# LESSON FIVE: PROBABILITY DISTRIBUTIONS

## 5.1 Introduction

- Probability is the likelihood or chance that a particular event will occur.
- In probability and statistics the term experiment refers to any procedure that gives rise to a collection of outcomes which cannot be predetermined.
- In tossing a coin, the possible outcomes are as follows:

  Tossing 1 coin: $\{H, T\}$

  Tossing 2 coins: $\{HH, HT, TH, TT\}$

  Tossing 3 coins: $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

- ✓ The set of all possible outcomes in an experiment is called a **sample space**.
- ✓ An **event** is a subset of the sample space.

**EXAMPLE**

Let the set of all outcomes (sample space) in the experiment of tossing two coins be $\{HH, HT, TH, TT\}$. Then

A= $\{HT, TH\}$ is the event of getting just one head/tail

B= $\{HH, HT, TH\}$ is the event of getting atleast one head

$\phi$= $\{\quad\}$ is the the impossible event

S = $\{HH, HT, TH, TT\}$ is the sure event

- An **elementary event or simple event** is the event containing only one point of the sample space E.G: In the Toss of two coins, the following are elementary events:

  $\{HH\}, \{HT\}, \{TH\}, \{TT\}$.

- A **random variable** is a function which assigns a numerical value to each simple event in a sample space.

  **Example**

  Suppose that three students are selected at random from a class and each is asked whether he smokes (S) or he does not (N). Then the sample space of this experiment is given by

  $$S = \{SSS, SSN, SNS, SNN, NSS, NSN, NNS, NNN\}$$

- Let $X$ denote the number of smokers among the three students chosen. Then:

| Simple event in S | Random variable $X$ |
|---|---|
| SSS | 3 |
| SSN | 2 |
| SNS | 2 |
| SNN | 1 |
| NSS | 2 |
| NSN | 1 |
| NNS | 1 |
| NNN | 0 |

Thus $X$ is a random variable which takes the values 0, 1, 2, or 3.

- If a random variable can assume only a countable number of distinct values, it is called a **discrete random variable**.

  E.G: The number of children in a family, the number of telephone calls at a switchboard in ten minutes period etc.

- A **continuous random variable** is one that can assume any value within a given time interval.

  E.G: Lifetime of an electric bulb, weight of a person etc.

## 5.3 Probability distribution function of a discrete random variable

- The probability distribution of a random variable can be described by using all the values that a random variable can together with the corresponding probabilities. Such a listing is called a **probability distribution** or **probability mass function** of the random variable.

  **Example**

  Suppose $X$ represents the number of heads in a random experiment of tossing three coins.

  The sample space is:

  $$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

  The probability distribution of the random variable $X$ defined as the "number of heads" is

| $x$ | $P(X=x)$ |
|---|---|
| 0 | $\frac{1}{8}$ |
| 1 | $\frac{3}{8}$ |
| 2 | $\frac{3}{8}$ |
| 3 | $\frac{1}{8}$ |

- In general, suppose $X$ is a random variable that assumes the values $x_1$ , $x_2$, …, $x_k$. if we represent the probability that $X$ assumes the value $x_i$ by $P(X=x_i)$, then the probability function can be given in the form of a table as

| $X$ | $P(x)$ |
|---|---|
| $x_1$ | $p(x_1)$ |
| $x_2$ | $P(x_2)$ |
| . | . |
| . | . |
| . | . |
| $x_k$ | $P(x_k)$ |
| | **Sum = 1** |

- The sum of the probabilities, i.e. $\sum_{i=1}^{k} p(x_i) = p(x_1) + P(x_2) + ... + P(x_k)$ is one.

**Conditions for a function to be a probability function**

i) The probability that a random variable assumes a value $x_i$ is always between 0 and 1,

i.e. $0 \le p(x_i) \le 1$

ii) The sum of all probabilities is equal to one, i.e. $\sum_{i=1}^{k} p(x_i) = 1$

**Example**

The number of telephone calls received in an office between 9 – 10 am has the probability distribution as shown below:

| Number of calls ($X$) | Probability, $P(x)$ |
|---|---|
| 0 | 0.05 |
| 1 | 0.20 |
| 2 | 0.25 |
| 3 | 0.20 |
| 4 | 0.10 |
| 5 | 0.15 |
| 6 | 0.05 |

a) Verify that it is a probability function

b) Find the probability that there will be 3 or more calls

c) Find the probability that there will be an even number of calls

**The Mean or Expected Value of a Discrete Random Variable**

- It is obtained by multiplying each possible value of the random variable by the corresponding probability and summing the terms. That is, if $x_1, x_2, ...x_n$ are the values assumed by a random variable with respective probabilities $p(x_1), p(x_2), ...p(x_n)$, then its mean $\mu$ (also called the expected value) is given by

$$\mu = x_1 p(x_1) + x_2 p(x_2) + ...x_n p(x_n)$$
$$= \sum_{i=1}^{n} x_i p(x_i)$$

The mean $\mu$ is also referred to as the expected value is denoted by $E(x)$.

**The Variance of a Discrete Random Variable**

- The variance of a discrete random variable is defined as

$$Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 .p(x_i)$$

- The positive square root of the variance is called the **standard deviation** of the random variable. The variance is commonly denoted as $\delta^2$, hence the standard deviation equals $\delta$.

**Example**

Suppose we are given the following data relating to the breakdown of a machine in a certain company during a given week, where *x* represents the number of breakdowns of the machine and *P(x)* represents the probability value of *x*.

| *x* | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| *P(x)* | 0.12 | 0.20 | 0.25 | 0.30 | 0.13 |

Find the mean and the variance of the number of breakdowns per week for this machine

**NB:** The computations of $\delta^2$ can be simplified by using the following version of the formulae:

$$\delta^2 = \sum x^2.P(x) - \mu^2$$

## 5.4 Discrete Probability Distributions

**Binomial Probability Distribution**

**Characteristics**

i) An outcome on each trial of an experiment is classified into one of two mutually exclusive categories; a success or a failure.

i) The probability of a success (*p*) remains the same from trial to trial and so does the probability of a failure (*q*), where $p + q = 1$.

ii) The trials are independent i.e. the outcome of one trial does not affect the outcome of any other trial.

- We are interested in the random variable *x*, where *x* is the number of successes in *n* trials.
- It is common to refer to each trial as a Bernoulli trial and to refer to the entire experiment as a binomial experiment.
- Given a Bernoulli Process where the probability of success in any trial equals *p* and the probability of a failure equals *q*, the probability of *x* successes in *n* trials is calculated as

$$p(n,x) = \binom{n}{x} p^x q^{n-x}$$

The mean of a Binomial distribution $\mu = np$

The variance $\delta^2 = npq$

**Example**

There are five flights daily from Moi International airport to Jomo Kenyatta International airport. Suppose the probability that any flight arrives late is 0.2. What is the probability that: -

i) None of the flights are late today?

ii) Exactly one of the flights is late today?

## 5.5 Continuous Probability Distributions

**Normal Probability Distribution**

**Characteristics**

- It is bell shaped and has a single peak at the center of the distribution
- The arithmetic mean, median and mode of the distribution are equal and located at the peak.
- Half of the area under the curve is above this center point and the other half is below it.
- It is symmetrical about its mean i.e. if it is cut vertically at the central value, the two halves will be mirror images
- It is asymptotic i.e. the curve gets closer and closer to the x-axis but never actually touches it.
- Since the normal distribution is a continuous distribution, the probabilities are given in terms of appropriate areas, and the total area under the curve is equal to 1. Thus the probability that a random variable $X$ having a normal distribution will assume a value between two numbers **a** and **b** is equal to the area under the curve between $\boldsymbol{x} = \mathbf{a}$ and $\boldsymbol{x} = \mathbf{b}$, as shown below:

**The standard normal probability distribution (μ = 0, δ = 1)**

- The standard normal curve describes the distribution of a normal random variable with mean zero and standard deviation 1. The random variable itself is called the standard normal variable and is denoted by Z.

  E.G: To find the area between z = 0 and z = 1.73, we go to 1.7 in the column and 0.03 in the row and read the corresponding entry as 0.4582. Hence the area between 0 and 1.73 is 0.4582 and $P(0 \le z \le 1.73) = 0.4582$

**NB:**

i) The curve is symmetrical w.r.t the vertical axis through zero

ii) It is strongly recommended that we sketch the curves and identify the areas under the curve and the values along the horizontal axis.

**EXAMPLES**

1. If $P(0 \leq z \leq c) = 0.3944$. Find *c*.
2. Find $P(-2.42 \leq z \leq 0.8)$
3. Find a) $P(1.8 \leq z \leq 2.8)$ b) $P(-2.8 \leq z \leq -1.8)$
4. Find a) $P(z > -2.13$ b) $P(z < -1.81$
5. Suppose *z* is a standard normal variable. In each of the following cases find *c* for which
   a) $P(z \leq c) = 0.1151$
   b) $P(z \leq c) = 0.8238$
   c) $P(1 \leq z \leq c) = 0.1525$
   d) $P(-c < z < c) = 0.8164$

- Having considered areas under the standard normal curve, we now consider the general case of a normal distribution with any mean $\mu$ and any standard deviation $\delta$, where $\delta > 0$.
- If *X* is a normal random variable with mean $\mu$ and standard deviation $\delta$, then *X* can be converted into a standard normal variable *z* by setting $z = \dfrac{X - \mu}{\delta}$

**EXAMPLE 6**

Suppose *X* has a normal distribution with $\mu$ = 30 and $\delta = 4$. Find

a) $P(30 < X < 35)$ b) $P(X > 40)$ c) $P(X < 22)$

## 5.6 Activities

1. A salesman who sells cars for General Motors claims that he sells the largest number of cars on Saturday. He has the following probability distribution for the number of cars he expects to sell on a particular Saturday.

| No. of cars (x) | Probability P(x) |
|---|---|
| 0 | .1 |
| 1 | .2 |
| 2 | .3 |
| 3 | .3 |
| 4 | .1 |
| **Total** | **1.0** |

i) On a typical Saturday, how many cars does the salesman expect to sell?

ii) What is the variance of the distribution?

2. In a recent survey, 90% of the homes in a city were found to have colored TV's. In a sample of nine homes, what is the probability that:
   i. All nine have colored TV's?
   ii. Less than five have colored TV's?
   iii. More than five have colored TV's?
   iv. At least seven homes have colored TV's?
3. The life times of electric components manufactured by Raman Industries Ltd are normally distributed with mean of 2500 hours and standard deviation of 600 hours. If the daily production is 500 components, how many are expected to have a life time of:
   i) Less than 2600 hours
   ii) Between 2350 hours and 2580 hours
   iii) More than 2380 hours