Probability and Probability Distribution

Adejumo Ridwan Suleiman

2022-06-28

What is probability

- It is simply the study of uncertainty.
- Example the possibility of raining, tossing a coin or rolling a die.
- It is the measuring of how likely an event will occur.
- Mathematically defined as:

$$Probability = \frac{\text{Number of Required outcomes}}{\text{Number of Possible outcomes}}$$

Terms in Probability

- **Experiment:** An uncertain situation e.g tossing a coin
- **Outcome:** Result of a trial in an experiment.
- **Event:** One or more outcome from a experiment
- **Sample Space:** The collection of possible outcomes of an experiment.

Random Variable

- Outcome of an event expressed in numbers
- For example in the coin toss experiment we can either have a **Head** or **Tail** which can be numerically expressed as $\mathbf{1}$ or $\mathbf{0}$ respectively.
- Let's call a set containing these two numbers **X** where; $\mathbf{X} = \{\mathbf{1,0}\}$.
- **X** represents the Random Variable
- What's the random variable of a Six face die?

The Two Coin Toss Experiment

First Coin

Second Coin	H	T
H	(H,H)	(T,H)
T	(H,T)	(T,T)

Figure 1: Tabular representation of the sample space of the two coin toss

The Two Coin Toss Experiment

- $\mathbf{S = \{HH,HT,TH,TT\}}$

$$\mathit{Probability} = \frac{\text{Number of Required outcomes}}{\text{Number of Possible outcomes}}$$

- Number of possible outcomes $= 4$
- Probability of getting a head in both coins is:

$$= \frac{\text{Number of Required outcomes(HH)}}{\text{Number of Possible outcomes(HH,HT,TH,TT)}} = \frac{1}{4}$$

- Probability of getting a head in the first coin and a tail in the second coin:

$$= \frac{\text{Number of Required outcomes(HT)}}{\text{Number of Possible outcomes(HH,HT,TH,TT)}} = \frac{1}{4}$$

- Probability of getting a head and a tail in both coins.

$$= \frac{\text{Number of Required outcomes(HT,TH)}}{\text{Number of Possible outcomes(HH,HT,TH,TT)}} = \frac{2}{4}$$

The Two Die Experiment

		White Die					
		1	2	3	4	5	6
Red Die	1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
	2	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
	3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
	4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
	5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
	6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

Figure 2: Tabular representation of the sample space of rolling two die

The Two Die Experiment

- $\mathbf{S = \{(1,1),(1,2),(1,3),\ldots,(6,6)\}}$

$$Probability = \frac{\text{Number of Required outcomes}}{\text{Number of Possible outcomes}}$$

- Number of possible outcomes $= 36$
- Probability of getting a `one` in both die:

$$= \frac{\text{Number of Required outcomes(1,1)}}{\text{Number of Possible outcomes}} = \frac{1}{36}$$

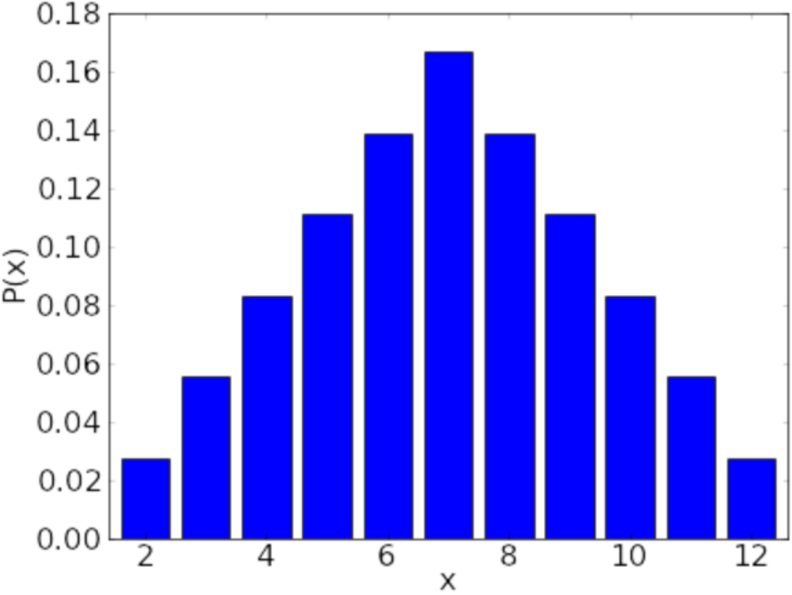
- Probability of getting a `one` in the first die and a `two` in the second die:

$$= \frac{\text{Number of Required outcomes(1,2)}}{\text{Number of Possible outcomes}} = \frac{1}{36}$$

- Probability of getting a one and a two:

$$= \frac{\text{Number of Required outcomes(1,2) or (2,1)}}{\text{Number of Possible outcomes}} = \frac{2}{36}$$

The Two Die Experiment



The Candy Jar

- A Candy Jar has 11 candies, **5 Green** candies, **2 Yellow** candies and **4 Red** candies.

1. If your sibling wants a candy what is the probability of you picking a **Green** Candy.
2. If your sibling **rejects** the Green candy, and decides he wants **Two Yellow** Candies what's the probability.
3. If you want a **Green** and a **Red** Candy whats the probability of you picking.

   Note all these scenarios are in order

The Candy Jar

- Number of Candies $= 11$
- Probability of Selecting Green $= \dfrac{5}{11}$
- Probability of Selecting Yellow $= \dfrac{2}{11}$
- Probability of Selecting Red $= \dfrac{4}{11}$

The Candy Jar

1. Probability of Green Candy $= \dfrac{5}{11}$
2. Probability of Two Yellow Candy given the Green Candy is returned $= \dfrac{2}{11} \times \dfrac{1}{10} = \dfrac{1}{55}$
3. Probability of Green Candy and a Red Candy given no more Yellow Candy $= \dfrac{5}{9} \times \dfrac{4}{8} = \dfrac{5}{18}$

Bernoulli Distribution

- A single trial with only two possible outcomes is called as Bernoulli distribution.
- Example is a coin tossed once or a fight between me and Mayowa(DevNet) where the probability of I winning is $\mathbf{0.9}$ and him losing is $\mathbf{0.1.}$

Distribution of a Bernoulli Experiment

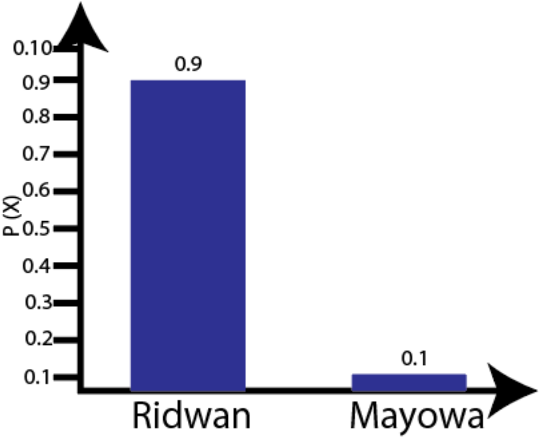


Figure 4: Distribution of a Bernoulli experiment

Binomial Distribution

- Unlike the Bernoulli Distribution, the binomial distribution has **n** number of trials.
- A distribution is said to be Binomial if the following are satisfied;
  - A trial with two outcomes and repeated **n** number of trials
  - Each trial is independent
  - A total numbers of **n** trials are conducted
  - The probability of **Success** and **Failure** is same for all trials.

Distribution of a Binomial Experiment

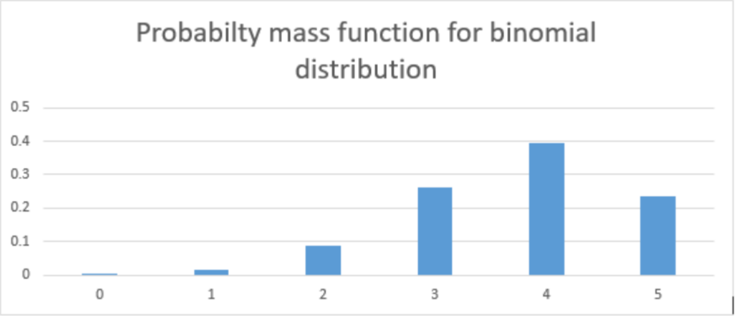
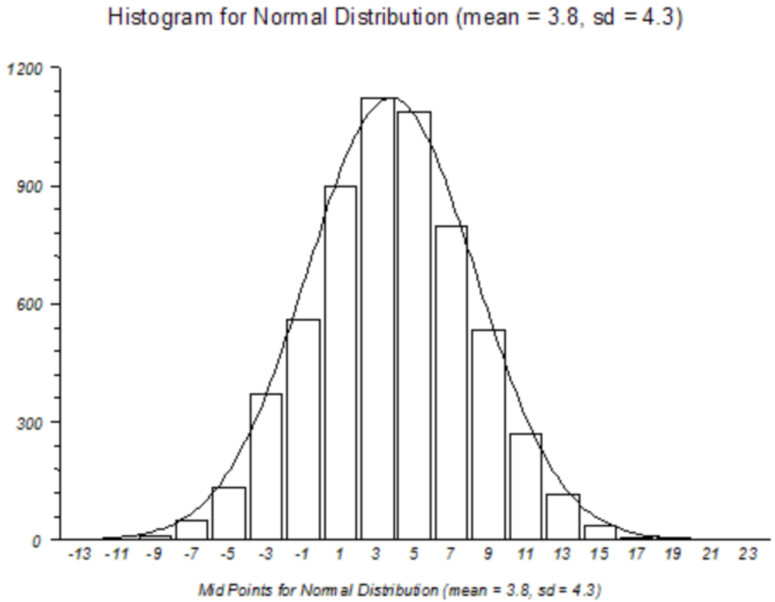


Figure 5: Distribution of a Binomial experiment

Normal Distribution

- A distribution is said to be normally distributed if it satisfies the following conditions;
  - The **Mean, Median and Mode** of the distribution are the same.
  - The curve of the distribution is bell shaped
  - Half of the value are left of the center and the other half at the right.

Normal Distribution



Central Limit Theorem

- Regardless of the distribution of a variable's population, if we have a sufficiently large sample size, the mean and standard deviation of that variable will be approximately normally distributed.

Central Limit Theorem

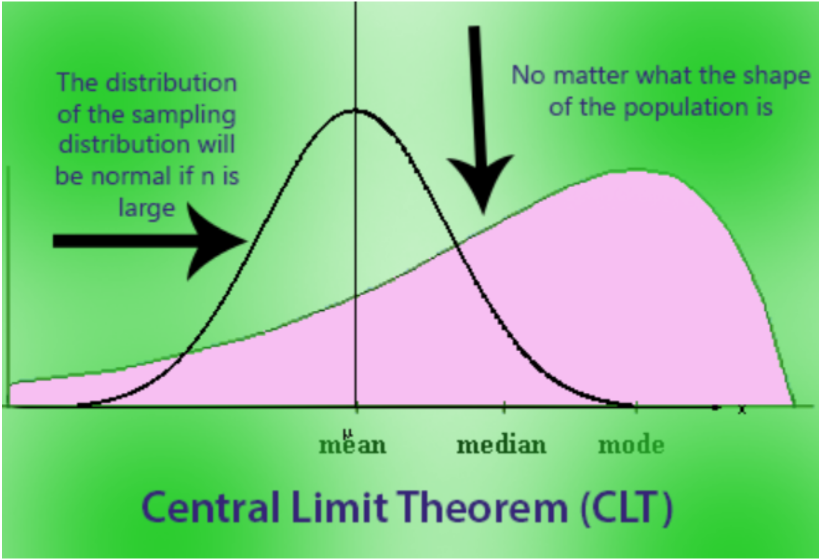


Figure 7: Central Limit Theorem

Challenge

- Use the table below to answer the following questions.

```
## # A tibble: 5 x 2
## # Groups:   cut [5]
##   cut           n
##   <ord>     <int>
## 1 Fair       1610
## 2 Good       4906
## 3 Very Good 12082
## 4 Premium   13791
## 5 Ideal     21551
```

- Imagine you are to pick a diamond from a virtual box online, what is the probability of you picking a **Premium** diamond?
- If you are to pick two **Premium** diamonds what is the probability?

References

- *Basics of Probability for Data Science explained with examples in R*
- *Central Limit Theorem Simplified!*