to the ground due to the force of gravitation.

Definition: An experiment is called a ***random experiment*** if, when repeated under the same conditions, it is such that the outcome can not be predicted with certainty but all possible out comes can be determined prior to the performance of the experiment.

Now onwards, an experiment means a random experiment.

Definition: The collection of all possible outcomes of a random experiment is called the ***Sample space***, denoted by S. The elements of Sample space are ***Sample points***

Example: In the toss of a single coin let the outcome 'Head turning up' be denoted by H and the outcome 'Tail turning up' be denoted by T. The coin is repeatedly tossed under the same conditions. Then the toss of the coin results in the outcome H or T, thus yielding the sample space $S = \{H, T\}$.

Example: A die is numbered with 1, 2, 3, 4, 5, 6 on the faces. When this die is thrown the Sample space is $S = \{1, 2, 3, 4, 5, 6\}$

Example: In the random experiment of drawing one card from a pack of 52 cards, outcome is any particular card and hence the Sample space S consists of all individual cards and $n(S) = 52$.

Definition: A Sample space is called ***finite Sample space***, if its sample points are finite

Definition: A Sample space is called an ***infinite Sample space***, if its sample points are infinite

Example: A bulb is allowed to burn continuously till it expires. Then the sample point (outcome) e may be expressed in number of hours (minutes and seconds ignored) and $e > 0$. Then the sample space is $S = \{1, 2, \ldots\}$ which is an infinite set, implying that it is an infinite sample space.

In order to model and analyze a random experiment, we must understand the collection of possible outcomes of the experiment. For this, in this introduction to probability, we use the basic concepts of sets and operations on sets, because a sample space of an experiment is the set S of possible outcomes of the experiment.

1.2.1 Axioms of Probability:

Axiomatic definition of probability

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties.

If S is the sample space and E is any event in a random experiment,

1. $0 \leq P(E) \leq 1$ for each event E is S.
2. $P(S) = 1$
3. If E_1 and E_2 are any mutually exclusive events in S, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

The first axiom states that probability are real numbers in the interval from 0 to 1.

The second axiom states that the sample space as a whole is assigned a probability of 1 and this expresses the idea that the probability of a certain event, is equal to 1.

The Third axiom states that the probability of the sum of two mutually exclusive events is equal to the sum of their probabilities.

Example: A broker feels that the probability that a given stock will go up in value during the day's trading is 0.3 and the probability that it will go down in value is 0.1. what is the probability that it will go up or down?

Solution: By noting the two events E_1: The stock goes up in value, E_2: The stock goes down in value, constitute a finite collection of mutually exclusive events. The events are mutually exclusive, since the closing price of the stock can't be both above and below its starting price simultaneously.

Then $P(E_1 \cup E_2) = P(E_1) + P(E_2) = 0.3 + 0.1 = 0.4$.

Axioms of probability do not determine probabilities. But the axioms restrict the assignments of probabilities in a manner that enables us to interpert probabilities as relative frequencies without inconsistencies.

1.3. Some Elementary Theorems:

The basic axioms can be used to derive a series of rules or theorems that help us calculate the probabilities of various events. With the use of mathematical induction the third axiom of probability can be extended or generalized to include any number of mutually exclusive events.

Theorem (Generalization of third axiom of probability):

If $E_1, E_2, \ldots, E_n$ are mutually exclusive events in a sample space S,

then $P(E_1 \cup E_2 \cup E_3 \ldots \cup E_n) = P(E_1) + P(E_2) + \ldots + P(E_n)$.

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

Example: A police department needs new tires for its patrol cars and the probabilities are 0.17, 0.22, 0.03, 0.29, 0.21 and 0.08 that it will buy Uniroyal tires, Goodyear tires, Michelin tires, General tires, Goodrich tires or Armstrong tires. Find the probabilities that it will buy

(a) Goodyear or Goodrich tires;
(b) Uniroyal, General or Goodrich tires;
(c) Michelin or Armstrong tires;
(d) Goodyear, General or Armstrong tires

Solution: Since the possibilities are all mutually exclusive

(a) $P(\text{Good year or Good rich tires}) = P(\text{Good year}) + P(\text{Good rich tires})$

$= 0.22 + 0.21$

$= 0.43$

Similarly (b) 0.67 (c) 0.11 (d) 0.59

Corollary: For any infinite sequience of mutually exclusive events $E_1, E_2, \ldots E_n, \ldots$

$$P(E_1 \cup E_2 \cup \ldots \cup E_n \cup \ldots) = P(E_1) + P(E_2) + \ldots + P(E_n) + \ldots$$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Theorem: (Rule of calculating probability of an event):

If A is an event in the finite sample space S, then P(A) equals the sum of the probabilities of the individual outcomes comprising A.

Proof: Let $E_1, E_2, \ldots, E_n$ be the n outcomes comprising event A,

then $A = E_1 \cup E_2 \cup \ldots \cup E_n$

Since $E's$ are individual outcomes, they are mutually exclusive

$$\begin{aligned} \therefore\ P(A) &= P(E_1 \cup E_2 \cup \ldots \cup E_n) \\ &= P(E_1) + P(E_2) + \ldots + P(E_n) \quad \text{[by axiom 3]} \end{aligned}$$

Hence the theorem. □

Example: When we roll a pair of balanced dice, what are the probabilities of getting 2, 3 or 12 ?

Solution: In throwing a pair of balanced dice simultaneously, there are 6×6 elementary events like (1, 1), (1, 2) ... (1, 6), ... (6, 6) in the sample space which are equally likely. Let

E_1 = event of getting 2 = $\{(1, 1)\}$

E_2 = event of getting 3 = $\{(1, 2), (2, 1)\}$ ∵; $E_1 \cap E_2 = \phi$

E_3 = event of getting 12 = $\{(6, 6)\}$; $E_1 \cap E_3 = \phi$, $E_2 \cap E_3 = \phi$

$$P(E_1) = \frac{1}{36}, P(E_2) = \frac{2}{36}. P(E_3) = \frac{1}{36}$$

clearly A = event of getting 2, (or) 3 (or) 12 = $E_1 \cup E_2 \cup E_3$

$$\begin{aligned} \therefore\ P(A) &= P(E_1) + P(E_2) + P(E_3) \\ &= \frac{1}{36} + \frac{2}{36} + \frac{1}{36} \\ &= \frac{4}{36} \\ &= \frac{1}{9}. \end{aligned}$$

□

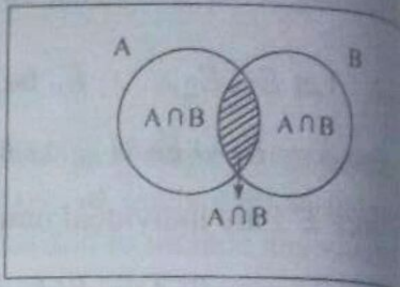
We know that the third axiom of probability can be extended to include more tha mutually exclusive events. Now we are going to see the extension of axiom 3 that al us to find the probability of the union of ***any two events*** in S regardless of whether are mutually exclusive.

Theorem: (General Addition Rule):

If A and B are any events in S, then

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

Proof. $A \cup B$ can be written as the union of two mutually exclusive events A and $\bar{A} \cap B$, since A and $(\bar{A} \cap B)$ are disjoint,



$$P(A \cup B) = P(A) + P(\bar{A} \cap B)$$

$$= P(A) + \left[P(\bar{A} \cap B) + P(A \cap B) - P(A \cap B)\right]$$

$$= P(A) + P\left[(\bar{A} \cap B) \cup (A \cap B)\right] - P(A \cap B)$$

$$(\because \bar{A} \cap B \text{ and } A \cap \bar{B} \text{ are disjoint})$$

$$= P(A) + P(B) - P(A \cap B).$$

Hence the theorem.

When A and B are mutually exclusive, then $P(A \cap B) = 0$, then the above theo reduces to the third axiom of probability, that's why, axiom 3 can be referred as **Special addition Rule**.

Theorem: The impossible event ϕ has probability zero i.e; $P(\phi) = 0$

Proof: For any event E we have $E \cup \phi = E$, $E \cap \phi = \phi$ i.e., E and ϕ are mutu exclusive

$$\text{Then} \quad P(E \cup \phi) = P(E)$$

$$\Rightarrow \quad P(E) + P(\phi) = P(E) \qquad \text{by axiom 3}$$

$$\Rightarrow \quad P(\phi) = 0.$$

Hence the theorem.

Theorem: (Probability of complement event):

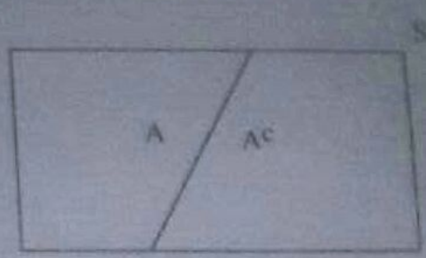
If A is any event in S, then $P(A^c) = 1 - P(A)$

Proof: Let A be any event in S, then A and A^c are mutually exclusive events i.e., $A \cap A^c = \phi$, also $A \cup A^c = S$

$$\Rightarrow P(A \cup A^c) = P(S) = 1$$

$$\Rightarrow P(A) + P(A^c) = 1 \qquad (\because \text{ by axiom 2 and 3})$$

so that $\quad P(A^c) = 1 - P(A).$



Hence the theorem. □

Aliter: As we know $S^c = \phi$ or $\phi^c = S$, $P(\phi) = P(S^c) = 1 - P(S) = 0.$

1.4. Conditional Probability:

In many cases, the probabilities of two or more events depend on one another. This means, the happening of one event depends on the happening of another event.

Example: To a merchant of umbrellas, the probability to get profit on a rainy day is more than the probability of gettting profit on any other day. Clearly, ***the event of "getting profit"*** depends on the **event of raining.**

Definition: If A, E are any two events of a smaple space S, then the event of "happening of E, after the happening of A" is called ***conditional event*** and is denoted by $P(E|A)$

However, it is necessary to find the probability of an event E, given the supplementary condition that an event A has preceded it and it has positive probability.

Let us now make the following general definition:

Conditional probability

If E and A are any events in S, $P(A) > 0$, the ***conditional probability*** of E given A is

$$P(E|A) = \frac{P(E \cap A)}{P(A)} = \frac{P(\text{both events E and A occur})}{P(\text{given event A occurs})}$$

Example: A die is rolled. If the outcome is an odd number. What is the probability that it is prime?

Solution: when a die is rolled, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$

Let A = event of getting an odd number = $\{1, 3, 5\}$

Let E = event of getting a prime number = $\{2, 3, 5\}$

Then $E \cap A = \{3, 5\}$

$\therefore P(A) = \frac{3}{6} = \frac{1}{2}$; $P(E) = \frac{3}{6} = \frac{1}{2}$ and $P(E \cap A) = \frac{2}{6} = \frac{1}{3}$, P(getting a prime, already which is an odd number)

$$= P(\text{getting a prime}|\text{getting an odd number})$$
$$= P(E|A)$$
$$= \frac{P(E\cap A)}{P(A)}$$
$$= \frac{1}{3}\times\frac{2}{1}$$
$$= \frac{2}{3}$$

□

Theorem: (General multiplication Rule):

If A and B are any events in S, then

$$P(A\cap B) = P(A)\cdot P(B|A) \text{ if } P(A) > 0$$
$$= P(B)\cdot P(A|B) \text{ if } P(B) > 0$$

Proof: The above statement can be obtained directly from the definition of conditional probability. □

Definition:

Two events are said to be ***independent*** if the occurreence (or) non occurrence of one event has no influence on the occurrence (or) nonoccurrence of the other event.

(OR)

If A and B are any two events in a sample S. Events A and B such that $P(A|B) = P(A)$ or $P(B|A) = P(B)$ are said to be independent events.

Example: Consider the experiment of rolling a pair of fair dice, one white and the other red. Are the events A: Doubles is rolled and B: the red is shown a 2 or a 3, independent?

Solution: Clearly $P(A) = \frac{6}{36}$, $P(B) = \frac{12}{36}$, $P(A and B) = P(A\cap B) = \frac{2}{36}$. By the definition conditional probability

$$P(A|B) = \frac{P(A\text{and}B)}{P(B)} = \frac{2/36}{12/36} = \frac{2}{12} = \frac{1}{6}$$

$$\text{and } P(A) = \frac{6}{36} = \frac{1}{6}$$

$$\therefore P(A|B) = P(A)$$

$\therefore$ A and B are independent.

Now, using definition of conditional probability and independent events, we can h
the following Rule:

Theorem: (Special multiplication Rule for independent Events)

If A and B are independent events, then

$$P(A \text{and} B) = P(A \cap B) = P(A) \cdot P(B)$$

Proof: By conditional probability

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

By the definition independent events

$$P(A|B) = P(A)$$

By using (1) and (2):

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

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

Hence the theorem.

Example: What is the probability of getting two heads in two tosses of a balance coin?

Solution: Since the probability of getting head is $\frac{1}{2}$ for each toss and the two toss are independent, the probability is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

The above rule is sometimes used as definition of independent events; In general we hav the following definition:

Definition: Two events are ***independent*** if, and only if, any one of the followi equivalent statements is true.

1. $P(A|B) = P(A)$
2. $P(B|A) = P(B)$
3. $P(A \cap B) = P(A) \cdot P(B)$

In any case, multiplication rule for independent events, is used to determine wheth wo events are independent. Also, it is useful as a tool for calculating the probability he occurrence of a sequence of events that are known to be independent.

Example: In India approximately 46% of the population has type 'O' blood. Approximately 39% has a negative Rh factor. If a person is selected at random, what is the probability that he or she will have type O- negative blood? (Individual's blood type is independent of the Rh factor).

Solution: Let A : The Blood group is O; B : The Rh factor is negative

Then

$$\begin{aligned} P(\text{type '}O\text{' and negative}) &= P(A \text{and} B) \\ &= P(A) \cdot P(B) \\ &= .46 \times .39 \\ &= .179 \end{aligned}$$

That is, approximately 17.9% of all individuals in India have type O- negative blood. □

Example: Find the probability of drawing a spade on each of two consecutive draws from a well shuffled pack of cards, without replacement of the first card.

Solution: Define the events A and B as follows:

A : drawing spade on first draw

B : drawing spade on second draw

Then $P(A) = \frac{13}{52} = \frac{1}{4}$

Given that a spade occurred on the first draw.

Then $P(B) = \frac{12}{51} = \frac{4}{17}$

Then probability that the two consecutive spades are drawn is

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B|A) \\ &= \frac{1}{4} \cdot \frac{4}{17} \\ &= \frac{1}{17}. \end{aligned}$$

□

When considering three or more independent events, we can extend the definition of independence, in the following way.

Definition: The Events $E_1, E_2, \ldots E_n$ are **independent** events if and only if

$$P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1) \times P(E_2) \times \ldots \times P(E_n)$$

Example: We often assume that successful flips of a coin are independent and that the probability of a head in a flip is 0.5. Therefore, the probability of five tosses resulting in the sequence [head, head, head, tail, tail] is $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{32}$.

Definition: If $P(A \cap B) \neq P(A) \cdot P(B)$, then the events A, B are said to be ***dependent events***

Example: If $P(A) = 0.65$, $P(B) = 0.40$ and $P(A \cap B) = 0.24$ are the events A and B dependent?

Solution: From the given values

$$P(A) \cdot P(B) = 0.65 \times 0.40 = 0.26$$

$$\text{But } P(A \cap B) = 0.24$$

$\therefore P(A \cap B) \neq P(A) \cdot P(B)$

$\therefore A$ and B are dependent events

The general multiplication Rule can be treated as the multiplication rule for Non independent events (dependent events).

For three or more dependent events the multiplication rule becomes more complicated. For instance, for three events we can write

Multiplication Rule for three dependent events

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B) \text{ provided } P(A \cap B) > 0$$

In general, for more events which are not independent, the probability that all will occur can be obtained by multiplying the probability that one of the events will occur times the probability that a second of the events will occur given that the first event has occured times the probability that a third of the events will occur given that the first two events have occured, and so on.

Example: Find the probability of drawing without replacement three aces in a row from an ordinary pack of 52 playing cards.

Solution: $\frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} = \frac{1}{5,525}$.

That is for n events $E_1, \ldots, E_n$

$P(E_1 \cap E_2 \cap \ldots E_n) = P(E_1)P(E_2|E_1) \ldots P(E_n|E_1 \cap E_2 \cap \ldots E_{n-1})$ provided $P(E_1 \cap E_2 \cap \ldots E_{n-1}) > 0$.

1.4.1 Bayes Theorem

In some problems, the outcome of an experiment is due to a particular one of the possible 'causes' (causing events) of the outcome. Even though the general multiplication rule is useful in solving these type of problems, we need a systematic developed formula:

For the purpose of obtaining a formula, let the sample space be divided into 'n' disjoint sets whose union is the sample space S. These 'n' sets will be denoted by $E_1, E_2, \ldots, E_n$. They represent the n possible causes (causing events) of an experimental outcome. Next, let A be an event that occurred when the experiment was performed, and consider the problem of calculating the probability that E_i is the cause of the occurrence of A. In the words of probability, the problem is to calculate the conditional probability $P(E_i|A)$.

Before going to get the required formula, consider the following theorem called 'Theorem of total probability', which is useful in getting the probability of A ie., $P(A)$, when the event A is due to a particular one of the possible events $E_1, E_2 \ldots E_n$:

Theorem: (of total probability):

If the event A can occur only along with the event E. Suppose also E can occur only in n mutually exclusive ways $E_1, E_2, \ldots, E_n$. Then

$$P(A) = \sum_{i=1}^{n} P(E_i) \cdot P(A|E_i) \text{ provided } P(E_i) > 0 \forall i$$

Proof: The events $A \cap E_1, \ldots, A \cap E_n$ are mutually exclusive.

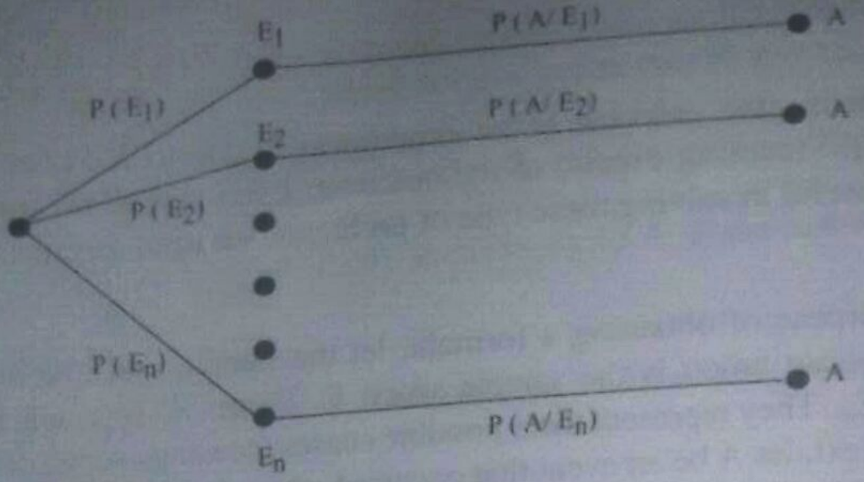
Thus

$$P(A) = P(A \cap E) = \sum_{i=1}^{n} P(A \cap E_i)$$
$$= \sum_{i=1}^{n} P(E_i) \cdot P(A|E_i)$$

Hence the theorem. □

The above rule of total probability sometimes called the rule of elimination. To visualize this rule, we have to construct the following tree diagram:

Where the probability of the final outcome is again given by the sum of the products of the probabilities corresponding to each branch of the tree.



Example: A certain product is manufactured at two plants I and II. Plant I makes 70% of the requirement and plant II makes 30%. From plant I, 90% meets a particular standard and plant II only 80% meet such standard. Evaluate, out of every 100 items purchased by a customer, how many will be up to the standard?

Solution: Let E denotes the event that the product is up to standard, F_1 and F_2 respectively, denote the events that the item is manufactured by plant I and plant II. Thus we have $P(E|F_1) = 0.90$, $P(E|F_2) = 0.80$ and $P(F_1) = 0.7$, $P(F_2) = 0.3$, since $E = (E \cap F_1) \cup (E \cap F_2)$

$$\therefore\ P(E) = P(E \cap F_1) + P(E \cap F_2) = P(F_1)P(E|F_1) + P(F_2)P(E|F_2)$$
$$= 0.70 \times 0.90 + 0.30 \times 0.80$$
$$= 0.87$$

and therefore out of every 100 items purchased, 87 will be up to standard. □

Now consider the following formula called 'Baye's Rule', to calculate $P(E_i|A)$

Theorem: (Baye's theorem):

If $E_1, E_2, \ldots, E_n$ are mutually disjoint events with $P(E_i) \neq 0$ for each i then for any arbitrary event A which is a subset of $\bigcup_{i=1}^{n} E_i$ with $P(A) > 0$, we have

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{\sum_{i=1}^{n} P(E_i)P(A|E_i)}, \quad i = 1 \text{ to } n$$

Proof: since $A \subset \bigcup_{i=1}^{n} E_i$, we have

2. Probability Distributions

2.0. Introduction

We have already observed that, in a chance experiment, it is often not the actual outcome that concerns us but some quantity that depends upon the outcome. In a random experiment, we may be interested quite often in the numerical measure of the different outcomes. It is true that in some experiments, the outcomes are directly expressed in quantitative measures. For example, consider tosses of n unbiased coins. There will be 2^n elementary events, each giving n-tuples like(HHHTTH...). We may be more interested to know the number of heads (or tails) in each outcome. For this purpose, we introduce a function X in the sample space S, associated with the each elementary event s_i ($i = 1$ to n) such that corresponding to each s_i in S, the function $X(s_i)$ gives a value. Such a function, $X(s)$ will be termed as *'Random Variable'* in our subsequent discussions.

Through the notion of random variable, we can develop methods for the study of experiments whose outcomes may be described numerically. Besides this convenience, Random variables also provide a more compact description of an experiment. The notion of random variable provides us the power of abstraction and thus allows us to discard unimportant details in the outcome of an experiment. All serious probabilistic computations are performed in terms of random variable. The description of a sample space of an experiment, which gives the information about the events along with their associated probabilities by random variable can be termed as "***Probability distribution***" of the concerned experiment (or) of the random variable.

2.1. Basic Definitions:

Definition: A ***random variable*** is a function that assigns a real number to each sample point in the sample space of a random experiment

A random variable is denoted by an uppercase letter, such as X and a corresponding lowercase letter, such as 'x' is used to denote a possible value of X.

In other words,

A real valued function X, defined on a sample space S, of a random experiment i.e., $X : S \to \mathbb{R}$ is called a ***random variable***

We refer to the set of possible numbers of a random variable X as the values of (***range of X***).

$X = x$	0	1	2
$f(x) = P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

□

We are interested not only in the probability $f(x)$, for the value of a random variable 'x', but also in the probability $F(x)$ that the value of a random variable is less than (or) equal to x. We refer to the function that assigns a value $F(x)$ to each x within the range of a random variable as the cumulative distribution Function

Definition: The ***Cumulative distribution function*** for a random variable X is defined by $F(x) = P(X \leq x)$, where x is any real number (i.e., $-\infty < x < \infty$)

Now, two important properties of F(x) are given by:

Result: (i) If $a < b$ then $P(a < X \leq b) = F(b) - F(a)$

(ii) $P(a \leq X \leq b) = P(X = a) + F(b) - F(a)$

Proof: (i) The events $a < X \leq b$ and $X \leq a$ are disjoint and their union is the event $X \leq b$

Hence $P(a < X \leq b) + P(X \leq a) = P(X \leq b)$

Therefore $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$

(ii) Since $P(a \leq X \leq b) = P(X = a) + P(a < X \leq b)$

$= P(X = a) + F(b) - F(a)$ by (i)

Hence the result. □

Clearly, by the above result,

The cumulative distribution function $F(x)$ can be obtained from probability distribution function $f(x)$.

Random variables are usually classified according to the number of values which they can take. In general, they are of two types.

(i) Discrete random variable
(ii) Continuous random variable

Definition: A ***discrete*** random variable is a random variable with a finite (or countably infinite) range

In other words, a discrete random variable is a random variable that can assume almost a finite or a countably infinite number of possible values.

To determine whether a random variable X is discrete, we ask the question, what are the possible values for X? If we can count the possibilities, then X is discrete. If we can begin to count the possibilities, but soon realize that the set of possible values for X is unending, then also X is discrete.

Notice that in either case a discrete random variable assumes its values only at isolated points.

Example: The random variable X = Sum of the spots on two dice is discrete, since X can assume only the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12. The Number of possibilities is finite, (11).

Example: The random variable W = Number of trials necessary to win in roulette is discrete, since W can assume any of the possible values 1, 2, 3, 4, 5, 6, 7, ... This collection of possibilities is countably infinite.

Definition: If the range of a random variable X is an interval of real numbers, then X is a ***continuous*** random variable.

In otherwords, a continuous random variable is one that can assume any value in some interval or interval of real numbers. Sometimes when we ask, what are the possible values for X? We are forced to admit that, conceivably, X can assume any value in some interval of real numbers, when this occurs, X is not discrete because any interval contains an uncountable number of possible values. In this case, we say that X is continuous.

Example: The random variable T = time that the first call of the day is received at a local police station is continuous, since the value of T lies in the interval (7:30, 8:00).

Example: The random variable L = length of time that it takes for a computer program to run is continuous. Let us assume that it is reasonable to expect that the value of L is less than four minutes. That is, the value of L lies in the interval (0, 4).

Note that counting problems usually give rise to discrete random variables. Measuring problems generate continuous random variables.

According to the type of random variable, in general we have two types of probability distributions, viz.,

1. Discrete probability distributions
2. Continuous probability distributions

Now Consider

2.2. Discrete Probability Distributions:

Definition: Let X be a discrete random variable. The ***discrete probability function*** $f(x)$ for X is given by

$$f(x) = P(X = x) \text{ for real } x$$

Note that the random variable is denoted by capital X and the assumed numerical value of X by a lowercase x. The probability function $f(x)$ in the discrete case is a table or an equation, that gives the possible values for X, together with the probability that X assumes those values.

Example: In tossing three coins, if $X =$ Number of heads, we have the probability distribution as:

$X = x$	0	1	2	3
$f(x) = P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Since probabilities cannot be negative, a probability function $f(x)$ cannot assume negative values. The probability associated with a sample space is 1. Thus if we add the values of $f(x)$ over all possible values oif X, the total should be 1. Infact, these two properties completely characterize the probability function of a discrete random variable.

Properties that identify a Probability function for a discrete random variable

1. $f(x) \geq 0$ for each real number x
2. $\sum_{\text{all } x} f(x) = 1$

Note that the discrete probability function $f(x)$ can also be called as

Probability mass function(p.m.f)

Any function $f(x)$ satisfying above properties 1 and 2 above will automatically be a discrete probability function or probability mass function.

Example: Check whether the following can serve as (discrete) probability function.

(a) $f(x) = \dfrac{x-2}{2}$ for $x = 1, 2, 3, 4$

(b) $h(x) = \dfrac{x^2}{25}$ for $x = 0, 1, 2, 3, 4$

Solution:

(a) This function cannot serve as a probability distribution because $f(1)$ is nega...

(b) This function cannot serve as a probability distribution because the sum of the five probabilities is $\frac{6}{5}$ and not 1.

Now coming to the cumulative distribution Function $F(x)$ for discrete random varia...

Definition: The **Cumulative distribution Function** $F(x)$ of a discrete rando... variable X is defined by $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$

where $f(x)$ is the discrete probability function.

Clearly, by the above definition, $F(x)$ can be obtained from $f(x)$ as follows:

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ f(x_1) + \ldots + f(x_n) & x_n \leq x < \infty \end{cases}$$

where $x_1, x_2, \ldots\ldots\ldots, x_n$ are the values of the discrete random variable X.

Example: The discrete probability function $f(x)$ of X of the random experime... consists of 3 independent tosses of a fair coin

$X = x_i$	0	1	2	3
$f(x_i) = P(X = x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Then the Cumulative probability function of X is given by

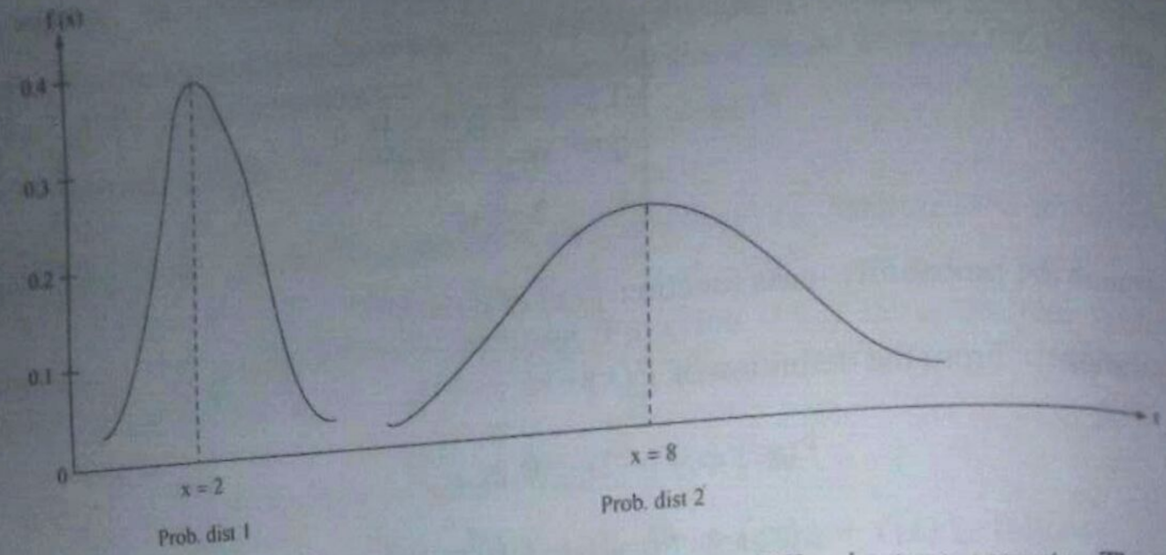
$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{8} & \text{if } 0 \leq x < 1 \\ \frac{1}{2} & \text{if } 1 \leq x < 2 \\ \frac{7}{8} & \text{if } 2 \leq x < 3 \\ 1 & \text{if } 3 \leq x < \infty \end{cases}$$

Conversely $f(x)$ can be obtained from $F(x)$. The following example show how F... can be used to determine $f(x)$ of a d.r.v.

General Properties of a Discrete Probability Distribution

The mean and the variance of a discrete probability distribution:

Now, We discuss some general characteristics of discrete probability distributions. For this consider the following graphs of two discrete probability distributions:



The above two discrete probability distributions differ in two aspects. The first probability distribution is centered about $x = 2$ where as the other is centered about $x = 8$, and we say that the two distributions differ in their ***Location***. Another distinction is that the second distribution is more spread out, and we say that two distributions differ in ***Variation***. To make such comparisons more specific, we shall introduce in this section two of the most important statistical measures, describing respectively the Location and the Variation of a probability distribution - the ***mean*** and the ***Variance***

Definition: The ***Mean*** of a discrete probability distribution is simply the mathematical expectation of a discrete random variable. If X is a discrete random variable takes the values $x_1, x_2, \ldots, x_k$ with the probabilities $f(x_1), f(x_2), \ldots$ and $f(x_k)$, then its mean or mathematical expectation or expected value is $\sum_{i=1}^{k} x_i\ f(x_i)$ or simply $\mu = \sum_{\text{all } x} x.f(x)$ where the mean is denoted by μ (Greek letter mu)(Some authors, use E(x)). The mean of a probability distribution measures its center in the sense of an average.

Example: Let X be the d.r.v taking the values $1, 2, \ldots, 6$ with probabilities $f(x_i) = \frac{1}{6}$

Then $\mu = \sum_{i=1}^{6} x_i f(x_i) = \frac{1}{6}(1 + 2 + \ldots + 6) = \frac{7}{2}$

It is easy to verify the following observations on mean

If X and Y are random variables then

(i) Mean $(c) = c$ where c is a constant

(ii) Mean $(cX) = c$ Mean (X)

(iii) Mean $(g(X)) = g(\text{Mean}(X))$, where g is a function of X

(iv) Mean $(X + Y) = \text{Mean}(X) + \text{Mean}(Y)$

(v) Mean $(XY) = \text{Mean(X)Mean(Y)}$ if X and Y are independent random variables.

Now, we define the second of the two properties i.e., ***Variance*** of a discrete probability distribution or that of discrete random variable X as:

Definition: Let X be a discrete random variable. The Variance of X, denoted by σ^2 is $\sigma^2 = \sum_{\text{all x}} (x-\mu)^2 f(x)$

However, the above definition is not usually used because it is tedious (to compute) and requiring many subtractions and squarings. The following formula often provides an easier method for computing σ^2

Result: The computational shortcut for finding the variance of X is

$$\sigma^2 = \sum_{\text{all x}} x^2 f(x) - \mu^2$$

Proof:

$$\text{Consider } \sigma^2 = \sum_{\text{all x}} (x-\mu)^2 f(x)$$

$$= \sum_{\text{all x}} (x^2 + \mu^2 - 2x\mu) f(x)$$

$$= \sum_{\text{all x}} x^2 f(x) + \mu^2 \sum_{\text{all x}} f(x) - 2\mu \sum_{\text{all x}} x f(x)$$

$$= \mu_2' + \mu^2 - 2\mu \cdot \mu \qquad \text{(because } \sum_{\text{all x}} f(x) = 1, \sum_{\text{all x}} x f(x) = \mu$$

$$= \mu_2' - \mu^2 \qquad \text{where } \mu_2' = \sum_{\text{all x}} x^2 f(x)$$

Hence the result.

and $\sum_{\text{all } x} x^2 f(x) = (1^2+2^2+3^2+4^2+5^2+6^2)\frac{1}{6} = \frac{91}{6}$

and hence $\sigma^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$.

It is easy to verify the following observations on Variance

Let X and Y be independent random variables and Let c be any real number. Then

1. Variance $(c) = 0$
2. Variance $(cX) = c^2\,\text{Var}(X)$
3. Var $(X+Y) = \text{Var}X + \text{Var}Y$

The variance is mainly useful for comparative purposes when dealing with two similar variables, it is useful to know which exhibits the greater variability. But variance is usually reported as a unitless number. However, since the standard deviation is the square root of the variance, the unit attached will always be the same as that of the original random variable. Usually, it will be physically meaningful.

Now, the standard deviation is defined as:

Definition: The Standard deviation of a discrete probability distribution or that of a discrete random variable X denoted as σ is the positive square root of σ^2 i.e.
$\sigma^2 = \sqrt{\sum_{\text{all } x}(x-\mu)^2 \cdot f(x)} = \sqrt{\sum_{\text{all } x} x^2 f(x) - \mu^2}$

Example: If the variance $\sigma^2 = 2.0979$ then standard deviation $\sigma = \sqrt{2.0979} = 1.448$
Suppose the d.r.v. X is the number of courses in which a student is enrolled, then the physical unit associated with σ is the number of courses.

Moments of a Discrete Probability Distribution:

Definition: Let X be a discrete random variable, then k^{th} moment about the origin is $\mu'_k = \sum_{\text{all } x} x^k f(x)$ and the k^{th} moment about the mean is $\mu_k = \sum_{\text{all } x}(x-\mu)^k f(x)$

Thus, the mean μ is the first moment about the origin, and the Variance σ^2 is the second moment about the mean.

Higher moments are often used in statistics to give further descriptions of probability distributions. For instance, the third moment about the mean is used to describe the symmetry or skewness of a distribution; the fourth moment about the mean is used to describe its peakedness or Kurtosis.

Result: Prove that $\mu_3 = \mu'_3 - 3\mu'_2 \cdot \mu + 2\mu^3$

Proof:

Consider $\mu_3 = \sum_{\text{all } x} (x-\mu)^3 f(x)$

$$= \sum_{\text{all } x} (x^3 - \mu^3 - 3x^2\mu + 3x\mu^2) f(x)$$

$$= \sum x^3 f(x) - \mu^3 \sum f(x) - 3\mu \sum x^2 f(x) + 3\mu^2 \sum x f(x)$$

$$= \mu_3' - \mu^3 - 3\mu\, \mu_2' + 3\mu^2 \cdot \mu$$

$$= \mu_3' - 3\mu\, \mu_2' + 2\mu^3$$

(Since $\sum f(x) = 1, \sum x^2 f(x) = \mu_2', \Sigma x f(x) = \mu$)

Hence the result. □

In previous pages, we studied the general properties of discrete probability distributions. We defined the idea of a probability distribution function in the discrete case and saw how to compute μ, σ^2 and σ using this function. Here we present two important specific discrete probability distributions that arises often in practice:

(i) The Binomial distribution
(ii) The Poisson distribution

Now consider

2.2.1 The Binomial Distribution

Many statistical problems deal with situations referred to as ***repeated trials*** and we int to develop a model that is useful in representing the probability distributions pertai to the number of occurrence of an event in repeated trials of an experiment. We consider those repeated trials in which each trial can have only two possible outco We label them arbitrarily by '**Success**' and '**failure**' without inferring that a su necessarily a desirable outcome. Situations that involve only two possible outc in an experimental trial arise in the diverse fields like, inspection of an item fr production line for defective and non defective criteria, reaction of a new drug subject being positive (or) negative, the sex determination of mammals etc.,

The repeated trials with only two possible outcomes are referred as:

Definition: A Bernouli trial is a random experiment with the following pos proposed by Bernouli:

Solution: The arrival of passengers at the checkout counter is a poisson process with $\lambda = 1.5$

(i) $P(X \leq 4) = \sum_{x=0}^{4} \frac{e^{-\lambda}\lambda^x}{x!} = 0.981$

(ii) $P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.423 = 0.577$, Since $\lambda = 1.5 \times 2 = 3$.

□

Example: If a bank receives on the average $\alpha = 6$ bad checks per day. What are the probabilities that it will receive

(a) Four bad checks on any given day
(b) 10 bad checks over any two consecutive days?

Solution: Substituting $x = 4$ and $\lambda = \alpha T = 6.1 = 6$ into the formula for the poisson distribution, we get

(a) $P(X = 4) = f(4; 6) = \frac{6^4 \cdot e^{-6}}{4!} = 0.134$

(b) here $\lambda = \alpha \cdot 2 = 12$ so to find $P(X = 10)$ with $\lambda = 12$

$$\text{i.e.,} \quad f(10; 12) = F(10; 12) - F(9; 12) = 0.347 - 0.242 = 0.105$$

□

The poisson distribution has many important applications in ***Queueing problems***, where we may be interested, for example, in the number of customers arriving for service at a cafe, the number of ships or trucks arriving to be unloaded at a receiving dock, the number of air craft arriving at an airport, and so forth.

2.3. Continuous Probability Distributions

The Binomial and the poisson distributions are examples of discrete probability distributions i.e; distributions is which the variate can take only integral values. In many distributions, however, the variate can theoretically take all values within a given range. In otherwords random variables take values on a continuous scale-for instance, the speed of car, the amount of alcohol in a persons blood, the efficiency of a solar collector or the tensile strength of a new alloy etc; are examples of continuous distributions Now, we shall first study the following concepts of continuous probability distribution.

Once a random variable has been recognized as continuous it is natural to ask quastions concerning the probability of its assuming various values. To answer such questions, a continuous probability function is needed. That is, we need a function f that can in some way be used to compute probabilities. Recall that in the discrete case $f(x) = P(X = x)$. It is obvious that this definition will not work for continuous random variables, since if X is continuous $P(X = x) = 0\ \forall\ x$. In the continuous case, we use a curve that can be used to calculate probabilities via areas. The characteristics of such a curve (Continuous probability function) are given in the following definition:

Definition: Let X be a continuous random variable. A function $f(x)$ is a *continuous probability function* of X if for any interval of real numbers $[a, b]$ such that

(i) $f(x) \geq 0$ for all x

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$ i.e., the total area bounded by the graph $f(x)$ and the horizontal axis is 1.

(iii) $P(a \leq X \leq b) = \int_{a}^{b} f(t)dt$, where a and b are any two values of x satisfying $a < b$.
i.e., the probability of observing X with a value between any two points a and b is equal to the area of the region bounded by the graph $f(x)$, $x = a, x = b$ and the horizontal axis.

Continuous probability function $f(x)$ also known as **probability density function** (p.d.f). Any function $f(x)$ satisfying above (i) and (ii) properties will automatically a continuous probability function or probability density function.

Example: Is the function defined by $f(x) = \begin{cases} 0 & x < 2 \\ \dfrac{3+2x}{18} & 2 \leq x \leq 4 \\ 0 & x > 4 \end{cases}$ a probability density function?

Solution:

(i) clearly $f(x) \geq 0$ for every x

(ii) $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{2} 0 \cdot dx + \int_{2}^{4} \frac{3+2x}{18} dx + \int_{4}^{\infty} 0 \cdot dx = 1$

Hence the given function is a density function.