Introduction to Probability and Statistics

Biswajit Roy

Dept. of Physics, Chakdaha College

What is Probability

To answer this question, it is better to start with an examples

Two person A and B toss a coin and look at their coins but not each other's.

(a) What is the probability that both coins show heads?

Answer.

The probability is either 0 and ½ depending on what A and B see their coins.

Suppose A see tails then probability is 0 because A don't see B's coin.

Suppose B see heads the probability is ½ because B don't see A's coin.

And vice versa

What is Probability

(b) Now suppose neither A and B looks at either coin, but a 3rd person C looks at both coins then what is the probability that both coins show heads?

Answer: Since there are 4 possibilities, namely hh, tt, th, ht

So, there is 1 chance of having both heads that is hh. That is one favorable outcome out of all possible outcomes. Hence, the probability is ¼.

(c) Suppose the 3rd person C looks at both coins and gives them information that at least one is heads then what is the probability that both coins show heads?

Answer: Now there is 3 possibilities hh, th, ht (tt is ruled out)

So, the probability is 1/3

Definition of Probability

So, we see that the answer to a probability problem depends on the state of knowledge (or ignorance) of the person giving the answer.

So, to define the probability

If there are several equally likely, mutually exclusive, and collectively exhaustive outcomes of an experiment, the probability of an event E is

$$p(E) = \frac{number\ of\ outcomes\ favourable\ to\ E}{total\ number\ of\ outcomes}$$

Outcomes refer to the possible results of a random experiment or event.

Each outcome represents a single, specific result that can happen when the experiment is conducted

Definition of Probability

Equally likely outcomes: Outcomes are said to be equally likely if each outcome has the same probability of occurring

For example: In rolling a fair six-sided die, each face (1, 2, 3, 4, 5, 6) has an equal chance of appearing, i.e., 1/6.equal chance of appearing)

Mutually exclusive outcomes: Outcomes are mutually exclusive if the occurrence of one outcome prevents the occurrence of the others.

For example: In rolling a die, the outcomes 1, 2, 3, 4, 5, and 6 are mutually exclusive because you cannot roll more than one number at the same time.

Collectively exhaustive: Outcomes are collectively exhaustive if they include all possible outcomes of an experiment.

For example: In rolling a die, the outcomes 1, 2, 3, 4, 5, and 6 are collectively exhaustive because they cover all the possible results of a roll

Sample Space

A set of all possible mutually exclusive outcomes is called a sample space and each individual outcome is called a point of the sample space.

There are two types of sample space

Uniform sample space: Sample space of equal probabilities (equally likely outcomes)

Non-uniform sample space: Sample space of unequal probabilities (not equally likely outcomes)

The tossing of the two coins can have the following sample spaces

1. 2 heads, 2 tails, 1 head & 1 tail, 1 tail & 1 head (Uniform)
2. 2 heads, 1 head, no heads (Non-uniform)
3. No heads, at least 1 head (Non-uniform)

Sample Space

The probabilities in the following space spaces are

2 heads, 2 tails, 1 head & 1 tail, 1 tail & 1 head (Uniform)

$\boxed{1/4}$ $\boxed{1/4}$ $\boxed{1/4}$ $\boxed{1/4}$

2 heads, 1 head, no heads (Non-uniform)

$\boxed{1/4}$ $\boxed{1/2}$ $\boxed{1/4}$

No heads, at least 1 head (Non-uniform)

$\boxed{1/4}$ $\boxed{3/4}$

More General Definition of Probability

Given any sample space (uniform or not) and the probabilities associated with the points, we find the probability of an event by adding the probabilities associates with all the sample points favorable to the event.

Example:

In the tossing of two coins problem the probability of at least one head is the probability of one head plus the probability of two heads = 1/2 +1/4 =3/4

So, The probability of an event E can be defined as the sum of the ratios of the number of favorable outcomes for E to the total number of possible outcomes, considering all sample points favorable to E.

Probability Theorem

Probability theorem shorten the task of calculating probabilities without going into details of sample space.

Suppose there are

5 black balls and 10 white balls in a box. Now drawing one ball at random and then without replacing the 1^{st} ball drawing another ball. What is the probability that the probability that the 1^{st} ball is white and the 2^{nd} ball is black?

Ans:

The probability of drawing a white ball the 1^{st} time is $\frac{10}{15}$.

The probability of drawing a black ball the 2^{nd} time is $\frac{5}{14}$(since there are 14 balls left)

So, the probability of drawing 1^{st} a white ball and 2^{nd} a black ball is $= \frac{10}{15} . \frac{5}{14}$

Probability Theorem

Since there are total 15 balls. Let's assume the balls are numbered from 1 to 15. For the 1st drawing there are 15 choices and 14 choices for the 2nd drawing to select either black or white balls from the box. Thus, there are 15.14 points in the sample space. And there are 10.5 sample points favorable to the event. So, the probability is $\frac{10.5}{15.14}$

In general

$$P(AB) = P(A).P_A(B)$$

$P(AB)$ = The probability that both the events A and B will happen.

$P(A)$ = The probability that both the events A will happen

$P_A(B)$= The probability that B will happen if know that A has happened.

Here the two events A and B are dependent.

Probability Theorem

If the two events A and B are independent of each other

Suppose in the previous problem: 5 black balls and 10 white balls in a box. Now draw one ball from the box, write down its color, and put it back into the box (with replacement). Then, shuffle the box again and draw a second ball. What is the probability that the probability that the 1st ball is white and the 2nd ball is black?

Ans:

Probability of the first ball being white: $P(A) = \frac{10}{15}$

Probability of the first ball being black: $P(B) = \frac{5}{15}$ (Since the 1st ball put back to the box(with replacement), the second draw happens independently of the first draw.)

So the combined probability is $= \frac{10}{15} \cdot \frac{5}{15}$

So, in general

$$P(AB) = P(A)P(B)$$

Methods of Counting (Permutations and Combinations)

Suppose there are 5 people A, B, C, D, E and 5 chairs (S1, S2, S3, S4, S5).

How many ways the 5 people can seat on 5 chairs (permute) ?

Ans:

1st person can sit on anyone of the 5 chairs.

2nd person can sit on anyone of the 4 chairs (since there are 4 chairs left)

3rd person can sit on anyone of the 3 chairs (since there are 3 chairs left)

4th person can sit on anyone of the 2 chairs (since there are 2 chairs left)

5th person don’t have any choice will sit on 1 chair (since 1 chair is left now)

So, the number of ways the 5 people can seat on 5 chairs is = 5 x 4 x 3 x 2 x 1=5!=120

In general, If there are n number of people and n locations then the number of ways the n number people are allocated on n locations is

$$P(n, n) = n(n-1)(n-2)(n-3) \ldots \ldots = n!$$

Methods of Counting (Permutations and Combinations)

Now, there are 5 people A, B, C, D, E but 3 chairs (S1, S2, S3) then how many ways the 3 people out of 5 people can seat on 3 chairs (permute) ?

Ans:

The 1st chair (S1) can select anyone from the 5 people

The 2nd chair (S2) can select anyone from the 4 people

The 3rd chair (S3) can select anyone from the 3 people

So, the many ways the 5 people can seat on 3 chairs is = 5 x 4 x 3 =60

It can be written as $\mathrm{P}(5,3) = \frac{5\times4\times3\times2\times1}{2\times1} = \frac{5!}{(5-3)!}$

In general, If there are n number of people and r(<n) locations then the number of ways the n number people are allocated on r locations is

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

Methods of Counting (Permutations and Combinations)

Next questions is how many ways 3 people can be selected from the 5 people to seat on 3 chairs (Combination) ?(Ignore ordering of the people that is ABC, CAB, BAC,…. Are same that is distinguishable)

Ans:

The List of Combinations C(n,r) are

1. A, B, C
2. A, B, D.
3. A, B, E
4. A, C, D
5. A, C, E
6. A, D, E
7. B, C, D
8. B, C, E
9. B, D, E
10. C, D, E

So, there are 10 combinations that 3 out of 5 people can seat on 3 chairs.

Methods of Counting (Permutations and Combinations)

Next questions is if the order is considered that is they are indistinguishable then how many ways 3 people can be selected from the 5 people to seat on 3 chairs (Combination) ? Ans:

Answer:

One particular combination can have =3 x 2=3! 6 choices suppose the 1st combination

(A,B,C),(A,C,B), (B,A,C), (B,C,A), (C,A,B), (C,B,A)

So for 10 combinations there are total = 10 x 6 = 60 ways that 3 out of 5 people can seat on 3 chairs. Which is equal to $P(n,r) = C(5,3) \times 3!$

In general

$$P(n,r) = C(n,r) \times r!$$

$$C(n,r) = \frac{P(n,r)}{r!}$$

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

Methods of Counting (Permutations and Combinations)

Example1: A club consists of 50 members. In how many ways can a president, vice president, secretary and treasurer be chosen. In how many ways can a committee of 4 members be chosen?

Answer:

The number of ways of selecting the officers is $P(50,4) = \frac{50!}{(50-4)!} = \frac{50!}{46!}$

The number of ways of selecting committees of 4 people(4 committee members) is $C(50,4) = \frac{50!}{46!4!}$

Example2: Find the coefficient of x^8 in the binomial expansion of $(1 + x)^{15}$?

Answer:

The coefficient of x^8 is $C(15,8) = \frac{15!}{8!7!}$

Methods of Counting (Permutations and Combinations)

Example3(Basic statistical mechanics problem): Suppose we have 15 balls and 6 boxes and the numbers of balls to put into the various boxes are:

Number of balls:	3	1	4	2	3	2
In box number	1	2	3	4	5	6

Calculate the total number of ways of putting the required numbers of balls into the boxes ?

Answer:

The number of ways to select 3 balls to go to the 1st box is $C(15,3)$, similarly

The number of ways to select 1 ball to go to the 2nd box is $C(12,1)$ (since 12 balls left), similarly

The number of ways to select 4 ball to go to the 3rd box is $C(11,4)$

The number of ways to select 2 ball to go to the 4th box is $C(7,2)$

The number of ways to select 3 ball to go to the 5th box is $C(5,3)$

The number of ways to select 2 ball to go to the 6th box is $C(2,2)$

Total number of ways of putting all the balls into the boxes is

$$C(15,3).C(12,1).C(11,4).C(7,2).C(5,3).C(2,2)$$

Problems

Problem-1: Let two dice be thrown; the 1st die can show any number from 1 to 6 and similarly for the 2nd die. Set up the sample space?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Problems

Problem-2: What is the probability that the sum of the numbers on the dice will be 5 ?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Answer: The probability is 4/36=1/9

Problems

Problem-3: What is the probability that the sum of the numbers on the dice is divisible by 5 ?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Answer: The probability is 7/36

Problems

Problem-4: Set up sample space in which the points corresponds to the possible sums of the two numbers on the dice and find the probabilities associated with the points ?

Sum	Outcomes	Count	Probability
2	(1,1)	1	1/36
3	(1,2),(2,1)	2	2/36
4	(1,3), (2,2), (3,1)	3	3/36
5	(1,4), (2,3), (3,2), (4,1)	4	4/36
6	(1,5), (2,4), (3,3), (4,2), (5,1)	5	5/36
7	(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)	6	6/36
8	(2,6), (3,5), (4,4), (5,3), (6,2)	5	5/36
9	(3,6), (4,5), (5,4), (6,3)	4	4/36
10	(4,6), (5,5), (6,4)	3	3/36
11	(5,6), (6,5)	2	2/36
12	(6,6)	1	1/36

Random Variables and Probability Functions

Random Variables:

In the problem of tossing two dice. Let's assume the x is the sum of the numbers on the two dice. Then for each point of the sample space x has a value. Such a variable, x, which has a definite value for each sample point, is called a random variable. So, we may say that a random variable x is a function defined on a sample space.

Probability Functions:

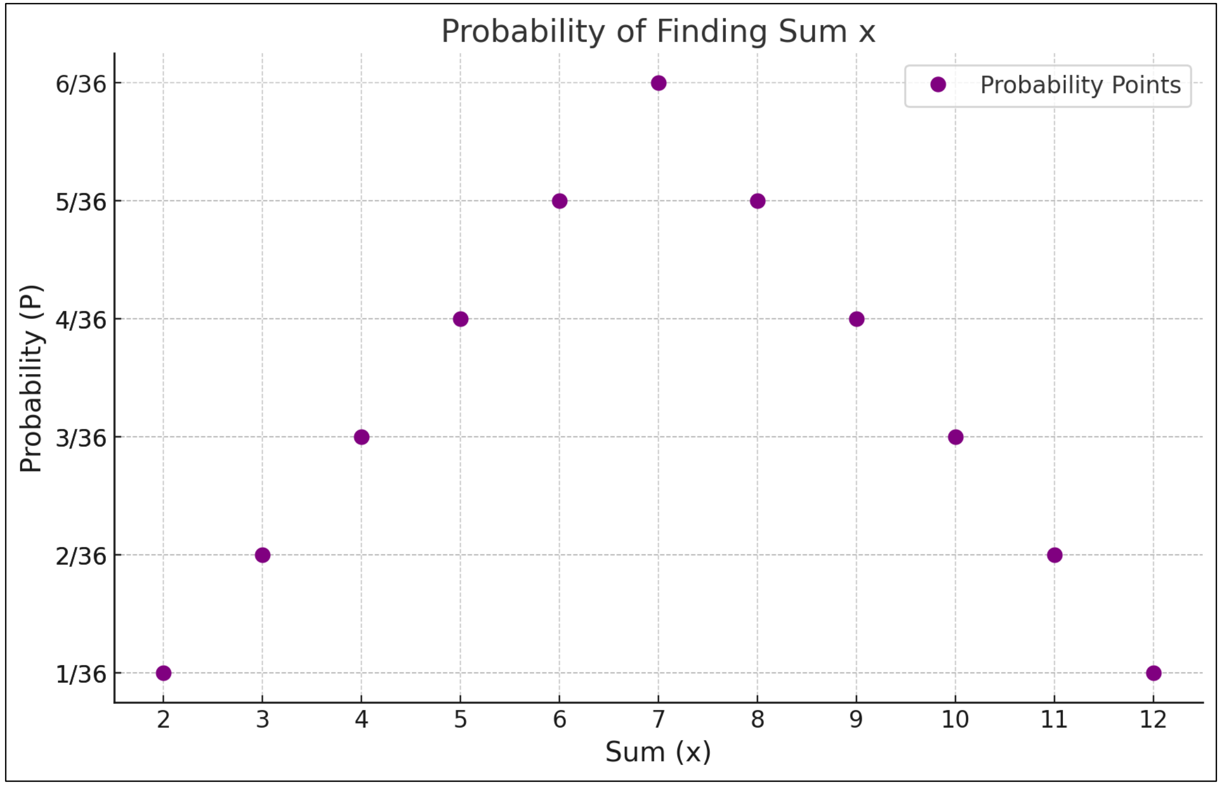
In the two dice problems, the x has values 2, 3, 4, 5, 6 ,7, 8, 9, 10, 11, 12 and has probabilities 1/36, 2/36, 3/36, 436, 5/36, 6/36,5/36,4/36,3/36,2/36,1/36.

The probabilities of occurrence x is called probability functions. The above numbers are the values of the probability functions.

Random Variables and Probability Functions

We can say that x is a random variable if it takes various values of x_i with probabilities $p_i = p(x_i)$. The value of x is unknown in advance that is why x is called a random variable.

The plot of the probability functions of finding sum x of the two dice problem



Average of a Random Variable

The average value of x is 7

$$Average\ value\ of\ x = \frac{sum\ of\ the\ product\ of\ each\ value\ of\ x\ and\ the\ count\ for\ each\ x}{total\ count\ for\ x\ (total\ outcomes\ of\ x)}$$

$$= \frac{(2\times 1)+(3\times 2)+(4\times 3)+(5\times 4)+(6\times 5)+(7\times 6)+(8\times 5)+(9\times 4)+(10\times 3)+(11\times 2)+(12\times 1)}{1+2+3+4+5+6+5+4+3+2+1}$$

$$= \frac{(2\times 1)+(3\times 2)+(4\times 3)+(5\times 4)+(6\times 5)+(7\times 6)+(8\times 5)+(9\times 4)+(10\times 3)+(11\times 2)+(12\times 1)}{36}$$

$$= \frac{252}{36} = 7$$

The average value of x can be written as

$$= 2\left(\frac{1}{36}\right)+3\left(\frac{2}{36}\right)+4\left(\frac{3}{36}\right)+5\left(\frac{4}{36}\right)+6\left(\frac{5}{36}\right)+7\left(\frac{6}{36}\right)+8\left(\frac{5}{36}\right)+9\left(\frac{4}{36}\right)+10\left(\frac{3}{36}\right)+11\left(\frac{2}{36}\right)+12\left(\frac{1}{36}\right)$$

$$= \sum_{i=1}^{11} x_i\, p(x_i)$$

Where $p(x_i)$ is the probability of each x

Variance of a Random Variable

The variance of a probability distribution measures how spread out the values are from the average.

The average value of x is

$$\mu = < x > = \sum_{i=1}^{N} x_i \, p(x_i)$$

Similarly, the average value of square of x is

$$< x^2 > = \sum_{i=1}^{N} x_i^2 \, p(x_i)$$

The variance of is

$$Var(x) = < x^2 > - < x >^2$$

The standard deviation of x is

$$\sigma_x = \sqrt{Var(x)}$$

Variance of a Random Variable

The variance and standard deviation in the two dice problem is

$$< x^2 > = 2^2\left(\frac{1}{36}\right) + 3^2\left(\frac{2}{36}\right) + 4^2\left(\frac{3}{36}\right) + 5^2\left(\frac{4}{36}\right) + 6^2\left(\frac{5}{36}\right) + 7^2\left(\frac{6}{36}\right) + 8^2\left(\frac{5}{36}\right) + 9^2\left(\frac{4}{36}\right) + 10^2\left(\frac{3}{36}\right) + 11^2\left(\frac{2}{36}\right) + 12^2\left(\frac{1}{36}\right) = 54.83$$

$$Var(x) = < x^2 > - < x >^2 = 54.83 - 49 = 5.83$$

$$\sigma_x = \sqrt{Var(x)} = 2.415$$

The variance can also be written as

$$Var(x) = \sum_{i=1}^{N} (x_i - \mu)^2 \, p(x_i) = 5.83$$

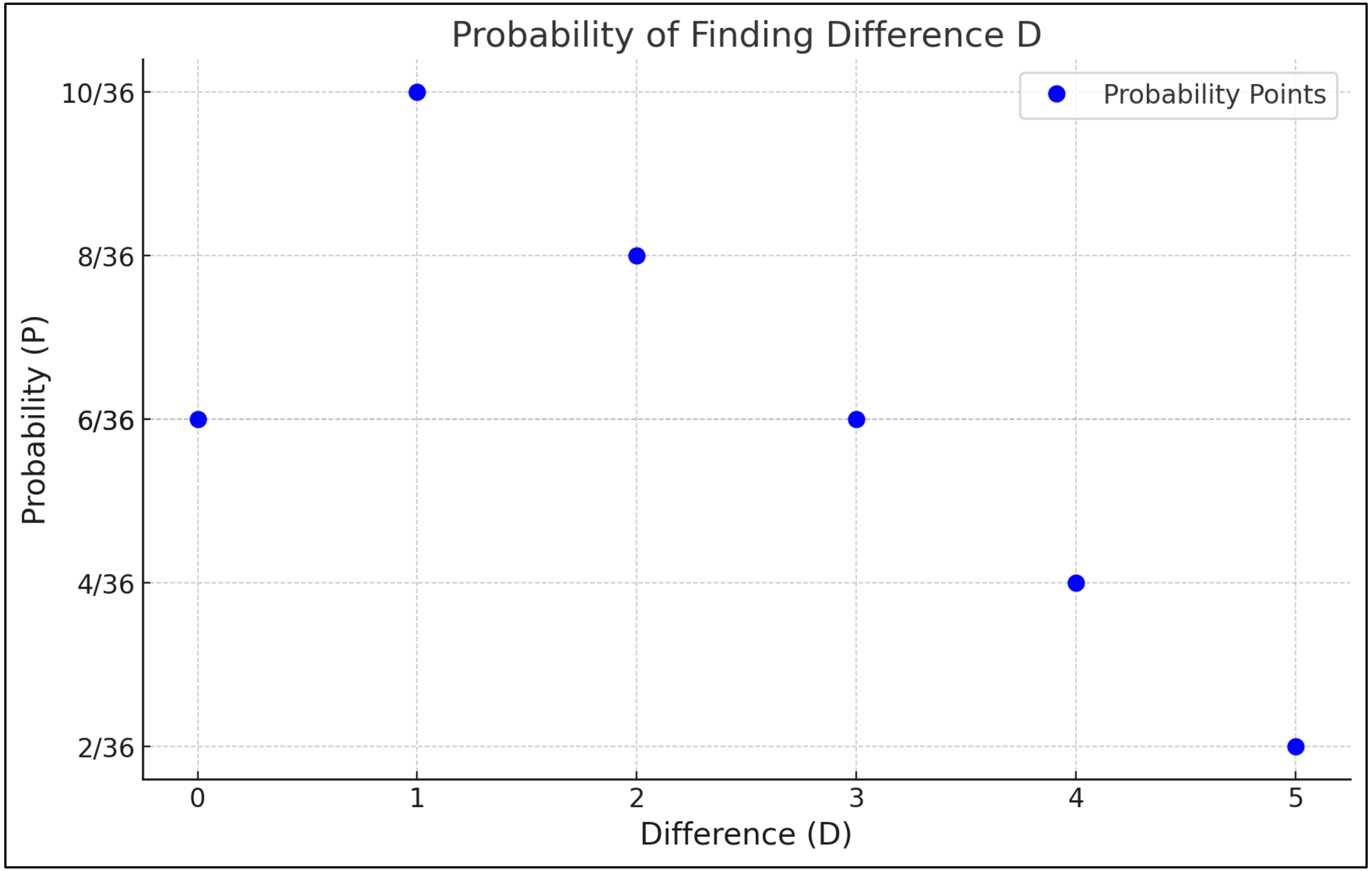
Variance of a Random Variable

Problem-5: Set up sample space of the two dice tossing problem in which the points corresponds to the possible absolute difference of the two numbers on the dice and find the probabilities associated with the points ?

Difference	Outcomes	Count	Probability
0	(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)	6	6/36
1	(2,1), (3,2), (4,3), (5,4), (6,5), (1,2), (2,3), (3,4), (4,5), (5,6)	10	10/36
2	(3,1), (4,2), (5,3), (6,4), (1,3), (2,4), (3,5), (4,6)	8	8/36
3	(4,1), (5,2), (6,3), (1,4), (2,5), (3,6)	6	6/36
4	(5,1), (6,2), (1,5), (2,6)	4	4/36
5	(6,1), (1,6)	2	2/36

Variance of a Random Variable

The plot of the probability function of finding absolute difference in the two dice problem



Variance of a Random Variable

Problem-6: Calculate the average, variance and standard deviation of absolute difference of the two numbers on the dice?

Answer:

$$< x > = 0\left(\frac{6}{36}\right) + 1\left(\frac{10}{36}\right) + 2\left(\frac{8}{36}\right) + 3\left(\frac{6}{36}\right) + 4\left(\frac{4}{36}\right) + 5\left(\frac{2}{36}\right) = 1.04$$

$$< x^2 > = 0^2\left(\frac{6}{36}\right) + 1^2\left(\frac{10}{36}\right) + 2^2\left(\frac{8}{36}\right) + 3^2\left(\frac{6}{36}\right) + 4^2\left(\frac{4}{36}\right) + 5^2\left(\frac{2}{36}\right) = 5.83$$

$$Var(x) =< x^2 > - < x >^2 = 5.83 - 1.04^2 = 2.052$$

$$\sigma_x = \sqrt{Var(x)} = \sqrt{2.052} = 1.432$$

Variance of a Random Variable

Let's interpret the results of the two problems

Variance of possible sum:

1. The sum of two dice has a larger variance, meaning the values are more spread out from the mean
2. This is expected because the possible sums range from 2 to 12, covering a broader range.
3. The most likely sum is 7, but the probability distribution is symmetric around it, leading to a wider spread.

Variance of Possible difference:

1. The difference has a much smaller variance, meaning the values are more concentrated around the mean.
2. The possible differences range from 0 to 5, which is a much smaller range than the sum.
3. Since rolling two dice often results in similar numbers (e.g., 3 and 4, 5 and 6), smaller differences (like 0 or 1) are more common, leading to less spread.

Cumulative Distribution Functions

In the two dice problem, the probability that the value of x(sum) less than or equal to a particular value x_i can be obtained by adding all the probabilities of values of x less than or equal to x_i. The resulting probability function is called cumulative probability function or cumulative distribution.

$$P(x_i) = probability\ that\ x \leq x_i = \sum_{x_j \leq x_i} p(x_j)$$

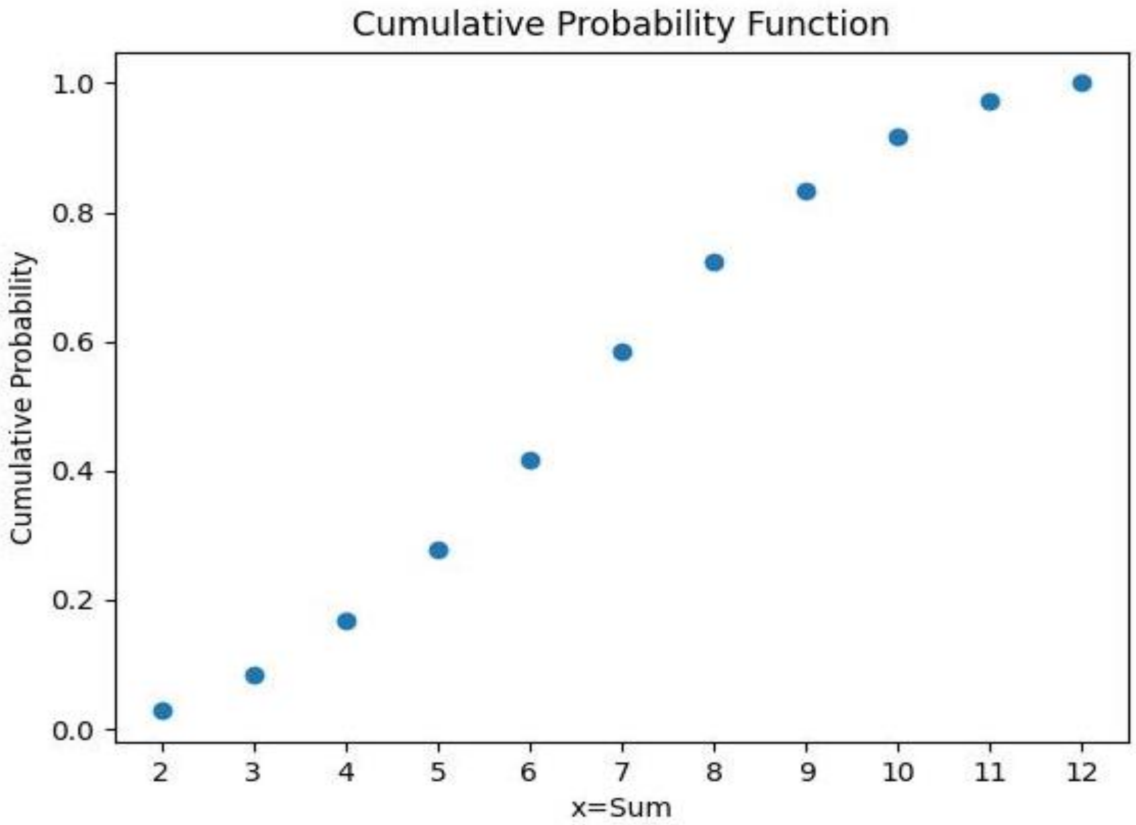
$P(x_j)$ is the cumulative probability function. That is the sum of the probability up-to a particular value.

The probability that x is say, less than or equal to 4 is the sum of the probabilities that x is 2 or 3 or 4 that is $\frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36} = \frac{1}{6}$

Cumulative Distribution Functions(CDF)

The values of cumulative probability function and its plot up-to all values of x is show below

$P(x_i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1
x_i	2	3	4	5	6	7	8	9	10	11	12



The final value of the of the CDF is always 1, because the sum of all the probabilities is 1 that means one of the possible outcomes must always happen.

Average and Variance for Continuous Distribution

The average value of x is

$$\mu = \int_{-\infty}^{\infty} xp(x)dx$$

The variance of x is

$$Var(x) = \int_{-\infty}^{\infty} (x - \mu)^2 p(x)dx$$

The standard deviation of x is

$$\sigma_x = \sqrt{Var(x)}$$

The total probability or Cumulative probability function for all values of x 1

$$P(x) = \int_{-\infty}^{\infty} p(x)dx = 1$$

Probability Density

The probability function or probability distribution p(x) is often called probability density.

Suppose the mean of x is the center of mass and p(x) is the density (mass per unit length) of a thin rod.

Then the center of mass of the rod is given by

$$\bar{x} = \frac{\int_{-\infty}^{\infty} xp(x)dx}{\int_{-\infty}^{\infty} p(x)dx}$$

Since total probability is 1 that is $\int_{-\infty}^{\infty} p(x)dx = 1$

So, $\bar{x} = \int_{-\infty}^{\infty} xp(x)dx$, this is the definition of the average value of x.

Joint Distribution

Suppose we have two random variables x and y and $f(x,y)$is their joint probability density function.

Then the probability that the point (x,y) is in a given region of the (x,y) plane, is the integral of $f(x,y)$ over that area.

The average values of x and y, the variances and standard deviations of x and y, and the covariance of (x,y) are given by

$$\bar{x} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy \,, \bar{y} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy$$

$$Var(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})f(x,y)dxdy = \sigma_x^2$$

$$Var(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\bar{y})f(x,y)dxdy = \sigma_y^2$$

$$Cov(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})(y-\bar{y})f(x,y)dxdy$$

Binomial Distribution

The Binomial Distribution applies to situations when there are n number of repeated independent trials, where each trial having two possible outcomes for example heads or tails, success or failure, good or defective etc.

Suppose a coin is tossed 3 times. Then what is the probability of getting exactly 2 heads?

Ans:

The tosses are independent and the possible outcomes of with 2 heads are HHT, HTH, THH

The probability of head is $\frac{1}{2}$ and tail is $\left(1-\frac{1}{2}\right)=\frac{1}{2}$

So, the probability of each outcome is $=\frac{1}{2}.\frac{1}{2}.\frac{1}{2}=\left(\frac{1}{2}\right)^3=0.125$

But there are 3 ways of getting exactly 2 heads, so the probability of getting exactly 2 heads is = 3 x 0.125 = 0.375

The probability can be written as $C(3,2)\left(\frac{1}{2}\right)^3$

Binomial Distribution

Let's do a similar problem where a die is tossed 5 times. Then what is the probability of getting 3 aces (aces refers to the side with only one dot, while "non-aces" refers to any of the other sides with more than one dot)

Ans:

If A is ace and N is not ace.

The tosses are independent, and the probability of ace (A) is $\frac{1}{6}$ and not-ace(N) is $\frac{5}{6}$.

The probability of a particular outcome A N N A A is = $\frac{1}{6}.\frac{5}{6}.\frac{5}{6}.\frac{1}{6}.\frac{1}{6} = \left(\frac{1}{6}\right)^3.\left(\frac{5}{6}\right)^2$

Since there are C(5,3) ways of getting 3 aces. So, the probability of getting 3 aces is

$C(5,3)\left(\frac{1}{6}\right)^3.\left(\frac{5}{6}\right)^2$

Binomial Distribution

Suppose there are n number of independent trials then the binomial distribution function can be written as

$$f(x) = C(n,x)p^x(1-p)^{n-x}$$

Where x is a random variable(heads or success) and p is the probability of heads or success etc.

The binomial cumulative distribution function up-to x is written as

$$F(x) = \sum_{u=0}^{x} C(n,u)p^u(1-p)^{n-u}$$

Mean of binomial distribution function

$$E = \sum_{x=0}^{n} xC(n,x)p^x(1-p)^{n-x} = \sum_{x=0}^{n} \frac{x\, n!}{x!\,(n-x)!} p^x(1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!\,(n-x)!} p^x(1-p)^{n-x} = np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,((n-1)-(x-1))!} p^{x-1}(1-p)^{(n-1)-(x-1)}$$

$$= np \sum_{x=1}^{n} C(n-1,x-1)p^{x-1}(1-p)^{n-x} = np.1$$

The summation is the sum of the binomial distribution function for (n-1) trial which is 1.

Binomial Distribution

The variance of the binomial distribution function is $Var(x) = E(x^2) - \left(E(x)\right)^2$

$$E(x^2) = \sum_{x=0}^{n} x^2 C(n,x) p^x (1-p)^{n-x} = \sum_{x=0}^{n} \frac{(x(x-1)+x)\, n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \frac{x(x-1)\, n!}{x!\,(n-x)!} p^x (1-p)^{n-x} + \sum_{x=0}^{n} \frac{x\, n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!\,((n-2)-(x-2))!} p^{x-2} (1-p)^{(n-2)-(x-2)}$$

$$+np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,((n-1)-(x-1))!} p^{x-1} (1-p)^{(n-1)-(x-1)}$$

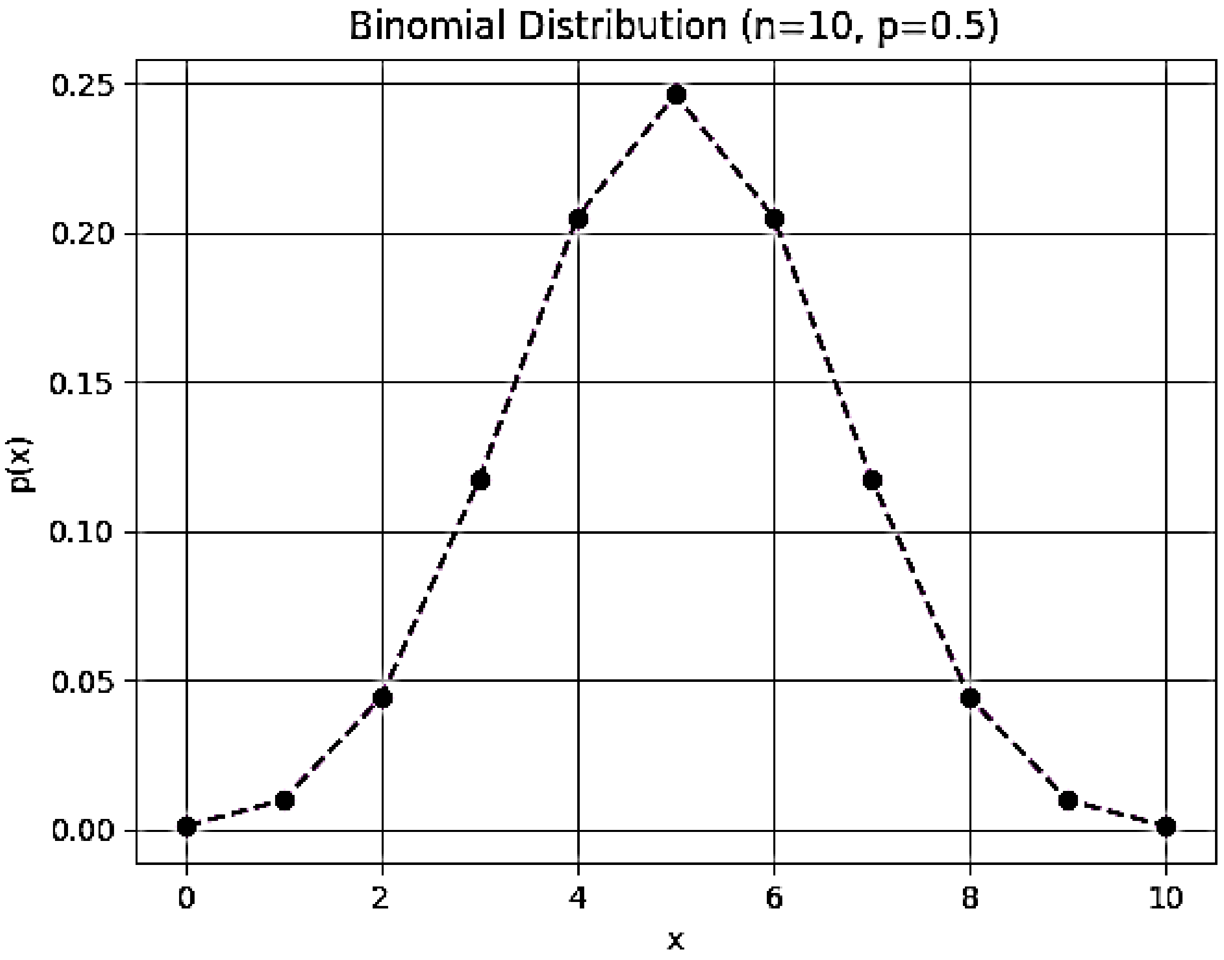
$$= n(n-1)p^2 \sum_{x=2}^{n} C(n-2, x-2) p^{x-2} (1-p)^{n-x} + np. \sum_{x=1}^{n} C(n-1, x-1) p^{x-1} (1-p)^{n-x}$$

$$= n(n-1)p^2 + np$$

$$Var(x) = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

Binomial Distribution

Plot of probability density of binomial distribution for n=10 and p=1/2



Normal or Gaussian Distribution

Binomial Distribution becomes the Normal Distribution as the number of trials goes to infinity

The probability of getting k successes in n independent trials is:

$$f(x) = C(n, x)p^x(1-p)^{n-x} = \frac{n!}{(n-x)!\,x!}p^x(1-p)^{n-x}$$

Using Stirling's approximation

By definition $n! = 1.2.3.4 \ldots \ldots n$

Taking log on both sides $\ln(n!) = \ln(1) + \ln(2) + \ln(3) + \cdots \ln(n)$

For large n the summation can be approximated by integral,

$$\ln(n!) \approx \int_1^n \ln(x)\,dx = [xln(x) - x]_1^n = nln(n) - n + 1$$

Stirling found that this approximation works even better if a small correction is added

$$\ln(n!) \approx nln(n) - n + \frac{1}{2}\ln(2\pi n)$$

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Normal or Gaussian Distribution

When n becomes large the Binomial Distribution looks like a bell-shaped curve (Normal Distribution).

The most likely outcome is always around: x = np, for example flipping a fair coin n=1000 times the most likely number of heads is 50% of 1000 = 500 heads, that's exactly at x=np

So, it can be written as $x \approx np$ and $(n-x) \approx n - np = n(1-p)$

So, the probability density becomes

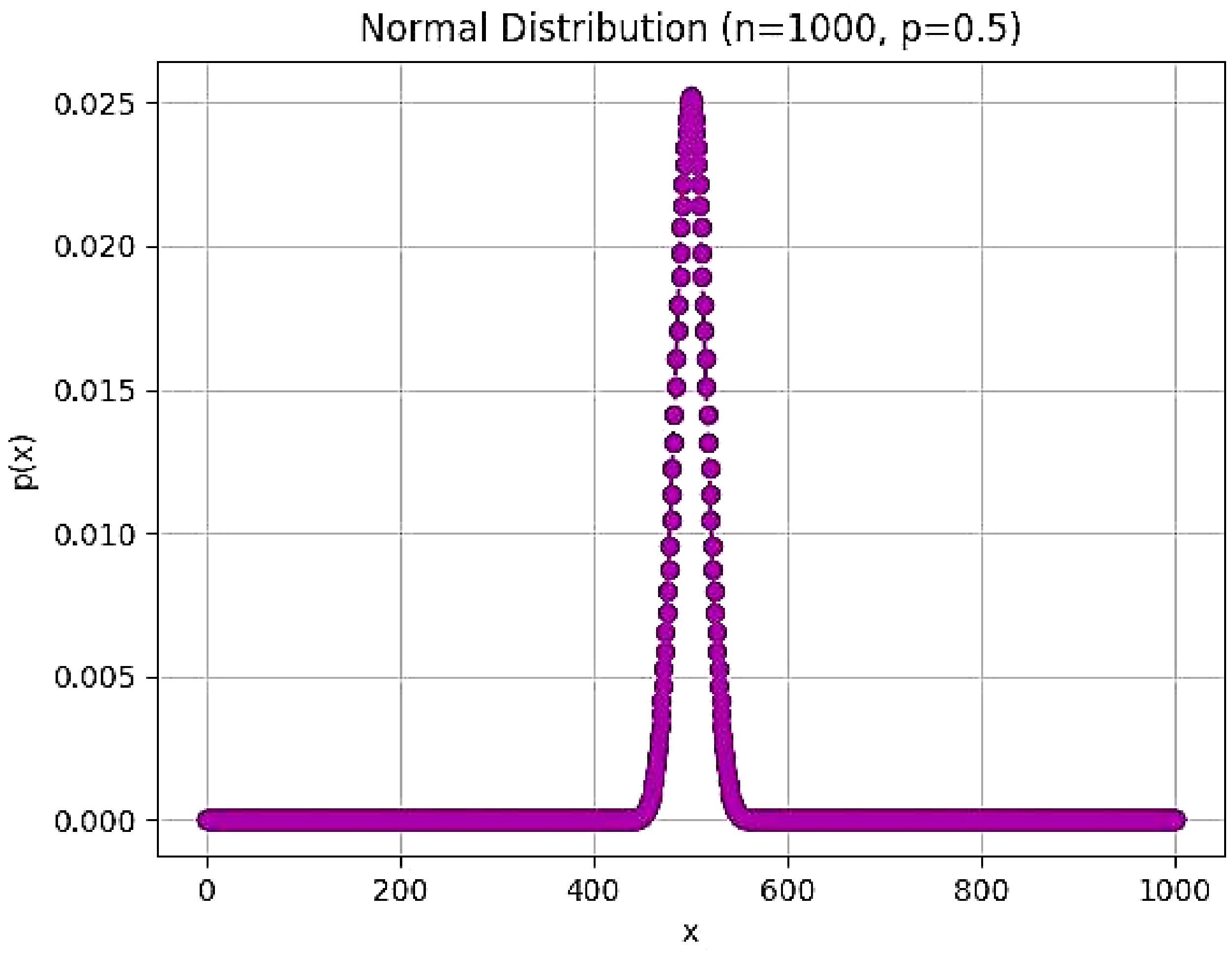
$$f(x) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}$$

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

Where $\mu = np\ is\ average\ value$
$\sigma = \sqrt{np(1-p)} = Standard\ deviation.$

Normal or Gaussian Distribution

The plot of the probability density of Normal distribution for n=1000 and p=1/2 is



Normal or Gaussian Distribution

Find the probability f exactly 52 heads in 100 tosses of a coin using the binomial distribution and using the normal approximation?

Ans:

Given n=100, and p=0.5 probability according to the binomial distribution function is

$$p(52) = C(100{,}52) \times 0.5^{52} \times 0.5^{100-52} = \frac{100!}{(100-52)!\,52!} \times 0.5^{100} \times 0.5^{100-52}$$

$$= 007353$$

For the normal approximation

$$\sigma = \sqrt{np(1-p)} = 5, \qquad \mu = 50$$
$$p(52) = 0.07365$$