

Unit 07 - The probability distribution

1. The probability distribution

The probability distribution for a random variable describes how the probabilities are distributed over the values of the random variable. In the development of the probability function for a discrete random variable, two conditions must be satisfied: (1) $p(x)$ must be nonnegative for each value of the random variable, and (2) the sum of the probabilities for each value of the random variable must equal one.

X	x_1	x_2	x_3	- - -	x_n
$P(X=x)$	p_1	p_2	p_3	- - -	p_n

Mean = Expected value $\mu = E(x) = \sum x p(x) = x_1 p_1 + x_2 p_2 - - - + x_n p_n$

Var(x) $= \sum x^2 p(x) - (E(x))^2$

Exercise:

1. A fair 4 sided die, numbered 1, 2, 3, and 4 is rolled twice. The random variable X is the sum of the two outcomes on which the die comes to rest.

(i) Show that $P(x = 8) = \frac{1}{8}$

(ii) Draw up the probability distribution table for X , and find $p(x > 6)$

2.

A bag contains 7 orange balls and 3 blue balls. 4 balls are selected at random from the bag, without replacement. Let X denote the number of blue balls selected.

(i) Show that $P(X = 0) = \frac{1}{6}$ and $P(X = 1) = \frac{1}{2}$.

(ii) Construct a table to show the probability distribution of X .

(iii) Find the mean and variance of X .

3.

A fair cubical die with faces numbered 1, 1, 1, 2, 3, 4 is thrown and the score noted. The area A of a square of side equal to the score is calculated, so, for example, when the score on the die is 3, the value of A is 9.

(i) Draw up a table to show the probability distribution of A .

(ii) Find $E(A)$ and $\text{Var}(A)$.

4.

The discrete random variable X has the following probability distribution.

x	1	3	5	7
$P(X = x)$	0.3	a	b	0.25

(i) Write down an equation satisfied by a and b .

(ii) Given that $E(X) = 4$, find a and b .

5.

Two fair dice are thrown. Let the random variable X be the smaller of the two scores if the scores are different, or the score on one of the dice if the scores are the same.

(i) Copy and complete the following table to show the probability distribution of X .

x	1	2	3	4	5	6
$P(X = x)$						

(ii) Find $E(X)$.

2. Binomial /Bernoulli Distribution:

A binomial distribution is one kind of discrete probability distribution that has **two possible outcomes** (Success or failure / Pass or Fail)

Properties/Criteria

Binomial distributions must also meet the following three criteria:

- The number of observations or trials is fixed
- Trails are independent
- The probability of success or pass) is exactly the same from one trial to another.

When the random variable X , satisfies these conditions we denote it by

$$X \sim \beta(n, p)$$

The random variable X , which represents the number of successes in the n trials of this experiment, has a probability distribution given by

$$P(X = r) = nC_r p^r q^{n-r} \quad \text{where } n = 0, 1, 2, 3, \dots, n \text{ (numbers of trails)}$$

$p = \text{probability of success}$

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

Mean (Expected Value) and Variance of Binomial Distribution

If $X \sim B(n, p)$, then mean $\mu = E(x) = n \times p$ Variance $\sigma^2 = npq$

Exercise:

1. A driving test is passed by 70% of people at their attempt. Find the probability that
 - (i) exactly 5 people out of 10 people will pass the driving test.
 - (ii) More than 1 person out of 8 people will pass the driving test.

2.

65% of all watches sold by a shop have a digital display and 35% have an analog display.

- (i) Find the probability that, out of the next 12 customers who buy a watch, fewer than 10 choose one with a digital display. [4]

3.

- (i) A garden shop sells polyanthus plants in boxes, each box containing the same number of plants. The number of plants per box which produce yellow flowers has a binomial distribution with mean 11 and variance 4.95.

- (a) Find the number of plants per box. [4]
- (b) Find the probability that a box contains exactly 12 plants which produce yellow flowers. [2]

3. Poisson Distribution

In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.

Conditions / Properties

- Occurs singly
- The average rate at which events occur is always the same
- Events are independent

The random variable X , satisfying Poisson distribution $X \sim P_0()$, then probability distribution given by

$$P(X = r) = \frac{e^{-\mu} \times \mu^r}{r!}$$

Mean (Expected Value) and Variance of Poisson Distribution

$$E(x) = \text{Variance } \sigma^2 = \mu$$

Exercise:

1.

Computer breakdowns occur randomly on average once every 48 hours of use.

(i) Calculate the probability that there will be fewer than 4 breakdowns in 60 hours of use.

(ii) there will be no breakdowns in 24 hours of use.

2.

Between 7 p.m. and 11 p.m., arrivals of patients at the casualty department of a hospital occur at random at an average rate of 6 per hour.

(i) Find the probability that, during any period of one hour between 7 p.m. and 11 p.m., exactly 5 people will arrive. [2]

(ii) A patient arrives at exactly 10.15 p.m. Find the probability that at least one more patient arrives before 10.35 p.m. [3]

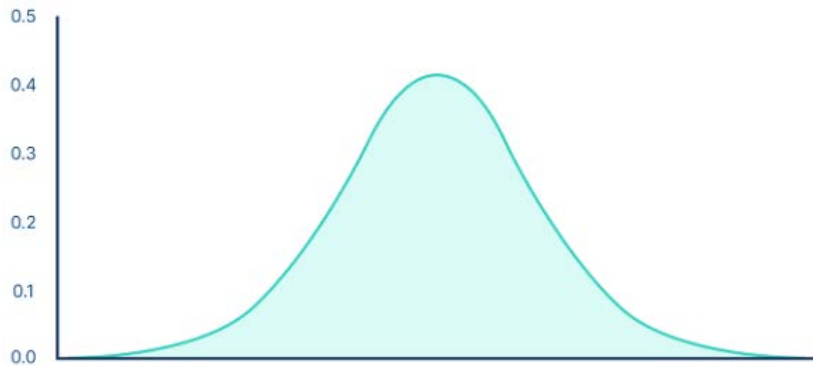
3.

The number of radioactive particles emitted per second by a certain metal is random and has mean 1.7. The radioactive metal is placed next to an object which independently emits particles at random such that the mean number of particles emitted per second is 0.6. Find the probability that the total number of particles emitted in the next 3 seconds is 6, 7 or 8. [4]

4. Normal Distribution

A normal distribution or Gaussian distribution is one of continuous probability distributions.

In a normal distribution, data are symmetrically distributed with no skew. Most values cluster around a central region, with values tapering off as they go further away from the center. The measures of central tendency (mean, mode, and median) are the same in a normal distribution. In graphical form, the normal distribution appears as a **"bell curve"**.



What are the properties of normal distributions?

Normal distributions have key characteristics that are easy to spot in graphs:

- The mean, median and mode are the same.
- The distribution is symmetric about the mean—half the values fall below the mean and half above the mean and is non-zero over the entire real line
- The distribution can be described by two values: the mean and the standard deviation.

If a variable X is normally distributed then, it is symbolized as $X \sim N(\mu, \sigma)$. The parameters by using the Greek symbol μ (mu) for the population mean and σ (sigma) for the population standard deviation.

Empirical rule

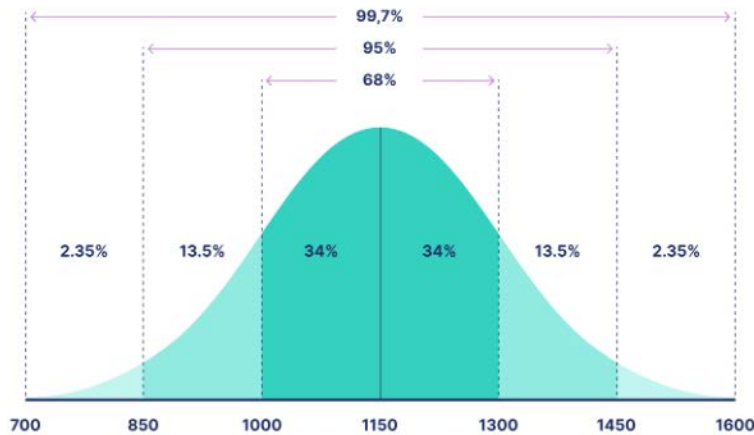
The empirical rule, or the 68-95-99.7 rule, tells you where most of your values lie in a normal distribution:

Around 68% of values are within 1 standard deviation from the mean.

Around 95% of values are within 2 standard deviations from the mean.

Around 99.7% of values are within 3 standard deviations from the mean.

Using the empirical rule in a normal distribution



Why do normal distributions matter?

All kinds of variables in natural and social sciences are normally or approximately normally distributed. Height, birth weight, reading ability, job satisfaction, or SAT scores are just a few examples of such variables. Because normally distributed variables are so common, many statistical tests are designed for normally distributed populations.

Normal probability density formula	Explanation
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$	<ul style="list-style-type: none"> • $f(x)$ = probability • x = value of the variable • μ = mean • σ = standard deviation • σ^2 = variance