

Probability

Basic Definitions

Random Experiment:

- An experiment, whose all possible outcomes are known in advance but the outcome of any specific performance cannot predicted before the completion of the experiment
- Eg: Tossing of a coin.

Sample-space

- A set of all possible outcomes associated with same random experiment
- Denoted by 'S'.
- Eg: In the experiment of tossing a die, If we are interested in the number that shows on the top face, then sample space would be $S = \{1, 2, 3, 4, 5, 6\}$

Experiment or Trial

- It is a series of action where the outcomes are always uncertain.
- Eg: - Tossing of a coin, Selecting a card from deck of cards, throwing a dice.

Event

- Subset of sample – space.
- In any sample space we may be interested in the occurrence of certain events rather than in the occurrence of a specific element in the sample space.

Simple Event

- If an event is a set containing only one element of the sample-space

Compound Event

- A compound event is one that can be represented as a union of sample points
- Eg:
Event of drawing a heart from a deck of cards is the subset $A = \{\text{heart}\}$ of the sample space $S = \{\text{heart, spade, club, diamond}\}$.
Therefore A is a simple event. None the event B of drawing a red card is a compound event since $B = \{\text{heart} \cup \text{diamond}\} = \{\text{heart, diamond}\}$.

Probability

- If a random experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes favours to A,
Then the probability of event A, $P(A) = n/N$ i.e. favourable cases/total no. of cases.
- Remarks:
 - If the probability of certain event is one, it doesn't mean that event is going to happen with certainty
 - It's just predicting that, the event is most likely to occur in comparison to other events. Predictions depend upon the past information and of course also on the way of analysing the information at hand
 - Similarly if the probability of certain event is zero, it doesn't mean that, the event can never occur!

Mutually exclusive Event

- If two events are mutually exclusive they cannot occur simultaneously.

Independent Events

- Events are said to be independent if the occurrence or non-occurrence of one does not affect the occurrence or non-occurrence of other.

Exhaustive Event

- A set of events is said to be exhaustive if the performance of random experiment always result in the occurrence of at least one of them

Conditional Probability

- The conditional probability of an event B is the probability that the event will occur given the knowledge that an event A has already occurred.
- This probability is written $P(B|A)$, notation for the probability of B given A.
- In the case where events A and B are independent (where event A has no effect on the probability of event B), the conditional probability of event B given event A is simply the probability of event B, that is $P(B)$.
- If events A and B are not independent, then the probability of the intersection of A and B (the probability that both events occur) is defined by $P(A \text{ and } B) = P(A) P(B|A)$.
- If E and F are two events associated with the same sample space of a random experiment, the conditional probability of the event E given that F has occurred,

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \text{ provided } P(F) \neq 0$$

I.e. $P(E|F)$ is given by

- Eg: If $P(A) = \frac{7}{13}$, $P(B) = \frac{9}{13}$ and $P(A \cap B) = \frac{4}{13}$, evaluate $P(A|B)$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{4}{13}}{\frac{9}{13}} = \frac{4}{9}$$

Properties of conditional probability

Let E and F be events of a sample space S of an experiment, then we have

Property 1

$$P(S|F) = P(F|F) = 1$$

We know that

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

Also

$$P(F|F) = \frac{P(F \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

Thus

$$P(S|F) = P(F|F) = 1$$

Property 2

If A and B are any two events of a sample space S and F is an event of S such that $P(F) \neq 0$, then

$$P((A \cup B)|F) = P(A|F) + P(B|F) - P((A \cap B)|F)$$

In particular, if A and B are disjoint events,

$$P((A \cap B)|F) = P(A|F) + P(B|F)$$

We have

$$\begin{aligned} P((A \cup B)|F) &= \frac{P[(A \cup B) \cap F]}{P(F)} \\ &= \frac{P[(A \cap F) \cup (B \cap F)]}{P(F)} \end{aligned}$$

(by distributive law of union of sets over intersection)

$$\begin{aligned} &= \frac{P(A \cap F) + P(B \cap F) - P(A \cap B \cap F)}{P(F)} \\ &= \frac{P(A \cap F)}{P(F)} + \frac{P(B \cap F)}{P(F)} - \frac{P[(A \cap B) \cap F]}{P(F)} \\ &= P(A|F) + P(B|F) - P((A \cap B)|F) \end{aligned}$$

When A and B are disjoint events, then

$$P((A \cap B)|F) = 0$$

$$\Rightarrow P((A \cup B)|F) = P(A|F) + P(B|F)$$

Property 3

$$P(E'|F) = 1 - P(E|F)$$

From Property 1, we know that $P(S|F) = 1$

$$\Rightarrow P(E \cup E'|F) = 1 \quad \text{since } S = E \cup E'$$

$$\Rightarrow P(E|F) + P(E'|F) = 1 \quad \text{since } E \text{ and } E' \text{ are disjoint events}$$

$$\text{Thus, } P(E'|F) = 1 - P(E|F)$$

Multiplication Theorem on Probability

- Let E and F be two events associated with a sample space S.
- Conditional probability of event E given that F has occurred is denoted by $P(E|F)$ and is given by

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, P(F) \neq 0$$

- From this result, we can write

$$P(E \cap F) = P(F) \cdot P(E|F) \quad \dots (1)$$

Also, we know that

$$P(F|E) = \frac{P(F \cap E)}{P(E)}, P(E) \neq 0$$

or

$$P(F|E) = \frac{P(E \cap F)}{P(E)} \quad (\text{since } E \cap F = F \cap E)$$

Thus,

$$P(E \cap F) = P(E) \cdot P(F|E) \quad \dots (2)$$

Combining (1) and (2), we find that

$$\begin{aligned} P(E \cap F) &= P(E) P(F|E) \\ &= P(F) P(E|F) \text{ provided } P(E) \neq 0 \text{ and } P(F) \neq 0. \end{aligned}$$

The above result is known as the **Multiplication rule of probability**.

• **Example:**

- An urn contains 10 black and 5 white balls. Two balls are drawn from the urn one after the other without replacement. What is the probability that both drawn balls are black?
- **Solution:**
 - Let E and F denote respectively the events that first and second ball drawn are black.
 - $P(E) = P(\text{black ball in first draw}) = 10/15$
 - Given
 - First ball drawn is black, i.e., event E has occurred, now there are 9 black balls and five white balls left in the urn.
 - Therefore, the probability that the second ball drawn is black, given that the ball in the first draw is black, is nothing but the conditional probability of F given that E has occurred.
 - i.e. $P(F|E) = 9/14$
 - By multiplication rule of probability, we have

$$\begin{aligned} P(E \cap F) &= P(E) P(F|E) \\ &= 10/15 \times 9/14 = 3/7 \end{aligned}$$

Note:

- Multiplication rule of probability for more than two events If E, F and G are three events of sample space, we have

$$P(E \cap F \cap G) = P(E) P(F|E) P(G|(E \cap F)) = P(E) P(F|E) P(G|EF)$$
- Similarly, the multiplication rule of probability can be extended for four or more events.

Independent Events

- If E and F are two events such that the probability of occurrence of one of them is not affected by occurrence of the other. Such events are called independent events.
- Let E and F be two events associated with the same random experiment, then E and F are said to be independent if

$$P(E \cap F) = P(E) \cdot P(F)$$

Remarks

- Two events E and F are said to be dependent if they are not independent, i.e. if $P(E \cap F) \neq P(E) \cdot P(F)$
- Sometimes there is a confusion between independent events and mutually exclusive events.
 - Term 'independent' is defined in terms of 'probability of events' whereas mutually exclusive is defined in term of events (subset of sample space).
 - Mutually exclusive events never have an outcome common, but independent events, may have common outcome.
 - Two independent events having nonzero probabilities of occurrence cannot be mutually exclusive, and conversely, i.e. two mutually exclusive events having nonzero probabilities of occurrence cannot be independent.
- Two experiments are said to be independent if for every pair of events E and F, where E is associated with the first experiment and F with the second experiment, the probability of the simultaneous occurrence of the events E and F when the two experiments are performed is the product of $P(E)$ and $P(F)$ calculated separately on the basis of two experiments, i.e. $P(E \cap F) = P(E) \cdot P(F)$
- Three events A, B and C are said to be mutually independent, if

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

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

$$P(B \cap C) = P(B) P(C)$$
 and

$$P(A \cap B \cap C) = P(A) P(B) P(C)$$
 If at least one of the above is not true for three given events, we say that the events are not independent.
- **Example**
 - A die is thrown. If E is the event 'the number appearing is a multiple of 3' and F be the event 'the number appearing is even' then find whether E and F are independent ?
 Solution:
 W.k.t the sample space is $S = \{1, 2, 3, 4, 5, 6\}$
 Now $E = \{3, 6\}$, $F = \{2, 4, 6\}$ and $E \cap F = \{6\}$
 Then
 - $P(E) = 2/6 = 1/3$
 - $P(F) = 3/6 = 1/2$
 - $P(E \cap F) = 1/6$

Clearly $P(E \cap F) = P(E) \cdot P(F)$. Hence E and F are independent events.

Bayes' Theorem: Description

- Also called as inverse probability theorem
- Consider that there are two bags I and II.
 - Bag I contains 2 white and 3 red balls
 - Bag II contains 4 white and 5 red balls.
 - One ball is drawn at random from one of the bags.

- Probability of selecting any of the bags (i.e. $1/2$) or probability of drawing a ball of a particular colour (say white) from a particular bag (say Bag I).
- Probability that the ball drawn is of a particular colour, if we are given the bag from which the ball is drawn.
- To find the probability that the ball drawn is from a particular bag (say Bag II), if the colour of the ball drawn is given we have to find the reverse probability of Bag II to be selected when an event occurred after it is known.
- Famous mathematician, John Bayes' solved the problem of finding reverse probability by using conditional probability.
- Hence named as 'Bayes theorem' which was published posthumously in 1763.

Definitions:

Partition of a sample space

- A set of events E_1, E_2, \dots, E_n is said to represent a partition of the sample space S if
 - $E_i \cap E_j = \emptyset, i \neq j, i, j = 1, 2, 3, \dots, n$
 - $E_1 \cup E_2 \cup \dots \cup E_n = S$ and
 - $P(E_i) > 0$ for all $i = 1, 2, \dots, n$.
- The events E_1, E_2, \dots, E_n represent a partition of the sample space S if they are pairwise disjoint, exhaustive and have nonzero probabilities.

Theorem of total probability

- Let $\{E_1, E_2, \dots, E_n\}$ be a partition of the sample space S ,
- Suppose that each of the events E_1, E_2, \dots, E_n has nonzero probability of occurrence.
- Let A be any event associated with S , then

$$P(A) = P(E_1) P(A|E_1) + P(E_2) P(A|E_2) + \dots + P(E_n) P(A|E_n)$$

$$= \sum_{j=1}^n P(E_j) P(A|E_j)$$

Proof

Given that E_1, E_2, \dots, E_n is a partition of the sample space S . Therefore,

$$S = E_1 \cup E_2 \cup \dots \cup E_n$$

and $E_i \cap E_j = \emptyset, i \neq j, i, j = 1, 2, \dots, n$

Now, we know that for any event A ,

$$\begin{aligned} A &= A \cap S \\ &= A \cap (E_1 \cup E_2 \cup \dots \cup E_n) \\ &= (A \cap E_1) \cup (A \cap E_2) \cup \dots \cup (A \cap E_n) \end{aligned}$$

Also $A \cap E_i$ and $A \cap E_j$ are respectively the subsets of E_i and E_j . We know that E_i and E_j are disjoint, for $i \neq j$, therefore, $A \cap E_i$ and $A \cap E_j$ are also disjoint for all $i \neq j, i, j = 1, 2, \dots, n$.

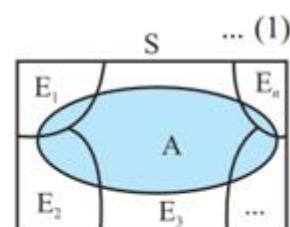


Fig 13.4

Thus,
$$P(A) = P[(A \cap E_1) \cup (A \cap E_2) \cup \dots \cup (A \cap E_n)]$$
$$= P(A \cap E_1) + P(A \cap E_2) + \dots + P(A \cap E_n)$$

Now, by multiplication rule of probability, we have

$$P(A \cap E_i) = P(E_i) P(A|E_i) \text{ as } P(E_i) \neq 0 \forall i = 1, 2, \dots, n$$

Therefore,
$$P(A) = P(E_1) P(A|E_1) + P(E_2) P(A|E_2) + \dots + P(E_n) P(A|E_n)$$

or
$$P(A) = \sum_{j=1}^n P(E_j) P(A|E_j)$$

Bayes' Theorem: Proof

If E_1, E_2, \dots, E_n are n non empty events which constitute a partition of sample space S , i.e. E_1, E_2, \dots, E_n are pairwise disjoint and $E_1 \cup E_2 \cup \dots \cup E_n = S$ and A is any event of nonzero probability, then

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{\sum_{j=1}^n P(E_j)P(A|E_j)} \text{ for any } i = 1, 2, 3, \dots, n$$

Proof:

By formula of conditional probability, we know that

$$\begin{aligned} P(E_i|A) &= \frac{P(A \cap E_i)}{P(A)} \\ &= \frac{P(E_i)P(A|E_i)}{P(A)} \text{ (by multiplication rule of probability)} \\ &= \frac{P(E_i)P(A|E_i)}{\sum_{j=1}^n P(E_j)P(A|E_j)} \text{ (by the result of theorem of total probability)} \end{aligned}$$

Remark

The following terminology is generally used when Bayes' theorem is applied.

- The events E_1, E_2, \dots, E_n are called hypotheses.
- The probability $P(E_i)$ is called the priori probability of the hypothesis E_i
- The conditional probability $P(E_i | A)$ is called a posteriori probability of the hypothesis E_i .
- Also called the formula for the probability of "causes". Since the E_i 's are a partition of the sample space S , one and only one of the events E_i occurs (i.e. one of the events E_i must occur and only one can occur). Hence, the above formula gives us the probability of a particular E_i , given that the event A has occurred.

Random Variables and its Probability Distributions

In most of the random experiments and Sample space, we were not only interested in the particular outcome that occurs but rather in some number associated with that outcomes as shown in following examples/experiments.

- Experiments
 - In tossing two dice, we may be interested in the sum of the numbers on the two dice.
 - In tossing a coin 50 times, we may want the number of heads obtained.
 - In the experiment of taking out four articles (one after the other) at random from a lot of 20 articles in which 6 are defective, we want to know the number of defectives in the sample of four and not in the particular sequence of defective and non-defective articles.
- In all the above experiments,
 - We have a rule which assigns to each outcome of the experiment a single real number.
 - This single real number may vary with different outcomes of the experiment. Hence, it is a variable.
 - Also its value depends upon the outcome of a random experiment and, hence, is called random variable.
 - A random variable is usually denoted by X .
 - A random variable can take any real value, therefore, its co-domain is the set of real numbers. Hence, a random variable can be defined as follows
 - A random variable is a real valued function whose domain is the sample space of a random experiment.
 - Eg: Consider the experiment of tossing a coin two times in succession
 - Sample space of the experiment is $S = \{HH, HT, TH, TT\}$.
 - If X denotes the number of heads obtained, then X is a random variable and for each outcome, its value is as given below :
 - $X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0$
 - Let Y denote the number of heads minus the number of tails for each outcome of the above sample space S .
 - $Y(HH) = 2, Y(HT) = 0, Y(TH) = 0, Y(TT) = -2$.
 - Hence, X and Y are two different random variables defined on the same sample space S .
 - Note: More than one random variables can be defined on the same sample space.

Probability distribution of a random variable

- Description giving the values of the random variable along with the corresponding probabilities is called the probability distribution of the random variable X .
- In general, the probability distribution of a random variable X is defined as follows:
 - The probability distribution of a random variable X is the system of numbers

$$\begin{array}{ccccccc} X & : & x_1 & x_2 & \dots & x_n \\ P(X) & : & p_1 & p_2 & \dots & p_n \end{array}$$

where, $p_i > 0, \sum_{i=1}^n p_i = 1, i = 1, 2, \dots, n$

The real numbers x_1, x_2, \dots, x_n are the possible values of the random variable X and p_i ($i = 1, 2, \dots, n$) is the probability of the random variable X taking the value x_i i.e., $P(X = x_i) = p_i$

- Also for all possible values of the random variable X , all elements of the sample space are covered. Hence, the sum of all the probabilities in a probability distribution must be one.
- If x_i is one of the possible values of a random variable X , the statement $X = x_i$ is true only at some point (s) of the sample space. Hence, the probability that X takes value x_i is always nonzero, i.e. $P(X = x_i) \neq 0$.

Mean of a random variable

- Mean is a measure of location or central tendency in the sense that it roughly locates a middle or average value of the random variable.
- Let X be a random variable whose possible values $x_1, x_2, x_3, \dots, x_n$ occur with probabilities $p_1, p_2, p_3, \dots, p_n$, respectively. The mean of X , denoted by μ , is the number $\sum_{i=1}^n x_i p_i$ i.e. the mean of X is the weighted average of the possible values of X , each value being weighted by its probability with which it occurs.

- The mean of a random variable X is also called the expectation of X , denoted by $E(X)$

$$E(X) = \mu = \sum_{i=1}^n x_i p_i = x_1 p_1 + x_2 p_2 + \dots + x_n p_n.$$

Thus

- The mean or expectation of a random variable X is the sum of the products of all possible values of X by their respective probabilities.

Variance of a random variable

- The mean of a random variable does not give us information about the variability in the values of the random variable.
- If the variance is small, then the values of the random variable are close to the mean. Also random variables with different probability distributions can have equal means, as shown in the following distributions of X and Y

X	1	2	3	4
$P(X)$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{4}{8}$

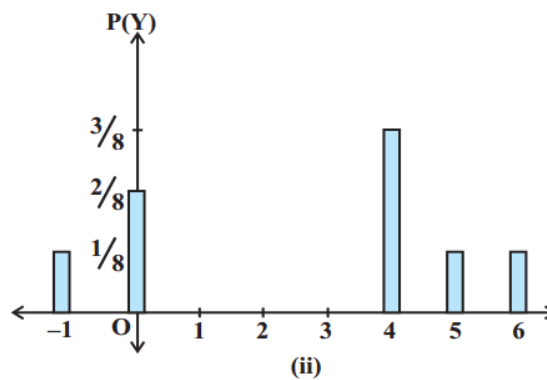
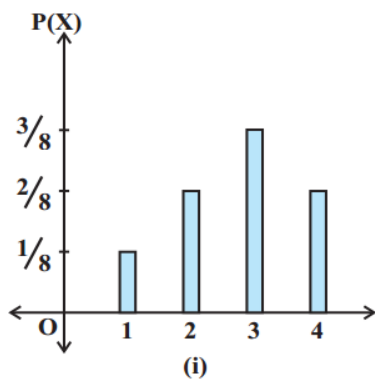
Y	-1	0	4	5	6
$P(Y)$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$$E(X) = 1 \times \frac{1}{8} + 2 \times \frac{2}{8} + 3 \times \frac{3}{8} + 4 \times \frac{4}{8} = \frac{22}{8} = 2.75$$

$$E(Y) = -1 \times \frac{1}{8} + 0 \times \frac{2}{8} + 4 \times \frac{3}{8} + 5 \times \frac{1}{8} + 6 \times \frac{1}{8} = \frac{22}{8} = 2.75$$

The variables X and Y are different, however their means are same.

The diagrammatic representation of these distributions are shown below:



- Let X be a random variable whose possible values x_1, x_2, \dots, x_n occur with probabilities $p(x_1), p(x_2), \dots, p(x_n)$ respectively.

Let $\mu = E(X)$ be the mean of X . The variance of X , denoted by $\text{Var}(X)$ or σ_x^2 is defined as

$$\sigma_x^2 = \text{Var}(X) = \sum_{i=1}^n (x_i - \mu)^2 p(x_i)$$

or equivalently $\sigma_x^2 = E(X - \mu)^2$

$$\sigma_x = \sqrt{\text{Var}(X)} = \sqrt{\sum_{i=1}^n (x_i - \mu)^2 p(x_i)}$$

The non-negative number is called the standard deviation of the random variable

Another formula to find the variance of a random variable. We know that,

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^n (x_i - \mu)^2 p(x_i) \\ &= \sum_{i=1}^n (x_i^2 + \mu^2 - 2\mu x_i) p(x_i) \\ &= \sum_{i=1}^n x_i^2 p(x_i) + \sum_{i=1}^n \mu^2 p(x_i) - \sum_{i=1}^n 2\mu x_i p(x_i) \\ &= \sum_{i=1}^n x_i^2 p(x_i) + \mu^2 \sum_{i=1}^n p(x_i) - 2\mu \sum_{i=1}^n x_i p(x_i) \\ &= \sum_{i=1}^n x_i^2 p(x_i) + \mu^2 - 2\mu^2 \left[\text{since } \sum_{i=1}^n p(x_i) = 1 \text{ and } \mu = \sum_{i=1}^n x_i p(x_i) \right] \\ &= \sum_{i=1}^n x_i^2 p(x_i) - \mu^2 \end{aligned}$$

$$\text{or} \quad \text{Var}(X) = \sum_{i=1}^n x_i^2 p(x_i) - \left(\sum_{i=1}^n x_i p(x_i) \right)^2$$

$$\text{or} \quad \text{Var}(X) = E(X^2) - [E(X)]^2, \text{ where } E(X^2) = \sum_{i=1}^n x_i^2 p(x_i)$$

Bernoulli Trials and Binomial Distribution

Bernoulli trials

- The outcome of any trial is independent of the outcome of any other trial. In each of such trials, the probability of success or failure remains constant. Such independent trials which have only two outcomes usually referred as 'success' or 'failure' are called Bernoulli trials.
- Trials of a random experiment are called Bernoulli trials, if they satisfy the following conditions :
 - There should be a finite number of trials.
 - The trials should be independent.
 - Each trial has exactly two outcomes: success or failure.
 - The probability of success remains the same in each trial.
 - Example: 30 Six balls are drawn successively from an urn containing 7 red and 9 black balls. Tell whether or not the trials of drawing balls are Bernoulli trials when after each draw the ball drawn is (i) replaced (ii) not replaced in the urn.

Solution

(i) The number of trials is finite. When the drawing is done with replacement, the probability of success (say, red ball) is $p = 7/16$ which is same for all six trials (draws). Hence, the drawing of balls with replacements are Bernoulli trials.

(ii) When the drawing is done without replacement, the probability of success (i.e., red ball) in first trial is $7/16$, in 2nd trial is $6/15$ if the first ball drawn is red or $7/15$ if the first ball drawn is black and so on. Clearly, the probability of success is not same for all trials, hence the trials are not Bernoulli trials

Binomial distribution

The probability distribution of number of successes in an experiment consisting of n Bernoulli trials may be obtained by the binomial expansion of $(q + p)^n$. Hence, this distribution of number of successes X can be written as

X	0	1	2	...	x	...	n
P(X)	${}^nC_0 q^n$	${}^nC_1 q^{n-1}p^1$	${}^nC_2 q^{n-2}p^2$		${}^nC_x q^{n-x}p^x$		${}^nC_n p^n$

The above probability distribution is known as binomial distribution with parameters n and p , because for given values of n and p , we can find the complete probability distribution.

The probability of x successes $P(X = x)$ is also denoted by $P(x)$ and is given by

$$P(x) = {}^nC_x q^{n-x}p^x, \quad x = 0, 1, \dots, n. \quad (q = 1 - p)$$

This $P(x)$ is called the probability function of the binomial distribution. A binomial distribution with n -Bernoulli trials and probability of success in each trial as p , is denoted by $B(n, p)$.