Department of Engg. Mathematics

Course : Engg. Mathematics-IV 15MAT41. Sem.: 4th (2017-18)

Course Coordinator:

Prof. S. L. Patil

Probability Theory

Contents

Probability

Probability is the likelihood that the event will occur.

Two Conditions:

- **Value is between 0 and 1.**
- **Sum of the probabilities of all events must be 1.**

Random Variable

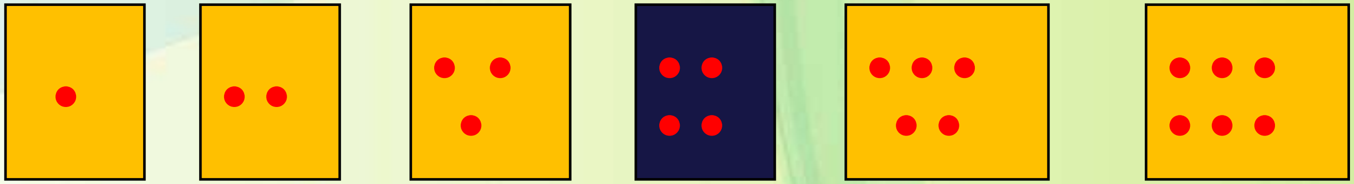
A numerical description of the outcome of an experiment

Random experiment is an experiment with random outcome.

Random variable is a variable related to a random event

Example: Discrete RV: countable # of outcomes

Throw a die twice: Count the number of times 4 comes up (0, 1, or 2 times)



- A random variable is a rule that assigns exactly one value to each point in a sample space for an experiment.
- A discrete random variable may assume either a finite number of values or an infinite sequence of values.
- A continuous random variable may assume any numerical value in an interval or collection of intervals.

Discrete Random Variable

- Obtained by Counting (0, 1, 2, 3, etc.)
- Usually finite by number of different values

Ex: Toss a coin 5 times. Count the number of tails.

(0, 1, 2, 3, 4, or 5 times)

Discrete Random Variables

- The number of throws of a coin needed before a head first appears
- The number of dots when rolling a dice
- The number of defective items in a sample of 20 items
- The number of customers arriving at a check-out counter in an hour
- The number of people in favor of nuclear power in a survey

- The probability distribution is defined by a probability function which provides the probability for each value of the random variable.
- A continuous random variable can assume any value in an interval on the real line or in a collection of intervals.
- Probability Density Function : For a continuous random variable X, a probability density function is a function such that

1) F(x) ≥ 0,   2) $\int_{-\infty}^{\infty} f(x)dx = 1$

Discrete - Continuous

- Random variable is discrete if it can take no more than countable number of values
- Random variable is continuous, if it can take any value in an interval

Continuous Random Variables

- The yearly income for a family
- The amount of oil imported into Finland in a particular month
- The time that elapses between the installation of a new component and its failure
- The percentage of impurity in a batch of chemicals

Expected Value

- Expected value is just like the mean in empirical distributions

Examples:

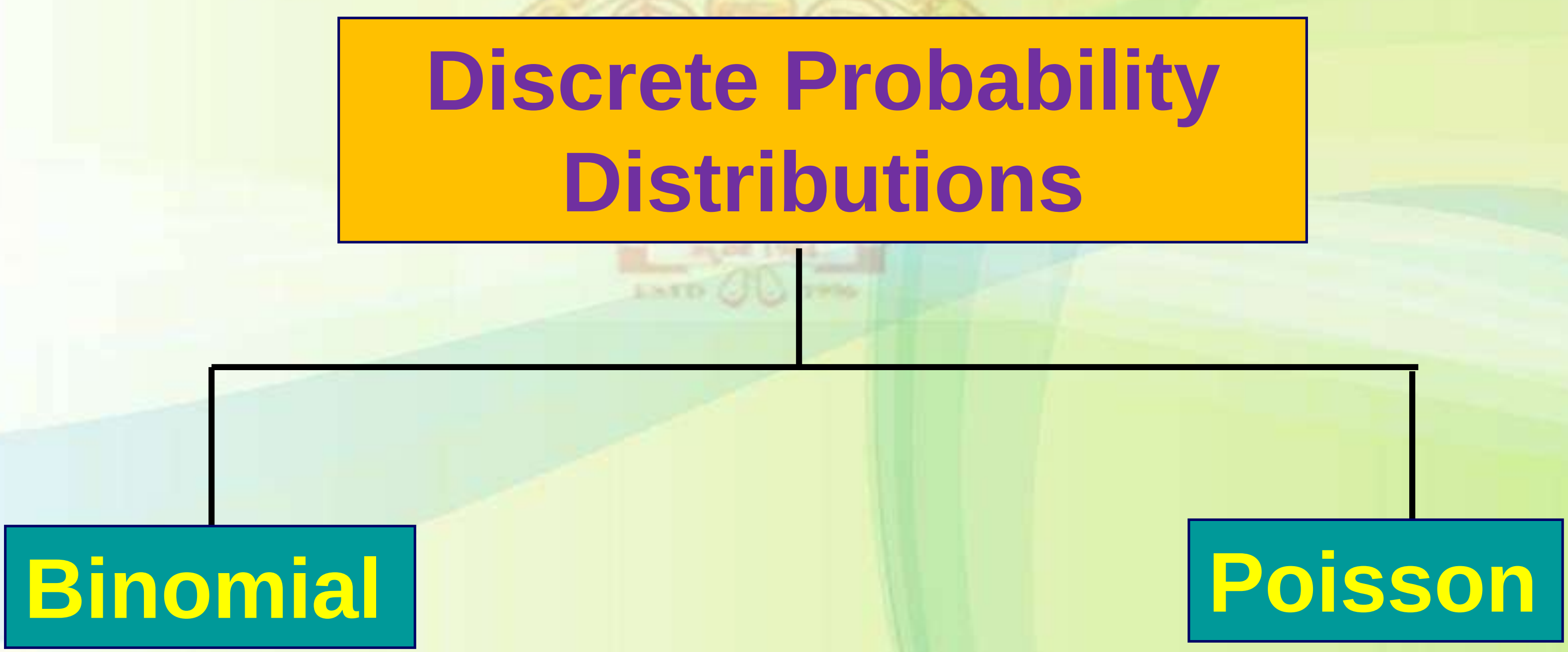
- When playing a dice the expected value equals 3,5
- Insurance company is interested in the expected value of indemnities
- Investor is interested in the expected value of portfolio’s revenue

Expected value calculation

- The expected value for a discrete random variable is obtained by multiplying each possible outcome by its probability and then sum these products

Discrete probability distribution

- Discrete random variable values and their probabilities.



Binomial Distribution

Binomial experiments satisfy the following:

- The experiment consists of a sequence of n identical trials
- All possible outcomes can be classified into two categories, usually called success and failure
- The probability of an success, p, is constant from trial to trial
- The outcome of any trial is independent of the outcome of any other trial

Binomial Distribution Random Variables

- The number of heads when tossing a coin for 50 times
- The number of reds when spinning the roulette wheel for 15 times
- The number of defective items in a sample of 20 items from a large shipment
- The number of people in favour of nuclear power in a survey

Poisson distribution

Poisson experiments satisfy the following

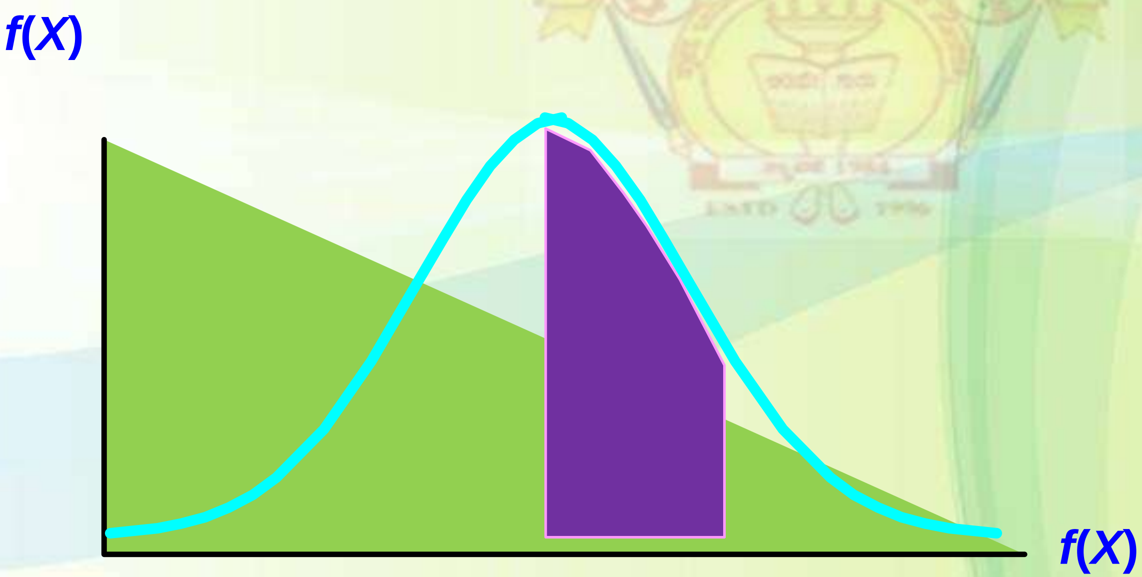
- The probability of occurrence of an event is the same for any two intervals of equal length
- The occurrence or non-occurrence of the event in any interval is independent of the occurrence or non-occurrence in any other interval
- The probability that two or more events will occur in an interval approaches zero as the interval becomes smaller

Poisson Distribution Random Variables

- The number of failures in a large computer system during a given day
- The number of ships arriving at a loading facility during a six-hour loading period
- The number of delivery trucks to arrive at a central warehouse in an hour
- The number of dents, scratches, or other defects in a large roll of sheet metal
- The number of accidents at a crossroads during one year

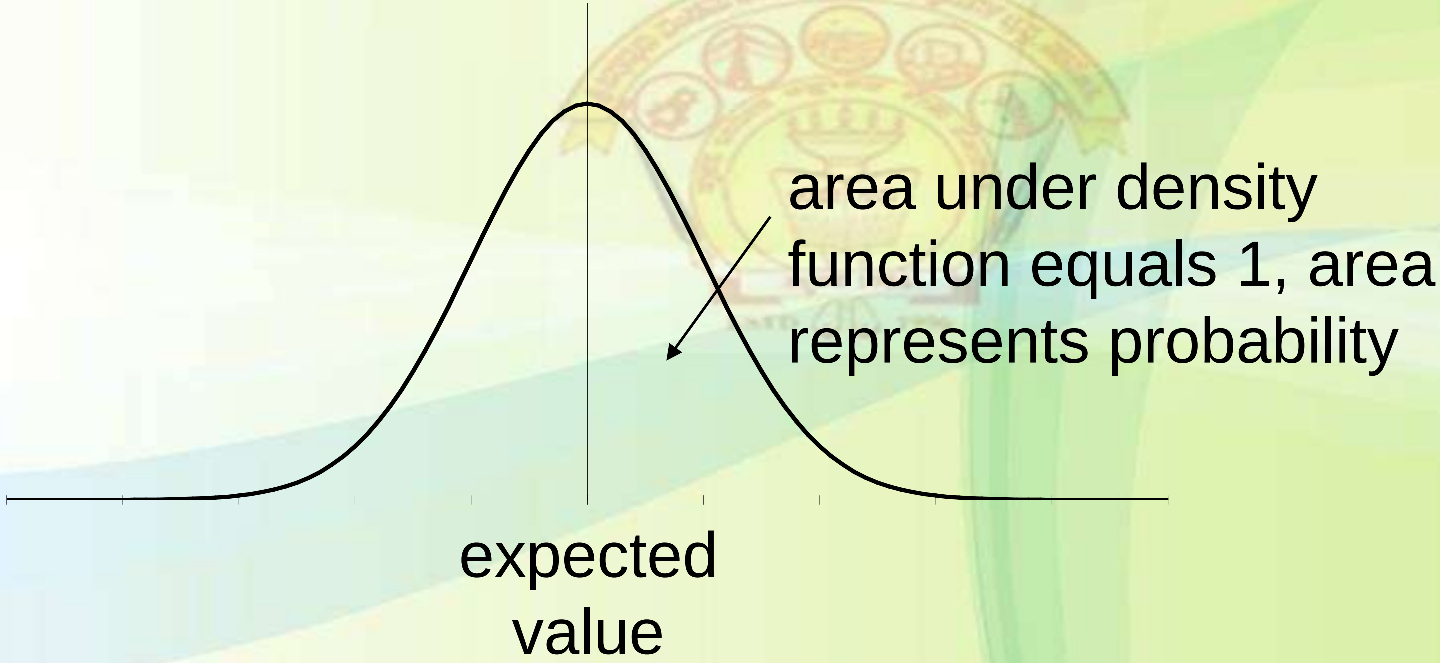
Normal Distribution

Probability is the area under the curve!



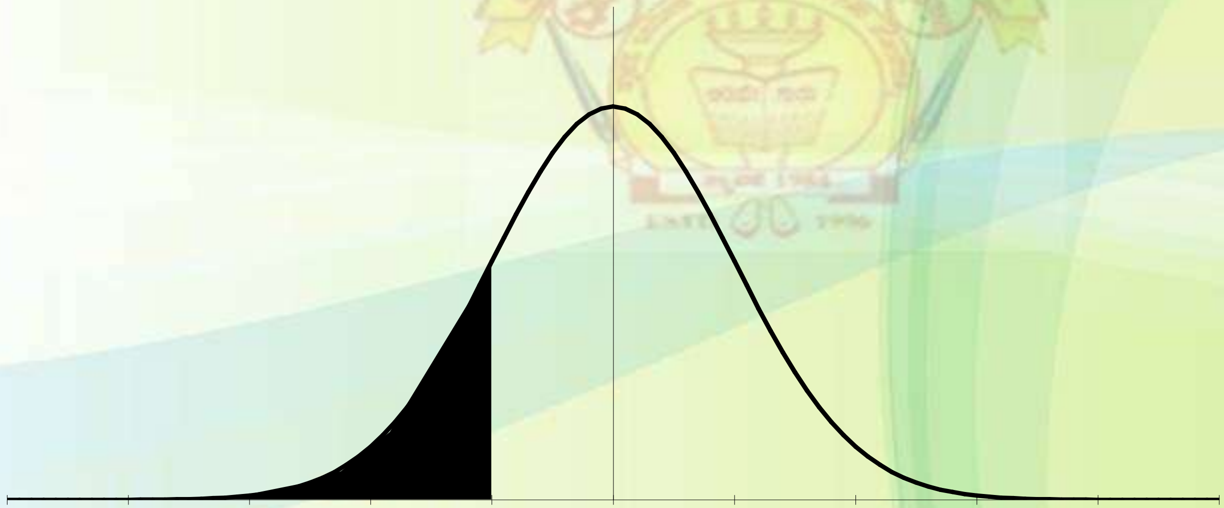
Normal Distribution

Normal distribution is defined by density function



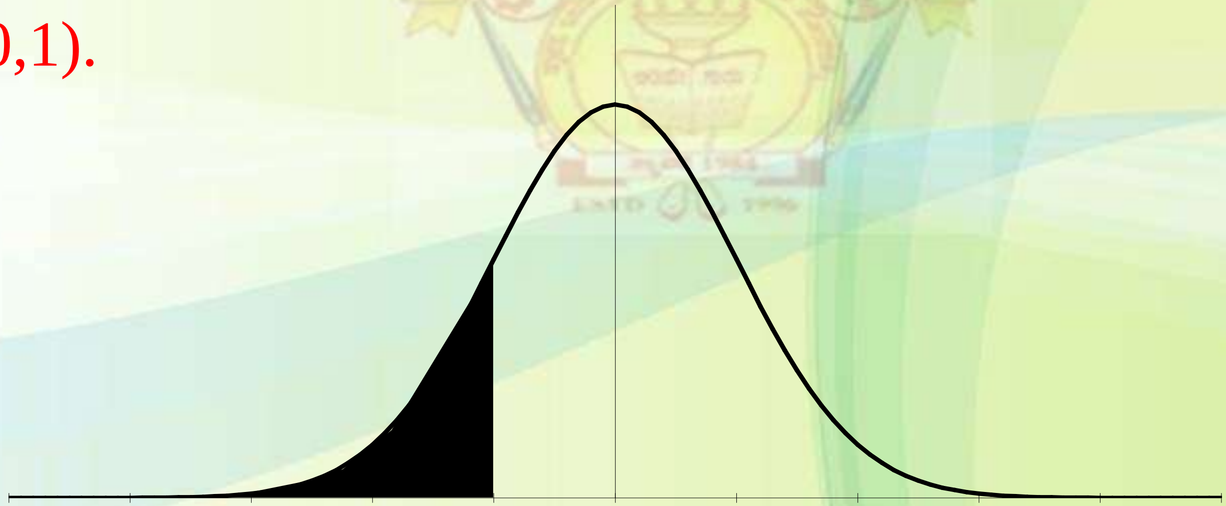
Cumulative Probability Function

- Cumulative function for x = area to the left of x = probability to get at most x:



Standardized Distribution N(0,1)

- Cumulative function values have been tabulated (in most statistics textbooks) for normal distribution with expected value 0 and standard deviation 1
- This distribution is called standardized distribution and is denoted N(0,1).



Standardizing

You can standardize any normal distribution N(μ,σ) variable to a standardized distribution N(0,1) variable

$$z = \frac{x - \mu}{\sigma}$$

Variance/standard deviation

"The average (expected) squared distance (or deviation) from the mean"

$$\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

Variance, formally

Discrete case:

$$Var(X) = \sigma^2 = \sum_{\text{all x}} (x_i - \mu)^2 p(x_i)$$

Continuous case:

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x_i - \mu)^2 p(x_i) dx$$

Two Discrete Random Variables

Joint Probability Distributions

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

(1) $f_{XY}(x, y) \geq 0$

(2) $\sum_x \sum_y f_{XY}(x, y) = 1$

(3) $f_{XY}(x, y) = P(X = x, Y = y)$ (5-1)

Two Discrete Random Variables

Marginal Probability Distributions

- The individual probability distribution of a random variable is referred to as its **marginal probability distribution**.

- In general, the marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables. For example, to determine $P(X = x)$, we sum $P(X = x, Y = y)$ over all points in the range of (X, Y) for which $X = x$. Subscripts on the probability mass functions distinguish between the random variables.

Two Continuous Random Variables

Conditional Probability Distribution

Definition

Continuous random variables $X_1, X_2, \ldots, X_p$ are **independent** if and only if

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2 \ldots, x_p) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_p}(x_p) \quad \text{for } all\ x_1, x_2, \ldots, x_p \qquad (5\text{-}24)$$

Covariance and Correlation

Definition

The **covariance** between the random variables X and Y, denoted as $\text{cov}(X, Y)$ or σ_{XY}, is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y \qquad (5\text{-}26)$$

Covariance is a measure of **linear relationship** between the random variables. If the relationship between the random variables is nonlinear, the covariance might not be sensitive to the relationship. This is illustrated in Fig. 5-13(d). The only points with nonzero probability are the points on the circle. There is an identifiable relationship between the variables. Still, the covariance is zero.

Covariance and Correlation

Definition

The **correlation** between random variables X and Y, denoted as ρ_{XY}, is

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{5-27}$$

For any two random variables X and Y

$$-1 \le \rho_{XY} \le +1 \tag{5-28}$$

Queries ...?