Probability & Statistical Modelling

AQ077-3-2-PSMOD and Version VD1

ASIA PACIFIC UNIVERSITY OF TECHNOLOGY & INNOVATION

Probability Distribution

Topic & Structure of The Lesson

- Introduction
- Binomial Distribution
- Poisson Distribution
- Normal Distribution

Learning Outcomes

- A the end of this topic, You should be able to:
  - Understand probability distribution and its relation to particular
    types of business situations and within these, how probabilities
    are calculated.

Key Terms You Must Be Able To Use

If you have mastered this topic, you should be able to use the following terms correctly in your assignments and exams:

- **Probability distribution**
- **Random variable**
- **Discrete**
- **Continuous**
- **Poisson Distribution**
- **Normal Distribution**
- **Z- value**
- **Continuity Correction**

Introduction

- **Probability Distribution**
  - **A listing of all the outcomes of an experiment and the probability associated with each outcome.**
  - **Related to frequency distributions by simply replaces the actual numbers (frequencies) with the proportion of the total at each level of frequency.**
  - **Graph of probability distribution would be the same as the graph of the frequency distribution, but with the vertical axis marked in proportions rather than in numbers.**

- **Area under the curve in a probability distribution is 100% or 1.**
- **Probability distribution are classified as either discrete (Binomial & Poisson) or continuous (Normal distribution)**

A · P · U
ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

Probability Distributions

- **Discrete** Probability Distributions
  - Binomial
  - Poisson
  - Hypergeometric
- **Continuous** Probability Distributions
  - Normal
  - Uniform
  - Exponential

Binomial Distribution

- **Binomial Distribution**
  - **It is useful for problems in which we are concerned with determining the number of times an event is likely to occur or not occur during a given number of trials and consequently the probability of the event occurring or not occurring.**

- **Characteristics**
  - **It is a discrete distribution of the occurrences of an event with two outcomes – success or failure, good or bad.**
  - **The trials must be independent of one another.**

  - **Main parameters are**

$$E(X) = \mu = np$$

$$Var(X) = npq \qquad \text{where } q = 1 - p$$

$$\text{Standard deviation, } \sigma = \sqrt{npq}$$

- **Binomial probability formula**
  - **Given a binomial situation, X~Bin(n, p) with p = probability of success at any trial and n = number of trails, the probability of obtaining x success is given by:**

$$\Pr(x) = {}^{n}C_{x}(p)^{x}(1-p)^{n-x}$$

OR

$$\Pr(x) = {}^{n}C_{x}(p)^{x}(q)^{n-x}$$

Example 1

At Sellitall Supermarket, 60% of customers pay by credit card. Find the probability that in a randomly selected sample of 10 customers,

(a) exactly 2 pay by credit card.

(b) more than seven pay by credit card.

Example 2

The random variable X is distributed B(7, 0.2).

Find, correct to three decimal places,

(a) $P(X = 3)$

(b) $P(1 < X < 4)$

(c) $P(X > 1)$

Example 3

The probability that it will be a fine day is 0.4. Find the expected number of fine days in a week and also the standard deviation.

Example 4

X is B(n, p) with mean 5 and standard deviation 2. Find the values of n and p.

Poisson Distribution

- **It describes the number of times some event occurring during a specified interval**
- **It is a discrete probability because it is formed by counting**
- **Based on two assumptions:**
  - **The probability is proportional to the length of the interval**
  - **The intervals are independent.**

  (That means the longer the interval the larger the probability, and the number of occurrences in one interval does not affect the other intervals)

- **Characteristics**
  - **It is a discrete distribution and is a limiting form of the binomial distribution when n is large and p or q is small.**
  - **Recognised by the existence of events that**
    - **occur at random**
    - **are ‘rare’**
  - **An interval (of time or space) is defined, within which events can occur.**

- Given a Poisson situation, X ~ P_0 (λ), with λ = mean number of events per interval, the probability of x events occurring is given by:

$$\Pr(x) = e^{-\lambda} \frac{\lambda^{x}}{x!}$$

- The letter 'e' represents a special mathematical constant (having approximate value 2.718)
- The variance of the Poisson is also equal to its mean and it is equal to np

If $X \sim P_0(\lambda)$, then $E(X) = \lambda = np$

And $Var(X) = \lambda$

Example 5

A student finds that the average number of bacteria in 10ml of pond water from a particular pond is 4. Assuming that the number of bacteria follows a Poisson distribution, find the probability that in a 10ml sample

(a) there are exactly 5 bacteria

(b) there are no bacteria

(c) there are fewer than 3 bacteria.

Example 6

On average the school photocopier breaks down 8 times during the school week (Monday to Friday). Assuming that the number of breakdowns can be modeled by a Poisson distribution, find the probability that it breaks down

(a) five times in a given week

(b) once on Monday

(c) eight times in a fortnight.

Example 7

- X follows a Poisson distribution with standard deviation 2. Find $P(X \geq 2)$.

- **A Poisson interval can be adjusted provided the mean is adjusted accordingly.**
- **In a binomial situation, the Poisson distribution can be used as an approximation if:**
  - **n is large (greater than 30)**
  - **p is small (less than 0.01)**

- **The Distribution of two independent Poisson variables**
  - The sum of two independent Poisson variables with parameters m,n, respectively, is a Poisson variable with parameter (m + n).

    i.e. if $X \sim P_o(m)$ and $Y \sim P_o(n)$, then $X + Y \sim P_o(m + n)$

Example 8

A company has 2 machines A and B. On average there are 0.8 breakdowns per week on Machine A and 1.2 breakdowns on Machine B. What is the probability of there being a total of 2 breakdowns on these two machines in a given week?

Example 9

A wholesaler supplies boxes of fireworks to each of two retailers when asked. The number of boxes asked for per week have Poisson distributions with mean 1 and 1.2 respectively. Calculate the probability that the number of boxes requested in a given week by both retailers together is

(a) exactly 3,

(b) more than 2.

Normal Distribution

- The normal distribution, $X \sim N(\mu, \sigma^2)$ is a continuous distribution that has a bell shape and is determined by its mean and standard deviation.
- Characteristics
  - it is symmetrical about its mean, with the greatest frequency at the mean itself.
  - The frequencies of the values taper away (symmetrically) either side of the mean, giving the curve a characteristics ‘ bell-shape’

- knowledge of mean and standard deviation is necessary to identify a specific normal distribution

$$\text{If } X \sim N(\mu, \sigma^2) \text{ then}$$

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

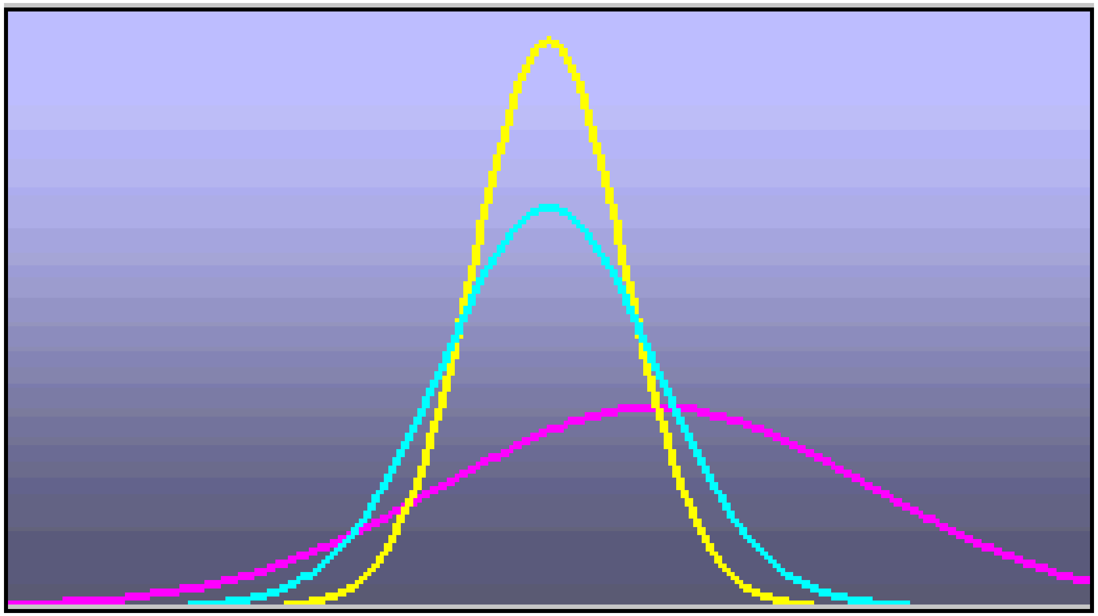
- It is asymptotic, meaning the curve approaches but never touches the X-axis.
- the total area under the curve is equal to 1 or 100%.

- **Probabilities**
  - **Steps that are required to take as follows:**
    - **Calculate the value z for the standard normal distribution if the given variable is normally distributed.**

$$z = \frac{x - \mu}{\sigma}$$

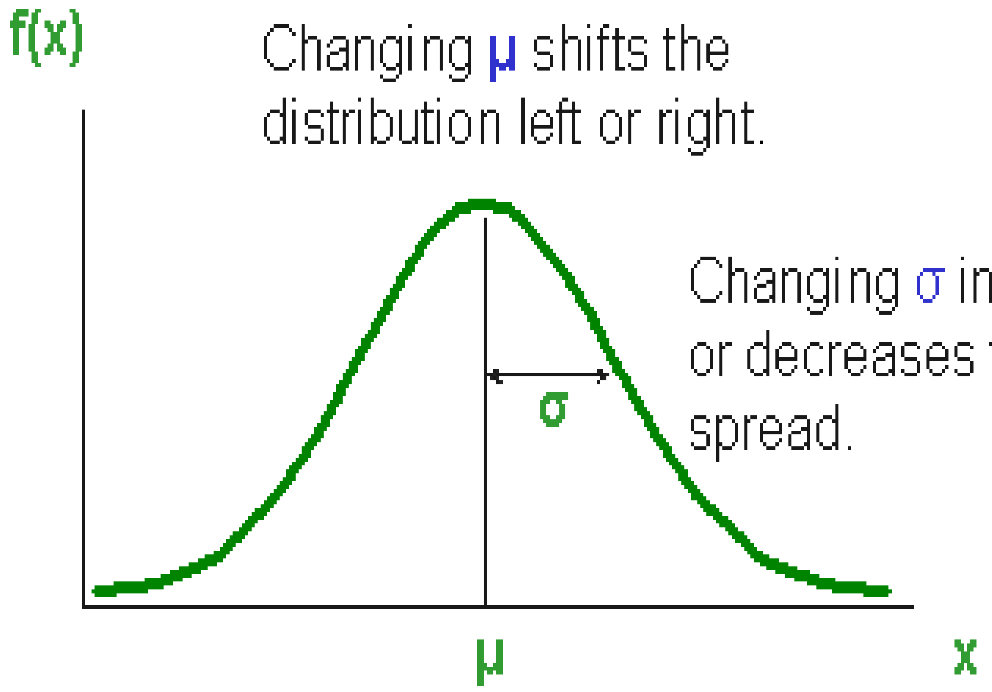
    - **Use the standard normal table to find the corresponding area under the curve for the z value.**
    - **Note that since the normal distribution is symmetrical about its mean, the left half of the curve is a mirror image of the right half.**

Many Normal Distributions



By varying the parameters μ and σ, we obtain different normal distributions

Normal Distribution shape



Let's practise..

Given $X \sim N(20, 3^2)$

Solve the followings:

(a) $P(X > 23)$

(b) $P(X < 24)$

(c) $P(X > 18)$

(d) $P(X < 15)$

Example 10

Lengths of metal strips produced by a machine are normally distributed with mean length of 150cm and a standard deviation of 10cm. Find the probability that the length of a randomly selected strip is

(a) longer than 140cm

(b) between 150cm and 160cm

(c) between 130cm and 155cm

Example 11

The heights of female students at a particular college are normally distributed with a mean of 169cm and a standard deviation of 9cm.

Given that 80% of the female students have a height less than h cm, find the value of h.

Approximation

- In a binomial situation, the Poisson distribution can be used as an approximation if:
  - n is large (greater than 30)
  - p is small (less than 0.01)
  - $np < 5$

Approximation

- $n \geq 30$
- $p < 0.01$
- $np < 5$

Binomial Distribution

Mean = np

Poisson Distribution

Mean = λ

$\lambda = np$

Example 12

Eggs are packed into boxes of 500. On average 0.7% of the eggs are found to be broken when eggs are unpacked. Find, using Poisson approximation, the probability that in a box of 500 eggs,

(a) exactly three are broken,

(b) at least two are broken.

- Normal distribution can be used as an approximation to the binomial distribution when:
  - n is large (greater than 30)
  - p is not too small or large (the closer to 0.5 the better)
  - $np \geq 5$
  - when used as an approximation, Normal distribution has

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

- **Continuity Correction**
  - **Used when Normal distribution are used for discrete variables rather than continuous variables.**
  - **When the range of values of the discrete variable is large, the continuity correction can be ignored because it will be significant.**
  - **However, it ought to be used when the range is small enough to make a difference of 0.5 potentially significant.**

- **How to apply the Correction Factor**
  - **For the probability at least X occur, use the area above (X – 0.5).**
  - **For the probability that more than X occur, use the area above (X + 0.5)**
  - **For the probability that X or fewer occur, use the area below (X + 0.5)**
  - **For the probability that fewer than X occur, use the area below (X – 0.5)**

Approximation

Continuity Correction

- $n \geq 30$
- $p \approx 0.5$
- $np \geq 5$

Normal Dist
(Continuous)
Mean = μ
Standard deviation = σ

$\mu = np$

$\sigma = \sqrt{npq}$

Binomial Dist
(Discrete)
Mean = np

- $n \geq 30$
- $p < 0.01$
- $np < 5$

Poisson Dist
(Discrete)
Mean = λ

$\lambda = np$

Example 13

(a) If $X \sim Bin(100, 0.4)$, find $P(X > 50)$ using the normal approximation.

(b) If 20% of loan applications received by a bank are rejected, what is the probability that among 225 loan applications, at least 50 will be rejected?

Summary of Main Teaching Points

- **A random variable is**
  a variable whose value is determined by the outcome of a random experiment.

- **A probability distribution is**
  a listing of all possible outcomes of an experiment and the probability associated with each outcome.

- A discrete probability distribution can assume only certain values. The main features are:
  - The sum of the probabilities is 1.00
  - The probability of a particular outcome is between 0.00 and 1.00

- A continuous distribution can assume an infinite number of values within a specific range

- Characteristics of a
  - Binomial distribution
  - Poisson Distribution
  - Normal Distribution

- **Calculating the probabilities of**
  - **Binomial distribution**
  - **Poisson Distribution**
  - **Normal Distribution**

- **Approximation of Binomial distribution using**
  - **Poisson distribution**
  - **Normal Distribution**

Question and Answer Session

Q & A

What we will cover next

- Estimation and Confidence Interval