Chapter- 4

**Probability Distributions**

**Objectives:**

* To introduce probability distributions most commonly used in decision making;
* To show which probability distribution to use and how to find its values; and
* To understand the limitations of each of the probability distributions you use.

**Chapter Contents:**

* Basic Terms Introduced in this Chapter
* What is Probability Distribution?
* Random Variables;
* Use of Expected Value in Decision Making;
* The Binomial Distribution;
* The Poisson Distribution;
* The Normal Distribution;
* Choosing the Correct Probability Distribution.

**Basic Terms Introduced in this Chapter**

***Probability Distribution:***

A list of the outcomes of an experiment with the probabilities we would expect to see associated with these outcomes is called probability distribution.

***Discrete Probability Distribution:***

A probability distribution in which the variable is allowed to take on only a limited number of values, which can be listed, is called discrete probability distribution.

***Random Variable:***

A variable whose values are determined by chance is called random variable.

***Continuous Random Variable:***

A random variable allowed to take on any value within a given range is called continuous random variable.

***Discrete Random Variable:***

A random variable that is allowed to take on only a limited number of values, which can be listed is called discrete random variable.

***Expected Value:***

A weighted average of the outcomes of an experiment is called expected value.

***Binomial Distribution:***

A discrete distribution describing the results of an experiment is known as binomial distribution.

***Poisson Distribution:***

A discrete distribution in which the probability of the occurrence of an event within a very small time period is a very small number, the probability that two or more such events will occur within the same time interval is effectively 0, and the probability of the occurrence of the event within one time period is independent of where that time period is.

***Normal Distribution:***

A distribution of a continuous random variable with a single- peaked, bell- shaped curve. The mean lies at the center of the distribution, and the curve is symmetrical around a vertical line erected at the mean. The two tails extend indefinitely, never touching the horizontal axis.

***Standard Normal Probability Distribution:***

A normal probability distribution, with mean μ = 0 and standard deviation σ = 1 is called standard normal probability distribution.

**Theoretical or Expected Frequency Distributions**

Following are various types of theoretical or expected frequency distributions:

1. Binomial Distribution,
2. Multinomial Distribution,
3. Negative Binomial Distribution,
4. Poisson Distribution,
5. Hypergeometric Distribution, and
6. Normal Distribution.

Amongst these the first five distributions are of discrete type and the last one is of continuous type. In these six distributions binomial, poisson and normal distributions have much more wider application in practice. So we shall discuss these three.

**Binomial Distribution**

The binomial distribution describes discrete, not continuous, data, resulting from an experiment known as a Bernoulli process, after the 17th century Swiss mathematician Jacob Bernoulli. The tossing of a fair coin a fixed number of times is a Bernoulli process, and the outcomes of such tosses can be represented by the binomial probability distribution. The success or failure of interviewees on an aptitude test may also be described by a Bernoulli process.

**Use of the Bernoulli Process:**

We can use the outcomes of a fixed number of tosses of a fair coin as an example of a Bernoulli process. We can describe this process as follows:

1. Each trial has only two possible outcomes: heads or tails, yes or no, success or failure.
2. The probability of the outcomes of any trial remains fixed over time. With a fair coin, the probability of heads remains 0.5 for each toss regardless of the number of times the coin is tossed.
3. The trials are statistically independent; that is, the outcome of one toss does not affect the outcome of any other toss.

**Binomial Formula:**

Probability of r successes in n trials = $\frac{n!}{r!(n-r)!}p^{r}q^{n-r}$

Where p = characteristic probability or probability of success

q = (1-p) = probability of failure

r = number of successes desired

n = number of trials undertaken.

**Example:** ***Calculate the chances (probability) of getting exactly two heads (in any order) on three tosses of a fair coin.***

**Solution:** We can use the above binomial formula to calculate desired probability. For this we can express the values as follows:

p = characteristic probability or probability of success = 0.5

q = (1-p) = probability of failure = 0.5

r = number of successes desired = 2

n = number of trials undertaken = 3

Probability of 2 successes (heads) in 3 trials = $\frac{3!}{2!(3-2)!}0.5^{2}0.5^{(3-2)}$

 = $\frac{3×2×1}{(2×1)(1×1)}0.5^{2}0.5^{1}$

 = $3×0.25×0.5$ = 0.375

Thus, there is a 0.375 probability of getting two heads on three tosses of a fair coin.

**Mean of a Binomial Distribution, *μ = np***

Where

n = number of trials

p = probability of success

**Standard Deviation of Binomial Distribution, *σ =*** $\sqrt{npq}$

Where

n = number of trials

p = probability of success

q = probability of failure = 1- p

**Example:** ***A packaging machine that produces 20 percent defective packages. If we take a random sample of 10 packages, what is the mean and standard deviation of the binomial distribution?***

**Solution:** Mean, μ = np = 10×0.2 = 2

 Standard Deviation, σ = $\sqrt{npq}=\sqrt{10×0.2×0.8}$ = $\sqrt{1.6}$ = 1.265

**The Poisson Distribution**

It is a discrete probability distribution developed by a French mathematician Simeon Denis Poisson. It may be expected in cases where the chance of any individual event being a success is small. This distribution is used to describe the behaviour of rare events such as the number of accidents on road, number of printing mistakes in a book, etc., and has been called “*the law of improbable events*”.

**Poisson Formula:**

Probability of exactly X occurrences, ***P(X) =*** $\frac{λ^{n}e^{-λ}}{x!}$

Where

 $λ^{n}= $lambda (the mean number of occurrences per interval of time) raised to the power x

$e^{-λ}= $e, or 2.71828 (the base of the Napierian, or natural, logarithm system), raised to the power negative lambda,

 $x!$ = x factorial.

**Example:** ***Suppose that we are investigating the safety of a dangerous intersection. Past police records indicate a mean of five accidents per month at this intersection. The number of accidents is distributed according to a Poisson distribution, and the Highway Safety Division wants us to calculate the probability in any month of exactly 0, 1, 2, 3, or 4 accidents.***

Solution: Using the Poisson formula, we can calculate the probability of no accidents:

P(0) = $\frac{λ^{n}e^{-λ}}{x!}$ = $\frac{5^{0}e^{-5}}{0!}$ = $\frac{(1)(0.0067)}{1}$ = 0.00674

 For exactly one accident:

P(1) = $\frac{λ^{n}e^{-λ}}{x!}$ = $\frac{5^{1}e^{-5}}{1!}$ = $\frac{(5)(0.0067)}{1}$ = 0.03370

For exactly two accidents:

P(2) = $\frac{λ^{n}e^{-λ}}{x!}$ = $\frac{5^{2}e^{-5}}{2!}$ = $\frac{(25)(0.0067)}{2×1}$ = 0.08425

For exactly three accidents:

P(3) = $\frac{λ^{n}e^{-λ}}{x!}$ = $\frac{5^{3}e^{-5}}{3!}$ = $\frac{(125)(0.0067)}{3×2×1}$ = 0.14042

For exactly four accidents:

P(4) = $\frac{λ^{n}e^{-λ}}{x!}$ = $\frac{5^{4}e^{-5}}{4!}$ = $\frac{(625)(0.0067)}{4×3×2×1}$ = 0.17552

 Our calculations will answer several questions. If we want to know the probability of 0, 1, or 2 accidents in any month, we can add these probabilities as:

P(0,1, or 2) = P(0) + P(1) + P(2)

 = 0.00674 + 0.03370 + 0.08425 = 0.12469

For, P(3 or fewer) = P(0, 1, 2, or 3) = P(0) + P(1) + P(2) + P(3)

 = 0.00674 + 0.03370 + 0.08425 + 0.14042 = 0.26511

If we want to calculate the probability of more than three then we must be 0.73489 (1- 0.26511).

**Important Point:**

*The poisson distribution is a good approximation of the binomial distribution when n is greater than or equal to 20 and p is less than or equal to 0.05.*

**The Normal Distribution**

It is a continuous probability distribution developed by Karl Gauss. The normal probability distribution is often called Gaussian distribution.

The normal curve is represented in several forms. The following is the basic form relating to the curve with mean μ and standard deviation σ:

**The Normal Distribution, *P(X) =*** $\frac{1}{σ√2π}$***.***$e^{\frac{-(x-μ)}{2σ^{2}}}$

Where

*X = values of the continuous random variable*

*μ = mean of the normal random variable*

*e = mathematical constant (= 2.7183)*

*π = mathematical constant (= 3.1416)*

**Characteristics (Graph) of Normal Probability Distribution:**

* The curve has a single peak; thus it is unimodal.
* The normal curve is” bell- shaped” and symmetric.
* For a normal probability distribution, mean median and mode all are equal.
* The two tails of the normal probability distribution extend indefinitely and never touch the horizontal axis.

**Areas under the Normal Curve:**

No matter what the values of mean (π) and standard deviation (σ) are for a normal probability distribution, the total area under the normal curve is 1.00. Mathematically it is true that-

* Approximately 68% of all the values in a normally distributed population lie within ±1σ from mean (π);
* Approximately 95.5% of all the values in a normally distributed population lie within ±2σ from mean (π); and
* Approximately 99.7% of all the values in a normally distributed population lie within ±3σ from mean (π). These are shown in the following graph:



**Formula for measuring distances under normal curve:**

**Standardizing a Normal Random Variable,**

*Z =* $\frac{X-μ}{σ}$

Where

*x = value of the random variable with which we are concerned;*

*μ = mean of the distribution of this random variable;*

*σ = standard deviation of this distribution;*

*z = number of standard deviations from x to the mean of this distribution.*

**Example 1: *What is the probability that a participant selected at random will require more than 500 hours to complete the training program?***

**Solution:** We can see that half of the area under the curve is located on either side of the mean of 500 hours. Thus, we can deduce that the probability that the random variable will take on a value higher than 500 is half, or 0.5.

**Example 2: *What is the probability that a candidate selected at random will take between 500 and 650 hours to complete the training program?***

**Solution:** The probability that will answer this question is the area between the mean (π = 500 hours) and the x value in which we are interested (650 hours). Using equation, we get a z value of

*Z =* $\frac{X-μ}{σ}$ *=* $\frac{650-500}{100} $*=* $\frac{150}{100}$ *=1.5 standard deviation*

If we look up z = 1.5 in Z- table, we find a probability of 0.4332. Thus, the chance that a candidate selected at random would require between 500 and 650 hours to complete the training program is slightly higher than 0.4.

**Example 3: *What is the probability that a candidate selected at random will take more than 700 hours to complete the training program?***

**Solution:** This situation is different from the above example 2. We are interested in the area to the right of the value 700 hours. So, first we will find out z value by using the formula-

Z = $\frac{X-μ}{σ}$ = $\frac{700-500}{100} $= $\frac{200}{100}$ =2.0 standard deviation

Looking in the Z- table for z value of 2.0, we find a probability of 0.4772. That represents the probability the program will require between 500 and 700 hours. But, we have to find out the probability that take more than 700 hours. Because the right half of the curve (between the mean and the right- hand tail) represents a probability of 0.5, we can get our answer (the area to the right of the 700- hour point) if we subtract 0.4772from 0.5; 0.5000- 0.4772 = 0.0228. Therefore, there are just over 2 percent chances (or 2 out of 100) that a participant chosen at random would take more than 700 hours to complete the course.

**Example 4:** ***Suppose the training program director wants to know the probability that a participant chosen at random would require between 550 and 650 hours to complete the required work.***

**Solution:** First calculate a z value for the 650 hour point, as follows:

Z = $\frac{X-μ}{σ}$ = $\frac{650-500}{100} $= $\frac{150}{100}$ =1.5 standard deviation

When we look up a z of 1.5 in z table, we see a probability value of 0.4332 (the probability that the random variable will fall between the mean and 650 hours). Now we calculate a z value for 550 hours as follows:

Z = $\frac{X-μ}{σ}$ = $\frac{150-500}{100} $= $\frac{50}{100}$ =0.5 standard deviation

When we look up a z of 0.5 in z table, we see a probability value of 0.1915 (the probability that the random variable will fall between the mean and 550 hours). To answer our question, we must subtract as follows to get probability that the random variable will lie between 550 and 650 hours = 0.4332 – 0.1915 = 0.2417.

Thus, the chance of a candidate selected at random taking between 550 and 650 hours to complete the program is 24 in 100.

**Example 5: *What is the probability that a candidate selected at random will require fewer than 580 hours to complete the program?***

**Example 6: *What is the probability that a candidate chosen at random will take between 420 and 570 hours to complete the program?***