INDIAN INSTITUTE OF TECHNOLOGY PATNA

Continuing Education Programme
Program: Executive M. Tech in Artificial Intelligence & Data Science Engineering
Curriculum and Syllabus-2024

Course Number	**EMC 5103**
Course Credit (L-T-P-C)	L-T-P-C: 3-0-2-4
Course Title	**Probability and Statistics**
Learning Mode	Online
Learning Objective	To understand the basic concepts in Probability Theory and Statistics through practical examples.
Course Description	The course is divided into two parts: In first part, basic concepts of probability theory are introduced. In the second part, different problems in classical statistics are discussed.
Course Outline	Conditional probability, Bayes' rule, Total probability law, Independence of events. Random variables (discrete and continuous), probability mass functions, probability density functions, Expectation, variance, moments, cumulative distribution functions, Function of random variables, Multiple random variables, joint and marginal, conditioning and independence, Markov and Chebyshev inequalities, Different notions of convergence. Weak law of large number, Central limit theorem. Estimation: Properties, Unbiased Estimator, Minimum Variance Unbiased Estimator, Rao-Cramer Inequality and its attainment, Maximum Likelihood Estimator and its invariance property, Efficiency, Mean Square Error. Confidence Interval: Coverage Probability, Confidence level, Sample size determination. Testing of Hypotheses: Null and Alternative Hypotheses, Test Statistic, Error Probabilities, Power Function, Level of Significance, Neyman-Pearson Lemma.
Learning Outcome	Students will become familiar with principal concepts probability theory and statistics. This helps them to handle, mathematically, various practical problems arising in uncertain situations.
Assessment Method	Quiz / Assignment / ESE

Text Books:

1. Ross, S.M.(2008) Introduction to Probability Models, Ninth edition, Academic Press.
2. Statistical Inference (2007), G. Casella and R.L. Berger, Duxbury Advanced Series.

Reference Book:

1. An Introduction to Probability and Statistics, V.K. Rohatgi and A.K.Md. Ehsanes Saleh, John Wiley, 2nd Ed, 2009.

Probability

If an experiment is repeated under ~~at~~ essentially homogeneous & similar conditions, we generally come across two types of situations —

(i) The result or what is usually known as the 'outcome' is unique or certain.

(ii) The result is not unique but may be one of the several possible outcomes.

The phenomenon covered by (i) are known as 'deterministic' ~~or~~ phenomenon. By deterministic phenomenon we mean one in which the result can be predicted with certainty.

eg-(a) For a perfect gas, $V \propto \frac{1}{P}$

i.e. $PV = \text{constt}$.

(b) Newton law, $v = u + at$

(c) Ohm's Law, $C = \frac{E}{R}$

(d) $\text{Speed} = \frac{\text{Distance}}{\text{Time}}$

However, there are phenomenon which do not lend themselves to deterministic approach & are known as probabilistic phenomenon.

eg(a) In a random toss of a uniform coin we are not sure of getting the head or tail.

(b) A manufacturer cannot ascertain the future demand of his product with certainty.

(c) Possibly, it will rain tonight.

(d) There is a high chance of my getting the job next month.

(e) The odds are 3:2 in favour of getting the contract applied for.

All the above sentences, with words like 'possibly', 'high chance', 'likely' & 'odds' are expressions indicating a degree of uncertainty about the happening of the event.

A numerical measures of uncertainty is provided by important branch of Mathematics called the "Theory of Probability". Broadly, there are three possible states of expectation - 'certainty', 'impossibility' & 'uncertainty'. The probability theory describes certainty by 1, impossibility by 0 & the various grades of uncertainties by coefficients ranging between 0 & 1.

Basic Terminology

1. Random Experiment :- If in each trial of an experiment conducted under identical conditions, the outcome is not unique, but may be any one of the possible outcomes, then such an experiment is called a random experiment.

2. Outcome :- The result of a random experiment will be called an outcome.

3. Trial & Event :- Any particular performance of a random experiment is called a trial & outcome or combination of outcomes are termed as events.

4. Exhaustive Events or Cases :- The total number of possible outcomes of a random experiment is known as the exhaustive events or case.

5. Favourable Events :- The number of cases favourable to an event in a trial is the number of outcomes which entail the happening of the event.

Short Story

Galileo (1564–1642), an Italian mathematician, was the first to attempt at a quantitative measure of probability while dealing with some problems related to the theory of dice in gambling.

But the first foundation of the mathematical theory of prob. was laid in the mid-seventeenth century of two French mathematician, B. Pascal (1623–62) & P. Fermat (1601–65), while solving a number of problems posed by French gambler & noble man Chevalier-De-Mere to Pascal. The famous 'problem of points' posed by De-Mere to Pascal is:

"Two persons play a game of chance".

"Two persons play a certain number of points wins the person who first gains a certain number of points wins the stake. They stop playing before the game is completed. How is the stake to be decided on the basis of the number of points each has won?" The two mathematicians, after a correspondence between themselves ultimately solved this problem & this correspondence laid the first foundation of the science of probability.

J. Bernoulli (1654–1705) → Treatise on Probability

→ T. Bayes (Inverse prob.)

→ P.S. Laplace

→ Chebyshev

→ Markoff

→ Liapounoff

→ Khintchine

→ A. Kolmogorov

6. <u>Mutually Exclusive Events</u> :- Events are said to be mutually exclusive if the happening of any one of them precludes the happening of all the others i.e. if no two or more of them can happen simultaneously in the same trial.

7. <u>Equally Likely Events</u> :- Outcomes of trial are said to be equally likely if taking into consideration all the relevant evidences, there is no reason to expect one in preference to the others.

8. <u>Independent Events</u> :- Several events are said to be independent if the happening of an event is not affected by the supplementary knowledge concerning the occurence of any number of the remaining events.

Mathematical (or Classical or "A Priori") Probability :-

<u>Def^n</u>:- If a random experiment or trial results in 'n' exhaustive, mutually exclusive & equally likely outcomes, out of which 'm' are favourable to the occurence of an event E, then the probability 'p' of occurence of E, is denoted by P(E), is given by —

$$p = P(E) = \frac{\text{Number of favourable cases}}{\text{Total no. of exhaustive cases}} = \frac{m}{n} \quad — (*)$$

↳ This definition was given by James Bernoulli who was the first person to obtain quantitative measure of uncertainty.

Remarks

(i) Since $m \geq 0$, $n > 0$ & $m \leq n$ —

$$P(E) \geq 0 \text{ \& } P(E) \leq 1$$

$$\Rightarrow \boxed{0 \leq P(E) \leq 1}$$

(ii) Sometimes we express (*) by saying that odds in favour of E are $m : (n-m)$

odds against E are $(n-m) : m$.

(iii) The non-happening of the event E is called the complementary event of E & is denoted by $\bar{E}$ or E^c.

$$p + q = 1$$

$$P(E) + P(\bar{E}) = 1$$

(iv) $P(E) = 1$, E is called a certain event

$P(E) = 0$, E is called an impossible event

Limitation of Classical Def^n

This definition of classical probability break down in the following cases —

(i) If the various outcomes of the random experiment are not equally likely.

(ii) If the exhaustive number of outcomes of the random experiment is infinite or unknown.

.(i)

eg. (a) The prob. that a candidate will pass in a certain test is not 50% since two possible outcomes, viz, success & failure (excluding the possible of a compartment) are not equally likely.

(b) The probability that a ceiling fan in a room will fall is $\frac{1}{2}$, since the events of the fan 'falling' & 'not falling' though mutually exclusive & exhaustive, are not equally likely. In fact, the prob. of the fan falling will be almost zero.

(c) If a person jumps from a running train, then the prob. of his survival will not be 50%, since in this case the events survival & death, though exhaustive and mutually exclusive, are not equally likely.

Numericals

① What is the chance that a leap year selected at random will contain 53 Sundays?

Sol. In a leap year (which consist of 366 days), there are 52 complete weeks and 2 days over. The following are the possible combinations for these two over days.

(i) Sunday & Monday ✓
(ii) Monday & Tuesday
(iii) Tuesday & Wednesday
(iv) Wednesday & Thursday
(v) Thursday & Friday
(vi) Friday & Saturday
(vii) Saturday & Sunday ✓

$\therefore$ Required Probability $= \frac{2}{7}$

② Two unbiased dice are thrown. Find the probability that

(i) both the dice show the same number
(ii) the first die shows 6
(iii) the total of the number on the dice is 8
(iv) the total of the number on the dice is greater than 8.
(v) the total of the number on the dice is 13,
(vi) the total of the number on the dice is any number from 2 to 12.

③ (a) Among the digits 1, 2, 3, 4, 5 at first one is chosen & then a second selection is made among the remaining four digits. Assuming that all twenty possible outcomes have equal probabilities, find the probability that an odd digit will be selected (i) the first time, (ii) the second time, & (iii) both times

(b) From 25 tickets, marked with first 25 numerals, one is drawn at random. Find the chance that (i) it is multiple of 5 or 7
(ii) it is multiple of 3 or 7.

Sol.

(i) Now there are 12 cases in which the first digit is odd, i.e.

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

$\therefore$ Required Probability $= \frac{12}{20} = \frac{3}{5}$

(ii) Also there are 12 cases in which the second digit drawn is odd i.e.

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

$\therefore$ R.P. $= \frac{12}{20} = \frac{3}{5}$

(iii) There are six cases in which both the digits drawn are odd i.e.

(1, 3), (1, 5), (3, 1), (3, 5), (5, 1) & (5, 3)

R.P. $= \frac{6}{20} = \frac{3}{10}$

b. (i) Multiple of 5 are 5, 10, 15, 20 & 25 i.e. 5

& multiple of 7 are 7, 14, 21 i.e. 3

$\therefore$ RP $= \frac{5+3}{25} = \frac{8}{25}$

(ii) Multiple of 3 are 3, 6, 9, 12, 15, 18, 21, 24 i.e. 8

Multiple of 7 i.e. 3

$\therefore$ RP $= \frac{8+3}{25}$

Que 4 What is the probability of getting 9 cards of the same suit in one hand at a game of bridge.

Sol.

Que. 5 A man is dealt 4 spade cards from an ordinary ~~par~~ pack of 52 cards. If he is given three more cards, find the probability p at least one of the additional card is also a spade.

Sol.

Statistical (or Empirical) Probability

Defn :- If an experiment is performed repeatedly under essentially homogeneous & identical conditions, then the limiting value of the ratio of the number of times the event occurs to the number of trials, as the number of trials becomes indefinitely large, is called the probability of happening of the event, it being assumed that the limit is finite & unique.

Symbolically, if in N-trials an event E happens M times, then the probability of the happening of E, denoted by P(E), is given by :

$$P(E) = \lim_{N \to \infty} \frac{M}{N}$$

Limitation

(i) If an experiment is repeated a large number of times, the experimental conditions may not remain identical & homogenous.

(ii) The limit may not attain a unique value, however large N may be.

Subjective Probability :- The probabilities of occurence of the corresponding events are assigned by individuals and are based on their personal judgement, wisdom, intuition & expertise. These probabilities are called the subjective probabilities & represent the degree of belief & the confidence, one has in the occurence of the respective event.

The set theory was developed by the German mathematician, G. Cantor (1845 – 1918).

Set :- A set is a well-defined collection of all possible objects having given properties & ~~so~~ specified according to well-defined rule.

The objects comprising a set ~~by~~ are called elements.

Sets are often denoted by capital letters, viz, A, B, C, etc.

eg → $A = \{ x : x \text{ rational}, x \geq 0, x^2 < 2 \}$

Remarks

(1) Every set is a subset of itself.

(2) An empty set is subset of every set.

Axiomatic Approach to Probability

It closely relates the theory of probability with the modern theory of functions & also ~~the~~ set theory, was proposed by A.N. Kolmogorov, a Russian mathematician, in 1933.

The axiomatic definition of probability includes both the classical & the statistical definitions as particular cases & overcomes the deficiencies of each of them.

A purely mathematical definition of probability cannot give the actual value of $P(A)$, the probability of occurrence of the event A & this must be considered as a function defined on all events.

Accordingly, as in the case of any function we need a domain space which is the σ-field B of the events, generated by S; a range space which is the closed interval $[0,1]$ on the real line, and a rule which assigns a value to every element of the domain space B.

Def^n:- $P(A)$ is the probability function defined on a σ-field B of events if the following properties or axioms holds –

(1) For each $A \in B$, $P(A)$ is defined, is real & $P(A) \geq 0$. (Axiom of non-negativity)

(2) $P(S) = 1$ (Axiom of certainty)

(3) If $\{A_n\}$ is any finite or infinite sequence of disjoint events in B, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i)$$ (Axiom of additivity)

Some Theorems on Probability

① Probability of complementary event $\bar{A}$ of A is given by

$$P(\bar{A}) = 1 - P(A)$$

Proof – A & $\bar{A}$ are mutually disjoint events, so that

$$A \cup \bar{A} = S$$

$$\Rightarrow P(A \cup \bar{A}) = P(S)$$

Hence, from Axioms 2 & 3 of probability, we have –

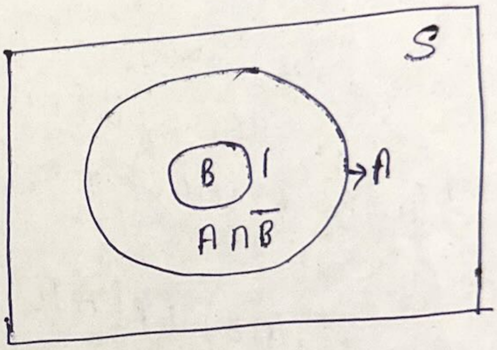
$$P(A) + P(\bar{A}) = P(S) = 1$$

$$\Rightarrow \boxed{P(\bar{A}) = 1 - P(A)}$$

② If $B \subset A$, then

(i) $P(A \cap \bar{B}) = P(A) - P(B)$

(ii) $P(B) \leq P(A)$



Proof: (i) When $B \subset A$, B & $A \cap \bar{B}$ are mutually exclusive events so that

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

$$\Rightarrow P(A) = P[B \cup (A \cap \bar{B})]$$

$$\Rightarrow P(A) = P(B) + P(A \cap \bar{B}) \quad \text{[By Axiom 3]}$$

$$\Rightarrow P(A \cap \bar{B}) = P(A) - P(B)$$

(ii) $P(A \cap \bar{B}) \geq 0$

$$\Rightarrow P(A) - P(B) \geq 0$$

$$\Rightarrow P(B) \leq P(A)$$

Hence $B \subset A$

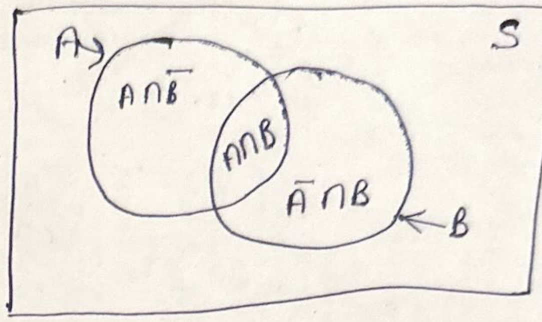
$$\Rightarrow P(B) \leq P(A)$$

(3) For any two events A & B, we have

(i) $P(\bar{A} \cap B) = P(B) - P(A \cap B)$

(ii) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

Proof (i) From the Venn Diagram, we get



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

where $\bar{A} \cap B$ and $A \cap B$ are disjoint events

Hence by Axiom (3), we get

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

$$\Rightarrow P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

(ii) Similarly, we have

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

where $(A \cap B)$ & $(A \cap \bar{B})$ are disjoint events. Hence, by Axiom 3, we get –

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

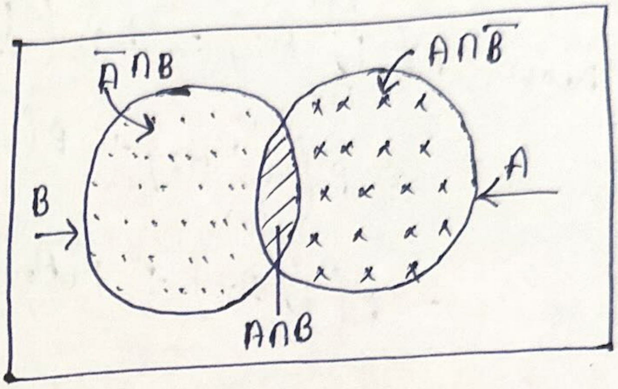
$$\Rightarrow P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

Addition Theorem of Probability

Theorem- If A & B are any two events (subset of sample space S) and are not disjoint, then

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

Proof From the Venn Diagram,



we have -

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

where A & $\bar{A} \cap B$ are mutually disjoint.

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

If the events A & B are mutually disjoint, then -

$$A \cap B = \phi$$
$$\Rightarrow P(A \cup B) = P(A) + P(B)$$

Note For three non-mutually exclusive events A, B & C, we have-

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

Extension of Addition ~~Theory~~ Theorem of Probability to n Events

Theorem- For n events $A_1, A_2, \ldots A_n$, we have -

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum\sum_{1 \le i < j \le n} P(A_i \cap A_j) + \sum\sum\sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n) \quad \text{--- (*)}$$

Proof- For two events A_1 & A_2, we have -

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \quad \text{--- (*)}$$

Hence (*) is true for n = 2

Let us now suppose that (*) is true for $n=r$, so that —

$$P\left(\bigcup_{i=1}^{r} A_i\right) = \sum_i P(A_i) - \sum_{1\le i<j\le r}\sum P(A_i \cap A_j) + \cdots + (-1)^{r-1} P(A_1 \cap A_2 \cap \cdots \cap A_r) \quad \text{—}(**)$$

$$\text{Now, } P\left(\bigcup_{i=1}^{r+1} A_i\right) = P\left\{\left(\bigcup_{i=1}^{r} A_i\right) \cup A_{r+1}\right\}$$

$$= P\left(\bigcup_{i=1}^{r} A_i\right) + P(A_{r+1}) - P\left\{\left(\bigcup_{i=1}^{r} A_i\right) \cap A_{r+1}\right\} \quad \{\text{Using } *\}$$

$$= P\left(\bigcup_{i=1}^{r} A_i\right) + P(A_{r+1}) - P\left\{\bigcup_{i=1}^{r} (A_i \cap A_{r+1})\right\}$$

{By Distributive Law}

$$= \sum_{i=1}^{r} P(A_i) - \sum_{1\le i<j\le r}\sum P(A_i \cap A_j) + \cdots + (-1)^{r-1} P(A_1 \cap A_2 \cap \cdots \cap A_r) + P(A_{r+1}) - P\left\{\bigcup_{i=1}^{r} (A_i \cap A_{r+1})\right\}$$

[From (**)]

$$= \sum_{i=1}^{r+1} P(A_i) - \sum_{1\le i<j\le r}\sum P(A_i \cap A_j) + \cdots + (-1)^{r-1} P(A_1 \cap A_2 \cap \cdots \cap A_r) - \left\{\sum_{i=1}^{r} P(A_i \cap A_{r+1}) - \sum_{1\le i<j\le r}\sum P(A_i \cap A_j \cap A_{r+1}) + \cdots + (-1)^{r-1} P(A_1 \cap A_2 \cap \cdots \cap A_r \cap A_{r+1})\right\}$$

[From(**)]

$$\Rightarrow P\left(\bigcup_{i=1}^{r+1} A_i\right) = \sum_{i=1}^{r+1} P(A_i) - \left\{\sum_{1\le i<j\le r}\sum P(A_i \cap A_j) + \sum_{i=1}^{r} P(A_i \cap A_{r+1})\right\} + \cdots + (-1)^r P\{(A_1 \cap A_2 \cap \cdots \cap A_{r+1}\}$$

$$= \sum_{i=1}^{r+1} P(A_i) - \sum_{1\le i<j\le (r+1)}\sum P(A_i \cap A_j) + \cdots + (-1)^r P(A_1 \cap A_2 \cap \cdots \cap A_{r+1})$$

Hence (X) is true for $n=r$, it is also true for $n=r+1$. But we have proved in (*) that (X) is true for $n=2$. Hence by the principle of mathematical induction, it follows that (X) is true for all positive ~~integers~~ integral values of n.

Eg - (1) A die is loaded (not all outcomes are equally likely) such that the probability that the number i shows up is K_i, $i = 1, 2, ---, 6$ where K is constant. Find

(a) the value of k.

(b) the probability that a number greater than 3 shows up.

Sol. (a) Here the sample space S has six outcomes $\{1, 2, ---, 6\}$. Hence, using axioms (2) & (3) we have

$$P(1) + P(2) + --- + P(6) = 1$$

Since $P(i) = K_i$, we have -

$$K.(1) + K.(2) + K.(3) + K.(4).\ K.(5) + K.(6) = 1$$

$$\Rightarrow K(1 + 2 + 3 + 4 + 5 + 6) = 1$$

$$\Rightarrow K.(21) = 1$$

$$\Rightarrow K = \frac{1}{21}$$

(b) Let A be the event that a number greater than 3 show up. Then the outcomes in A are $\{4, 5, 6\}$ & they are mutually exclusive. Therefore -

$$P(A) = P(4) + P(5) + P(6)$$

$$= \frac{4}{21} + \frac{5}{21} + \frac{6}{21} = \frac{15}{21}$$

(2) In a large university, the freshman profile for one year's fall admission says that 40% of the students were in the top 10% of their high school class, & that 65% are white, of whom 25% were in the top 10% of their high school class. What is the prob. that a freshman student selected randomly from this class either was in the top % of his or ~~high~~ her high school class or is white?

Sol. Let E_1 be the event that a person chosen at random was in the top 10% of his or her high school class, & let E_2 be the event that the student is white.

We are given —

$P(E_1) = 0.40$

$P(E_2) = 0.65$

& $P(E_1 \cap E_2) = 0.25$

Then the event that the student chosen is white or was in the top 10% of his or her high school class is represented by $E_1 \cup E_2$.

Thus —

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

$$= 0.40 + 0.65 - 0.25$$

$$= 0.80$$

Multiplication Theorem of Probability

Theorem :- For two events A & B,

$$\left.\begin{aligned} P(A\cap B) &= P(A).\ P(B/A), \quad P(A)>0 \\ &= P(B).\ P(A|B), \quad P(B)>0 \end{aligned}\right\} \quad — (☆)$$

where $P(B|A)$ represents conditional probability of occurence of B when the event A has already happened & $P(A|B)$ is the conditional probability of happening of A, given that B has already happened.

Proof. In the usual notations, we have –

$$P(A) = \frac{n(A)}{n(S)} \quad ; \quad P(B) = \frac{n(B)}{n(S)} \quad \& \quad P(A\cap B) = \frac{n(A\cap B)}{n(S)} \quad — (*)$$

For the conditional event $A|B$, the favourable outcomes must be one of the sample points of B, i.e., for the event $A|B$, the sample space is B & out of the $n(B)$ sample points, $n(A\cap B)$ pertain to the occurence of the event A. Hence –

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

Rewriting (*), we get

$$P(A\cap B) = \frac{n(B)}{n(S)} \times \frac{n(A\cap B)}{n(B)} = P(B).\ P(A|B) \quad — (**)$$

Similarly, we get from (*)

$$P(A\cap B) = \frac{n(A)}{n(S)} \times \frac{n(A\cap B)}{n(A)} = P(A).\ P(B|A) \quad — (***)$$

From (**) & (***) we get ☆

Thus, we have proved that " the prob. of the simultaneous occurence of two events A & B is equal to the product of the prob. of one of these events & the conditional probability of the other, given that the first one has occured.

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

Thus the conditional probabilities $P(B|A)$ & $P(A|B)$ are defined iff $P(A)\neq 0$ & $P(B)\neq 0$ respectively

Conditional Probability

It is the probability of an event occuring given that another event has already occured. If A & B are two events, the conditional probability of A given B (denoted as $P(A|B)$) is defined as :-

$$P(A|B) = \frac{n(A\cap B)}{n(B)}, \cancel{\text{where } P(B)>0}$$

- $n(A\cap B)$:- The ~~probability~~ event that both A & B occur
- $P(B)$:- The ~~probability~~ event that B occurs (the condition)

The formula essentially measures how likely A is, considering that B has already happened.

Remark (1) $P(A|B) = \dfrac{P(A\cap B)}{P(B)}$ & $P(B|A) = \dfrac{P(A\cap B)}{P(A)}$

Thus the conditional probabilities $P(B|A)$ & $P(A|B)$ are defined iff $P(A)\neq 0$ & $P(B)\neq 0$ respectively.

Bayes Theorem :-

Prior Probabilities → New Information → Application of Bayes' Theorem → Posterior Probabilities

In the discussion of conditional probability we indicated that revising when new information is obtained is an important phase of probability analysis. Often, we begin our analysis with initial or prior probability estimates for specific events of interests. Then, from sources such as a sample, a special report, a product test, & so on we obtain some additional information about the events. Given this new information, we update the prior probability values by calculating revised probabilities, referred to as posterior probabilities.

Extension of Multiplication Theorem of Probability to n Events

<u>Theorem</u> For n events $A_1, A_2, \ldots A_n$, we have

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \ldots P(A_n | A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \quad \text{—} (*)$$

where $P(A_i | A_j \cap A_k \cap \ldots \cap A_l)$ represents conditional probability of the event A_i given that the events $A_j, A_k, \ldots A_l$ have already happened.

<u>Proof</u> For two events A_1 & A_2, $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1)$

We have three events A_1, A_2 & A_3

$$P(A_1 \cap A_2 \cap A_3) = P\{A_1 \cap (A_2 \cap A_3)\}$$
$$= P(A_1) \cdot P\{(A_2 \cap A_3) | A_1\}$$
$$= P(A_1) \cdot P(A_2 | A_1) \cdot P\{(A_3 | (A_1 \cap A_2)\}$$

Thus we have find that $(*)$ is true for $n=2$ & $n=3$. Let us suppose that $(*)$ is true for $n=k$, so that

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \ldots P(A_k | A_1 \cap A_2 \cap \ldots \cap A_{k-1})$$

Now,
$$P[(A_1 \cap A_2 \cap \ldots \cap A_{k+1})] = P(A_1 \cap A_2 \cap \ldots \cap A_k) \times P[A_{k+1} | A_1 \cap A_2 \cap \ldots \cap A_k)$$
$$= P(A_1) \cdot P(A_2 | A_1) \ldots P(A_k | A_1 \cap A_2 \cap \ldots \cap A_{k-1}) \times P(A_{k+1} | A_1 \cap A_2 \cap \ldots \cap A_k)$$

Thus it is true for $n = k+1$.

Therefore by the principle of mathematical induction, it is true for all positive integral values of n.

Independence of events :- Two or more events are said to be independent if the happening or non-happening of any one of them, does not, in any way, affect the happening of them.

Def^n:- An event A is said to be independent of another event B, if the conditional probability of A given B i.e. $P(A|B)$ is equal to the unconditional probability of B i.e. if

$$\boxed{P(A|B) = P(A)} \quad ; \quad P(B) \neq 0$$

Similarly, $\boxed{P(B|A) = P(B)}$; $P(A) \neq 0$

Theorem - If the events A & B are such that $P(A) \neq 0$, $P(B) \neq 0$ & A is independent of B, then B is independent of A.

Proof. Since the event A is independent of B, we have

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

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

$$\Rightarrow P(A|B) = \frac{P(A).\cancel{P(B)}}{\cancel{P(B)}} \quad \{\because P(A \cap B) = P(A).P(B)\}$$

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

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

Bayes Theorem

If $E_1, E_2, \ldots E_n$ are mutually disjoint events with $P(E_i) \neq 0$ $(i = 1, 2, \ldots n)$ then for any arbitrary event A which is subset of $\bigcup_{i=1}^{n} E_i$ such that $P(A) > 0$, we have —

$$P(E_i \mid A) = \frac{P(E_i) \cdot P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i) \cdot P(A \mid E_i)}$$

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

$$A = A \cap \left(\bigcup_{i=1}^{n} E_i \right) = \bigcup_{i=1}^{n} (A \cap E_i)$$

(by distributive law)

Since $(A \cap E_i) \subset E_i$, $(i = 1, 2, \ldots n)$ are mutually disjoint events, we have by addition theorem of probability —

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

by multiplication theorem of probability.

Also we have, $P(A \cap E_i) = P(A) \cdot P(E_i \mid A)$

$$\Rightarrow P(E_i \mid A) = \frac{P(A \cap E_i)}{P(A)} = \frac{P(E_i) \cdot P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i) \, P(A \mid E_i)}$$

Remark 1. The probabilities $P(E_1), P(E_2), \ldots P(E_n)$ are termed as the 'a prior probabilities' because they exist before we gain any information from the experiment itself.

2. The probabilities $P(A \mid E_i)$; $i = 1, 2, \ldots n$ are called 'likelihoods' because they indicate how likely the event A under consideration is to occur, given each & every a prior probability.

3. The probability $P(E_i \mid A)$, $i = 1, 2, \ldots n$ are called 'posterior probabilities' because they are determined after the results of the experiment are known.

<u>The Tabular Approach</u> – The posterior probabilities, using Bayes theorem can be obtained conveniently in a tabular form, which involves the following steps –

(1) Prepare the following three columns –

Column 1 – The list of all mutually exclusive events E_i, $i = 1, 2, \dots n$ occur in the problem

Column 2 – The prior probabilities $P(E_i)$; $i = 1, 2, \dots n$ for the events

Column 3 – The conditional probabilities of the new information (A) given each event, viz $P(A|E_i)$; $i = 1, 2, \dots, n$

(2) In column 4, compute joint probabilities for each event using the formula

$$P(E_i \cap A) = P(E_i) \cdot P(A|E_i)$$

(3) Sum the joint probability column to find the probability of the new information, i.e. $P(A)$.

(4) In column 5, compute the posterior probabilities using the basic relationship of conditional probability

$$P(E_i|A) = \frac{P(E_i \cap A)}{P(A)} \; ; \; i = 1, 2, \dots n.$$

Que. From a vessel containing 3 white & 5 black balls, 4 balls are transferred into an empty vessel. From this vessel a ball is drawn & is found to be white. What is the probability that out of four balls transferred 3 are white & 1 is black?

Sol. Let us define the following events:

E_1 : Transfer of 0 white & 4 black balls

E_2 : Transfer of 1 white & 3 black balls

E_3 : Transfer of 2 white & 2 black balls

E_4 : Transfer of 3 white & 1 black balls

(Since the urn contains 3 white balls, more than 3 white balls cannot be transferred from the vessel)

E = Drawing of a white ball from the second vessel

Then $P(E_1) = \dfrac{{}^5C_4}{{}^8C_4} = \dfrac{1}{14}$

$$P(E_2) = \frac{{}^3C_1 \times {}^5C_3}{{}^8C_4} = \frac{3}{7}$$

$$P(E_3) = \frac{{}^3C_2 \times {}^5C_2}{{}^8C_4} = \frac{3}{7}$$

$$P(E_4) = \frac{{}^3C_3 \times {}^5C_1}{{}^8C_4} = \frac{1}{14}$$

Also $P(E|E_1) = 0$, $P(E|E_2) = \dfrac{1}{4}$, $P(E|E_3) = \dfrac{2}{4}$ & $P(E|E_4) = \dfrac{3}{4}$

Hence, by Bayes Theorem, the probability that out of four balls transferred, 3 are white & 1 black is —

$$P(E_4|E) = \frac{P(E_4).P(E|E_4)}{P(E_1).P(E|E_1) + P(E_2).P(E|E_2) + P(E_3).P(E|E_3) + P(E_4).P(E|E_4)}$$

$$= \frac{\frac{1}{14} \times \frac{3}{4}}{\frac{1}{14} \times 0 + \frac{3}{7} \times \frac{1}{4} + \frac{3}{7} \times \frac{1}{2} + \frac{1}{14} \times \frac{3}{4}} = \frac{3}{6+12+3} = \frac{1}{7}$$

$= 0.14$