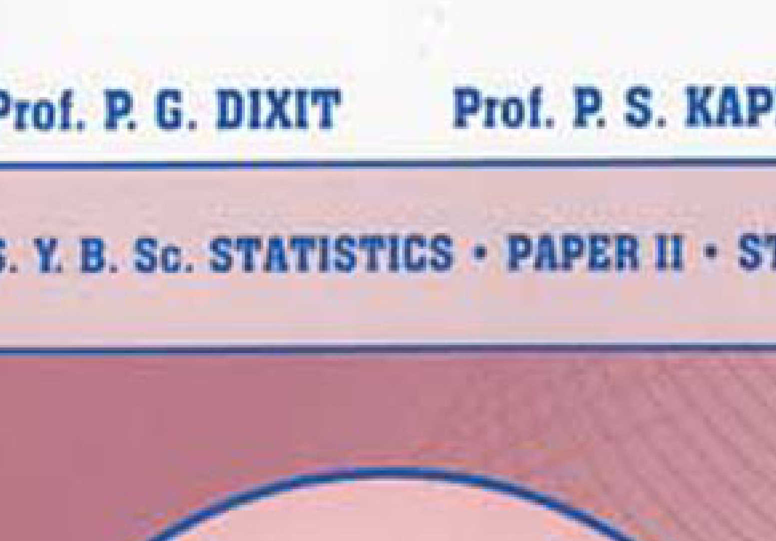
NEW SYLLABUS

A TEXT BOOK OF

CONTINUOUS PROBABILITY DISTRIBUTIONS-I

Prof. P. G. DIXIT Prof. P. S. KAPRE

S. Y. B. Sc. STATISTICS • PAPER II • ST-212 • SEMESTER I



NIRALI PRAKASHAN

A TEXT BOOK OF

STATISTICS

S.Y.B.Sc. & B.A. PAPER – II, SEMESTER – I

ST-212 : CONTINUOUS PROBABILITY DISTRIBUTIONS - I

As Per New Revised Syllabus of

Savitribai Phule Pune University, June 2014

Dr. P. G. DIXIT

M. Sc., M. Phil., Ph.D.
Vice Principal and
Head, Department of Statistics
Modern College, Shivajinagar,
Pune - 5.

Prof. P. S. KAPRE

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

Price ₹ 195.00

NIRALI PRAKASHAN
ADVANCEMENT OF KNOWLEDGE

N2267

(S.Y.B.Sc. & B.A.) STATISTICS P-II (SEMESTER-I) **ISBN 978-93-5164-016-5**

Fourth Edition : July 2018

Published By : **Polyplate** **Printed By :**

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

YOGIRAJ PRINTERS AND BINDERS
Works: Sr. No. 10\1,Ghule Industrial Estate,
Nanded Village Road,
Tal-Haveli, Dist-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

Preface ...

We feel indeed very happy to present the book of Statistics Paper – II, **'Continuous Probability Distributions-I'** to the students of S.Y.B.Sc./S.Y.B.A. It is written according to the revised syllabus of Savitribai Phule Pune University, Pune with effect from June 2014.

The overwhelming response for text book at F.Y.B.Sc./B.A. encouraged to write this book.

The main purpose of the book is to provide foundation as well as comprehensive background of continuous probability distributions to the 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.

The inclusion of MS-EXCEL commands and R commands is an additional feature of the book. It will help the students to compute the probabilities and fit the distributions.

Applications of Statistics in real life situations is emphasized through illustrative examples. Ample number of graded problems, theoretical as well as numerical, objective type questions are provided at the end of each chapter along with hints and answers. The numerical problems will also be useful for the S.Y.B.Sc. students to prepare for Paper – III : practicals. A list of practicals is given in the syllabus. The appendix compiles some important mathematical results which are needed during the entire course and normal probability paper. A specimen paper is set for students self assessment.

This book will also serve the purpose of reference book for M.B.A., C.A., I.C.W.A., T.Y.B.Com., F.Y.B.Sc. (Computer Science).

It will be also useful to prepare for statistics quiz competition, science quiz and competitive examinations.

We especially thank Prof. Pawagi V. R. for valuable help in bridging out this book.

We are also thankful to Mr. Dineshbhai Furia and the staff of Nirali Prakashan for bringing out this book in short time. Mr. M. P. Munde and Mrs. Angha Medhekar deserve special thanks for the co-operation they have extended to us. Finally, our family members deserve special thanks for the support, encouragement and tolerance.

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

AUTHORS

Syllabus ...

1. Continuous Univariate Distributions **(12 L)**

1.1 Continuous sample space : Definition, Illustrations, Continuous random variable : Definition, Probability density function (p.d.f.), Cumulative distribution function (c.d.f.), Properties of c.d.f. (without proof), Probabilities of events related to random variable.

1.2 Expectation of continuous r.v., Expectation of function of r.v. E [g(x)], Mean, Variance, Geometric mean, Harmonic mean, Raw and central moments, Skewness, Kurtosis.

1.3 Moment generating function (M.G.F.) : Definition and properties, Cumulant generating function (C.G.F.) : Definition, Properties.

1.4 Mode, Median, Quartiles.

1.5 Probability distribution of function of r.v. : Y = g(X) using (i) Jacobian of transformation for g(·) monotonic function and one-to-one, on to functions, (ii) Distribution function for $Y = X^2$, $Y = |X|$ etc., (iii) M.G.F. of g(X).

2. Continuous Bivariate Distributions **(12 L)**

2.1 Continuous bivariate random vector of variable (X, Y) : Joint p.d.f., Joint c.d.f., Properties (without proof), Probabilities of events related to r.v. (events in terms of regions bounded by regular curves, Circles, Straight lines). Marginal and conditional distributions.

2.2 Expectation (r.v.), Expectation of function of r.v. E [g(X, Y)], Joint moment, Cov (X, Y), Corr (X, Y), Conditional mean, Conditional variance, E [E(X|Y = y)] = E(X), Regression as a conditional expectation.

2.3 Independence of r.v. (X, Y) and its extension of k dimensional r.v. Theorems on expectation : (i) E (X + Y) = E (X) + E (Y), (ii) E (XY) = E (X) E (Y), if X and Y are independent, Generalization to k variables E (aX + bY = c), Var (aX + bY + c).

2.4 M.G.F. : $M_{X,Y}(t_1, t_2)$, Properties, M.G.F. of marginal distribution of r.v.s., Properties.

(i) $M_{X,Y}(t_1, t_2) = M_X(t_1, 0)\, M_Y(0, t_2)$, if X and Y are independent r.v.s.

(ii) $M_{X+Y}(t) = M_{X,Y}(t, t)$.

(iii) $M_{X+Y}(t) = M_X(t)\, M_Y(t)$ if X and Y are independent r.v.s.

2.5 Probability distribution of transformation of bivariate r.v. $U = \phi_1(X, Y)$, $V = \phi_2(X, Y)$.

3. Standard Univariate Continuous Distributions

3.1 Uniform of Rectangular Distribution **(03 L)**

Probability density function (p.d.f.) $f(x) = \begin{cases} \dfrac{1}{b-a} & ; \quad a \le x \le b \\ 0 & ; \quad \text{otherwise} \end{cases}$

Notation : $X \to U\,[a, b]$.

p.d.f., Sketch of p.d.f., c.d.f., Mean, Variance, Symmetry, Distribution of (i) $\dfrac{X-a}{b-a}$, (ii) $\dfrac{b-X}{b-a}$, (iii) Y = F(X), where F(X) is the c.d.f. of continuous r.v. X.

Application of the result of model sampling. (Distributions of X + Y, X − Y, XY and X/Y are not expected.)

3.2 **Normal Distribution** **(12 L)**

Probability density function (p.d.f.).

$$f(x) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} & ; \quad -\infty < x < \infty; -\infty < \mu < \infty; \sigma > 0 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Notation : $X \rightarrow N(\mu, \sigma^2)$.

p.d.f. curve, Identification of scale and location parameters, Nature of probability curve, Mean, Variance, M.G.F., C.G.F., Central moments, Cumulants, β_1, β_2, γ_1, γ_2, Median, Mode, Quartiles, Mean deviation, Additive property, Computations of normal probabilities using normal probability integral tables, Probability distribution of : (i) $\frac{X-\mu}{\sigma}$, Standard normal variable (S.N.V.), (ii) aX + b, (iii) aX + bY + c, (iv) X^2, where X and Y are independent normal variates. Probability distribution of $\bar{X}$, the mean of n i.i.d. N (μ, σ^2) r.vs. Normal probability plot, q-q plot to test normality. Model sampling from normal distribution using (i) Distribution function method and (ii) Box-Muller transformation as an application of simulation.

Statement and proof of central limit theorem (CLT) for i.i.d. r.vs with finite positive variance. (Proof should be using M.G.F.). Its illustration for Poisson and Binomial distributions.

3.3 **Exponential Distribution** **(04 L)**

Probability density function (p, d, f). $f(x) = \begin{cases} \alpha e^{-\alpha x} & ; \quad x \geq 0; \alpha > 0 \\ 0 & ; \quad \text{otherwise} \end{cases}$

Notation : $X \rightarrow E(\alpha)$.

Nature of p.d.f., Density curve, Interpretation of α as rate and $1/\alpha$ as mean, Mean, Variance, M.G.F., C.G.F., c.d.f., Graph of c.d.f., Lack of memory property, Median, Quartiles. Distribution of min (X, Y) with X, Y i.i.d. exponential r.v.s.

3.4 **Gamma Distribution** **(05 L)**

Probability density function (p.d.f.) $f(x) = \begin{cases} \dfrac{\alpha^2}{\Gamma\lambda}\, x^{\lambda-1}\, e^{-\alpha x} & ; \quad x \geq 0; \alpha > 0 \\ 0 & ; \quad \text{otherwise} \end{cases}$

Notation : $X \rightarrow G(\alpha, \lambda)$.

Nature of probability curve, Special cases : (i) $\alpha = 1$, (ii) $\lambda = 1$, M.G.F., C.G.F., Moment, Cumulants, β_1, β_2, γ_1, γ_2, Mode, Additive property. Distribution of sum of n i.i.d. exponential variables. Relation between distribution function of Poisson and Gamma variables.

❑❑❑

Contents ...

❑❑❑

Chapter **1**...

Continuous Univariate Probability Distributions

Pierre-Simon Laplace

Pierre-Simon, marquis de Laplace (23 March 1749 – 5 March 1827) was a French mathematician and astronomer whose work was pivotal to the development of mathematical astronomy and statistics. He summarized and extended the work of his predecessors in his five-volume *Mécanique Céleste* (Celestial Mechanics) (1799–1825). This work translated the geometric study of classical mechanics to one based on calculus, opening up a broader range of problems. In statistics, the Bayesian interpretation of probability was developed mainly by Laplace.

Laplace formulated Laplace's equation, and pioneered the Laplace transform which appears in many branches of mathematical physics, a field that he took a leading role in forming. The Laplacian differential operator, widely used in mathematics, is also named after him. He restated and developed the nebular hypothesis of the origin of the solar system and was one of the first scientists to postulate the existence of black holes and the notion of gravitational collapse.

Contents ...

Key Words :

Continuous sample space, continuous random variable, probability density function (p.d.f.), cumulative distribution function (c.d.f.), expectation, moments, mode, median, geometric mean, harmonic mean partition values, moment generating function (M.G.F.), cumulant generating function (C.G.F.), probability distribution of transformation of a r.v.

Objectives :

Conversion of a non-numeric continuous sample space to set of real numbers with the help of continuous r.v. To obtain summary statistics of a continuous r.v. To study the nature of r.v. such as central value, spread, symmetry. To develop tools such as M.G.F., C.G.F. for further study. To obtain probability of events related to continuous r.v. To obtain the probability distribution of a r.v., which is a function of r.v.

1.0 Introduction

We have seen that, there are three types of sample spaces (i) finite, (ii) countably infinite and (iii) continuous. In this chapter we discuss the random variables defined on continuous sample space.

A sample space which is finite or countably infinite is called as **denumerable** or **countable.** If the sample space is not countable then it is called **continuous.** In other words for a continuous sample space Ω we can not have one to one correspondence between Ω and set of natural numbers {1, 2, ... }. *Illustrations of uncountable sample space :*

(1) Suppose, weight of an oil bag having the capacity of 1 kg filled by an automatic filling machine is noted.

The sample space will be an interval in the neighbourhood of 1 kg such as Ω = (0.980, 1.005).

(2) Suppose in an experiment, life of an electronic component in hours is recorded.

The sample space in this case may be an interval as a part of R^+ such as Ω = (0, 5000)

Note : A continuous sample space is a subset of real line.

1.1 Continuous Random Variable

In general, we define a random variable X (ω) as a real valued function on domain Ω. If the range set of X (ω) is continuous the r.v. is *continuous.* The range set will be a subset of real line.

Illustrations of continuous r.v. :

(1) Weight of a person in kg.

(2) Consumption of electricity of a town in a specific month.

(3) Daily rainfall in cm. at a particular place.

(4) Instrumental error (measured in suitable units) in the measurement.

(5) Life in hours of an electrical component.

Note : The distinction between continuous random variable and discrete random variable is as follows :

(1) A continuous r.v. takes all possible values in a range set. The set is in the form of interval. On the other hand discrete r.v. takes only specific or isolated values.

(2) Since, a continuous r.v. takes uncountably infinite values no probability mass can be attached to a particular value of r.v. X. Therefore, P (X = x) = 0 for all x. However in case of a discrete r.v., probability mass is attached to individual values taken by r.v.

In case of continuous r.v. probability is attached to an interval which is a subset of R.

1.2 Continuous Probability Distribution

In case of discrete r.v. using p.m.f. we get probability distribution of r.v., however in case of continuous r.v. probability mass is not attached to any particular value. It is attached to an interval. The probability attached to an interval depends upon its location.

For example, P (a < X < b) varies for the different values of a and b. In other words, it will not be uniform. In order to obtain the probability mass associated with any interval, we need to take into account the concept of probability density.

A function f (x) which is to be treated as a probability density function, should be a non-negative and continuous function of x. The probability that a variable X takes values in a small interval $\left(x - \frac{\delta x}{2}, x + \frac{\delta x}{2}\right)$ will be the product of length of interval and the value of density function f (x) at the centre of interval.

$$\therefore \quad P\left(x - \frac{\delta x}{2} < X < x + \frac{\delta x}{2}\right) \approx f(x) \, . \, \delta x$$

Note : Here, we assume that the probability density is constant over the interval $\left(x - \frac{\delta x}{2}, x + \frac{\delta x}{2}\right)$. This assumption will not be valid for large interval. To overcome this difficulty we integrate f (x) w.r.t. x over the given interval. (In case of discrete r.v. we take summation). Thus, $P(a < X < b) = \int_a^b f(x)\, dx$.

(2) The above probability is a definite integral, hence geometrically it is the area under curve y = f (x) bounded by X axis and the ordinates at a and b.

Precise definition of probability density function is as follows.

Definition : A real valued function f (x) is called as a probability density function (p.d.f.) of a continuous random variable X if,

(i) $f(x) \geq 0 \,;\; -\infty < x < \infty$ (ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

Note : (1) Since, probability is associated with any individual value is zero, $P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b) = \int_a^b f(x)\,dx$ = Area under the curve f (x) bounded between X axis and the ordinates at a and b.

It is shown in the following figure by shaded region.

(ii) $\int_{-\infty}^{\infty} f(x)\,dx$, can be interpreted as $P(-\infty < X < \infty)$. It also represents the total area under the density curve bounded by X-axis. Thus, the total area under the curve $= P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1$.

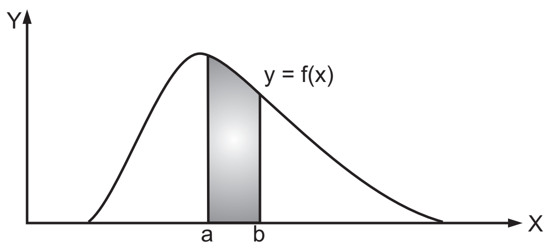


Fig. 1.1

(iii) If $A \subset R$, then $P(X \in A) = \int_A f(x)\,dx$.

(iv) f(x) need not be less than 1.

Example : Let X be r.v. with following p.d.f.

$$f(x) = 2 \quad , \quad 0 \le x \le 1/2$$
$$= 0 \quad , \quad \text{otherwise}$$

It is a p.d.f.,

since $f(x) > 0$ for $x \in [0, 1/2]$ and $\int_0^{1/2} f(x)\,dx = 1$.

Example 1.1 : Verify which of the following functions are p.d.f.s

(i) $f(x) = 3x^2 \quad ; \quad 0 \le x \le 1$
$\quad = 0 \quad ; \quad \text{otherwise}$

(ii) $f(x) = 2e^{-x} \quad ; \quad x \ge 0$
$\quad = 0 \quad ; \quad \text{otherwise}$

Solution : To verity whether f (x) is p.d.f. we need to verify the following two conditions (a) $f(x) \ge 0 \ \forall \ x$ and (b) $\int_{-\infty}^{\infty} f(x)\,dx = 1$,

(i) $f(x) = 3x^2 \ge 0 \ \forall \ x$ and $\int_{-\infty}^{\infty} 3x^2\,dx = \int_0^1 3x^2\,dx = [x^3]_0^1 = 1$

Therefore, f (x) is a p.d.f.