Probability in Computing

LECTURE 7

Last time

- Three kinds of sample spaces
- Random Variables

Today

- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)
- Probability Density Function (PDF)

Reminders

- HW 3 due on Thursday

Reading

- 19.3 (MIT)
- 3.1.3 (Supplement)

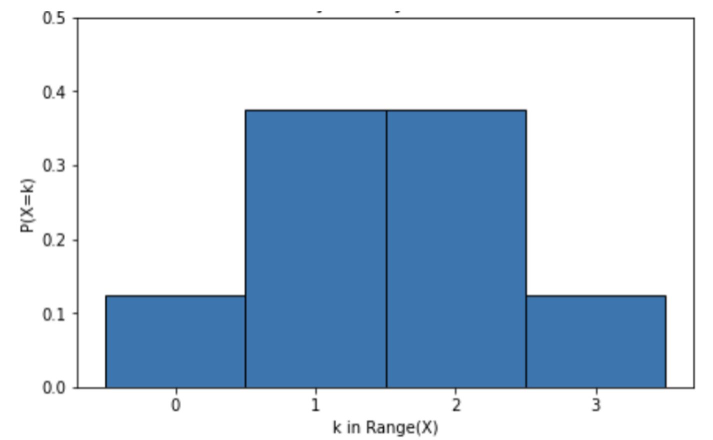
Probability Mass Function

For discrete random variables, the **Probability Mass Function** (PMF) is also called the probability distribution. Thus, when asked to find the probability distribution of a discrete random variable X, we can do this by finding its PMF.

C = "Flip three coins and return the number of heads"

Range(C) = { , , , }

Probability Mass Function P_C



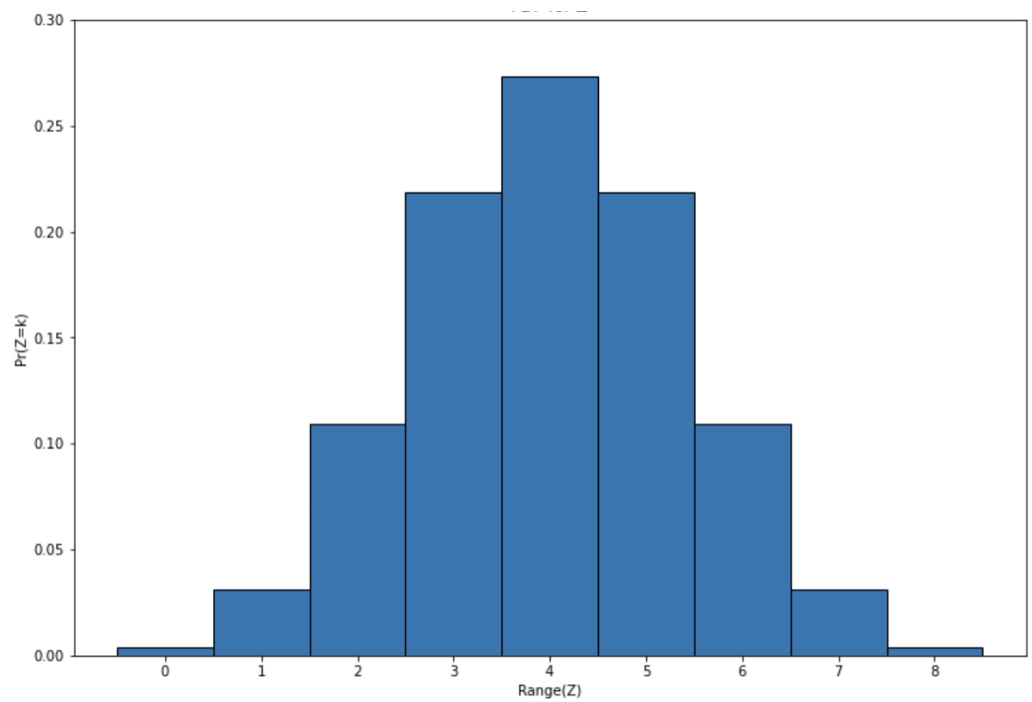
A PMF contains all the information you need to know about a Random Variable and is often illustrated as a histogram.

Probability Mass Function

Example 2:

Z = ”Toss 8 fair coins and count the number of heads"

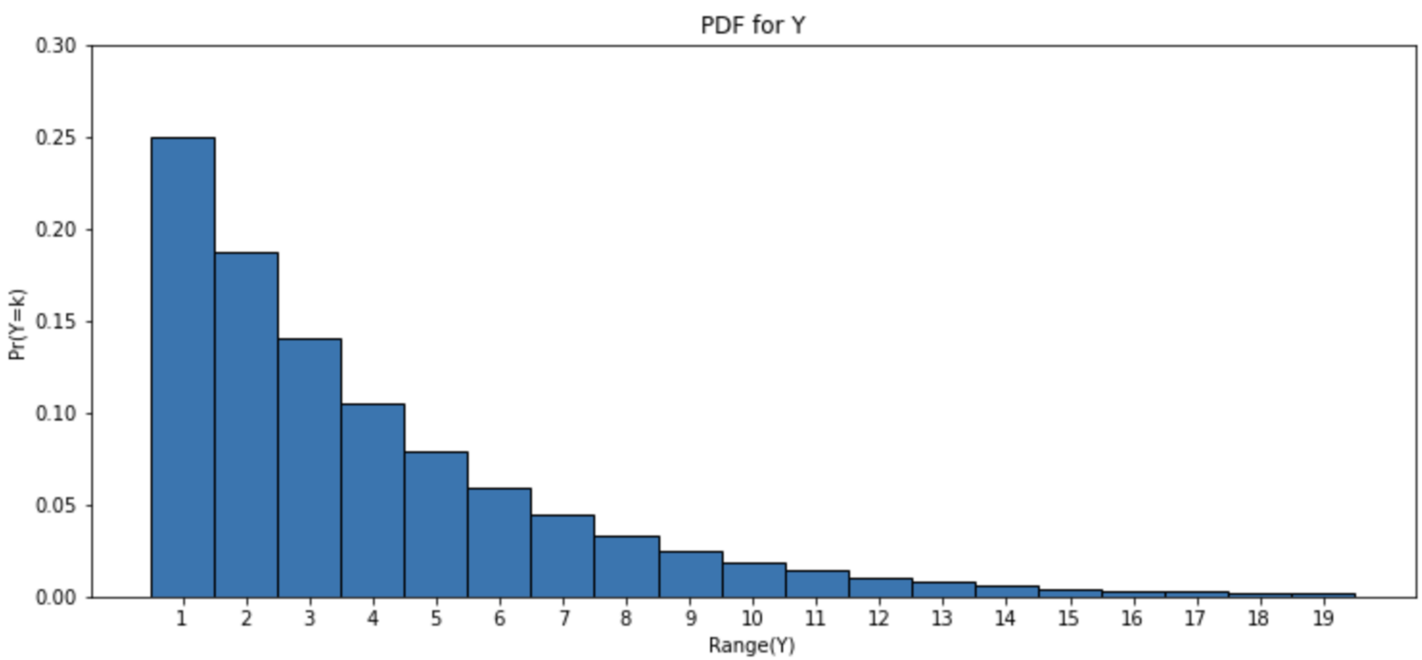
Range(Z) = { , , , , , , , }



Probability Mass Function

Example 3:

Y = “Flip a 4-sided die and count the number of flips until it lands a 4.”



Properties of PMFs

Consider a discrete random variable X with Range(X)=R_X.

Note that by definition the PMF is a probability measure, so it satisfies all properties of a probability measure.

Example

I have an unfair coin for which $P(H)=p$, where $0<p<1$. I toss the coin repeatedly until I observe a heads for the first time. Let Y be the total number of coin tosses. **Find the PMF of Y.**

Top Hat question one (Join: 033357)

What is the value of $P_Y(k)$?

A. $(1-k)p$
B. $(1-p^k)p$
C. $(1-p)^{k-1}p$
D. $(1-p)^k p$
E. None of the above

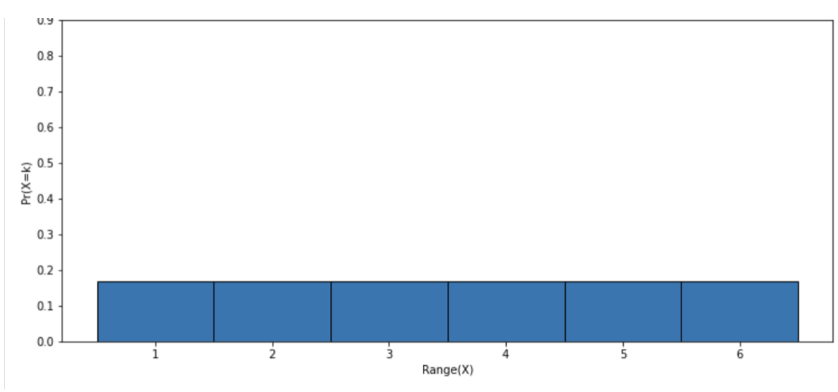
Example

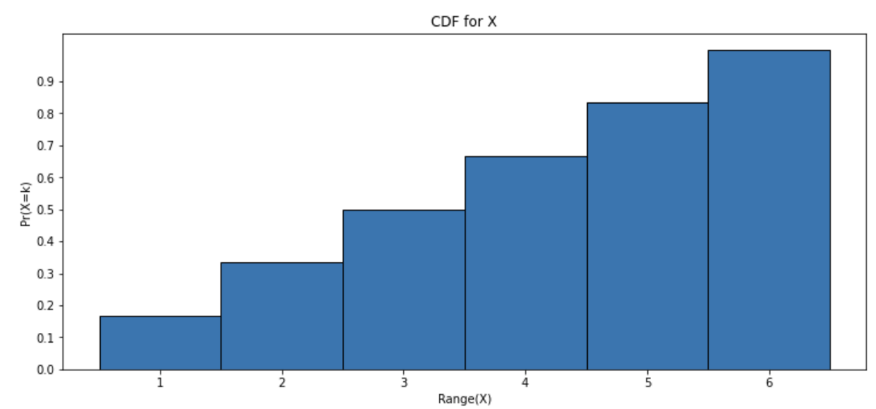
Your turn: If $p = \frac{1}{2}$, what is $P(2 \leq Y < 5)$?

Cumulative Distribution Functions

The Cumulative Distribution Function (CDF) for a random variable X shows what happens when we keep track of the sum of the probability distribution from left to right over its range.

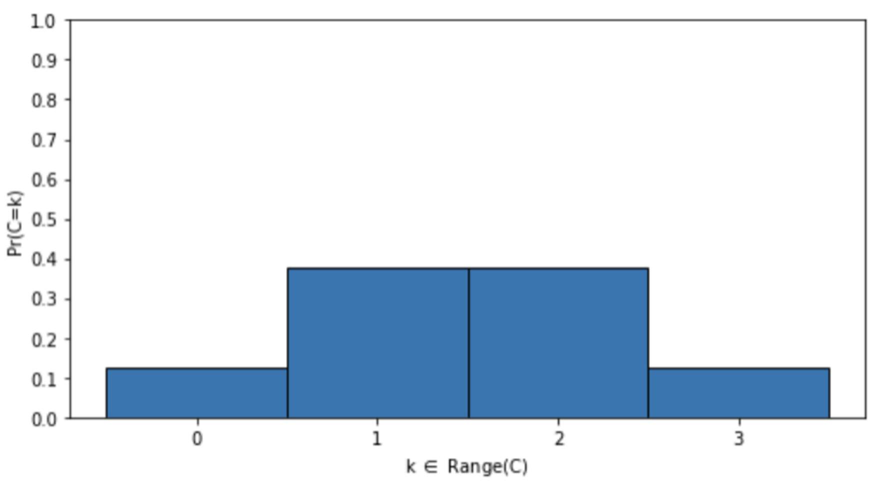
Example 1: X = “Count the dots on a thrown die”

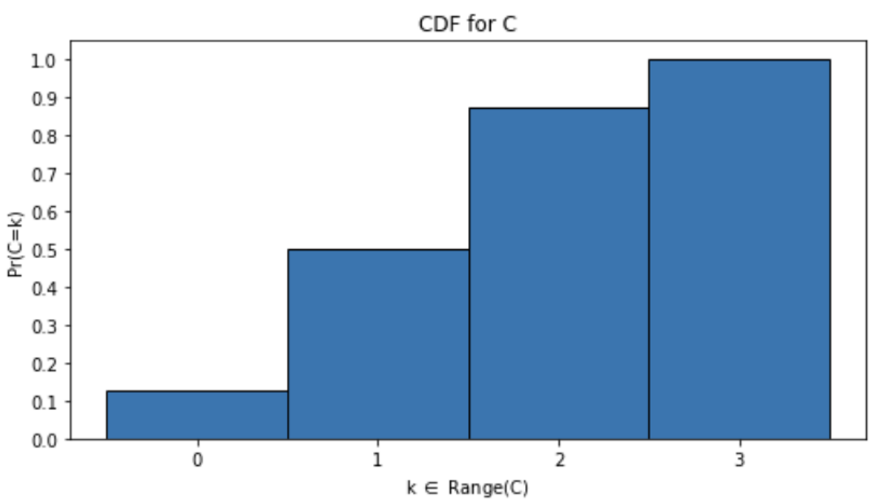




Cumulative Distribution Functions

Example 2: C = "Flip three coins and return the number of heads"

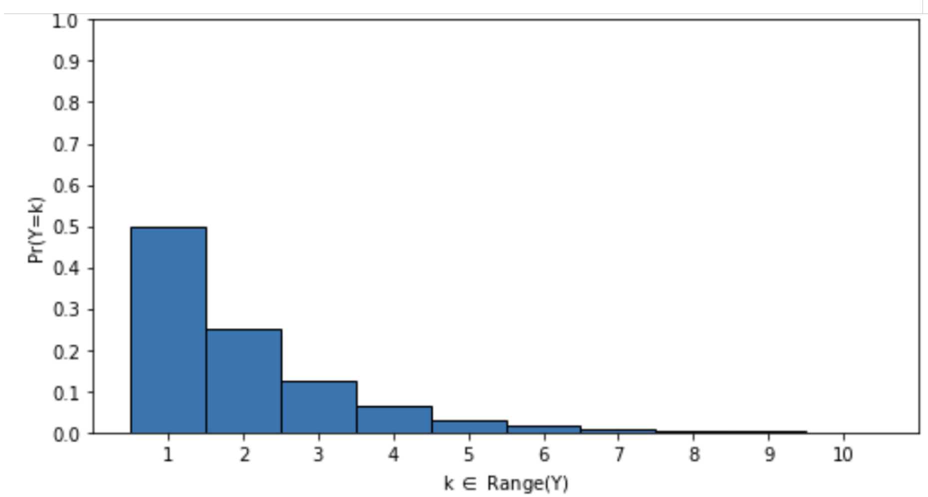


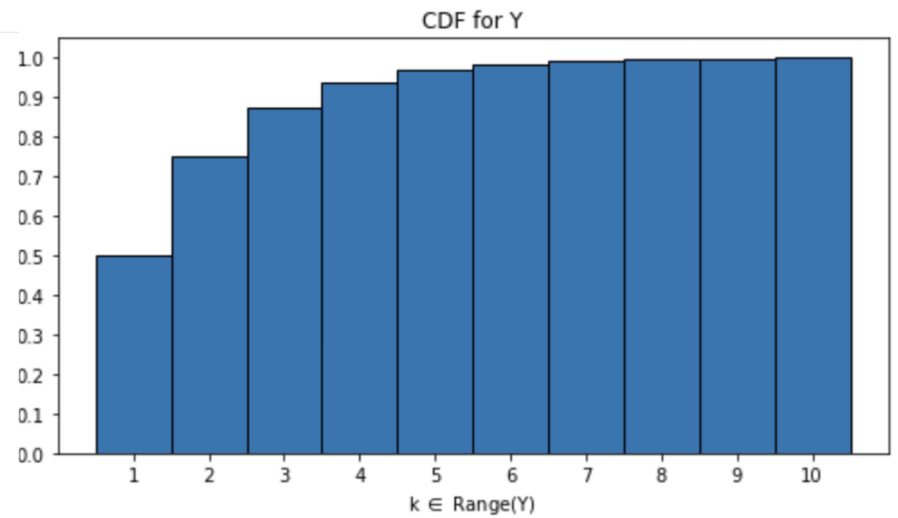


Cumulative Distribution Functions

Example 3:

Y = “Count how many flips until a fair coin shows a head.”





Properties of CDFs

- CDF is always a non-decreasing function

$$\text{if } y \geq x \text{ then } F_X(y) \geq F_X(x)$$

- CDF approaches 1 as x becomes large

$$\lim_{x \to \infty} F_X(x) = 1$$

- $P(a < X \leq b) = F_X(b) - F_X(a)$

Top Hat question (Join Code: 033357)

Suppose we have a random variable X with probability distribution:

k:	0	1	2	3
Pr(X = k):	0.1	0.4	0.2	0.3

Which of these is a possible list of probabilities for Pr(X <= x) ?

A. { 0.1, 0.4, 0.2, 0.3 }
B. { 0.1, 0.5, 0.8, 1.0 }
C. { 0.1, 0.5, 0.7, 1.0 }
D. { 0.1, 0.3, 0.6, 1.0 }

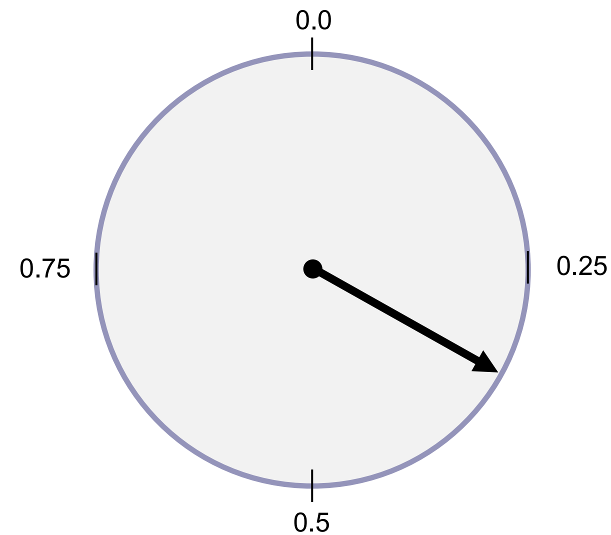
Probability Density Function

- Why PMF does not work for continuous random variables?

- For continuous random variables, the CDF is well-defined so we can provide the CDF.
- The concept is very similar to mass density in physics: its unit is probability per unit length.

Probability Density Function

Consider the spinner example again....

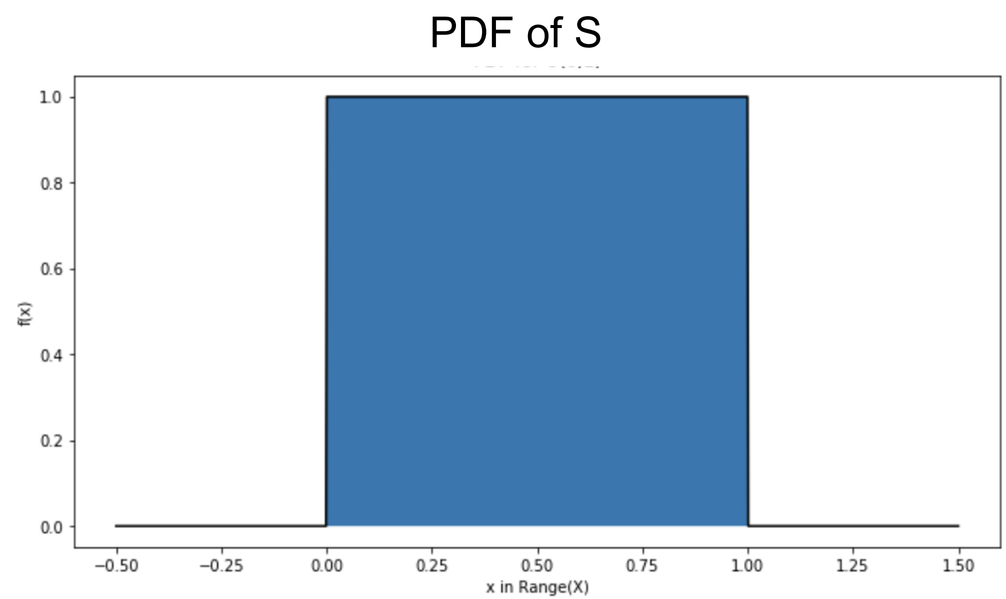


X = "choose a real number uniformly from [0..1]"

The PDF of X is uniform (same probability across the range) and has an area of 1.0, and is given by:

$$\text{Range}(X) = [0..1]$$

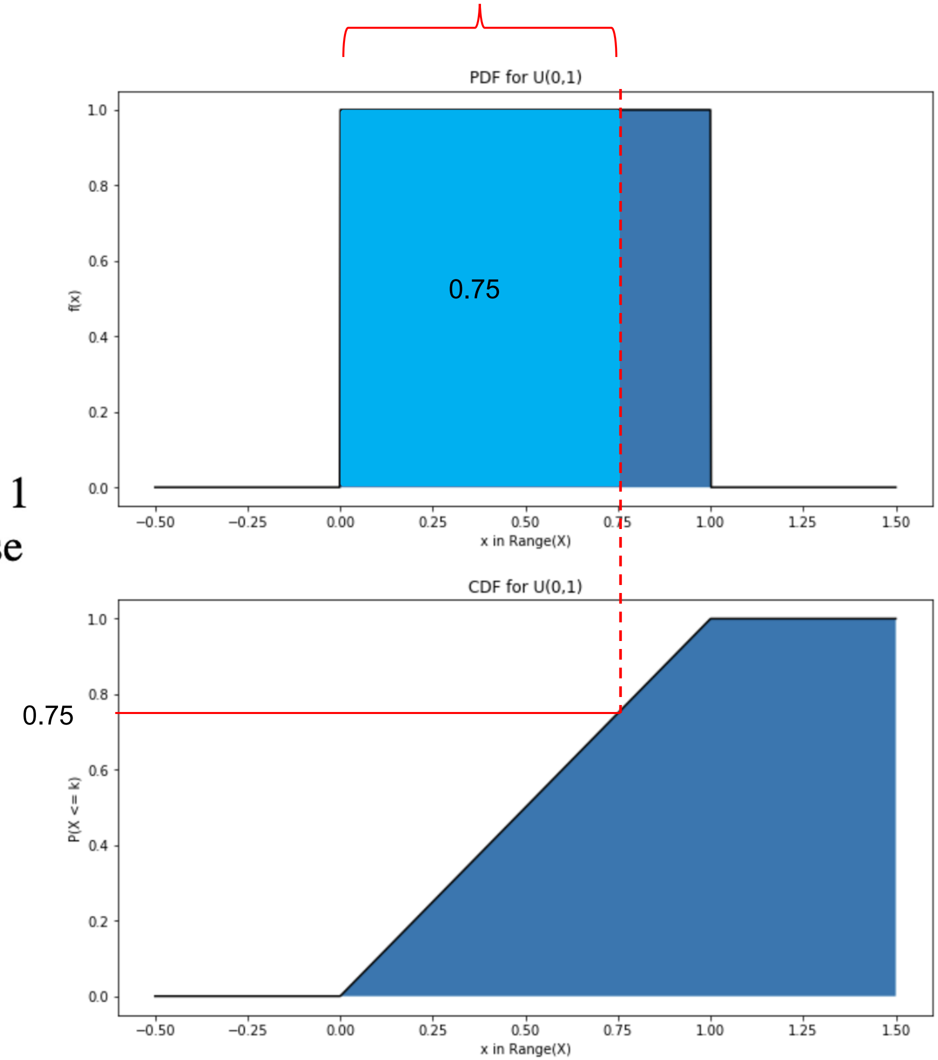
$$PDF_X(a) = \begin{cases} 1.0 & 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



PDFs and CDFs of Continuous Distributions

For continuous probability, we use extent (e.g., area) for probability. To calculate areas, we will use the CDF, which gives the area in the PDF to the left of given real number.

$$Pr(X < 0.75) = CDF_X(0.75) = 0.75$$



$$\text{Range}(X) = [0..1]$$

$$PDF_X(a) = \begin{cases} 1.0 & 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$F(a) = \int_0^a 1 \, dx = x \Big|_0^a = a$$

$$CDF_X(a) = \begin{cases} a & 0 \leq a \leq 1 \\ 1 & a \geq 1 \\ 0 & a \leq 0 \end{cases}$$

PDFs and CDFs of Continuous Distributions

To calculate probability of intervals, we use Inclusion/Exclusion:

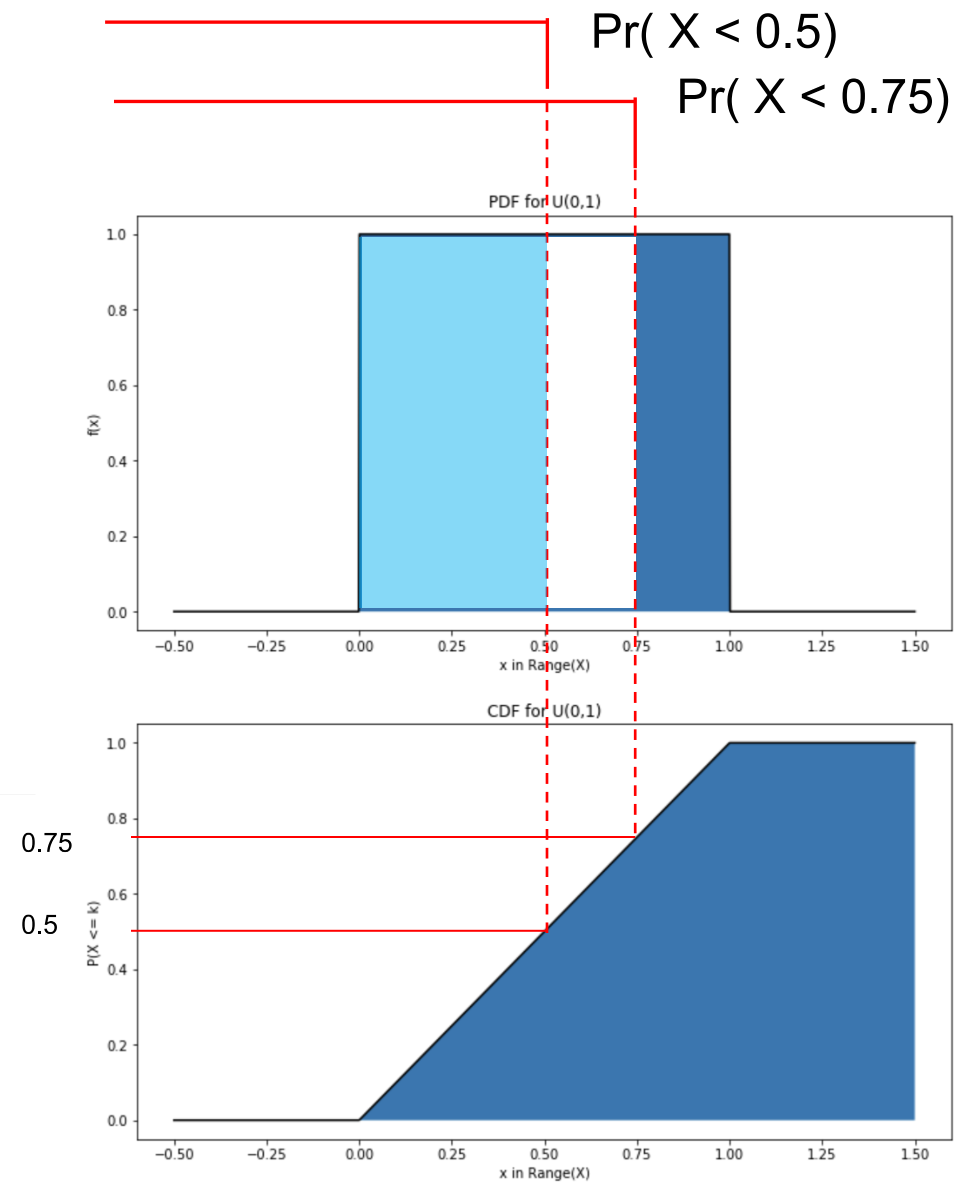
Pr([a..b]) = Pr(X < b) – Pr(X < a)

$$\text{Range}(X) = [0..1]$$

$$PDF_X(a) = \begin{cases} 1.0 & 0 \le a \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\int_0^a 1\, dx = x \Big|_0^a = a$$

$$CDF_X(a) = \begin{cases} 0 & \text{if } a < 0 \\ a & \text{if } 0 \le a \le 1 \\ 1 & \text{if } a > 1 \end{cases}$$

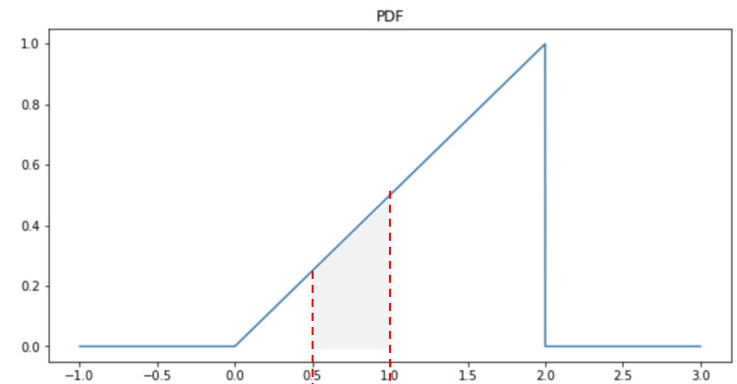


PDFs and CDFs of Continuous Distributions

Bottom Line: In order to deal with continuous distributions, you have to either calculate areas using geometric techniques , or do integrals.

Example: Suppose our PDF looked like this

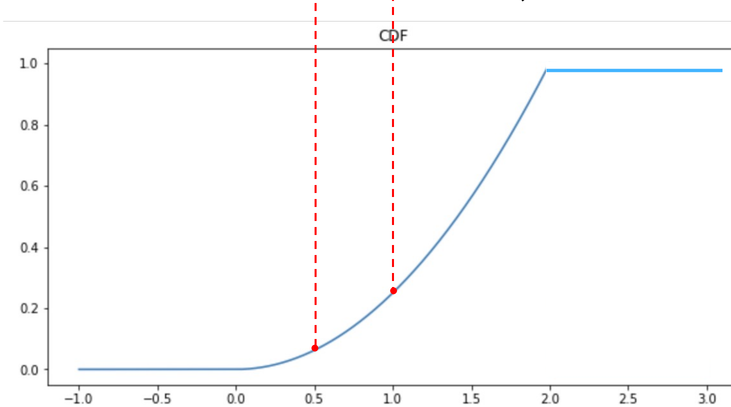
$$f(x) = \begin{cases} \dfrac{x}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$



To calculate the probability of intervals, we need to determine the CDF, which means doing the following integral:

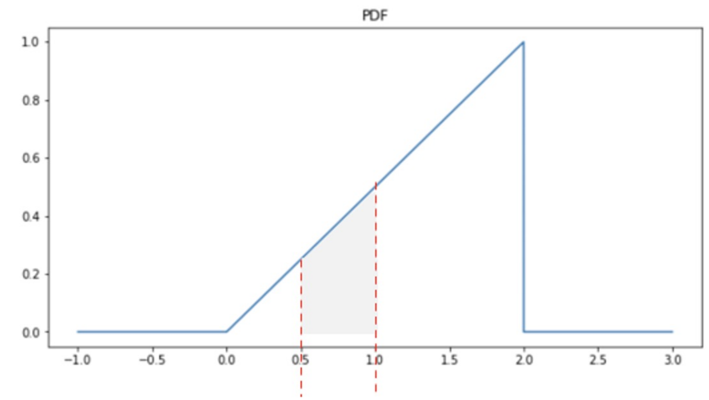
$$\int_0^a \frac{x}{2} dx = \frac{a^2}{4}$$

$$CDF_X(a) = \begin{cases} \dfrac{a^2}{4} & 0 \leq a \leq 2 \\ 1.0 & a > 2 \\ 0 & a < 0 \end{cases}$$



PDFs and CDFs of Continuous Distributions

$$F(a) = \int_{-\infty}^{a} \frac{x}{2}\, dx = \frac{a^2}{4}$$



$$P(0.5 \leq X \leq 1) =$$