F.Y.B.Sc. & B. A. : Statistic : Paper-II

DISCRETE PROBABILITY AND PROBABILITY DISTRIBUTIONS

Dr. (Mrs.) V. R. PRAYAG
Prof. P. G. DIXIT
Prof. P. S. KAPRE

NIRALI PRAKASHAN
the way to excellence

A Book Of

DISCRETE PROBABILITY AND PROBABILITY DISTRIBUTIONS

STATISTICS

Paper – II
For
F.Y.B.Sc./B.A.
As Per Pune University's Revised Syllabus
Effective from June 2013

Dr. (Mrs.) V. R. PRAYAG

M.Sc., Ph.D.
Associate Director (Statistics)
Cytel Statistical Software, Pune, India

Prof. P. G. DIXIT

M.Sc., M.Phil. (Statistics)
Head of Statistics Department,
Modern College, Pune - 5.

Prof. P. S. KAPRE

M.Sc., M.Phil. (Statistics)
Department of Statistics,
Modern College, Pune - 5.

Price ₹ 240.00

NIRALI PRAKASHAN
ADVANCEMENT OF KNOWLEDGE

N0927

F.Y.B.Sc./B.A. (STATISTICS) : DISCRETE PROBABILITY AND PROBABILITY DISTRIBUTIONS

ISBN 978-93-83073-78-8

Fifth Edition : June 2018

Published By : (–Ve) **Printed By :**

NIRALI PRAKASHAN YOGIRAJ PRINTERS AND BINDERS

Abhyudaya Pragati, 1312, Shivaji Nagar,
Off J.M. Road, PUNE – 411005
Tel - (020) 25512336/37/39, Fax - (020) 25511379
Email : niralipune@pragationline.com

Works: Sr. No. 10\1,Ghule Industrial Estate,
Nanded Village Road,
TAL-HAVELI, DIT-PUNE 411041.
Mobile - 9850046517, 9404233041

➢ DISTRIBUTION CENTRES

PUNE

Nirali Prakashan : 119, Budhwar Peth, Jogeshwari Mandir Lane, Pune 411002, Maharashtra
Tel : (020) 2445 2044, 66022708
Email : bookorder@pragationline.com, niralilocal@pragationline.com

Nirali Prakashan : S. No. 28/27, Dhyari, Near Pari Company, Pune 411041
Tel : (020) 24690204 Fax : (020) 24690316
Email : dhyari@pragationline.com, bookorder@pragationline.com

MUMBAI

Nirali Prakashan : 385, S.V.P. Road, Rasdhara Co-op. Hsg. Society Ltd.,
Girgaum, Mumbai 400004, Maharashtra
Tel : (022) 2385 6339 / 2386 9976, Fax : (022) 2386 9976
Email : niralimumbai@pragationline.com

➢ DISTRIBUTION BRANCHES

JALGAON

Nirali Prakashan : 34, V. V. Golani Market, Navi Peth, Jalgaon 425001,
Maharashtra, Tel : (0257) 222 0395, Mob : 94234 91860
Email : niralijalgoan@pragationline.com

KOLHAPUR

Nirali Prakashan : New Mahadvar Road, Kedar Plaza, 1st Floor Opp. IDBI Bank
Kolhapur 416 012, Maharashtra. Mob : 9850046155
Email : niralikolhapur@pragationline.com

NAGPUR

Pratibha Book Distributors : Above Maratha Mandir, Shop No. 3, First Floor,
Rani Jhanshi Square, Sitabuldi, Nagpur 440012, Maharashtra
Tel : (0712) 254 7129

DELHI

Nirali Prakashan : 4593/15, Basement, Agarwal Lane, Ansari Road, Daryaganj
Near Times of India Building, New Delhi 110002 Mob : 08505972553
Email : niralidelhi@pragationline.com

BANGALURU

Nirali Prakashan : Maitri Ground Floor, Jaya Apartments, No. 99, 6th Cross, 6th Main,
Malleswaram, Bangaluru 560 003, Karnataka
Mob : +91 9449043034
Email: niralibangalore@pragationline.com

niralipune@pragationline.com | www.pragationline.com

Also find us on www.facebook.com/niralibooks

Statistical Thinking will one day be necessary for effective citizenship as the ability to read and write

– H.G. Wells

The day has come now !

Preface …

We feel indeed very happy to present this text-book of 'Statistics Paper-II' - 'Discrete Probability and Probability Distributions' to the students of F.Y.B.Sc. and F.Y.B.A. The book is written according to the revised syllabus of University of Pune and University Grants commission with effect from June, 2013.

The main purpose of the book is to provide foundation as well as comprehensive background of 'Discrete Probability Theory' to beginners in simple and interesting manner. In order to make the contents of the book easier to comprehend, we have included a requisite number of illustrations, remarks, figures, diagrams etc. to elucidate statistical concepts. Application of Statistics in real life situations is emphasized through illustrative examples. Ample number of graded problems, theoretical as well as numerical, are provided at the end of each chapter along with hints and answers. The numerical problems will also be useful for the F.Y.B.Sc. students to prepare for Paper – III : Practicals. A list of practicals is given in the syllabus. Appendix A compiles some important mathematical results which are needed during the entire course. Values of individual terms of binomial probabilities are given in Appendix B. A specimen paper is set for student's self assessment.

MS-Excel commands to obtain probabilities is an additional feature of the book.

This book will also serve the purpose of reference book for F.Y.B.Sc. Computer Science, M.B.A., C.A., I.C.W.A., M.P.M. classes.

We are thankful to Mr. D. K. Furia and the staff of Nirali Prakashan for bringing out this book in short time. Mrs. Anagha Medhekar and Mr. Santosh Bare deserve special thanks for the work done with atmost care and sincerity. Finally, our families deserve special thanks for their support, encouragement and tolerance.

We request our colleagues, teaching Statistics to offer their criticism and suggestions, for further improvement of the book.

– Authors

Guru Pournima

22 July 2013

Syllabus ...

1. Review of Probability, Conditional Probability, Independence (6 L)

1.1 Experiments/Modes, ideas of deterministic and non-deterministic models, Random experiment, concept of statistical regularity.

1.2 Definitions of : (i) Sample space, (ii) Discrete sample space : finite and countably infinite, (iii) Event, (iv) Elementary event, (v) Complement of an event, (vi) Certain event, (vii) Impossible event.

1.3 Concept of occurrence of an event.

1.4 Algebra of events and its representation in set theory notation. Occurrence of following events :

(i) at least one of the given event,

(ii) none of the given events,

(iii) all of the given events,

(iv) mutually exclusive events,

(v) mutually exhaustive events,

(vi) exactly one event out of the given events.

1.5 Classical definition of probability and its limitations.

1.6 Probability model, probability of an event, equiprobable and non-equiprobable sample space.

1.7 Axiomatic definition of probability.

1.8 Definition of conditional probability of an event.

1.9 Definition of independence of two events $P(A \cap B) = P(A) \cdot P(B)$.

1.10 Pairwise independence and mutual independence for three events.

1.11 Multiplication theorem $P(A \cap B) = P(A) \cdot P(B|A)$.

Generalization of $P(A \cap B \cap C)$.

2. Bayes' Theorem (3 L)

2.1 Partition of the sample space

2.2 Proof of Bayes' theorem. Applications of Bayes' theorem in real life

3. Univariate Probability Distributions (Defined on Discrete Sample Space) (4 L)

3.1 Concept and definition of a discrete random variable

3.2 Probability mass function (p.m.f.) and cumulative distribution function (c.d.f.), $F(\cdot)$ of discrete random variable, properties of c.d.f.

3.3 Mode and median of a univariate discrete probability distribution

4. Mathematical Expectation (Univariate Random Variable) (8 L)

4.1 Definition of expectation (mean) of a random variable, expectation of a function of a random variable, m.g.f. and c.g.f. Properties of m.g.f. and c.g.f.

4.2 Definitions of variance, standard deviation (s.d.) and coefficient of variation (c.v.) of univariate probability distribution, effect of change of origin and scale on mean, variance and s.d.

4.3 Definition of raw, central and factorial raw moments of univariate probability distributions and their interrelations (without proof)

4.4 Coefficients of skewness and kurtosis based on moments.

5. Some Standard Discrete Probability Distributions – I (15 L)

5.1 Degenerate distribution (one point distribution), P(X = c) = 1, mean and variance

5.2 Uniform discrete distribution on integers 1 to n : p.m.f., c.d.f., mean, variance, real life situations, comments on mode and median.

5.3 Bernoulli' distribution : p.m.f., mean, variance.

5.4 Binomial distribution : p.m.f.

$$p(x) = \binom{n}{x} p^x q^{n-x}; \quad x = 0, 1, \ldots, n; \ 0 < p < 1, q = 1 - p$$
$$= 0 \qquad , \text{ otherwise}$$

Notation : X ~ B(n, p).

Recurrence relation for successive probabilities, computation of probabilities of different events, mode of the distribution, mean, variance, m.g.f. and c.g.f. moments, skewness (comments when p = 0.5, p > 0.5, p < 0.5). Situations where this distribution is applicable.

5.5 Hypergeometric distribution : p.m.f.

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}; \quad x = 0, 1, \ldots, \min(M, n)$$
$$= 0 \text{ otherwise}$$

Notation : X ~ H (N, M, n).

Computation of probability, situations where this distribution is applicable, binomial approximation to hypergeometric probabilities, mean and variance of the distribution.

6. Some Standard Discrete Probability Distribution – II (16 L)

6.1 Poisson distribution : Notation : X ~ P(m).

$$p(x) = \frac{e^{-m} m^x}{x!}, \quad x = 0, 1, 2, \ldots, m > 0$$

m.g.f. and c.g.f. Moments, mean, variance, skewness and kurtosis. Situations where this distribution is applicable.

6.2 Geometric distribution : Notation : X ~ G(p).

Geometric distribution on support (0, 1, 2, …,) with p.m.f. $p(x) = pq^x$.

Geometric distribution on support (1, 2, …) with p.m.f. $p(x) = pq^{x-1}$.

$0 < p < 1, q = 1 - p$.

Mean, variance, m.g.f. and c.g.f. Situations where this distribution is applicable.

7. Bivariate Discrete Probability Distribution **(6 L)**

7.1 Definition of two-dimensional discrete random variable, its joint p.m.f. and its distribution function and their properties, concept of identically distributed r.v.s.

7.2 Computation of probabilities of events in bivariate probability distribution.

7.3 Concepts of marginal and conditional probability distributions.

7.4 Independence of two discrete random variables based on joint and marginal p.m.f.s.

8. Mathematical Expectation (Bivariate Random Variable) **(14 L)**

8.1 Definition of raw and central moments, m.g.f., c.g.f.

8.2 Theorems on expectations of sum and product of two jointly distributed random variables

8.3 Conditional expectation

8.4 Definitions of conditional mean and conditional variance.

8.5 Definition of covariance, coefficient of correlation, independence and uncorrelatedness of two variables.

8.6 Variance of linear combination of variables Var (aX + bY).

8.7 Additive property for binomial and Poisson distributions.

8.8 Introduction of negative binomial distribution as sum of k i.i.d. geometric random variables. Statement of p.m.f., mean and variance.

8.9 Conditional distribution of X given (X + Y) for binomial and Poisson distributions.

•••

List of Practicals

1. Diagrammatic representation of statistical data: simple and subdivided bar diagrams, multiple bar diagram, percentage bar diagram, pie diagram.
2. Graphical representation of statistical data : histogram, frequency curve and ogive curves. Determination of mode and median graphically.
3. Use of random number tables to draw SRSWOR, SRSWR, stratified sample and systematic sample.
4. Computation of measures of central tendency and dispersion (ungrouped data). Use of an appropriate measure and interpretation of results and computation of partition values.
5. Computation of measures of central tendency and dispersion (grouped data). Use of an appropriate measure and interpretation of results and computation of partition values.
6. Measures of skewness and kurtosis, Box plot.
7. Scatter diagram, correlation coefficient (ungrouped data). Fitting of line of regression , residual plot.
8. Fitting of second degree curve, exponential curve of type $y = ab^x$, $y = ax^b$ compression, finding the best fit using residual s.s. and coefficient of determination.
9. Fitting of binomial distribution and computation of expected frequencies.
10. Fitting of Poisson distribution and computation of expected frequencies.
11. Problems on bivariate probability distribution.
12. Applications of binomial and hypergeometric distributions.
13. Applications of Poisson and geometric distributions.
14. Model sampling from Poisson and binomial distributions.
15. Index numbers.
16. Graphical and diagrammatic representation of statistical data using Excel.
17. Use of random numbers to draw SRSWOR, SRSWR, stratified sample and systematic sample using MS-Excel.
18. Computation of summary statistics using MS Excel.

19. Scatter diagram, correlation coefficient, fitting a line of regression, fitting of second degree curve using MS-Excel.
20. Project equivalent to 5 practicals.

Contents …

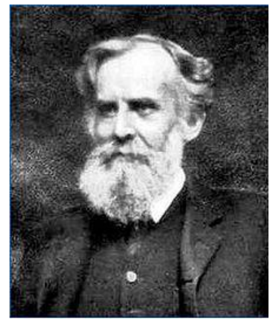
University Question Papers	**P.1 – P.10**

●●●

Chapter 1…

Sample Space and Events

John Venn (1834-1923), a British mathematician introduced the frequency interpretation of probability in his *'The Logic of Chance'* in 1866. Venn had other skills and interests too, including a rare skill in building machines. He used his skill to build a machine for bowling cricket balls which was so good that when the Australian Cricket team visited Cambridge in 1909, Venn's machine clean bowled one of its top stars four times.



John Venn

Contents …

Key Words :

Random experiment, Uncertainity, Sample space, Discrete sample space, Continuous sample space, Events.

Objectives :

- To distinguish between deterministic and non-deterministic experiments.
- To write the sample space and related events.
- To express the events in set notations.

1.1 Introduction

You are familiar with the word 'experiment'. You perform experiments in Physics, Chemistry or Biology.

For example, in Chemistry, you estimate the exact amount of alkali required to neutralize acid using titration method. Or in Physics, velocity (v) of a particle at time (t) can be determined using $v = u + at$, where u is the initial velocity and a is the acceleration. In Biological experiments, a type of diet is fed to animals and increase in their weights are recorded.

However, in Statistics, the word 'experiment' is used in a wider sense. It is not necessarily restricted to laboratory experiments.

Experiment : An experiment is virtually any operation that results in one or more outcomes.

For instance,

(i) Appearing for F.Y. B.Sc. examination is an experiment with possible outcomes as PASS or FAIL.

(ii) Casting a vote in the election is an experiment with outcomes; the party you voted for wins or looses.

(iii) Releasing a stone from hand is an experiment with the outcome that 'it will fall on the ground'.

(iv) Tossing a coin is an experiment with two possible outcomes, 'Head up' or 'Tail up'.

Head Tail

Fig. 1.1

(v) Rolling a six faced die; outcomes are 1, 2, 3, 4, 5, 6.

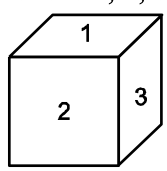


Fig. 1.2

Trial of an experiment : A trial of an experiment is nothing but performing the experiment once. So 'n' trials of an experiment means that the experiment is performed 'n' times, either sequentially or simultaneously. Experiments are classified into (i) Deterministic Experiments and (ii) Non-deterministic or Random Experiments.

(i) Deterministic Experiments : A deterministic experiment is an experiment for which the outcome is unique; hence certain. The outcome can therefore be predicted before performing the experiment. In other words, deterministic experiments are predictable phenomena. Following are some examples of deterministic experiments.

(i) Throwing a ball in the sky; Outcome : It falls down.

(ii) Cooling water below O^oC; Outcome : It will freeze.

(iii) Determining the pressure of a gas using Boyle's law.

$$\text{Outcome : } P = \frac{\text{Constant}}{\text{Volume of the gas}}$$

Observe that, in all the above experiments, the outcome is certain, unique and predictable.

Such experiments are known as deterministic experiments. All the deterministic experiments can be described by mathematical formulae. These mathematical formulae are called as deterministic models. For example, PV = constant, $S = \frac{1}{2} gt^2$ are deterministic models used in Physics. Since, there is no uncertainty in the result of the experiment, probability theory does not play any role.

(ii) Non-deterministic (Random) Experiments : A non-deterministic experiment or a random experiment is an experiment, for which there are more than one possible outcomes and the result of the experiment cannot be predicted in advance.

For instance,

(i) Sex of a new born baby is recorded. Outcomes : Male or Female.

(ii) Rolling a die. Outcomes : 1, 2, 3, 4, 5, 6.

(iii) Tossing of a coin. Outcomes : Head or Tail.

(iv) Blood group of a person recorded in a blood donation camp. Outcomes : O, A, B, AB.

In random experiments, 'chance' element plays a vital role in determining the outcome. No mathematical formula can describe these experiments. These are the experiments we are interested in.

Statistical Regularity :

Though, we cannot predict the outcome of a single trial of the experiment, we can get some knowledge about the pattern among the outcomes, when the experiment is performed repeatedly for a large number of times.

For example, we can't say whether 'head' or 'tail' will come-up when a fair coin is tossed. However, if we toss the coin say 1000 times, what will we expect ? *About* 500 times 'head' will turn up and *about* 500 times 'tail' will turn up. If we further increase the number of tossings, we expect that the proportion of getting 'head' should approach the value $\frac{1}{2}$.

Fig. 1.3 (a) exhibits the statistical regularity when a coin is tossed 100 times (b) 1000 times and the number of heads and tails noted.

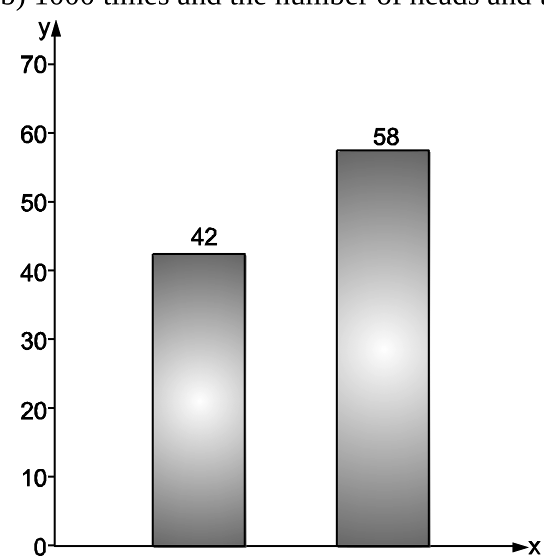


Fig. 1.3 (a)

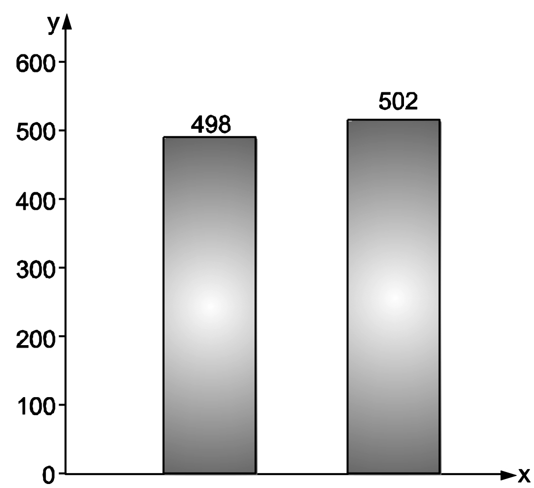


Fig. 1.3 (b)

Thus the proportion of getting head approaches 0.5 as the number of tosses increase.

Thus, though the outcome of any particular experiment may be uncertain, there exists a long term regularity. This is the basis for the development of random models. These models help in predicting the outcome of the non-deterministic experiment in probabilistic terms. Therefore, these models are also known as probabilistic models.

For example, using a non-deterministic i.e. random model, one can make the statement as 'Probability of death at the age of 80 years is 0.9.

These models are helpful in dealing with groups of cases. Random models are used in almost all branches of physical and social sciences. In fact, application of random models in different fields have given rise to several branches of Statistics such as Industrial Statistics, Econometrics, Statistical Ecology, Medical Statistics, Biometry, Acturial Science, Demography etc.

We cite here some situations where random models are applied.

(i) In industries, quality of the product manufactured is maintained.

(ii) Decision on whether to start an additional booking counter at railway station is taken by using random models in queueing theory.

(iii) A physicist studies the motion of particles emitted by a radioactive substance.

(iv) An economist, using random models can study the changes in price levels and construct index numbers.

(v) In a sociological survey, these models may be used to investigate the relationship between women literacy and success of family planning programmes.

(vi) In ecology, changes in the population of endangered species can be studied with a view to provide remedial measures.

Probability theory deals with non-deterministic or random models which describe and study the various phenomena happening in the world. We shall learn some of the simplest types of random models in this course in later chapters.

We are now in the position of learning some terminologies and concepts used in the theory of probability.

1.2 Sample Space

In the previous section, we talked about random experiments which have more than one possible outcome. In order to develop the probability theory, naturally, the first step is to *group* all possible outcomes of the experiment in a set. This set is called as 'Sample Space'.

Definition : Sample Space : The set of all possible distinct outcomes of an experiment is called a sample space. **(April 2015)**

Sample space is denoted by Ω or S. Thus, a sample space is nothing but the universal set concerned with the experiment.

For example, consider the experiment of tossing a coin. The corresponding sample space will be;

$$\Omega = \{\text{Head, Tail}\} = \{H, T\}$$