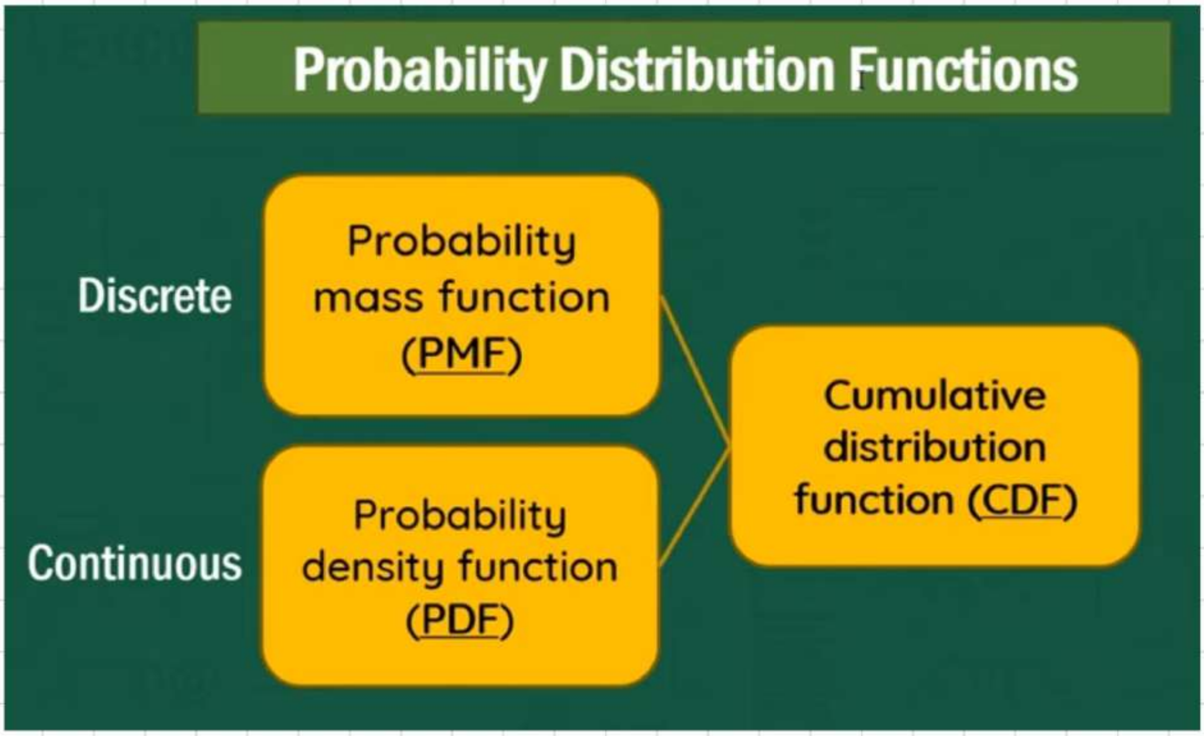
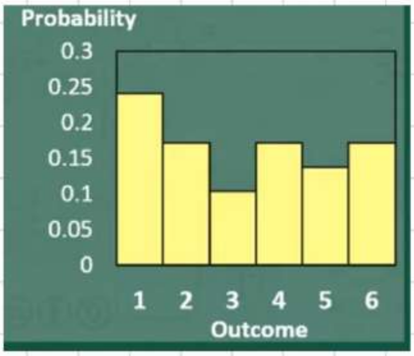
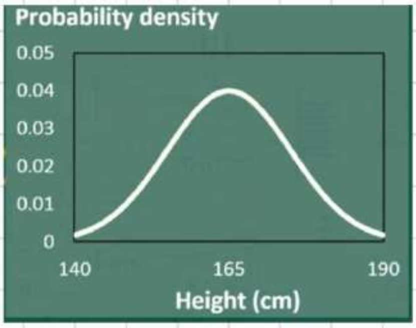


For discrete variables we have something called the probability mass function. It is just the probability of each discrete outcome.



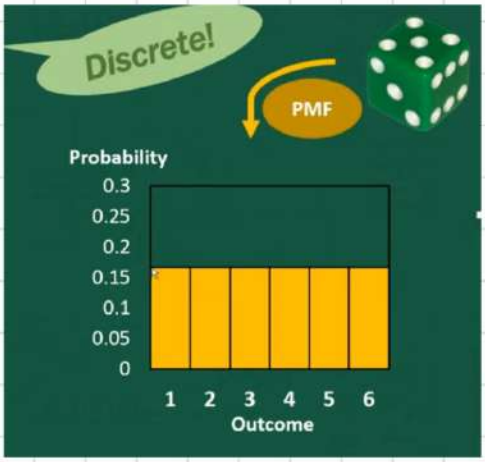
For continuous variables we use probability density function. It is slightly different than the previous one. We'll see how later.



Both discrete and continuous variables can construct a cumulative dist. function.

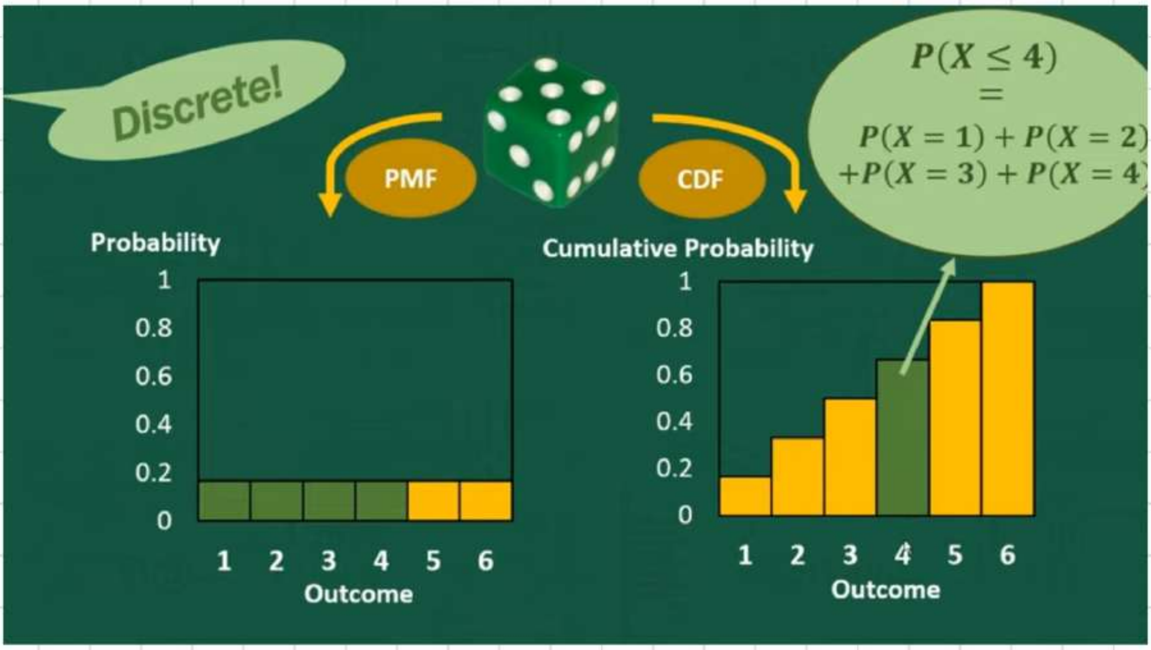
PROBABILITY MASS FUNCTION (DISCRETE VARIABLES)

Let's say we have a dice. Each outcome has the prob. of 1/6. This would be called a probability mass function.



It's discrete because outcomes are whole numbers.

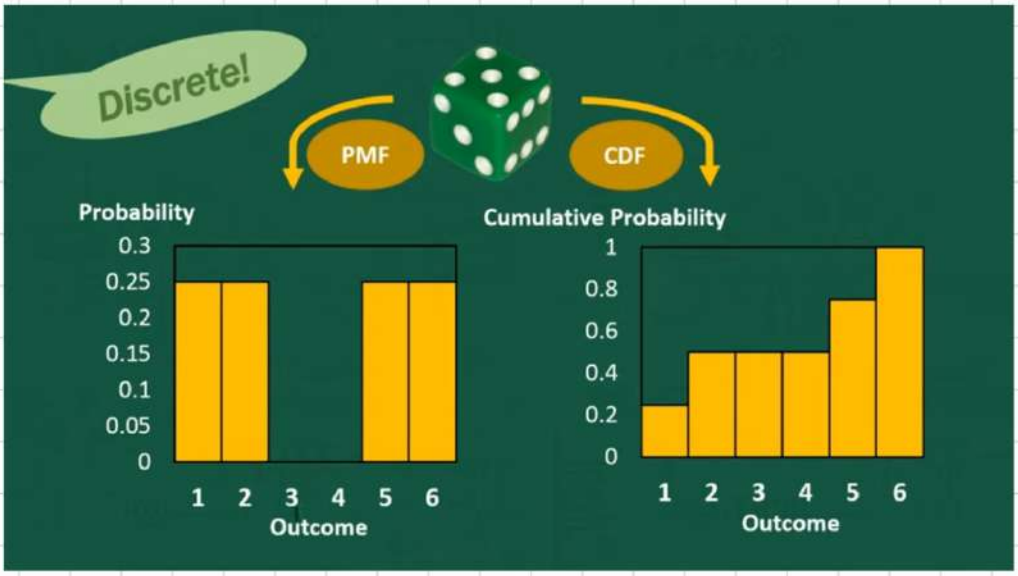
However we can also construct a cumulative probability distribution function with this:



In this example, the height of the 4th column does not represent the prob. of rolling a 4. Rather it reprsnts the prob. of rolling a no. less than or equal to 4.

Notice, the final bar needs to get to 1.

Let's pretend now that the dice is rigged such that $P(X=3) = P(X=4) = 0$. What will the two PDFs look like?



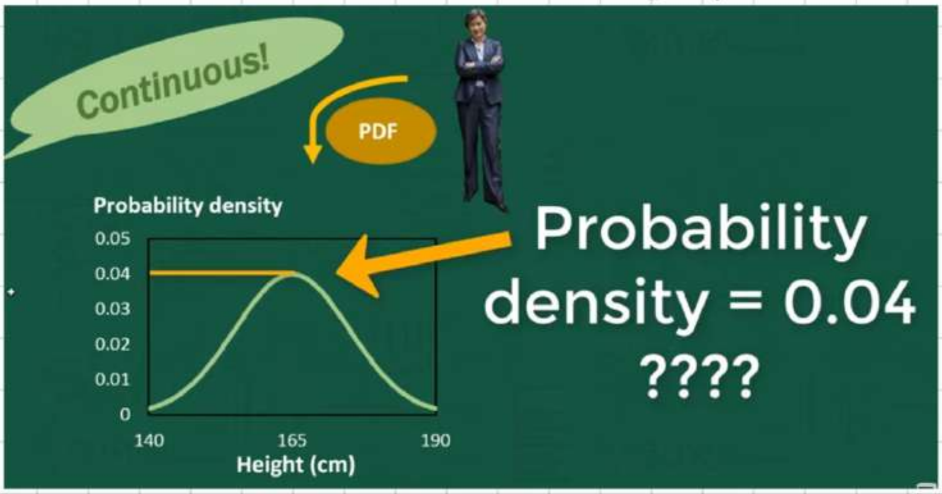
* Flatness in the cumulative distribution function means that there is no mass for some outcomes before.

PROBABILITY DENSITY FUNCTION (CONTINUOUS VARIABLES)

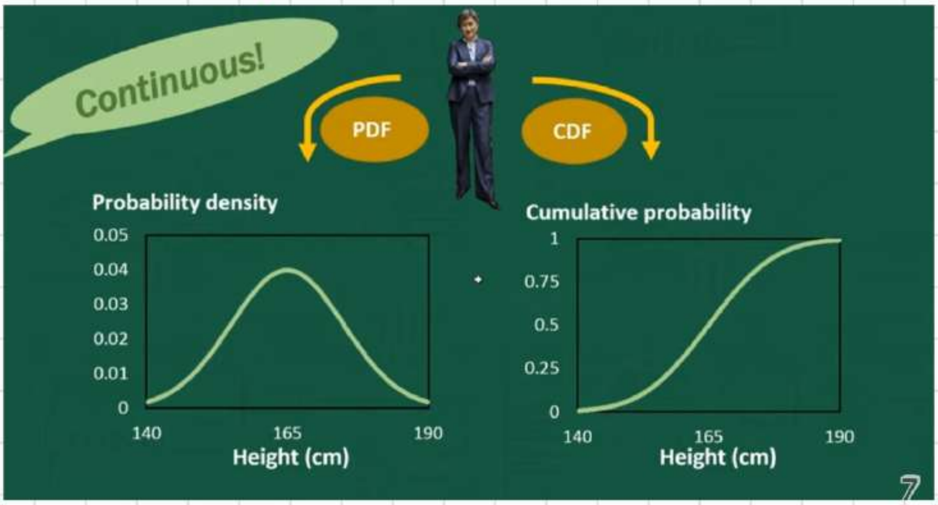
The example we are going to use for continuous variable is the height of women.

The distribution function we are going to get is called the probability density function :

But what does a prob. density of 0.04 at 165 cm mean.



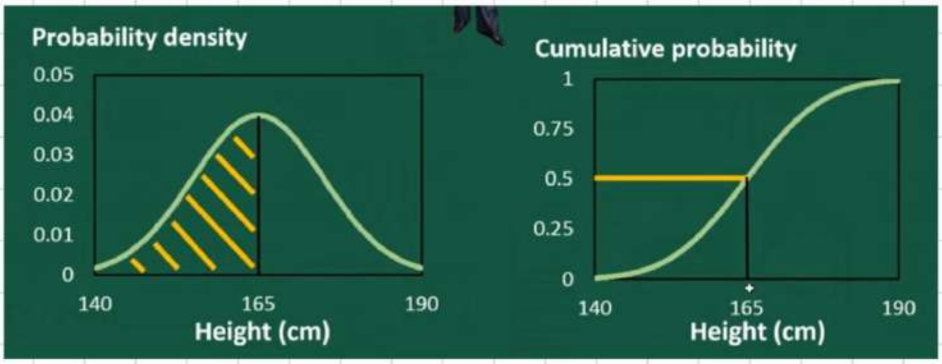
Hold this thought. Let's look at the cumulative probability distribution first.



also called S-curve.

Anything that has a bell curve as PDF becomes an S-curve as CDF

Let's see if we can get one from the other.



To the left of mean we know there lies half the probability. This is represented in the S-curve. At 165, the prob. is 0.5.

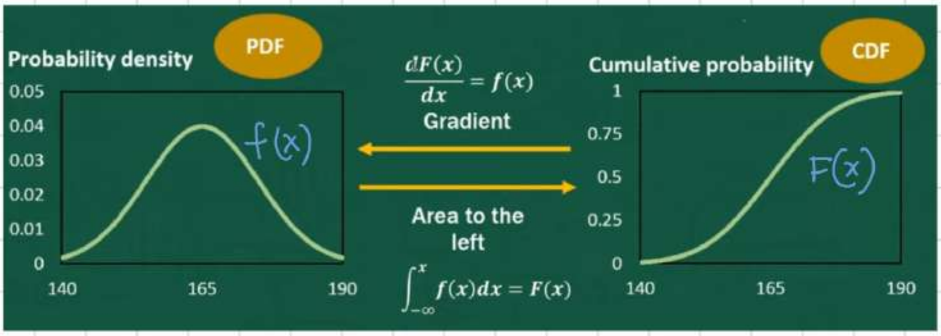
Same goes for any point. We can indeed derive the CDF from the PDF.

Can we go in the reverse way?
Can we get the PDF from the CDF.

The answer is that we can!
The gradient of this curve in the CDF gives us the PDF curve.

* Prob. density curve effectively represents the gradient of the CDF.

Summary :-



↳ Integrate PDF to get CDF
↳ Differentiate CDF to get PDF