Introduction to Statistics

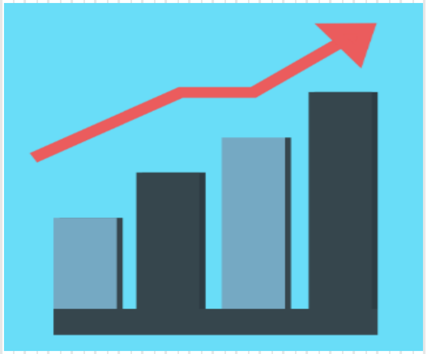
Dr. Farman Ali

Assistant Professor

DEPARTMENT OF SOFTWARE

SEJONG UNIVERSITY

Lecture-7



Course Syllabus

- Introduction to the course
- Sampling and data presentation
- Basic of probability
- Distributions
- Confidence intervals
- Hypothesis testing
- Correlation and simple linear regression
- Multiple regression

Outline

- **Random variables**
- **Discrete random variables**
- **Probability distribution of discrete random variables**
- **Distribution functions**
- **Examples of probability distribution of discrete random variables**
- **Continuous random variables**

Random Variable

- The outcome of an experiment need not be a number, for example, the outcome when a coin is tossed can be 'heads' or 'tails'.

- However, we often want to represent outcomes as numbers.

- A **random variable** is a function that associates a unique numerical value with every outcome of an experiment. The **value** of the random variable will vary from trial to trial as the experiment is repeated.

Examples

- A coin is tossed ten times. The random variable X is the number of tails that are noted. X can only take the values 0, 1, ..., 10.

- A light bulb is burned until it burns out. The random variable Y is its lifetime in hours. Y can take any positive real value.

Random Variable

- A **random variable** is abbreviated as r.v.
- The random variables are usually denoted by capital letters such as X, Y, Z; while the values taken by them are represented by the corresponding small letters such as x, y, z.
- It should be noted that more than one r.v. can be defined on the same sample space.

There are two types of r.v's:

- Discrete Random Variable
- Continuous Random Variable

Discrete Random Variable

- A random variable X is said to be discrete if it can assume values which are finite.
- When X takes on a finite number of values, they may be listed as x1, x2, …, xn

Examples:

- The number of heads obtained in coin tossing experiments
- The number of defective items observed in a consignment
- The number of fatal accidents

Probability Distribution of a Discrete Random Variable

- The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values.

Probability Distribution of a Discrete Random Variable

- The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values.

- More formally, Let X be a discrete r.v. taking on distinct values x1, x2, …., xn, ….

Then probability distribution function (pdf) of the r.v. X, denoted by p(x) or f(x), defined as:

$$f(x_i) = \begin{cases} P(X = x_i) & for\, i = 1, 2, \ldots, n, \ldots \\ 0 & for\, x \neq x_i \end{cases}$$

Note: The probability distribution is also called the probability function or the probability mass function.

Probability Distribution of a Discrete Random Variable

Properties:

1. $f\left(x_i\right) \geq 0, \qquad \text{for all } i$
2. $\sum_{i=1}^{n} f\left(x_i\right) = 1$

Distribution Function

Distribution Function

- It is a function giving the probability that the random variable X is less than or equal to x, for every value x.

- More formally, the distribution function of a random variable X taking value x, denoted by F(x) is defined as: $F(X=x)=P(X\leq x)$.
- The distribution function is abbreviated as d.f. and is also called Cumulative Distribution Function (c.d.f.) as it is the cumulative probability function of the X from the smallest up to specific value of x.

Since F(x) is a probability, so $F(-\infty) = F(\phi) = 0 \ \ and \ \ F(+\infty) = F(S) = 1$

If a and b are any two real numbers such that a<b. Then $P(a\leq X\leq b)=P(X\leq b)-P(X\leq a)=F(b)-F(a)$

Cumulative Distribution Function

Properties of Cumulative Distribution Function (CDF)

- $F(-\infty) = 0 \;\; and \;\; F(+\infty) = 1$
- F(x) is a non-decreasing function of x, i.e. $F(x_1) \leq F(x_2)$ if $x_1 \leq x_2$

Note: All random variables (discrete and continuous) have a cumulative distribution function.

Cumulative Distribution Function

Cumulative Distribution Function of a Discrete Random Variable is:

$$F(x) = P(X \le x_i) = \sum_{i} f(x_i)$$

Probability Distribution of a Discrete Random Variable

Example: Find the probability distribution and distribution function for the number of heads when 3 balanced coins are tossed.

Construct a probability histogram and a graph of the CDF.

Solution: Here S={HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

Let X= number of heads then x=0,1,2,3

$$f(0) = P(X=0) = P\left[\{TTT\}\right] = 1/8$$

$$f(1) = P(X=1) = P\left[\{HTT, THT, TTH\}\right] = 3/8$$

$$f(2) = P(X=2) = P\left[\{HHT, HTH, THH\}\right] = 3/8$$

$$f(3) = P(X=3) = P\left[\{HHH\}\right] = 1/8$$

Probability Distribution of a Discrete Random Variable

Example: Find the probability distribution and distribution function for the number of heads when 3 balanced coins are tossed.

Construct a probability histogram and a graph of the CDF.

Solution: Here S={HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

Let X= number of heads then x=0,1,2,3

$$f(0)=P(X=0)=P\left[\{TTT\}\right]=1/8$$

$$f(1)=P(X=1)=P\left[\{HTT,THT,TTH\}\right]=3/8$$

$$f(2)=P(X=2)=P\left[\{HHT,HTH,THH\}\right]=3/8$$

$$f(3)=P(X=3)=P\left[\{HHH\}\right]=1/8$$

No of heads (xi)	Probability P(xi) or f(xi)
0	1/8
1	3/8
2	3/8
3	1/8
Total	**1**

Probability Distribution of a Discrete Random Variable

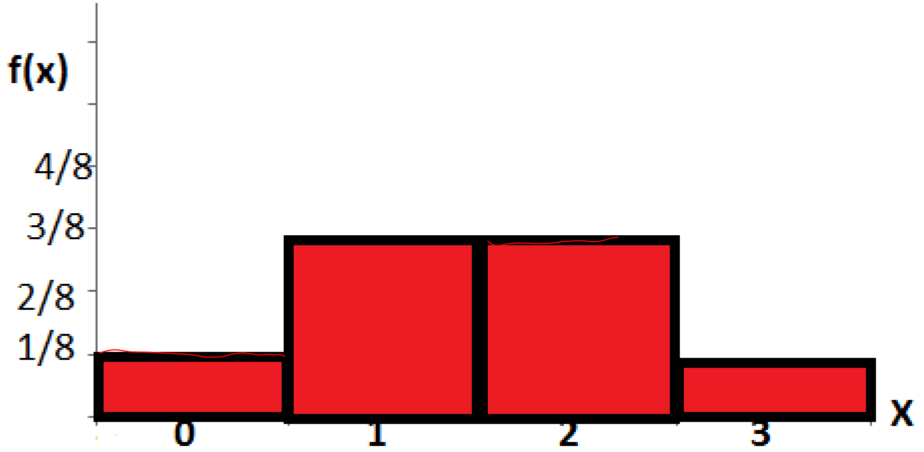
Example: Find the probability distribution and distribution function for the number of heads when 3 balanced coins are tossed.

Construct a probability histogram and a graph of the CDF.

Solution: Here S={HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

Let X= number of heads then x=0,1,2,3

The probability histogram is given by:



No of heads (xi)	Probability P(xi) or f(xi)
0	1/8
1	3/8
2	3/8
3	1/8
Total	**1**

Probability Distribution of a Discrete Random Variable

Example: Find the probability distribution and distribution function for the number of heads when 3 balanced coins are tossed.

Construct a probability histogram and a graph of the CDF.

Solution:

The CDF of X is:

(xi)	f(xi)	Distribution Function F(xi)=P(X<=xi)
0	1/8	P(X<=0)=1/8
1	3/8	P(X<=1)=P(X=0)+P(X=1)=1/8+3/8=4/8
2	3/8	P(X<=2)=P(X=0)+P(X=1) +P(X=2)=1/8+3/8+3/8=7/8
3	1/8	P(X<=3)=P(X=0)+P(X=1)+P(X=2) +P(X=3)=1/8+3/8+3/8+1/8=8/8=1

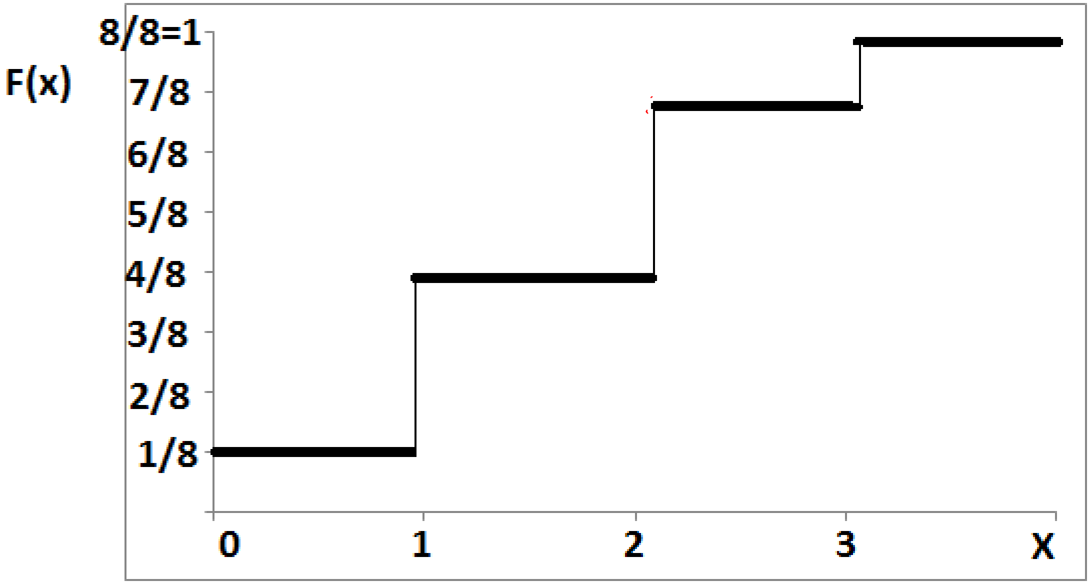
Probability Distribution of a Discrete Random Variable

Example: Find the probability distribution and distribution function for the number of heads when 3 balanced coins are tossed.

Construct a probability histogram and a graph of the CDF.

Solution:

The graph of CDF of X is:



(xi)	f(xi)	F(xi)=P(X<=xi)
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1

Probability Distribution of a Discrete Random Variable

Example: Find the probability distribution and distribution function for the sum of dots when two fair dice are thrown.

Solution: Here n(S)=36

Let X= Sum of dots,

then x=2, 3, 4, …, 11, 12

$$S=\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),$$
$$(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),$$
$$(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),$$
$$(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),$$
$$(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),$$
$$(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

Probability Distribution of a Discrete Random Variable

Example: Find the probability distribution and distribution function for the sum of dots when two fair dice are thrown.

x_i	$f(x_i)$	$F(x_i)=P(X<=x_i)$
2	1/36	1/36
3	2/36	3/36
4	3/36	6/36
5	4/36	10/36
6	5/36	15/36
7	6/36	21/36
8	5/36	26/36
9	4/36	30/36
10	3/36	33/36
11	2/36	35/36
12	1/36	36/36=1

$$
\begin{aligned}
S = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\
&(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\
&(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\
&(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\
&(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\
&(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}
\end{aligned}
$$

Continuous Random Variable

A random variable X is said to be continuous if it can assume every possible value in an interval [a, b], a<b.

Examples:

- The height of a person
- The temperature at a place
- The amount of rainfall

Probability Density Function of a Continuous Random Variable

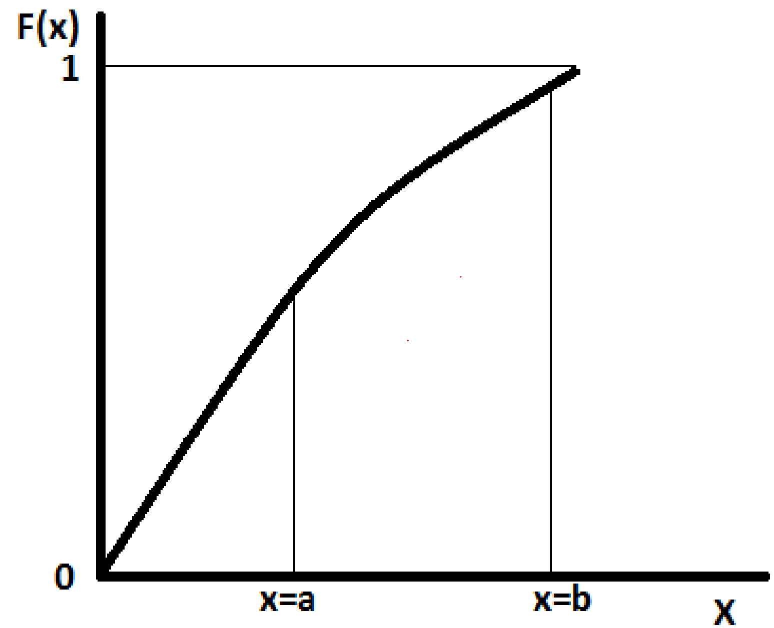
- The probability density function of a continuous random variable is a function which can be integrated to obtain the probability that the random variable takes a value in a given interval.

- More formally, the probability density function, f(x), of a continuous random variable X is the derivative of the cumulative distribution function F(x), i.e.

$$f(x) = \frac{d}{dx} F(x)$$

Where,

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$



$$P(a \le X \le b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a)$$

Probability Density Function of a Continuous Random Variable

Properties:

1. $f\left(x_i\right) \geq 0, \qquad for\ all\ x_i$

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

Note: The probability of a continuous r.v. X taking any particular value ‘k’ is always zero.

$$P\left(X = k\right) = \int_{k}^{k} f\left(x\right) dx$$

That is why probability for a continuous r.v. is measurable only over a given interval.

Probability Density Function of a Continuous Random Variable

Example: Find the value of k so that the function f(x) defined as follows, may be a density function.

$$f(x)=\begin{cases} kx & , \quad 0\le x\le 2 \\ 0 & , \quad otherwise \end{cases}$$

Solution:

Probability Density Function of a Continuous Random Variable

Example: Find the value of k so that the function f(x) defined as follows, may be a density function.

$$f(x) = \begin{cases} kx & , \quad 0 \le x \le 2 \\ 0 & , \quad otherwise \end{cases}$$

Solution: Since we have, $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

So,

$$\int_0^2 (kx)\,dx = 1 \Rightarrow k\int_0^2 (x)\,dx = 1 \Rightarrow k\left|\frac{x^2}{2}\right|_0^2 = 1 \Rightarrow k\left(\frac{(2)^2}{2} - \frac{(0)^2}{2}\right) = 1$$

$$\Rightarrow 2k = 1 \Rightarrow k = 1/2$$

Hence the density function becomes, $f(x) = \begin{cases} \dfrac{1}{2}x & , \quad 0 \le x \le 2 \\ 0 & , \quad otherwise \end{cases}$

Probability Density Function of a Continuous Random Variable

Example: Find the distribution function of the following probability density function.

$$f(x)=\begin{cases}\frac{1}{2}x & , \quad 0\le x\le 2\\ 0 & , \quad otherwise\end{cases}$$

Solution: The distribution function is: $F(x)=P(X\le x)=\int_{-\infty}^{x} f(x)\,dx$

So,

$$For -\infty < x \le 0, F(x)=\int_{-\infty}^{x} f(x)\,dx=\int_{-\infty}^{x}(0)\,dx=0$$

$$For\ 0 < x \le 2, F(x)=\int_{-\infty}^{0} f(x)\,dx+\int_{0}^{x} f(x)\,dx=\int_{-\infty}^{0}(0)\,dx+\int_{0}^{x}\frac{x}{2}\,dx=\frac{x^2}{4}$$

$$For\ x > 2, F(x)=\int_{-\infty}^{0} f(x)\,dx+\int_{0}^{2} f(x)\,dx+\int_{2}^{x} f(x)\,dx=\int_{-\infty}^{0}(0)\,dx+\int_{0}^{2}\frac{x}{2}\,dx+\int_{2}^{x}(0)\,dx=1$$

Probability Density Function of a Continuous Random Variable

$$For\ -\infty < x \le 0, F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{x} (0)\,dx = 0$$

$$For\ 0 \le x \le 2, F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{x} f(x)\,dx = \int_{-\infty}^{0} (0)\,dx + \int_{0}^{x} \frac{x}{2}\,dx = \frac{x^2}{4}$$

$$For\ x > 2, F(x) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{2} f(x)\,dx + \int_{2}^{x} f(x)\,dx = \int_{-\infty}^{0} (0)\,dx + \int_{0}^{2} \frac{x}{2}\,dx + \int_{2}^{x} (0)\,dx = 1$$

So the distribution function is:

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ \dfrac{x^2}{4} & , \quad 0 \le x \le 2 \\ 1 & , \quad x > 2 \end{cases}$$

Probability Density Function of a Continuous Random Variable

Example: A r.v. X is of continuous type with p.d.f.

$$f(x) = \begin{cases} 2x & , \quad 0 < x < 1 \\ 0 & , \quad otherwise \end{cases}$$

Calculate:

- P(X=1/2)
- P(X<=1/2)
- P(X>1/4)
- P(1/4<=X<=1/2)
- P(X<=1/2 | 1/3<=X<=2/3)

Summary

- Random variables
- Discrete random variables
- Probability distribution of discrete random variables
- Distribution functions
- Examples of probability distribution of discrete random variables
- Continuous random variables

Thank You !