# TYCHR

*Friend, Philosopher, Guide*

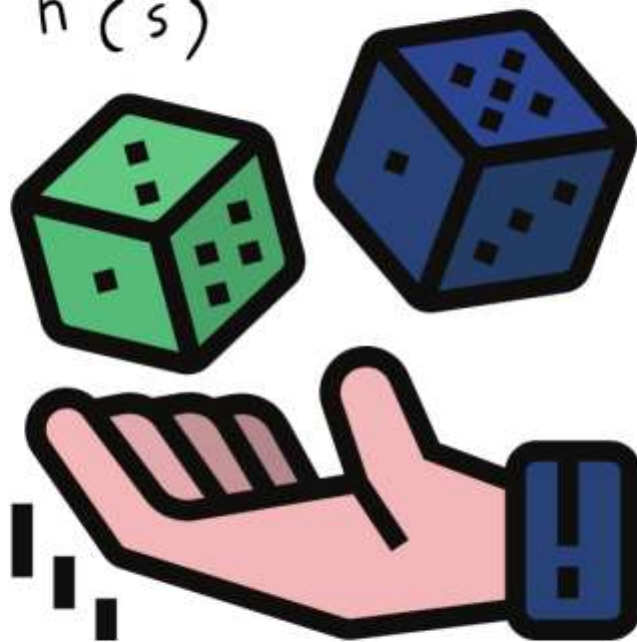# Probability Distributions

$$P(A) = \frac{n(A)}{n(s)}$$

**By the end of this chapter, you will be familiar with:**
- Random variables and their probability distributions
- The normal distribution
- Standardizing normal variables
- Inverse normal calculations
- The binomial distribution

# RANDOM VARIABLES
## DISCRETE AND CONTINUOUS RANDOM VARIABLE
- A random variable, usually written X, is a variable whose possible values are numerical outcomes of a random phenomenon.
- Random variables are represented by capital letters. The actual measured values which the random variable can take are represented by lower case letters.
- P(X=x)=Probability of x.
- Ex. A dice is rolled 3 times, and X represents number of sixes, write P(X=x) to represent "the probability that the number of sixes is x" where x can take the values 0, 1, 2 and 3.
- There are two types of random variables, discrete and continuous.
- A random variable that may assume only a finite number or an infinite sequence of values is said to be discrete.
- One that may assume any value in some interval on the real number line is said to be continuous.
- For instance, a random variable representing the number of automobiles sold at a particular dealership on one day would be discrete, while a random variable representing the weight of a person in kilograms (or pounds) would be continuous.

## PROBABILITY DISTRIBUTION
- A probability distribution is a list of all of the possible outcomes of a random variable along with their corresponding probability values.
- Suppose you flip a coin two times. This simple statistical experiment can have four possible outcomes: HH, HT, TH, and TT.
- Now, let the variable X represent the number of heads that result from this experiment.
- The variable X can take on the values 0, 1, or 2. In this example, X is a random variable; because its value is determined by the outcome of a statistical experiment.
- A probability distribution is a table or an equation that links each outcome of a statistical experiment with its probability of occurrence.

- The table below, which associates each outcome with its probability, is an example of a probability distribution.

| Number Of Heads (X) | Probability P(X=X) |
| --- | --- |
| 0 | 0.25 |
| 1 | 0.50 |
| 2 | 0.25 |

- A cumulative probability refers to the probability that the value of a random variable falls within a specified range.
- Like a probability distribution, a cumulative probability distribution can be represented by a table or an equation.
- In the table below, the cumulative probability refers to the probability than the random variable X is less than or equal to x.

| Number Of Heads (X) | Probability P(X=X) | Cumulative Probability P(X≤X) |
| --- | --- | --- |
| 0 | 0.25 | 0.25 |
| 1 | 0.50 | 0.75 |
| 2 | 0.25 | 1 |

- Uniform distributions are probability distributions with equally likely outcomes. Ex. When a die is tossed, there are 6 possible outcomes represented by: S = {1, 2, 3, 4, 5, 6} and each outcome is equally likely to occur.
  $P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = P(X = 6) = 1/6$

**Ex.** A random variable X has the following probability distribution:

| X | 1 | 2 | 3 | 4 | 5 |
| --- | --- | --- | --- | --- | --- |
| P(X=X) | 3k | 2k | k | 5k | 3k |

1. Find the value of k.
2. Find P(X≥3)

We have PX=x=1
3k+2k+k+5k+3k=1
14k=1
**k = 1/14**
$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$
$P(X \geq 3) = k + 5k + 3k \ P(X \geq 3) = 9k$
**$P(X \geq 3) = 9/1$**


# EXPECTATION, VARIANCE AND STANDARD DEVIATION

- The **expected value** (or mean) of X, where X is a discrete random variable, is a weighted average of the possible values that X can take, each value being weighted according to the probability of that event occurring.
- The expected value of X is usually written as E(X).
- $\mu = E(X) = \sum x \, P(x)$

**Ex.** What is the expected value when we roll a fair die?
There are six possible outcomes: 1, 2, 3, 4, 5, 6. Each of these has a probability of 1/6 of occurring. Let X represent the outcome of the experiment.
$P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = P(X = 6) = 1/6$
$E(X) = \sum x \, P(x)$
$E(X) = (1 \times P(1)) + (2 \times P(2)) + (3 \times P(3)) + (4 \times P(4)) + (5 \times P(5)) + (6 \times P(6))$
$E(X) = (1/6) + (2/6) + (3/6) + (4/6) + (5/6) + (6/6)$
$E(X) = 7/2 = 3.5$

So the expectation is 3.5 . If we think about it, 3.5 is halfway between the possible values the die can take and so this is what we should have expected.

- The **variance** of a random variable tells us something about the spread of the possible values of the variable. For a discrete random variable X, the variance of X is written as Var(X) or V(X) or $\sigma^2$.
- $\sigma^2 = \sum E(X^2) - [E(X)]^2$
  $\sigma 2 = \sum (x - E(X))^2 \, P(x)$
- The standard deviation will be the square root of the variance.

$$\sigma = \sqrt{\sum E(X^2) - [E(X)]^2} \quad \sigma = \sqrt{\sum (x - E(X))^2 P(x)}$$

**Ex.** Consider the following probability distribution and calculate the variance and standard deviation.

| X | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| P(X=X) | 0.07 | 0.41 | 0.19 | 0.07 | 0.12 | 0.14 |

$E(X) = \Sigma x P(x)$
$E(X) = (0 \times 0.07) + (1 \times 0.41) + (2 \times 0.19) + (3 \times 0.07) + (4 \times 0.12) + (5 \times 0.14)$
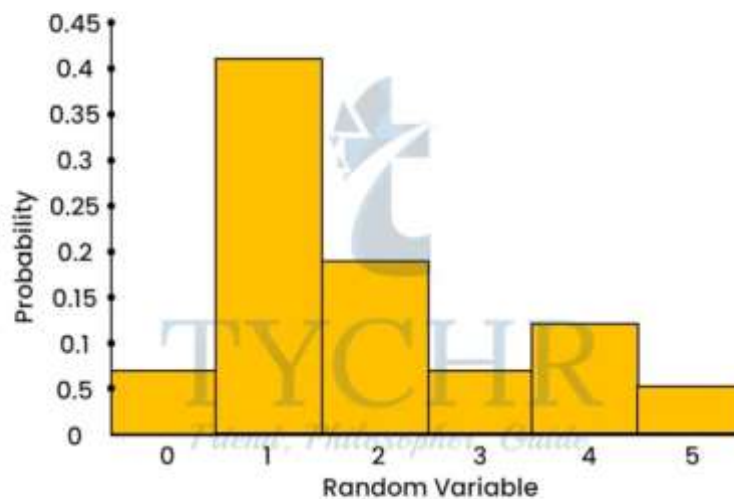$E(X) = 2.18$

Variance $= \sigma^2 = \Sigma(x - E(X))^2$
$P(x) = (0 - 2.18)^2 (0.07) + (1 - 2.18)^2 (0.41) + (2 - 2.18)^2 (0.19) + (3 - 2.18)^2 (0.07) + (4 - 2.18)^2 (0.12) + (5 - 2.18)^2 (0.14) = 2.138$
Standard deviation $= \sqrt{2.138} = 1.462$

Variance $=\sigma^2$ $\quad = \sigma^2 = \Sigma (x - E(X))^2 P(x)$ The Probabilty distribution graph:

# THE BINOMIAL DISTRIBUTION

- A binomial distribution can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times.
- The binomial is a type of distribution that has two possible outcomes (the prefix "bi" means two, or twice).

**Ex.** You sell sandwiches. 70% of people choose chicken, the rest choose something else. What is the probability of selling 2 chicken sandwiches to the next 3 customers?
We are familiar with drawing a tree diagram.
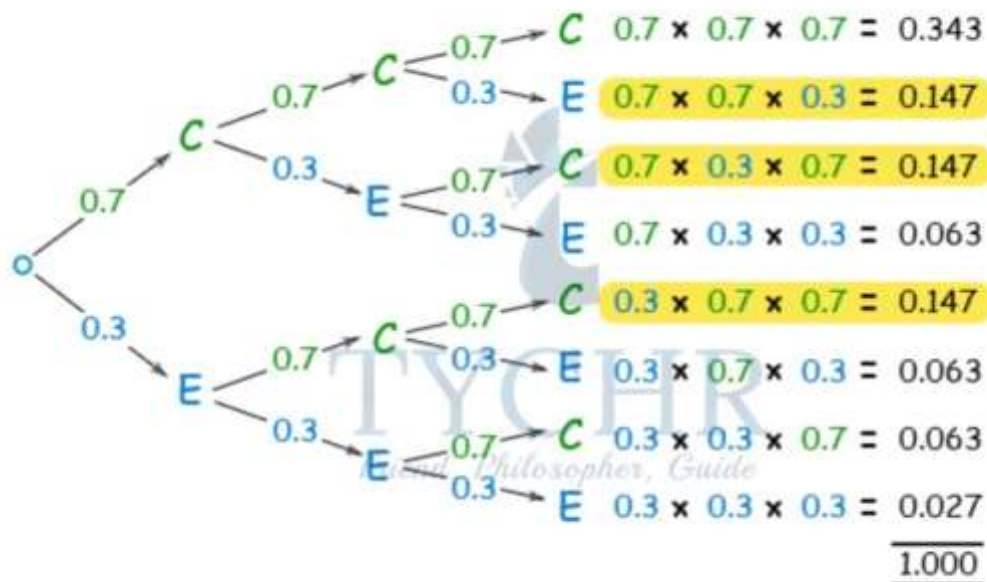


We cannot draw a tree diagram when there are a large number of cases. So, we will find a general formula for such problems. It is known as the binomial function.
The probabilities for "two chickens are 0.147. In each case, we multiply 0.7 two times and 0.3 one time.
0.7 is the probability of 'SUCCESS'
while, 0.3 is the probability of 'FAILURE'
Let p=0.7

Now, 0.3 is the opposite choice, hence, 1-p=0.3
We have considered 3 customers- n=3
We need the probability of choosing 2 chicken sandwiches, let r=2
There were 3 customers, the number of opposite choice is 1.
n-r=1

Now, there are 3 ways in which it can happen:
(chicken, chicken, other) or (chicken, other, chicken) or (other, chicken, chicken)

So, the total number of two chicken outcomes is 23C = = 3!/2!(3−2)! =
3 Required Probability according to tree
diagram =0.7×0.7×0.3+0.7×0.3×0.7+0.3×0.7×0.7=0.441
Now, let us calculate the probability in a more general way-
Required Probability =Number of outcomes Probability of each outcome
= 3×0.147
= $23C \times (0.7 \times 0.7 \times 0.3)$
= $23C \times (0.7^2 \times 0.3^1)$

Replacing n, r and p for a general result: $= {}^{n}_{r}C p^r (1-p)^{n-r}$
Now,

- If X is a discrete random variable which is binomially distributed, X~B(n,p), then the probability of obtaining r successes out of n independent trials, when p is the probability of success for each trial, is $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$

- $\binom{n}{r} = {}^{n}_{r}C$ is called the binomial coefficient.
- The expectation of X for a binomial distribution is given by $E(X) = np$
- The variance of a binomial function is given by $Var(X) = \sigma^2 = np(1-p)$
- For binomial distribution as well:
  - $0 \le P(x) \le 1$
  - $\sum_{r=0}^{n} P(x)\, n = 1$

# CONTINUOUS DISTRIBUTIONS

## THE NORMAL DISTRIBUTION

- Normal distribution is a continuous probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

- In graph form, normal distribution will appear as a bell curve.
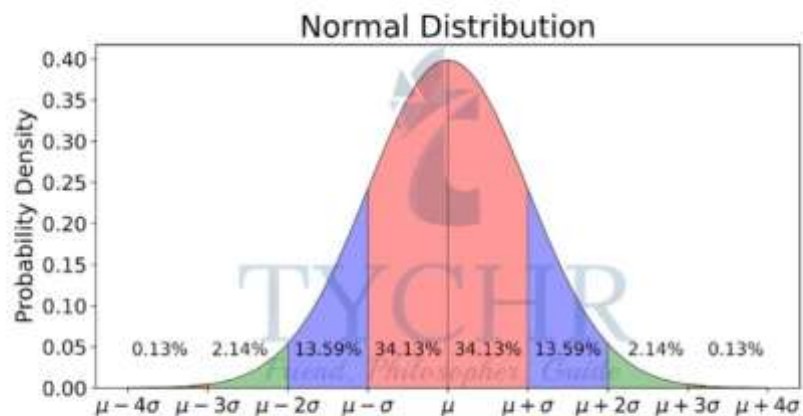


- Normal distributions have the following features:
    1. symmetric bell shape
    2. mean and median are equal; both located at the center of the distribution
    3. ≈68% of the data falls within 1 standard deviation of the mean
    4. ≈95% of the data falls within 2 standard deviation of the mean
    5. ≈99.7% of the data falls within 3 standard deviation of the mean
- The normal distribution is fully determined by its mean and its standard deviation .

## PROBABILITY DENSITY FUNCTION

- Probability density function (PDF), is a function whose integral is calculated to find probabilities associated with a continuous random variable.
- For a continuous random variable X, the probability density function f(x), has the following properties:
    1. fx>0
    2. The area under the probability density function over all values of X is equal to 1. $\int_{-\infty}^{\infty} f(x)dx = 1$
- The probability that a random variable X takes on values in the interval a≤X≤b is defined as-

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

- The probability density function for a normally distributed random variable X is-

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

## THE STANDARD NORMAL DISTRIBUTION

- It is a normal distribution where $\mu=0$ and $\sigma=1$.
- It enables us to read areas under any normal distribution through the standardisation process.
- We use Z to describe a random variable with a standard normal distribution and we write $Z\sim N(0,1)$.
- Every normal random variable X can be transformed into a Z score via the following equation: $Z = (X-\mu)/\sigma$
- We can use a GDC to calculate the areas under the standard normal distribution curve for values of Z between Z=a and Z=b. This allows us to calculate P(a<Z<b).

**Ex.** Molly earned a score of 940 on a national achievement test. The mean test score was 850 with a standard deviation of 100. What proportion of students had a higher score than Molly? (Assume that test scores are normally distributed)

First, we transform Molly's test score into a Z-score $Z = (X-\mu)/\sigma$

$Z = (940-850)/100$

$Z = 0.9$

Now, using GDC we will find the required area under the standard normal curve and hence the probability.

PZ<0.9=0.18159

Required probability =PZ>0.9=1-PZ<0.9=1-0.8159=0.1841

Thus, 18.41% of the students tested had a higher score than Molly.

## THE INVERSE NORMAL DISTRIBUTION

- An inverse normal distribution is a way to work backwards from a known probability to find an X-value.
- The approach is to find the standard inverse normal number and then to de-standardise it. That is, to find the value from the original data that corresponds to the z-value at hand.

**Ex.** Given that X$\sim$N(15, 3$_2$), find the value of x for which PX<x=0.75

We know that $Z = (X-\mu)/\sigma$

$Z = (X-15)/3$

Now, $P (Z < (x-15)/3 ) = 0.75$

Using inverse normal function on GDC: $(x-15)/3 = 0.6744$

x = 17

# TYCHR

*Friend, Philosopher, Guide*

**WWW.TYCHR.COM**

**+91 9540653900**