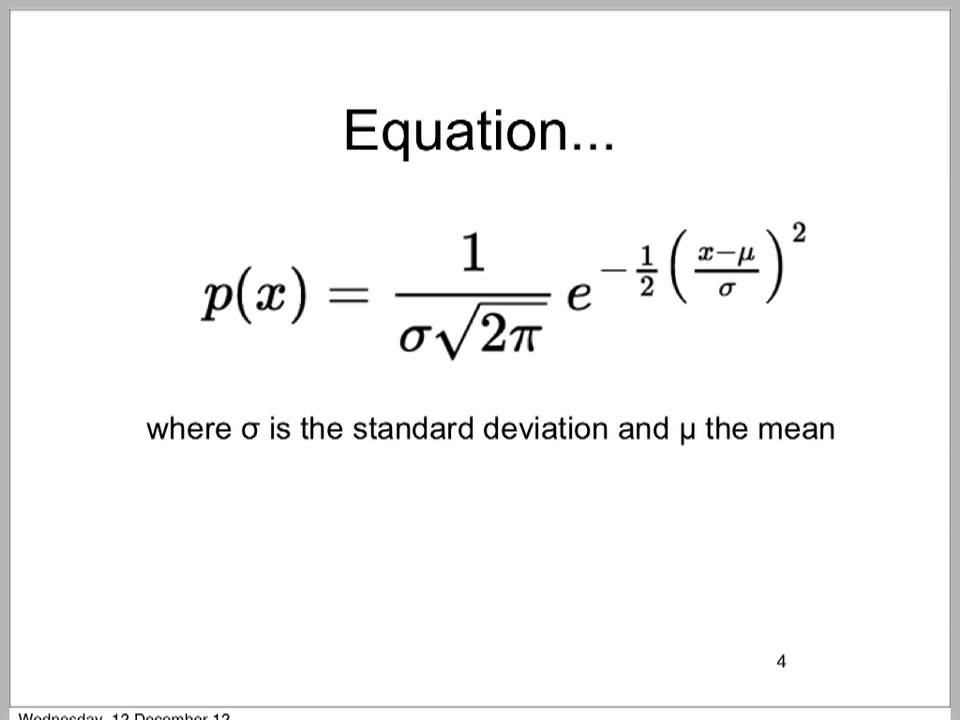
Description automatically generated

You might recognize this shape; it's oftentimes called a bell-shaped curve. Notice that the y-axis says "probability density function" just like the uniform continuous distribution.

The x-axis currently says "x-value"; however, the x-axis could take on any continuous variable like temperature, height, or velocity.

**Gaussian Equation**

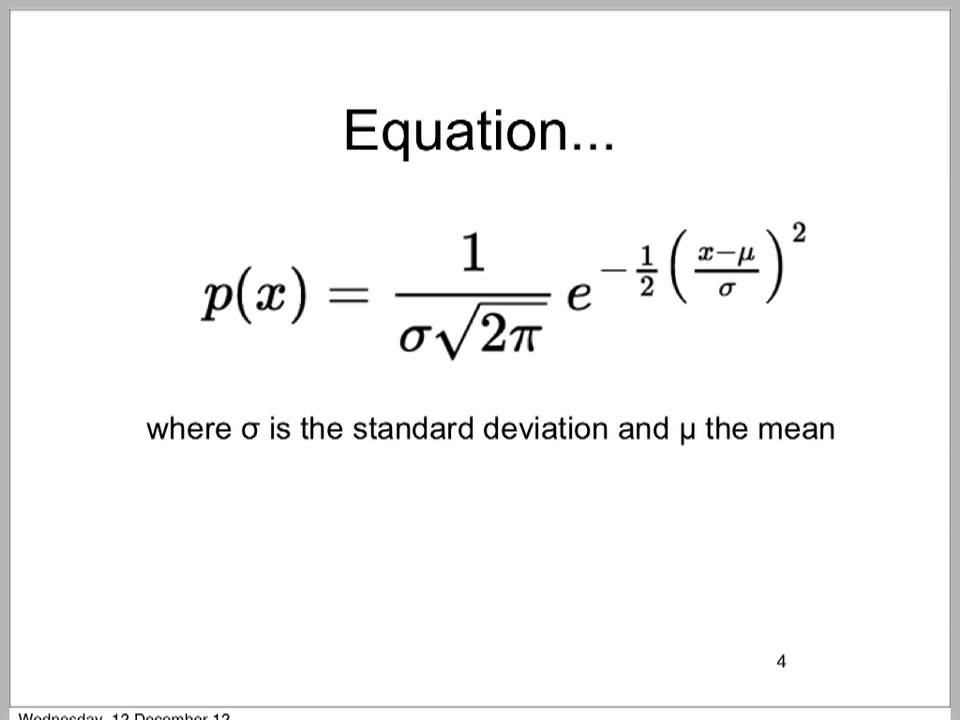
While this graph might look difficult to draw, there is actually an equation that maps the x variable to the y variable. Here is the probability density function for the Gaussian distribution:



This equation might look a bit intimidating, so let's break it down in the next part of the lesson.

# Gaussian Equation

Here again is the Gaussian probability density function.



Don't let this equation intimidate you! This equation only has three input variables:

* \large \mu*μ*
* \large \sigma*σ*
* \large x*x*

The symbol \mu*μ* represents a population mean. The symbol \sigma*σ* is the standard deviation of the distribution, which represents the spread. If you remember from statistics, the mean and standard deviation are constants when dealing with a population. So for a specific population, the only value that varies is x.

# Example

Let's make this more concrete. Say you are looking at the probabilities associated with winter temperatures in San Francisco. Assume that the minimum daily temperature follows a Gaussian distribution. This might or might not actually be true if you were to measure winter temperatures experimentally, but for this example, assume that it is true.

Say that on average, the minimum winter temperature in San Francisco is 50 degrees Fahrenheit. In other words, If you measured the minimum temperature every day of winter over all winters ever, the average value would be 50. Let's say the standard deviation is 10 degrees.

Now, substituting the mean and standard deviation into the Gaussian equation gives:

Just substitute it!

After substituting in the mean and standard deviation into the equation, the equation doesn't look so bad. You could take a range of x values, which in this case represents temperature, and then you could calculate y for each x value.

So, get out your calculator because it's time for a quiz.

# Mean

What exactly does the mean value represent in a Gaussian distribution? The mean value is the center of the bell curve.

Take a look at the visualization below. Staying with the San Francisco temperature example, it shows three distributions with three different means. Everything else stays the same including the standard deviation.

A close up of a map

Description automatically generated

A close up of a map

Description automatically generated

A close up of a mans face

Description automatically generated

You can see that changing the mean, while keeping everything else constant, shifts the curve to the left or to the right.

# Standard Deviation

What effect does the standard deviation have on the Gaussian distribution? Let's repeat the same experiment. Take a look at what happens as the standard deviation changes but everything else remains the same.

A screenshot of a cell phone

Description automatically generated

A close up of a map

Description automatically generated

A close up of a map

Description automatically generated

What do you notice is happening as the standard deviation increases? Answer the quiz:

As the standard deviation increases, uncertainty increases as well. Compare the charts where standard deviation is 5 and standard deviation is 15. When the standard deviation is five, the distribution looks tall and skinny, which implies that the temperature is more likely to be near 50 degrees.

When the standard deviation increases to 15, the distribution gets flat and wide; the probability that the temperature is near 50 goes down while the probability that the temperature is farther to the left or to the right on the x-axis is increasing.

# Area Under the Curve

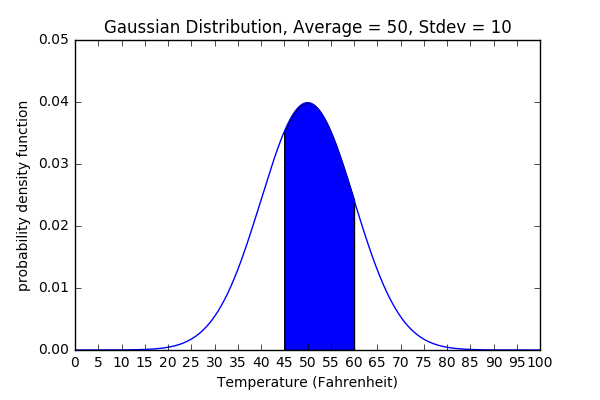
As a reminder, the y-axis of a continuous probability distribution shows the probability density function. The x-axis represents a continuous variable like height, velocity, etc. For reference, here is a Gaussian distribution that we were looking at earlier with mean = 50 and standard deviation = 10.

A close up of a map

Description automatically generated

But the probability density function on the y-axis is not a direct measurement of probability. The **area** under the probability density function is the probability.

Like if you wanted to know the probability that the temperature was between 45 and 60 degrees Fahrenheit, you'd need to calculate the area under the curve from 45 to 60 degrees:



Area Under the Curve from 45 to 60

# Area Under the Gaussian Probability Density Function

Yet, how can you calculate the area under a curve like the Gaussian distribution? If you are familiar with calculus, the area under a curve is called an integral. However, taking the integral of the Gaussian probability density function is not an easy task!

Instead, you could use something called the [**standard normal table**](https://en.wikipedia.org/wiki/Standard_normal_table). This table contains data for calculating the area underneath a Gaussian probability density function. How to use these tables goes beyond the scope of this course; it's generally taught in an introductory statistics course like [**Udacity's Statistics 101**](https://classroom.udacity.com/courses/st101).

Lucky for us, Python has a library that can calculate the area underneath the Gaussian probability density function. In the next section, you'll learn how to use this library.

**Calculating Area Under the Curve Solution**

**from** scipy.stats **import** norm

**def** **gaussian\_probability**(mean, stdev, x\_low, x\_high):

**return** norm(loc = mean, scale = stdev).cdf(x\_high) - norm(loc = mean, scale = stdev).cdf(x\_low)

# Central Limit Theorem

The information from here until the end of the Gaussian Distributions lesson is optional.

This part of the lesson gives insight into where the Gaussian distribution comes from. If you would like to learn about the theory behind Gaussian distributions, then read on! Otherwise, you can move on to the Robot Localization lesson.

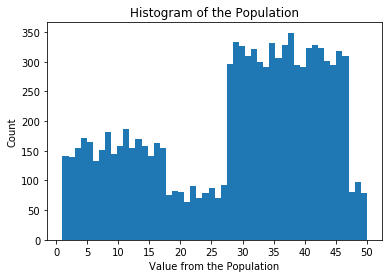
# What is the Central Limit Theorem?

The central limit theorem is quite interesting. It says that if you take large enough samples from a population and then calculate the sample means, these means will be normally distributed. The theorem should hold as long as the sample size is large enough and the variable in question is [**independent**](http://stattrek.com/statistics/dictionary.aspx?definition=independent) and [**random**](http://stattrek.com/statistics/dictionary.aspx?definition=random_variable).

This might all sound theoretical at this point. So in the next part of the lesson, we'll use Python to show you how the theorem works.

# A Population

A population consists of all of the values of a data set. For this lesson, we're going to use a data that looks like this:



Population Distribution

For example, the value 15 shows up in the population about 160 times. The value 50 shows up in the population about 70 times. In total, this population has 10,000 data points.

Consider randomly grabbing 100 data points from this distribution. Call these 100 data points a sample. Then calculate the mean value of the sample. If you did this random sampling over and over again, the mean values would have a Gaussian distribution.

It's amazing that a population distribution that does not look Gaussian at all becomes Gaussian as you take the mean of many samples.

In the following part of the lesson, we'll show you how this works using Python code.

# Conclusion

In this lesson you've learned what the Gaussian distribution is and where the distribution comes from. As you progress in your studies, you'll notice that the Gaussian distribution appears in a few different contexts.

Uncertainty in sensor measurements like radar or lidar are often modeled with Gaussian distributions. Gaussian distributions are also sometimes used to represent uncertainty in the location or velocity of objects surrounding an autonomous vehicle.

In the next lesson, you'll learn about how a robot figures out its location. The probability lessons and programming exercises provide a base for understanding the next part of the lesson.