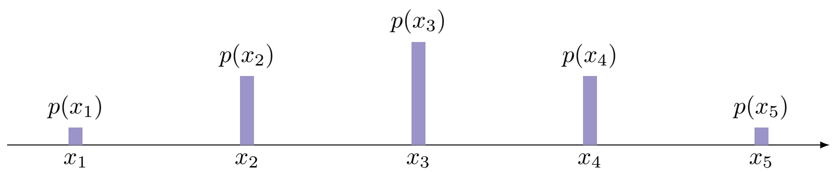
Continuous Random Variables

The next thing we need to familiarize ourselves with is continuous random variable. This describes random variables that can take on an *infinite* or *non-countable* range of values. Some examples are $(-\infty, \infty)$, $[0, 5]$, and $[0, \infty)$.

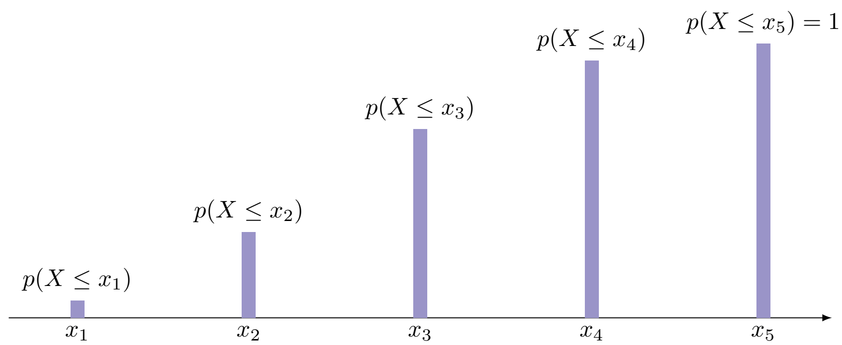
Discrete random variables take on a finite, countable number of values which allowed us to visualize the probability distribution easily using tables and graphs. For continuous random variables we must formulate a new plan to visualize the distribution for an infinite number of values.

3.5.1 Probability Density Functions

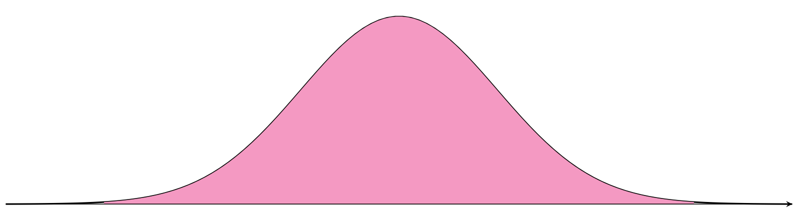
In the last section we discussed how discrete random variables have *probability mass functions* or pmf's for short. Consider the random variable X that can take on x_1 x_2, x_3, x_4 and x_5.



We also discovered that we can find cumulative probability by taking the sum of all probabilities up to some value of x.



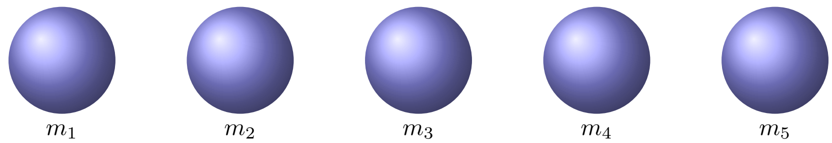
Imagine now the continuous case. Let's call our continuous random variable Y. The distribution for Y can be represented with a smooth curve called the **probability density function**. Probability density functions are analogous to probability mass functions, but for continuous random variables.



The area under the probability density function represents probability and total area must always be one. However we are now talking about probability **density**, not probability **mass**. When someone refers to the **distribution** of a continuous random variable, they are referring to the pdf. Before we go further let's take a deeper dive into these two types of functions.

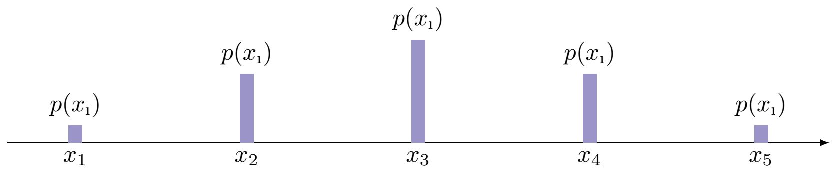
3.5.2 Mass and Density: A Splash of Physics

Let's take a moment to describe the relationship between mass and density. Consider five point masses. m_1, m_2, m_3, m_4, and m_5. If we want the total mass, just take the sum of each mass.



$$\text{Total Mass} = m_1 + m_2 + m_3 + m_4 + m_5$$

This is similar to a probability **mass** function for discrete random variables where probability is analogous to mass. Let's consider a random variable X that can take on five values; x_1, x_2, x_3, x_4, and x_5 with with probabilities $p(x_1)$, $p(x_2)$, $p(x_3)$, $p(x_4)$, and $p(x_5)$.

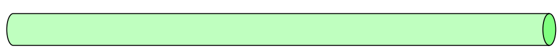


In this case "total probability" will always be 1, as $p(x_1) + p(x_2) + p(x_3) + p(x_4) + p(x_5) = 1$. We could find the probability of multiple 'pieces' by summing them.

Next up let's talk about **density**. Linear density refers to a mass per length ratio.

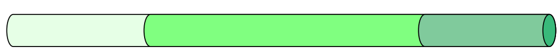
$$\text{Linear Density} = \frac{\text{Mass}}{\text{Length}}$$

Consider a rod with uniform linear density and and a specific length. We can calculate the mass of the rod by taking Mass = Density · Length.



$$\text{Mass} = \text{Density} \cdot \text{Length}$$

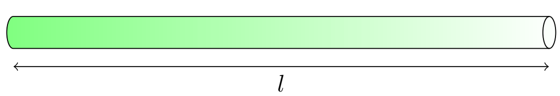
Now suppose we have a rod with five sections of lengths l_1, l_2, and l_3 and densities d_1, d_2, and d_3. We can find the mass of each piece, as well as the total mass.



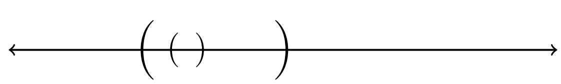
$m_1 = d_1 l_1$ $\quad$ $m_2 = d_2 l_2$ $\quad$ $m_3 = d_3 l_3$

$$\text{Total Mass} = d_1 l_1 + d_2 l_2 + d_3 l_3$$

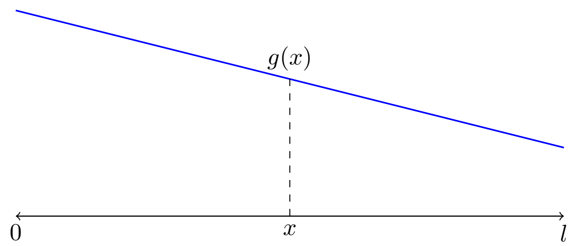
But now let's look at something more challenging. What if the density is unique for every infinite point along a rod of length l?



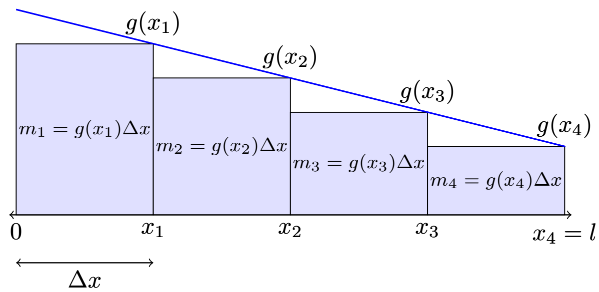
This is more difficult because we need to find the mass for every point along the wire. The problem is there are an infinite number of points. In fact on any interval of wire we choose, there are an infinite number of points with different densities!



Let's let $g(x)$ be the **density** of the rod at a particular value of x. Here x can take on any value from 0 to l.



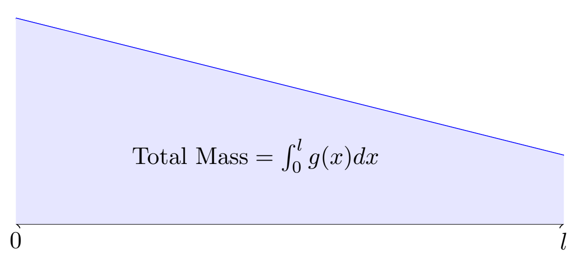
We can approximate the mass of the rod by splitting the wire into several pieces of width Δx and finding the approximate mass of each piece. The mass of each piece is represented by the area of each bar. We are approximating the density for each section of rod with the density at each endpoint.



$$
\begin{aligned}
\text{Total Mass} &\approx m_1 + m_2 + m_3 + m_4 \\
&= g(x_1)\Delta x + g(x_2)\Delta x + \cdots + g(x_4)\Delta x
\end{aligned}
$$

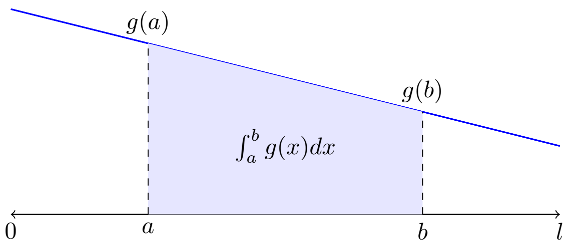
However this is still an approximation, it is not the true mass. The more sections we split our rod into, the better our approximation will be. What if we could split the rod into an infinite amount of tiny pieces? The width of each interval would be infinitesimally small. We call the infinitesimal interval width dx. We are also summing an *infinite* amount of these tiny pieces. We call this infinite sum an **integral**. It represents the **area under a curve**, and in this context the **exact mass of the rod.**

$$
\begin{aligned}
\text{Total Mass} &= g(x_1)dx + g(x_2)dx + \cdots + g(x_n)dx \\
&= \int_0^l g(x)dx
\end{aligned}
$$



Notice for a friendly linear density function like this one, we could find the area under the curve geometrically as well.

We can also find the mass of a certain segment of the rod, let's say from a to b, by taking the integral over the interval. This represents the area under the curve from a to b.



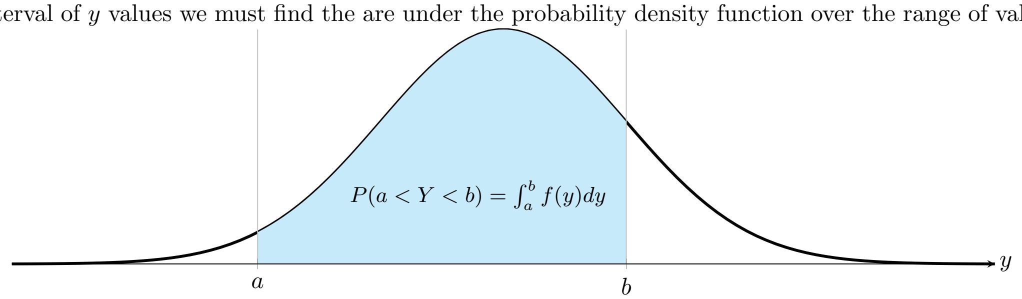
Also notice that the mass at a specific point will always be 0 because no area we will be created if you don't have any specific length of the rod.

> Don't be scared by the fancy new notation. A deeper exploration of integration is left for calculus. For statistics we have **three important take aways:**
>
> **i.** The area under the density curve represents the mass of the rod.
>
> **ii.** Mass for a certain segment of the rod can be computed by finding the area under the density function for the segment.
>
> **iii.** The mass at a specific point will always be 0.

Now let's talk about **probability density functions.**. They work the exact same way as the rod but now instead of mass we have probability, and instead of density we have probability density.

$$\text{Linear Density} = \frac{\text{Mass}}{\text{Length}} \qquad \text{Probability Density} = \frac{\text{Probability}}{y}$$

Where y is the unit of the random variable. For a continuous random variable Y the probability density function, $f(y)$, shows the probability density at a particular point. To find probability over an interval of y values we must find the are under the probability density function over the range of values.



We also know the total area under the curve (total probability) must always be one, and the probability at a specific point will always be zero. Note that this also means

- $P(Y \leq a) = P(Y < a)$
- $P(a \geq Y) = P(a > Y)$

> ***Properties of Probability Density Functions***
>
> 1. $f(y) \geq 0$ for all values of y
> 2. $\int_{-\infty}^{\infty} f(y)\, dy = 1$ (Just a fancy way of saying the entire area under the curve must always be one)

We may also utilize complimentary probability for continuous random variables.

$$P(Y > y) = 1 - P(Y \leq y)$$

3.5.3 Mean for Continuous Random Variables

Recall that for a discrete random variable X we found the expected value using the weighted mean for every value of x

$$E(X) = \sum_{\text{all } x} p(x) \cdot x$$

We have a simple extension using integration for continuous random variables. Suppose Y is a continuous random variable

$$E(X) = \int_{-\infty}^{\infty} f(y) \cdot y \, dy$$

We are more interested in the theoretical meaning of expected value. It represents the exact same thing as with discrete case; the average value of Y.

3.5.4 Variance for Continuous Random Variabes

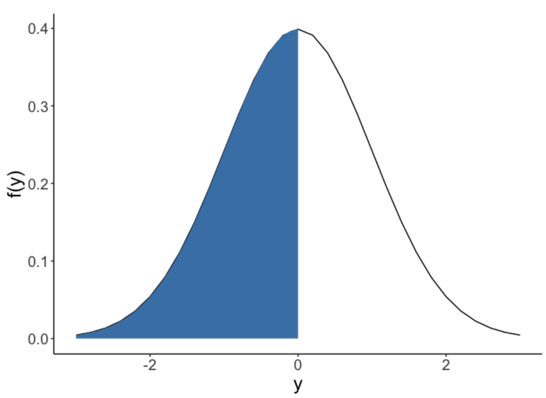
We can also extend variance to continuous random variables:

$$Var(X) = E(X - \mu_Y)^2 = E(X^2) - E(X)^2$$

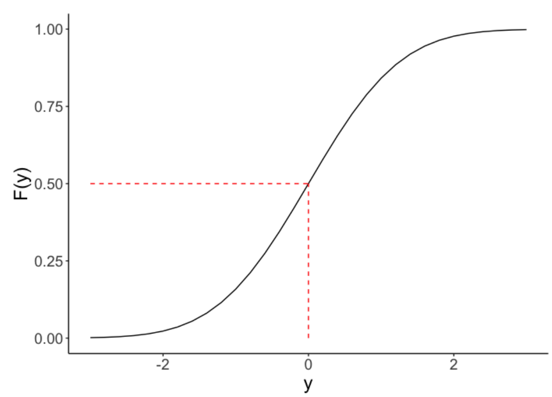
Just like with expectation, we are more interested in the theoretical meaning for variance rather than computation. Its meaning is also the exact same as the discrete case; how spread out values of Y are about μ_Y, the mean of Y.

3.5.5 Cumulative Distribution Functions

Cumulative Distribution Functions are similar for both the continuous and discrete case. It is a function that represents the cumulative probability up to a value of $y = a$, for a random variable Y. In other words it's a function that inputs $y = a$, and outputs $P(Y \leq a)$. For continuous random variables the pdf is often denoted with a lower case letter, while the cdf is denoted with a capitol letter. Here the pdf is $f(y)$, and the cdf is $F(y)$.



PDF for Y



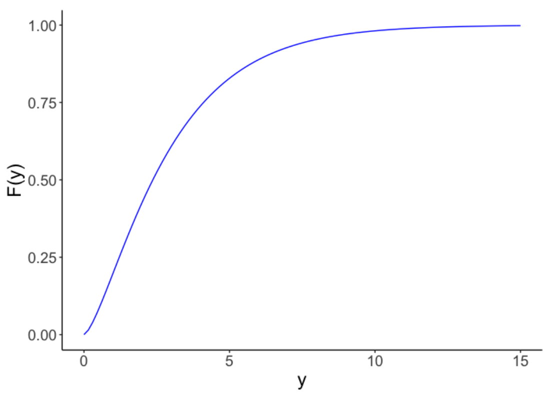
CDF for Y

Notice that if you look at the slope of the CDF at a single point it gives you the change in probability for a change in y. This is probability density at the particular point. In Calculus we would describe this relationship by saying the derivative of $F(y)$ is $f(y)$ (Don't worry about the notation if you are unfamiliar with calculus).

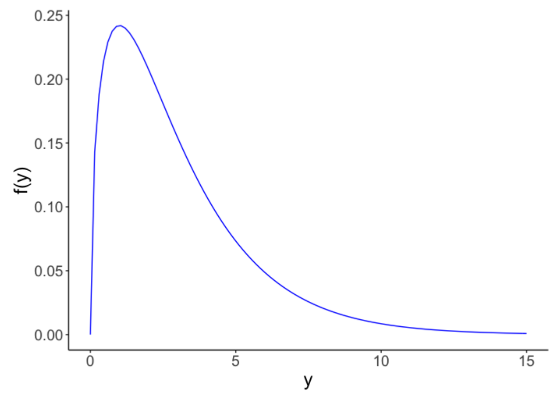
$$\frac{d}{dx} F(y) = f(y) \qquad \text{and} \qquad F(y) = \int_{-\infty}^{y} f(t) dt$$

Let's explore this relationship a little deeper. What if we want to know what the pdf looks like simply by looking at the cdf?

Example 1: For the following cdf, sketch the approximate pdf



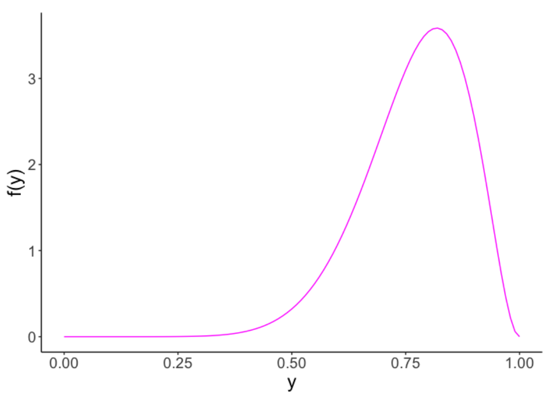
Cumulative Distribution Function for Y



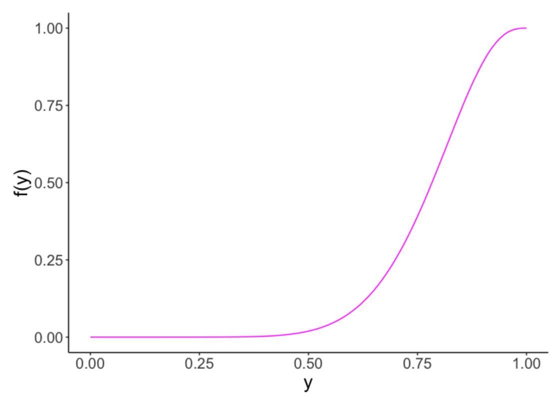
Probability Density Function for Y

Think about the pdf in terms of the **slope** of the cdf. It looks like the slope starts out flat, and then quickly rises to it's steepest at $y \approx 2$. The curve then starts to flatten out resulting in a smaller slope. As y gets large it seems the slope is approaching zero (a flat line). We call this a **skew right** density function. We will talk about the shape of probability distributions in more detail later.

Example 2: For the following pdf, sketch the approximate cdf



Probability Density Function for Y



Cumulative Distribution Function for Y

Now think about the area under the pdf. You can see that the area will continue to increase at an accelerated rate up to $y \approx 0.80$. After this the area continue to increase but at a decelerated rate until $y \approx 1.00$. Also notice that the acceleration in the first part of the graph is more moderate than the deceleration in the second part of the graph. The result is what we call a **skew left** density function.

Percentiles

Cumulative distribution functions also allow us to easily find *percentiles*. A percentile is the value of $Y = y$ such that a certain percentage of all Y values fall below y. You've probably heard before of someone scoring in the 99% percentile on an exam. This describes a score that is greater than or equal to 99% of all scores. For a random variable Y with cdf $F(y)$, the 99% percentile would simply be the value of y such that $F(y) = 0.99$.