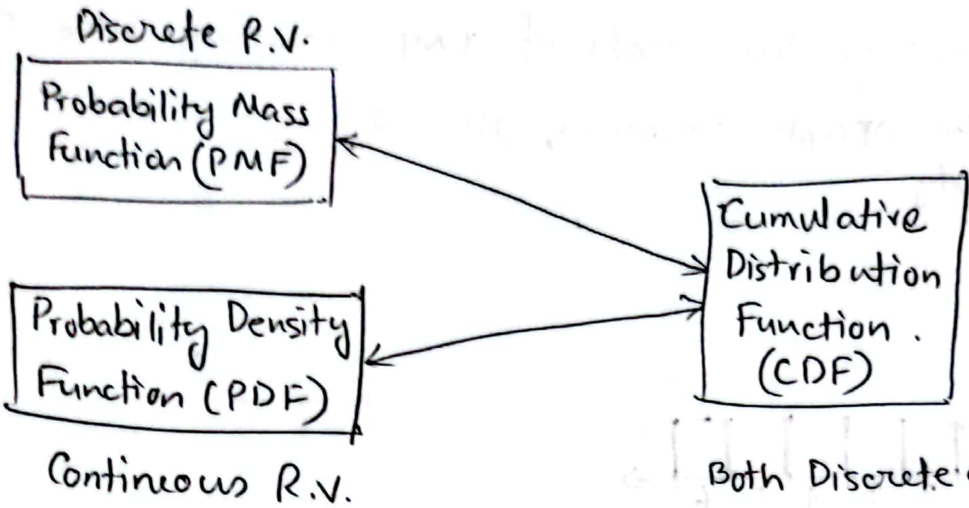
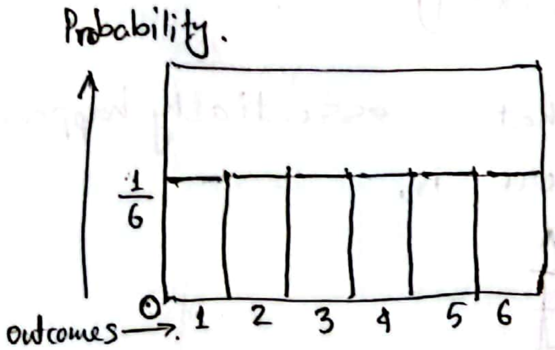
PMF, PDF, CDF

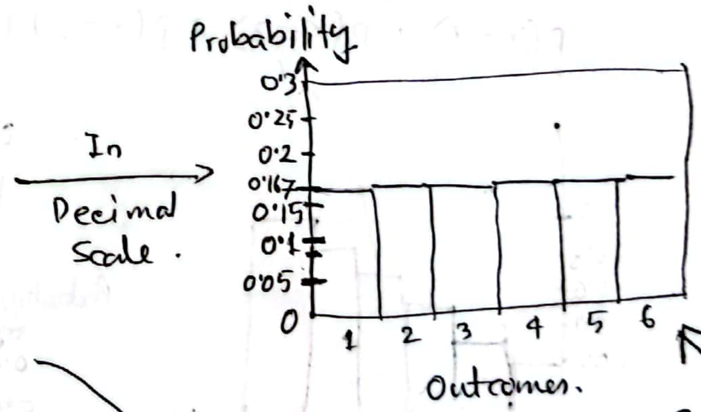


Both Discrete and Continuous R.V. can construct a CDF.

Discrete R.V.

Take an example of a fair die tossing experiment. The sample space, $S = \{1, 2, 3, 4, 5, 6\}$ and each of these outcomes will ~~be~~ have equal probability, which is $\frac{1}{6}$.





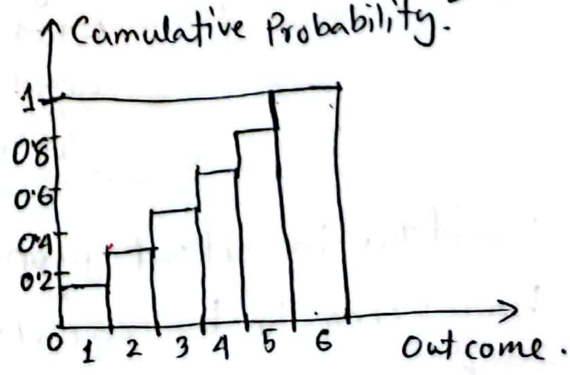
This is "PMF"

Scales are different

This is discrete because the outcomes are countable discretely. You cannot roll 2.5 or 4.6.

$\frac{1}{6} = 0.167$

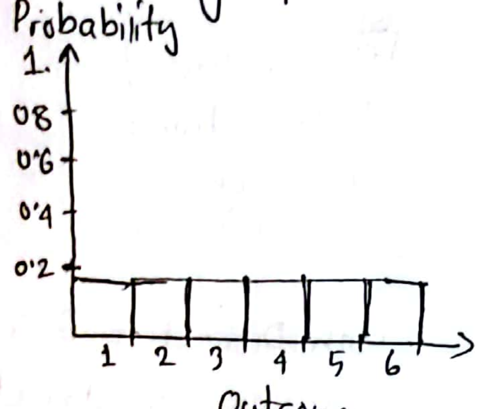
The CDF of this example will be,



In Y-axis we plot the Cumulative probability, not just the probability.

CDF

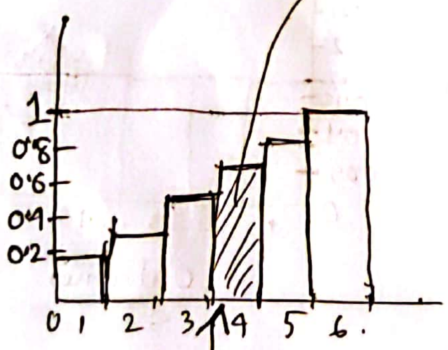
If we change the scale of PMF in a way like CDF. Then the graph becomes, like,



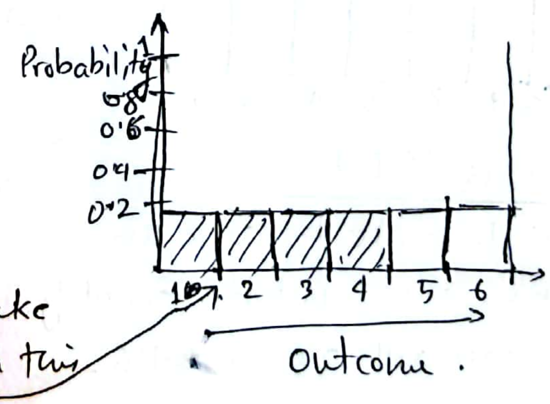
If we take the outcome 4 from the CDF diagram, the hight of 4 does not represents the probability of getting a 4, it is actually the probability of rolling a 4 or less. That is,

$P(X \leq 4)$. which is actually the sum up of,

$P(X=1) + P(X=2) + P(X=3) + P(X=4)$



what is essentially happening here; is,

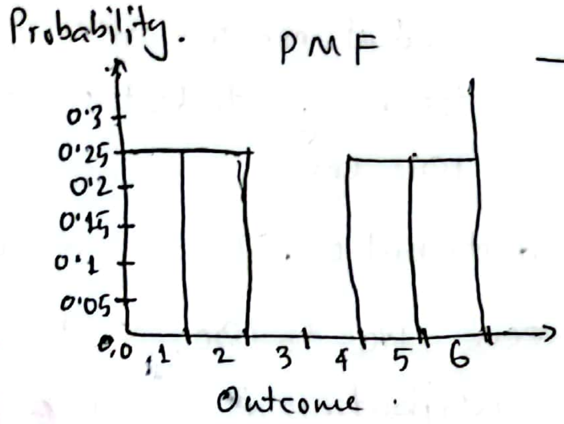


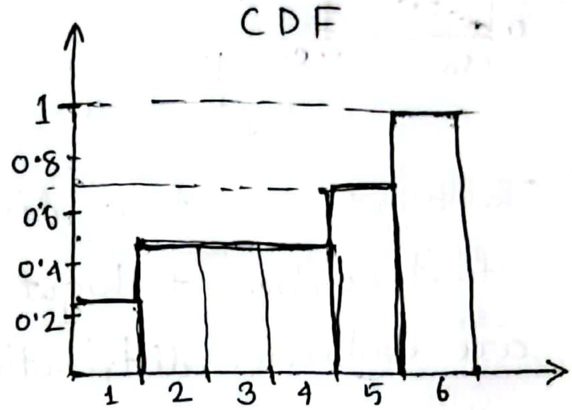
If you stake 1 to 4 from this graph.

You will get the hight of 4 in CDF graph.

• One of the important property of CDF is, that the final bar needs to be equal to 1. Because the probability of getting a 6 or less, in case of a fair dice has to be 100%. You cannot roll a 1·5 or 3·7. or 7.

Now consider the dice is modified such that it cannot roll 3, and 4. So ~~tha~~ how will it be its PMF.





The probability of rolling a, 1, 2, 5 or 6 is 0·25.

So here, $P(X \leq 4) = P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= \underbrace{P(X=1) + P(X=2)}$

This is same as the probability of getting 2 or less.

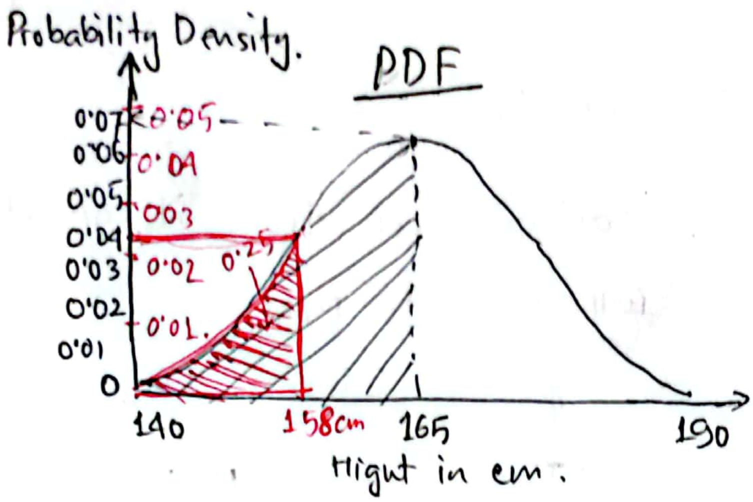
The flatum in CDF indicates that there is no Mass around 3 and 4.

<u>Contincous R.V.</u>

Let us consider the higut of your class. Let us say that the higut of the students of your class is distributed with mean, 165 cm. Majority of the students' higut will be around 165 cm.

Probability Density.

PDF



It has some kind standard deviation such that, by about 190 cm you will not get too many student nor will you get too many student up at 190 cm.

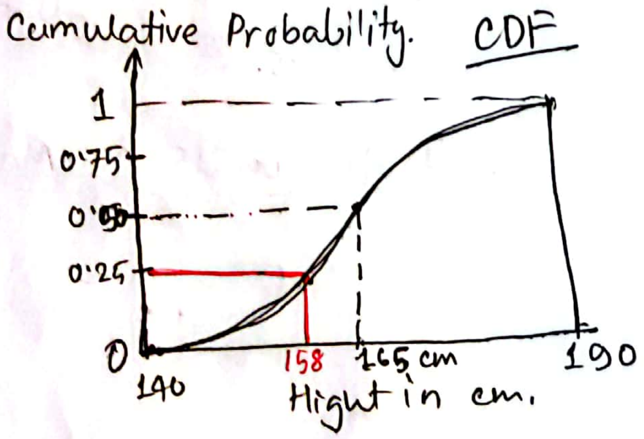
This hight is not a discrete distribution. There can be infinit number of hight in between 140 to 190. So these are contineous distribution. This distribution is called a PDF.

Now you might ask what does that curve mean? Does the pick point mean ~~that~~ correspond to 0.07 (0.05) mean that is there a 7% (0.07) (5% (0.05)) chance of being 165 cm tall?

Not exactly.

Now what will be the CDF of this PDF?

Cumulative Probability. CDF



Anything which is bell shaped curve for PDF will give you a S shaped ~~cumulative~~ curve in its cumulative probability.

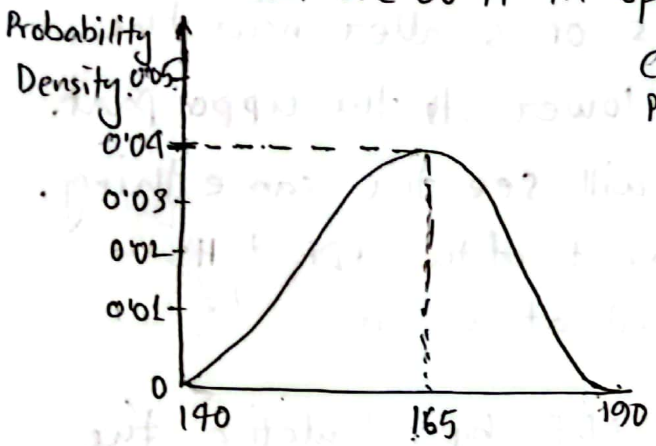
Now incase of PDF, the mean value is 165. In ~~case~~ case of any bell shaped curve, the proportion of the distribution to the left of the mean, is going to be 50%. The same thing is represented by the curve of CDF.

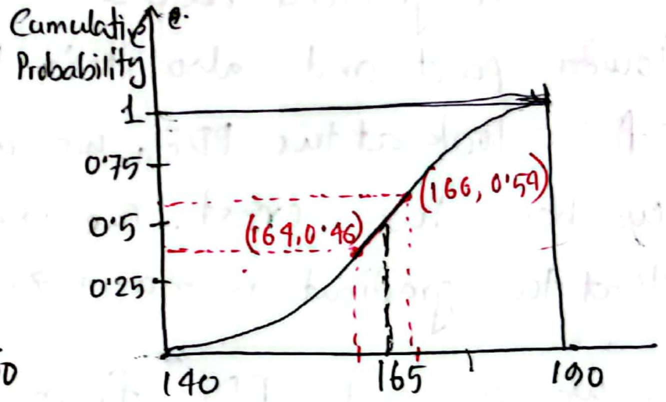
At 165 cm it goes upto a value of 0.5. This tells us that 50% or 0.5 of the distribution has ellapsed at this point of 165 cm.

In other way, we can tell that we have "ACCUMULATE" half of the distribution by the time we get to the point 165 cm. Now if we chose a point before 165 cm, may be that point corresponds to 25% of the distribution. At that point in case of PDF, the hight corresponds to 158 cm. Similarly, in case of CDF, we can see that 158 cm corresponds to the value 0'25.

Therefore, in case of CDF, the numbers on the Y-axis, actually telling us, How much of the distribution is the left of a given hight/or a given point. This is how we can derive CDF from PDF.

Can we do it in opposit way?





Question – How much of the distribution is going to be around 165 cm, and can we guess that from our CDF? Yes we can. and we can do it by finding out the gradient of the curve The higher the gradient, the more ~~will be~~ of the distribution will be hovering near 165 cm. If the line is flat that means none of the distribution is around there. Remember in case of discrete R.V. at 3, and 4 the curve was flat.

Now let us find the gradient at 165 cm. Construct an interval around 165 cm and pick up two points.

Say that the points are, $(164, 0{\cdot}46)$ and $(166, 0{\cdot}54)$. S

$$\text{Gradient} = \frac{\text{Rise}}{\text{Run}} = \frac{0{\cdot}08}{2} = 0{\cdot}04$$. This is the gradient at 165 cm.

Now have a look at the PDF. You can find out that in Y-axis, the highest point corresponds to $0{\cdot}04$. So the numbers in the Y-axis of PDF, is actually the Gradient of CDF.

In the CDF curve, we can see and understand that the gradient is the highest right at the middle of the CDF curve.

The gradient decreases or smaller near the lower part and also it is lower at the upper part. If we look at the PDF, we will see the same thing. We have the crest. The crest of the PDF tells us that the gradient is maximized at 165 cm.

We can get PDF from CDF by calculating the Gradient.

$$\begin{tikzcd} \boxed{\text{PDF}} \arrow[r, bend right, "\text{Area to the left.}"'] & \boxed{\text{CDF}} \arrow[l, "\text{Gradient.}"'] \end{tikzcd}$$

The gradient of CDF is the PDF. If we can find out the area to the left of a given point.

$$\begin{tikzcd} \boxed{\text{PDF}} \arrow[r, bend right, "\int_{-\infty}^{x} f(x)\cdot dx = F(x)"'] & \boxed{\text{CDF}} \arrow[l, bend right, "\frac{dF(x)}{dx} = f(x)"'] \end{tikzcd}$$

%

PDF = Differential of CDF

CDF = Itegral of PDF from $-\infty$ upto x