MODULE-V

- DISCRETE AND CONTINUOUS RANDOM VARIABLES
- CUMULATIVE DISTRIBUTION FUNCTION
- PROBABILITY MASS FUNCTION
- PROBABILITY DENSITY FUNCTION
- EXPECTATION
- VARIANCE
- MOMENT GENERATING FUNCTION
- INTRODUCTION TO BINOMIAL,POISSON AND NORMAL DISTRIBUTION.

Random Variables

- If a real variable X be associated with the outcome of a random experiment, then since the values which X takes depend on chance, it is called a random variable.
- For example,if a random experiment E consists of tossing a pair of dice,the sum X of the two numbers which turn up have values 2,3,…,12 depending on chance.Then X is the random variable.It is a function whose values are real numbers and depend on chance.

Discrete Random Variables

- If a random variable takes a finite set of values it is called discrete random variables.
- Example the number of heads in 4 tosses of a coin is a discrete random variable as it cannot assume values other than 0,1,2,3,4.
- Another example where the number of aces in a draw of 2 cards from a well shuffled deck is a random variable as it can takes the values 0,1,2 only.

Discrete Probability Distribution(DPD)

- Suppose a discrete variate X is the outcome of some experiment.If the probability that X takes the values x_i is p_i,then

  $P(X=x_i)=p_i$ or $p(x_i)$ for i=1,2,...

  where

  i)$p(x_i)\geq 0$ for all values of i,

  ii)$\sum p(x_i)=1$

  The set of values x_i with their probabilities p_i constitute a discrete probability distribution of the discrete variable X.

  Example:The DPD for X,the sum of the numbers which turn on tossing a pair of dice is given by the following table:

X=xi	2	3	4	5	6	7	8	9	10	11	12
P(xi)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Since there are 6x6=36 equally likely outomes and each has the probability 1/36.When X=2 for 1 outcome,i.e.(1,1);X=3 for two outcomes (1,2) and (2,1) and so on.

Question

A random variable X has the following probability function:

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. Evaluate $P(X<6)$, $P(X\geq6)$, $P(3<X\leq6)$

iii. Find the maximum value of x so that $P(X\leq x) > 1/2$.

SOLUTION

$$i) \quad \sum_{x=0}^{7} p(x) = 1$$

$0+k+2k+2k+3k+k^2+2k^2+7k^2+k = 1$

$10k^2+9k-1=0, k=1/10$

ii) $P(X<6)=P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)$

$=$ **$0+k+2k+2k+3k+k^2$**

$=81/100$

$P(X\geq 6)=P(X=6)+P(X=7)=2k^2+7k^2+k = 19/100$

$P(3<X\leq 6)=P(X=4)+P(X=5)+P(X=6)=3k+k^2+2k^2=33/100$

iii. $P(X\leq 1)=k=1/10<1/2$;

$P(X\leq 2)=k+2k=3/10<1/2$; $P(X\leq 3)=k+2k+2k=5/10=1/2$; $P(X\leq 4)=k+2k+2k+3k$ $=8/10>1/2$.

The maximum value of x so that $P(X\leq x) >1/2$ is 4.

DISTRIBUTION FUNCTION OR CUMULATIVE DISTRIBUTION FUNCTION

The distribution function F(x) of the discrete variate X is defined by

$F(x)=P(X\leq x)=\sum_{i=1}^{x} p(x_i)$ where x is any integer.

MEAN AND VARIANCE OF RANDOM VARIABLES

Let X :$x_1, x_2, \ldots, x_n$

P(X):$p_1, p_2, \ldots, p_n$ be a discrete probability distribution.

Mean $= \mu = \sum x_i p_i / \sum p_i$

Variance$= \sigma^2 = \sum p_i (x_i - \mu)^2$

Standard deviation$= \sqrt{(\text{Variance})}$

CONTINUOUS RANDOM VARIABLES

- A continuous random variable is one which can assume any value within an interval.
- Examples:The height/weight of a group of individuals.
- The probability distribution of a continuous variate x is defined by a function f(x) such that the probability of the variate x falling in the small interval $x-(1/2)dx$ to $x+(1/2)dx$.
- $P(x- (1/2)dx \le x \le x+(1/2)dx)=f(x)dx$.Then f(x) is called the probability density function.

PROBABILITY DENSITY FUNCTION

- The density function f(x) is always positive and

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

(The total area under the probability curve and the x-axis is unity)

DISTRIBUTION FUNCTION

- If F(x)=P(X≤x)= $\int_{-\infty}^{\infty} f(x)dx$
- F(x) is defined as the cumulative distribution function or the distribution function of the continuous variate X. It is the probability that the value of the variate X will be ≤x.
- The distribution function F(x) has the following properties:

i.F'(x)=f(x) ≥0,so that F(x) is non-decreasing function.

ii.F(-∞)=0; iii)F(∞)=1.

iv)P(a≤x≤b)= $\int_{a}^{b} f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = F(b) - F(a)$

PROBLEMS

Is the function defined as follows a density function?

$f(x) = e^{-x}, x \geq 0$

$= 0, x < 0$

Clearly $f(x) \geq 0$ for every x in (1,2) and

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} 0.dx + \int_{0}^{\infty} e^{-x}dx = 1$$

The function f(x) satisfies the requirements for a density function.

EXPECTATION,VARIANCE AND MOMENT GENERATING FUNCTION

- The mean value (μ) of the probability distribution of a variate X is commonly known as its expectation and denoted by E(X).
- If f(x) is the probability density function of the variate X,then

$$E(X) = \sum_{i} x_i f(x_i)$$ Dicrete Distribution

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

VARIANCE

Discrete Distribution

$$\sigma^2 = \sum_I (x_i - \mu)^2 f(x_i)$$

Continuous Distribution

$$\sigma^2 = \int_{-\infty}^{\infty} (x_i - \mu)^2 f(x_i)$$

Problems

A variate X has the probability distribution

x:-3,6,9

P(X=x):1/6,1/2,1/3

Find $E(X), E(X^2), E(2X+1)^2$

$E(X)=-3*1/6+6*1/2+9*1/3=11/2$

$E(X^2)=9*1/6+36*1/2+81*1/3=93/2$

$E(2X+1)^2 = E(4X^2+4X+1)=4E(X^2)+4E(X)+1$

$=4*93/2+4*11/2+1=209$

MOMENT GENERATING FUNCTION

- The moment generating function of the discrete probability distribution of the variate X about the value x=a is defined as the expected value of $e^{t(x-a)}$ and is denoted by

  $M_a(t)=\sum p_i\ e^{t(x_i - a)}$ which is a function of the parameter t only.
- $M_a(t)= e^{-at}\sum p_i\ e^{tx_i}$
- If f(x) is the density function of a continuous variate X then m.g.f of this continuous probability distribution about x=a is given by:

  $M_a(t)= \int_{-\infty}^{\infty} e^{t(x-a)} f(x)dx$

PROBLEMS

Find the m.g.f of the exponential distribution

f(x)= $\frac{1}{c}e^{\frac{-x}{c}}, 0 \leq x \leq \infty, c > 0$

Find its mean and standard deviation.

$M_o(t)$= $\int_0^\infty e^{tx}\frac{1}{c}e^{\frac{-x}{c}}dx = \frac{1}{c}\int_0^\infty e^{(t-1/c)x}dx$

=$1+ct+c^2t^2+c^3t^3+\ldots$

mean= $\frac{d}{dt}[M_o(t)]_{t=0}$

=c

Standard deviation= $\frac{d^2}{dt^2}[M_o(t)]_{t=0}$

=c^2

BINOMIAL DISTRIBUTION

- It is concerned with trials of a repetitive nature in which only the occurrence or non-occurrence, success or failure , acceptance or rejection , yes or no of a particular event is of interest.
- Let there be n independent trials in an experiment . Let a random variable X denote the number of successes in these n trials . Let p be the probability of a success and q that of a failure in a single trial so that p+q =1.The probability of r successes in n trials

$$P(X=r)= {}^{n}C_{r}\, p^{r} q^{n-r}, \text{ where } p+q=1 \text{ and } r=0,1,2,\ldots,n.$$

The above distribution is called binomial probability distribution and X is called the binomial variate.

NOTE

- In binomial distribution:

  1.n,the number of trials is finite.

  2.Each trial has only two possible outcomes called success and failure.

  3.All the trials are independent.

  4.p(and hence q) is constant for all the trials.

MEAN AND VARIANCE OF THE BINOMIAL DISTRIBUTION

- Mean μ= $\sum_{r=0}^{n} rP(r) = \sum_{r=0}^{n} r.{}^{n}C_{r} q^{n-r} p^{r} = np$
- Variance σ^2 = $\sum_{r=0}^{n} r^{2} P(r) - \mu^{2} = npq$

Problem

- In 256 sets of 12 tosses of a coin, in how many cases one can expect 8 heads and 4 tails.

P(Head)=1/2 and P(tail)=1/2.

P(X=8)= $^{12}C_8 (\frac{1}{2})^8 (\frac{1}{2})^4 = \frac{495}{4096}$

The expected number of such cases in 256 sets=256*P(X=8)=30.9=31(approx)

Problem

- In sampling a large number of parts manufactured by a machine , the mean number of defectives in a sample of 20 is 2.Out of 1000 such samples, how many would be expected to contain at least 3 defective parts.

Here mean number of defectives 2=np=20p

Probability of defective part is p=2/20=.1

Probability of non-defective is q=.9

The probability of at least three defectives in a sample of 20=1-(prob. That either none,or one,or two are non-defective parts)

$=1-[{}^{20}C_0(0.9)^{20} + {}^{20}C_1(0.1)(0.9)^{19} + {}^{20}C_2(0.1)^2(0.9)^{18}]$

=.323

POISSON DISTRIBUTION

If the parameters n and p of a binomial distribution are known , we can find the distribution.But when n is very large and p is very small,binomial distribution is very labourious.As $n \rightarrow \infty$ and $p \rightarrow 0$ such that np always remains finite say λ,we get the Poisson approximation to the binomial distribution.

$$P(X=r) = \frac{\lambda^r e^{-\lambda}}{!r}$$ r=0,1,2,...where λ=np.

Mean and Variance of the Poisson Distribution

- Mean μ = $\sum_{r=0}^{\infty} rP(r) = \lambda$

- Variance σ^2 = $\sum_{r=0}^{\infty} r^2 P(r) = \lambda$

The mean and variance of the Poisson distribution are equal to λ.

Problems

If the probability of a bad reaction from a certain injection is 0.001,determine the chance that out of 2000 individuals more than two will get a bad reaction.

Mean=λ=np=2000*0.001=2

Probability that more than 2 will get a bad reaction=1-[prob. That no one gets a bad reaction+prob that one gets a bad reaction +prob that two get bad reaction]

$=1-[e^{-m}+me^{-m}+m^{2}e^{-m}/!2]=1-5/e^{2}=.32$

Problem

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 on which neither car is used and proportion of days on which some demand is refused.($e^{-1.5}$=.2231)

λ=1.5

Proportion of days on which neither car is used=Probability of there being no demand for the car=$m^0e^{-m}/!0=e^{-1.5}$=.2231.

Proportion of days on which some demand is refused=Probability for the number of demands to be more than two=1-P(x≤ 2)=1-($e^{-m}+me^{-m}+m2e^{-m}$ /!2)

$=1-e^{-1.5}(1+1.5+(1.5)^2/2)$=.1912625

NORMAL DISTRIBUTION

- The normal distribution is a continuous distribution.
- It can be derived from the binomial distribution in the limiting case when n,the number of trials is very large and probability of a success is close to ½.
- The general equation of the normal distribution is given by:

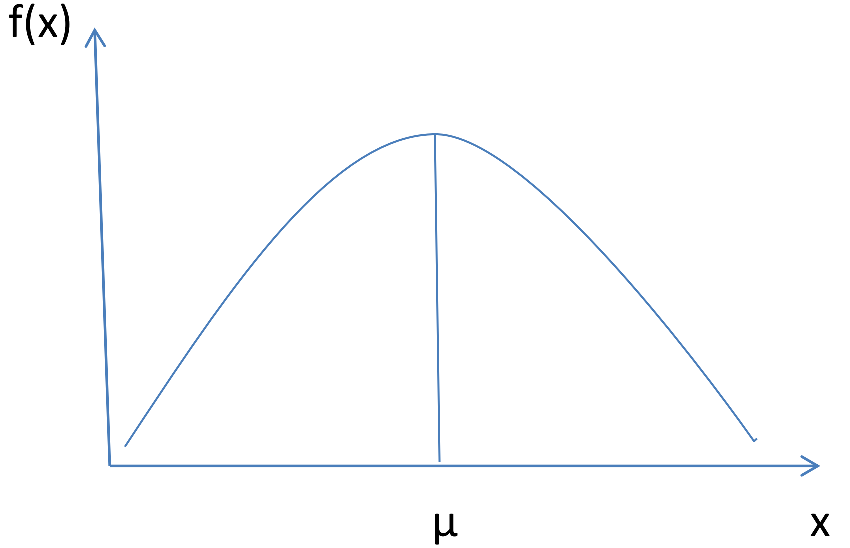
$$f(x)= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1(\frac{x-\mu}{\sigma})^2}{2}}$$

Where the variable x can assume all values from -∞ to ∞.μ and σ called the parameters of the distribution are the mean and standard deviation of the distribution and $-\infty<\mu<\infty,\sigma>0$.x is called the normal variate and f(x) is called the probability density function of the normal distribution.

If a variable x has the normal distribution with mean μ and standard deviation σ,we briefly write $x:N(\mu,\sigma^2)$

NORMAL CURVE

- The graph of the normal distribution is called the normal curve.
- It is bell shaped and symmetrical about the mean μ.



Normal Curve

- The two tails of the curve extend from +∞ to -∞ towards the positive and negative directions of the x-axis and gradually approaches the x-axis without ever meeting it.
- The curve is unimodal and the mode of the normal distribution coincides with its mean μ.The line $x=\mu$ divides the area under the normal curve above x-axis into two equal parts.
- The median of the distribution also coincides with its mean and mode.
- The area under the normal curve between any two given ordinates $x=x_1$ and $x=x_2$ represents the probability of values falling into the given interval.
- The total area under the normal curve above the x-axis is 1.

STANDARD FORM OF THE NORMAL DISTRIBUTION

- If X is a normal random variable with mean μ and standard deviation σ ,then the random variable Z is

$$Z = \frac{X - \mu}{\sigma}$$

The random variable Z is called the standardized normal random variable.

- The probability density function for the normal distribution in standard form is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2} * z^2}$$

NOTE

- If f(z) is the probability density function for the normal distribution,then

  $P(z_1 \leq Z \leq z_2) = \int_{z_1}^{z_2} f(z)dz = F(z_2) - F(z_1)$

  where $F(z) = \int_{-\infty}^{z} f(z)dz = P(Z \leq z)$

  The function F(z) defined above is called the distribution function for the normal distribution.

- The probabilities $P(z_1 \leq Z \leq z_2)$, $P(z_1 < Z \leq z_2)$ and $P(z_1 \leq Z < z_2)$ are all regarded to be the same.
- $F(-z_1) = 1 - F(z_1)$

Problems

A sample of 100 dry battery cells tested to find the length of life produced the following results:

$$\overline{x} = 12 hours, \sigma = 3 hours$$

Assuming the data to be normally distributed,what percentage of battery cells are expected to have life

i)more than 15 hours ii)less than 6 hours

iii)between 10 and 14 hours

Solution

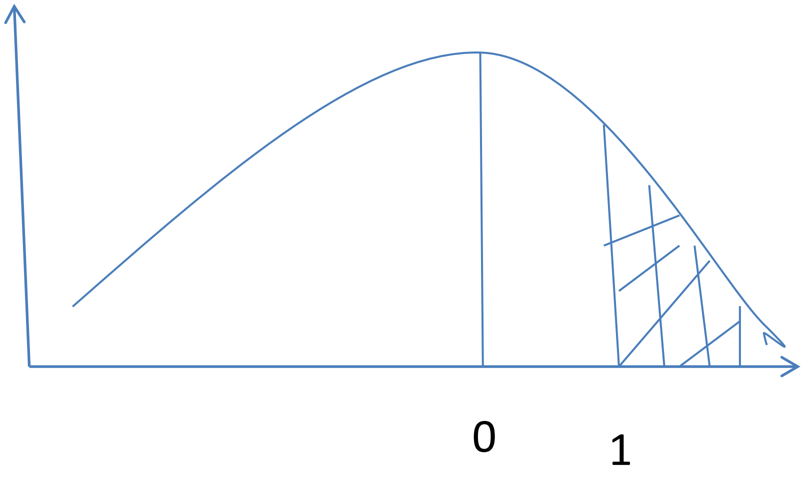
- Let x be the length of life of battery cells

$z=(x-\bar{x})/\sigma=(x-12)/3$

i) $x=15, z=1$

$P(x>15)=P(z>1)=P(0<z<\infty)-P(0<z<1)$

$=0.5-.3413=.1587$



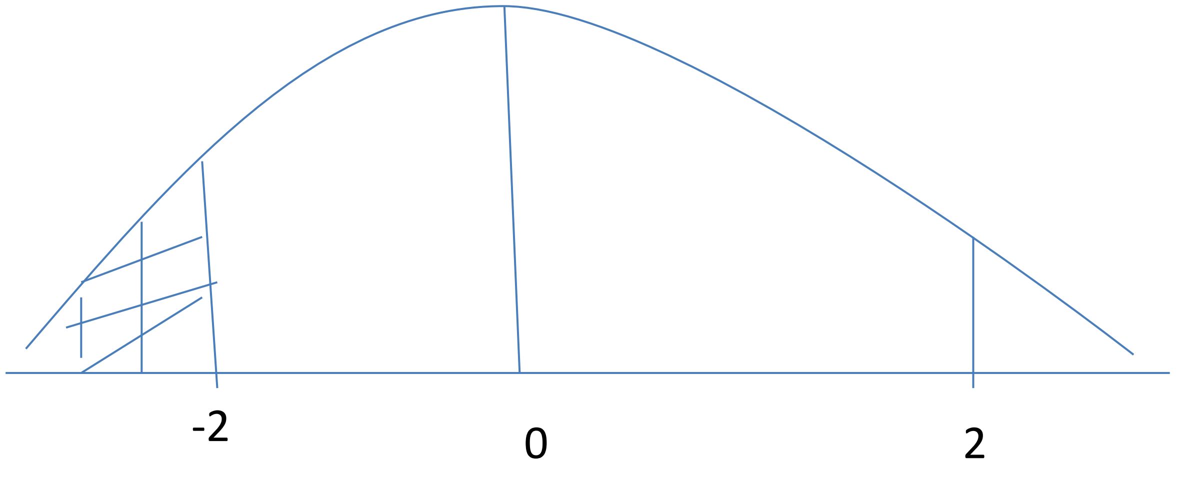
Criss cross lines show shaded area

Solution

ii)When x=6,z=-2

$P(x<6)=P(z<-2)=P(z>2)=P(0<z<\infty)-P(0<z<2)$

$=0.5-.4772=0.0228$



solution

iii.When x=10,z=-2/3=-.67

when x=14,z=2/3=.67

P(10<x<14)= P(-.67<z<.67)

=2P(0<z<.67)

=2*.2487

=.4974