

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/268870344>

Probability and Statistical Applications – Distributions

Chapter · November 2014

DOI: 10.13140/2.1.4295.3284

CITATIONS

0

READS

81,227

1 author:



Nimmagadda Venkata Nagendram

Kakinada Institute of Technology & science KITS

129 PUBLICATIONS 682 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



I am doing research work on near-fields and extensions simultaneously non-commutativity of the same area under algebra of mathematics advanced [View project](#)



Description of the project near-field space with right sub near-field space over a near-field [View project](#)

Probability and Statistical Applications

By Dr. N V Nagendram

UNIT – I

Probability Theory: Sample spaces Events & Probability; Discrete Probability; Union, intersection and compliments of Events; Conditional Probability; Baye's Theorem .

UNIT – II

Random Variables and Distribution; Random variables Discrete Probability Distributions, continuous probability distribution, Mathematical Expectation or Expectation Binomial, Poisson, Normal, Sampling distribution; Populations and samples, sums and differences. Central limit Elements. Theorem and related applications.

UNIT – III

Estimation – Point estimation, interval estimation, Bayesian estimation, Text of hypothesis, one tail, two tail test, test of Hypothesis concerning means. Test of Hypothesis concerning proportions, F-test, goodness of fit.

UNIT – IV

Linear correlation coefficient Linear regression; Non-linear regression least square fit; Polynomial and curve fittings.

UNIT – V

Queing theory – Markov Chains – Introduction to Queing systems- Elements of a Queuing model – Exponential distribution – Pure birth and death models. Generalized Poisson Queuing model – specialized Poisson Queues.

Text Book: Probability and Statistics By T K V Iyengar S chand, 3rd Edition, 2011.

References:

1. Higher engg. Mathematics by B V Ramana, 2009 Edition.
2. Fundamentals of Mathematical Statistics by S C Gupta & V K Kapoor Sultan Chand & Sons, New Delhi 2009.
3. Probability & Statistics by Schaum outline series, Lipschutz Seymour, TMH, New Delhi 3rd Edition 2009.
4. Probability & Statistics by Miller and freaud, Prentice Hall India, Delhi 7th Edition 2009.

Planned Topics

UNIT - II

1. Random Variables - Introduction
2. Discrete and Continuous Random Variables, Distribution Function
3. Mathematical Expectations, Examples
4. Problems
5. Binomial Distribution – Mean, Variance, Mode
6. Problems
7. Poisson Distribution – Mean, Variance, Mode
8. Tutorial
9. Normal Distribution – Properties, Mean, Variance
10. Area under standard normal curve, Problems
11. Problems
12. Sampling distribution of mean
13. Sampling distribution of proportion
14. Sampling distribution of sum and differences
15. Central limit Theorem and Applications
16. Tutorial

Chapter 2

Lecture 1

By Dr. N V Nagendram

Probability Distributions

Introduction: In random experiments, we are interested in the numerical outcomes i.e., numbers associated with the outcomes of the experiment. For example, when 50 coins are tossed, we ask for the number of heads. Whenever we associate a real number with each outcome of trial, we are dealing with a function whose range is the set of real numbers we ask for such a function is called a random variable (r. v.) chance variable, stochastic variable or simply a variable.

Definition: Quantities which vary with some probability are called random variables.

Definition: By a random variable we mean a real number associated with the outcomes of a random experiment.

Example: Suppose two coins are tossed simultaneously then the sample space is $S = \{HH, HT, TH, TT\}$. Let X denote the number of heads, then if $X = 0$ then the outcome is $\{TT\}$ and $P(X = 0) = \frac{1}{4}$.

If X takes the value 1, then outcome is $\{HT, TH\}$ and $P(X = 1) = \frac{2}{4}$. Next if X takes the value 2 then the outcome is $\{HH\}$ and $P(X = 2) = \frac{1}{4}$. The probability distribution of this random variable X is given by the following table:

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

Example: out of 24 mangoes 6 are rotten, 2 mangoes are drawn. Obtain the probability distribution of the number of rotten mangoes that can be drawn:

Let X denote the number of rotten mangoes drawn then X can take values 0, 1, 2.

$$P(X = 0) = \frac{{}^{18}C_2}{{}^{24}C_2} = \frac{18 \times 17}{24 \times 23} = \frac{51}{92} ; P(X = 1) = \frac{{}^{18}C_1 \times {}^6C_1}{{}^{24}C_2} = \frac{18 \times 6}{24 \times 23} = \frac{9}{23} \text{ and}$$

$$P(X = 2) = \frac{{}^6C_2}{{}^{24}C_2} = \frac{6 \times 5}{24 \times 23} = \frac{5}{92}$$

$X = x$	0	1	2	Total
$P(X = x)$	$\frac{51}{92}$	$\frac{9}{23}$	$\frac{5}{92}$	1

Types of Random Variables:

There are two types of random variables:

- (i) Discrete random variables (ii) Continuous random variables

Distribution function: Let X be a one-dimensional random variable. The function F defined for all x , by the equation $F_X(x) = P(X \leq x)$ is called the cumulative distribution function of X .

Note: 1. We write c. d. f. For cumulative distribution function. Only d. f is written instead of c. d. f.

Note 2: suffix X in F_X is used to emphasize the fact that the distribution function is associated with the particular variate X . when the particular underlying variate is clear from the context, we shall simply write $F(x)$ instead of $F_X(x)$.

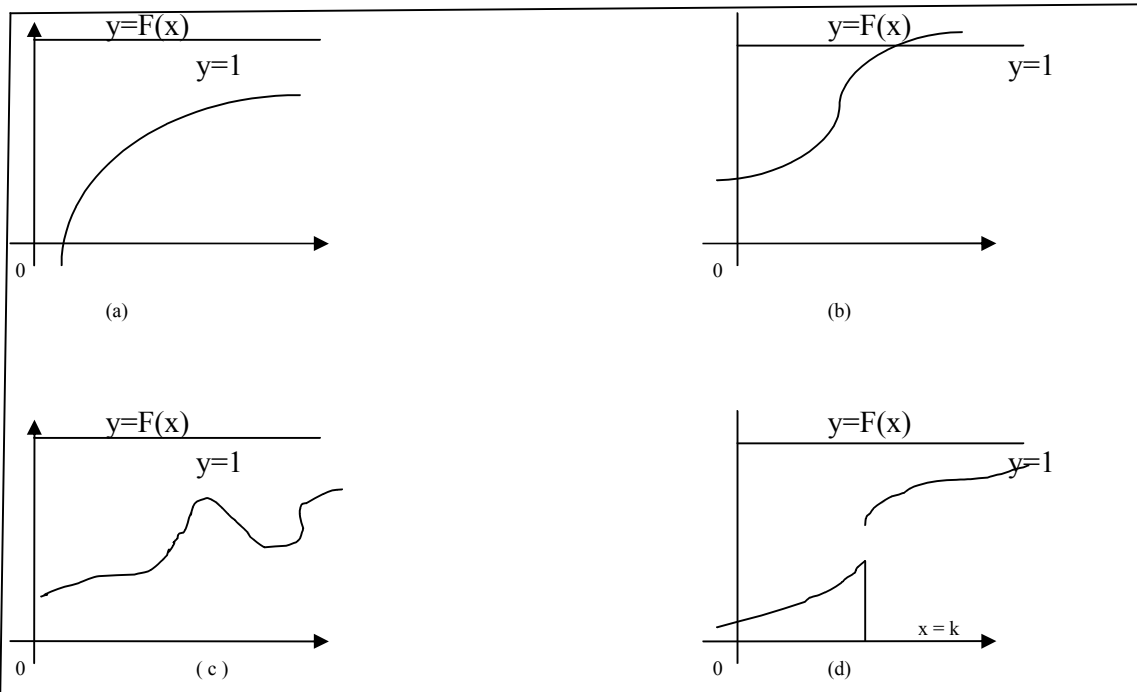
Note 3: tail events let ' x ' be any real number then the events $|X < x|$ and $|X > x|$. $|X \geq x|$ are called tail events. For distinction, we may label them open, closed, upper and lower tails. Often, simple r.v.'s are expanded as linear combination of tail events.

Some Properties of a c. d. f.:

1. $P\{a < X \leq b\} = F(b) - F(a)$ Interval property
2. $0 \leq F(x) \leq 1, \forall x \in R$ Boundedness property
3. F is non-decreasing i.e., if $x \leq y$, then $F(x) \leq F(y)$ Monotone increasing property
4. $F(-\infty) = 0, F(+\infty) = 1$ i.e.,
 $\lim_{n \rightarrow -\infty} F(x_n) = 0; \lim_{n \rightarrow \infty} F(x_n) = 1$ Limits property
5. F is continuous from the right at each point Right continuous property
i.e., $F(a+) = F(a)$
 $F(a+) - F(a-) = P(X = a)$ Jump discontinuity

Conditions (3),(4) and (5) are necessary as well as sufficient for F to be c.d.f. on R .

Problem 1: Give reasons why each of the graphs of F given below does not represent a distribution function.



Solution: (a) $F(x) < 0$ – ve for some x (b) $F(x) > 1$ for some x
 (c) F is non-decreasing i.e., some times F is decreasing also
 (d) F is not right continuous at $x = k$ infact it is left continuous.

Definition: Discrete Random variables:

Quantities which are capable of taking only integral values are called discrete random variables.

Example: The number of children in a family of a colony.

Example: The number of rooms in the houses of a township.

Probability mass function: Probability distribution

Definition: Let X be a discrete random variable taking value x , $x = 0, 1, 2, 3, \dots$ then $P(X = x)$ is called the probability mass function of X and it satisfies the following (i) $P(X = x) \geq 0$

(ii) $\sum_{x=0}^{\infty} P(X = x) = 1$

Definition: Discrete distribution function:

A r. v. X is said to be discrete, if there exist a countable number of points $x_1, x_2, x_3, \dots$ and

number $p(x_i) \geq 0$, $\sum_{x=1}^{\infty} p(x_i) = 1$ such that $F(x) = \sum_{x_i \leq x} p(x_i)$.

Definition: Finite equiprobable space (Uniform space)

A finite equiprobable space is finite probability distribution where each sample point $x_1, x_2, x_3, \dots, x_n$ has the same probability for all i

i.e., $P(X = x_i) = p_i = a$ constant for all i and $\sum_{x_i \leq x} p_i = 1$.

Chapter 2 Probability Distributions

Tutorial 1

By Dr. N V Nagendram

Problem 1: Show that the average of the deviations of a variate about its mean is zero and sum of the squared deviations is minimum when they are taken about the mean.

$$[\text{Ans. } A = \bar{X} = A = \frac{\sum f_i x_i}{\sum f_i}]$$

Problem 2: A random variable X has the following probability distribution:

x	0	1	2	3	4	5	6	7	8
P(x)	k	3k	5k	7k	9k	11k	13k	15k	17k

- (a) Determine the value of k
- (b) find $P(X < 4)$, $P(X \geq 5)$, $P(0 < x < 4)$
- (c) find the c.d.f.
- (d) find the smallest value of x for which $P(X \leq x) < 0.5$

$$[\text{Ans. } k = \frac{1}{81}, \frac{16}{81}, \frac{15}{81}, \frac{15}{81}, \text{ and } F(x) = 0.5, F(5) = 0.44, F(6) = 0.61]$$

Problem 3: Given the discrete random variable X has mass function.

$$P(X=x) = \begin{cases} \frac{x}{6} & \text{where } x = 1, 2, 3 \\ 0 & \text{elsewhere} \end{cases} \quad \text{Describe and graph its cumulative distribution function,}$$

F(x).

$$[\text{Ans. } 1/6, 2/6 \text{ and } 3/6]$$

Problem 4: If $f(x) = \frac{1}{4}$, $x = 0, 1, 2, 3$ is a probability mass function, find F(x), the cumulative

distribution function and sketch its graph.

$$[\text{Ans. } F(x) = \sum_{x=0}^3 f(x) = 1 \quad \forall 3 \leq x]$$

Problem 5: A random variables X has the following probability function.

X = x	0	1	2	3	4	5	6	7
P(X=x)	0	k	2k	2k	3k	k ²	2k ²	7k ² +1

- (i) Find the value of k
- (ii) $P(X \leq 5)$, $P(X > 5)$, $P(X < 6)$, $P(X \geq 6)$
- (iii) $P(0 < X < 6)$, $P(0 < X < 5)$

$$[\text{Ans. (i) } k = -1, \frac{1}{10}; \text{ (ii) } 0.81, 0.19 \text{ (iii) } 0.81, 0.8]$$

Problem 1: A random variable X takes values 0, 1, 2, 3, 4, . . . with probability proportional to $(x+1)\left(\frac{1}{5}\right)^x$. find $P(X \leq 5)$ [Ans. 0.9997]

Problems 2: A random variable X has the following probability function:

Value of x	-2	-1	0	1	2	3
P(x)	0.1	k	0.2	2k	0.3	K

(i) Find the value of k, and calculate mean and variance.

(ii) Construct the c.d.f. F(x) and draw its graph.

[Ans. (i). 0.1, 0.8 and 2.16 (ii). F(x) = 0.1, 0.2, 0.4, 0.6, 0.9, 1.0]

Problem 3: If a variable X assumes three values 0, 1, 2 with probabilities $\frac{1}{3}, \frac{1}{6}, \frac{1}{2}$ respectively, find the c.d.f. of X and show that $P(X \leq 1) = \frac{1}{2}$.

Problem 4: A random variable X assumes the values -3, -2, -1, 0, 1, 2, 3 such that $P(X > 0) = P(X = 0)$; $P(X < -3) = P(X = -2) = P(X = -1)$; $P(X = 1) = P(X = 2) = P(X = 3)$ write down the distribution of X and show that $P(X \leq 3) = \frac{2}{3}$.

Definition: Expectation: The behaviour of r.v. either discrete or continuous is completely characterized by the distribution function F(x) or density f(x) [P(x_i) in discrete case]. instead of a function, a more compact description can be made by a single numbers such as mean (expectation), median and mode known as measures of central tendency of the r.v. X.

Expectation or median or expected value of a r.v. X denoted by E[X] or μ , is defined as

$$E[X] \equiv \begin{cases} \sum_i x_i f(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{if } X \text{ is continuous} \end{cases}$$

Definition: Variance: variance characterizes the variability in the distributions, since two distributions with same mean can still have different dispersion of data about their means,

Variance of r.v. X is $\sigma^2 \equiv E[(X - \mu)^2] \equiv \sum (X - \mu)^2 f(x)$ for X discrete

$$\sigma^2 \equiv E[(X - \mu)^2] \equiv \int_{-\infty}^{\infty} (X - \mu)^2 f(x) dx \text{ for } X \text{ is continuous.}$$

Definition: Standard Deviation: standard deviation denoted by σ (S.D.) is the positive square root of variance.

$$\begin{aligned} \sigma^2 \equiv E[(X - \mu)^2] &\equiv \sum (X - \mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x) \\ &= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x) \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 \cdot 1 = E(X^2) - \mu^2 \end{aligned}$$

since $\mu = \sum x f(x)$, $\sum f(x) = 1$.

Probability Density function (p. d. f)

Let X be a continuous random variable taking value x , $a \leq x \leq b$ then $f(x) = P(X = x)$ is defined as the p. d. f. of X and satisfies the following

$$(i) f(x) \geq 0 \quad (ii) \int_a^b f(x) dx = 1.$$

- Note:**
1. For a continuous variate, point probabilities are zero.
 2. Area under the probability curve $y = f(x)$ is unity; the fact $f(x) \geq 0$ implies the graph $f(x)$ is above x -axis.
 3. Area under the probability curve $y = f(x)$ bounded by $x = a$, $x = b$ is simply $P(a \leq x \leq b)$.
 4. Relation between p. d. f. and c. d. f.: The density f and c. d. f. F are always

$$\text{connected by (a) } F(x) = \int_{-\infty}^x f(x) dx \quad \forall x \in R \quad (b) \frac{d}{dx}[F(x)] = f(x) \quad \forall x \in R.$$

Moments: If the range of the probability density function is from $-\infty$ to ∞ , the r^{th} moment about origin is defined as $\mu'_r = \int_{-\infty}^{\infty} x^r f(x) dx$.

The r^{th} moment about any arbitrary origin 'a' is $\mu'_r = \int_{-\infty}^{\infty} (x-a)^r f(x) dx$

The mean is given by (taking moment about $x = 0$) $\mu'_1 = \int_{-\infty}^{\infty} x^1 f(x) dx$

The variance σ^2 is given by $\sigma^2 = \mu'_2 - \mu_1'^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - \left[\int_{-\infty}^{\infty} x f(x) dx \right]^2$

Jointly Distributed Random Variables:

Introduction: When the outcome of a random experiment can be characterized in more than one way, the probability density is a function of more than one variate.

Example: When a card is drawn from an ordinary deck, it may be characterized according to its suit in some order viz., say clubs, diamonds, hearts and spades and Y be a variate that assumes the values 1, 2, 3, . . . , 13 which correspond to the denominations: Ace, 2, 3, . . . , 10, J, Q, K. Then (X, Y) is a 2 – dimensional variate. The probability of drawing a particular card will be denoted by $f(x, y)$ and if each card is equi-probable of being drawn, the density of

$$(X, Y) \text{ is } f(x, y) = \frac{1}{52} \quad \forall 1 \leq x \leq 4 \\ \forall 1 \leq y \leq 13$$

Trials whose outcomes can be characterized by two (three) variates give rise to bivariate (tri-variate) distributions etc. Extensions to n-variate distributions are fairly straight forward.

We study about the types of Distribution Functions as mentioned below:

- **Joint distribution function - properties**
- **Joint discrete distribution function**
- **Individual or Marginal Probability functions**
- **Bivariate Probability distribution function**
- **Conditional probability functions**
- **Some properties of Joint density**
- **Individual or Marginal distribution functions**
- **Conditional distribution function**

Joint distribution Function and its properties:

Let (X, Y) be a random vector or random variable on the probability space. The joint c. d. f. of X and Y is denoted by $F_{X,Y}$ and is defined by $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$, $x, y \in \mathbb{R}$.

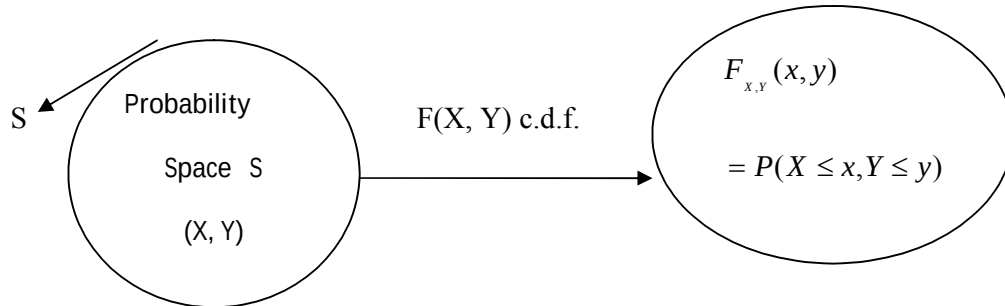


Fig.

A joint c. d. f. of two variates has the following properties:

1. **Non-negativity and Boundedness:**

$$0 \leq F_{X,Y}(x, y) \leq 1, \quad \text{for every } x, y \in \mathbb{R}.$$

2. **Monotonicity:** the c. d. f. “F” is monotonically non-decreasing function in each of the individual variables, i.e., (i) $F(a, y_2) \geq F(a, y_1)$, if $y_2 \geq y_1$ (ii) $F(x_2, b) \geq F(x_1, b)$, if $x_2 \geq x_1$.

3. **Rectangle rule:** Let a, b, c, d be any real numbers with $a < b$ and $c < d$.

$$\text{Then, } P(a < X \leq b, c < Y \leq d) = F(b, d) + F(a, c) - F(b, c) - F(a, d).$$

4. **Individual limits:** (i) $\lim_{x \rightarrow -\infty} F(x, y) = F(-\infty, y) = 0$ as $x \rightarrow -\infty$; (ii) $\lim_{y \rightarrow -\infty} F(x, y) = F(x, -\infty) = 0$

5. **Double limit:** $\lim_{x \rightarrow \infty, y \rightarrow \infty} F(x, y) = 1$ i.e., $F(\infty, \infty) = 1$ formally.

Note: we don't claim $F(\infty, y) = 1$ or $F(x, \infty) = 1$.

6. **Individual continuity:** F is continuous from the right in each of its individual variables.

$$\text{i.e., (i) } \lim_{x \rightarrow a^+} F(x, y) = F(a, y), \text{ (ii) } \lim_{y \rightarrow b^+} F(x, y) = F(x, b)$$

7. If the density function $f(x, y)$ is continuous at (x, y) , then $\frac{\partial^2 F}{\partial x \partial y} = f(x, y)$

Joint discrete Distribution Function:

Definition: The joint c. d. f. of X and Y is said to be discrete if there exists a non-negative function P such that P vanishes everywhere except a finite or countably infinite number of points in the plane and at such points (x, y) so that $P(x, y) = P(X = x, Y = y)$, for all $x, y \in \mathbb{R}$.

Definition: Let X and Y have a joint discrete distribution. A function P which does not vanish on the set $\{(x_i, y_j) \text{ such that } i, j = 1, 2, 3, \dots\}$ and satisfies the following properties:

(i) $P(x_i, y_j) \geq 0$ for all $i, j = 1, 2, 3, \dots$ and (ii) $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(x_i, y_j) = 1$ is called joint probability (mass) function of X and Y or simply the joint probability function.

Individual and Marginal Probability Functions:

Let X and Y be two jointly distributed variables with joint discrete density $P(x, y)$, the individual variates X and Y themselves are random variables.

The individual distributions of X and Y are called marginal distributions of X and Y

(i) The Marginal probability function for X is denoted by $P_X(x)$ or $P(x)$ and is given by

$$P(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y P(x, y)$$

(ii) The marginal probability function for Y is denoted by $P_Y(y)$ and is given by

$$P(y) = P(Y = y) = \sum_x P(X = x, Y = y) = \sum_x P(x, y).$$

Note: It is convenient to display the probability function of a bivariate distribution in a rectangular array, in which the row totals and column totals provide the marginal probability functions of X and Y respectively.

$\begin{matrix} x \backslash y \\ y \end{matrix}$	y_1	y_2	y_3	$\dots$	y_j	$\dots$	y_m	$\dots$	$P(X = x_i)$
x_1	P_{11}	P_{12}	P_{13}	$\dots$	P_{1j}	$\dots$	P_{1m}	$\dots$	$P(x_1)$
x_2	P_{21}	P_{22}	P_{23}	$\dots$	P_{2j}	$\dots$	P_{2m}	$\dots$	$P(x_2)$
x_3	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_i	P_{i1}	P_{i2}	P_{i3}	$\dots$	P_{ij}	$\dots$	P_{im}	$\dots$	$P(x_i)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	P_{n1}	P_{n2}	P_{n3}	$\dots$	P_{nj}	$\dots$	P_{nm}	$\dots$	$P(x_n)$
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$P(Y = y_i)$	$P(y_1)$	$P(y_2)$	$P(y_3)$	$\dots$	$P(y_j)$	$\dots$	$P(y_m)$	$\dots$	1

We have here, $P_{ij} = P(X = x_i, Y = y_j)$; $P(x_j) = \sum_i P_{ij}$ and $P(y_j) = \sum_i P_{ij}$.

Conditional Probability functions (cond. P. f.):

Let X and Y have joint discrete distribution with associated probability function P. Let the possible values of X be $\{x_1, x_2, x_3, \dots, x_i, \dots\}$ and those of Y be $\{y_1, y_2, y_3, \dots, y_j, \dots\}$ respectively.

The conditional probability function of X, given $Y = y_j$ denoted by $P_{X/y_j} \left(\frac{x_i}{y_j} \right)$ is defined

$$\text{by } P_{X/y_j} \left(\frac{x_i}{y_j} \right) = P(x_i, y_j) / P_Y(y_j) \quad \text{for } i = 1, 2, 3, \dots \\ = 0 \quad \text{if } P_Y(y_j) = 0$$

The conditional probability function of Y, given $X = x_i$ denoted by $P_{Y/x_i} \left(\frac{y_j}{x_i} \right)$ is defined

$$\text{by } P_{Y/x_i} \left(\frac{y_j}{x_i} \right) = P(x_i, y_j) / P_X(x_i) \quad \text{for } j = 1, 2, 3, \dots \\ = 0 \quad \text{if } P_X(x_i) = 0$$

Therefore, $P(x_i, y_j) = P(X = x_i, Y = y_j)$; $P(Y = y_j) = P_Y(y_j)$ and $P(X = x_i) = P_X(x_i)$

Independent and dependent Random Variables:

Definition: Two random variables x and Y are called independent if for every pair of real number “x” and “y”, the two events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent. That is we can express as $P\{X \leq x, Y \leq y\} = P\{X \leq x\} P\{Y \leq y\}$ ----- (1)

Condition (1) in terms of distribution functions is $F(x, y) = F_X(x) F_Y(y)$ ----- (2)

and also $f(x, y) = f_X(x) f_Y(y)$ [if densities exists]----- (3)

Conversely, if (2) or (3) is true, then (1) follows.

Definition: dependent Variates: Variates which are not independent are called dependent variates or dependent random variables.

Definition: continuous random Variates: A 2-dimensional random vector (X, Y) is called a continuous random vector if there exists a function $f(x, y) \geq 0$ such that for $-\infty < x, y < \infty$,

the c. d. f. of (X, Y) given by $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du$ is continuous. The function $f(x, y)$

is called the joint p. d. f of (X, Y).

Some properties of joint density: Let $f(x, y) \geq 0$ be the joint p. d. f of continuous random vector (X, Y) and $F(x, y)$ be the c. d. f. of (X, Y) then it holds the following properties:

$$(i) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x, y) dx \right] dy = 1 \quad (ii) P\{a < X \leq b, c < Y \leq d\} = \int_a^b \int_c^d f(x, y) dy dx$$

$$(iii) f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Individual or Marginal Distributions: Let (X, Y) be a continuous random vector with joint c. d. f. “F” and joint p. d. f. “f”. Then $F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du$.

Definition: Let (X, Y) be a 2-dimesnional continuous random vector with joint p. d. f. $f(x, y)$. Then the individual or marginal distribution of X and Y are defined by the p. d. f.’s

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \text{ and } f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

$$\text{On observation, we have } P(a \leq X \leq b) = \int_a^b f_X(x) dx = \int_a^b \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx.$$

Conditional Distribution Function: The conditional c. d. f. of a variate X , given $Y = y$, written $F_{X/Y}(x/y) = \lim_{\epsilon \rightarrow 0^+} P\{X \leq x/y - \epsilon < Y \leq y + \epsilon\}$ -----(1)

Provided that the limit in (1) exists.

The conditional p. d. f. of X given $Y = y$, written $f_{X/Y}(x/y) \quad \forall x \in R$ is a non-negative

function satisfying $F_{X/Y}(x/y) = \int_{-\infty}^x f_{X/Y}(t/y) dt \quad \forall x \in R$.

Note: The conditional p. d. f. $f(x/y)$ is given by $f_{X/Y}(x/y) = f(x, y)/f_Y(y)$ where f_Y is the marginal p. d. f. of Y , $f_Y(y) > 0$, and is continuous.

Chapter 2

Probability Distributions

Lecture 3

By Dr. N V Nagendram

Mathematical Expectation:

Let X be a discrete random variable taking value x , $x = 0, 1, 2, 3, \dots$ then the mathematical expectation of X is denoted by $E(x)$ [read it as expected value of X] and is defined as

$$E(X) = \sum_x x P(X = x)$$

Similarly if k is any positive integer then, $E(X^k) = \sum_x x^k P(X = x)$.

Similarly if X is continuous random variable taking value x , $-\infty < x < \infty$ with $f(x)$ as the probability density function then $E(X) = \int_{-\infty}^{\infty} x f(x) dx$ and $E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx$.

Definition: $E(X)$ is also called mean or arithmetic mean of X denoted by μ .

Definition: If X is a random variable then the variance of X is denoted by $V(X)$ and is defined as $V(X) = E[(X - E(X))^2]$.

This can be simplified as $V(X) = E(X^2) - [E(X)]^2$.

Notation: The variance is denoted by $\sigma^2 = V(X)$.

Standard deviation: The positive square root of variance is defined as standard deviation and is denoted by σ . Therefore, $\sigma = \sqrt{V(X)}$.

Theorem of Expectation of a sum and product:

Expectation of the sum of the random variables:

Theorem: The mathematical expectation of a sum of a number of random variables is equal to the sum of their expectations.

Proof: follows

Expectation of the product of the random variables:

Theorem: The mathematical expectation of the product of a number of independent random variates is equal to product of their expectations.

Theorem: The mathematical expectation of a sum of a number of random variables is equal to the sum of their expectations.

Proof: Let us consider the random variables x and y . Let x assume the values x_i for all $i = 1, 2, 3, \dots, m$ and y the values y_j for all $j = 1, 2, 3, \dots, n$ with respective probabilities P_i and P_j . The sum $x + y$ is a random variable which can take $m \cdot n$ values,

$$x_i + y_j \quad \text{for } i = 1, 2, 3, \dots, m \\ \text{for } j = 1, 2, 3, \dots, n \text{ with probabilities } P_{ij}.$$

Hence its Expectation is $E(x + y) = \sum_{j=1}^n \sum_{i=1}^m (x_i + y_j) P_{ij}$

$$= \sum_{j=1}^n \sum_{i=1}^m x_i P_{ij} + \sum_{j=1}^n \sum_{i=1}^m y_j P_{ij} \\ = \sum_{i=1}^m x_i \left(\sum_{j=1}^n P_{ij} \right) + \sum_{j=1}^n y_j \left(\sum_{i=1}^m P_{ij} \right) \\ = \sum_i x_i P_{ij} + \sum_j y_j P_{ij} = E(x) + E(y)$$

$$\therefore E(x + y) = E(x) + E(y)$$

Since $\sum_{j=1}^n P_{ij} = P_j$ and $\sum_{i=1}^m P_{ij} = P_i$.

By generalization of the above theorem, we have

$$E(x_1 + x_2 + x_3 + \dots + x_n) = E(x_1) + E(x_2) + E(x_3) + \dots + E(x_n).$$

This completes the proof of the theorem. $\square$

Expectation of the product of the random variables:

Theorem: The mathematical expectation of the product of a number of independent random variates is equal to product of their expectations.

Proof: Let us use the notation as $E(xy) = \sum_j \sum_i x_i y_j P_{ij}$

Since the variates are independent, by the law of compound probabilities we have

$$P_{ij} = P_i \cdot P_j$$

$$\sum_j \sum_i x_i P_i y_j P_j = \sum_i x_i P_i \sum_j y_j P_j = \sum_i P_i x_i E(y) = E(y) \sum_i P_i x_i = E(x) E(y)$$

The theorem can be generalized for a number of independent random variates such that

$$E(x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n) = E(x_1) \cdot E(x_2) \cdot E(x_3) \cdot \dots \cdot E(x_n).$$

This completes proof of the theorem. $\square$

Note: $E(x, y) = E(x) E(y)$ does not guarantee the independent of x and y .

One can easily verify that the following expectations:

$$(a) E(a) = a \quad (b) E(aX) = a E(X) \quad (c) E(aX \pm bY) = aE(X) \pm bE(Y)$$

$$(d) E(aX + b) = a E(X) + b \quad (e) V(a) = 0 \quad (f) V(aX \pm b) = a^2 V(X)$$

$$(g) V(x) = E(x^2) - [E(x)]^2 \quad (h) V(aX + bY) = a^2 \sigma_X^2 + b^2 \sigma_Y^2 + 2ab \sigma_{XY}.$$

Chapter 2 Probability Distributions

Tutorial 3
By Dr. N V Nagendram

Problem 1: Two coins are tossed simultaneously. Let X denote the number of heads, Find E(X) and V(X)?

Solution:

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

$$\text{Mean: } \mu = E(X) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = 1$$

$$\begin{aligned}\text{Variance: } \sigma^2 = V(X) &= E(X^2) - [E(X)]^2 \\ &= 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{2}{4} + 2^2 \cdot \frac{1}{4} - (1)^2 \\ &= \frac{2}{4} + 1 - 1 \\ &= \frac{1}{2}\end{aligned}$$

Hence the solution.

Problem 2: If it rains, a dealer in rain coats earns Rs. 500/- per day and if it is fair, he loses Rs.50/- per day. If the probability of a rainy day is 0.4. Find his average daily income?

Solution:

X = x	500	-50	Total
P(X = x)	0.4	0.6	1

$$\begin{aligned}\text{Average} &= E(X) = 500 (0.4) + (-50) (0.6) \\ &= 200 - 30 \\ &= \text{Rs. } 170/-\end{aligned}$$

Hence the solution.

Chapter 2

Probability Distributions

Lecture 4

By Dr. N V Nagendram

Binomial Distribution:

This distribution was discovered by James Bernoulli. This is a discrete distribution. It occurs in cases of repeated trials such as students writing an examination, births in a hospital etc. Here all the trials are assumed to be independent and each trial has only two outcomes namely success and failure.

Let an experiment consist of “n” independent trials. Let it succeed “x” times. Let “p” be the probability of success and “q” be the probability of failure in each trial.

$$\therefore p + q = 1$$

$$\text{The probability of getting } x \text{ successes} = p.p.p.\dots\dots p(x \text{ times}) = p^x$$

$$\text{The probability of getting } (n - x) \text{ failures} = q.q.q.\dots\dots q[(n - x) \text{ times}] = q^{(n-x)}$$

$\therefore$ From multiplication theorem, the probability of getting x successes and (n – x) failures is $p^x q^{(n-x)}$.

This is the probability of getting x successes in one combination. There are such nC_x mutually exclusive combinations each with probability $p^x q^{(n-x)}$.

$\therefore$ From addition theorem the probability of getting x success in ${}^nC_x p^x q^{(n-x)}$.

Notation: $b(x; n, p)$ denotes a binomial distribution with x successes, n trials and with p as the probability of success.

$$\therefore b(x; n, p) = {}^nC_x p^x q^{(n-x)}, x = 0, 1, 2, 3, \dots, n.$$

Parameters of Binomial distribution:

In $b(x; n, p)$ there are 3 constants viz., n, p and q. Since $q = 1 - p$, hence there are only 2 independent constants namely n and p. These are called the parameters of binomial distribution.

Note: since $b(x; n, p)$ is same as the $(x + 1)^{\text{th}}$ term in the binomial expansion of $(q + p)^n$, hence this distribution is called the “Binomial Distribution”.

Mean of the Binomial distribution:

$$\begin{aligned}\text{Mean} = \mu &= \sum_{x=0}^n x \cdot b(x; n, p) \\&= \sum_{x=0}^n x \cdot \frac{n!}{x!(n-x)!} p^x q^{(n-x)} \\&= \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x q^{(n-x)} \\&= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{(n-x)}\end{aligned}$$

Put $y = x - 1$, $\therefore x = 1 + y$

When $x = 1$ implies $y = 0$

$x = 1$ implies $y = x - 1$

$$\begin{aligned}&= np \sum_{y=0}^{n-1} {}^{(n-1)}C_y p^y q^{(n-y-1)} \\&= np (q + p)^{(n-1)}\end{aligned}$$

So, $\mu = np$ is Mean of Binomial distribution.

Variance of the Binomial Distribution:

$$\begin{aligned} E(X)^2 &= \sum_{x=0}^n x^2 b(x, n, p) \\ &= \sum_{x=0}^n [x(x-1) + x] b(x, n, p) \\ &= \sum_{x=0}^n x(x-1) b(x, n, p) + \sum_{x=0}^n x b(x, n, p) \\ &= \sum_{x=0}^n x(x-1) \frac{n!}{x! (n-x)!} p^x q^{(n-x)} + \mu \\ &= \sum_{x=2}^n \frac{n!}{(x-2)! (n-x)!} p^x q^{(n-x)} + np \\ &= n(n-1) p^2 \sum_{x=2}^n \frac{(n-2)!}{(x-2)! (n-x)!} p^{(x-2)} q^{(n-x)} + np \end{aligned}$$

$$\text{Put } y = x - 2 \quad \therefore x = 2 + y$$

$$\text{When } x = 2 \text{ implies } y = 0$$

$$\text{When } x = n \text{ implies } y = n - 2$$

$$\begin{aligned} &= n(n-1) p^2 \sum_{y=0}^{n-2} \frac{(n-2)!}{y! (n-2-y)!} p^y q^{(n-2-y)} + np \\ &= n(n-1) p^2 \sum_{y=0}^{n-2} {}^{(n-2)}C_y p^y q^{(n-2-y)} + np \\ &= n(n-1) p^2 (q + p)^{n-2} + np \end{aligned}$$

$$E(X^2) = n(n-1) p^2 + np$$

$$\sigma^2 = E(X^2) - [E(X)]^2$$

$$= n(n-1) p^2 + np - (np)^2$$

$$= n^2 p^2 - np^2 + np - n^2 p^2$$

$$= np(1 - p)$$

$$\sigma^2 = npq$$

$\therefore$ The variance of binomial distribution is npq . The standard deviation is $\sigma = +\sqrt{npq}$.

Moments of Binomial distribution:

$$\begin{aligned}\text{Mean } \mu'_1 = \mu = E[X] &= \sum_x x P(x) \\&= \sum_{x=0}^n x {}^nC_x p^x q^{(n-x)} \\&= \sum_{x=0}^n x \frac{n!}{x! (n-x)!} p^x q^{(n-x)} \\&= \sum_{x=1}^n \frac{n(n-1)!}{(x-1)! (n-x)!} p p^{(x-1)} q^{(n-x)} \\&= np \sum_{x=1}^n \frac{n(n-1)!}{(x-1)! (n-x)!} p^{(x-1)} q^{(n-x)} \\&= np(p+q)^{(n-1)} = np(1)^{(n-1)} = np\end{aligned}$$

$$\begin{aligned}E[X^2] = \mu'_2 &= \sum_x x^2 P(x) \text{ by definition} \\&= \sum_{x=0}^n x^2 {}^nC_x p^x q^{(n-x)} \\ \mu'_2 &= \sum_{x=0}^n \{x(x-1) + x\} {}^nC_x p^x q^{(n-x)} \\ \mu'_2 &= \sum_{x=0}^n x(x-1) {}^nC_x p^x q^{(n-x)} + \sum_{x=0}^n x {}^nC_x p^x q^{(n-x)} \\ \mu'_2 &= \sum_{x=0}^n x(x-1) {}^nC_x p^x q^{(n-x)} + np\end{aligned}$$

Consider

$$\begin{aligned}\mu'_2 &= \sum_{x=0}^n x(x-1) {}^nC_x p^x q^{(n-x)} \\ \mu'_2 &= \sum_{x=0}^n x(x-1) \frac{n!}{x! (n-x)!} p^x q^{(n-x)} \\ \mu'_2 &= \sum_{x=2}^n \frac{n!}{(x-2)! (n-x)!} p^x q^{(n-x)} \\ \mu'_2 &= \sum_{x=2}^n \frac{n(n-1)(n-2)!}{(x-2)! (n-x)!} p^2 p^{(x-2)} q^{(n-x)} \\ \mu'_2 &= n(n-1)p^2 (q+p)^{(n-2)} \\ \mu'_2 &= n(n-1)p^2 \\ \therefore \mu'_2 &= n(n-1)p^2 + np\end{aligned}$$

$$\begin{aligned}
\mu_2 \text{ (variance)} &= \mu'_2 - \mu'^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= n^2 p^2 - np^2 + np - n^2 p^2 \\
&= np(1-p) \\
&= npq \quad \text{[since, } 1-p=q\text{]} \\
\therefore np &> npq \quad \text{[since } q \text{ is a fraction]}
\end{aligned}$$

Mean > Variance

Similarly $\mu_3 = npq(1-2p)$

$$\text{Hence } \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{n^2 p^2 q^2 (1-2p)^2}{(npq)^3} = \frac{(1-p-p)^2}{npq} = \frac{(q-p)^2}{npq}$$

$$\text{When } p = \frac{1}{2} = q$$

$$\text{Therefore } \beta_1 = 0$$

$$\text{Case (i) when } p = \frac{1}{2}, \beta_1 = 0$$

$$\text{Case (ii) when } n \rightarrow \infty \text{ then } \beta_1 = 0$$

$$\text{Standard deviation} = \sqrt{npq}$$

$$\text{Skewness} = \frac{1-2p}{\sqrt{npq}}$$

Moment Generating Function of Binomial Distribution:

Let X be a variable following binomial distribution, then

$$\begin{aligned}
 M_x(t) &= E(e^{tx}) = \sum_{x=0}^n e^{tx} p(x) \\
 &= \sum_{x=0}^n e^{tx} {}^nC_x p^x q^{(n-x)} \\
 &= (q + pe^t)^n
 \end{aligned}$$

M.G.F about Mean of binomial Distribution:

$$\begin{aligned}
 E\{e^{t(x-np)}\} &= E(e^{tx - tnp}) \\
 &= e^{-tnp} E(e^{tx}) \\
 &= e^{-tnp} M_x(t) \\
 &= e^{-tnp} (q + pe^t)^n \\
 &= (qe^{-pt} + pe^{t-tp})^n \\
 &= (qe^{-pt} + pe^{t-tp})^n \\
 &= (qe^{-pt} + pe^{tq})^n \\
 &= \left[q \left\{ 1 - pt + \frac{p^2 t^2}{2!} - \frac{p^3 t^3}{3!} + \frac{p^4 t^4}{4!} - \dots \right\} + p \left\{ 1 + tq + \frac{t^2 q^2}{2!} + \frac{t^3 q^3}{3!} + \dots \right\} \right]^n \\
 &= \left[\left\{ 1 + \frac{t^2}{2!} pq(p+q) + \frac{t^3}{3!} pq(q^2 - p^2) + \frac{t^4}{4!} pq(q^3 + p^3) + \dots \right\} \right]^n \\
 &= \left[1 + {}^nC_1 \left\{ \frac{t^2}{2!} pq + \frac{t^3}{3!} pq(q-p) + \frac{t^4}{4!} pq(1-3pq) + \dots \right\} + \right. \\
 &\quad \left. {}^nC_2 \left\{ \frac{t^2}{2!} pq + \frac{t^3}{3!} pq(q-p) + \frac{t^4}{4!} pq(1-3pq) + \dots \right\}^2 + \dots \right]
 \end{aligned}$$

Since we have, $a^3 + b^3 = (a+b)(a^2 - ab + b^2)$; $p^3 + q^3 = (p+q)(p^2 - pq + q^2)$
 $= (1)(p+q)^2 - 3(pq) = (1 - 3pq)$

Now $\mu_2 = \text{coefficient of } \frac{t^2}{2!} = npq$; $\mu_3 = \text{coefficient of } \frac{t^3}{3!} = npq(q-p)$;

$$\begin{aligned}
 \mu_4 &= \text{coefficient of } \frac{t^4}{4!} \\
 &= npq(1-3pq) + \frac{n!}{2!(n-2)!} 6 \frac{t^4}{4 \times 3 \times 2} p^2 q^2 \\
 &= npq(1-3pq) + \frac{n(n-1)(n-2)!}{2!(n-2)!} 6 p^2 q^2 \\
 &= npq(1-3pq) + 3n(n-1)p^2 q^2
 \end{aligned}$$

Additive Property of Binomial Distribution:

Let $X \sim b(n_1, p_1)$ and $Y \sim b(n_2, p_2)$ be independent random variables.

Then $M_X(t) = (q_1 + p_1 e^t)^{n_1}$; $M_Y(t) = (q_2 + p_2 e^t)^{n_2}$ what is distribution of $X + Y$

We have $M_{X+Y}(t) = M_X(t) M_Y(t)$ [since X and Y are independent]

$$= (q_1 + p_1 e^t)^{n_1} (q_2 + p_2 e^t)^{n_2}$$

Since (2) cannot be expressed in the form $(q + pe^t)^n$, from uniqueness theorem of m.g.f it follows that $X + Y$ is not a binomial variate. Hence, in general the sum of two independent binomial variates is not a binomial variate.

In other words, binomial distribution does not possess the additive or reproductive property.

However, if we take $p_1 = p_2 = p$ say then from (2), we get

$$M_{X+Y}(t) = (q + pe^t)^{n1+n2}$$

which is the m.g.f of a binomial variate with parameters (n_1+n_2, p) . Hence, by uniqueness theorem of m.g.f's $X + Y \sim b(n_1+n_2, p)$. Thus the binomial distribution possesses the additive or reproductive property if $p_1 = p_2$.

Generalization: If X_i for all $i = 1, 2, 3, \dots, k$ then their sum $\sum_{i=1}^k X_i \approx B\left(\sum_{i=1}^k n_i p\right)$

Recurrence Relation for the probabilities of Binomial Distribution: (Fitting of Binomial Distribution)

We have $P(x+1) = {}^nC_{x+1} p^{x+1} q^{n-x-1}$ and $P(x) = {}^nC_x p^x q^{n-x}$

$$\therefore \frac{P(x+1)}{P(x)} = \frac{{}^nC_{x+1} p^{x+1} q^{n-x-1}}{{}^nC_x p^x q^{n-x}} = \frac{\frac{n!}{(x+1)!(n-x-1)!} p}{\frac{n!}{x!(n-x)!} q}$$

$$\frac{P(x+1)}{P(x)} = \frac{x!(n-x)!}{(x+1)!x!} \frac{(n-x-1)!}{(n-x-1)!} \frac{p}{q} = \left(\frac{n-x}{1+x}\right) \frac{p}{q}$$

$$\therefore P(x+1) = \left\{ \left(\frac{n-x}{x+1} \right) \frac{p}{q} \right\} P(x) \text{ which is the required recurrence formula.}$$

*** *** *** *** *** *** *** *** *** *** ***

Problem 1: It has been claimed that in 60% of all solar heat installations the utility bills is reduced by at least one third. Accordingly what are the probabilities that the utility bill will be reduced by at least one third in (i) four or five installations (ii) at least four of five installations?

Problem 2: Two coins are tossed simultaneously. Find the probability of getting at least seven heads?

Problem 3: If 3 of 20 tyres are defective and 4 of them are randomly chosen for inspection. What is the probability that only one of the defective tyres will be included?

Problem 1: It has been claimed that in 60% of all solar heat installations the utility bills is reduced by at least one third. Accordingly what are the probabilities that the utility bill will be reduced by at least one third in (i) four or five installations (ii) at least four of five installations?

Solution: $n = 5, p = 0.6, q = 1 - p = 0.4$
 (i) $b(4; 5, 0.6) = {}^5C_4 (0.6)^4 (0.4)^1 = 5(0.6)^4(0.4) = 0.2592$
 (ii) at least 4 means 4 or 5
 $b(5; 5, 0.6) = {}^5C_5 (0.6)^5 (0.4)^0 = 0.0778$
 $\therefore$ Probability in at least four installations = $b(4; 5, 0.6) + b(5; 5, 0.6)$
 $= 0.2592 + 0.0778 = 0.337$

Hence the solution.

Problem 2: Two coins are tossed simultaneously. Find the probability of getting at least seven heads?

Solution: $n = 10, p = P(H) = \frac{1}{2}; q = 1 - p = \frac{1}{2}$
 $P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$
 $= {}^{10}C_7 ({}^1C_2)^7 ({}^1C_2)^3 + {}^{10}C_8 ({}^1C_2)^8 ({}^1C_2)^2 + {}^{10}C_9 ({}^1C_2)^9 ({}^1C_2)^1 + {}^{10}C_{10} ({}^1C_2)^{10} ({}^1C_2)^0$
 $= \frac{1}{2^{10}} [{}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}] = \frac{1}{2^{10}} [{}^{10}C_3 + {}^{10}C_2 + {}^{10}C_1 + {}^{10}C_0]$
 $= \frac{1}{2^{10}} \left[\frac{10.9.8}{1.2.3} + \frac{10.9}{1.2} + 10 + 1 \right] = \frac{1}{2^{10}} [120 + 45 + 10 + 1]$
 $= \frac{176}{2^{10}} = 0.172$

Hence the solution.

Problem 3: If 3 of 20 tyres are defective and 4 of them are randomly chosen for inspection. What is the probability that only one of the defective tyres will be included?

Solution: $n = 4, p = \frac{3}{20}, q = 1 - p = \frac{17}{20}$
 $P(x = 1) = {}^4C_1 (p)^1 (q)^{(4-1)}$
 $= 4 \cdot \frac{3}{20} \cdot \left(\frac{17}{20}\right)^3 = \frac{4 \cdot 3 \cdot 17^3}{20^4} = 0.368$

Hence the solution.

Chapter 2 Probability Distributions

Binomial distribution

Tutorial 5
By Dr. N V Nagendram

Problem 1: Determine the binomial distribution for which the mean is four and variance three. Also find its mode? [Ans.4.25or4]

Problem 2: If A and B play games of chess of which 6 are won by A, 4 are won by B and 2 end in draw. Find the probability that (i) A and B win alternatively (ii) B wins at least one game (iii) Two games end in draw? [Ans.5/36,19/27,5/72]

Problem 3: If the probability that a person will not like a new tooth paste is 0.20. what is the probability that 5 out of 10 randomly selected persons will dislike it? [Ans. 0.0264]

Problem 4: A shipment of 20 tape recorders contains 5 defectives find the standard deviation of the probability distribution of the number of defectives in a sample of 10 randomly chosen for inspection? [Ans, $\sigma=\sqrt{8/5}$ (S.D.)]

Problem 5: If A and B play game in which their chances of winning are in the ratio 3 : 2 Find A's chance of winning at least three games out of the five games played? [Ans. 0.68]

Problem 6: A department has 10 machines which may need adjustment from time to time during the day. Three of these machines are old, each having a probability of $\frac{1}{11}$ of needing adjustment during the day and 7 are new, having corresponding probabilities of $\frac{1}{21}$. Assuming that no machine needs adjustments twice on the same day, determine the probabilities that on a particular day. (i) just 2 old and no new machines need adjustment. (ii) if just 2 machines need adjustment, they are of the same type. [Ans. 0.016;0.028]

Problem 7: An irregular six faced die is thrown and the probability exception that in 10 throws it will give five even numbers is twice, the probability expectation that it will give four even numbers. How many times in 10000 sets of 10 throws each, would you expect it to give no even number? [Ans. 1 approxly]

Problem 8: The mean of binomial distribution is 3 and variance is 4? [Ans. $\frac{4}{3}$]

Problem 9: The mean and variance of binomial distribution are 4 and $\frac{4}{3}$ respectively. Find $P(X \geq 1)$? [Ans.0.9983]

Chapter 2 Probability Distributions

Binomial distribution

Tutorial 6
By Dr. N V Nagendram

Problem 01: Find a binomial distribution for the following data and compare the theoretical frequencies with the actual ones:

x:	0	1	2	3	4	5
f:	2	14	20	34	22	8

[Ans. 100(0.432 + 0.568)]

Problem 02: The probability that a bomb dropped from a plane will strike the target is $\frac{1}{5}$. If six bombs are dropped, find the probability that (i) exactly two will strike the target, (ii) at least two will strike the target.
[Ans. (i) 0.246 (ii) 0.345]

Problem 03: If the probability that a new-born child is a male is 0.6, find the probability that in a family of 5 children there are exactly 3 boys?
[Ans. 0.3456]

Problem 04: Find the probability of guessing correctly at least 6 of the 10 answers on a true-false examination?
[Ans. $\frac{193}{512}$]

Problem 05: Out of 800 families with 5 children each, how many would you expect to have (i) 3 boys (ii) 5 girls and (iii) either 2 or 3 boys? Assuming that equal probabilities for girls and boys.
[Ans. (i) 250 (ii) 25 (iii) 500]

Problem 06: If the probability of a defective bolt is 0.1, find (i) the mean and (ii) the standard deviation for the distribution of defective bolts in a total of 400? [Ans. (i) 40 (ii) 6]

Problem 07: Find the probability that in five tosses of a fair die a 3 appears (i) at no times (ii) four times?
[Ans. (i) $\frac{3125}{7776}$ (ii) $\frac{25}{7776}$]

Problem 08: Find the probability that in a family of 4 children there will be (i) at least 1 boy and (ii) at least 1 boy and 1 girl?
[Ans. (i) $\frac{15}{16}$ (ii) $\frac{7}{8}$]

Problem 09: Find the probability of getting at least 4 heads in 6 tosses of a fair coin?
[Ans. $\frac{11}{32}$]

Chapter 2 Probability Distributions

Binomial distribution

Tutorial 7
By N V Nagendram

Problem 1: The following data due to Weldon shows the results of throwing 12 dice 4096 times, a throw of 4, 5 or 6 being called success (x).

X	0	1	2	3	4	5	6	7	8	9	10	11	12
V	-	7	60	198	430	731	948	847	536	257	71	11	-

Fit a Binomial distribution and calculate the expected frequency?

$$[\text{Ans. } \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{25145}{4096} = 6.14]$$

Problem 2: Fit a Binomial distribution to the following data and test for goodness of fit

X	0	1	2	3	4
F	28	62	46	10	4

$$[\text{Ans. } \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{200}{15} = 1.33; p = 0.333; q = 0.667]$$

Problem 3: In 256 sets of 12 tosses of a coin, in how many cases one can expect eight heads and 4 tails?

$$[\text{Ans. } P(X=8) = {}^{12}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4; \text{no of cases} = 256 \times P(X=8) = 256 \times 11.58 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4 = 31 (\text{apprxly})]$$

Problem 4: The mean and variance of a binomial variate X with parameters “n” and p are 16 and 8. Find (i) p(X = 0) (ii) p(X = 1) and (iii) p(X ≥ 2).

$$[\text{Ans. (i) } p(X=0) = {}^{32}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{32}; \text{ (ii) } P(X=1) = {}^{32}C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{31};$$

$$\text{And (iii) } P(X \geq 2) = 1 - \left\{ {}^{32}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{32} + {}^{32}C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{31} \right\}$$

Problem 5: Seven coins are tossed and the number of heads are noted. The experiment is repeated 128 times and the following distribution is obtained:

No of heads	0	1	2	3	4	5	6	7	Total
Frequencies	7	6	19	35	30	23	7	1	128

Fit a binomial distribution(B.D.) assuming (i) the coin is unbiased

(ii) the nature of the coin is not known?

Chapter 2

Lecture 5

Chebyshev's theorem

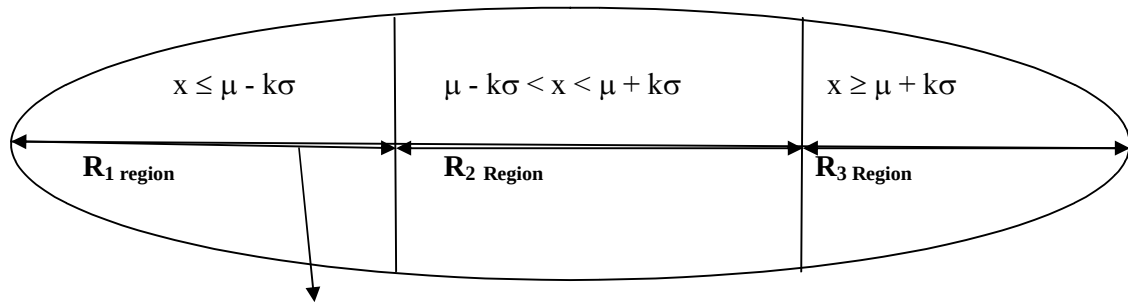
Probability Distributions by Dr. N. V. Nagendram

Chebyshev's theorem: Let X be a random variable with mean μ and standard deviation σ then $P(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2}$.

Proof: Let $f(x)$ be the probability mass function of a random variable having mean μ and variance σ^2 .

$$\text{Now } \sigma^2 = \sum_x (x - \mu)^2 f(x) \quad \dots\dots\dots(1)$$

Let R_1 be the region in which $x \leq \mu - k\sigma$, R_2 the region in which $\mu - k\sigma < x < \mu + k\sigma$ and R_3 be the region in which $x \geq \mu + k\sigma$.



Values of x

$$\therefore \sigma^2 = \sum_{R_1} (x - \mu)^2 f(x) + \sum_{R_2} (x - \mu)^2 f(x) + \sum_{R_3} (x - \mu)^2 f(x) \quad \dots\dots\dots(2)$$

Since $(x - \mu)^2 \geq 0$ i.e., non-negative, hence $\sum_{R_2} (x - \mu)^2 f(x) \geq 0$ also non-negative.

$$\text{So equation (2) implies } \sigma^2 \geq \sum_{R_1} (x - \mu)^2 f(x) + \sum_{R_3} (x - \mu)^2 f(x) \quad \dots\dots\dots(3)$$

$$\begin{aligned} \text{In } R_1 \quad & x \leq \mu - k\sigma \\ \Rightarrow \quad & x - \mu \leq -k\sigma \quad \dots\dots\dots(4) \end{aligned}$$

$$\begin{aligned} \text{In } R_3 \quad & x \geq \mu + k\sigma \\ \Rightarrow \quad & x - \mu \geq k\sigma \quad \dots\dots\dots(5) \end{aligned}$$

Equations (4) and (5) $\Rightarrow |x - \mu| \geq k\sigma$

Hence both R_1 and R_3 , $(x - \mu)^2 \geq k^2 \sigma^2$.

$$\therefore \text{From (3)} \quad \sigma^2 \geq \sum_{R_1} k^2 \sigma^2 f(x) + \sum_{R_3} k^2 \sigma^2 f(x) \quad \dots\dots\dots(6)$$

$$\text{i.e., } \sigma^2 \geq k^2 \sigma^2 \left[\sum_{R_1} f(x) + \sum_{R_3} f(x) \right] \Rightarrow \frac{1}{k^2} \geq \left[\sum_{R_1} f(x) + \sum_{R_3} f(x) \right] \quad \dots\dots\dots(7)$$

Now $\left[\sum_{R_1} f(x) + \sum_{R_3} f(x) \right]$ represents the probability assigned to the region $R_1 \cup R_3$.

$$\therefore \left[\sum_{R_1} f(x) + \sum_{R_3} f(x) \right] = P[|x - \mu| \geq k\sigma] \dots\dots\dots(8)$$

$$\therefore \text{From equations (7) and (8) } \frac{1}{k^2} \geq P[|x - \mu| \geq k\sigma]$$

$$\Rightarrow P[|x - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

This completes the proof of the theorem

$$\textbf{Note: } P[|x - \mu| \geq k\sigma] \leq \frac{1}{k^2} \Rightarrow P[|x - \mu| < k\sigma] \leq 1 - \frac{1}{k^2}.$$

Chapter 2 Probability Distributions

Chebyshev's theorem

Tutorial 8
By Dr. N V Nagendram

Problem 1: X is random variable such that $E(X) = 3$ and $E(X^2) = 13$. Determine a lower bound for $P(-2 < x < 8)$, using Chebyshev's inequality? [Ans. $\sigma = 2$; lower bound = $\frac{21}{25}$]

Problem 2: 500 articles were selected at random out of a batch containing 10000 articles and 30 were found to be defective. How many defective articles would you reasonably expect to have in the whole batch? [Ans. $E(X) = Np = 10000 \times \frac{3}{50} = 600$]

Problem 3: A symmetric die is thrown 600 times. Find the lower bound for the probability of getting 80 to 120 sixes? [Ans. $P(80 \leq x \leq 120) = \frac{19}{24}$]

Problem 4: Given that the discrete random variable X has density function $f(x)$ given by $f(-1) = \frac{1}{8}$, $f(0) = \frac{6}{8}$, $f(1) = \frac{1}{8}$ use Chebyshev's inequality to find the upper bound when $k = 2$? [Ans. $\frac{1}{4}$]

Problem 5: For geometric distribution $P(x) = 2^{-x}$; $x = 1, 2, \dots$. Prove that Chebyshev's inequality gives $P(|x - 2| \leq 2) > \frac{1}{2}$ while the actual probability is $\frac{15}{16}$.

Problem 6: Two unbiased dice are thrown. If X is the sum of the numbers showing up. Prove that $P(|x - 7| \geq 3) \leq \frac{35}{54}$. Also compare this with actual probability?

Problem 7: Suppose that X assumes the values 1 and -1 , each with probability 0.5. Find and compare the lower bound on $P[-1 < X < 1]$ given by Chebyshev's inequality and the actual probability that $-1 < X < 1$?

Problem 8: Find a lower bound on $P[-3 < X < 3]$ where $\mu = E(X) = 0$ and variance $= \sigma^2 = 1$. [Ans. L.b = $\frac{8}{9}$]

Problem 9: Use Chebyshev's inequality to find a lower bound (l. b.) on $P[-4 < X < 20]$ where the random variable X has a mean $\mu = 8$ and variance $\sigma^2 = 9$. [Ans. $\frac{15}{16}$]

Problem 10: If X is the number appearing on a die when it is thrown, show that the Chebyshev's theorem gives $P[|x - \mu| > 2.5] < 0.47$ while the actual probability is zero.

Problem 11: The number of customers who visit a car dealer show room on a certain day is a random variable with mean 18 and standard deviation 2.5. With what probability can it be asserted that there will be between 8 and 28 customers? [Ans. $\frac{15}{16}$]

Chapter 2 Probability Distributions

Chebyshev's theorem

Tutorial 9
By Dr. N V Nagendram

Problem 1: A random variable X has density function given by $f(x) = \begin{cases} 2e^{-2x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

(a) Find $P[|x - \mu| > 1]$; (b) Use Chebyshev's inequality to obtain an upper bound on $P[|x - \mu| > 1]$ and compare with the result in (a). [Ans. (a) $e^{-3} = 0.04979$ (b) 0.25]

Problem 2: Prove Chebyshev's inequality for a discrete variable X?

Problem 3: Let $X_1, X_2, X_3, \dots, X_n$ be n independent random variables each having density

function $f(x) = \begin{cases} \frac{1}{2} & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$. If $S_n = X_1 + X_2 + X_3 + \dots + X_n$ then show that

$$P\left[\left|\frac{S_n}{n} - \mu\right| < \epsilon\right] = 1.$$

Problem 4: A random variable X has mean 3 and variance 2. Use Chebyshev's inequality to obtain an upper bound for (a) $P[|X - 3| \geq 2]$ (b) $P[|X - \mu| \geq 1]$ [Ans. $1, \frac{1}{4}$]

Problem 5: A random variable X has the density function $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$ then

(a) find $P[|X - \mu| \geq 2]$ (b) use Chebyshev's inequality to obtain an upper bound on $P[|X - \mu| \geq 2]$ and compare with the result in (a). [Ans. (a) e^{-2} , (b) 0.5]

Chapter 2

Lecture 6

Poisson's theorem

Probability Distributions

by Dr. N. V. Nagendram

Definition: A random variable X is said to follow Poisson distribution if its probability mass

function is given by $f(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$, for $x = 0, 1, 2, 3, \dots$
 $= 0$ otherwise.

Poisson Approximation to Binomial Distribution Theorem:

Statement: As $n \rightarrow \infty$ and $p \rightarrow 0$ so that $np = \lambda$ where λ is a finite non-zero constant then

$$b(x, n, p) \rightarrow \frac{\lambda^x}{x!} e^{-\lambda}.$$

Proof: Let us consider $b(x, n, p)$ so that $b(x, n, p) = {}^nC_x p^x q^{n-x}$

$$= \frac{n(n-1)(n-2)\dots(n-(x-1))}{x!} p^x q^{n-x}$$

and given $np = \lambda \Rightarrow p = \frac{\lambda}{n}$ also $q = 1 - p = 1 - \frac{\lambda}{n}$

$$= \frac{n(n-1)(n-2)\dots(n-(x-1))}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!} \frac{n(n-1)(n-2)\dots(n-(x-1))}{n^x} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x}$$

$$= \frac{\lambda^x}{x!} \left(\frac{n}{n}\right) \left(\frac{n-1}{n}\right) \left(\frac{n-2}{n}\right) \dots \left(\frac{n-(x-1)}{n}\right) \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x}$$

$$b(x, n, p) = {}^nC_x p^x q^{n-x} = \frac{\lambda^x}{x!} \left(\frac{n}{n}\right) \left(\frac{n-1}{n}\right) \left(\frac{n-2}{n}\right) \dots \left(\frac{n-(x-1)}{n}\right) \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x} \dots\dots\dots (1)$$

Now as $n \rightarrow \infty$, $\frac{1}{n}, \frac{2}{n}, \dots, \frac{x-1}{n} \rightarrow 0$

$$\left(1 - \frac{\lambda}{n}\right)^x \rightarrow 1 \quad \text{and} \quad \left(1 - \frac{\lambda}{n}\right)^n = \left[\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda} \rightarrow e^{-\lambda}$$

$\therefore$ from equation (1) $b(x, n, p) \rightarrow \frac{\lambda^x}{x!} e^{-\lambda}$.

This completes the proof of the Poisson's Approximation to Binomial distribution theorem.

Note: 1. $e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!}$

2. Show that $\sum_{x=0}^{\infty} f(x, \lambda) = 1$

For that consider $\sum_{x=0}^{\infty} f(x, \lambda) = \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^\lambda = 1$

3. $\lambda > 0$ is called the parameter of the Poisson Distribution.

4. $P(X=0) = \frac{e^{-\lambda} \lambda^0}{0!} = e^{-\lambda}$

Applications of Poisson distribution:

Poisson distribution is applicable when n is very large and p is very small. Hence some of the applications of Poisson distribution are as follows:

1. Number of faulty blades produced by a reputed firm
2. Number of deaths from a disease such as heart attack or cancer.
3. Number of telephone calls received at a particular telephone exchange.
4. Number of cars passing a crossing per minute.
5. Number of printing mistake in a page of a book.

Mean and Variance of Poisson distribution:

$$\text{Mean } \mu = E(X)$$

$$= \sum_{x=0}^{\infty} x P(X=x) = \sum_{x=0}^{\infty} x f(x, \lambda)$$

$$= \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{e^{-\lambda} \lambda^x}{(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} e^{\lambda}$$

Therefore Mean = $\mu = \lambda$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 f(x, \lambda)$$

$$= \sum_{x=0}^{\infty} [x(x-1) + x] f(x, \lambda)$$

$$= \sum_{x=0}^{\infty} x(x-1) f(x, \lambda) + \sum_{x=0}^{\infty} x f(x, \lambda)$$

$$= \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda} \lambda^x}{x!} + \lambda$$

$$= \sum_{x=2}^{\infty} \frac{e^{-\lambda} \lambda^x}{(x-2)!} + \lambda$$

$$= \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda$$

$$\therefore E(X^2) = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda$$

$$E(X^2) = \lambda^2 + \lambda$$

$$\text{Variance} = V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad \therefore \text{variance} = \lambda$$

$$\text{Standard Deviation} = \text{S.D.} = \sqrt{\text{Variance}} = \sqrt{\lambda}$$

Note : In a Poisson distribution mean always equal to the variance.

Chapter 2

Lecture 7

Poisson's m. g. f. **Probability Distributions** by Dr. N. V. Nagendram

Moment Generating Function of Poisson Distribution:

$$\begin{aligned} M_X(t) &= E[e^{tx}] \\ &= \sum_{x=0}^{\infty} e^{tx} f(x, \lambda) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} e^{tx} \frac{(e^t \lambda)^x}{x!} = e^{-\lambda} \cdot e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

Additive Property of Poisson Variates:

Theorem: If x and Y are two independent random Poisson variates with parameters λ and μ then $X + Y$ is also a Poisson variate with parameter $\lambda + \mu$.

Proof: Since X is a Poisson variate with parameter $\lambda \Rightarrow M_X(t) = e^{\lambda(e^t - 1)}$

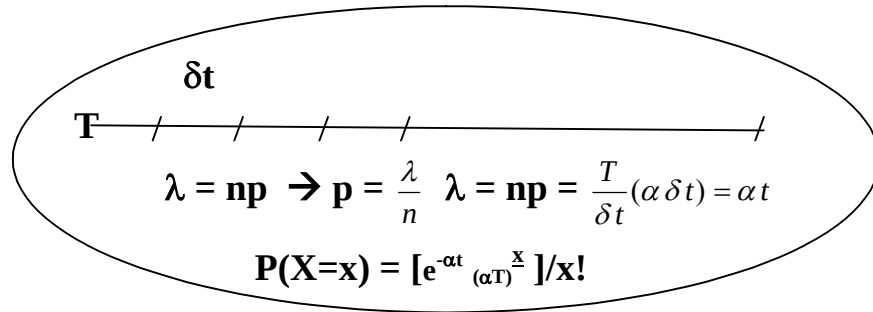
Similarly, since Y is Poisson variate with parameter $\mu \Rightarrow M_Y(t) = e^{\mu(e^t - 1)}$

$$\begin{aligned} \text{From the additive property of the moment generating function } M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= e^{\lambda(e^t - 1)} \cdot e^{\mu(e^t - 1)} \\ &= e^{(\lambda + \mu)(e^t - 1)} \end{aligned}$$

Which is the moment generating function of a Poisson variate with parameter $\lambda + \mu$.

$\therefore X + Y$ is also a Poisson variate with parameter $\lambda + \mu$.

POISSON PROCESS:



Suppose we have to find the probability of x successes during a time interval T . Divide the time interval T into n equal parts of width δt . Therefore $T = n \cdot \delta t$.

Here we make the following assumptions:

- (a) The probability of success during an interval δt is given by $\alpha \cdot \delta t$.
- (b) The probability more than one success in a small interval δt is negligible.
- (c) The probability of success in interval $(t, t+\delta t)$ is independent of the actual time t and also of all previous successes.

Here the assumptions of binomial distribution are satisfied and the probability $b(x, n, p)$

where $n = \frac{T}{\delta t}$ and $p = \alpha \cdot \delta t$.

$\therefore$ As $n \rightarrow \infty$ binomial distribution approaches to Poisson distribution and here parameter λ is

$$\lambda = np \rightarrow p = \frac{\lambda}{n} \quad \lambda = np = \frac{T}{\delta t} (\alpha \delta t) = \alpha t$$

$$\therefore P(X=x) = [e^{-\alpha t} (\alpha t)^x] / x!$$

Chapter 2

Lecture 8

Normal distribution **Probability Distributions** by Dr. N. V. Nagendram

Normal Distribution (N.Dⁿ):

Normal distribution is also a continuous distribution. A random variable X is said to follow normal distribution (N. Dⁿ) with mean μ and variance σ^2 if its probability density function is

$$\text{given by } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]}, -\infty < x < \infty; -\infty < \mu < \infty; \sigma > 0$$

$$= 0, \text{ otherwise.}$$

The corresponding distribution function is $f(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(t-\mu)^2}{2\sigma^2}\right]} dt$

Let $Z = \frac{x-\mu}{\sigma}$ then the mean of Z is 0 and the variance is 1.

The corresponding probability density function is $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}}, -\infty < z < \infty$

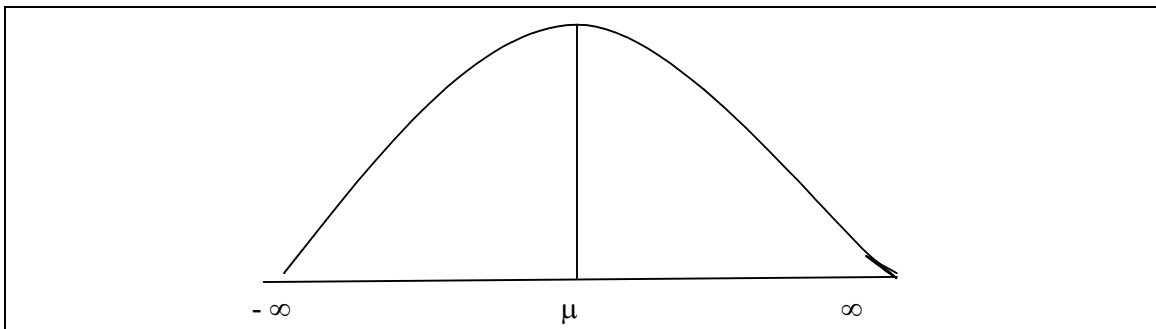
Z is called standard normal variate.

Notation : 1. $X \sim N(\mu, \sigma^2)$ denotes that X is a normal variate with mean μ and variance σ^2 .

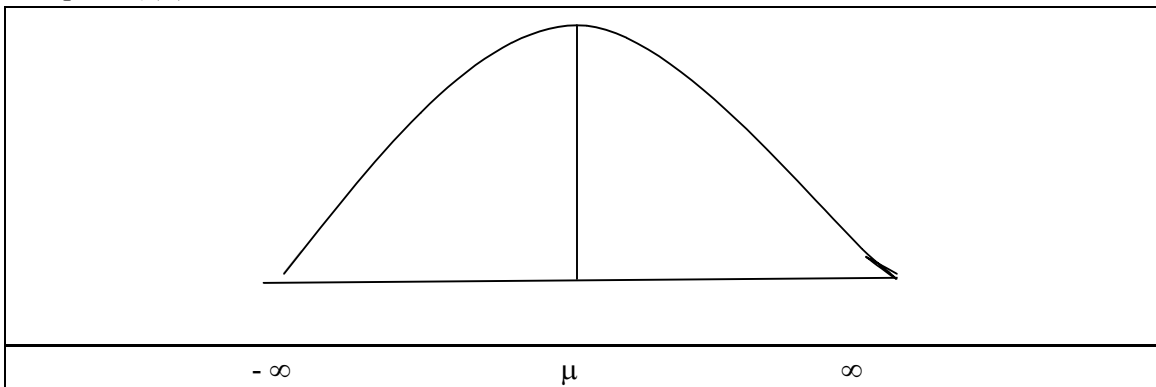
2. $Z \sim N(0, 1)$.

Features of Normal Distribution curve:

The graph of f(x) is a bell shaped curve extending from $-\infty$ to ∞ with its peak at μ .

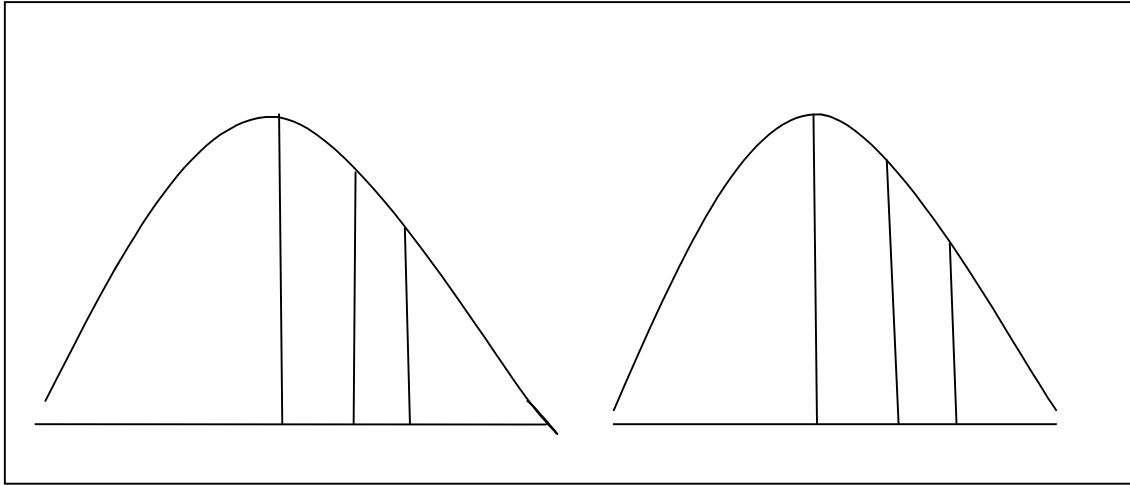


Graph of $\phi(Z)$:



Note 1. The mode of normal distribution is μ .

2. The median of normal distribution is also μ . Hence for a normal distribution the mean, median and mode coincide.



The area under the normal curve between the ordinates $x = a$ and $x = b$ gives the probability that the random variable X lies between a and b .

$$\therefore P(a < X < b) = \int_a^b f(x) dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$\text{Put } Z = \frac{x - \mu}{\sigma}$$

$$\text{So } dz = \frac{dx}{\sigma} \Rightarrow dx = \sigma \cdot dz$$

$$\text{When } x = a, z = \frac{a - \mu}{\sigma} = c \text{ (say)}$$

$$\text{When } x = b, z = \frac{b - \mu}{\sigma} = d \text{ (say)}$$

$$\therefore P(a < X < b) = \int_c^d \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} \sigma \cdot dz = \int_c^d \phi(z) dz = P(c < z < d).$$

So,

$$1. \sum_{-\infty}^{\infty} \phi(z) dz = 1$$

$$2. \int_{-\infty}^{\infty} \phi(z) dz = \int_0^{\infty} \phi(z) dz = 0.5$$

$$3. \int_{-a}^a \phi(z) dz = 2 \int_0^a \phi(z) dz$$

$$4. \int_{-a}^a \phi(z) dz = \int_0^a \phi(z) dz$$

The $\sum_0^z \phi(z) dz$ are available in the table 1.

Problem 1# Prove that normal distribution is a limiting form of Binomial distribution?

Problem 2# If 20% of the memory chips made in a certain plant are defective what are the probabilities that in a lot of 100 randomly chosen for inspection (i) at most 15 will be defective (ii) exactly 15 will be defective. [Ans. i) 0.1292 ii) 0.0454]

Problem 3# The mean weight of 500 male students at a certain college is 75 kg and the standard deviation is 7 kg. Assuming that the weights are normally distributed. Find how many students weigh (i) between 60 and 78 kg (ii) more than 92 kg.
[Ans. $0.4838+0.1664=0.6502$ ii) $0.5000-0.4925 = 0.0075$]

Problem 4# Find the probability of getting 3 and 6 heads inclusive in 10 tosses of a fair coin by using (i) Binomial distribution (ii) the normal approximation to the binomial distribution.
[Ans. 0.773 ; 0.6337]

Problem 5# If the masses of 300 students are normally distributed with mean 68.0 kg and standard deviation 3.0 kg, how many students have masses:
(i) 72 kgs (ii) ≤ 64 kgs (iii) $65 \leq X \leq 71$ kg inclusive
[Ans. i) 0.0918 28 students ii) 0.0918 28 students iii) 0.6826 205 students]

Chapter 2 Probability Distributions

Poisson's

Tutorial 11
By Dr. N V Nagendram

Problem 1# Define Poisson process with example and show that mean = variance for a Poisson distribution?

Solution: Definition: Poisson process: The Poisson process is the method of obtaining Poisson distribution independently without considering it as a limiting case of binomial distribution. It will be a Poisson distribution with parameter λt .

Example: 1. No. of telephones were Poisson process at a telephone exchange
2. No. of deaths due to heart attack or cancer.

To show that mean = variance in a Poisson distribution. For that Consider $\mu = E(X) =$

$$\sum_{x=0}^{\infty} x P(X=x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \cdot e^{\lambda}$$

$$\therefore \mu = \lambda$$

$$\text{Consider } E(X^2) = \sum_{x=0}^{\infty} x^2 P(X=x) = \sum_{x=0}^{\infty} (x^2 - x + x) P(X=x) = \sum_{x=0}^{\infty} [x(x-1) + x] P(X=x)$$

$$= \sum_{x=0}^{\infty} x(x-1) P(X=x) + \sum_{x=0}^{\infty} x P(X=x)$$

$$= \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda} \lambda^x}{x!} + \sum_{x=1}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = \lambda^2 e^{-\lambda} \sum_{x=2}^{\infty} \frac{e^{\lambda} \lambda^{x-2}}{(x-2)!} + \mu$$

$$= \lambda^2 e^{-\lambda} \cdot e^{\lambda} + \lambda$$

$$\therefore E(X^2) = \lambda^2 + \lambda \text{ and } \sigma^2 = V(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2$$

$$\therefore \sigma^2 = \lambda. \quad \therefore \mu = \sigma^2 \text{ i.e., mean = variance}$$

Hence the solution.

Problem 2# If the probability that an individual suffers a bad reaction due to a certain injection is 0.001, determine the probability that out of 2000 individuals (i) exactly 3 (ii) more than 2 individuals will suffer a bad reaction?

Solution: Given $p = 0.001$; $n = 2000$; $\lambda = np = 2$

$$(i) \text{ to find } P(\text{Exactly } 3) = P(X=3) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-2} 2^3}{3!} = \frac{4}{3e^2} = 0.1804 \text{ since } e=2.086, 2 < e < 3$$

$$(ii) P(\text{more than 2 individuals}) = P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X=0) + P(x=1) + P(x=2)]$$

$$= 1 - \left[\frac{e^{-\lambda} \lambda^0}{0!} + \frac{e^{-\lambda} \lambda^1}{1!} + \frac{e^{-\lambda} \lambda^2}{2!} \right]$$

$$= 1 - e^{-\lambda} \left[1 + \lambda + \frac{\lambda^2}{2} \right]$$

$$= 1 - 5e^{-2} = 0.323. \quad \text{Hence the solution.}$$

Problem 3# A manufacturer of cotter pins knows that 5% of his product is defective. If he sells cotter pins in boxes of 100 and guarantees that not more than 10 pins will be defective, what is the approximate probability that a box will fail to meet the guaranteed quality?

Solution: We are given $n = 100$, $p = \text{probability of defective pin} = 5\% = 0.05$

And $\lambda = \text{mean number of defective pins in a box of } 100 = np = 100 \times 0.05 = 5$

Since p is small, we may use Poisson distribution probability of 'x' defective pins in a box of

$$100 \text{ is } P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-5} 5^x}{x!} \text{ for all } x = 0, 1, 2, \dots$$

Probability that a box will fail to meet the guaranteed quality is $P(X > 10) = 1 - P(X \leq 10)$

$$\begin{aligned} &= 1 - \sum_{x=0}^{10} \frac{e^{-5} 5^x}{x!} \\ &= 1 - e^{-5} \sum_{x=0}^{10} \frac{5^x}{x!} \end{aligned}$$

Hence the solution.

Problem 4# 10% of the bolts produced by a certain machine turn out to be defective. Find the probability that in a sample of 10 tools selected at random exactly two will be defective using (i) binomial distribution (ii) Poisson distribution and comment upon the result?

Solution: Given $p = \frac{10}{100} = 0.1$, $n = 10$, $\lambda = np = 1$

(i) Using binomial distribution

$$\text{Let } q = 1 - p = 1 - 0.1 = 0.9$$

$$P(X=2) = {}^{10}C_2 p^2 q^{(n-2)} = \frac{10 \times 9}{1 \times 2} (0.1)^2 (0.9)^8 = 0.194$$

(ii) Using Poisson distribution

$$P(X=2) = \frac{e^{-\lambda} \lambda^2}{2!} = \frac{e^{-1} 1^2}{2} = \frac{1}{2e} = 0.184$$

Comment : There is a difference between the two probabilities because of the fact that Poisson distribution (P.D.) is an approximation to binomial distribution (B.D.) and it is applicable for large n . Hence the solution.

Problem 5# A hospital switch board receives an average of 4 emergency calls in a 10 min. interval. What is the probability that (i) there are at the most 2 emergency calls and (ii) there are exactly 3 emergency calls in a 10 min. interval?

Solution: Given $\lambda=4$, (i) $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = \frac{e^{-\lambda} \lambda^0}{0!} + \frac{e^{-\lambda} \lambda^1}{1!} + \frac{e^{-\lambda} \lambda^2}{2!} =$

$$e^{-\lambda} \left[1 + \lambda + \frac{\lambda^2}{2} \right] = e^{-4} [1 + 4 + 8] = 13 e^{-4} = 0.238.$$

(ii) $P(X=3) = \frac{e^{-\lambda} \lambda^3}{3!} = \frac{e^{-4} 4^3}{3!} = 0.195$ Hence the solution.

Chapter 2 Probability Distributions

Poisson's

Tutorial 12
By Dr. N V Nagendram

Problem 6# A rent a car firm has two cars which it hires from day to day. The number of demands for a car on each day is distributed as a Poisson variate with mean 1.5. Calculate the proportion of days on which (i) neither car is used (ii) some demand is refused?

Problem 7# In a Poisson distribution (P.D.), $P(X = 0) = 2 P(X = 1)$, then find $P(X = 2)$?

Problem 8# In a factory which turns out razor blades, there is a chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10 each. Using Poisson distribution, Calculate the approximate number of packets containing no defective, one defective and two defective blades if there are 10,000 such packets?

Problem 9# the probability of getting no misprint in a page of a book is e^{-4} . Determine the probability that a page of a book contains more than 2 misprints?

Problem 10# Obtain the Poisson distribution (P.D.) as a limiting case of Binomial distribution?

Problem 11# Fit a Poisson distribution to the following data and calculate the theoretical frequencies:

x	0	1	2	3	4
y	46	38	22	9	1

Solution: Mean $\mu = E(X) = \lambda$ and Variance $V(X) = \sigma^2 = E(X^2) - [E(X)]^2$

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$
0	46	0	0	0
1	38	38	1	38
2	22	44	4	88
3	9	27	9	81
4	1	4	16	16
	$\sum_{x=0}^4 f_i = 116N$	$\sum_{x=0}^4 f_i x_i = 113$		$\sum_{x=0}^4 f_i x_i^2 = 223$

$$\text{Mean} = \bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{113}{116} = 0.974;$$

$$\text{Variance} = \frac{\sum f_i x_i^2}{\sum f_i} - (\bar{x})^2 = \frac{223}{116} - (0.974)^2 = 1.9224 - 0.0487 = 0.974$$

$\therefore \text{Mean} = \text{Variance} = \lambda = 0.974.$

The theoretical frequencies are $f(x) = N \cdot P(X=x)$
 $f(0) = 116 \cdot P(X=0) = 116 \cdot e^{-0.974} = 44$

$$f(1) = 116 \cdot P(X=1) = 116 \cdot e^{-0.974} (0.974) = 42$$

$$f(2) = 116 \cdot P(X=2) = \frac{116 \cdot e^{-0.974} (0.974)^2}{2!} = 21$$

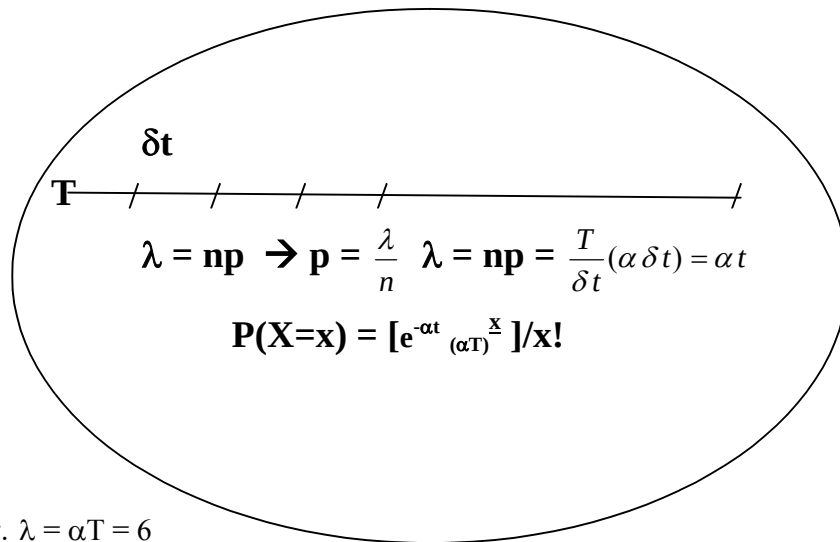
$$f(3) = 116 \cdot P(X=3) = \frac{116 \cdot e^{-0.974} (0.974)^3}{3!} = 7$$

$$f(4) = 116 \cdot P(X=4) = 116 - \{f(0) + f(1) + f(2) + f(3)\} = 116 - 114 = 2$$

Hence the solution.

Problem 12# If a bank receives on an average 6 bad cheques per day, what are the probabilities that it will receive (i) four bad cheques on any given day (ii) 10 bad cheques on any two consecutive days.

Solution: Let



$$\begin{aligned} & \delta t \\ & \text{---} / \text{---} / \text{---} / \text{---} / \text{---} \\ & \lambda = np \rightarrow p = \frac{\lambda}{n} \quad \lambda = np = \frac{T}{\delta t} (\alpha \delta t) = \alpha t \\ & P(X=x) = \frac{e^{-\alpha t} (\alpha t)^x}{x!} \end{aligned}$$

$$\alpha = 6, T = 1 \text{ and } \therefore \lambda = \alpha T = 6$$

$$f(4,6) = \frac{e^{-6} \cdot 6^4}{4!} = 0.1339$$

$$F(10; \lambda) = \frac{e^{-\lambda} \lambda^{10}}{10!} = \frac{e^{-12} 12^{10}}{10!} = 0.105$$

Hence the solution.

Chapter 2 Probability Distributions

Poisson's Process

Tutorial 13
By Dr. N V Nagendram

Problem 1# Fit a Poisson distribution to the following

x:	0	1	2	3	4
y:	46	38	22	9	1

Problem 2# Fit a Poisson distribution to the set of observations as below

x:	0	1	2	3	4
y:	122	60	15	2	1

Problem 3# The incidence of occupational disease in an industry is such that the workmen have a 10% chance of suffering from it. What is probability of 7, five or more will suffer from it?

Problem 4# A car hire firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson distribution with mean 1.5. calculate the proportion of days. (i) on which there is no demand (ii) on which demand is refused ($e^{-5} = 0.2231$)?
[Ans. i) 0.2231 ii) 0.1913]

Problem 5# If a random variable has a Poisson distribution such that $P(1) = P(2)$ find (i) mean of the distribution (ii) $P(4)$?
[Ans. i) 2 ii) $(2/3) \cdot e^{-2}$]

Problem 6# If the probability of a bad reaction from a certain injection is 0.001, determine the chance that out of 2,000 individuals more than two will get a bad reaction?[Ans. 0.32]

Problem 7 # If 3 % of the electric bulbs manufactured by a company are defective, find the probability that in a sample of 100 bulbs
(i) 0 (ii) 1 (iii) 4
[Ans. i) 0.04979 ii) 0.1494 iii) 0.1008]

Problem 8# Ten percent of the tools produced in a certain manufacturing process turn out to be defective. Find the probability that in a sample of 10 tools chosen at random exactly two will be defective by using the Poisson approximation to the binomial distribution?[Ans. 0.18]

Chapter 2 Probability Distributions

Normal Distributions

Tutorial 14
By Dr. N V Nagendram

Problem 1# Show that the mean deviation from the mean for normal distribution (N.Dⁿ) is equal to $4/5$ of standard deviation approximately? [Ans. M.D= $4/5\sigma$]

Problem 2# X is normally distributed with mean 12 and S.D = 4 then find (i) $P(0 \leq X \leq 12)$ (ii) $P(X \leq 20)$ (iii) $P(X \geq 20)$ (iv) if $P(X > C) = 0.24$.
[Ans. i) 0.4896 ii) 0.9772 iii) 0.0228 iv) 0.24 and C= 14.84]

Problem 3# Show that the mean deviation from the mean for the normal distribution [N.Dⁿ] is $4/5$ of standard deviation approximately. [Ans. $\eta = 0.79\sigma = 4/5\sigma$]

Problem 4# X is a normal variate with mean 30 and standard deviation 5. Find the probabilities that (i) $26 \leq X \leq 40$ (ii) $X \geq 45$. [Ans. i) $0.2882 + 0.4772 = 0.7653$ ii) 0.0013]

Problem 5# A random variable has normal distribution with $\mu = 62.4$. find its standard deviation if the probability is 0.20 that it will take on a value greater than 79.2. [Ans. $\sigma = 20$]

Problem 6# find the probabilities that a random variable having a standard normal distribution will take on a value (i) between 0.87 and 1.28 (ii) between -0.34 and 0.62 .
[Ans. i) 0.0919 ii) $0.1443 + 0.2343 = 0.3767$]

Problem 7# In a normal distribution (N.Dⁿ) 31% of the items are under 45 and 8% are over 63. Find the mean and variance of the distribution. [Ans. $\mu = 50, \sigma = 10$]

Problem 8# In a normal distribution (N.Dⁿ), 7% of the items are under 35 and 89% are over 64. Find the mean and variance of the distribution. [Ans. $\mu = 50.3, \sigma = 10.33$]

Sampling Distributions: Introduction

The field of statistics deals with the collection, presentation, analysis and use of data to make decision and solve problems. The main objective of any statistical study is to draw conclusions about a collection of objects under study. This collection is called the Population. Instead of examining this population, which may be difficult or impossible to do, one may arrive at the idea of examining only a small part of this population, which is called a sample. This can be done with the aim of drawing inferences about the population by using information from the sample, this process is known as statistical inference. The process of drawing samples is called sampling. A sample is a true or good representative of the population, if the sampling method is probabilistic. The most important of all probabilistic samplings is the random sampling, in which each member of the population has the equal chance of being included in the sample. Samples will be used to draw inferences about population, by estimating the parameters of population, such as mean (μ), standard deviation (σ) etc., Estimation of population parameters is possible only by studying some relevant statistical quantities computed from a sample of the population called sample statistics (or) simply statistic is often used for the random variable or for its value, the particular sense being clear from the context.

Let us consider all possible samples of a population and calculate a statistic for instance sample mean. Then the set of all such values, one for each sample, is called the sampling distribution of the statistic.

Now we can compute the statistics mean variance etc., for this sampling distribution.

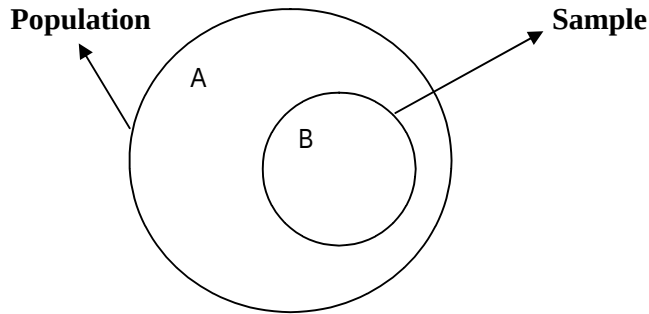
In most statistic problems, it is necessary to use the information from sample to draw inferences about the population.

Definition: Population

The population in a statistical study is the set or collection or totality of observations about which inferences are to be drawn. Thus the population consists of sets of numbers, measurements or observations. Population size N is the number of objects or observations in the population.

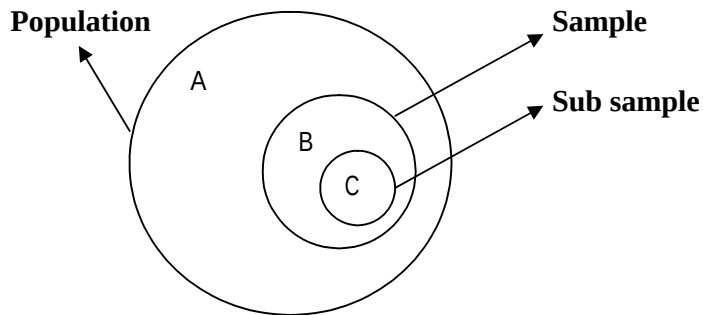
Population is said to be finite or infinite depending on the size N being finite or infinite. Since it is impracticable to examine the entire population, a finite subset of the population known as sample is studied. Sample size n is the number of objects or observations in the sample.

Example: (i) Engineering graduate students in A.P. (Population), Engineering graduate students of a college (Sample)



Example: Total production of items in a month (Population), Total production of items in one day (Sample)

Example: Budget of India (Population), Budget of A.P. (Sample), budget of a district (sub sample)



Definition: Population parameter:

A statistical measure or constant obtained from the population is called population parameter.

Example: population mean (μ), population variance (σ^2).

Definition: (Sample) statistic: A statistical measurement computed from sample observations is called a (sample) statistic.

Example: sample mean ($\bar{x}$), sample variance (s^2) clearly, parameters are to population while statistics are to sample.

μ , σ , p represent the population mean, population standard deviation, population proportion, similarly $\bar{X}$, s , p denote sample mean, sample standard deviation(s. s. d.), sample proportion.

Note: The samples must be a true or good representative of the population, sampling should be random or probabilistic.

Definition: Sampling: The process of drawing or obtaining samples is called sampling.

Definition: Large sampling: If $n \geq 30$, then the sampling is known as large sampling.

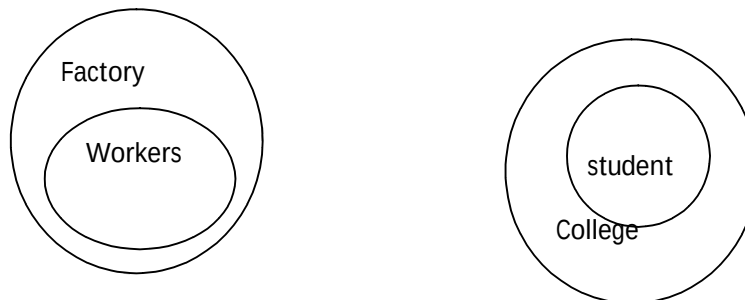
Definition: Small sampling: If $n < 30$, then the sampling is known as small or exact sampling.

Note: The simplest and most commonly used type of probabilistic sampling is the random sampling.

Definition: Random Sampling: Each member of the population has equal chances or probability of being included in the sample. The sample obtained by this method is termed as a random sample.

Definition: Finite Population: Population may be finite or infinite. If the number of items or observations consisting the population is fixed and limited, it is called as finite population.

Example: The workers in a factory, student in a college etc.,



Definition: Infinite Population: If the number of items or observations consisting the population is infinite (not fixed and not limited), it is called as infinite population.

Example: The population of all real numbers lying between 0 and 1. The population of stars or astral bodies in the sky.

Definition: Sampling with replacement: If the items are selected or drawn one by one such a way that an item drawn at a time is replaced back to the population before the next or subsequent draw, it is known as (random) sampling with replacement.

In this type of sampling from a population of size N , the probability of a selection of a unit at each draw remains $\frac{1}{N}$. Thus sampling from finite population with replacement can be considered theoretically as sampling from infinite population. In this, N^n samples will be drawn.

Definition: In Sampling without replacement:

An item of the population cannot be chosen more than once, as it is not replaced. In this N_{Cn} samples will be drawn. Hence the probability of drawing a unit from a population of N items at r^{th} draw is $\frac{1}{N-r+1}$.

Statistic is a real-valued function of the random sample. So it is a function of one or more random variables not involving any unknown parameter. Thus statistic is a function of samples observations only and is itself a random variable. Hence a statistic must have a probability distribution.

Definition: Sample mean: Let $x_1, x_2, x_3, \dots, x_n$ be a random, sample of size n from a

population. Then sample mean $= (\bar{x}) = \frac{\sum_{i=1}^n x_i}{n}$.

Definition: sample Variance: Then sample variance $= s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$.

Sample standard deviation is the positive square root of sample variance. Sample mean and sample variance are two important statistics which are statistical measures of a random sample of size n .

Chapter 2

Lecture 8

Sampling

Sampling Distributions

by Dr. N. V. Nagendram

Sampling Distribution:

Let us consider all possible samples of size n , from a finite population of size N . Then the total number of all possible samples of size n , which can be drawn from the population is $N_{Cn} = m$.

Compute a statistic θ [such as mean, variance /s.d, proportion] for each of these sample using the sample data $x_1, x_2, x_3, \dots, x_n$ by $\theta = \theta(x_1, x_2, x_3, \dots, x_n)$

Sample number	1	2	3	...	m
Statistic θ	θ_1	θ_2	θ_3	...	θ_m

Sampling distribution of the statistic θ is the set of values $\{\theta_1, \theta_2, \theta_3, \dots, \theta_m\}$ of the statistic θ Obtained, one for each sample. Thus sampling distribution describes how a statistic θ will vary from one sample to the other of the same size. Although all the m samples are drawn from the given population, the items included in different samples are different.

If the statistic θ is mean, then the corresponding distribution of the statistic is known as sampling distribution of means, thus if θ is variance, proportion etc., the corresponding distribution is known as sampling distribution of variances, sampling distribution of proportions etc.,

Then Mean of sampling distribution of $\theta = (\bar{\theta}) = \frac{\sum_{i=1}^m \theta_i}{m}$.

And Variance of sampling distribution of $\theta = \frac{\sum_{i=1}^m (\theta_i - \bar{\theta})^2}{m}$.

Similarly we can have mean of sampling distribution of means, variance of sampling distribution of means, variance of the sampling distribution of variances etc.,

Standard Error:

The standard deviation of the sampling distribution of a statistic is known as standard error (SE). The standard error gives some idea about the precision of the estimate of the parameters. As the sample size n increases, S.E. decreases. S.E. plays a very important role in large sample decision theory and forms the basis in hypothesis testing.

Sampling distribution of a statistic enables us to know information about the corresponding population parameter.

Degrees of freedom (v):

The number of degrees of freedom usually denoted by greek alphabet v , is a positive integer equals to $n - k$ where n is the number of independent observations of the random sample and k is the number of population parameters which are calculated using the sample data. The degrees of freedom $v = n - k$ is the difference between n the sample size and k the number of independent constraints imposed on the observations in the sample.

The sampling distribution of the Mean (σ known):

To answer any questions related to sampling distribution of the mean ($\bar{X}$) we need to consider a random sample of n observations and determine the value $\bar{x}$ for each sample, then by various values of $\bar{X}$, it may be possible to get an idea of the nature of the sampling distribution. Also we have to consider the following theorem for the mean $\mu_{\bar{x}}$ and the variance $\sigma_{\bar{x}}^2$ of sampling distribution of the mean ($\bar{X}$).

Theorem: If a random sample of size n is taken from a population having the mean μ and the variance σ^2 , then ($\bar{X}$) is a random variable whose distribution has the mean μ .

Proof: For samples from infinite population the variance of this distribution is $\frac{\sigma^2}{n}$.

For samples from finite population the variance of this distribution is $\frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$.

By above statement, population is infinite then sampling with replacement

$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

And when the population is finite, size N (sampling without replacement)

$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\left(\frac{N-n}{N-1} \right)}.$$

Note: The factor $\left(\frac{N-n}{N-1} \right)$ is known as finite population correction factor.

In sampling with replacement, we will have N^n samples each with probability $\frac{1}{N^n}$.

In sampling without replacement we will have N_{Cn} samples each with probability $\frac{1}{N_{Cn}}$.

Note: The factor $\left(\frac{N-n}{N-1} \right)$ can be neglected if N is too large compared to the sample size n .

Chapter 2

Lecture 9

Sampling

Sampling Distributions

by Dr. N. V. Nagendram

Central limit theorem:

Whenever n is large, the sampling distribution of $\bar{x}$ approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ regardless of the form of the parent population distribution, as the following theorem states [without proof]

Theorem: If $\bar{x}$ is the mean of a random sample of size n drawn from a population with mean μ and finite variance σ^2 then the standardized sample mean $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ is a random variable

whose distribution function approaches that of the standard normal distribution $N(0, 1)$ as $n \rightarrow \infty$.

Normal distribution provides a good approximation to the sampling distribution for almost all the populations for $n \geq 30$.

For $n < 30$ small samples, sampling distribution of $\bar{x}$ is normally distributed, provided sampling is from normal population.

Sampling distribution of proportions:

Suppose that a population is infinite and that the probability of occurrence of an event called its success is p , while the probability of non-occurrence of the event is $q = 1 - p$. Consider all possible samples of size N drawn from this population, and for each sample compute the proportion p of successes. Then, we can have a sampling distribution of proportions whose

mean μ_p and standard deviation σ_p are given by $\mu_p = p$ and $\sigma_p^2 = \frac{p(1-p)}{n} = \frac{pq}{n} \dots\dots\dots(1)$

While population is binomially distributed, the sampling distribution of proportion is normally distributed whenever n is large ≥ 30 . Equation (1) are also valid for a finite population in which sampling is with replacement.

For finite population sampling without replacement of size N

$$\mu_p = p \text{ and } \sigma_p^2 = \frac{pq}{n} \left(\frac{N-n}{N-1} \right)$$

Sampling distributions of differences and sums:

Let μ_{S_1} and σ_{S_1} be the mean and standard deviation of a sampling distribution of statistic S_1 obtained by calculating S_1 for all possible samples of size n_1 drawn from population 1. This yields a sampling distribution of the statistic S_1 .

In a similar manner, μ_{S_2} and σ_{S_2} be the mean and standard deviation of sampling distribution of statistic S_2 obtained by calculating S_2 for all possible samples of size n_2 drawn from another different population 2.

Now we can have a distribution of differences $S_1 - S_2$, called the sampling distribution of differences of the statistics, from the two population 1 and 2. Then the mean $\mu_{S_1 - S_2}$ and the standard deviation $\sigma_{S_1 - S_2}$ the sampling distribution of differences are given by

$$\mu_{S_1 - S_2} = \mu_{S_1} - \mu_{S_2}$$

and

$$\sigma_{S_1 - S_2} = \sqrt{(\sigma_{S_1}^2 + \sigma_{S_2}^2)} \text{ provided the samples are independent.}$$

Sampling distribution of sum of statistics has mean $\mu_{S_1 + S_2} = \mu_{S_1} + \mu_{S_2}$ and

$$\sigma_{S_1 + S_2} = \sqrt{(\sigma_{S_1}^2 + \sigma_{S_2}^2)} \text{ provided the samples are independent.}$$

For infinite population the sampling distribution of the differences of means has mean $\mu(\bar{X}_1 - \bar{X}_2)$ and $\sigma(\bar{X}_1 - \bar{X}_2)$ given by

$$\mu(\bar{X}_1 - \bar{X}_2) = \mu(\bar{X}_1) - \mu(\bar{X}_2) = \mu_1 - \mu_2 \text{ and}$$

$$\sigma(\bar{X}_1 - \bar{X}_2) = \sqrt{\sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

For infinite population the sampling distribution of sums of means has mean

$$\mu(\bar{X}_1 + \bar{X}_2) \text{ and } \sigma(\bar{X}_1 + \bar{X}_2) \text{ given by}$$

$$\mu(\bar{X}_1 + \bar{X}_2) = \mu(\bar{X}_1) + \mu(\bar{X}_2) = \mu_1 + \mu_2 \text{ and}$$

$$\sigma(\bar{X}_1 + \bar{X}_2) = \sqrt{\sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Sampling distribution of mean σ unknown: t-distribution:

To estimate or infer on a population mean or the difference between two population means, it was assumed that the population standard deviation σ is known. When σ is unknown, for large $n \geq 30$, σ can be replaced by the sample standard deviation s , calculated using the

sample mean $\bar{x}$ by the formula $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$.

For small sample of size $n < 30$ the unknown σ can be substituted by s , provided we make an assumption that the sample is drawn from a normal population.

A random variable having the t-distribution:

Let $\bar{x}$ be the mean of a random sample of size n drawn from a normal population with mean μ and variance σ^2 then $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ is a random variable having the t-distribution with $v = n - 1$

degrees of freedom. Where $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$.

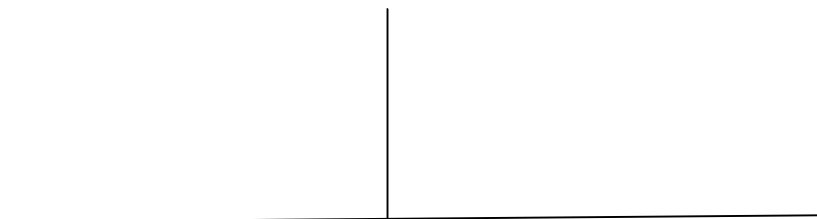
This result is more general than previous theorem CLT in the sense that it does not require knowledge of σ : on the other hand, it is less general than the previous theorem CLT in the sense that it requires the assumption of normal population.

Thus for all small samples $n < 30$ and with σ unknown a statistic for inference on population

mean μ is $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ With the underlying assumption of sampling from normal population.

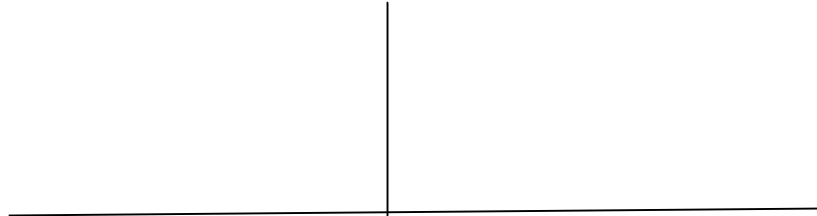
The t-distribution curve is symmetric about the mean 0, bell shaped and asymptotic on both sides of horizontal t-axis.

Thus t-distribution curve is similar to normal curve. The variance for the t-distribution is more than 1 as it depends on the parameter $v = n - 1$ degrees of freedom.



but it approaches 1 as $n \rightarrow \infty$. In essence, as $v = (n - 1) \rightarrow \infty$, t-distribution tends to the standard normal distribution. Clearly for $n \geq 30$, standard normal distribution provides a good approximation to the t-distribution.

Critical values of t-distribution is denote by t_{α} , which is such that the area under the curve to the right of t_{α} equals to α . Since the t-distribution is symmetric, it follows that $t_{1-\alpha} = -t_{\alpha}$ **i.e.**, the t-value leaving an area of $1 - \alpha$ to the right and therefore an area α to its left, is equal to the negative t-value which leaves an area α in the right tail of the distribution.



Please observe critical values of t_{α} for values of the parameter v . In tables the left-hand column contains values of v , the column headings are area α in the right hand tail of the t-distribution, the entries are values of t_{α} .

Definition: χ^2 (chi squared) distribution is a continuous probability distribution of a c.r.v. X

with probability density function given by

$$f(x) = \begin{cases} \frac{1}{2^{v/2} \Gamma(v/2)} x^{v/2-1} e^{-x/2}, & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

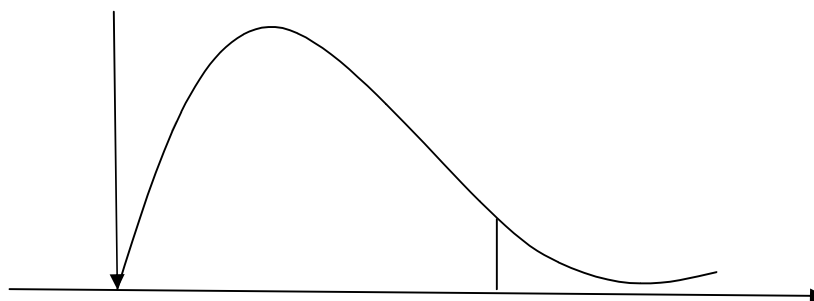
Where v is a +ve integer is the only single parameter of the distribution, also known as degrees of freedom.

Properties of χ^2 - Distribution:

(i) χ^2 - Distribution curve is not symmetrical, lies entirely in the first quadrant. And hence not a normal curve, since χ^2 varies from 0 to ∞ .

(ii) It depends only on the degrees of freedom v .

(iii) If X_1^2 and X_2^2 are two independent distributions with v_1, v_2 degrees of freedom then $\chi_1^2 + \chi_2^2$ will be chi- squared distributions with $(v_1 + v_2)$ degrees of freedom – i.e, it is additive.



Hence α denotes the area under the chi-squared distribution to the right of χ_{α}^2 .

So χ_{α}^2 represents the χ^2 -value such that the area under the χ^2 -curve to its right is equal to α .

χ^2 -distribution is very important in estimation and hypothesis testing χ^2 -distribution is used in sampling distributions, analysis of variance mainly, it is used as a measure of goodness of fit and in analysis of $r \times c$ tables.

For various values of α and v , the values of χ_{α}^2 are tabulated.

In χ^2 - table the left-hand column contains values of v (degrees of freedom), the column headings are areas α in the right hand tail of χ^2 -distribution curve, the entries are χ^2 - values. It is necessary to calculate values of χ_{α}^2 for $\alpha > 0.50$, since χ_{α}^2 curve or distribution is not symmetrical.

Sampling distribution of Variance s^2 :

From the earlier discussions, the sample mean is used to estimate the population mean.

Similarly, the sample variance is used to estimate the population variance (σ^2). The sample

variance is usually denoted by s^2 and is given by $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

A random variable having the χ^2 -distribution:

Theorem: If s^2 is the variance of a random sample of size n from a normal population having

the variance σ^2 then $\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$ is a random variable having the χ^2 -distribution with $v = n - 1$ dof.

Exactly 95% of χ^2 -distribution lies between $\chi^2_{0.975}$ and $\chi^2_{0.025}$ when σ^2 is too small. χ^2 -value falls to the right of $\chi^2_{0.025}$ and when σ^2 is too large, χ^2 falls to the left of $\chi^2_{0.975}$. thus when σ^2 is correct χ^2 -value falls to the left of $\chi^2_{0.975}$ or to the right of $\chi^2_{0.025}$.

critical region for testing: $H_0: \sigma^2 = \sigma_0^2$

Alternate hypothesis	Reject H_0 if
$\sigma^2 < \sigma_0^2$	$\chi^2 \leq \chi^2_{1-\alpha}$
$\sigma^2 > \sigma_0^2$	$\chi^2 \geq \chi^2_{\alpha}$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 \leq \chi^2_{(1-\alpha)/2}$ or $\chi^2 \geq \chi^2_{\alpha/2}$

F-Distribution (sampling distribution of the ratio of two sample variances):

Definition: Another important continuous probability distribution which plays an important role in connection with sampling from normal population is the F-distribution.

If s_1^2 and s_2^2 are the variances of independent random samples of size n_1 and n_2 from normal populations with variances σ_1^2 and σ_2^2 .

To determine whether the two samples come from two populations having equal variances, consider the sampling distribution of the ratio of the variances of the two independent random

samples defined by $F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} = \frac{\sigma_2^2 s_1^2}{\sigma_1^2 s_2^2}$ which follows F-distribution with $v_1 = n_1 - 1$ and

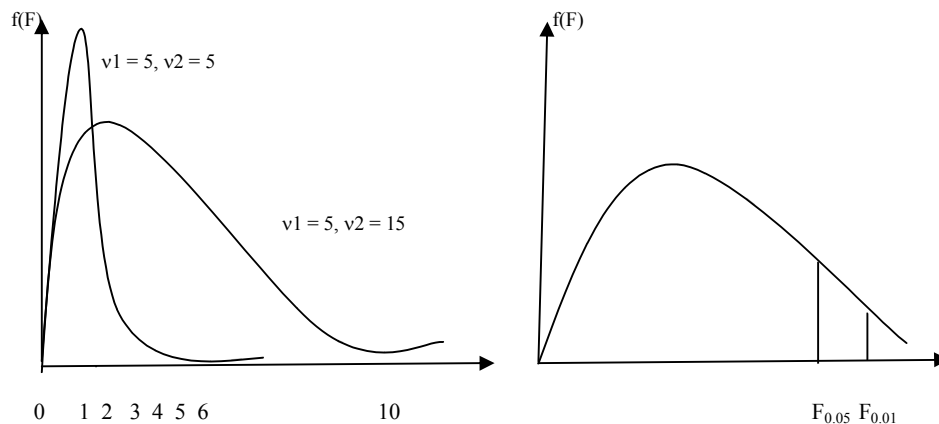
$v_2 = n_2 - 1$ degrees of freedom.

Uses: F-distribution can be used for testing the quality of several population means, comparing sample variances, and analysis of variance completely depends on F-distribution.

Under the hypothesis that two normal populations have the same variance : $\sigma_1^2 = \sigma_2^2$, we

have $F = \frac{s_1^2}{s_2^2}$.

F determines whether the ratio of two sample variances s_1 and s_2 is too small or too large. When F is close to 1, the two sample variances s_1 and s_2 are almost same. F is always a positive number whenever the larger sample variance as the numerator.



Probability density functions of several F distributions

Tabulated values of F

Properties of F-distribution:

(i) F-distribution curve lies entirely in first quadrant.

(ii) The F-curve depends not only on the two parameters ν_1 and ν_2 but also on the order in which they are stated.

(iii) $F_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{F_{\alpha}(\nu_2, \nu_1)}$ where $F_{\alpha}(\nu_2, \nu_1)$ is the value of F with ν_1 and ν_2 degrees of

Freedom such that the area under the F-distribution curve to the right of F_{α} is α .

Note:

Critical regions for testing the null hypothesis: $\sigma_1^2 = \sigma_2^2$

Alternate hypothesis	Test statistic	Reject H_0 if:
$\sigma_1^2 < \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$	$F > F_{\alpha}(n_2 - 1, n_1 - 1)$
$\sigma_1^2 > \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$	$F > F_{\alpha}(n_1 - 1, n_2 - 1)$
$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{s_M^2}{s_m^2}, s_M^2 > s_m^2$	$F > F_{\alpha/2}(n_M - 1, n_m - 1)$

Chapter 2 Probability Distributions

Sampling - Population PROBLEMS

Tutorial 15
by Dr. N V Nagendram

Problem 1# Find the value of the finite population correction factor for (i) $n = 10$ and $N = 1000$ (ii) $n = 100$ and $N = 1000$?

Problem 2# A random sample of size 2 is drawn from the population 3,4,5. Find (i) population mean (ii) Population S.D. (iii) Sampling distribution (SD) of means (iv) the mean of SD of means (v) S.D of SD means?

Problem 3# A random sample of size 2 is drawn from the population 3,4,5. Find (i) population mean (ii) Population S.D. (iii) Sampling distribution (SD) of means (iv) the mean of SD of means (v) S.D of SD means? Solve the problem without replacement?

[Ans.0.4082]

Problem 4# Determine the mean and s.d of sampling distributions of variances for the population 3,7,11,15 with $n = 2$ and with sampling (i) with replacement and (ii) without replacement?

[Ans. 11.489]

Problem 5# Find $P(\bar{X} > 66.75)$ if a random sample size 36 is drawn from an infinite population with mean $\mu = 63$ and s.d. $\sigma = 9$.

[Ans. 0.0062]

Problem 6# Determine the probability that mean breaking strength of cables produced by company 2 will be (i) at least 600N more than (ii) at least 450 N more than the cables produced by company 1, if 100 cables of brand 1 and 50 cables of brand 2 are tested.

company	Mean breaking strength	s.d.	Sample size
1	4000 N	300 N	100
2	4500 N	200 N	50

[Ans. 0.8869]

Problem 7# Let $\bar{X}_1$ and $\bar{X}_2$ be the average drying time of two types of oil paints 1 and 2 for samples size $n_1 = n_2 = 18$. Suppose $\sigma_1 = \sigma_2 = 1$. Find the value of $P(\bar{X}_1 - \bar{X}_2 > 1)$, assuming that mean drying time is equal for the two types of oil paints.

[Ans. 0.0013]

Problem 8# A company claims that the mean life time of tube lights is 500 hours. Is the claim of the company tenable if a random sample of 25 tube lights produced by th company has mean 518 hours and s.d. 40 hours.

[Ans. 2.492]

Problem 9# Determine the probability that the variance of the first sample of size $n_1 = 9$ will be at least 4 times as large as the variance of the second sample of size $n_2 = 16$ if the two samples are independent random samples from a normal population.

[Ans. 0.01]

Problem 10# Is there reason to believe that the life expected of group A and Group B is same or not from the following data

GroupA	34	39.2	46.1	48.7	49.4	45.9	55.3	42.7	43.7	56.6
Group B	49.7	55.4	57.0	54.2	50.4	44.2	53.4	57.5	61.9	58.2

[Ans. 1.63]

Problem 11# A random sample of size 25 from a normal population has the mean $\bar{x}=47.5$ and the standard deviation $s = 8.4$. does this information tend to support of refute the claim that the mean of the population is $\mu = 42.1$?
[Ans. $t=3.21$]

Problem 12# In 16 hour ten runs, the gasoline consumption of an engine averaged 16.4 gallons with a. s. d. of 2.1 gallons. Test the claim that the average gasoline consumption of this engine is 12.0 gallons per hour.
[Ans. $t=8.38$]

Problem 13# Suppose that the thickness of a part used in a semiconductor is its critical dimension, and that process of manufacturing these parts is considered to be under control if the true version among the thickness of the parts is given by a standard deviation not greater than $\sigma = 0.60$ thousandth of an inch. To keep a check on the process, random samples of size $n = 20$ are taken periodically, and is regarded to be “out of control” if the probability that s^2 will take on a value greater than or equal to the observed sample value is 0.01 or less even though $\sigma = 0.60$ what can one conclude about the process if the standard deviation of such a periodic random sample is $s = 0.84$ thousandth of an inch?
[Ans.37.24]

Problem 14# A soft-drink vending machine is set so that the amount of drink dispensed is a random variable with a mean of 200 millilitres and a standard deviation of 15 millilitres'. What is the probability that the average (mean) amount dispensed in a random sample size of 36 at least 204 millilitres?

Problem 15# If two independent random sample of size $n_1 = 7$ and $n_2 = 13$ are taken from a normal population what is the probability that the variance of the first sample will be at least three times as large that of the second sample?

Problem 16# The claim that the variance of a normal population is $\sigma^2 = 21.3$ is rejected if the variance of a random sample of size 15 exceeds 39.74. What is the probability that the claim will be rejected even though $\sigma^2 = 21.3$?
[Ans.0025]

Problem 17# An electronic company manufactures resistors that have a mean resistance of 100 Ω and a standard deviation of 10 Ω . The distribution of resistance is normal. Find the probability that a random sample 25 resistors will have an average resistance less than 95 Ω ?
[Ans. 0.0062]

Problem 18# The mean voltage of a battery is 15 volt and s.d.is 0.2 volt. What is the probability that four such batteries connected in series will have a combined voltage of 60.8 or more volts? [Ans. 0.0228]

Problem 19# Certain ball bearings have a mean weight of 5.02 ounces and standard deviation of 0.30 ounces. Find the probability that a random sample of 100 ball bearings will have a combined weight between 496 and 500 ounces? [Ans. 0.2318]

Problem 20# A manufacturer of fuses claims that with a 20% overload, the fuses will blow in 12.40 minutes on the average. To test the claim, a sample of 20 of the fuses was subjected to a 20% overload, and the times it took them to blow had a mean of 10.63 minutes and a s.d. of 2.48 minutes. If it can be assumed that the data constitute a random sample from a normal population, do they tend to support or refute the manufacturer's claim? [Ans.- 3.19]

Problem 21# show that for random samples of size n from a normal population with the variance σ^2 , the sampling distribution of σ^2 has the mean σ^2 and the variance $\frac{2\sigma^4}{n-1}$?

Problem 22# If S_1^2 and S_2^2 are the variances of independent random samples of size $n_1 = 10$ and $n_2 = 15$ from normal population with equal variances find $P(S_1^2 / S_2^2 < 4.03)$? [Ans. 0.99]

Problem 23# A random sample of size $n = 25$ from a normal population has the mean $\bar{X} = 47$ and the standard deviation $\sigma = 7$. If we base our decision on the statistic, can we say that the given information supports the conjecture that the mean of the population is $\mu = 42$?

Problem 24# The claim that the variance of a normal population is $\sigma^2 = 4$ is to be rejected if the variance of a random sample of size 9 exceeds 7.7535. What is the probability that this claim will be rejected even though $\sigma^2 = 4$? [Ans. 0.5]

Problem 25# A random sample of size $n = 12$ from a normal population $\bar{x} = 27.8$ has the mean and the variance $\sigma^2 = 3.24$. If we base our decision on the statistic can we say that the given information supports the claim that the mean of the population is $\mu = 28.5$? [Ans.-1.347]

Problem 26# The distribution of annual earnings of all bank letters with five years experience is skewed negatively. This distribution has a mean of Rs.19000 and a standard deviation of Rs.2000. If we draw a random sample of 30 tellers, what is the probability that the earnings will average more than Rs.19750 annually? [Ans. 0.0202]

Problem 27# If a gallon can of paint covers on the average 513.3 square feet(Ft^2) with a standard deviation(s.d.) of 31.5 square feet(Ft^2). what is the probability that the mean area covered by a sample of 40 of these 1 gallon cans will be anywhere from 510 to 520 square feet(Ft^2)? [Ans.0.6553]

Problem 28# A random sample of 100 is taken from an infinite population having the mean $\mu = 76$ and the variance $= \sigma^2 = 256$. Find the probability that $\bar{X}$ will be between 75 and 78?
[Ans. 0.6268]

Problem 29# If two independent random samples of size $n_1 = 13$ and $n_2 = 7$ are taken from a normal population. What is the probability that the variance of the first sample will be atleast four times as that of the second sample?
[Ans. 4.00]

Problem 30# If two independent random samples of size $n_1 = 26$ and $n_2 = 8$ are taken from a normal population. What is the probability that the variance of the second sample will be atleast 2.4 times as that of the first sample?
[Ans. 0.05]

Problem 31# If the actual amount of instant coffee which a filling machine puts into “6-ounce” jars is r. v. having a normal distribution with s.d. 0.05 ounce and if only 3% of the jars are to contain less than 6 ounces of coffee, what must be the mean fill of these jars?
[Ans. $\mu=6.094$]

Problem 32# A manufacturer of a certain type of synthetic fishing line has found from long experience of testing that the breaking strength of his product has an approximate normal distribution with a mean of 30 pounds(lbs.) and a standard deviation of 4 pounds(lbs.). A time and money saving change in the manufacture process of the product is tried. A sample of 25 testing length pieces of the new process line is taken and tested with a resulting sample mean of 28 pounds(lbs.) What is the probability of obtaining a mean as low as 28 if the process has had no harmful effect on breaking strength?
[Ans. 0.006]

Problem 33# An Urn contains 1000 white and 2000 black balls. If X denotes the number of white balls when 300 balls are drawn without replacement, then find $P(180 < X < 120)$?
[Ans. 0.9858]

Problem 34# Two movie theatres compete for 900 visitors. Suppose each visitor chooses one of the two balls independent of the choice of the other visitors; how many seats should each theatre have so that the probability of turning away any visitor for lack of seats is less than 1%?
[Ans. 489]

Problem 35# Let X be a random variable where μ_x is unknown as $\sigma_x^2 = 0.25$ i.e., 1/4 Find out how large a random sample must be taken in order that the probability will be at test 0.95 and the sample mean $\bar{x}$ will lies within 0.25 of the population mean?
[Ans. 80]

Problem 36# If a random sample of size n is selected from the finite population that consists of the integers $1, 2, 3, \dots, N$ show that (i) the mean $\bar{X}$ is $\frac{N+1}{2}$ (ii) the variance of $\bar{X}$ is $\frac{(N+1)(N-n)}{12n}$ (iii) the mean and the variance of $Y = n \bar{X}$ are $E(Y) = \frac{n(N+1)}{2}$ and the $\text{var}(Y) = \frac{n(N+1)(N-n)}{12}$?

Problem 37# How many different samples of size $n=3$ can be drawn from a finite population of size (a) $N=12$ (b) $N=20$ (c) $N=50$ [Ans. a) 220, b) 1140 c) 19600]

Problem 38# What is the probability of each possible sample if (i) a random sample of size $n=4$ is to be drawn from a finite population of size $N=12$ (ii) a random sample of size $n=5$ is to be drawn from a finite population of size $N=22$? [Ans. a) $1/495$ b) $1/77$]

Problem 39# Independent random samples of size $n_1 = 30$ and $n_2 = 50$ are taken from two normal populations having the means $\mu_1 = 78$ and $\mu_2 = 78$ and the variances σ_1^2 and σ_2^2 . Find the probability that the mean of the first sample will exceed that of the second sample by at least 4.8? [Ans. 0.2743]

Problem 40# If S_1 and S_2 are the variances of independent random samples of size $n_1 = 61$ and $n_2 = 31$ from normal population with $\sigma_1^2 = 12$ and $\sigma_2^2 = 18$ Find $P\left(\frac{S_1^2}{S_2^2} > 1.16\right)$ [Ans. 0.05]

Chapter 2 Probability Distributions

Sampling - Population

Tutorial 15
by Dr. N V Nagendram

Problem 1# Find the value of the finite population correction factor for (i) $n = 10$ and $N = 1000$ (ii) $n = 100$ and $N = 1000$?

Solution: (i) $\frac{N-n}{N-1} = \frac{1000-10}{1000-1} = \frac{990}{999} = 0.991$

(ii) $\frac{N-n}{N-1} = \frac{1000-100}{1000-1} = \frac{900}{999} = 0.900$

Hence the solution.

Problem 2# A random sample of size 2 is drawn from the population 3,4,5. Find (i) population mean (ii) Population S.D. (iii) Sampling distribution (SD) of means (iv) the mean of SD of means (v) S.D of SD means?

Solution:

(i) Population mean $= \mu = \frac{3+4+5}{3} = 4$

(ii) s.d. of population $= \sigma = \sqrt{\frac{(3-4)^2 + (4-4)^2 + (5-4)^2}{3}} = \sqrt{\frac{2}{3}} = \sqrt{0.6666} = 0.8164$

(iii) sampling with replacement (infinite population): The total number of samples with replacement is $N^n = 3^2 = 9$ here N = population size and n = sample size. Listing all possible samples of size 2 from population 3,4,5 with replacement, we get 9 samples as below:

(3,3) (3,4) (3,5)

(4,3) (4,4) (4,5)

(5,3) (5,4) (5,5)

Now compute the statistic the arithmetic mean for each of these 9 samples the set of 9 samples means $\bar{X}$, gives rise to the distribution of means of the sample known as sampling distribution of means

3 3.5 4
3.5 4 4.5
4 4.5 5

This sampling distribution of means can also be arranged in the form of frequency distribution

Sample mean $\bar{X}_i$	3	3.5	4	4.5	5
Frequency f_i	1	2	3	2	1

(iv) Mean of the sampling distribution of means =

$$\mu_{\bar{X}} = \frac{3 + 2(3.5) + 3(4) + 2(4.5) + 5}{9} = 36/9 = 4$$

Showing $\mu_{\bar{X}} = \mu = 4$

$$(v) \sigma^2_{\bar{X}} = \frac{(3-4)^2 + 2(3.5-4)^2 + 3(4-4)^2 + 2(4.5-4)^2 + (5-4)^2}{9} = \frac{3}{9} = \frac{1}{3}$$

therefore $\sigma_{\bar{X}} = 0.5773$

Problem 3# A random sample of size 2 is drawn from the population 3,4,5. Find (i) population mean (ii) Population S.D. (iii) Sampling distribution (SD) of means (iv) the mean of SD of means (v) S.D of SD means? Solve the problem without replacement?

[Ans.0.4082]

Solution:

$$(i) = 4 \quad (ii) \sigma = 0.8164$$

(iii) Sampling without replacement finite population the total number of samples without replacement is $N_{cn} = {}^3C_2 = 3$ the three samples are (3,4), (3,5) (4,5) and their means are 3.5, 4, 4.5

$$(iv) \mu_{\bar{X}} = \text{mean of sampling distribution of means} = \frac{(3.5) + 4 + 4.5}{3} = \frac{12}{3} = 4 = \mu$$

$$(iv) \sigma^2_{\bar{X}} = \frac{(3.5-4)^2 + (4-4)^2 + (4.5-4)^2}{3} = \frac{2 \cdot (0.5)^2}{3}$$

$$\sigma_{\bar{X}} = 0.4082.$$

Hence the solution.

Problem 4# Determine the mean and s.d of sampling distributions of variances for the population 3,7,11,15 with $n = 2$ and with sampling (i) with replacement and (ii) without replacement?

[Ans. 11.489]

Solution: (i) $N^n = 4^2 = 16$ samples

(3,3),(3,7), . . . , (15,11), (15,15)

With

Means	3	5	7	9	11	13	15
Frequency	1	2	3	4	3	2	1
Variances	0	4	16	36			

$$\mu_{s^2} = 10; \sigma^2_{s^2} = \sqrt{\frac{2112}{16}} = 11.489$$

Hence the solution.

Problem 5# Find $P(\bar{X} > 66.75)$ if a random sample size 36 is drawn from an infinite population with mean $\mu = 63$ and s.d. $\sigma = 9$. [Ans. 0.0062]

Solution: let $z = \frac{66.75 - 63}{9/\sqrt{36}} = 2.5$ Hence $P(\bar{X} > 66.75) = P(Z > 2.50) = 0.0062$.

Hence the solution.

Problem 6# Determine the probability that mean breaking strength of cables produced by company 2 will be (i) at least 600N more than (ii) at least 450 N more than the cables produced by company 1, if 100 cables of brand 1 and 50 cables of brand 2 are tested.

company	Mean breaking strength	s.d.	Sample size
1	4000 N	300 N	100
2	4500 N	200 N	50

[Ans. 0.8869]

Solution: $\mu(\bar{X}_2 - \bar{X}_1) = \mu(\bar{X}_2) - \mu(\bar{X}_1) = 4500 - 4000 = 500$ N

$$\sigma(\bar{X}_2 - \bar{X}_1) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(200)^2}{50} + \frac{(300)^2}{100}} = \sqrt{1700} = 41.23$$

$$(i) P(\bar{X}_2 - \bar{X}_1 > 600) = P(Z > \frac{600 - 500}{41.23}) = P(Z > 2.4254) = 0.0078$$

$$(ii) P(\bar{X}_2 - \bar{X}_1 > 450) = P(Z > \frac{450 - 500}{41.23}) = P(Z > -1.2127) = 0.8869.$$

Hence the solution.

Problem 7# Let $\bar{X}_1$ and $\bar{X}_2$ be the average drying time of two types of oil paints 1 and 2 for samples size $n_1 = n_2 = 18$. Suppose $\sigma_1 = \sigma_2 = 1$. Find the value of $P(\bar{X}_1 - \bar{X}_2 > 1)$, assuming that mean drying time is equal for the two types of oil paints. [Ans. 0.0013]

$$\text{Solution: } \sigma^2(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}$$

$$P(\bar{X}_1 - \bar{X}_2) = P(Z > \frac{1 - (\mu_1 - \mu_2)}{\sigma(\bar{X}_1 - \bar{X}_2)}) = P(Z > \frac{1}{\sqrt{1/9}}) = P(Z > 3) = 1 - 0.9987 = 0.0013$$

Hence the solution.

Problem 8# A company claims that the mean life time of tube lights is 500 hours. Is the claim of the company tenable if a random sample of 25 tube lights produced by the company has mean 518 hours and s.d. 40 hours. [Ans. 2.492]

Solution: Given $\bar{x} = 518$ hrs. $n = 25$, $s = 40$, $\mu = 500$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{518 - 500}{40/\sqrt{25}} = 2.25 \text{ since, } t = 2.25 < t_{0.01, v=24} = 2.492$$

Accept the claim of the company. Hence the solution.

Problem 9# Determine the probability that the variance of the first sample of size $n_1 = 9$ will be at least 4 times as large as the variance of the second sample of size $n_2 = 16$ if the two samples are independent random samples from a normal population. [Ans. 0.01]

Solution: From table $F_{0.01} = 4$ for $v_1 = n_1 - 1 = 9 - 1$
 $v_2 = n_2 - 1 = 16 - 1 = 15$, the desired probability is 0.01 [from $F_{0.01}$ tables]
Hence the solution.

Problem 10# Is there reason to believe that the life expected of group A and Group B is same or not from the following data

GroupA	34	39.2	46.1	48.7	49.4	45.9	55.3	42.7	43.7	56.6
Group B	49.7	55.4	57.0	54.2	50.4	44.2	53.4	57.5	61.9	58.2

[Ans. 1.63]

Solution: Given data $S^2_A = \frac{1}{8} \left(18527.78 - \frac{(405)^2}{9} \right) = 37.848$

$$S^2_B = \frac{1}{10} \left(32799.918 - \frac{(598.5)^2}{11} \right) = 23.607$$

$$F = \frac{S^2_A}{S^2_B} = \frac{37.848}{23.607} = 1.63 \text{ clearly, variances expectancy is same for}$$

Group A and Group B. Hence the solution.

Problem 11# A random sample of size 25 from a normal population has the mean $\bar{x} = 47.5$ and the standard deviation $s = 8.4$. does this information tend to support or refute the claim that the mean of the population is $\mu = 42.1$? [Ans. $t = 3.21$]

Solution: given $n = 25$, $\bar{x} = 47.5$, $\mu = 42.1$, $s = 8.4$ we have from t-distribution

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{47.5 - 42.1}{8.4 / \sqrt{25}} \text{ (or) } t = 3.21. \text{ This value of } t \text{ has 24 degrees of freedom. From the}$$

table of t-distribution for $v = 24$, we get probability that t will exceed 2.797 is 0.005. Then the probability of getting a value greater than 3.21 is negligible. Hence we conclude that the information given in the data of this example tend to refute the claim that the mean of the population is $\mu = 42.1$. Hence the solution.

Problem 12# In 16 hour ten runs, the gasoline consumption of an engine averaged 16.4 gallons with a. s. d. of 2.1 gallons. Test the claim that the average gasoline consumption of this engine is 12.0 gallons per hour. [Ans. $t=8.38$]

Solution: substituting $n = 16$, $\mu=12.0$, $\bar{x} = 16.4$ and $s = 2.1$ into the formula for $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{16.4 - 12.0}{2.1/\sqrt{16}} = 8.38$, but from the table for $v = 15$ the probability of getting a value

of t greater than 2.947 is 0.005. the probability of getting a value greater than 8 must be negligible. Thus, it would seem reasonable to conclude that the true average hourly gasoline consumption of the engine exceeds 12.0 gasoline. Hence the solution.

Problem 13# Suppose that the thickness of a part used in a semiconductor is its critical dimension, and that process of manufacturing these parts is considered to be under control if the true version among the thickness of the parts is given by a standard deviation not greater than $\sigma = 0.60$ thousandth of an inch. To keep a check on the process, random samples of size $n = 20$ are taken periodically, and is regarded to be “out of control” if the probability that s^2 will take on a value greater than or equal to the observed sample value is 0.01 or less even though $\sigma = 0.60$ what can one conclude about the process if the standard deviation of such a periodic random sample is $s = 0.84$ thousandth of an inch? [Ans. 37.24]

Solution: The process will be declared “out of control” if $\frac{(n-1)s^2}{\sigma^2}$ with $n = 20$ and $\sigma = 0.60$

exceeds $\chi^2_{0.01,19} = 36.91$, since $\frac{(n-1)s^2}{\sigma^2} = \frac{19(0.84)^2}{(0.60)^2} = 37.24$ exceeds 36.91, the process is

declared out of control. Of course it is assumed here that the sample may be regarded as a random sample from a normal population. Hence the solution.

Problem 14# A soft-drink vending machine is set so that the amount of drink dispensed is a random variable with a mean of 200 millilitres and a standard deviation of 15 millilitres'. What is the probability that the average (mean) amount dispensed in a random sample size of 36 at least 204 millilitres?

Solution: The distribution of $\bar{X}$ has the mean $\mu(\bar{X}) = 200$ and the standard deviation $\sigma(\bar{X}) = \frac{15}{\sqrt{36}} = 2.5$, and according to the central limit theorem, this distribution is

approximately normal. And $Z = \frac{204 - 200}{2.5} = 1.6$.

Then $P(\bar{x} \geq 204) = P(Z \geq 1.6) = 0.5000 - 0.4452 = 0.0548$ Hence the solution.

Problem 15# If two independent random sample of size $n_1 = 7$ and $n_2 = 13$ are taken from a normal population what is the probability that the variance of the first sample will be at least three times as large that of the second sample?

Solution: $F_{0.05}(v_1 = 6, v_2 = 12) = 3$ thus the desired probability is 0.05. Hence the solution.

Problem 16# The claim that the variance of a normal population is $\sigma^2 = 21.3$ is rejected if the variance of a random sample of size 15 exceeds 39.74. What is the probability that the claim will be rejected even though $\sigma^2 = 21.3$? [Ans.0025]

Solution: $n = 15, \sigma^2 = 21.3, s^2 = 39.74$

$$\chi^2 = \frac{14(39.74)}{21.3} = 26.120$$

And $\chi^2_{0.025, 14} = 26.119$

$$\chi^2 > \chi^2_{0.05/2, 14} (26.119) = 0.025$$

Therefore, probability that the claim will be rejected is 0.0025. Hence the solution.

Problem 17# An electronic company manufactures resistors that have a mean resistance of 100 Ω and a standard deviation of 10 Ω . The distribution of resistance is normal. Find the probability that a random sample 25 resistors will have an average resistance less than 95 Ω ?

[Ans. 0.0062]

Solution: $n = 25, \mu = 100 \Omega, \sigma = 10 \Omega$ so $\mu(\bar{x}) = 100$ and $\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$

$$\text{For } \bar{x} = 95, z = \frac{x - \mu}{\sigma} = \frac{95 - 100}{2} = -2.5$$

$$\text{Hence } P(\bar{X} < 95) = P(Z < -2.5) = F(-2.5) = 1 - F(2.5) = 1 - 0.9938 = 0.0062$$

Hence the solution.

Problem 18# The mean voltage of a battery is 15 volt and s.d.is 0.2 volt. What is the probability that four such batteries connected in series will have a combined voltage of 60.8 or more volts? [Ans. 0.0228]

Solution: Let, mean voltage of a batteries 1,2,3,4 be $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$ the mean of the series of the four batteries connected is

$$\mu(\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4) = \mu(\bar{x}_1) + \mu(\bar{x}_2) + \mu(\bar{x}_3) + \mu(\bar{x}_4) = 15 + 15 + 15 + 15 = 60$$

$$\sigma(\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4) = \sqrt{\sigma^2(\bar{x}_1) + \sigma^2(\bar{x}_2) + \sigma^2(\bar{x}_3) + \sigma^2(\bar{x}_4)} = \sqrt{4(0.2)^2} = 0.4$$

$$\text{Let } X \text{ be the combined voltage of the series. When } x = 60.8, z = \frac{x - \mu}{\sigma} = \frac{60.8 - 60}{0.4} = 2$$

Then the probability that the combined voltage is more than 60.8 is given by $P(X \geq 60.8) = P(Z \geq 2) = 0.0228$. Hence the solution.

Problem 19# Certain ball bearings have a mean weight of 5.02 ounces and standard deviation of 0.30 ounces. Find the probability that a random sample of 100 ball bearings will have a combined weight between 496 and 500 ounces? [Ans. 0.2318]

Solution: $\mu = 5.02$, $\sigma = 0.30$, $n = 100$

$$\begin{aligned}\mu(\bar{X}) &= \mu = 5.02, \quad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{0.30}{\sqrt{100}} = 0.03 \\ P(4.96 < \bar{X} < 5.05) &= P\left(\frac{4.96 - 5.02}{0.03} < Z < \frac{5.05 - 5.02}{0.03}\right) = P(-2 < Z < 0.66) \\ &= F(0.66) - F(-2) \\ &= F(0.66) - [1 - F(2)] \\ &= F(0.66) + F(2) - 1 \\ &= 0.7454 + 0.9772 - 1 \\ &= 0.7226\end{aligned}$$

Hence the solution.

Problem 20# A manufacturer of fuses claims that with a 20% overload, the fuses will blow in 12.40 minutes on the average. To test the claim, a sample of 20 of the fuses was subjected to a 20% overload, and the times it took them to blow had a mean of 10.63 minutes and a s.d. of 2.48 minutes. If it can be assumed that the data constitute a random sample from a normal population, do they tend to support or refute the manufacturer's claim? [Ans.- 3.19]

Solution: $n = 20$, $\mu = 12.40$, $\bar{x} = 10.63$, $s = 2.48$ then $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{10.63 - 12.40}{2.48/\sqrt{20}} = -3.19$

Date refutes the producer's claim since $t = -3.19 < -2.861$ with probability $\alpha = 0.005$.
Hence the solution.

Problem 21# show that for random samples of size n from a normal population with the variance σ^2 , the sampling distribution of s^2 has the mean σ^2 and the variance $\frac{2\sigma^4}{n-1}$?

Solution: We have $E\left(\frac{(n-1)s^2}{\sigma^2}\right) = n-1 \Rightarrow E(s^2) = \left(\frac{(n-1)\sigma^2}{n-1}\right) = \sigma^2$

$$Var\left(\frac{(n-1)s^2}{\sigma^2}\right) = 2(n-1)$$

$$Var(s^2) = \frac{\sigma^4 \times 2(n-1)}{(n-1)^2} = \frac{2\sigma^4}{n-1}$$

Hence the solution.

Problem 22# If S_1^2 and S_2^2 are the variances of independent random samples of size $n_1 = 10$ and $n_2 = 15$ from normal population with equal variances find $P(S_1^2 / S_2^2 < 4.03)$? [Ans. 0.99]

Solution: Let $F = \frac{S_1^2}{S_2^2}$ and $P\left(\frac{S_1^2}{S_2^2} < 4.03\right) = 1 - P(F > 4.03)$ with 9 and 14 d.o.f.

From table $F_{0.01, 9, 14} = 4.03$ then the probability $= 1 - 0.01 = 0.99$ Hence the solution.

Problem 23# A random sample of size $n = 25$ from a normal population has the mean $\bar{X} = 47$ and the standard deviation $\sigma = 7$. If we base our decision on the statistic, can we say that the given information supports the conjecture that the mean of the population is $\mu = 42$?

Solution: $f = \frac{47 - 42}{7/\sqrt{25}} = 3.57$ since, 3.57 exceeds $t_{0.005, 24} = 2.797$ for $v = 24$

Clearly that the result is highly unlikely and conjecture is probably false.
Hence the solution.

Problem 24# The claim that the variance of a normal population is $\sigma^2 = 4$ is to be rejected if the variance of a random sample of size 9 exceeds 7.7535. What is the probability that this claim will be rejected even though $\sigma^2 = 4$? [Ans. 0.5]

Solution: given $\sigma^2 = 4$, $n = 9$, $y = \frac{8s^2}{4} = 2s^2$

$P(y \geq 2(7.7535)) = P(y \geq 15.507)$ with 8 d.o.f. $= 0.5$ (table v)

Hence the solution.

Problem 25# A random sample of size $n = 12$ from a normal population $\bar{x} = 27.8$ has the mean and the variance $\sigma^2 = 3.24$. If we base our decision on the statistic can we say that the given information supports the claim that the mean of the population is $\mu = 28.5$? [Ans. -1.347]

Solution: The statistic is $\frac{27.8 - 28.5}{1.8/\sqrt{12}} = -\frac{0.7}{1.8/3.464} = -1.347$ since this is fairly small and close to $-t_{0, 10, 11}$ the data tend to support the claim. Hence the solution.

Problem 26# The distribution of annual earnings of all bank letters with five years experience is skewed negatively. This distribution has a mean of Rs.19000 and a standard deviation of Rs.2000. If we draw a random sample of 30 tellers, what is the probability that the earnings will average more than Rs.19750 annually? [Ans. 0.0202]

Solution: $\bar{X} = 19750$, $\mu = 19000$, $n = 30$, $\sigma = 2000$, standard error of the mean (σ_x) $= \sigma/\sqrt{n} = 2000/\sqrt{30} = 365.16$ consider the standard normal probability distribution, as follows: $Z = \frac{\bar{x} - \mu}{\sigma_x} = \frac{19750 - 19000}{365.16} = 2.05$

Now $P(\text{earnings will average more than Rs.19750 annually})$
 $= P(\bar{X} > 19750)$

$$\begin{aligned}
&= P(Z > 2.05) = 1 - P(Z \leq 2.05) \\
&= 1 - F(2.05) \\
&= 1 - 0.9798 = 0.0202
\end{aligned}$$

Therefore we have determined that there is slightly more than a 2% chance of average earnings more than Rs.19750 annually in a group of 30 letters. Hence the solution.

Problem 27# If a gallon can of paint covers on the average 513.3 square feet(Ft^2) with a standard deviation(s.d.) of 31.5 square feet(Ft^2). what is the probability that the mean area covered by a sample of 40 of these 1 gallon cans will be anywhere from 510 to 520 square feet(Ft^2)? [Ans.0.6553]

Solution: $n = 40$, $\mu = 513.3$ and $\sigma = 31.5$

$$\text{Let } Z = \frac{\bar{x} - \mu}{\sigma_x} = \frac{510 - 513.3}{31.5/\sqrt{40}} = -0.66$$

$$\text{And } Z = \frac{\bar{x} - \mu}{\sigma_x} = \frac{520 - 513.3}{31.5/\sqrt{40}} = 1.34$$

$$\begin{aligned}
P(510 < \bar{X} < 520) &= P(-0.66 < Z < 1.34) = F(1.34) - F(-0.66) = F(1.34) - 1 + F(0.66) \\
&= 0.9099 - 1 + 0.7454 \\
&= 0.6553
\end{aligned}$$

We obtain the probability 0.6553 note that if $\bar{x}$ turned out to be much less than 513.3, say less than 500 this might cause serious doubt whether the sample actually came from a population having $\mu = 513.3$ and $\sigma = 31.5$. the probability of obtaining such a small value i.e., $Z < -2.67$ is only 0.0038. Hence the solution.

Problem 28# A random sample of 100 is taken from an infinite population having the mean $\mu = 76$ and the variance $= \sigma^2 = 256$. Find the probability that $\bar{X}$ will be between 75 and 78? [Ans. 0.6268]

Solution: $n = 100$, $\mu = 76$ and $\sigma = 256$

$$\begin{aligned}
P(75 < \bar{X} < 78) &= P\left(\frac{75 - 76}{1.6} < Z < \frac{78 - 76}{1.6}\right) = P(-0.625 < Z < 1.25) \\
&= F(1.25) - F(-0.625) \\
&= F(1.25) - 1 + F(0.625) \\
&= 0.8944 - 1 + 0.7324 \\
&= 0.6268
\end{aligned}$$

Hence the solution.

Problem 29# If two independent random samples of size $n_1 = 13$ and $n_2 = 7$ are taken from a normal population. What is the probability that the variance of the first sample will be atleast four times as that of the second sample? [Ans. 4.00]

Solution: Given $n_1 = 13$ and $v_1 = n_1 - 1 = 12$; $n_2 = 7$ and $v_2 = n_2 - 1 = 6$

$$S_1^2 = 4S_2^2 \quad \text{Now } F = \frac{S_1^2}{S_2^2} = \frac{4S_2^2}{S_2^2} = 4.00 \quad \text{This value of } F \text{ follows}$$

F-distribution with $v_1 = n_1 - 1 = 12$ and $v_2 = n_2 - 1 = 6$ degrees of freedom Hence from tables we get $F_{0.05}(12,6) = 4.00$ Hence the required probability is 0.05. Hence the solution.

Problem 30# If two independent random samples of size $n_1 = 26$ and $n_2 = 8$ are taken from a normal population. What is the probability that the variance of the second sample will be atleast 2.4 times as that of the first sample? [Ans. 0.05]

Solution: Given $n_1 = 26$ and $n_2 = 8$, $v_1 = 25$ and $v_2 = 7$

$$S_1^2 = (2.4)S_2^2$$

Now $F_{\alpha(25,7)} = 2.4 \Rightarrow \alpha = 0.05$

Hence the solution.

Problem 31# If the actual amount of instant coffee which a filling machine puts into “6-ounce” jars is r. v. having a normal distribution with s.d. 0.05 ounce and if only 3% of the jars are to contain less than 6 ounces of coffee, what must be the mean fill of these jars?

[Ans. $\mu=6.094$]

Solution: Let X be the actual amount of coffee put into the jars, $X \sim N(\mu, 0.05)$

Given $P(X < 6) = 0.03$

$$P(-\infty < \frac{X - \mu}{0.05} < \frac{6 - \mu}{0.05}) = 0.03$$

$$P(-\infty < Z \leq -z) = 0.5 - P(0 < Z \leq z)$$

$$0.5 - P(0 < Z < \frac{\mu - 6}{0.05}) = 0.03$$

$$P(0 < Z < \frac{\mu - 6}{0.05}) = 0.47 \quad \text{from table of areas } P(0 < Z < 1.808) = 0.47$$

$$\text{Implies } \frac{\mu - 6}{0.05} = 1.808 \Rightarrow \mu = 6.094 \text{ ounces. Hence the solution.}$$

Problem 32# A manufacturer of a certain type of synthetic fishing line has found from long experience of testing that the breaking strength of his product has an approximate normal distribution with a mean of 30 pounds(lbs.) and a standard deviation of 4 pounds(lbs.). A time and money saving change in the manufacture process of the product is tried. A sample of 25 testing length pieces of the new process line is taken and tested with a resulting sample mean of 28 pounds(lbs.) What is the probability of obtaining a mean as low as 28 if the process has had no harmful effect on breaking strength? [Ans. 0.006]

Solution: Let X be the breaking strength of a randomly selected piece of line and if $X \sim N(30, 4)$ and $n = 25$, $\mu_{\bar{X}}$ (or $\bar{x}$) = 30, $\sigma_{\bar{x}}$ (or s) = $\frac{\sigma}{\sqrt{n}} = 0.8$

Then $P(\bar{X} \leq 28) = P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}} < \frac{28 - 30}{0.8}\right) = P(Z \leq -2.5) = F(-2.5) = 1 - F(2.5) = 1 - 0.9938$
 $= 0.006$ Thus there is a very small chance of obtaining a sample mean as low as 28 if there had been no change in the quality of the line due to the new process. Hence the solution.

Problem 33# An Urn contains 1000 white and 2000 black balls. If X denotes the number of white balls when 300 balls are drawn without replacement, then find $P(80 < X < 120)$?

[Ans. 0.9858]

Solution: clearly $X \sim B.D^n = (300, 1/3)$

If $p = P(\text{the ball drawn is white}) = 1/3$

Mean $\mu = np = 300 \times 1/3 = 100$

Variance $= \sigma^2 = npq = 200/3$

Since $n = 300$ is large the required probability is

$$P(80 < X < 120) = P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{x}}} = \frac{80 - 100}{200/\sqrt{3}} < Z < \frac{120 - 100}{200/\sqrt{3}}\right) = P(-2.45 < Z < 2.45) = 0.9858$$

Hence the solution.

Problem 34# Two movie theatres compete for 900 visitors. Suppose each visitor chooses one of the two balls independent of the choice of the other visitors; how many seats should each theatre have so that the probability of turning away any visitor for lack of seats is less than 1%?
 [Ans. 489]

Solution: clearly $X \sim B.D^n = (900, 1/2)$

If $p = P(i=1 \text{ to } 900) = 1/2$

Mean $\mu = np = 900 \times 1/2 = 450$

Variance $= \sigma^2 = npq = 900/2 \times 2 = 225$

Since $n = 900$ is large the required probability is a $N(0,1)$ random variable

Now

$$P(-2.58 \leq \frac{X - 450}{15} \leq 2.58) = 0.9902$$

$P(411.3 < X < 488.7) = P(-2.45 < Z < 2.45) = 0.9858$. So the required number of seats is 489. Hence the solution.

Problem 35# Let X be a random variable where μ_x is unknown as $\sigma_x^2 = 0.25$ i.e., $1/4$ Find out how large a random sample must be taken in order that the probability will be at test 0.95 and the sample mean $\bar{x}$ will lie within 0.25 of the population mean? [Ans. 80]

Solution: we have $\sigma_x^2 = 0.25$, $\varepsilon = 0.25$ and $1 - \frac{\sigma^2}{n\varepsilon^2} > 0.95$

Therefore $0.05 > \frac{\sigma^2}{n\varepsilon^2}$ and $n > \frac{\sigma^2}{0.05 \times \varepsilon^2} = \frac{1/4}{0.05 \times (0.25)^2} = 80$ Hence the solution.

Problem 36# If a random sample of size n is selected from the finite population that consists of the integers $1, 2, 3, \dots, N$ show that (i) the mean $\bar{X}$ is $\frac{N+1}{2}$ (ii) the variance of $\bar{X}$ is $\frac{(N+1)(N-n)}{12n}$ (iii) the mean and the variance of $Y = n \cdot \bar{X}$ are $E(Y) = \frac{n(N+1)}{2}$ and the $\text{var}(Y) = \frac{n(N+1)(N-n)}{12}$?

Solution: (i) $\mu = \frac{1+2+3+\dots+N}{N} = \frac{N(N+1)}{2N} = \frac{N+1}{2}$

$$\therefore \mu = \frac{N+1}{2}$$

(ii) $\text{Variance}(\sigma^2) = \frac{1^2 + 2^2 + 3^2 + \dots + N^2}{N} - \frac{(N+1)(N+1)}{4}$

$$= \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4}$$

$$\therefore \sigma^2 = \frac{N^2 - 1}{12}$$

$$\therefore \text{Var}(\bar{X}) = \frac{N^2 - 1}{12n} \cdot \frac{N-n}{N-1} = \frac{(N+1)(N-n)}{12n}$$

(iii) $\mu_y = \frac{n(N+1)}{2}$

$$\text{Var}(Y) = \frac{n^2(N+1)(N-n)}{12n} = \frac{n(N+1)(N-n)}{12}$$

$$\therefore \text{Var}(Y) = \frac{n(N+1)(N-n)}{12}$$

Problem 37# How many different samples of size $n=3$ can be drawn from a finite population of size (a) $N=12$ (b) $N=20$ (c) $N=50$ [Ans. a) 220, b) 1140 c) 19600]

Solution: a) ${}^{12}C_3 = \frac{12 \cdot 11 \cdot 10}{3!} = 220$; b) ${}^{20}C_3 = \frac{20 \cdot 19 \cdot 18}{3!} = 1140$;

c) ${}^{50}C_3 = \frac{50 \cdot 49 \cdot 48}{3!} = 19600$;

Hence the solution.

Problem 38# What is the probability of each possible sample if (i) a random sample of size $n=4$ is to be drawn from a finite population of size $N=12$ (ii) a random sample of size $n=5$ is to be drawn from a finite population of size $N=22$? [Ans. a) $1/495$ b) $1/77$]

Solution: (i) $\frac{1}{N_c} = \frac{1}{{}^{12}C_4} = \frac{1}{495}$ (ii) $\frac{1}{N_c} = \frac{1}{{}^{22}C_5} = \frac{1}{77}$

Hence the solution.

Problem 39# Independent random samples of size $n_1 = 30$ and $n_2 = 50$ are taken from two normal populations having the means $\mu_1 = 78$ and $\mu_2 = 78$ and the variances σ_1^2 and σ_2^2 . Find the probability that the mean of the first sample will exceed that of the second sample by at least 4.8? [Ans. 0.2743]

Solution: clearly $\mu(\bar{x}_1 - \bar{x}_2) = 78 - 75 = 3$

$$\sigma(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{150}{30} + \frac{200}{50}} = 3$$

$$P((\bar{x}_1 - \bar{x}_2) > 3) = P(Z > \frac{4.8 - 3.0}{3}) = P(Z > 0.6) = 0.2743.$$

Hence the solution.

Problem 40# If S_1 and S_2 are the variances of independent random samples of size $n_1 = 61$ and $n_2 = 31$ from normal population with $\sigma_1^2 = 12$ and $\sigma_2^2 = 18$ Find $P(\frac{S_1^2}{S_2^2} > 1.16)$

[Ans. 0.05]

Solution: Let $F = \frac{S_1^2/12}{S_2^2/18} = 1.5 \frac{S_1^2}{S_2^2}$

Consider $P\left(\frac{S_1^2}{S_2^2} > 1.16\right) = P\left(\frac{1.5 \times S_1^2}{S_2^2} > 1.16 \times 1.5\right)$

$$= P(F > 1.74) \text{ for } 60 + 30 \text{ d.o.f.} = 0.05 \text{ Hence the solution.}$$

Chapter 2

Objective bits III

Sampling Distributions

By Dr. N. V. Nagendram

01. A sample consists of _____ [Ans. any part of population]
02. Another name of population is _____ [Ans. Universe]
03. The number of possible samples of size n out of N population units without replacement is _____ [Ans. ${}^N C_n$]
04. The number of possible samples of size n from a population of N units with replacement is _____ [Ans. $\frac{1}{N} C_n$]
05. Probability of anyone sample of size n being drawn out of N units is _____ [Ans. $\frac{1}{N} C_n$]
06. Probability of including a specified unit/ item in a sample of size n selected out of N units is _____ [Ans. $\frac{1}{N}$]
07. Having sample observations $x_1, x_2, x_3, \dots, x_n$ the formula for variance is _____ [Ans. $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$]
08. Sample mean formula _____ [Ans. $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$]
09. $\frac{N-n}{N-1}$ is called _____ [Ans. Finite population correction factor]
10. The discrepancies between sample estimate and population parameter is the _____ [Ans. Sampling Error]
11. If the observations recorded on five sampled items are 3,4,5,6,7 the sample variance is _____ [Ans. 2.5]
12. A population consisting of all real numbers is an example of [Ans. An infinite population]
13. Standard deviation of all possible estimate from samples of fixed size is called _____ [Ans. Standard error]

14. A population parameter is a _____ associated with the entire population
[Ans. descriptive or statistical]

15. If $\bar{x}$ is the mean of a random sample size n taken from a population nearly normal having mean μ and the finite variance σ^2 then $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

Is a random variable following as n tends to infinite i.e. $n \rightarrow \infty$
[Ans. standard normal distribution]

16. Standard error of the statistic sample mean $\bar{x}$ _____ [Ans. $\frac{\sigma}{\sqrt{n}}$]

17. If $x_1, x_2, x_3, \dots, x_n$ constitute a random sample from an infinite population with the mean μ and the variance σ^2 then $\mu(\bar{x}) =$ _____ and $\sigma^2(\bar{x}) =$ _____ [Ans. $\mu, \frac{\sigma^2}{n}$]

18. If $\bar{x}$ is the mean of a random sample from a finite population size N with the mean μ and the variance σ^2 then $\mu(\bar{x}) =$ _____ and $\sigma^2(\bar{x}) =$ _____ [Ans. $\mu, \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$]

19. $t_{1-\alpha} =$ _____ [Ans. $-t_\alpha$]

20. $F_{1-\alpha}(v_1, v_2) =$ _____ [Ans. $\frac{1}{F_\alpha(v_2, v_1)}$]

Chapter 1

PROBABILITY DISTRIBUTION

Tutorial – 16

Probability Density Function Problems **REVISION**

By Dr. N.V.Nagendram

Problem #1 If $E(X) = 1$, $E(X^2) = 4$, find the mean and variance of $Y = 2x - 3$? [Ans. Var = 12]

Problem #2 A continuous random variable X has the p.d.f. given by $f(x) = kx^2$, $0 \leq x \leq 1$. Find the value of k . with this value of k find $P(x < \frac{1}{2})$ and $P(x \geq \frac{3}{4})$? [Ans. $\frac{1}{8}, \frac{37}{64}$]

Problem #3 The probability density $p(x)$ of a continuous random variable is given by $p(x) = y_0 e^{-|x|}$, $-\infty < x < \infty$, prove that $y_0 = \frac{1}{2}$ find the mean and variance of the distribution? [Ans. var = 2]

Problem #4 A continuous random variable X has the p.d.f. given by

$$\begin{aligned} f(x) &= kx, & 0 \leq x \leq 1 \\ &= k, & 1 \leq x \leq 2 \\ &= -x+3k, & 2 \leq x \leq 3 \end{aligned}$$

$= 0$ otherwise. Find the value of k . Also calculate $P(X \leq 1.5)$? [Ans. $\frac{1}{2}$]

Problem #5 Given that $f(x) = \frac{k}{2^x}$ is a probability distribution function for a random variable X , that can take on the values $x = 0, 1, 2, 3$ and 4 (i) find k (ii) mean and variance of x ? [Ans. $\mu=0.839$ $\sigma^2 = 1.168$]

Problem #6 (a) is the function $f(x)$, defined as follows, a density function?

$$\begin{aligned} f(x) &= 0 & x < 2 \\ &= \frac{1}{8}(3 + 2x) & -2 \leq x \leq 4 \\ &= 0, & x > 4 \end{aligned}$$

(b) Find the probability that a variate having this density will fall in the interval $2 \leq x \leq 3$? [Ans. a) 1b) $\frac{4}{9}$]

Problem #7 Find the constant k so that function $F(x)$ is defined as follows may be a density

function:
$$f(x) = \begin{cases} \frac{1}{k} & a \leq x \leq b \\ 0 & \text{elsewhere.} \end{cases}$$

Find also the cumulative distribution function of the random variable X and K satisfies the requirements for $f(x)$ to be a density function? [Ans. $k = b-a$, $F(x) = 1$]

Problem #8 If X is a continuous random variable with p. d. f. given by

$$\begin{aligned} F(x) &= kx & 0 \leq x \leq 2 \\ &= 2k & 2 \leq x \leq 4 \\ &= -kx + 6k & 4 \leq x \leq 6 \end{aligned}$$

Find the value of k and mean value of X.

$$[\text{Ans. } k = \frac{1}{8}, \mu = E(X) = 3]$$

Problem #9 (a) verify that the following is a distribution function:

$$\begin{aligned} F(x) &= 0 & x < -a \\ &= \frac{1}{2} \left(\frac{x}{a} + 1 \right) & -a \leq x \leq a \\ &= 1 & x > a \end{aligned}$$

$$\begin{aligned} \text{(b) show that } F(x) &= 0 & -\infty < x < 0 \\ &= 1 - e^{-x} & 0 \leq x < \infty \end{aligned} \text{ is possible distribution}$$

function and find the density function?

$$[\text{Ans. a) 1 b) 1}]$$

Problem #10 A random process gives measurements X between 0 and 1 with a probability density function $f(x) = 12x^3 - 21x^2 + 10x$, $0 \leq x \leq 1$

$$= 0 \text{ otherwise. (i) find } P(X \leq \frac{1}{2}) \text{ and } P(X > \frac{1}{2}) \text{ (ii) Find a number}$$

$$k \text{ such that } P(X \leq k) = \frac{1}{2} ?$$

$$[\text{Ans. a) } \frac{9}{16}, \text{ b) } \frac{7}{16}, \frac{1}{2} k = 0.45]$$

Problem #11 The probability distribution function of a random variable X is

$$\begin{aligned} f(x) &= x & 0 \leq x \leq 1 \\ &= 2 - x & 1 \leq x \leq 2 \\ &= 0 & x \geq 2 \end{aligned}$$

compute the cumulative distribution function of X?

$$[\text{Ans. } F(x) = 1]$$

Problem #12 The frequency function of a continuous random variable is given by $f(x) = y_0 x(2 - x)$, $0 \leq x \leq 2$. Find the value of y_0 , mean and variance of X ?

$$[\text{Ans. } y_0 = 3/4, \text{ var} = 1/5]$$
