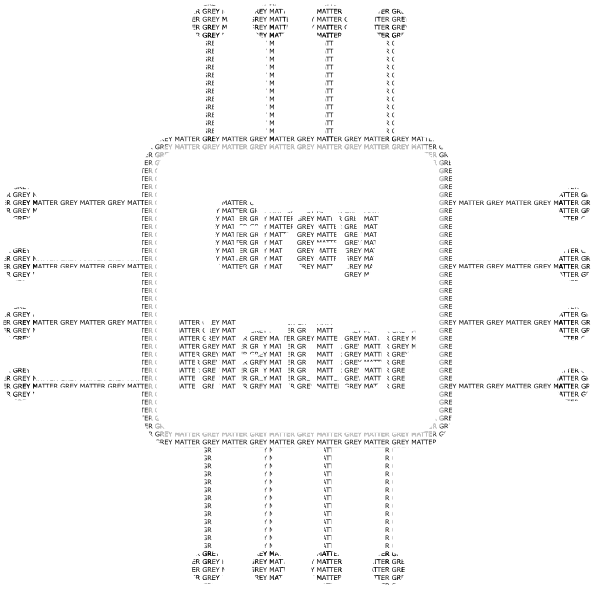
PROJECT - Grey Matter

A collection of Notes

Abirami Sukumaran
Divyansh Chahar



Tuesday 17th November, 2020

Meet Thy Contributors

Abirami Sukumaran

Senior Data Architect at Citrix

See my codes here

For professional collaboration

Divyansh Chahar

Budding Data Scientist | CFD Aficionado |External Flows Expert

See my codes here

For professional collaboration

Nomenclature

Greek Characters

$\bar{x}$	sample mean
$\widetilde{x}$	sample median
C_v	Cofficient of variance/Relative Standard Deviation
$corr(x, y)$	Correlation between x and y
$cov(x, y)$	covarince of x and y
IQR	Interquartile Range
N	number of memers in the population
n	number of members in samlpe dataset
P	Probability
Q_1	First Quartile
Q_2	Second Quartile
Q_3	Third Quartile
R	range
R_p	Rank of Percentile
S	Satndard Deviation of Sample
S^2	sample variance
w	Percent of Values
x_i	memebr of dataset
Z	Standard Score

Roman Characters

μ	population mean
σ	Standard DEviation of Population
σ^2	population variance

1 | Introduction to Statistics and Measures used in Descriptive Statistics

1.1 Measure of Central Tendency

1.1.1 Mean

Mean

Mean is simply the average of the data.

It can be mathematically expressed as follows

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

$$\boxed{\mu = \frac{1}{N}\sum_{i=1}^{n} x_i}$$

1.1.2 Median

Median

Median is the central value of the data when the values are arranged in increasing or decreasing order

It can be mathematically expressed as

$$\boxed{\widetilde{x} = \begin{cases} x_{(n+1)/2}, \text{if n is odd} \\ \frac{1}{2}\left(x_{(n)/2} + x_{(n+1)/2}\right), \text{if n is even} \end{cases}}$$

1.1.3 Mode

Mode

Mode is the most frequently occurring value in the dataset

1.2 Measure of Dispersion

If data points are represented on number line than measure of dispersion will be an indication of how far are data points from each other.

1.2.1 Range

Range

It is the interval over which the data is spread. It is the difference between the highest and lowest value of the data

It is mathematically expressed as

$$\boxed{R = Max(x_i) - Min(x_i)}$$

1.2.2 Quartiles and Inter-Quartile Range

Quartiles

Quartiles divide the ordered dataset into 4 parts with equal members

- The **second quartile** or $\boldsymbol{Q_2}$ is the value at the middle position when the data is arranged in ascending order, it is also the **median value**.
- The first quartile or $\boldsymbol{Q_1}$ divides the first half of the data into 2 parts, we can say that the **first quartile is the median of the first half of the data**.
- Similarly the **third quartile** or $\boldsymbol{Q_3}$ is the **median of the second half of the data**.

Interquartile Range

Interquartile range is the differance between the median of the first and second half of the data, or we can say that inter quartile range is the differance between Q_1 and Q_3.

$$\boxed{IQR = Q_3 - Q_1}$$

1.2.3 Variance

Variance

Variance is a measure of how far the data points are from the mean.

Variance is one of the methods of measuring how spread out the values are in a dataset. Variance can be mathematically calculated as

$$\boxed{S^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n-1}}$$

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}}$$

Due to squaring of difference, the figure obtained could be pretty large and insignificant as the unit of measurement is squared.

1.2.4 Standard Deviation

Standard Deviation

Standard Deviation is the square root of Variance

Standard Deviation also measure the spread of data. Since the units are not squared in Standard Deviation and the values are not too large either, it is much more used than variance. Standard Deviation can be calculated as :

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

1.2.5 Coefficient of Variation or Relative Standard Deviation

Standard Deviation is the most common measure of the variability of a single data-set. However when we want to compare the standard deviation of two or more data-sets the issue becomes a little more complicated, we need a relative measure such as Coefficient of Variation. To better understand the concept of Variance, Standard Deviation and Coefficient of Variance, let us consider the following example.

Mean, Variance, Satndard Deviation

Lets us consider stock price of a random company over a week. Let us consider two investors, one from U.S.A. and another one from India. although the price of Stock is same but the Indian Investor has to pay in Indian Rupee at an exchange rate of 73.51 INR per USD.

	USD	INR
Day 1	514.79	37842.213
Day 2	498.79	36666.053
Day 3	504.72	37101.968
Day 4	508.57	37384.981
Day 5	504.05	37052.7155
Mean	506.184	37209.586
Variance	28.225	152519.7
Standard Deviation	5.313	390.538

- Mean, variance and standard deviation for USD are **506.184 USD**, **28.225 USD^2** and **5.313 USD** respectively.
- Mean, variance and standard deviation for INR are **37209.586 INR**, **152519.7 INR^2** and **390.538 INR** respectively.

From the above data we can observe:

- Variance values are comparatively larger than Standard Deviation
- Even though we have same variability for INR and USD values, Standard Deviation varies a lot.

To compare the variability of two different datasets we use Cofficient of Variance.
For population it can be calculated as

$$\boxed{C_v = \frac{\sigma}{\mu}}$$

For Sample it can be calculated as

$$\boxed{C_v = \frac{S}{\bar{x}}}$$

Coefficient of Variance or Related Standard Deviation

Let us calculate the cofficient of variance for the above data using the population formulae

- Cofficient of variance for USD is $C_v = \frac{5.313}{506.184} = 0.010$
- Cofficient of variance for INR is $C_v = \frac{390.538}{32709.586} = 0.010$

1.3 Measure of Position

Measure of Position is the measure of the location of a data point in the population or sample.

1.3.1 Percentile

Percentile

N^{th} percentile will represent a datapoint which is larger than N percent of values in the dataset i.e. 99^{th} percentile represents a value which is greater than 99 percent of the values. Percentile divide the ordered data into 100 equal parts in terms of number of members.

Rank of Percentile can be mathematically measured as

$$\boxed{R_p = \frac{w}{100}N}$$

The above formulae works fine if R_p is a positive integer. But the situation can become bit complicated if R_p also has decimal value.

Rank of Percentile

For example, let us consider a dataset with following values: $P = 35$, $w = 90$. Therefore R_p can be calculate as follows:

$$R_p = \frac{35}{90}200 = 31.5$$

As we can observe that 31.5 rank could be difficult to calculate. However this can be achieved as follows, Let us consider the $\mathbf{30^{th}}$ value is **60** and $\mathbf{31^{st}}$ value is **65**. Thus we will proceed as follows.

- **STEP 1:** Break the rank i.e. R_p into integer and decimal part. In this case the integer part is 31 and decimal part is 0.5.
- **STEP 2:** Find the Value at integer rank and integer rank + 1. In this case the values are 60 and 65 respectively
- **STEP 3:** Calculate the difference between the value at integer rank and integer rank +1

$$65 - 60 = 5$$

- **STEP 4:** Multiply the difference with the decimal part.

$$0.5 \times 5 = 2.5$$

- **STEP 5:** Add the two values to get the final value

$$60 + 2.5 = 62.5$$

1.3.2 Quartile

As mentioned in the above section quartiles divide the dataset into 4 equal parts in terms of number of members. Rank of the three quartiles can be calculated as:

$$\boxed{Q_1 = \frac{25}{100}N}$$

$$\boxed{Q_2 = \frac{50}{100}N}$$

$$\boxed{Q_3 = \frac{75}{100}N}$$

- Second Quartile is the median of the dataset.
- If the Rank of a quartile is a not a poitive integer, than it can be resolved just like the percentile shown in the example above.

1.3.3 Standard Score

Standard Score

Standard Score measure the differance between the datatpoints and the mean in terms of standard deviation

It can be mathematically expressed as

$$\boxed{Z = \frac{x - \bar{x}}{\sigma}}$$

2 | Exploratory Data Analysis

As we proceed further, we will observe that the selection of Statistical and Machine Learning Algorithms depends on the task at hand, but the nature and type of data also plays a very important part while selecting these algorithm

Data forms the backbone of any statistical algorithm, hence it is is imperative to understand the data from a statistical point of view. Thus what is precursor to feature engineering(which is a precursor to algorithm deployment) is ***Exploratory Data Analysis***

Exploratory Data Analysis

Exploratory Data Analysis is an approach that employs variety of graphical and non-graphical techniques for uncovering the following aspects of data:

- Data Set Understanding
- Data Structure
- Important Variables
- Outliers and Anomalies
- Assumptions
- Dependencies
- Relationships
- Correlation
- Patterns

In simple words, EDA is a process to understand your data.

Based on the type of techniques used, EDA can be classified into two categories:

- ***Graphical EDA:*** This approach uses various visualization techniques like Bar Graphs, Histograms, Pie Charts etc to understand the data
- ***Non-Graphical EDA:*** Any approach that makes use of any technique that is not graphical in nature can be classified as non-graphical EDA.

Based on the number of variables being examined we can classify EDA techniques into two categories:

- ***Univariate EDA:*** When EDA techniques are apllied to a single variable it is called univariate EDA.

- ***Multivariate EDA:*** When more than one variables are examined it is called multivariate EDA.

Before we proceed further we need to familiarize our self with the definition of outliers.

Outliers

Any data point with a z-score of more than or equal to 3 or -3 is considered a outlier

2.1 Box and Whisker Plot

A box and whisker plot is a simple graphical method used to convey the measures of spread and measure of center. It has the following components:

- Minimum : the lowest data point excluding any outliers
- Maximum : the largest data point excluding any outliers
- Median (Q_2 or 50^{th} percentile)
- First quartile (Q_1 or 25^{th} percentile)
- Third Quartile (Q_3 or 75^{th} percentile)
- Interquartile Range

Any data outside the maximum and minimum value is plotted as outlier, represented by dots before and after the minimum and maximum mark.

Due to the five quantities a Box and Whisker Plot represents, it is often used for five-number summary approach to visualize the data set.

We will now demonstrate the procedure of drawing a box and whisker plot with the help of an example

Box and Whisker Plot

Let us consider an fictitious data set with the following characteristics:

- $Q_1 = 10$
- $Q_2 = 20$
- $Q_3 = 30$

To plot the above data set as box and whisker plot we need to follow the following steps:

1. **Plot Q_2 or median:** The first step in drawing a box and whisker plot is plotting the median. The median is represented by a line.
2. **Plot Q_1 or first quartile:** After plotting median we need to plot Q_1 or first quartile, this is also representd by a line. It must be noted that median can be plotted anywhere, however when plotting Q_1 or first qurtile, the distance between the median and Q_1 should be representative of the difference between the two characteristics
3. **Plot Q_3 or third quartile:** In a similar manner plot Q_3. We have now completed the box plot.
4. **Plot Minimum:** In order to calculate the minimum whisker we need to calculate the IQR, the minimum marker lies 1.5 times before the first quartile.
5. **Plot maximum:** The maximum whisker lies 1.5 times from the Q_3.
6. **Plotting Outliers:** Any data point which is larger than the maximum whisker is plotted outside the maximum whisker as outlier. Similarly any data point which is smaller than the minimum whisker is plotted before the minimum whisker as outlier.

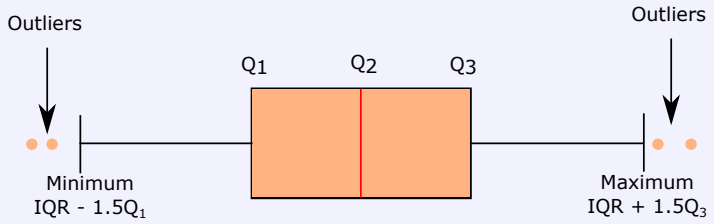


Figure 2-1: Box and Whisker Plot

2.2 Histogram

- The purpose of a histogram (Chambers) is to graphically summarize the distribution of a univariate data set.
- The variable is divided into several bins or intervals, the number of observation per bin or interval is represented by the height of the bar
- A simple method to roughly calculate the number of bins or interval, is to take the square root of the total number of values in your distribution.

- The histogram graphically shows the following
  - Center of the data
  - Spread of the data
  - Skewness of the data
  - Presence of Outliers
  - Presence of multiple Modes in the data

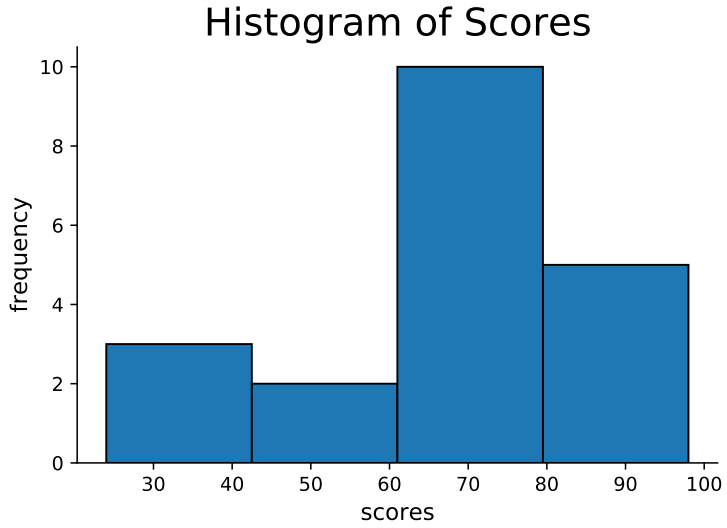


Figure 2-2: Example of histogram

2.3 Bar Graph

A bar graph looks similar to histogram, however there is one key difference between them, bar graphs are used to visualize discrete data where as histograms are used for continuous data. A bar graph is shown in figure 2-3

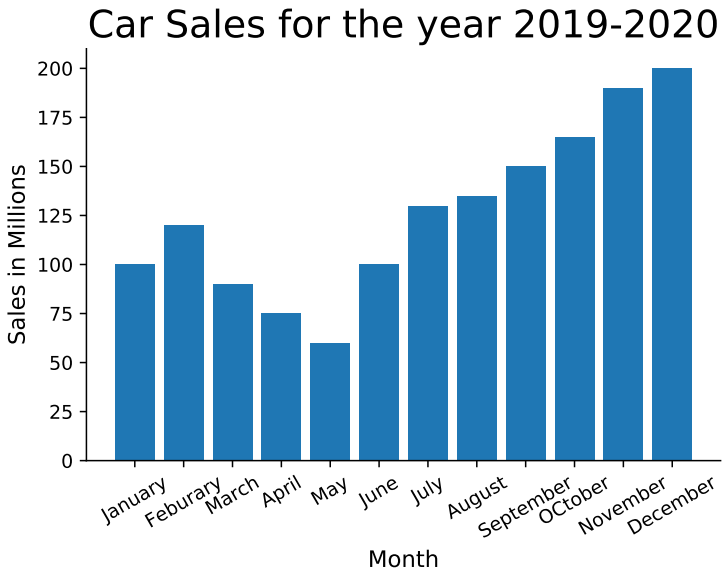
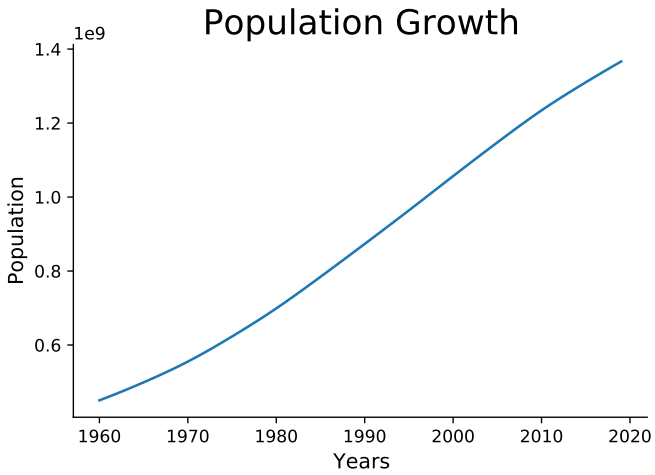


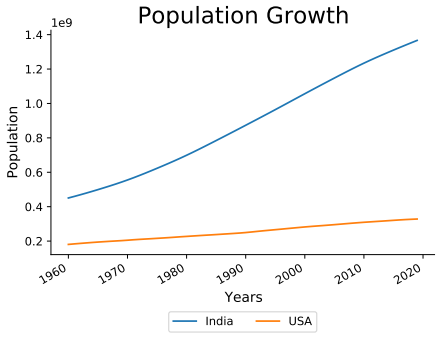
Figure 2-3: Example of bar plot

2.4 Line Graphs

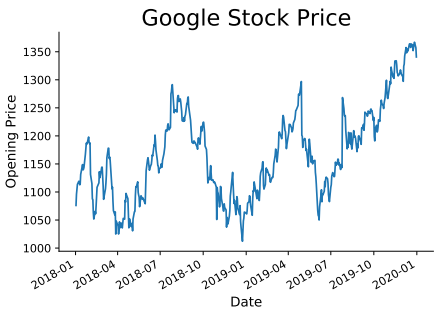
- A line graph is typically used to visualize the change in variable over a period of time.
- It could also be used to represent the relationship between two variables
- When a Line Graph is used to visualize the change in variable over time it is known as ***Run Sequence Plot***
- With run sequence plots, shifts in location and scale are typically quite evident
- In Run Sequence Plot, outliers can be easily identified



(a) Univariate Line Graph



(b) Multivariate Line Graph



(c) Run Sequence Plot

Figure 2-4: Various Examples of Line Graphs

2.5 Pie Chart

- Unlike other charts and graphs, a pie is circular in nature
- It is often used to visualize proportions
- Various parameters like angle subtended by the arc at the center, length of the arc and area of the slice are proportional to the quantity it represents

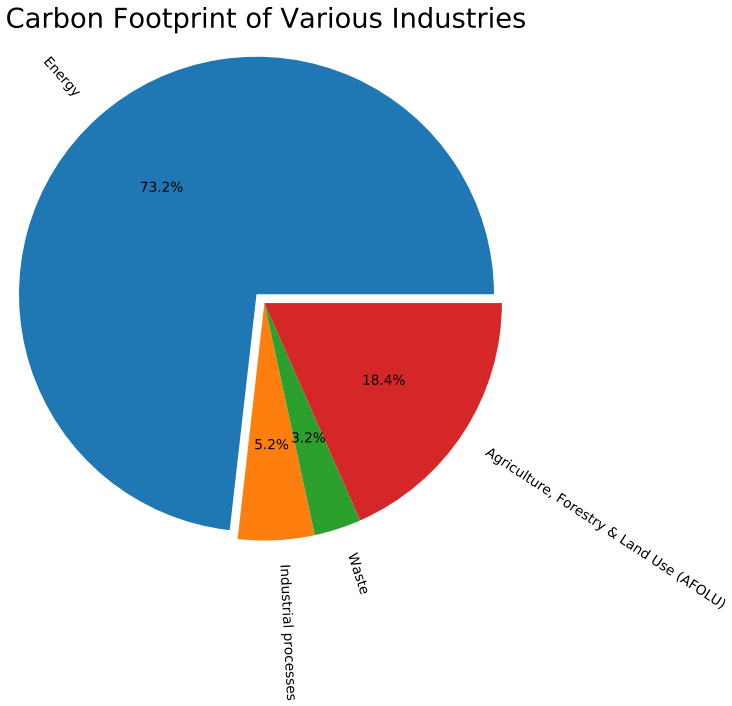


Figure 2-5: Example of Pie Chart

2.6 Scatter Plot

Before proceeding further, we must familiarize ourselves with **covariance** and **correlation**

2.6.1 What is Covariance ?

Covariance

Covariance is a measure of relation between the two variables which indicates the direction of linear relationship between variables. Covariance can be calculated as:

$$\boxed{cov(x,y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n}}$$

Covariance can indicate the following about the data:

- A positive covariance indicates that increasing x will result in a higher value of y and decreasing x will result in a lower value of y
- A negative correlation indicates that increasing x will result in a lower value of y and decreasing x will result in a higher value of y
- If the covariance is zero it indicates that either decreasing or increasing the value of x does not have any effect on y

Covariance is not often used in statistics since it is hard to interpret due to the following reasons:

- It does not gives much information about the slope of relationship
- It does not gives much information about relative position of data points

2.6.2 What is Correlation ?

Covariance

Correlation is a measure of strength and direction between two variables. Correlation can be calculated as:

$$\boxed{corr(x,y) = \frac{cov(x,y)}{\sigma_x \sigma_y}}$$

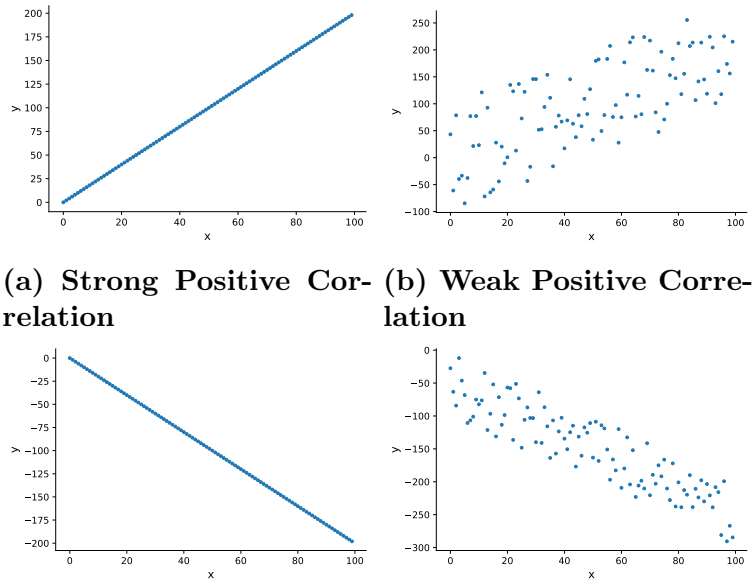
For a better understanding of correlation, let us imagine a following scenario. Let us imagine a straight line given by equation $y = mx + c$, where x and y represents the x and y coordinate respectively, m is the slope and c represents y-intercept. For this scenario the following holds true:

- For every point on the line there is a strong correlation between the points on x and y
- The nature of correlation depends on the slope of the line

For the above scenario the correlation between the points on x and y-axis will be either -1 or 1 based on the nature of the slope. A correlation of -1 or 1 means for any given value of x the corresponding y value will be linearly related. Now let us imagine that few points are slightly displaced from their location such that they do not follow strict linear relationship, for this dataset the value of correlation will not be 1 or -1 the value will decrease based on the amount of displacement from from the original location.

2.6.3 What are Scatter Plots?

- Scatter plots are used to observe relationships between variables
- A scatter plot (aka scatter chart, scatter graph) uses dots to represent values for two different numeric variables
- The position of each dot on the horizontal and vertical axis indicates values for an individual data point



(a) Strong Positive Correlation **(b) Weak Positive Correlation**

(c) Strong Negative Correlation **(d) Weak Negitive Correlation**

Figure 2-6: Examples of Scatter Plots

2.7 Heat Map

Heat Map

A Heat Map is a method of visualizing the values in 2D matrix. The magnitude of the value is represented by a color taken from a colormap.

Heat maps are most widely used for visualizing correlation between variables, hence it usually has the same number of rows and columns. ***It must be noted that when a heat map is used for visualizing correlation, the diagonals always have a value of 1 because they represent the relation of a column with itself.***

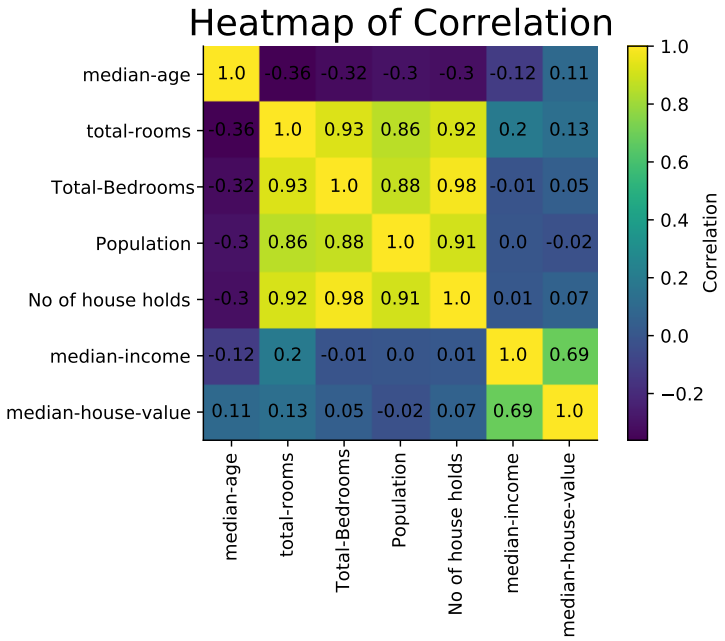


Figure 2-7: Example of Heat Map

We will now proceed with a slightly more advance concept of ***Pareto Chart***

2.8 Pareto Chart

Pareto Charts are based on ***Pareto Principal***, hence it imperative that we first familiarize ourself with Pareto Principal

Pareto Principal

The Pareto principle (also known as the 80/20 rule, the law of the vital few, or the principle of factor sparsity) states that, for many events, roughly 80% of the effects come from 20% of the causes

A ***Pareto Chart*** is a combination of Line Graph and Bar Plot as we will see in the example below.

Pareto Chart

Let us suppose we have the data given in table 2.1. This data can be converted to a Pareto Chart using the following steps:

1. **Arrange the data in descending order**
2. **Plot the bar graph**
3. **Calculate cumulative percentage :** To calculate cumulative percentage, we first need to calculate cumulative frequency, then cumulative frequency for each category is divided by total and multiplied by 100. For ex: To calculate the cummulative frequency of Electrical Defects we will divide 687 by 754 and multiply by 100, i.e. cumulative frequency of Electrical Defects = $\frac{687}{754} \times 100 = 91.11$
4. **Plot line graph :** Line graph is plotted to represent cumulative percentage of each category

Types of Defects	Frequency	Cummulative Frequency	Cummulative Percentage
Panel Gaps	200	200	26.53
Paint Defects	123	323	42.84
Electrical Defects	364	687	91.11
Major Mechanical Defects	10	697	92.44
Minor Fixture Issues	57	754	100.00

Table 2.1: Types of Defects reported by a car manufacturer

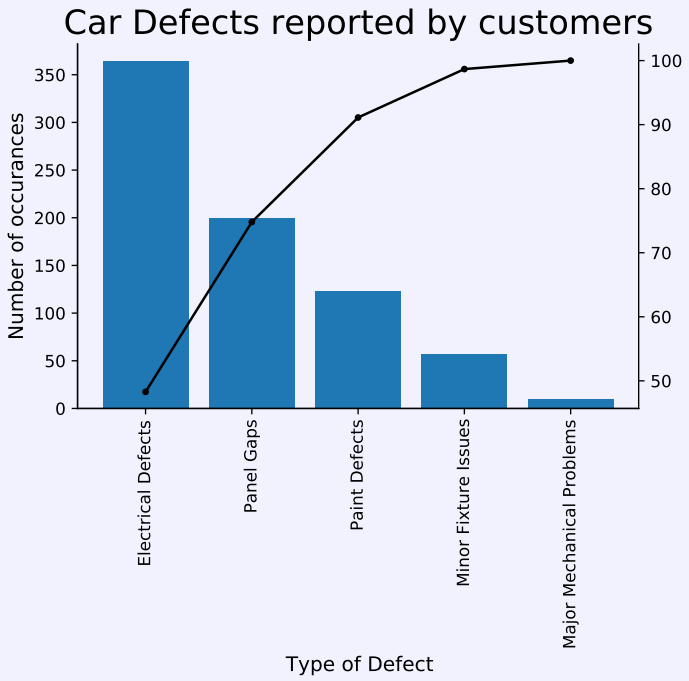


Figure 2-8: Example of Pareto Chart

3 | Probability

3.1 What is Probability ?

Probability

Probability means chance. It can be mathematically expressed as

$$P = \frac{\text{favourable outcome}}{\text{total number of outcomes}}$$

We can also say that Probability is the ratio of favorable outcome to possible outcomes

Let us take an example to understand this concept

Probability

What is the probability of drawing a club from a deck of cards.

$$\text{Favourable outcome} = \text{number of club cards in a deck} = 13$$

$$\text{Total Number of outcome} = \text{Total number of cards in a deck} = 52$$

$$P = \frac{13}{52} = \frac{1}{4} = 0.25$$

There is 25% probability that the card drawn will a club card

3.2 Basic Terminology of Probability

To fully understand probability the reader first need to familiarize itself with the basic terminology used in in Probability. Before proceeding further we need to familiarize ourselves with the following terms:

- Probability Experiment
- Events
- Outcomes
- Sample Space

3.2.1 Probability Experiment

Probability Experiment

A repeatable experiment with defined set of outcomes is called a Probability Experiment

3.2.2 Outcomes

Outcomes

The result of a probability experiment is known as outcome

3.2.3 Sample Space

Sample Space

Collection of all possible outcomes of a probability experiment is known as Sample Space

3.2.4 Events

Events

A collection of a few outcomes from sample space is known as an event. It is a subset of sample space.

Probability Experiment, Outcomes, Sample Space, Events

Let us consider that a fair coin is tossed 2 times, then we can say that

Tossing of coin itself is the **probability experiment**.

Occurrence of Heads or Tails are the two **outcomes**

The **Sample Space** can be expressed as {HH}, {TT}, {HT}, {TH}.

We can have the following events:

Event	Outcomes
1 Head	{HT}, {TH}
1 Tail	{HT}, {TH}
2 heds	{HH}
2 Tails	{TT}

Table 3.1: Possible Events and Outcomes form tossing of a coin

3.3 Union and Intersection of Events

Union of Events

In case of two distinct events A and B, the union of the event will contain all the outcomes that are present in either A or B. The union of two events is represented as $A \bigcup B$.

Intersection of Events

In case of two distinct events A and B, the intersection of the events will contain all the elements present in both A and B. The union of two events is represented as $A \bigcap B$

Union and Intersection of Events

Let us consider two events A and B having the following outcomes from rolling of a dice:
Event A = $\{2, 4, 6\}$
Event B = $\{3, 6\}$
Event(A $\bigcup$ B) = $\{2, 3, 4, 6\}$
Event(A $\bigcap$ B) = $\{6\}$

3.4 Types of Events

3.4.1 Mutually Exclusive Events

Mutually Exclusive Events

Two events are said to be mutually exclusive if the occurrence of one event implies that other event event will not occur.

Mutually Exclusive Events

Occurrence of head or tail are mutually exclusive events because occurance of one confirms that the other will not occur.

It must be noted that for mutually exclusive events A and B, $P(A \bigcap B) = 0$

3.4.2 Non Mutually Exclusive Events

Non Mutually Exclusive Events

Two events are said to be non-mutually exclusive if they can occur together

Non-Mutually exclusive Event

A single card drawn from a deck of cards could be both an ace and a club card at the same time. Thus we can say that drawing a club card and drawing an ace are non-mutually exclusive events

3.4.3 Independent Events

Independent Events

Two Events are said to be independent if the occurrence of one event does not affects the probability of occurrence of the other

Independent Events

Two consecutive rolls of a fair dice.

3.4.4 Dependent Events

Dependent Events

Two events are said to be dependent if occurrence of one event does affects the probability of occurrence of another event

Dependent Events

Let us imagine a box of fruits containing 5 apples, 5 bananas and 5 oranges. Let us say we randomly pick a fruit from the box. Then we pick another fruit without replacing the fruit we picked earlier, this will affect the next outcome.

Let us look at this from a mathematical point of view.

Probability of picking an orange in the first go is $\frac{5}{15}$
Let us now randomly draw a fruit from the box, let this fruit be a banana, thus probability of drawing an orange in the second go is $\frac{5}{14}$ because we have not replaced the fruit we have drawn in the first draw.

3.5 Rules of Probability

1. $P(A^C) = 1 - P(A)$
2. $A \subset B \implies P(A) \leq P(B)$
3. $P(A \bigcup B) = P(\text{A or B}) = P(A) + P(B) - P(A \bigcap B)$
4. $P(A \bigcup B \bigcup C) = P(A) + P(B) + P(C) - P(A \bigcap B) - P(A \bigcap C) - P(B \bigcap C) + 2P(A \bigcap B \bigcap C)$

5. If A and B are independent events than $P(\text{A and B}) = P(A \bigcap B) = P(A) \times P(B)$

Union and Intersection of Two Sets

In a class of engineering freshmen, there is a 80 percent chance that a student will opt for advance physics during the next semester, there is 60 percent chance that a student will opt for advance maths and there is 50 percent chance that a student will opt for both. Given a student is selected at random what is the probability that he/she will opt for either advance physics or advance maths ?

Let A be the probability that student selects advance physics and let B be the probability that student selects advance maths i.e.

$$P(A) = \text{P(Advance Physics)} = \frac{80}{100}$$

$$P(B) = \text{P(Advance Maths)} = \frac{60}{100}$$

$$P(A \bigcap B) = \text{P(Advance Maths \& Advance Physics)} = \frac{50}{100}$$

$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$$

$$P(A \bigcup B) = \frac{80}{100} + \frac{60}{100} - \frac{50}{100}$$

$$P(A \bigcup B) = \frac{90}{100}$$

There is 90 percent chance that a student selected at random will select either Advance Physics or Advance Maths.

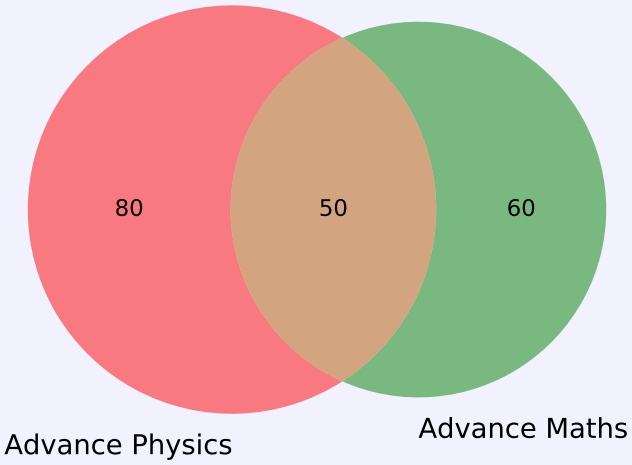


Figure 3-1: Venn Diagram of Probabilities for Two Events

Union and Intersection of Three Sets

In a survey conducted on high school class it was observed that if a student is selected at random, then there is 30 percent probability that he/she is studying physics as major, 45 percent probablity that he/she is studying chemistry as major and 60 percent probability that he/she is studying maths as major. There is 15 percent probability that a student is majoring in both physics and chemistry, 30 percent probability that a student is majoring in both physics and maths, and 20 percent probability that a student is majoring in chemistry and maths. If the there is 80 percent probability that a student selected at random is majoring in either in physics, chemistry or maths is 80 percent. Calculate the probability that a student selected at random will be majoring in all three subjects.

Let

$$P(A) = P(\text{Physics}) = \frac{30}{100}$$
$$P(B) = P(\text{Chemistry}) = \frac{45}{100}$$
$$P(C) = P(\text{Maths}) = \frac{60}{100}$$
$$P(A \bigcap B) = \frac{15}{100}$$
$$P(A \bigcap C) = \frac{30}{100}$$
$$P(B \bigcap C) = \frac{20}{100}$$
$$P(A \bigcup B \bigcup C) = \frac{80}{100}$$

We know that

$$P(A \bigcup B \bigcup C) = P(A) + P(B) + P(C) - P(A \bigcap B) - P(A \bigcap C) - P(B \bigcap C) + 2P(A \bigcap B \bigcap C)$$
$$\frac{80}{100} = \frac{30}{100} + \frac{45}{100} + \frac{60}{100} - \frac{15}{100} - \frac{30}{100} - \frac{20}{100} + 2P(A \bigcap B \bigcap C)$$
$$\frac{80}{100} = \frac{135}{100} - \frac{65}{100} + 2P(A \bigcap B \bigcap C)$$
$$\frac{80}{100} = \frac{70}{100} + 2P(A \bigcap B \bigcap C)$$
$$P(A \bigcap B \bigcap C) = \frac{5}{100}$$

Thus there is 5 percent probability that a student selected at random will be majoring in physics, chemistry and maths.

3.6 Types of Probability

3.6.1 Marginal Probability

Marginal Probability

The probability of occurrence of a single event is called marginal probability

Marginal Probability

The probability of occurrence of a an even number when a fair dice is rolled a single time.

3.6.2 Joint Probability

Joint Probability

The Probability of occurrence of two events at the same time is known as joint probability

Joint Probability has the following properties:

- Joint Probability of two independent events A and B can be calculated as

$$P(A \bigcap B) = P(A) \times P(B)$$

- Joint Probability of dependent events can't be calculated

Joint Probability

A box contains 5 red and 3 blue marbles, calculate the probability that the first marble drawn is red and the second marble is blue if the marbles are drawn successively without replacing.

$$P(A) = P(\text{Drawing a red marble}) = \frac{5}{8}$$

$$P(B) = P(\text{Drawing a blue marble}) = \frac{3}{7}$$

$$P(A \bigcap b) = P(A) \times P(B) = \frac{5}{8} \times \frac{3}{7} = \frac{15}{56}$$

3.6.3 Conditional Probability

Conditional Probability

Conditional Probability is the probability of a dependent event A occurring given that another event B has already occurred

Conditional probability of Event A occuring given that event B has already occurred is expressed as

$$\boxed{P(A|B) = \frac{P(\text{A and B})}{P(B)} = \frac{P(A \bigcap B)}{P(B)}}$$

Conditional Probability

Calculate the conditional probability of successively drawing two kings from a deck of cards without replacing.
Let

$$P(A) = P(\text{Drawing the first king}) = \frac{4}{52}$$

$$P(B) = P(\text{Drawing the second king}) = \frac{3}{51}$$

$$P(\text{A and B}) = P(A) \times P(B) = \frac{4}{52} \times \frac{3}{51}$$

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)} = \frac{\frac{4}{52} \times \frac{3}{51}}{\frac{4}{52}} = \frac{3}{51}$$

3.7 Odds and Log of Odds

3.7.1 Odds

Odds

Odds of an Event A are the ratio of outcomes . Odds in favour of Event A is the ratio of favorable outcomes to the ratio of unfavorable outcomes. Odds against Event A is the ratio of unfavorable outcomes to the ratio of favorable outcomes. Odds can be mathematically expressed as

$$\text{Odds in favour of A} = \frac{\text{Number of favourable outcomes}}{\text{Number of unfavourable outcomes}}$$

$$\text{Odds against A} = \frac{\text{Number of unfavourable outcomes}}{\text{Number of favourable outcomes}}$$

Odds

What are the odds of drawing an ace from a deck of cards ?

$$\text{Favourable outcomes} = 4$$

$$\text{Unfavourable Outcomes} = 52 - 4 = 48$$

$$\text{Odds in favour} = \frac{4}{48}$$

$$\text{Odds against} = \frac{48}{4}$$

3.7.2 Log of Odds

Log of Odds

Log of Odds is the logarithmic of odds

3.8 Baye's Theorem

A formal definition of Baye's theorem at this point will be complicated to interpret as it requires the understanding of certain concepts and principals which have not been covered, hence we will take a slightly different approach to understand the driving ideas behind Baye's Theorem.

In the previous section we discovered what is conditional probability. The notion $P(A|B)$ denotes the probability of event A occurring given that event B has already occurred.

Let us assume that there exist two events named A and B. Let us assume the following probabilities are known:

- $P(A \bigcap B)$
- $P(B)$

Using the above data we can clearly determine the probability of Event A given that Event B has already occurred. But what if we have to determine the probability of Event B given that Event A has already occurred.

This is where Baye's Theorem comes into picture. **Baye's Theorem is used to analyze how well a proof fits a theory.** Let us familiarize ourselves with this concept with the help of an example.

Let us assume a group of individuals who are inhabitants of different states of USA. It is theorized that 30 percent of individuals from the state of Texas are farmers, 20 percent of all the individuals are farmers and 40 percent individuals are from Texas.

We can mathematically express it as

$$P(A) = P(\text{Farmer}) = \frac{20}{100} = 0.2$$

$$P(B) = P(\text{Texas}) = \frac{40}{100} = 0.4$$

$$P(A|B) = P(Farmer|Texas) = \frac{30}{100} = 0.3$$

Now let us assume that we need to calculate the probability of a person being from Texas given that he is a farmer i.e. $P(B|A)$.

Thus using Baye's Theorem we can use the following relation

$$\boxed{P(B|A) = \frac{P(A|B)P(B)}{P(A)}}$$

Thus based on the above theorem we can say that

$$P(Texas|Farmer) = \frac{P(Farmer|Texas)P(Texas)}{P(Farmer)}$$

$$P(B|A) = \frac{0.3 \times 0.4}{0.2} = 0.6$$

As can be seen that there is 60 percent probability that if a farmer is selected he/she is a resident of Texas.

4 | Probability Distribution

Before exploring the more advance concepts of probability, familirity with the term ***Random Experiment*** and ***Random Variable*** is required. Hence we will first explore these concepts

Random Experiment

A random experiment is a experiment whose outcome cannot be predicted with absolute certainty.

Random Variable

A random variable is real valued function which can take up any value from the sample space of Random Experiment.

4.1 Types of Probability Distribution Functions

Before we proceed further we must familiarize ourselves with what is ***probability distribution function***

Probability Distribution Function

Probability distribution is a list of all outcomes of a random experiment along with their probabilities.

In order to better understand this we need to understand different kinds of probability distribution functions. Probability distribution function can be classified into three categories:

- Probability Mass Function
- Probability Density Function
- Cumulative Distribution Function

4.1.1 Probability Mass Function

Probability Mass Function

A probability distribution function can be classified as probability mass function if the random variable can take only discreet values

To better understand this let us take an example

Probability Mass Function

Let the random experiment be rolling of a dice . Then the random variable X can take values 1, 2, 3, 4, 5 and 6. Then Probability Mass function will look like this

$X = x$	$P(X = x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Table 4.1: Probability Mass Function for rolling of a dice

4.1.2 Probability Density Function

Probability Density Function

A probability distribution function can be classified as probability density function if the random variable can take only continuous values.

In simple words we can say that Probability Density Function can predict the probability of random variable X falling within a specific range within the sample space i.e. it can be mathematically expressed as

$$P(a \leq X \leq b) = \int_a^b f(x)$$

It must be noted that probability function for probability density function is an integral, this is because of the nature of random variable

Probability Density Function

Let us consider a dataset consisting of height of individuals as shown in table 4.2.

Height in ft	Frequency
2.5 - 3.5	23
3.5 - 4.5	135
4.5 - 5.5	340
5.5 - 6.5	340
6.5 - 7.5	135
7.5 - 8.5	23

Table 4.2: Probability Density Function for Height of people

The above table can be interpreted as follows, if we randomly select an individual from the dataset shown in table 4.2 then there is $\frac{640}{960}$ chance that the person will have a height between 4.5 and 6.5 ft.

4.1.3 Cumulative Distribution Function

To fully understand cumulative distribution function, we need to first understand how probability density function and probability mass function are visualized.

Visualizing Probability Mass Function

The probability mass function given in table 4.1 can be visualized as shown in figure 4-1

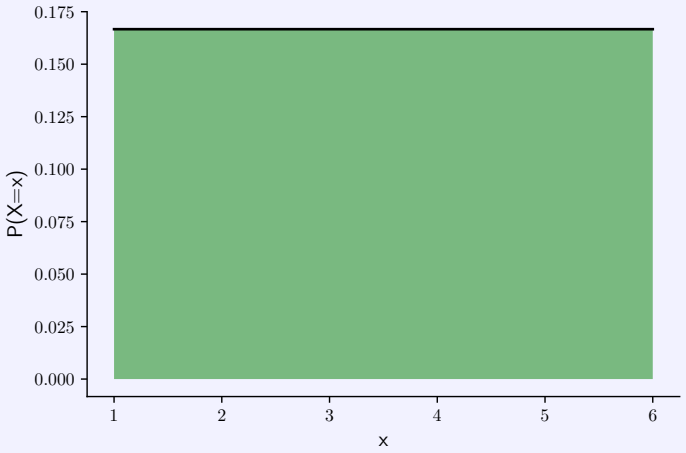


Figure 4-1: Probability Mass Function

Since the probability for all the random variables is same hence the graph is a flat line parallel to x-axis.

Now let us proceed with visualizing the probability density function

Visualizing Probability Density Function

Let us visualize the data given in table 4.2. Maximum individual have heights between 4.5 to 6 ft, this can be seen from the histogram in figure 4-2. Hence the probability for the same interval should also be highest. It can be observed from 4-2 that probability for this interval is the highest

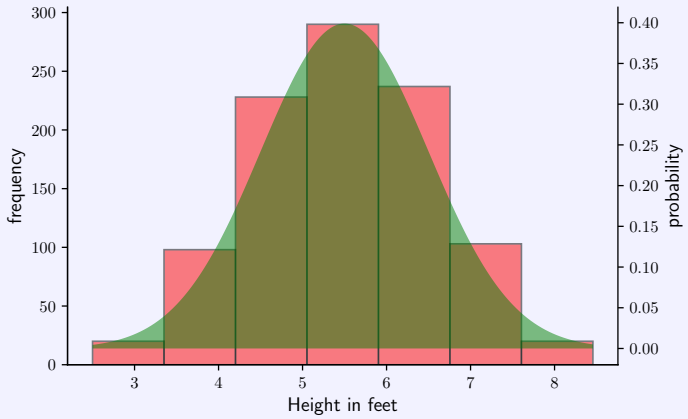


Figure 4-2: Probability Density Function

Since now we are familiar with how to interpret the visual representation of Probability Mass Function and Probability Distribution Function, we will proceed further with cumulative distribution function

Cumulative Distribution Function

Cumulative Distribution Function for a random variable X is used to define the sum of probabilities for $X \leq x$

If the random Variable X is discreet random variable then cumulative distribution function can be mathematically expressed as

$$F(x) = P(X \leq x) = \sum_{x_i}^{x} p(x)$$

The cumulative distribution function of a discreet random variable X will have the following properties:

- $0 \leq F(x) \leq 1$
- $\lim_{x \to \infty} F = 1$
- $\lim_{x \to -\infty} F = 0$
- $a<b \rightarrow F(a) < F(b)$

Cumulative Distribution Function for Discreet Random Variable

Let us take an example of rolling of a dice, the probability of getting a number less than or equal to 1 is $\frac{1}{6}$, However the probability of getting a number less than or equal to 2 is $P(X \leq 2) = p(1)+p(2) = \frac{1}{6}+\frac{1}{6} = \frac{2}{6}$

$\boldsymbol{X \leq x}$	$\boldsymbol{P(X \leq x)}$
$x \leq 1$	$\frac{1}{6}$
$x \leq 2$	$\frac{2}{6}$
$x \leq 3$	$\frac{3}{6}$
$x \leq 4$	$\frac{4}{6}$
$x \leq 5$	$\frac{5}{6}$
$x \leq 6$	1

Table 4.3: Cumulative Distribution Function for rolling of a dice

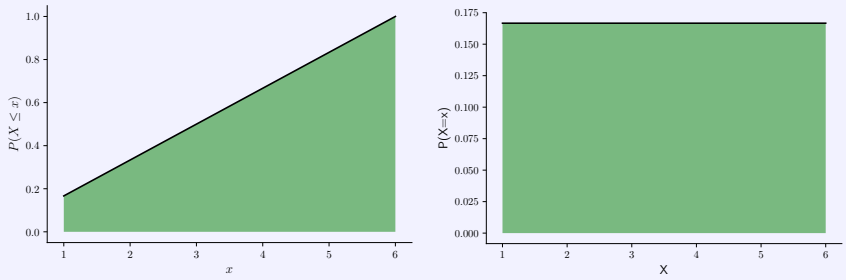


Figure 4-3: Comparison of Cumulative Distribution Function and Probability Mass Function for rolling of a dice

If the cumulative distribution function represents a continuous random variable X, then the cumulative distribution function can be mathematically represented as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

The cumulative distribution function of a continuous random variable X have the following characteristics:

- $0 \leq F(x) \leq 1$
- $\lim_{x \to \infty} F(x) = 1$
- $\lim_{x \to -\infty} F(x) = 0$
- $a < b \to F(a) \leq F(b)$

Cumulative Distribution Function for Probability Density Distribution

For understanding this we will refer back to the example shown in figure 4-2. We can clearly see a peak in the center of the graph indicating that if an individual is selected at random the probability of this individual being 5 to 6 feet tall is highest. However cumulative distribution increases as height increases, this happens because of the addition of probability as the height increases.

To understand it more intuitively, let us interpret it as follows, in our dataset the height of individuals range from 3 to 8 feet, the number of individuals who are very short i.e. around 3 ft and the number of individuals who are very tall i.e. around 8 feet are very less, hence there is very less probability that if an individual who is selected at random would be either very short or very tall. However when the concept of cumulative probability is applied to it the scenario changes. If we select the number of individuals who are less than or equal to 3 ft tall, then we cannot pick a lot of individuals as most of the people are taller than that i.e. cumulative frequency will be very less. However if we pick a the individuals who are more less than or equal to 8 ft tall, then this would include majority of people, hence the cumulative frequency will be high. This could be better visualized by figure 4-4

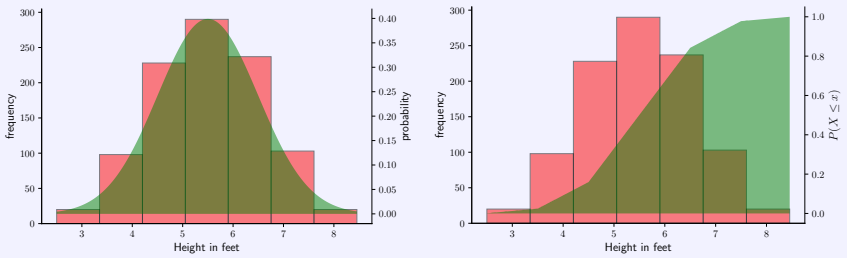


Figure 4-4: Comparison of Cumulative Distribution Function and Probability Density Function

4.2 Bernoulli Distribution Function

In order to understand bernoulli distribution we first need to understand bernoulli trails

Bernoulli Trial

A Bernoulli Trial is an experiment with only one trial and two possible outcomes.

Whenever we are dealing with the bernoulli distribution we either have the probability of a single event occuring or past data from repeated trials.

While using the Bernoulli Distribution we often have to take the following into consideration:

- Bernoulli Distribution has two outcomes p and $1-p$
- The event with higher probability is considered to be p
- We often have to assign which outcome is 0 and which outcome is 1, usually the outcome with higher probability is assigned 1 and the other is assigned 0, i.e. p is conventionally assigned 1 and $1-p$ is assigned 0 this makes sure that the expected outcome is p

Thus based on the above we can say that for a bernoulli distribution

$$\boxed{P(X=1)=p}$$

$$\boxed{P(X=0)=1-p}$$

Also

$$\boxed{\sigma^2=p(1-p)}$$

$$\boxed{\sigma=\sqrt{p(1-p)}}$$

$$\boxed{\mu=p}$$

Examples of a Bernoulli Trial include flipping of a coin

Bernoulli Distribution Function

Bernoulli Distribution Function is used to predict sucess i.e. p or failure i.e. q in a Bernoulli Trail

Bernoulli Distribution Function

Let us consider the flipping of a biased of coin, the weight distribution of the coin is such that it shows head 60 percent of the times. Calculate the probability of sucess and failure.

It is given that

$$P(X = 1) = p = 0.6$$

$$P(X = 0) = 1 - p = 0.4$$

$$\sigma^2 = 0.6 \times 0.4$$

$$\mu = 0.6$$

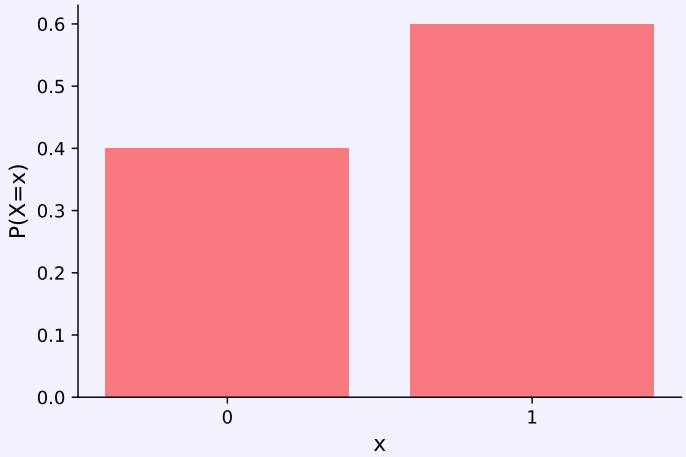


Figure 4-5: Bernoulli Distribution

4.3 Binomial Distribution Function

Binomial Distribution Function

Binomial Distribution Function is a probability mass function used to calculate the probability of occurrence of an event exactly x number of times in n attempts. If p is the probability of outcome y in a single independent trial then binomial probability function can be given as

$$\boxed{P(X = x) = C_x^n p^x (1-p)^{n-x}}$$

Variance, standard deviation and mean can be calculated as

$$\boxed{\sigma^2 = np(1-p)}$$

$$\boxed{\sigma = \sqrt{np(1-p)}}$$

$$\boxed{\mu = np}$$

Binomial Distribution Function

Let us consider a factory manufactures one car every minute and there is a 40 percent chance that the car produced will be defective. What is the probability that the factory will make 3 defective cars in next 5 minutes ?

Here we have,

$$x = 3$$

$$n = 5$$

$$p = 0.4$$

Thus

$$P(X = 3) = C_3^5 0.4^3 (1 - 0.4)^{5-3}$$

$$P(X = 3) = \frac{5!}{3!(5-3)!} 0.4^3 (1 - 0.4)^{5-3}$$

$$P(X = 3) = \frac{5 \times 4 \times 3!}{3!2!} 0.4^3 0.6^2$$

$$P(X = 3) = 10 \times 0.064 \times 0.36$$

$$P(X = 3) = 0.2304$$

Thus there is a 23.04 percent chance that the next 3 out of 5 cars will be defective

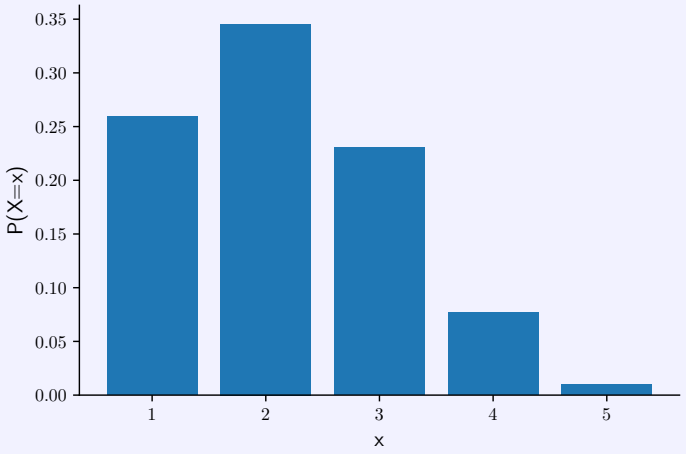
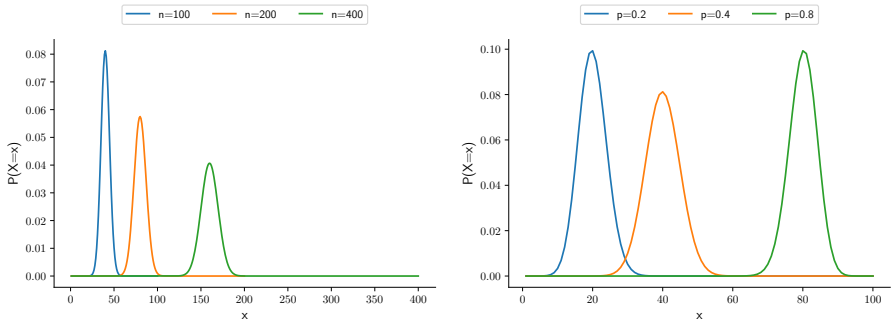


Figure 4-6: Binomial Distribution



(a) Binomial Probability Distribution for different values of n

(b) Binomial Probability Distribution for different values of p

Figure 4-7: Comparison of Binomial Probability Distribution with varying parameters

The formulae for binomial probability distribution depends on

- n
- p
- x

However the probability distribution is not affected by x. Binomial Distribution shows significant variation for different values of p and n. As the value of n increases for a given value of p the peak of the curve shifts towards right but decreases in height. It must be noted that peak of the curve is always at the mean of the distribution.

For a fixed value of n the peak of the curve shifts towards right as the value of p increases. However the height of the peak follows no significant pattern.

Thus we can say that the for Binomial Probability Distribution the peak of the curve always coincides with the mean of the distribution

4.4 Poisson Distribution Function

Poisson Distribution Function

Poisson Distribution is a Probability Mass Function used to predict the probability of specific number of occurrences of an event in a given time frame, distance or length by using the frequency of occurrence of that event λ from previous observations. If an event occurs λ times in a unit distance, time or length then the probability of that event occurring x times in another unit of time, length or distance can be given as

$$\boxed{P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}}$$

The mean and standard deviation of the distribution is given as

$$\boxed{\mu = \lambda}$$

$$\boxed{\sigma = \sqrt{\mu}}$$

Poisson Distribution Function

If 10 customers arrive at billing counter of a supermarket in an hour, what is the probability that 15 customers would come at the billing counter in the next 1 hour?

Here we know

$$\lambda = 10$$

$$x = 15$$

Therefore

$$P(X = 15) = \frac{2.718^{-10}10^{15}}{15!}$$

$$P(X = 15) = 0.03472$$

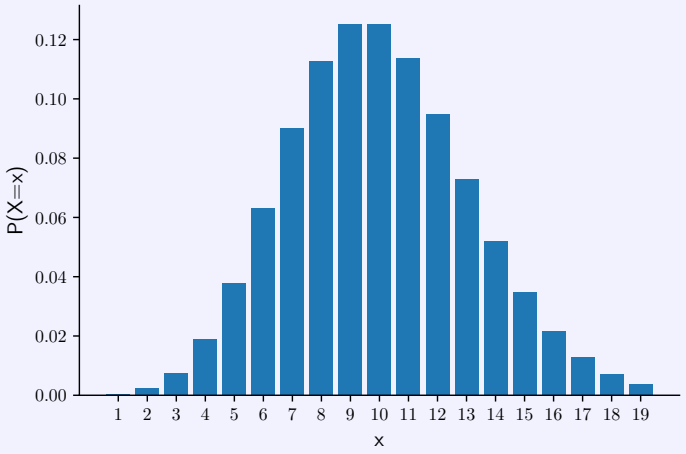


Figure 4-8: Poisson Distribution

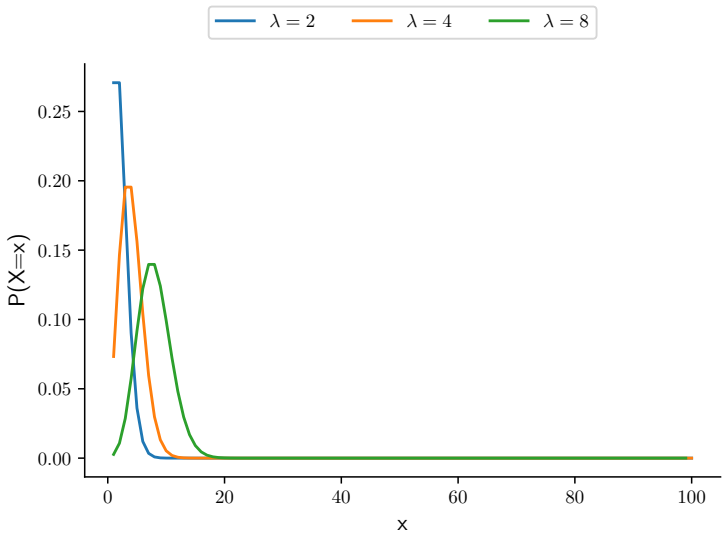


Figure 4-9: Poisson Distribution Function for different values of λ

As we can see from figure 4-9 probability distribution for Poisson Distribution Function depends on λ. As the value of λ increases the peak of the curve shifts towards right and becomes smaller.

4.5 Exponential Distribution Function

Exponential Distribution Function can be seen as inverse of Poisson Distribution Function. Poisson Distribution uses the number of events per unit time,length or distance where as exponential distribution is used to predict the time between two consecutive events. For example Poisson Distribution looks at number of people arriving at a billing counter in one hour where as exponential distribution will look at the time between two consecutive arrivals of costumer.

Exponential Probability Distribution

Exponential Distribution Function is a probability density function used to calculate the probability of time between events. It can be mathematically expressed as

$$P(X = x) = \frac{e^{\frac{-x}{\mu}}}{\mu}$$

where

$$\mu = \frac{1}{\lambda}$$

For a distribution to be classified as exponential probability distribution, the following must be true:

- Events must occur at a constant rate
- Events must be independent of each other.

The mean and variance of Exponential distribution function can be calculated as follows

$$\boxed{\mu = \frac{1}{\lambda}}$$

$$\boxed{\sigma^2 = \frac{1}{\lambda^2}}$$

The Cumulative Distribution Function for Exponential Probability Distribution is given as

$$\boxed{P(X < x) = 1 - e^{\frac{-x}{\mu}}}$$

Any Piosson Distribution can be converted to exponential distribution as follows

Exponential Distribution Function

Let us consider 6 unique visitors arrive at an e-commerce website every minute. What is the probability that another visitor is gonna arrive in

- within next 5 seconds
- after 10 second
- exactly at 15 seconds

Here

$$\lambda = 6 \text{ per minute}$$

Therefore

$$\mu = \frac{1}{6} \text{ minute} = 10 \text{ seconds}$$

Thus

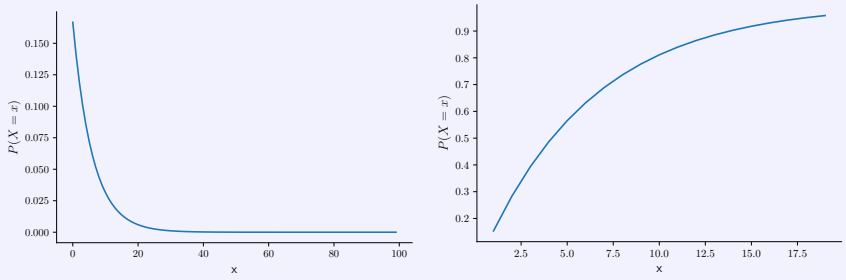
$$P(X < 5) = 1 - e^{\frac{-5}{20}} = 0.393$$

Similarly

$$P(X > 10) = 1 - (1 - e^{\frac{-10}{20}})$$

$$P(X > 10) = e^{\frac{-10}{20}} = 0.607$$

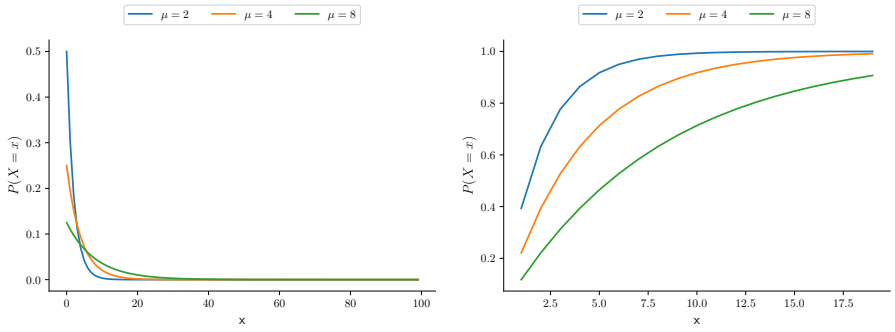
$$P(X = 15) = \frac{e^{\frac{-15}{10}}}{10} = 0.022$$



(a) Probability Density Function for Exponential Distribution

(b) Cumulative Distribution Function for Exponential Distribution

Figure 4-10: Probability Density Function and Cumulative Distribution Function for Exponential Distribution



(a) Effect of μ on PDF for Exponential PRobability Distribution

(b) Effect of μ on CDF for Exponential PRobability Distribution

Figure 4-11: Probability Density Function and Cumulative Distribution Function for Exponential Distribution

As can be seen from figure 4-11 that value of μ affects the probability distribution for both PDF and CDF of exponential function. As the value of μ increases the maximum probability of the event decreases. But this is not true for CDF, as the value of μ decreases the cumulative distribution tends to take a higher value for higher values of x.

4.6 Geometric Distribution Function

Geometric Distribution

Geometric Distribution Function is a Probability Mass Function used to calculate that your first success will be on the n^{th} try. It can be mathematically expressed as

$$\boxed{P(X = x) = (q)^{x-1}p}$$

The mean and variance of Geometric Distribution Function are given as

$$\boxed{\mu = \frac{1-p}{p}}$$

$$\boxed{\sigma^2 = \frac{1-p}{p^2}}$$

Let us elaborate with an example

Geometric Distribution

Let us consider that the probability of a basketball player making a successful free throw is 20 %. What is the probability that the first successful free throw will be after 10 failed trials ?
We have

$$p = \frac{20}{100}$$

$$q = 1 - \frac{20}{100} = \frac{80}{100}$$

Thus

$$P(X = 10) = (0.8)^{10-1} \times 0.2$$

$$P(X = 10) = 0.027$$

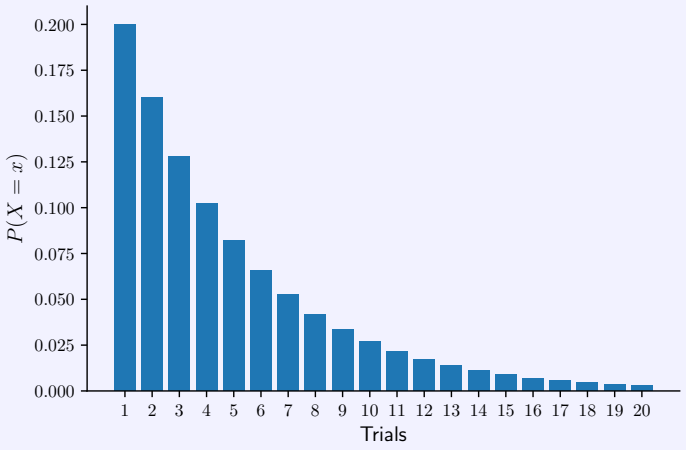


Figure 4-12: Geometric Distribution

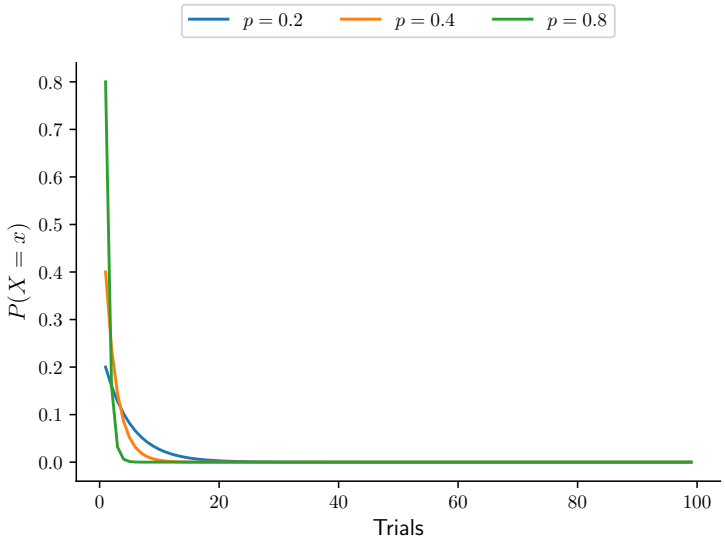


Figure 4-13: Geometric Distribution

The effect of probability of success on Geometric Probability Distribution can be seen in figure 4-13.

4.7 Negative Binomial Distribution Function

Negative Binomial Distribution

Negative Binomial Function is a Probability Mass Function used to calculate the number of trials(k) required to get x^{th} success.

$$\boxed{P(X = x) = C_{k-1}^{x-1} p^k (1-p)^{k-x}}$$

Let us illustrate with an example.

Negative Binomial Distribution

If a basketball player makes 3 successful free throws for every 5 attempts, what is the probability that 7^{th} successful free throw will be made in the 13^{th} attempt ?

We have

$$p = \frac{3}{5}$$

$$x = 7$$

$$k = 13$$

$$P(X = 13) = \frac{(13-1)!}{(7-1)!(13-7)!} \times \left(\frac{3}{5}\right)^7 \times \left(1 - \frac{3}{5}\right)^{13-7}$$

$$\frac{12!}{6! \times 6!} \times \left(\frac{3}{5}\right)^7 \times \left(\frac{2}{5}\right)^6 = 0.005$$

4.8 Normal Distribution Function

As we did in a previous section, here also we will refrain from providing a formal definition and will focus more on the idea of Normal Distribution.

Normal Distribution can be seen in many spectrum of life. Before we explore these examples let us briefly look at what is normal distribution.

Normal Distribution is a type of Probability Density Function. If a random variable is normally distributed than the data points will cluster around the mean i.e. most of the values will be close to the mean and as we move away from the mean the frequency of values will decrease. As shown in figure 4-14.

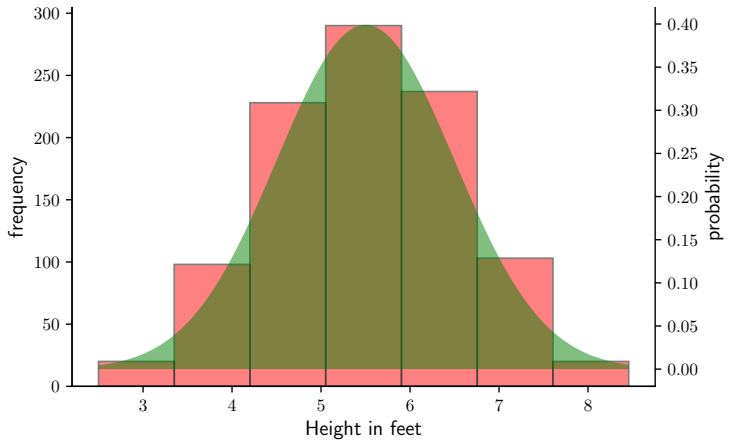


Figure 4-14: Normal Distribution

If we closely observe figure 4-14 we can observe that it is a distribution plot and an area graph. The mean of the distribution is 5.5 and most of the values lie in the same interval as that of mean hence we can observe that the height of the graph is highest since most of the values lie in the same interval as that of the mean thus if a value is picked at random the probability that it will lie in the same interval as that of the mean is highest, hence we can see that the probability curve peaks out at the mean.

In a normal distribution the probability of random variable X being equal to x can be calculated as

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

The above equation will give the area to the left of the curve for a given value of x.

Normal Distribution

If the heights of a group of individuals is normally distributed with a mean of 5.5 feet and standard deviation of 0.5 feet, what is the probability that an individual selected at random will have a height of 4.5 feet?

We know tha

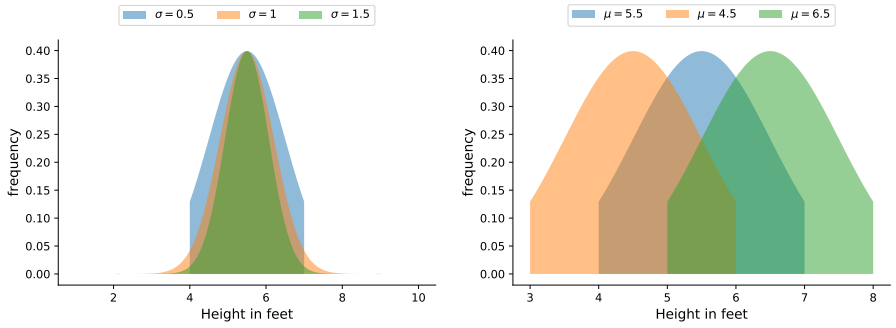
$$\mu = 5.5$$

$$\sigma = 0.5$$

$$P(X = 4.5) = \frac{1}{0.5 \times \sqrt{2\pi}} e^{\frac{-(4.5-5.5)^2}{2\times(0.5)^2}} = 0.107$$

As can be seen from above equation, the probability of random variable X in normally distributed data depends on σ and μ.

4.8.1 Effect of standard deviation and mean



(a) Effect of σ on Normal Probability Distribution

(b) Effect of μ on Exponential Probability Distribution

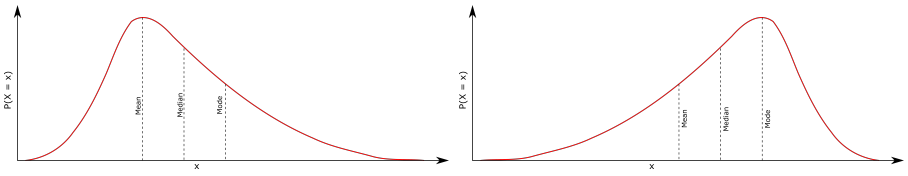
Figure 4-15: Effect of μ and σ on Normal Probability Distribution

As discussed above, the mean and standard deviation are the two driving factors in normal distribution. As mentioned above, in a normal distribution the data clusters around the mean, hence if we change the mean the peak of the bell curve will shift accordingly.

The mean of the distribution determines the location of the peak of the curve. Similarly Shape of the bell curve is determined by the standard deviation. The higher the value of standard deviation the more widespread will be the shape of the curve.

4.8.2 Skewness and Kurtosis

As can be seen in figure 4-15, the normal distribution is symmetrically distributed around the mean. But this is not always the case. may times the values tend to cluster towards one end of the spectrum. This clustering of values is what is known as skew.



(a) Right Skewed(Positive Skew) Normal Distribution Curve

(b) Left Skewed(Negitive Skew) Normal Distribution Curve

Figure 4-16: Skewness in Normal Distribution Curve

In a normal distribution the data is equally clustered around the mean i.e. there are equal number of observation which are either smaller or larger than mean.

If the number of observations greater than the mean have a higher frequency than the number of observation smaller than the mean, the data is said to be Right Skewewd or Positively Skewed.

If the data is positively skewed than the mean, median and mode will have the following relationship

$$mean < median < mode$$

If the number of observations smaller than the mean have a higher frequency than the number of observation greater than the mean, the data is said to be Left Skewed or Negatively Skewed.

For a left skewed data the following is true

$$mean > median > mode$$

If the data is skewed than there is a sepration between mean median and mode. This sepration has been made the basis of skewness calculation however now a moment based approach is followed based on the method of moments. MEthod of moments is not the focus of this chapter, hence we will not go deeper in to the method of moments.

Based on the method of moments, skewness for population can be calculated as

$$\boxed{skew_{sample} = \frac{n}{(n-1)(n-2)} \frac{\sum (x - \bar{x})^3}{s^3}}$$

The formulae for population skewness can be given as

$$\boxed{skew_{population} = \frac{1}{n} \frac{\sum (x - \mu)^3}{\sigma^3}}$$

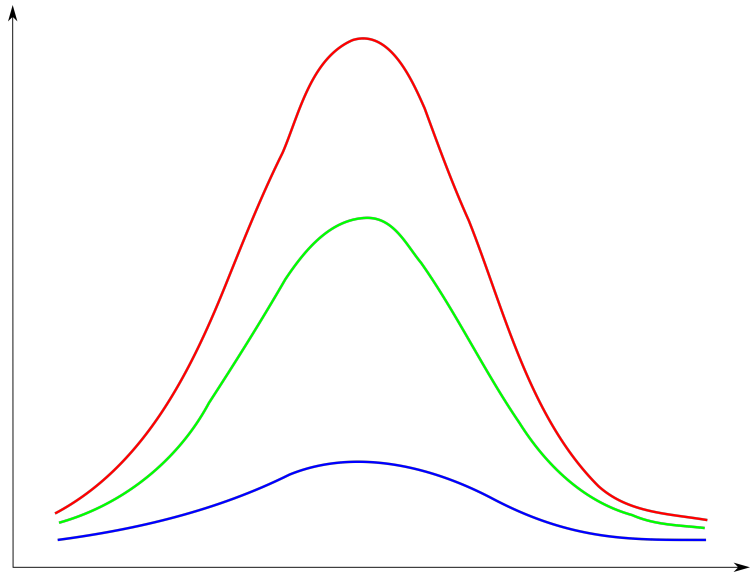


Figure 4-17: Normal Distribution with Kurtosis

As can be seen in figure 4-17 the three curves have the same spread but have different peaks. The curve with the highest peak will have the highest kurtosis. **Thus we can say that kurtosis is a measure of the peakedness of the data.**

At this point a further understanding of kurtosis is not required hence will proceed further without touching upon the formulae for kurtosis.

4.8.3 Standard Normal Distribution

Before the discovery of computers, a lot of calculation were done with hands, PDF for Normal Distribution were also one such calculations. However due to the complexity of the calculation, table were formed with pre-calculated values. However these tables are only valid for **standard normal distribution**.

A Standard Normal Distribution is Normal Distribution with 0 as mean and standard deviation of 1. To convert a normal distribution to standard normal distribution we need to calculate the z-score of each data point and plot the density against these z-scores instead of data points as shown in figure 4-18

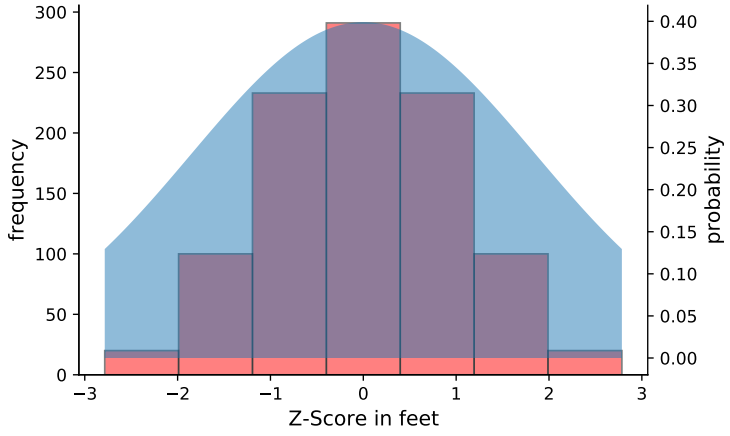


Figure 4-18: Standard Normal Distribution