Learning Probability and Statistics for Machine Learning and Data Science

Parth Sinha
Student of AI & ML
Oxford Brookes University
Email: sinhaparth555@gmail.com

***Abstract*—This is a note for Probability and Statistics for Machine Learning and Data Science, from YouTube channel Data Dissection. Link: https://www.youtube.com/watch?v=vEwe_1b1Df0.**

***Index Terms*—data science, machine learning, probability and statistics**

I. Probability

A. Introduction

Probability is the measure of the likelihood or chances that an event will occur.
For example, person A and person B are working in a company, so what are the changes for person A who is wealthier than person B. So we need to quantify how much person A earns more than person B? This quantification is known as probability.
The probability range between

$$0 \leq P(E) \leq 1$$

where 0 means that the probability of the event occurring is null and 1 means that the probability of the event occurring is true. For example $P(sun\ will\ rise\ for\ west) = 0$ and similarly $P(sun\ will\ rise\ from\ east) = 1$.
Sample Space: The sample space of an experiment or random trial is the set of all possible outcomes or results of that experiment.
Example:
Experiment: Rolling a standard six-sided die.
Sample Space: $\{1, 2, 3, 4, 5, 6\}$
This means that the possible outcomes are rolling a 1, 2, 3, 4, 5, or 6.
The sample space is the set of all possible outcomes.
Case $\Rightarrow$ Odd no: $\{1, 3, 5\}$
An event is something that happens, and event is always a subset of sample space. $E \subseteq S$ where E = event and S = sample space.
For example

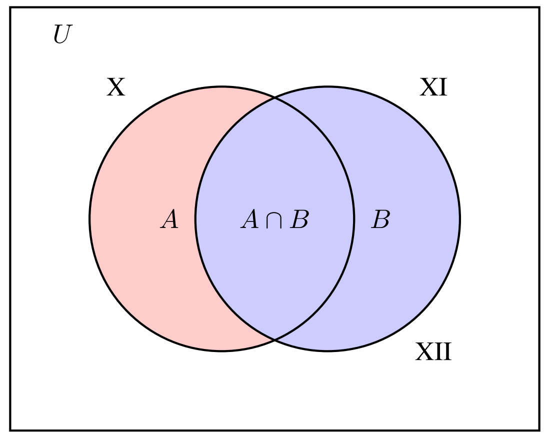
$$E = \{1, 3, 5\} \Rightarrow n(E) = 3$$

$$S = \{1, 2, 3, 4, 5, 6\} \Rightarrow n(S) = 6$$

$$P(E) = \frac{n(E)}{n(S)} = \frac{3}{6} = \frac{1}{2} = 0.5 = 50\%$$

If the dice is rolled fairly 100 times, 50% times it will give odd numbers.

B. Sets for probability



U: All Students
A: Class IX
B: Computer Science

Fig. 1: Venn diagram showing relationship between Class IX students and Computer Science students

1) Intersection: Students who are in class IX and the students who have opted for computer science $\Rightarrow n(A \cap B)$

2) Union: Students who are in class IX and the students who are in computer science $\Rightarrow n(A \cup B)$

$$n(A \cap B) = n(A) + n(B) - n(A \cap B)$$

3) Null Sets: Class XI students opted Economics and also Class XI Science batch student have opted for Economics $n(A \cap B) = \emptyset$

C. Fair Events

Fair events in probability are events where each possible outcome has an equal chance of occurring. This means that there is no bias or favoritism towards any particular outcome.
For example: Coin in which tails and heads equally likely to occurs $\Rightarrow P(H) = 0.5$ where H is head of the coin.
Take a fair coin and toss n times.

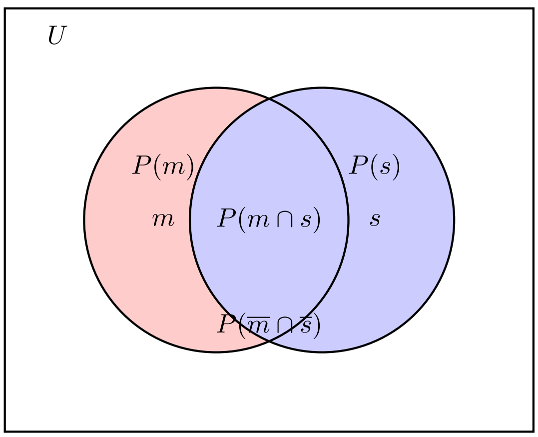
In table I when $n \to \infty$ the probability is head is 50% i.e., $P(H) \to 0.5$.

TABLE I: Occurrence of of Heads n times

n	Occurrence of head
10	6
100	52
1000	512
.	.
.	.
.	.
1 million	500010

D. Joint Probability

Joint probability is a fundamental concept in probability theory that measures the likelihood of two or more events occurring simultaneously. It's often represented as $P(A \cap B)$, where A and B are the events in question.
For example, U is the total malaria patient, m no. of patient getting medicine and s is no. of patient survive, refer to figure 2. So, $P(m \cap s) \rightarrow$ no. of malaria patient those are given medicine and also survived.



U: Total patients
m: Patients getting medicine
s: Patients who survive

Fig. 2: Joint probability distribution of malaria patients receiving medicine and survival

For example
So, if total malaria patient are 500, 50 of them got medicine and 200 survived, and patient who got medicine and survived are 25, then:

$$U = 500$$

$$m = 50$$

$$s = 200$$

$$m \cap s = 25$$

$$\text{Then, } P(m \cap s) = \frac{25}{500} = \frac{1}{20} = 0.05$$

Question: Fair dice throw:
$A : even\ no.$
$B : prime\ no.$
$C : odd\ no.$

Ans *a)* $P(A \cap B \cap C)$

$$S = \{1, 2, 3, 4, 5, 6\}$$

$$A : \{2, 4, 6\}$$

$$B : \{2, 3, 5\}$$

$$C : \{1, 3, 5\}$$

$$n(A \cap B \cap C) = 0$$

$$P(A \cap B \cap C) = \frac{0}{6} = 0 = \emptyset$$

Ans *b)* $P(even\ or\ prime)$

$$A : \{2, 4, 6\}$$

$$B : \{2, 3, 5\}$$

$$n(A \cup B) = n(5)$$

$$P(A \cup B) = \frac{n(A \cup B)}{n(U)} = \frac{5}{6} = 0.83$$

E. Conditional Probability

Conditional probability is a measure of the probability of an event occurring, given that another event has already occurred.[**?**] It essentially refines our understanding of probability by considering the impact of prior knowledge or conditions.
Equation,

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

For s is survived and m is medicine given to patient then, $P(s|m) \rightarrow$ Probability that the patient survives, given that he's been administrated the medicine. Then,

$$s = 200$$

$$m = 50$$

$$m \cap s = 25$$

$$U = 500$$

$$P(s|m) = \frac{P(s \cap m)}{P(m)}$$

$$\frac{\frac{n(s \cap m)}{n(U)}}{\frac{n(m)}{n(U)}}$$

$$\frac{n(s \cap m)}{n(m)} = \frac{25}{50} = \frac{1}{2} = 0.5$$

F. Types of events

1) Mutually Exclusive events: two events are considered mutually exclusive if they cannot both occur simultaneously. This means that if one event happens, the other event cannot happen.
Fair dice is throw:
$A : no.\ is\ even = \{2, 4, 6\}$
$B : no.\ is\ odd = \{1, 3, 5\}$

$$A \cap B = 0 = \emptyset$$
$$P(A \cap B) = \emptyset$$

2) Independent events: In probability theory, two events are considered independent if the outcome of one event does not affect the probability of the other event occurring.[**?**]
For example:
Total cards: 52
$A : card\ is\ a\ king$
$B : Card\ is\ a\ heart$

$$P(A) = \frac{4}{52} = \frac{1}{13}$$
$$P(B) = \frac{13}{52} = \frac{1}{4}$$
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{52}}{\frac{1}{4}} = \frac{4}{52} = \frac{1}{13}$$
$$P(A|B) = P(A)$$
$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{52}}{\frac{1}{13}} = \frac{13}{52} = \frac{1}{4}$$
$$P(B|A) = P(B)$$

If the occurrence of one event does not alter the changes of occurrence of another event knows as independent events.

$$P(E_1|E_2) = P(E_1)$$
$$P(E_2|E_1) = P(E_2)$$

G. Co-joint Probability

Joint probability is a fundamental concept in probability theory that measures the likelihood of two or more events occurring simultaneously.
Equation,

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

Question: n-coins toss, probability all heads?
$Sol^n : E_i : coin\ is\ head$

$$P(E_1 \cap E_2 \cap E_3 ...) = P(E_1) \times P(E_2) \times P(E_3) ...$$
$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} ...$$
$$= \left(\frac{1}{2}\right)^n$$

Lets see,

$$P(E_1 \cap E_2 \cap E_3) = P(E_1 \cap (E_2 \cap E_3))$$
$$= P(E_1) \cdot P(E_2 \cap E_3)$$
$$= P(E_1) \cdot P(E_2) \cdot P(E_3)$$

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$
$$P(E_1) \cdot P(E_2|E_1) = P(E_1 \cap E_2)$$
$$P(E_1)P(E_2) = P(E_1 \cap E_2)$$

H. Bayes' Theorem:

Bayes' Theorem is a mathematical formula that helps you update your beliefs (or probabilities) about something based on new evidence.[**?**] It allows you to go from a prior belief to a posterior belief after observing new data.[**?**]

Bayes' Theorem problem from the book Think Bayes'[**?**]
The Cookie problem
Suppose there are two bowls of cookies. Bowl 1 contains 30 vanilla cookies and 10 chocolate cookies. Bowl 2 contains 20 of each.
Now suppose you choose one of the bowls at random and, without looking select a cookie at random. The cookie is vanilla. What is the probability that it came from Bowl 1?
This is a conditional probability, we want $p(Bowl\ 1|vanilla)$, but it is not obvious how to compute it. If I asked a different question - the probability of a vanilla cookie given Bowl 1 - it would be easy:

$$p(vanilla|Bowl\ 1) = \frac{3}{4}$$

Sadly, $p(A|B)$ is *not* the same as $p(B|A)$, but there is a way to get from one to the other: Bayes' theorem.

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

Proof,

$$P(A \cap B) = P(B \cap A)$$
$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$
$$P(B \cap A) = P(A) \cdot P(B|A)$$
$$P(A \cap B) = P(A) \cdot P(B|A)$$
$$\Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$P(A|B) \cdot P(B) = P(A) \cdot P(B|A)$$
$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

TABLE II: True & false table

Question 1	Question 2	Question 3
True	True	True
True	True	False
True	False	True
True	False	False
False	True	True
False	True	False
False	False	True
False	False	False

Cookie problem continue:

$$p(Bowl\ 1|vanilla) = \frac{p(Bowl\ 1) \cdot p(vanilla|Bowl\ 1)}{p(vanilla)}$$
$$= \frac{\frac{1}{2} \cdot \frac{3}{4}}{\frac{50}{80}}$$
$$= \frac{3}{5}$$
$$P(Bowl\ 1|vanilla) = \frac{P(Bowl\ 1 \cap vanilla)}{P(vanilla)}$$

Why we use Bayes' theorem in Data Science and Machine Learning
Let say we want to know what are the chances of getting heart attack to a person, we will note down some conditions like, cholesterol, BMI, etc,.
Lets say $P(H) \Rightarrow P(H|Ch \uparrow)$ the probability of the heart attack will be high and if the $P(H|Ch \downarrow)$ the probability of the heart attack will be low, Similarly with BMI.

I. Combinatorics

Question $\Rightarrow$ Given that we have 3 questions of True and False, in how many ways we can answer them?
Sol^n: Each question has 2 possible answers (True or False). Since there are 3 questions, the total number of ways to answer them is $2 \times 2 \times 2 = 2^3 = 8$. Please see the table II.

J. Permutations

Permutations in mathematics refer to the different arrangements of a set of objects where the order of the objects matters. For example $(i, j) \neq (j, i)$, ex. Student 1 got rank 1 and student 2 is got rank 2 but it's not possible that student 2 get rank 1 and student 1 get rank 2.

$$(S_1, S_2) \neq (S_2, S_1)$$

Question $\Rightarrow$ There are 3 alphabets (a, b, c), how many ways to arrange them?
$Sol^n \Rightarrow$ There are 6 ways to arrange the alphabets (a, b, c), see table III.

Formula for factorial:

$$n! = n \times (n-1) \times (n-2) \times ... \times 1$$

Lets take an example, there are n items and we have to arrange them in r places. So lets take a look:
Lets say we have 5 objects i.e., (a, b, c, d, e) and we have to arrange them in 2 places, $n = 5$ and $r = 2$, Then using the formula we get $5 \times (5-1) = 5 \times 4 = 20$.
Lets say we have N objects and we need to arrange them in r places.
$(n)\ and\ (n-1)\ and\ (n-2)...(n-r+1)$
No. of ways $= n \times (n-1) \times (n-2)...(n-r+1)$

$$= \frac{(n \times (n-1) \times (n-2)...(n-r+1)) \times ((n-r) \times (n-r-1)...1}{((n-r) \times (n-r-1) \times ... \times 1}$$

$$\Rightarrow \frac{n!}{(n-r)!}$$

$$^nP_r = \frac{n!}{(n-r)!}$$

TABLE III: Ways to arrange aplphabets (a, b, c)

Position 1	Position 2	Position 3
a	b	c
a	c	b
b	a	c
b	c	a
c	a	b
c	b	a

TABLE IV: Students combinations

Combinations	*Student 1*	*Student 2*	*Student 3*
1	S_1	S_2	S_3
2	S_1	S_2	S_4
3	S_1	S_3	S_4
4	S_2	S_3	S_4

K. Combinations

Combinations in mathematics refer to the selection of a subset of items from a larger set, where the order of selection does not matter. The section of objects are equal i.e., $(i, j) = (j, i)$. For example, if we have to select student S_1 or S_2 in any combination it's equal, i.e., $(S_1, S_2) = (S_2, S_1)$
Lets take little complicated example, we need to select 3 students out of 4 students, and we represent all students with S_1, S_2, S_3 and S_4.
Sol^n: Total no. of arrangements = 24 and no. of selections = $\frac{24}{3!}$
$= 4$
You can also see the table IV.

No. of arrangements $\Rightarrow {}^nP_r = \frac{n!}{(n-r)!}$

$$\frac{\frac{n!}{(n-r)!}}{r!}$$

No. of selections = $\frac{n!}{(n-r)! \cdot r!}$

$$^nC_r = \frac{n!}{(n-r)! \cdot r!}$$

Lets take an example, we have set of alphabets $\{a, b, c, d, e\}$ and we want to select only 3 of them, lets solve it with the

P_1P_2	$P_3P_4P_5$
P_2P_3	$P_2P_4P_5$
P_1P_4	$P_2P_3P_3$
P_1P_3	$P_2P_4P_3$

TABLE V: Selection and Removing elements

equation.

$$^nC_r = \frac{n!}{(n-r)!r!} \Rightarrow \frac{5!}{(5-3)!3!} = 10$$

So we can have 10 selections.

Question $\rightarrow$ $^nC_0 +^n C_1 +^n C_2 + ... +^n C_n =?$
$^nC_r \rightarrow no.\ of\ selections$
$Sol^n \Rightarrow \{1, 2, 3\}$
Lets say I don't want to select any one from the set, so $\rightarrow r = 1$ and so on, so we have now

$$^nC_0 = 1 = \{\}$$

$$^nC_1 = \{1\},\ \{2\},\ \{3\}$$

$$^nC_2 = \{1, 2\},\ \{1, 3\},\ \{2, 3\}$$

$$^nC_3 = \{1, 2, 3\}$$

$S = \{1, 2, 3\}$
$\{1\},\ \{2\}...$
$^nC_0 \rightarrow^n C_n$
$S = 2^n$
$\{a_1,\ a_2,\ a_3, ...,\ a_n\}$
So basically solving this we get:

$$^nC_0 +^n C_1 +^n C_2 + ... +^n C_n = 2^n$$

Question $\rightarrow$ $^nC_n =^n C_{n-r}$

$$\frac{n!}{(n-r)!r!} = \frac{n!}{(n-(n-r))!(n-r)!}$$

$$= \frac{n!}{n!(n-r)!}$$

Then above equation doesn't work
Let $5 \rightarrow n$ and we need to select only 2 i.e., $2 \rightarrow r$ so it will become $^5C_2 = x$, where x is the result.
Now again, let's say $5 \rightarrow n$ and we need to removed 3 i.e., $3 \rightarrow r$ so it will become $^5C_3$, let take set $\{P_1,\ P_2,\ P_3,\ P_4,\ P_5\}$, now you can take a look at table V.
So as we look at table V we can say that

$$^3C_2 =^5 C_3$$

so we get

$$^nC_r =^n C_{n-r}$$

II. Probability Distribution

A probability distribution is a mathematical function that describes the likelihood of different possible values for a random variable within a given range.[**?**]

A. Random variables

A random variable is a function that assigns a numerical value to each possible outcome of a random experiment.[**?**]

For example, if we roll the dice i.e., $S = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$ and every number is countable, another we assign grades to students i.e.m $S = \{A,\ B,\ C,\ D,\ E,\ F\}$

1) Discrete: Set of countable.

2) Continuous: Set of all possible values within an interval.

3) Ordinal: Type of statistical data where the variables have a defined order or ranking.

Mean: the mean is the average of a set of numbers.
Let say we take 5 students age, and we have $10,\ 11,\ 12,\ 9$ and 8, so we can get the mean of the student's age.

$$Mean: \frac{10 + 11 + 12 + 9 + 8}{5} = \frac{50}{5} = 10$$

So the mean of the students age is 10.
But here there is a limit cause it can hide some important information, lets see it with an example, let's take 5 people, and 10, 12, 13, 10, 1003 respectively in there house, so if we take mean of the light bulb then

$$M = \frac{10 + 12 + 13 + 10 + 1003}{5} = 209.6$$

, so basically if you look the mean is $209.6 \approx 210$ bulbs in a house, but we know that's not true. To solve this problem we use another term which is $Median$. Now we again take the same example, first we need to sort the sample set in ascending order so it's $\{10,\ 10,\ 12,\ 13,\ 1003\}$, $Median$ suggest that we should take the middle value, so the middle value of the light bulb sample set is 12, The $Median$ is 12 so the average light bulb in a house is 12.
Now $Mode$, in this the answer is data which is has high frequency is know as $Mode$. Lets take a data set of light bulb, $B = \{10,\ 10,\ 11,\ 11,\ 13,\ 10,\ 12,\ 13,\ 12\}$ in the data set the frequency is $10 \rightarrow 3,\ 11 \rightarrow 2,\ 12 \rightarrow 2,\ 13 \rightarrow 2$, highest frequency is of 10 so it the $Mode$.

4) Standard deviations: Standard deviation is a statistical measure that quantifies the amount of variation or dispersion of a set of data values. In simpler terms, it tells you how spread out the numbers are in relation to the mean (average) of the dataset.[**?**]

5) Variance: is a statistical measure that quantifies the average squared deviation of each data point from the mean.[**?**] In simpler terms, it tells you how much the data points in a set are spread out around their average value.[**?**]

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

$$Standard\ deviance(\sigma): \sqrt{variance}$$

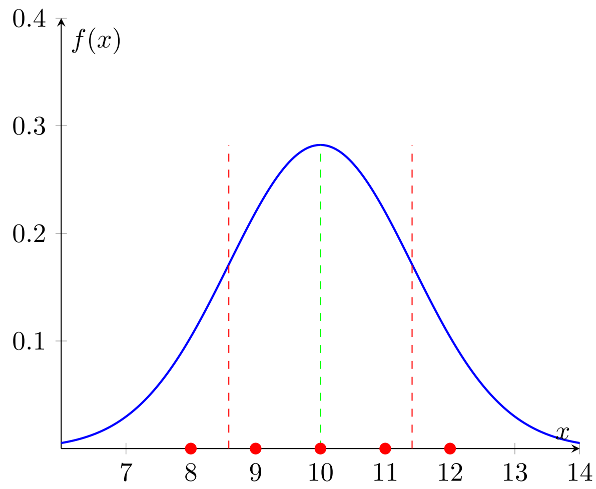


Fig. 3: Normal distribution of data points with mean $\mu = 10$ and standard deviation $\sigma = 1.414$

After take point $\{8,\ 9,\ 10,\ 11,\ 12\}$ the mean is 10 and solve it with the standard deviation:

$$\sigma^2 = \frac{(8-10)^2 + (9-10)^2 + 0 + (11-10)^2 + (12-10)^2}{5}$$

$$\sigma^2 = \frac{10}{5} = 2$$

$$\sigma = \sqrt{2} = 1.414$$

Look at figure 3.

6) Median absolute deviation: The Median Absolute Deviation (MAD) is a robust measure of statistical dispersion. It represents the median of the absolute deviations of the data points from the data's median.[**?**]

7) Mean absolute deviation: The Mean Absolute Deviation (MAD) is a statistical measure that quantifies the average absolute distance between each data point and the mean of the dataset. In simpler terms, it tells you how spread out the numbers are in relation to the mean.[**?**]

Formulas

$$MAD = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$$

$$MAD = median(|x_i - median(x)|)$$

8) Expectation: the expectation (or expected value) of a random variable is a central concept that represents the average value that the variable should take on over many trials. It's a weighted average of all possible values the variable can assume, where the weights are given by their respective probabilities.

$$\mathbb{E}[X] = \sum_x x \cdot P(X = x)$$

Let's say if we throw dice and when 3 is the result you will get ₹100 and when anything other than 3 you will give ₹10, Expected how much will you earn.
Let's take solve with the formula:

$$\mathbb{E}[x]\frac{1}{6}\cdot(-10) + ... + \frac{1}{6}\cdot(100) + ... + \frac{1}{6}\cdot(-10)$$

$$\mathbb{E}[x] = \frac{-10\times 5}{6} + \frac{100}{6}$$

$$\mathbb{E}[x] = \frac{50}{6} = 8.33$$

We can earn ₹8.33.

9) Expectation Properties: The expectation operator (E[X]) possesses several valuable properties that simplify calculations and provide insights into the behavior of random variables.
Property 1:

$$\mathbb{E}[aX] = a\mathbb{E}[X]\ where\ a \rightarrow constant$$

Proof:

$$\mathbb{E}[X] = \sum x_i P(x_i)$$

$$\mathbb{E}[aX] = \sum a x_i P(a x_i)$$

$$= a\sum x_i P(x_i)$$

$$\mathbb{E}[aX] = a\mathbf{E}[X]$$

Property 2:

$$\mathbb{E}[X * Y] = \mathbb{E}[X] * \mathbb{E}[Y] \quad \text{(if X and Y are independent)}$$

$$\mathbb{E}[X+Y] = \sum_i\sum_j (x_i + y_j)P(x_i, y_i)$$

$$= \sum_i\sum_j x_i P(x_i,\ y_j) + \sum_i\sum_j y_j P(x_i,\ y_j)$$

$$P(A\cap B) = P(A)P(B)$$

$$= \sum_i x_i(P(x_i)P(y_1) + P(x_i)P(y_2) + ... + P(x_i)P(y_n)$$

$$= \sum_i x_i P(x_i) = \mathbb{E}[X]$$

Similarly for the $\mathbb{E}[Y]$ and after this we get the proof.

$$= \mathbb{E}[X] + \mathbb{E}[Y]$$

Property 3:

$$\mathbb{E}[ax+b] = a\mathbb{E}[x] + b$$

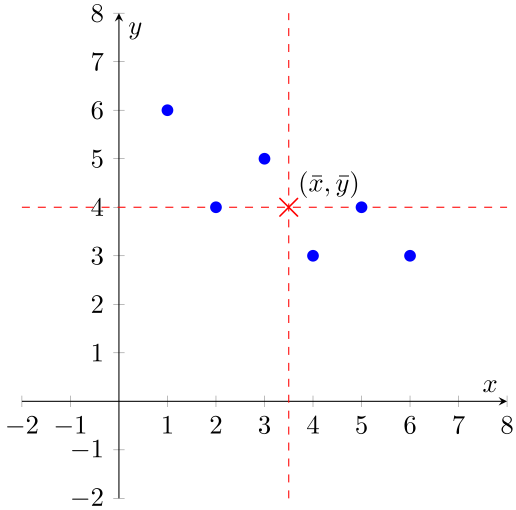


Fig. 4: Scatter plot with mean point $(\bar{x}, \bar{y})$

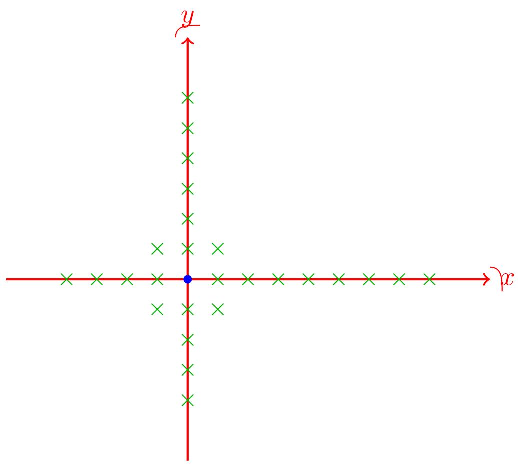


Fig. 5: Coordinate system with marked points including (±1, ±1)

10) Covariance: Variance of 2 Random variables(x,y) with respect to each other.
Formula:

$$Cov(x,\ y) = \frac{1}{n-1}\sum(x_i - \bar{x})(y_i - \bar{y})$$

Where $\bar{x}$ and $\bar{y}$ are the mean.

Question $\rightarrow$ Looking at the figure 5 what's the order:
a)$con(x,\ y) > var(x) > var(y)$
b) $var(y) > cov(x,\ y) > var(x)$
c) $var(x) > var(y) > cov(x,\ y)$
Sol^n: c) $var(x) > var(y) > cov(x,\ y)$

Question $\rightarrow$ Are following variables independent. Look at the fig 6
a) Yes
b) No

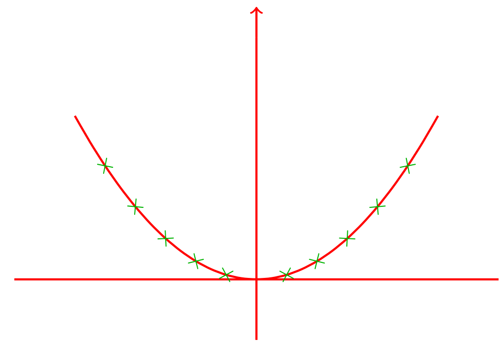


Fig. 6: Parabola with marked points

$$sol^n\ con(x,\ y) = 0$$

$$y = x^2$$

$\rightarrow$ If $x,\ y \Rightarrow$ independent, $cov = 0$
$\rightarrow$ If $cov(x,\ y) = 0 \Rightarrow$ then there is no surety that they are independent or not.
$\rightarrow$ Covariance can only detect linear relationships (limitations)

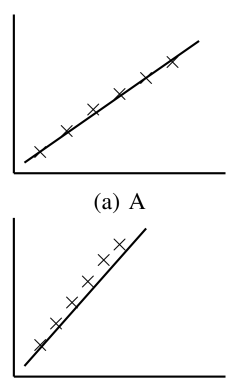
11) Correlations: Pearson's Correlation Coefficient: Pearson's correlation coefficient (often denoted by the letter "r") is a statistical measure that quantifies the strength and direction of the linear relationship between two continuous variables.[**?**]
Formula:

$$\rho_{x,\ y} = \frac{cov(x,\ y)}{\sigma_x \cdot \sigma_y}$$

$$\rho_{x,\ y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

Range of $cov(x,\ y) \rightarrow (-\infty,\ \infty)$
$corr \rightarrow [-1,\ 1]$

Question



(a) A

(b) B

Fig. 7: Two different linear patterns with marked points

i) Which has high co-variations
$a)\ A \quad b)\ B \quad c)\ A = B$
$Sol^n \rightarrow$ A

ii) Which has high correlation
$a)\ A \quad b)\ B \quad c)\ A = B$
$Sol^n \rightarrow$ C, cause when the graph is leaner the graph has correlation.

III. Binomial Distribution

The binomial distribution is a statistical concept that helps us understand the probability of a specific number of ”successes” occurring in a fixed number of independent trials, where each trial has the same probability of success.
For example, if we roll 5 dice and we need probability to get 2, and the probability of getting 2 is $\frac{1}{6}$ and rest of the probability is $\frac{5}{6}$ $\frac{5}{6}$ $\frac{5}{6}$ $\frac{5}{6}$ $\frac{5}{6}$
So, $P(exactly\ getting\ 1\ time\ 2$
Let take p which means it getting 2 on dice is true and $(1-p)$ means getting not getting 2 on dice roll.
So, $(p)^2(1-p)^5$
Second dice roll, we get $\frac{5}{6} \times \frac{1}{2} \times \frac{5}{6} \times ... \times \frac{5}{6}$
$= (1-p) \times p \times (1-p)^3$
Third dice roll, we get $\frac{5}{6} \times \frac{5}{6} \times \frac{1}{2} \times ... \times \frac{5}{6}$
$= (1-p)^2 \times p \times (1-p)^2$

$$P(2\ excalty\ 1\ time) = ^S C_1(p)^1(1-p)^4$$

Binomial generalisation formula:

$$n \rightarrow events$$

$$r \rightarrow true$$

$$P(x=r) = ^n C_r(p)_r(1-p)^{n-r}$$

→ Discrete events (True or False)
→ If $n = 1$ (Bernoulli’s events)

IV. Probability Distribution

A. Probability Mass function (PMF)

Of a discrete random variable X is a function that gives the probability that X will take on a particular value.

B. Probability Distribution Function (PDF)

is a mathematical function that describes the likelihood of a continuous random variable taking on a specific value within a 1 given range.[?] Look at the fig 8 to understand PDF.

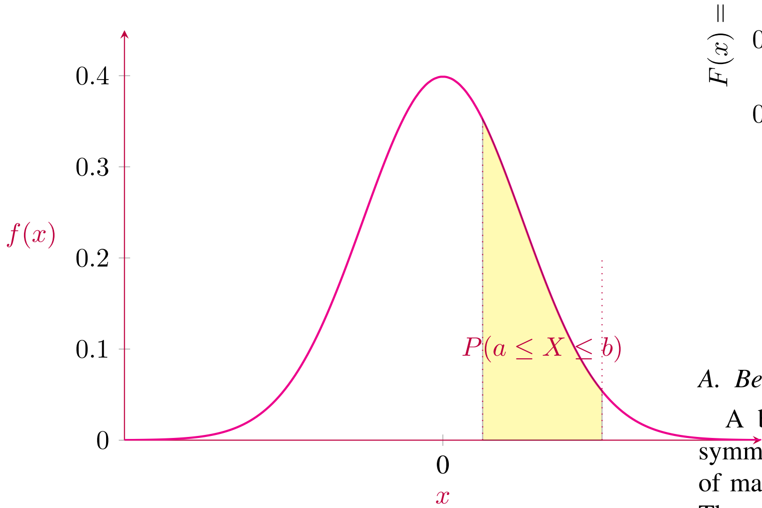


Fig. 8: Normal distribution with probability region between a and b

TABLE VI: Probability Distribution of Red Balls

Number of Red Balls (x)	Probability $P(X=x)$	Cumulative $P(X \leq x)$
0	0.05	0.05
1	0.10	0.15
2	0.25	0.40
3	0.30	0.70
4	0.20	0.90
5	0.10	1.00
Total	**1.00**	–

Find the area for the integration:

$$PDF = \frac{e^{\left(\frac{-1}{2}\right)\left(\frac{x-\nu}{\sigma}\right)^2}}{\sqrt{2r \times \sigma_2}}$$

C. Cumulative Distributive Function

The CDF of a random variable X gives the probability that X will take a value less than or equal to a specific value ’x’.

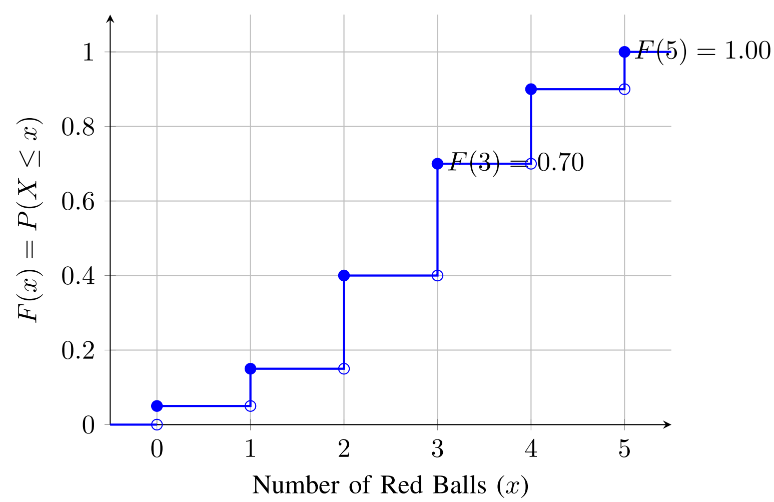
$$P(X=r) = ^n C_r(p)^r(1-p)^{n-r}$$

Let’s say total ball = m, red balls = l and blue balls = 5

$$P(x=1) = ^5 C_1(p)^1(1-p)^{5-1}$$

Probability of getting red ball as you can see the table VI and the graph for the values in table is given in fig 9

Fig. 9: Cumulative Distribution Function (CDF) for Red Balls Distribution



V. Z-score and Bell Curve

A. Bell Curve

A bell curve, also known as a normal distribution, is a symmetrical, bell-shaped curve that describes the distribution of many natural phenomena.
The standard normal distribution shows:

- Within $\pm 1\sigma$: 68.27% of data (light blue)
- Within $\pm 2\sigma$: 95.45% of data (lighter blue)

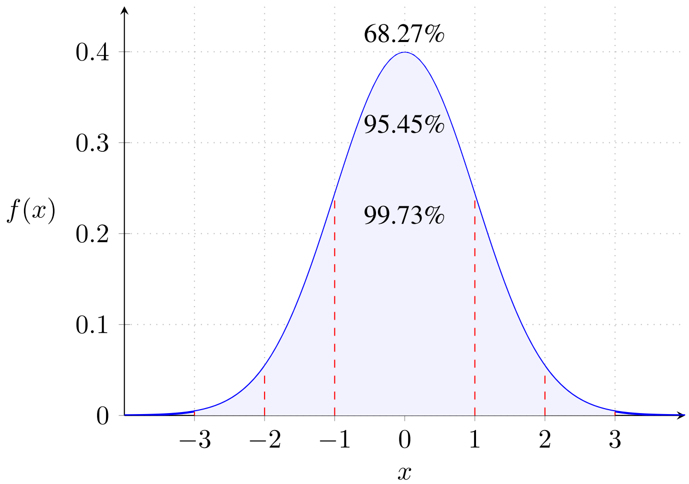


Fig. 10: Standard Normal Distribution showing 1σ, 2σ, and 3σ intervals

- Within $\pm 3\sigma$: 99.73% of data (lightest blue)

Question $\rightarrow \ \nu = 35, \ \sigma = 5$
$i) \ P(25 < x < 45)$
$Sol^n \ 25 \Rightarrow 35 - 2 \times 5$
$45 \Rightarrow 35 + 2 \times 5$
$\Rightarrow (\nu - 2\sigma, \ \nu + 2\sigma)$
$P(E) = 95\% = 0.95$

$ii) P(25 < x < 50)$
$25 \Rightarrow 35 - 2 \times 5$
$50 \Rightarrow 35 + 3 \times 5$
$Sol^n \Rightarrow \nu \pm 2\sigma \Rightarrow 95\%$
$\nu \pm 3\sigma \Rightarrow 99.7\%$

$$95\% + \frac{99.7\% - 95\%}{2}$$

$$95\% + \frac{4.7\%}{2} = 97.3\%$$

B. Z-Score

A z-score, also known as a standard score, is a statistical measure that tells you how many standard deviations a data point is away from the mean of a dataset.
Formula:

$$\mathbb{Z} = \frac{x - \nu}{\sigma}$$

Where $\nu \rightarrow mean$
$\sigma \rightarrow standard \ deviation$
$x \rightarrow value$

Question $\rightarrow$ On a dice roll, one will win the amount equal to the number on dice comes up. Find the variance using the expectation.

$$var = \frac{\sum |x : -\nu|^2}{n}$$

$$= \mathbb{E}[(x - \nu)^2]$$

$$\mathbb{E}[x_2 + \nu^2 - 2x\nu]$$

$$\mathbb{E}[x^2] + \nu^2 - 2\nu \times \nu$$

$$\mathbb{E}[x^2] - \nu^2$$

$$Var = \mathbb{E}[x^2] - \nu^2$$

$$Var = \mathbb{E}[x^2] - E[x]^2$$

$$\mathbb{E}[x] = \sum (xP(x))$$

$$= \left(\frac{1}{6} \times 1\right) + \left(\frac{1}{6} \times 2\right) + \left(\frac{1}{6} \times 3\right) ... \left(\frac{1}{6} \times 6\right)$$

$$\frac{21}{6} = 3.5$$

$$\mathbb{E}[x] = 3.5$$

Var = $\mathbb{E}[x^2] - (\mathbb{E}[x]^2)$

$$\sum \frac{1}{6} x^2$$

$$= \left((\frac{1}{6} \times 1^2) + (\frac{1}{6} \times 2^2) + ... + (\frac{1}{6} \times 6^2) \right)$$

$$\frac{91}{6} - (3.5)^2 = 2.916$$

$$Var = 2.916$$

Question $\rightarrow$ You asked your three friends whether it will rain or not and all three affirmed for rain. All three have the probability of $\frac{1}{3}$ of lying. What is the probability that it will actually rain?
$Sol^n \ E \rightarrow (rain)$
If they are lying the probability is $\frac{1}{3}$
If they are telling the truth, what's the probability?

$$1 - \frac{1}{3} = \frac{2}{3}$$

$$\frac{2}{3} \times \frac{2}{3} \times \frac{2}{3}$$

$$= \left(\frac{2}{3}\right)^3$$

$$= \frac{8}{27}$$

Question $\rightarrow$ A test has a true positive rate of 100% and a false-positive rate of 5%. There is a population with a $\frac{1}{1000}$ rate of having the condition the test identifies. Considering a positive test, what is the probability of having that condition?
$Sol^n \rightarrow$ Let consider getting positive test be $P(C|+)$
As it's given that the probability of condition is $P(C) = \frac{1}{1000}$ and we are also given that the test getting false-positive is $P(C|+) = \frac{50}{1000}$
So the total positive test is:

$$P(+) = \frac{50}{1000} + \frac{1}{1000} = \frac{51}{1000}$$

$$P(Cov|+) = \frac{P(Cov \cap +)}{P(+)} = \frac{\frac{1}{1000}}{\frac{51}{1000}} = \frac{1}{51} = 0.019$$

Question $\rightarrow$ Supposedly a friend of yours has children and at least one of them is a boy. What is probability that the other is also a boy?
Sol^2 $\rightarrow$ Possibility of having 2 children. Lets say possibility of have at least one boy and possibility of have both of them boy. Sample set of the children is $F = \{BB,\ BG,\ GB,\ GG\}$. Possibility of aleast one boy is 3 and possibility of both of them is boy is 1.

$$So,\ P(BB) = \frac{1}{3}$$

VI. Statistics

Statistics is the science of collecting, organizing, analyzing, interpreting, and presenting data.

A. *Population and sample*

There is election going on in Uttar Pradesh, and we take 3 cities as sample for survey to find which party is wining the election. Now let's say some place x_1 is wining by $\frac{1}{2}$, and somewhere x_2 is wining by $\frac{1}{3}$, and lets assume we take $x_3 \rightarrow \frac{2}{3}$, $x_4 \rightarrow \frac{1}{4}$ and so on ....

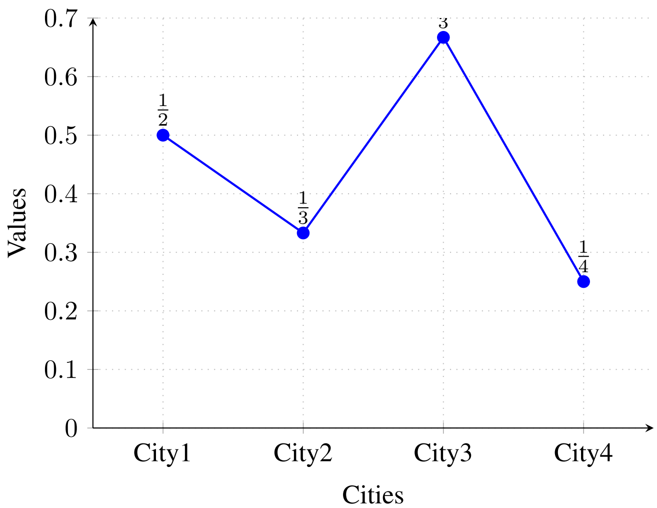


Fig. 11: Values for Different Cities

See fig 11 and fig 12 to understand the different value of x, and you increase the value of x from $100 \rightarrow 1000 \rightarrow 10000 \rightarrow 100000$.

B. *Standard deviation of sample mean*

$$Var(\bar{x}) = Var\left(\frac{x_1 + x_2 + x_3 + ... + x_n}{n}\right)$$

$$= \frac{1}{n^2} \cdot var(x_1 + x_2 + x_3 + ... + x_n)$$

$$= \frac{1}{n^2} \cdot (var(x_1) + var(x_2) + var(x_3) + ... + var(x_n))$$

$$= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \sigma^2 + ... + \sigma^2)$$

$$= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

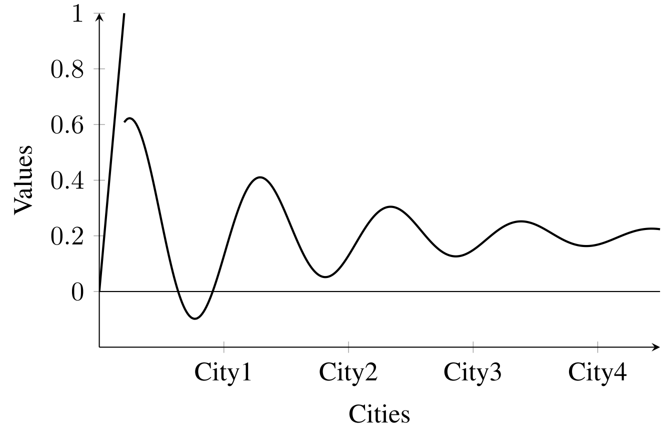


Fig. 12: Damped Oscillation Pattern Across Cities

$$var(\bar{x}) = \frac{\sigma^2}{n}$$

$$Standard\ deviation\ \rightarrow \sqrt{var(\bar{x})}$$

$$Standard\ deviation\ of\ sample\ means = \sqrt{\frac{\sigma^2}{n}}$$

$$Standard\ error = \frac{\sigma}{\sqrt{n}}$$

$\frac{\sigma}{\sqrt{n}} \Rightarrow$ how much means of sample means deviate from the population means.

C. *Sample variance and Central Limit Theorem (CLT)*

1) Sample variance: The sample variance is a measure of how spread out a set of data is. It is calculated by taking the average of the squared differences from the mean.[**?**]
Formula:

$$Sample\ variance = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

Where x_i is sample mean, $\bar{x}$ is means of sample means and n is no. of sample

2) Central Limit Theorem(CLT): is a fundamental concept in statistics. It states that, under certain conditions, the distribution of the sample means of a sufficiently large number of independent random variables will be approximately normal, regardless of the underlying distribution of the individual variables.[**?**]
Formula:

$$\nu = \bar{x} + \mathbb{Z}(standard\ error)$$

Where ν is population mean, $\bar{x}$ is mean

VII. QQ Plot

A QQ plot (Quantile-Quantile plot) is a graphical tool used to assess the similarity between the distributions of two datasets. It helps determine if a dataset follows a particular theoretical distribution (like the normal distribution) by plotting the quantiles of the two datasets against each other. If the datasets have similar distributions, the points on the QQ plot will fall approximately along a straight line.

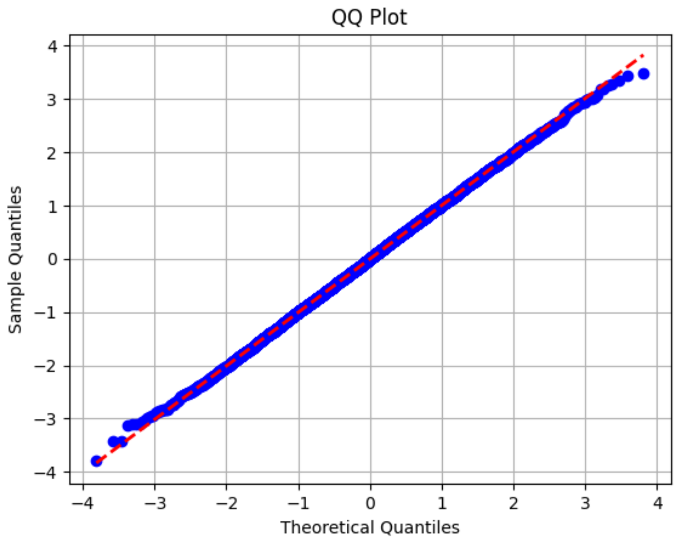


Fig. 13: Q-Q Plot for Normality Assessment

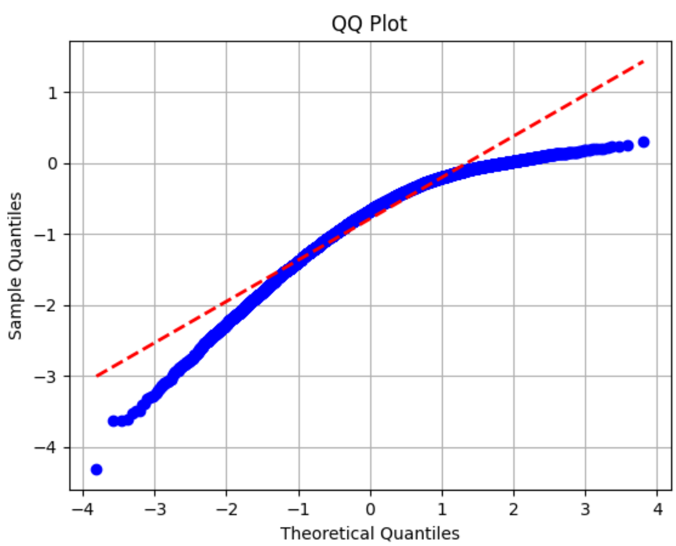


Fig. 14: Q-Q Plot for Skew data

Take a look at the fig 13 and 14 for example for QQ Plot with normal and skew data.

VIII. Hypothesis Testing

Suppose you work at a healthcare company and the company is looking for a new drug to lower sugar levels. A drug manufacturer comes to your company with a drug claims has 90% success rate of lowering sugar level in a person. Your job is to validate this claim.
$P(success) = 0.9$
$P(failure) = 0.1$
Lets collect the data, assuming we have 10 people in which 8 people's blood sugar level went down, again assuming we have 100 people 83 people's blood sugar level went down, again assuming we have 1000 people in which 850 people's blood sugar level went down. We can see that, when we assume 10, 100 and 1000 people we have seen the success rate of the people who's blood sugar level went down is 80%, 83% and 85%.
For such scenario we use Binomial distribution cause we have $n = 1000$, $P(success) = 0.9$ and $P(failure) = 0.1$

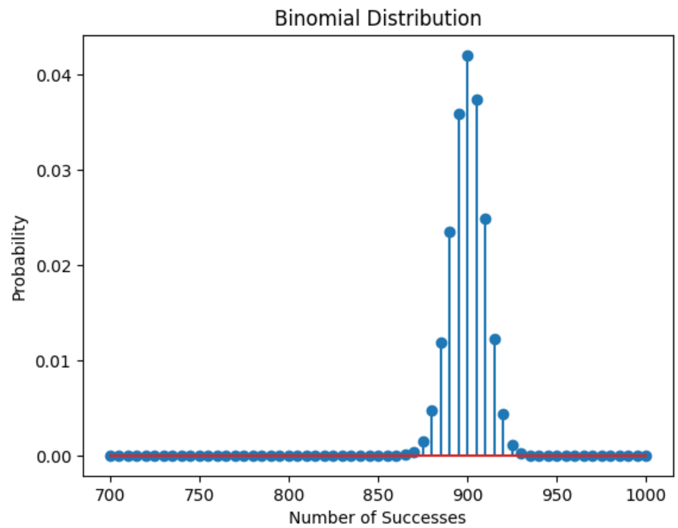


Fig. 15: Binomial Distribution

Check graph 15