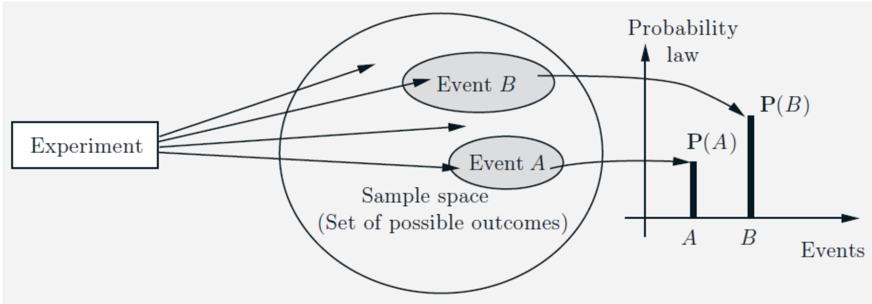
Lecture 06 -

Chapter 02: Random Variables

Math 4441: Probability and Statistics

Reference: Goodman & Yates – Introduction to Probability and Stochastic Process, 3rd Edition

Summary of Chapter 1: Simple Probability Models



Sample Space: S

- Elements of S can be anything
- Does not facilitate further processing

Probability of Outcomes/Events: $p[\cdot]$

- Lack a concise Representation of probabilities

Example 2.1: (**Example 1.12)**

Procedure: Send 3 packets from a sender to a receiver.

Observation: Number of successes.

$$S = \{FFF, FFD, FDF, FDD, DFF, DFD, DDF, DDD\}$$

Probabilities of Outcomes	
$P[FFF] = (1-p)^3$ $P[FFD] = p(1-p)^2$ $P[FDF] = p(1-p)^2$ $P[FDD] = p^2(1-p)$ $P[DFF] = p(1-p)^2$ $P[DFD] = p^2(1-p)$ $P[DDF] = p^2(1-p)$ $P[FDD] = p^3$	$P[\text{number of success is } 0] = (1-p)^3$ $P[\text{number of success is } 1] = 3p(1-p)^2$ $P[\text{number of success is } 2] = 3p^2(1-p)$ $P[\text{number of success is } 3] = p^3$

$p \triangleq$ probability that a single packet is delivered

Each of the deliveries is independent of the others

What do we need?

- Each element of S is a number
  - Define a function that converts each element $\omega \in S$ into a real number $x \in R$.
  - $X: S \to R$
- Probabilities in a mathematical way
  - Recalculate the probability of each real number or an interval of numbers as outcome
  - Represent both the real numbers and their probabilities mathematically

Probability Models

Random Variable:

Random variables express the outcome of an experiment by real numbers

- It is a function that generates values (numbers) on demand
- The values generated are random (Not kwon which one will appear)
- Values are related to the events of the experiment
  - Each value has its own chances of appearing
  - But it needs to be related to one event
- Function converts the events into real numbers

Distribution Function:

A distribution function represents a collection of probabilities

- Each probability is related to a real number, x
- Represents the chance of occurring the event represented by x

Random Variable:

How to define the functions?

- Identify the events related to the observations of the experiment
- Find an event space associated with the experiment (Why?)
- Assign a real number to each event – based on the problem statement

Type of Random Variables:

- Discrete random variables
  - Possible values are from a discrete set
  - Number of values are either finite or countably infinite
- Continuous random variables
  - Possible values are from an interval
  - Number of values are uncountable
- Mixed random variables

Example 2.1: (Continued)

Procedure: Send 3 packets from a sender to a receiver.

Observation: Number of successes.

$$S = \{FFF, FFD, FDF, FDD, DFF, DFD, DDF, DDD\}$$

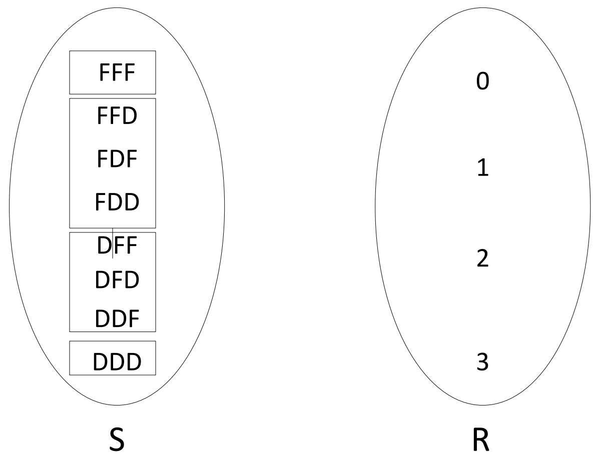
Events related to the observations

$E_i \triangleq \#$ of success(es) is i

$$i = 0, 1, \ldots, 3$$

$$E = \{E_0, E_1, E_2, E_3\}$$

$$S_X = \{ \qquad \qquad \}$$



Definition (Random Variable): The point function $X(\omega)$ is called a *random variable* if

(a) It is a finite real-valued function defined on the sample space *S* of a random experiment, and

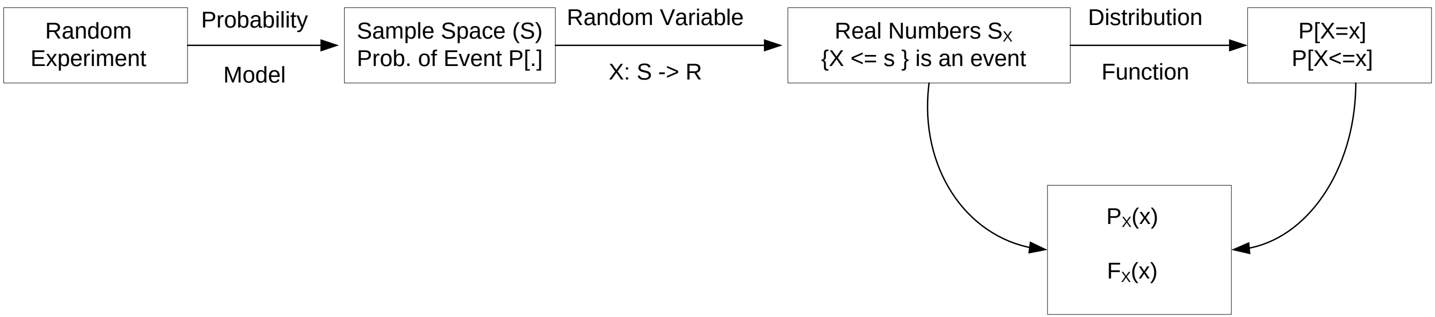
(b) For every real number x, the set $\{\omega : X(\omega) \leq x\}$ is an event.

$$X : S \rightarrow R$$

Distributions Functions:

- Represents the distribution of probabilities on the number line
- A probability is attached to each number on the number line
- Probabilities are non-zero for a number or an interval, only if the random variable can take on that value
- Probability that x is an outcome is related to the event corresponding to x defined by the random variable

Probability Models by Random Variables



Events associated with a random variable are:

- The random variable has a specific value
  - $\{X = x\}$ or more specifically $\{X(\omega) = x\}$
- The random variable has a value which is less than or equal to a specific value:
  - $\{X \leq x\}$ or $X(\omega) \leq x\}$
- The random variable has value which is greater than a specific value
  - $\{X \geq x\}$ or $\{X(\omega) \geq x\}$

<u>Possible distribution functions are</u>:

- Probability Mass function (PMF): The probability that a random variable X has a specific value x
  - $P[X = x]$

- Cumulative distribution function (CDF): The probability that a random variable has a value which is less than or equal to a specific value x
  - $P[X \leq x]$

- Complementary cumulative distribution function (CCDF): The probability that a random variable has a value which is greater than a specific value x
  - $P[X > x]$

Example 2.1: (Continued)

Procedure: Send 3 packets from a sender to a receiver.
Observation: Number of successes.

$$S = \{FFF, FFD, FDF, FDD, DFF, DFD, DDF, DDD\}$$

$$E = \{E_0, E_1, E_2, E_3\}$$

$X \triangleq$ Random variable that counts the number of successes

$$S_X = \{0, 1, 2, 3\}$$

	E_0	E_1	E_2	E_3
S	FFF	$FFD\ \ FDF\ \ DFF$	$FDD\ \ DFD\ \ DDF$	DDD
x				
$P[X = x]$				
$P[X \leq x]$				

Probability Mass Function (PMF): $P_X(x)$

If the set of possible values of $X, S_X = \{x_1, x_2, \dots, x_n\}$, then

1. $P_X(x_i) = 0$, if $x_i \notin S_X$

2. $P_X(x_i) = P[X = x_i]$, and hence, $P_X(x_i) > 0$, for $i = 1, 2, \dots, n$

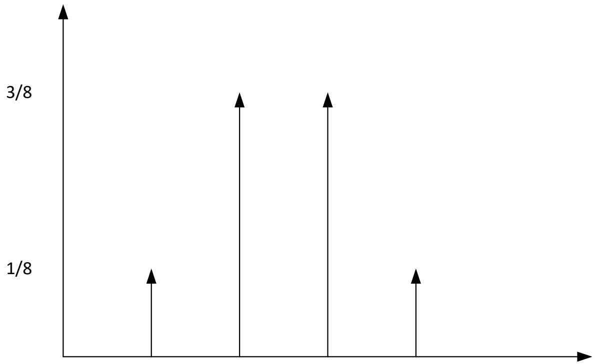
3. $\sum_i^n P_X(x_i) = 1$

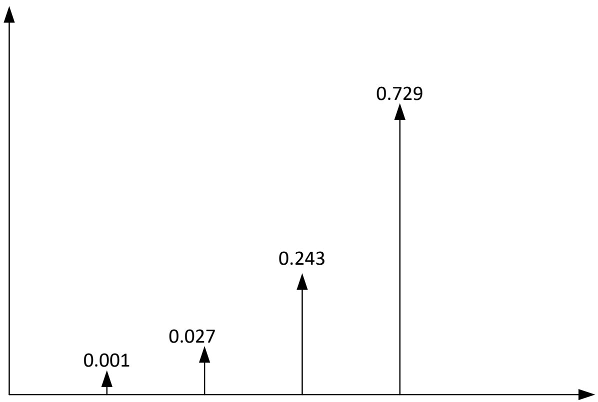
Example 2.1 (Continued)

$$P_X(x) = \begin{cases} \quad , & x = 0 \\ \quad , & x = 1 \\ \quad , & x = 2 \\ \quad , & x = 3 \\ 0, & \text{otherwise} \end{cases} \qquad P_X(x) = \begin{cases} \quad , & x = 0 \\ \quad , & x = 1 \\ \quad , & x = 2 \\ \quad , & x = 3 \\ 0, & \text{otherwise} \end{cases}$$

For $p = \frac{1}{2}$,

Graphical Representation of the PMF



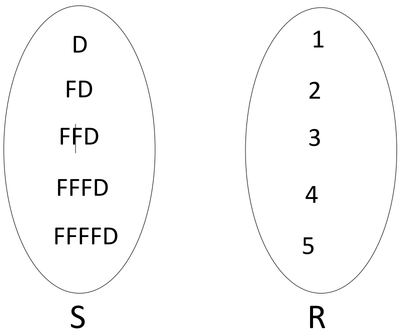


Example 2.2 (Example 1.13 Continued)

Procedure: Keep sending packets from a sender to a receiver until 1 packet is delivered

Observation: Number of attempts

$$S = \{D, FD, FFD, FFFD, \ldots\} \quad s_X = \{1, 2, 3, \ldots\}$$



	E_1	E_2	E_3	E_4	E_5
S	D	FD	FFD	$FFFD$	$FFFFD$
x					
$P[X = x]$					
$P[X \leq x]$					

$$F_X(x) = 1 - (1 - p)^x \quad x \geq 1$$

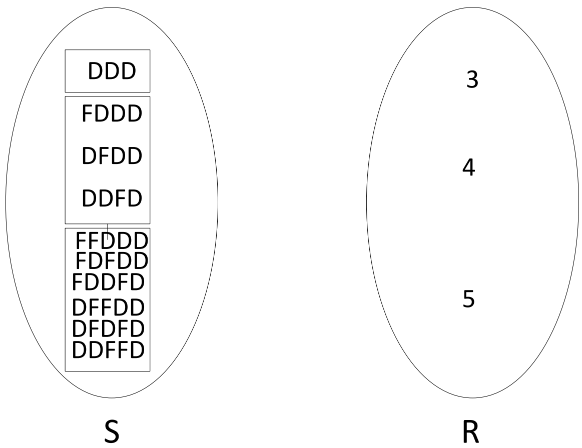
Example 2.3: (Example 1.14 Continued)

Example 1.14 (**Experiment 1.5)**

Procedure: Keep sending packets from a sender to a receiver until 3 packets are delivered
Observation: Number of attempts

$$S = \{DDD, FDDD, DFDD, DDFD, FFDDD, \dots\}$$

$$S_X = \{3, 4, 5, \dots\}$$



	E_3	E_4	E_5
S	DDD	$FDDD \quad DFDD \quad DDFD$	$FFDDD \quad FDFDD \quad FDDFD$ $DFFDD, DFDFD, DDFFD$
x			
$P[X = x]$			
$P[X \leq x]$			

Cumulative Distribution Function (CDF)

$F_X(x) = P[X \leq x]$

Probability that X will assume a value from the subset of S, where the subset is the point x and all the points to the left of x.

Properties of CDF:

1. It is applicable to both discrete and continuous RVs
2. It is nonnegative, non-decreasing function of x
3. For discreate random variables it is step function
    a. Jumps at the values of x where $P_X(x) > 0$
    b. For continuous random variables, it is continuous
4.
    a. $F_X(-\infty) = 0$
    b. $F_X(+\infty) = 1$
5. If a and b are two real numbers such that $a < b$

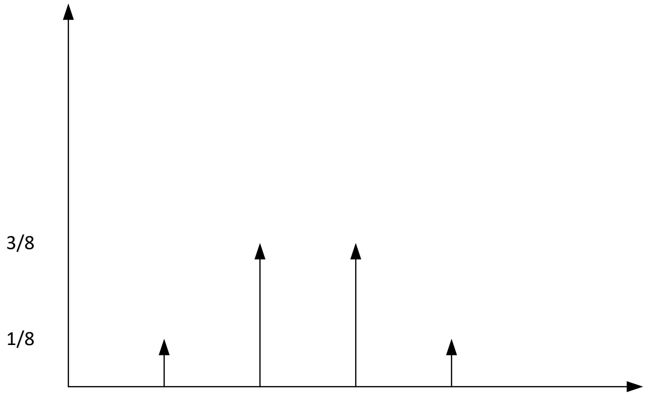
$$P[a < X \leq b] = F_X(b) - F_X(a)$$

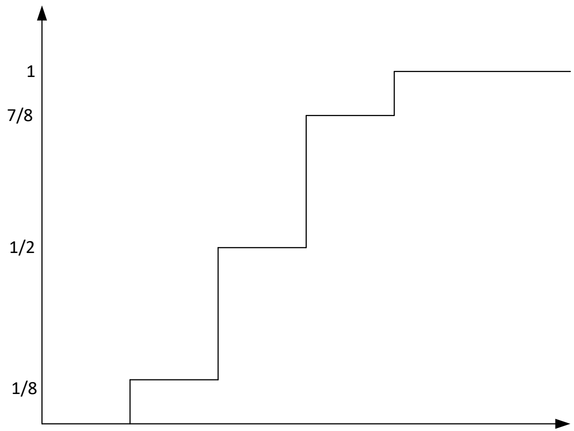
Which is a direct result of

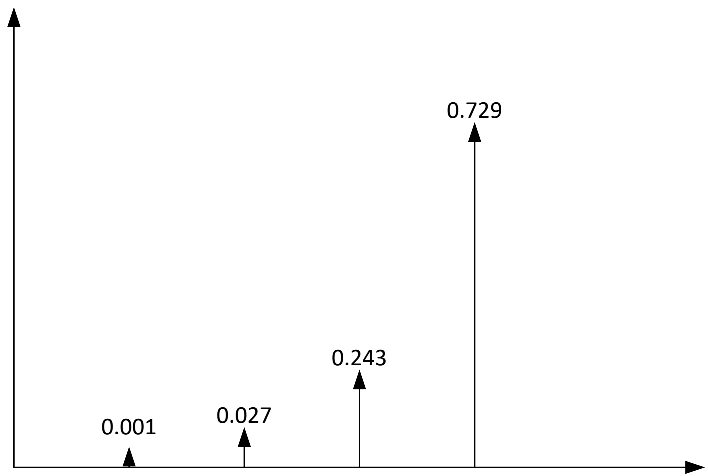
$$P[X \leq b] = P[X \leq z] + P[a < X \leq b]$$

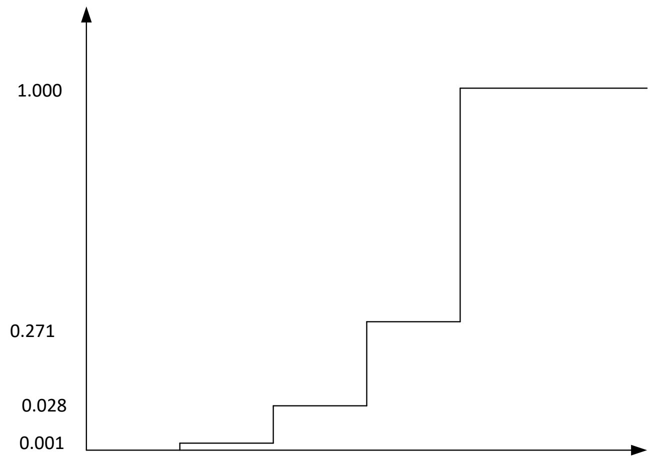
Example 2.4: Let a discrete random variable X assumes values $-1, 1, 2$, and 3, with probabilities 0.6, 0.3, 0.08, and 0.02 , respectively.

$$P_X(x) = \begin{cases} 0.6, & x = -1 \\ 0.3, & x = 1 \\ 0.008, & x = 2 \\ 0.002, & x = 3 \\ 0, & \text{otherwise} \end{cases}$$	$$F_X(x) = \begin{cases} 0, & x \leq -1 \\ 0.6, & -1 \leq x < 1 \\ 0.9, & 1 \leq x < 2 \\ 0.98, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$


3/8
1/8


1
7/8
1/2
1/8


0.729
0.243
0.027
0.001

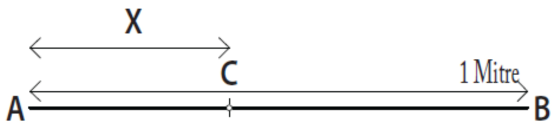

1.000
0.271
0.028
0.001

Continuous Random Variables

- Set of possible values of a random variable is uncountable or denumerable
- Set of values are represented by a range of values or by an interval
    - The delay of a packet to reach the destination from the source.
- The probability that a continuous random variable has a specific value is ?

Probability Models of Continuous Random variables

Consider a line AB of length 1 unit.

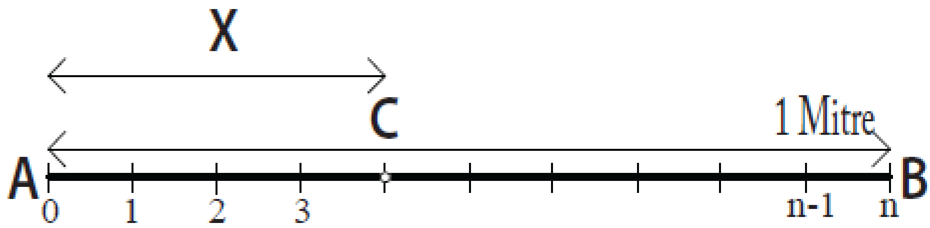


Suppose you randomly choose a point C within AB that divides the line into two parts AC and CB

Let, the length of the point C from A, AC, is a random variable and is denoted by X

Since $S_X = (0, 1)$ is an interval, X is a continuous random variable

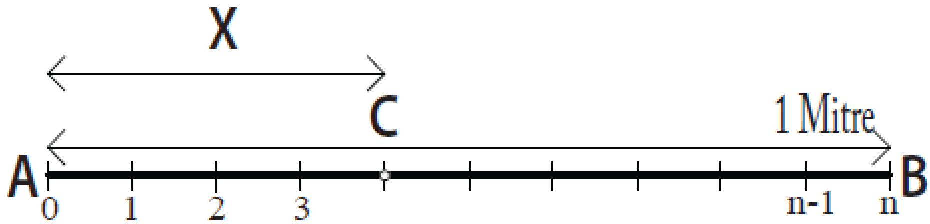
Further there are infinite possible points in between A and B, therefore, the probability that X has any specific value is $\frac{1}{\infty}$, i.e., intuitively it is zero.



To develop a probability model of X, let us consider a reasonable discrete approximation of X

Let us divide the line segment into n equal segments, each numbered from 1 to n

Since all segments are equal in length, if we randomly select a point from AB, it is equally likely that the selected point will be on any specific segment.



Let Y denote a discrete random variable, representing the number of the segment on which the random point lies.

The range of values of Y, S_Y, is

$$S_Y = \{1, 2, \dots, n\}$$

Two important questions to be answered are:

1. The relation between random variable X and the random variable Y
2. How well does Y approximate the value of X.

From the Figure, we can easily see that

$$Y = \lceil nX \rceil$$

If we denote $\{X = x\}$ and $\{Y = \lceil nX \rceil\}$ as two events, we have

$\{X = x\} \subset \{Y = \lceil nX \rceil\}$, and it implies

$$P[X = x] \leq P[Y = \lceil nX \rceil] = P[X = x] \leq \frac{1}{n}$$

$$P[X = x] \leq \lim_{n \to \infty} P[Y = \lceil nX \rceil = \lim_{n \to \infty} \frac{1}{n} = 0$$

$$P[X = x] \leq 0$$

However, according to the 1st Axioms of Probability $P[X = x] \geq 0$, hence

$$P[X = x] = 0$$

PMF as a Probability Model for Continuous Random Variables:

Since $P_X(x) = P[X = x]$ and $P[X = x] = 0$ for continuous random variables,

- The probability that a continuous RV has a specific value is always zero.
- The MPF is meaningless for a continuous random variable.

Distribution Functions for Continuous Random Variables

- The probability of a continuous random variable for a specific value is not defined.
- However, probability for a range of values, an interval, is well defined
  - $P[a < X \leq b]$ is well defined
- All points in $[0,1]$ are equally likely to be selected as point C
- Assume $a = 0$ and $c = 0.5$, intuitively we can say that 50% of the time point C will be selected in the interval $[0, 0.5]$
- Probability that X has a value between 0 and 0.5 is $P[a < X \leq b] = P[0 < X \leq 0.5] = \frac{0.5}{1.0} = 0.5$
- Further assume that $a = \infty$ and $b = x$ then

$$P[a \leq X \leq b] = P[-\infty < X \leq x] = P[X \leq x] = F_X(x)$$

Cumulative Distribution Function (CDF) of Continuous Random Variables

The cumulative distribution function of a continuous random variable X is denoted as $F_X(x)$, and is defined as

$$F_X(x) = P[X \leq x]. \tag{3.9}$$

The CDF of the example Random variable is

$$F_X(x) = \begin{cases} 0, & \text{for } x < 0; \\ x, & \text{for } 0 \leq x < 1; \\ 1, & \text{for } x \geq 1. \end{cases}$$

Properties of CDF for Continuous Random Variables

1. The CDF of a continuous random variable for a value of $-\infty$ is zero:

$$F_X(-\infty) = P[X \leq -\infty] = 0.$$

2. The CDF of a continuous random variable for a value of $+\infty$ is one, i.e.,

$$F_X(+\infty) = P[X \leq +\infty] = 1.$$

3. The probability that a continuous random variable for a value within the interval $(a, b]$ can be given in terms of its CDF

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

4. For continuous random variables, following four possibilities are equal due to the fact that $P[X = x] = 0$.

$$P[a \leq X \leq b] = P[a < X \leq b] = P[a < X < b] = P[a \leq X \leq b].$$

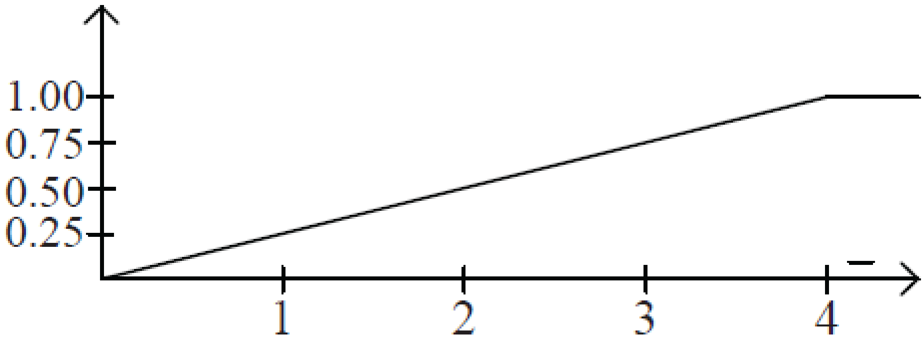
5. For continuous random variables, if $a < b$, then $F_X(a) < F_X(b)$.

6. The CDF of a continuous random variable is non-decreasing and continuous function of x.

Example 2. 10: The CDF of a continuous random variable is

$$F_X(x) = \begin{cases} 0, & \text{for } x < 0; \\ \frac{x}{4}, & \text{for } 0 \le x < 4; \\ 1, & \text{for } x \ge 1. \end{cases}$$

a) Draw the CDF curve.
b) Find the values of $F_X(-1)$, $F_X(1)$, $P[2 < X \le 3]$ and $F_X(1.5)$.


1.00
0.75
0.50
0.25
1
2
3
4

$F_X(-1) = 0.$

$F_X(1) = \frac{1}{4}.$

$P[2 < X \leq 3] = F_X(3) - F_X(2) = \frac{3}{4} - \frac{2}{4} = \frac{1}{4}.$

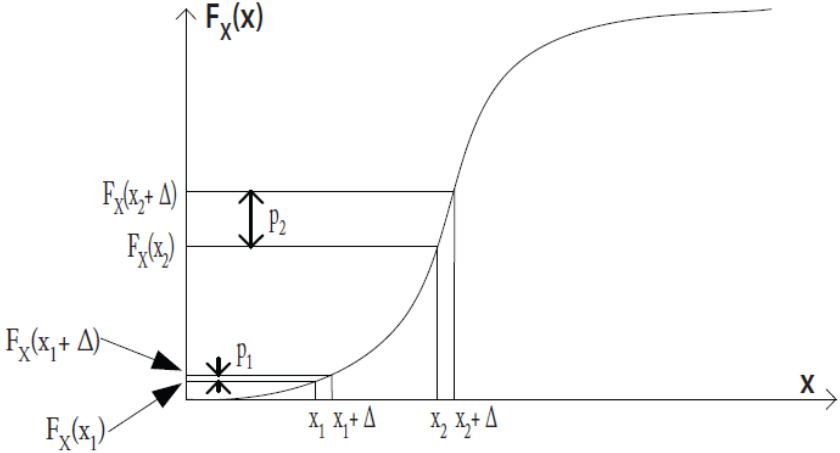
Probability Density Function (PDF)

- PMF: Distribution of probabilities (one unit) on the number line

For continuous Random variable defined in $(0, 1)$

- Distribute one unit of probability in the interval $[0, 1]$ on the number line
- Infinite points, cannot assign probability to a specific point, $P[X = x] = 0$
- Though, can assign probability for a range of values, e.g., 0.25 unit in $[0,0.25]$ 0.50 unit in $[0, 0.50]$
- Distribution can be uniform or non-uniform

We lack to quantify the amount of probability for a specific value of X

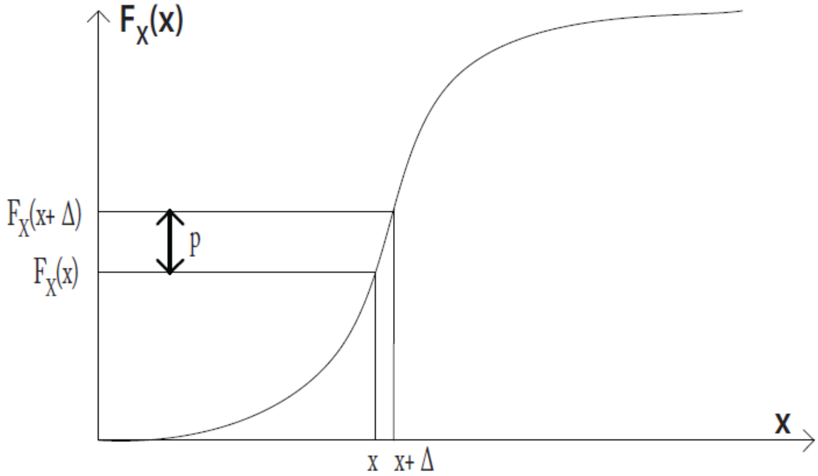


p_1: probability that X is within x_1 and $x_1 + \Delta$

p_2: probability that X is within x_2 and $x_1 + \Delta$

$$
\begin{aligned}
p_1 &= P[x_1 < X \leq x_1 + \Delta] \\
&= F_X(x_1 + \Delta) - F_X(x_1)
\end{aligned}
$$

$$
\begin{aligned}
p_2 &= P[x_2 < X \leq x_2 + \Delta] \\
&= F_X(x_2 + \Delta) - F_X(x_2)
\end{aligned}
$$



Rate of change of probability is higher if the curve is steeper

$$
\begin{aligned}
p &= P[x < X \leq x + \Delta] \\
&= F_X(x + \Delta) - F_X(x) \\
&= \frac{F_X(x + \Delta) - F_X(x)}{\Delta} \times \Delta
\end{aligned}
$$

<u>Average amount of Probability per unit length</u>

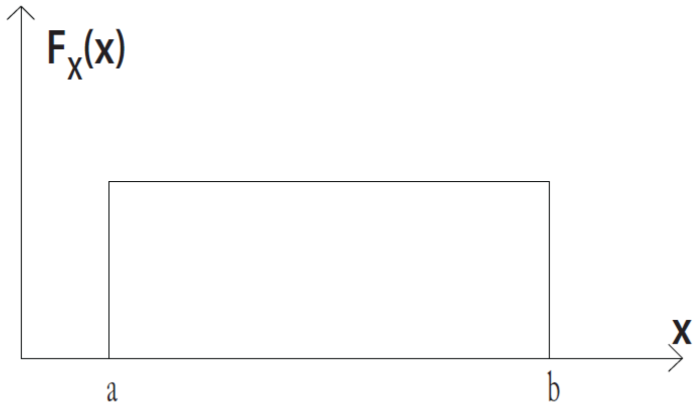
<u>Density:</u> measure of amount of mass in a given space (volume)

Probability Density: Measure of the amount of probability per unit length

$$f_X(x) = \lim_{\Delta \to 0} \frac{F_X(x + \Delta) - F_X(x)}{\Delta}$$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Area Under the Curve (AUC)



$$P[a < X \leq b] = \int_a^b f_X(x)dx$$

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

$$\begin{aligned} P[-\infty < X \leq x] &= \int_{-\infty}^{x} f_X(x)dx \\ &= P[X \leq x] \\ &= F_X(x). \end{aligned}$$

Example 2.11: For the CDF $F_X(x)$ given in Example 3.11, find the following:

1. $f_X(x)$, from the CDF
2. $F_X(x)$, from the PDF
3. $P[2 < X \leq 3] =?$
4. Draw the PDF curve

$$
\begin{aligned}
f_X(x) &= \frac{d}{dx} F_X(x) \\
&= \frac{d}{dx} \frac{x}{4} \\
&= \frac{1}{4}.
\end{aligned}
$$

$$
f_X(x) = \begin{cases} \frac{1}{4}, & \text{for } 0 \le x \le 4; \\ 0, & \text{ohterwise.} \end{cases}
$$

$$
\begin{aligned}
F_X(x) &= \int_{-\infty}^{x} f_X(x)dx \\
&= \int_{0}^{x} \frac{1}{4}dx \\
&= \frac{x}{4}\Big|_{0}^{x} \\
&= \frac{x}{4} - \frac{0}{4} = \frac{x}{4}
\end{aligned}
$$

$$
F_X(x) = \begin{cases} 0, & \text{for } x < 0; \\ \frac{x}{4}, & \text{for } 0 \leq x < 4; \\ 1, & \text{for } x \geq 1. \end{cases}
$$

$$
\begin{aligned}
P[2 < X \leq 3] &= \int_2^3 f_X(x)dx \\
&= \int_2^3 \frac{1}{4}dx \\
&= \frac{x}{4}\bigg|_{x=2}^{x=3} \\
&= \frac{3}{4} - \frac{2}{4} = \frac{1}{4}
\end{aligned}
$$

Example 2.13: The PDF of a continuous random variable X is

$$f_X(x) = \begin{cases} cxe^{-x/2}, & \text{for } x \geq 0; \\ 0, & \text{ohterwise.} \end{cases}$$

1. Find the value of the constant c.
2. Find the CDF of the random variable X.
3. Find the probability $P[2 \leq X \leq 5]$ from the PDF of X.
4. Find the probability $P[2 \leq X \leq 5]$ from the CDF of X.

$$
\begin{aligned}
\int_{-\infty}^{+\infty} f_X(x)dx &= \int_0^{\infty} cxe^{-x/2}dx \\
&= cx(-2)e^{-x/2}\Big|_{x=0}^{x=\infty} - \int_0^{\infty} 1.(-2)e^{-x/2}dx \\
&= 0 + 2c\int_0^{\infty} e^{-x/2}dx \\
&= 2c(-2)e^{-x/2}\Big|_{x=0}^{x=\infty} \\
&= 0 + 4c = 4c
\end{aligned}
$$

$4c = 1$ and $c = \frac{1}{4}$.

$$
\begin{aligned}
F_X(x) &= \int_{-\infty}^{x} \frac{1}{4} x e^{-x/2} dx \\
&= \frac{1}{4} x(-2) e^{-x/2} \Big|_{x=0}^{x=\infty} - \frac{1}{4} \int_0^{\infty} 1.(-2) e^{-x/2} dx \\
&= -\frac{1}{2} x e^{-x/2} + \frac{1}{2} \int_0^{\infty} e^{-x/2} dx \\
&= -\frac{1}{2} x e^{-x/2} + \frac{1}{2} (-2) e^{-x/2} \Big|_{x=0}^{x=\infty} \\
&= -\frac{1}{2} x e^{-x/2} - e^{-x/2} + 1 \\
&= 1 - e^{-x/2} - \frac{1}{2} x e^{-x/2}.
\end{aligned}
$$