

Advanced Statistics

Dr. Syed Faisal Bukhari

Associate Professor

Department of Data Science

Faculty of Computing and Information Technology

University of the Punjab

Textbooks

- ❑ **Probability & Statistics for Engineers & Scientists,**
Ninth Edition, Ronald E. Walpole, Raymond H.
Myer

References

Readings for these lecture notes:

- ❑ **Probability & Statistics for Engineers & Scientists**, Ninth edition, Ronald E. Walpole, Raymond H. Myer

These notes contain material from the above book.

Discrete Probability Distribution

□ The set of ordered pairs $(x, f(x))$ is a **probability function, probability mass function, or probability distribution** of the discrete random variable X if, for each possible outcome x ,

1. $f(x) \geq 0$,

2. $\sum_x f(x) = 1$,

3. $P(X = x) = f(x)$.

Example: A shipment of **20 similar laptop computers** to a retail outlet contains **3 that are defective**. If a school makes a random purchase of **2 of these computers**, **find the probability distribution** for the number of defectives.

$$N = 20$$

$$n = 2$$

$$k = 3$$

$$P(X = x) = h(x; N, n, k) = \frac{{}_k C_x {}_{N-k} C_{n-x}}{{}_N C_n}, \max\{0, n - (N - k)\} \leq x \leq \min\{n, k\}$$

Let X represent the number of defective computers

$$\begin{aligned} \max\{0, n - (N - k)\} &= \max\{0, 2 - (20 - 3)\} \\ &= \max(0, -17) = 0 \end{aligned}$$

$$\min\{n, k\} = \min(2, 3) = 2$$

Probability Distribution	
x	$P(X = x)$
0	$\frac{136}{191}$
1	$\frac{51}{190}$
2	$\frac{3}{190}$
	$\sum P(X) = 1$

Example : If a car agency **sells 50%** of its inventory of a certain foreign car equipped with side airbags, find a formula for the **probability distribution** of the number of cars with side airbags among the **next 4 cars** sold by the agency.

$$\mathbf{b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \dots, n}$$

Here $n = 4$, $p = 0.50$, $q = 0.50$

Let x denotes the number of cars with side airbags

$$\begin{aligned}\mathbf{b(x; 4, 0.50)} &= \binom{4}{x} (0.50)^x (0.50)^{4-x}, \quad x = 0, 1, 2, 3, 4 \\ &= \binom{4}{x} (0.50)^4, \quad x = 0, 1, 2, 3, 4\end{aligned}$$

or

$$\mathbf{b(x; 4, 0.50) = \frac{1}{16} \binom{4}{x}, \quad x = 0, 1, 2, 3, 4}$$

Cumulative Distribution Function

The **cumulative distribution function $F(x)$** of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \text{ for } -\infty < x < \infty$$

$$P(a < X < b) = F(b) - F(a)$$

Example: Let's consider an example of a **Cumulative Distribution Function (CDF)** for a simple discrete random variable: **rolling a fair six-sided die.**

Solution: Let X be the random variable representing the outcome of rolling the die. The CDF for X can be represented as:

$$F(x) = P(X \leq x)$$

x	1	2	3	4	5	6	
P(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\sum P(x) = \frac{6}{6} = 1$

$$\therefore F(x) = P(X \leq x)$$

Let X be the random variable representing the outcome of rolling the die.

$$F(x) = \begin{cases} 0, & \text{for } x < 1, \\ \frac{1}{6}, & \text{for } 1 \leq x < 2, \\ \frac{2}{6}, & \text{for } 2 \leq x < 3, \\ \frac{3}{6}, & \text{for } 3 \leq x < 4, \\ \frac{4}{6}, & \text{for } 4 \leq x < 5, \\ \frac{5}{6}, & \text{for } 5 \leq x < 6, \\ 1, & \text{for } x \geq 6. \end{cases}$$

Example A stockroom clerk returns **three safety helmets at random** to three steel mill employees who had previously checked them. If **Smith, Jones, and Brown**, in that order, receive one of the three hats, list the **sample points for the possible orders of returning the helmets**, and find the value **m** of the random variable **M** that represents the number of **correct matches**

If **S**, **J**, and **B** stand for **Smith's**, **Jones's**, and **Brown's** helmets, respectively, then the possible arrangements in which the helmets may be returned and the number of correct matches are

Sample space	m
SJB	3
SBJ	1
JSB	1
BJS	1
JBS	0
BSJ	0

$$\because F(x) = P(X \leq x)$$

For the **random variable M** , the number of correct matches in the previous example, we have

$$F(2) = P(M \leq 2) = f(0) + f(1) = \frac{2}{6} + \frac{3}{6} = \frac{5}{6}$$

The cumulative distribution function of M is

$$F(m) = \begin{cases} 0, & \text{for } m < 0, \\ \frac{2}{6} = \frac{1}{3}, & \text{for } 0 \leq m < 1, \\ \frac{5}{6}, & \text{for } 1 \leq m < 3, \\ 1, & \text{for } m \geq 3. \end{cases}$$

Example : Find the **cumulative distribution function** of the random variable X in $f(x) = \frac{1}{16} \binom{4}{x}$, $x = 0, 1, 2, 3, 4$. Using **$F(x)$** , verify that **$f(2) = 3/8$** .

$$f(x) = \frac{1}{16} \binom{4}{x}, \quad x = 0, 1, 2, 3, 4$$

$$f(0) = \frac{1}{16}$$

$$f(1) = \frac{4}{16}$$

$$f(2) = \frac{6}{16}$$

$$f(3) = \frac{4}{16}$$

$$f(4) = \frac{1}{16}$$

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \text{ for } -\infty < x < \infty$$

$$F(0) = P(X \leq 0) = f(0) = \frac{1}{16},$$

$$\begin{aligned} F(1) &= P(X \leq 1) = f(0) + f(1) \text{ -----(1)} \\ &= \frac{1}{16} + \frac{4}{16} = \frac{5}{16}, \end{aligned}$$

$$\begin{aligned} F(2) &= P(X \leq 2) = f(0) + f(1) + f(2) \text{ -----(2)} \\ &= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16}, \end{aligned}$$

$$\begin{aligned} F(3) &= P(X \leq 3) = f(0) + f(1) + f(2) + f(3) \\ &= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} \\ &= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{15}{16}, \end{aligned}$$

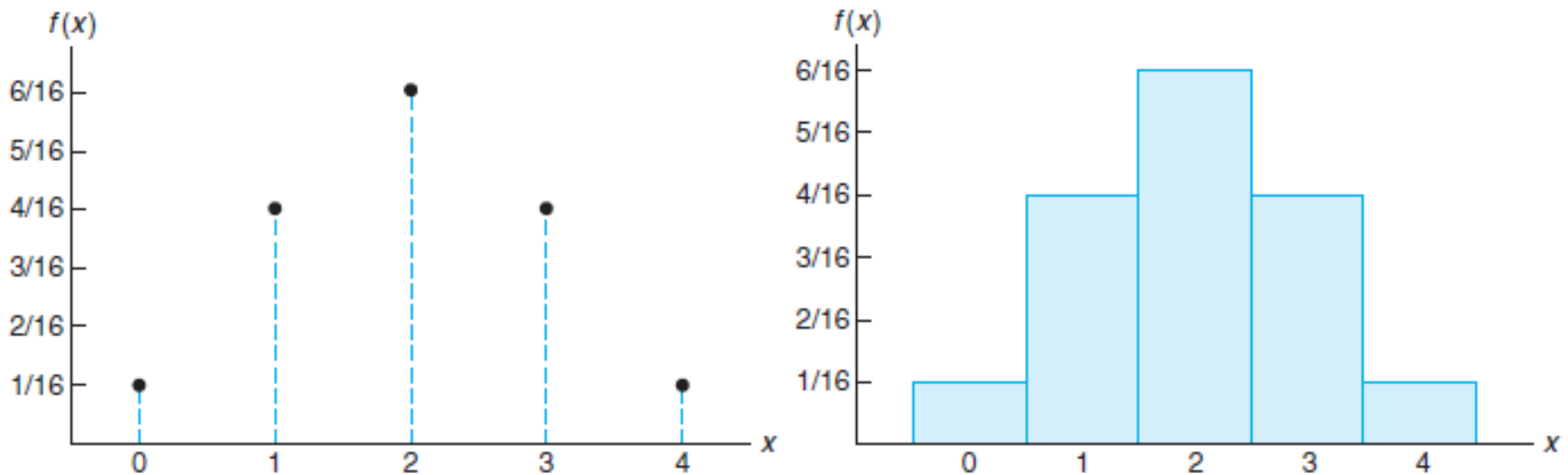
$$\begin{aligned}
 F(4) = P(X \leq 4) &= f(0) + f(1) + f(2) + f(3) + f(4) \\
 &= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} \\
 &= \frac{16}{16} = \mathbf{1}
 \end{aligned}$$

$$\therefore F(x) = \begin{cases} \mathbf{0}, & \text{for } x < 0, \\ \frac{\mathbf{1}}{\mathbf{16}}, & \text{for } 0 \leq x < 1, \\ \frac{\mathbf{5}}{\mathbf{16}}, & \text{for } 1 \leq x < 2, \\ \frac{\mathbf{11}}{\mathbf{16}}, & \text{for } 2 \leq x < 3, \\ \frac{\mathbf{15}}{\mathbf{16}}, & \text{for } 3 \leq x < 4, \\ \mathbf{1}, & \text{for } x \geq 4. \end{cases}$$

Using CDF, to find the probability
(2) – (1):

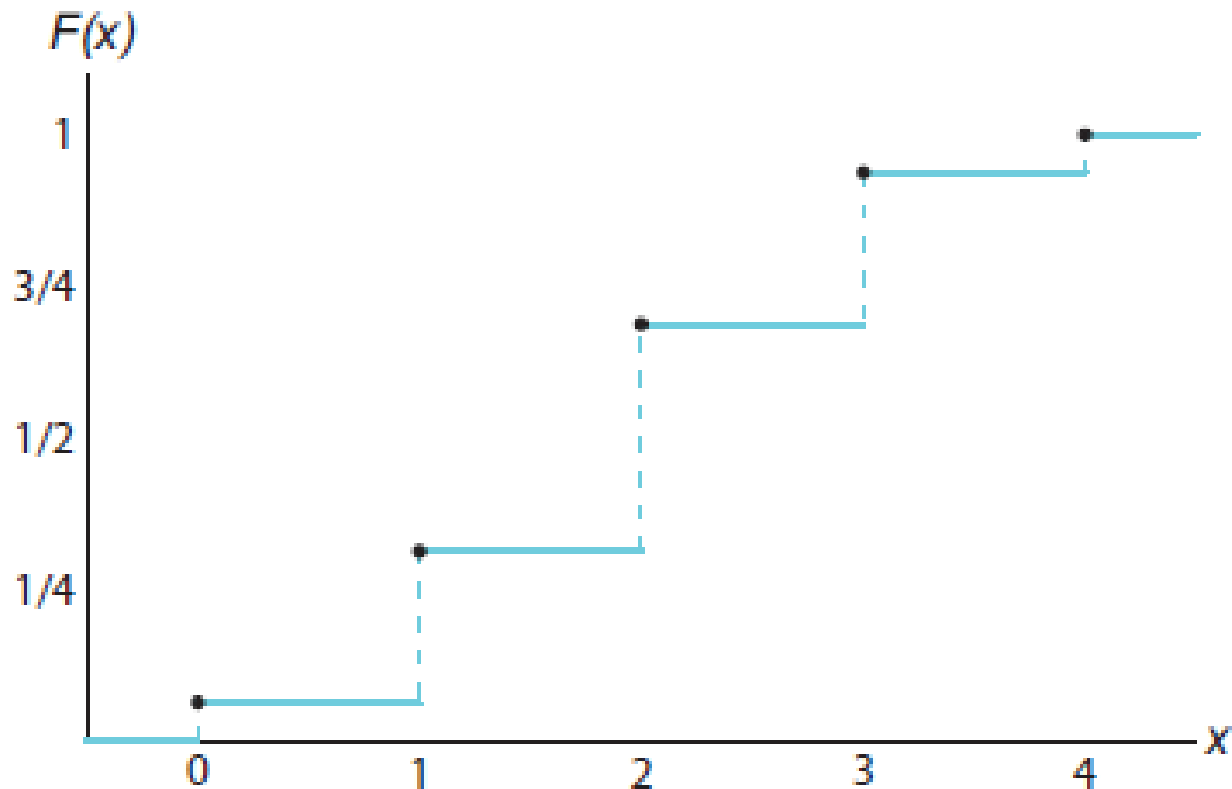
$$f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{6}{16} = \frac{3}{8}$$

Probability mass function plot vs. Probability histogram



Probability mass function plot vs. Probability histogram

Discrete cumulative distribution function



Discrete cumulative distribution function

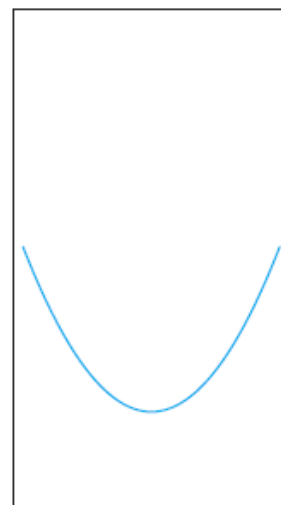
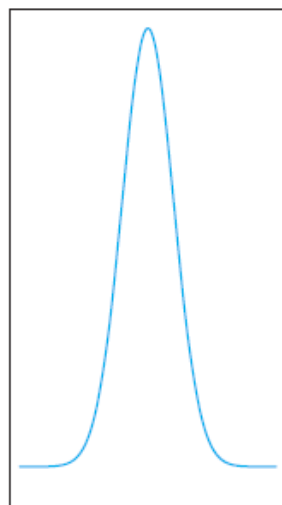
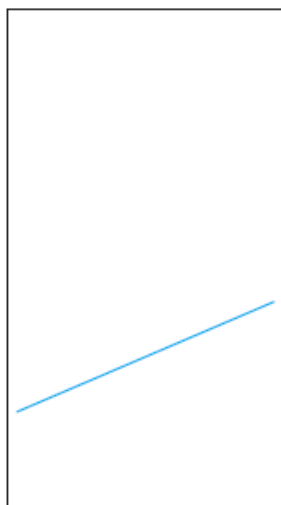
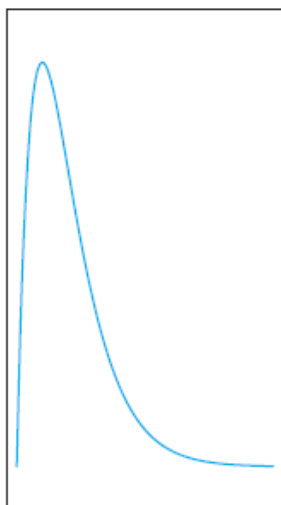
Continuous Probability Distributions

- ❑ A **continuous random variable** has a probability of **0** of assuming *exactly* any of its values.
- ❑ Consequently, its **probability distribution cannot** be given in **tabular form**.

Continuous Probability Distributions

- We shall concern ourselves with **computing probabilities for various intervals** of continuous random variables such as $P(a < X < b)$, $P(W \geq c)$, and so forth.
- Note that when X is continuous,
$$P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b).$$
- That is, it does not matter whether we include an endpoint of the interval or not.
- This is not true, though, when X is discrete.

- ❑ Because **areas** will be used to represent probabilities and probabilities are **positive numerical values**, the **density function** must lie entirely **above the x axis**.



Typical density functions.

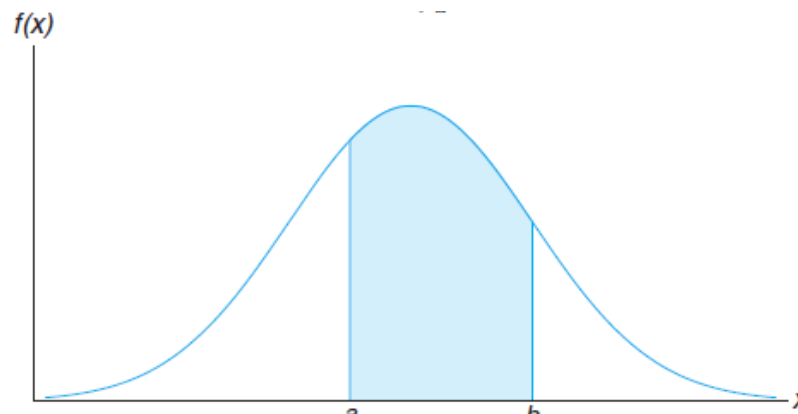
Probability Density Function

The function $f(x)$ is a **probability density function (pdf)** for the continuous random variable X , defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.

2. $\int_{-\infty}^{+\infty} f(x) dx = 1$.

3. $P(a < X < b) = \int_a^b f(x) dx$



Example: Suppose that the error in the reaction temperature, in $^{\circ}\text{C}$ for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere} \end{cases}$$

- (a) Verify that $f(x)$ is a **density function**.
- (b) Find $P(0 < X \leq 1)$.

Property 1: $f(x) \geq 0$

$$f(x) = \frac{x^2}{3}$$

$$f(-1) = \frac{(-1)^2}{3} = 0.33 > 0$$

$$f(2) = \frac{(2)^2}{3} = 1.33 > 0$$

Property 2: $\int_{-\infty}^{+\infty} f(x) dx = 1.$

$$\text{LHS} = \int_{-1}^2 \frac{x^2}{3} dx$$

$$= \left[\frac{x^3}{9} \right]_{-1}^2$$

$$= \frac{[(2)^3 - (-1)^3]}{9}$$

$$= 1$$

LHS = RHS

$$P(0 < X \leq 1) = \int_0^1 \frac{x^2}{3} dx$$

$$= \left[\frac{x^3}{9} \right]_0^1$$

$$= \frac{[(1)^3 - (0)^3]}{9}$$

$$= \frac{1}{9}$$

Cumulative Distribution Function

The **cumulative distribution function $F(x)$** of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \text{ for } -\infty < x < \infty$$

$P(a < X < b) = F(b) - F(a)$ and $f(x) = \frac{dF(x)}{dx}$, if the derivative exists.

Example: For the density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere} \end{cases}$$

, **find $F(x)$** , and use it to evaluate **$P(0 < X \leq 1)$** .

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \text{ for } -\infty < x < \infty$$

For $-1 < x < 2$,

$$\begin{aligned} F(x) &= \int_{-1}^x \frac{t^2}{3} dt \\ &= \left[\frac{t^3}{9} \right]_{-1}^x \\ &= \frac{[(x)^3 - (-1)^3]}{9} \\ &= \frac{x^3 + 1}{9} \end{aligned}$$

$$F(x) = \begin{cases} 0, & \text{for } x < -1, \\ \frac{x^3 + 1}{9}, & \text{for } -1 \leq x < 2, \\ 1, & \text{for } x \geq 2 \end{cases}$$

$$F(x) = \begin{cases} 0, & \text{for } x < -1, \\ \frac{x^3 + 1}{9}, & \text{for } -1 \leq x < 2, \\ 1, & \text{for } x \geq 2 \end{cases}$$

$$P(0 < X \leq 1) = F(1) - F(0)$$

$$F(1) = \frac{1^3 + 1}{9} = \frac{2}{9}$$

$$F(0) = \frac{0^3 + 1}{9} = \frac{1}{9}$$

$$P(0 < X \leq 1) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

Example: The **Department of Energy (DOE)** puts projects out on bid and generally estimates what a reasonable bid should be. Call the **estimate b** . The DOE has determined that the **density function** of the

winning (low) bid is $f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b, \\ 0, & \text{elsewhere} \end{cases}$

Find **$F(y)$** and use it to **determine the probability** that the **winning bid is less than** the DOE's preliminary **estimate b** .

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \text{ for } -\infty < x < \infty$$

$$\frac{2}{5}b \leq y \leq 2b$$

$$F(y) = \int_{\frac{2}{5}b}^y \frac{5}{8b} dt$$

$$= \left[\frac{5}{8b} t \right]_{\frac{2}{5}b}^y$$

$$= \frac{5}{8b} y - \frac{5}{8b} \left(\frac{2}{5} b \right)$$

$$= \frac{5}{8b} y - \frac{1}{4}$$

$$F(y) = \begin{cases} 0, & y < \frac{2}{5}b, \\ \frac{5}{8b}y - \frac{1}{4}, & \frac{2}{5}b \leq y \leq 2b \\ 1, & y \geq 2b. \end{cases}$$

To determine the probability that the **winning bid** is less than the **preliminary bid estimate b** , we have

$$F(y) = \frac{5}{8b} \mathbf{y} - \frac{1}{4}$$

$$\Rightarrow F(b) = \frac{5}{8b} \mathbf{b} - \frac{1}{4}$$

$$\Rightarrow F(b) = \frac{5}{8} - \frac{1}{4}$$

$$\therefore \mathbf{P(Y \leq b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}}$$

Consider a discrete uniform random variable X that takes values 1, 2, 3, 4, 5 with equal probability. The discrete uniform probability distribution is

$$P(x; 5) = \frac{1}{5}, x = 1, 2, 3, 4, 5$$

Find the CDF.

The cumulative distribution function (CDF) of discrete uniform distribution is

$$\therefore F(x) = \begin{cases} 0, & \text{for } x < 1, \\ \frac{1}{5}, & \text{for } 1 \leq x < 2, \\ \frac{2}{5}, & \text{for } 2 \leq x < 3, \\ \frac{3}{5}, & \text{for } 3 \leq x < 4, \\ \frac{4}{5}, & \text{for } 4 \leq x < 5, \\ 1, & \text{for } x \geq 5. \end{cases}$$