B. Tech Computer Science & Engineering (School of Technology)

MATH2361: PROBABILITY & STATISTICS

UNIT-II

Random Variables Probability functions Mathematical Expectations

Dr. Mallikarjuna Reddy Doodipala
Associate Professor
Department of Mathematics
GITAM University Hyderabad Campus

Learning Objectives

At the end of the module students able to learn:

- Know and differentiate between discrete and continuous random variables
- Understand the density and distribution functions for discrete and continuous variables
- Know about probability distributions binomial, Poisson and Normal
- Applications

Learning Outcomes

After completion of this unit, the student will be able to

- explain the notion of random variable, distribution functions and expected value (L3).
- apply Binomial and Poisson distributions to compute probabilities, theoretical frequencies (L3).
- explain the properties of normal distribution and its applications (L3).

Random Variable and Probability Distributions

- Random variables (Discrete and Continuous),
- Probability mass and density functions,
- Probability distributions

1. Binomial,
2. Poisson,
3. Normal distributions and their properties(Only Mean Variance)

- Applications on Binomial, Poisson and Normal Distributions

Prerequisites

Before you start reading this unit, you should:

- Have some knowledge on definite integrals
- Know about probability and calculating it for simple problems
- Binomial coefficients
- Exponential functions

Random Variables

- Random Variable (RV): A numeric outcome that results from an experiment
- For each element of an experiment's sample space, the random variable can take on exactly one value
- Discrete Random Variable: An RV that can take on only finite or countably infinite set of outcomes It can put 1-1 correspondence
- A R.V is a real valued function defined on a probability space (S,B, P) where S= Sample Space, B=Boral set, P=Probability

Random Variables

Continuous Random Variable : An RV that can assume all

- possible values in a certain interval (or) R.V cannot put 1-1 correspondence .
- Random Variables are denoted by upper case letters (X)
- Individual outcomes for RV are denoted by lower case letters (x)

Distribution Functions -Probability Distributions

- Probability Distribution: A R.V X is distributed according to some probability law
- In other words Table, Graph, or Formula that describes values a random variable can take on, and its corresponding probability (discrete RV) or density (continuous RV)
- Discrete Probability Distribution: Assigns probabilities (masses) to the individual outcomes
- Discrete Probabilities denoted by: $p(x) = P(X=x)$
- Cumulative Distribution Function: $F(x) = P(X \leq x)$

Distribution Functions -Probability Distributions

- Continuous Probability Distribution: Assigns density at individual points, probability of ranges can be obtained by integrating density function
- Continuous Densities denoted by: $f(x)$
- Cumulative Distribution Function: $F(x) = P(X \leq x)$

Discrete Probability Distributions(PMF&DF)

Probability (Mass) Function: p(x) is said to be p.m.f or probabity function

$p(x) = P(X = x)$

$p(x) \geq 0 \quad \forall x$

$$\sum_{\text{all } x} p(x) = 1$$

Distribution function or

Cumulative Distribution Function (CDF):

$F(x) = P(X \leq x)$

$$F(b) = P(X \leq b) = \sum_{x=-\infty}^{b} p(x)$$

$F(-\infty) = 0 \quad F(\infty) = 1$

$F(x)$ is monotonically increasing in x

Example – Rolling 2 Dice (Red/Green)

Y = Sum of the up faces of the two die. Table gives value of *y* for all elements in *S*

Red\Green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

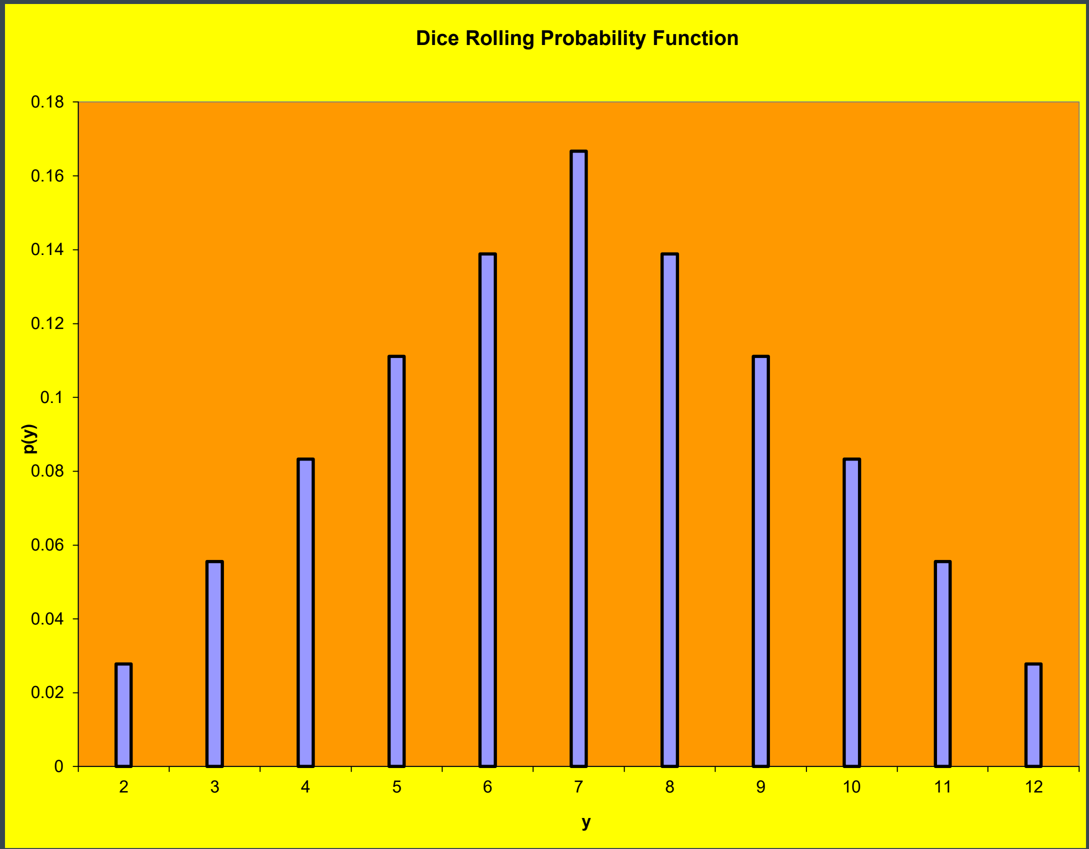
Rolling 2 Dice – Probability Mass Function & CDF

X	p(x)	F(x)
2	1/36	1/36
3	2/36	3/36
4	3/36	6/36
5	4/36	10/36
6	5/36	15/36
7	6/36	21/36
8	5/36	26/36
9	4/36	30/36
10	3/36	33/36
11	2/36	35/36
12	1/36	36/36

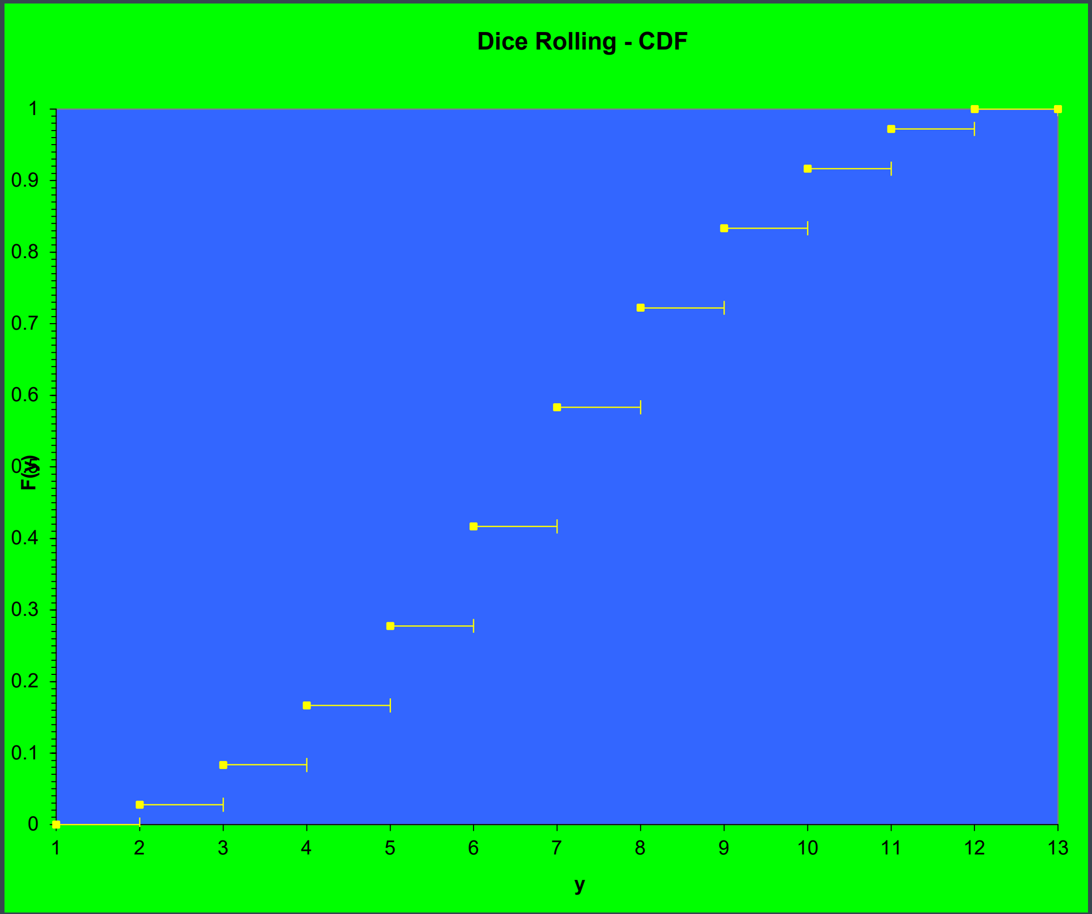
$$p(x) = \frac{\#\text{ of ways 2 die can sum to } x}{\#\text{ of ways 2 die can result in}}$$

$$F(x) = \sum_{x} p(X \leq x)$$

Rolling 2 Dice – Probability Mass Function Graph



Rolling 2 Dice – Cumulative Distribution Function



Expected Values of Discrete RV's

- Mean (Expected Value) – Long-Run average value an RV (or function of RV) will take on
- Variance – Average squared deviation between a realization of an RV (or function of RV) and its mean
- Standard Deviation(SD) – Positive Square Root of Variance (in same units as the data)
- Notation: Mean: $E(Y) = \mu$
- Variance: $V(Y) = \sigma^2$
- Standard Deviation: σ

Expected Values of Discrete RV's

Mean: $E(X) = \mu = \sum_{\text{all } x} xp(x)$

Mean of a function $g(X)$: $E[g(X)] = \sum_{\text{all } x} g(x)p(x)$

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

$$= \sum_{\text{all } x} (x - \mu)^2 p(x) = \sum_{\text{all } x} (x^2 - 2x\mu + \mu^2)p(x) =$$

$$= \sum_{\text{all } x} x^2 p(x) - 2\mu \sum_{\text{all } x} xp(x) + \mu^2 \sum_{\text{all } x} p(x) =$$

$$= E[Y^2] - 2\mu(\mu) + \mu^2(1) = E[X^2] - \mu^2$$

Standard Deviation: $\sigma = +\sqrt{\sigma^2}$

Expected Values of Linear Functions of Discrete RV's

Linear Functions: $g(Y) = aY + b \quad (a, b \equiv \text{ constants})$

$$E[aY + b] = \sum_{\text{all } y} (ay + b)p(y) =$$

$$= a \sum_{\text{all } y} yp(y) + b \sum_{\text{all } y} p(y) = a\mu + b$$

$$V[aY + b] = \sum_{\text{all } y} ((ay + b) - (a\mu + b))^2 p(y) =$$

$$\sum_{\text{all } y} (ay - a\mu)^2 p(y) = \sum_{\text{all } y} [a^2(y - \mu)^2] p(y) =$$

$$= a^2 \sum_{\text{all } y} (y - \mu)^2 p(y) = a^2\sigma^2$$

Sample Problems

x	-3	-1	3	5
p(x)	0.4	0.1	0.2	0.3

- The expectation or mean of the r.v. distributed as in the table below is:

$$E(X) = \sum_{x=-3}^{5} x.p(x)$$

- E[X] = (-3)(0.4)+(-1)(0.1)+(3)(0.2)+(5)(0.3)

  = -1.2-0.1+0.6+1.5

  = 0.8

Sample Problem -2 A discrete r.v X has the following probability distribution. Find the expected value

x	0	1	2	3	4
p(x)	0.1296	0.3456	0.3456	0.1536	0.0256

- To find the expected value of the given distribution by def.of expectation we know that

$$E(X) = \sum_{x=0}^{4} x.p(x)$$

$$E[X] = 0*0.1296 + 1*0.3456 + 2*0.3456 + 3*0.1536 + 4*0.0256$$

$$= 0.16$$

Sample Problem -3

x	0	1	2	3	4	5	6
p(x)	k	3k	5k	7k	9k	11k	13k

Find constant k, p(x<4),p(x>=4), p(3<x<=6)

Sol: To find constant k we know that total probability is one.

Thus

$$\sum_{x=0}^{6} p(x) = 1 \Rightarrow p(0) + p(1) + p(2) + p(3) + p(4) + p(5) + p(6) = 1$$

$$\Rightarrow k + 3k + 5k + 7k + 9k + 11k + 13k = 1 \Rightarrow k = 1/49$$

$$p(x < 4) = p(x = 0,1,2,3) = 1/49 + 3/49 + 5/49 + 7/49 = 16/49$$

$$p(x \geq 4) = 1 - p(x < 4) = 1 - 16/49 = 33/49$$

$$p(3 < x \leq 6) = p(x = 4,5,6) = p(x = 4) + p(x = 5) + p(6)$$

$$= 9/49 + 11/49 + 13/49 = 33/49$$

Example – Rolling 2 Dice: pmf, Mean Variance

y	p(y)	yp(y)	$y^2p(y)$
2	1/36	2/36	4/36
3	2/36	6/36	18/36
4	3/36	12/36	48/36
5	4/36	20/36	100/36
6	5/36	30/36	180/36
7	6/36	42/36	294/36
8	5/36	40/36	320/36
9	4/36	36/36	324/36
10	3/36	30/36	300/36
11	2/36	22/36	242/36
12	1/36	12/36	144/36
Sum	36/36 =1.00	252/36 =7.00	1974/36= 54.833

$$\mu = E(X) = \sum_{x=2}^{12} xp(x) = 7.0$$

$$\sigma^2 = E[X^2] - \mu^2 = \sum_{y=2}^{12} x^2 p(x) - \mu^2$$

$$= 54.8333 - (7.0)^2 = 5.8333$$

$$\sigma = \sqrt{5.8333} = 2.4152$$

Continuous Random Variables

- Recall that a random variable X is simply a function from a sample space S into the real numbers.
- The random variable is discrete is the range of X is finite or countably infinite.
- This refers to the number of values X can take on, not the size of the values.
- The random variable is continuous if the range of X is uncountably infinite and X has a suitable pdf (see below).
- Typically an uncountably infinite range results from an X that makes a physical measurement—e.g., the position, size, time, age, flow, volume, or area of something.

Definition and Basic Properties

- The pdf of a continuous random variable X must satisfy three conditions.
- It is a nonnegative function (but unlike in the discrete case it may take on values exceeding

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Its definite integral over the whole real line equals one. That is

Definition and Basic Properties

- The pdf of a continuous random variable X must satisfy three conditions.
  - Its definite integral over a subset B of the real numbers gives the probability that X takes a value in B. That is,

    $$\int_B f(x) = P(X \in B)$$

    for “every” subset B of the real numbers. As a special case (the usual case) for all real numbers a and b

    $$\int_a^b f(x)dx = P(a \le X \le b)$$

    Put simply, the probability is simply the area under the pdf curve over the interval [a,b].

Definition and Basic Properties

- If X has uncountable range and such a pdf, then X is a *continuous random variable*.
- In this case we often refer to f as a *continuous pdf*.
- Note that this means f is the pdf of a continuous random variable.
- It does not necessarily mean that f is a continuous function.

Sample Problems

- Note that by this definition the probability of X taking on a single value a is always 0. This follows from

$$P(X = a) = P(a \leq X \leq a)$$
$$= \int_a^a f(x)dx = 0$$

- since every definite integral over a degenerate interval is 0.
- This is, of course, quite different from the situation for discrete random variables.

Definition and Basic Properties

- Consequently, we can be sloppy about inequalities. That is

$$P(a < X < b) = P(a \leq X < b)$$
$$= P(a < X \leq b)$$
$$= P(a \leq X \leq b)$$

Remember that this is blatantly false for discrete random variables.

Sample Problems

- Examples
- Let X be a random variable with range [0,2] and pdf defined by f(x)=1/2 for all x between 0 and 2 and f(x)=0 for all other values of x. Note that since the integral of zero is z ero we get

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{2} 1/2\, dx = \frac{1}{2}x\Big|_{0}^{2} = 1 - 0 = 1$$

- That is, as with all continuous pdfs, the total area under the curve is 1. We might use this random variable to model the position at which a two-meter with length of rope breaks when put under tension, assuming "every point is equally likely". Then the probability the break occurs in the last half-meter of the rope is

$$P(3/2 \le X \le 2) = \int_{3/2}^{2} f(x)dx = \int_{3/2}^{2} 1/2dx = \frac{1}{2}x\Big|_{3/2}^{2} = 1/4$$

Sample Problems

- Examples
  - Let Y be a random variable whose range is the nonnegative reals and whose pdf is defined by

$$f(x) = \frac{1}{750} e^{-x/750}$$

  - for nonnegative values of x (and 0 for negative values of x). Then

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \frac{1}{750} e^{-x/750} dx = \lim_{t\to\infty} \int_{0}^{t} e^{-x/750} dx$$
$$= \lim_{t\to\infty} e^{-x/750}\Big|_{0}^{t} = \lim_{t\to\infty} \left(e^{-0} - e^{-750/t}\right) = 1 - 0 = 1$$

Sample Problems

- The random variable Y might be a reasonable choice to model the lifetime in hours of a standard light bulb with average life 750 hours.
- To find the probability a bulb lasts under 500 hours, you calculate

$$P(0 \le Y < 500) = \int_0^{500} \frac{1}{750} e^{-x/750} dx = -e^{-x/750}\Big|_0^{500} = -e^{-2/3} + 1$$
$$\approx 0.487$$

Cumulative Distribution Functions

- The cdf F of a continuous random variable has the same definition as that for a discrete random variable. That is,

$$F(x) = P(X \leq x)$$

- In practice this means that F is essentially a particular anti derivative of the pdf since

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

- Thus at the points where f is continuous F'(x)=f(x).

Cumulative Distribution Functions

- Knowing the cdf of a random variable greatly facilitates computation of probabilities involving that random variable since, by the Fundamental Theorem of Calculus,

$$P(a \leq X \leq b) = F(b) - F(a)$$

Cumulative Distribution Functions

- In the second example above, F(x)=0 if x is negative and for nonnegative x we have

$$F(x) = \int_0^x \frac{1}{750} e^{-t/750} dt = -e^{-t/750}\Big|_0^x = -e^{-x/750} + 1 = 1 - e^{-x/750}$$

- Thus the probability of a light bulb lasting between 500 and 1000 hours is

$$F(1000) - F(500) = (1 - e^{-1000/750}) - (1 - e^{-500/750}) = e^{-2/3} - e^{-4/3} \approx 0.250$$

Cumulative Distribution Functions

- In the first example above

F(x) =0 for negative x,

=1 for x greater than 2 and

=x/2 for x between 0 and 2 since for such x we have

$$F(x) = \int_0^x 1/2dt = \frac{1}{2}t\Big|_0^x = \frac{1}{2}x$$

- Thus to find the probability the rope breaks somewhere in the first meter we calculate

F(1)-F(0)=1/2-0-1/2,

which is intuitively correct.

Cumulative Distribution Functions

- If X is a continuous random variable, then its cdf is a continuous function. Moreover,

$$\lim_{x \to -\infty} F(x) = 0$$

- and

$$\lim_{x \to \infty} F(x) = 1$$

- Again these results are intuitive

Tutorial Problems for Practice

1. A random variable has the following probability distribution

Values of X	0	1	2	3	4	5	6	7	8
	a	$3a$	$5a$	$7a$	$9a$	$11a$	$13a$	$15a$	$17a$

Determine the value of a (2) Find (i) P(x < 3) (ii) P(x <= 3) (iii) P(x >7) (iv)P(2 <x < 5), (v) P(2 < x <5) (3) Find the cumulative distribution function of x.

Tutorial Problems for Practice

2. An urn contains 6 red and 4 white balls. Three balls are drawn at random. Obtain the probability distribution of the number of white balls drawn.

Tutorial Problems for Practice

3. Find the probability distribution of the number of sixes is a r.v in throwing two dice once. Also obtain distribution function.

Tutorial Problems for Practice

4. A random variable X has the following probability distribution

Value of x	0	1	2	3	4
P(X=x)	3a	4a	6a	7a	8a

(a) Determine the value of a

(b) Find p(1 < x < 4) (c) P(1 <= x <4)

(d) Find P(x >2)

(e) Find the distribution function of x

Tutorial Problems for Practice

5. A random variable X has the following probability function.

Values of X	0	1	2	3	4	5	6	7
p(x)	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

(i) Find k (ii) Find $p(0 < x < 5)$ (iii) Find $p(x <= 6)$

Tutorial Problems for Practice

6. Examine whether $f(x) = 5x^4$, $0 < x < 1$ can be a p.d.f of a continuous random variable x.

Tutorial Problems for Practice

7. A continuous random variable X has the probability density lawf(x) = Ax^2, $0 < x < 1$.
(i) Determine A (ii) $p(0<X<0.5)$, $p(X>=5)$ (iii) $p(1/4<X<1/2)$

Tutorial Problems for Practice

8. $f(x) = c(1-x)\,x^2$, $0 < x < 1$ be a probability density function of a random variable x.
   Find the constant c. (i) $P(1/4<X<1/2)$ (ii) CDF

Tutorial Problems for Practice

9. A random variable x has the density function $f(x) = 1/4$, $-2 < x < 2$, $= 0$, otherwise. Obtain (i) $P(-1 < x < 2)$ (ii) $P(x > 1)$

10. In a continuous distribution, whose probability density function is given by $f(x) = K\,x(2-x)$, $0 < x < 2$.

Find K (i) $p(0<X<1.5)$, $p(X>1.5)$ (iii) $p(1/2<X<2)$ (Ans $K=3/4$)

Mathematical expectation and Variance

- Definitions : Case Continuous
  - The expected value of a continuous random variable X is defined by

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

μ

  - Note the similarity to the definition for discrete random variables. Once again, we often denote it by . As in the discrete case this integral may not converge, in which case the expectation if X is undefined.

Mathematical expectation and Variance

- Definitions
  - As in the discrete case we define the variance by

    $$Var(X) = E((X - \mu)^2)$$

  - Once again, the standard deviation is the square root of variance.
  - Variance and standard deviation do not exist if the expected value by which they are defined does not converge.

Mathematical expectation and Variance

- Theorems
  - The Law of the Unconscious Statistician holds in the continuous case. Here it states

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx$$

  - Expected value still preserves linearity. That is $E(aX + b) = aE(X) + b$
  - The proof depends on the linearity of the definite integral (even an improper Riemann integral).

Mathematical expectation and Variance

- Theorems
  - Similarly the expected value of a sum of functions of X equals the sum of the expected values of those functions
  - The shortcut formula for the variance holds for continuous random variables, depending only on the two preceding linearity results and a little algebra, just as in the discrete case. The formula states

    $$Var(X) = E(X^2) - E(X)^2 = E(X^2) - \mu^2$$

  - Variance and standard deviation still act in the same way on linear functions of X. Namely $Var(aX + b) = a^2 Var(X)$ and $SD(aX + b) = |a| SD(X)$

Mathematical expectation and Variance

- Examples
  - In the two-meter-wire problem,
  - the expected value should be 1, intuitively.
  - Let us calculate: By the notation of Expectation

$$E(X) = \int_0^2 x\left(\frac{1}{2}\right) dx = \int_0^2 \frac{1}{4} x\, dx = \frac{1}{4} x^2 \Big|_0^2 = 1 - 0 = 1$$

Mathematical expectation and Variance

- Examples
  - In the same example the variance is

$$Var(X) = E(X^2) - 1^2 = \int_0^2 x^2 \left(\frac{1}{2}\right) dx - 1 = \frac{1}{6} x^3 \Big|_0^2 - 1 = \frac{1}{3}$$

  - and consequently

$$SD(X) = \sqrt{\frac{1}{3}} = \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3} \approx 0.577$$

Examples More....

1. Let X have the p.d.f. $f(x) = 4x^3$ when x is in [0,1] and

   zero elsewhere.

   $E[X] = {}_0\int^1 4x^4 dx = 4x^5/5\ {}_0|^1 = 4/5$

2. Let X be uniform ~ U[a,b].

   Then f(x) = 1/(b-a),

   $E[X] = {}_0\int^1 x\{1/(b-a)\}dx = \{1/2(b-a)\}x^2\ {}_0|^1$

   $= (b^2-a^2)/2(b-a) = (a+b)/2$

3. In particular if X ~U[0,1] then E[X] = 1/2

Feed Back to mallik.reddyd@gmail.com