Unit-V

Syllabus: Discrete Distribution : Binomial & Poisson Distribution with their constant , Moment Generating function , Continuous distribution : Normal Distribution, Properties , Constants ,Moments.
Curve Fitting using Least Square Method.

Random variable – has a single numerical value, determined by chance, for each outcome.

Probability distribution – represents all values of a random variable and probability of each value.

Theoretical Distributions
Definition : When frequency distribution of some universe are not based on actual observation or experiments , but can be derived mathematically from certain predetermined hypothesis , then such distribution are said to be theoretical distributions.

Types of Theoretical Distributions: Following two types of Theoretical Distributions are usually used in statistics:

1) Discrete Probability Distribution
   a) Binomial Distribution
   b) Poisson Distribution
2) Continuous Probability Distribution
   Normal Distribution

Binomial Distribution:

1. The procedure has a **fixed number of trials**. [n trials]
2. The trials must be **independent**.
3. Each trial is in **one of two mutually exclusive categories**.
4. The **probabilities remain constant** for each trial.

Notations:

P(success) = P(S) = p probability of success in one of the n trials
P(failure) = P(F) = 1 – p = q probability of failure in one of the n trials

n = fixed number of trials; x = number of successes, where $0 \leq x \leq n$

$P(x)$ = probability of getting exactly x successes among the n trials

$P(x \leq a)$ = probability of getting x-values less than or equal to the value of a.

$P(x \geq a)$ = probability of getting x-values greater than or equal to the value of a.

NOTE: Success (failure) does not necessarily mean good (bad).

Formula for Binomial Probabilities: $P(x) = \frac{n!}{(n-x)!x!} p^x q^{n-x} \; for \; x = 0, 1, 2, ..., n$

Factorial definition: $n! = n(n-1)(n-2)\cdots 2\cdot 1$; $0! = 1$; $1! = 1$

Example (Formula): Find the probability of 2 successes of 5 trials when the probability of success is 0.3.

$$P(x=2) = \frac{5!}{(5-2)!2!} 0.3^2 0.7^{5-2} = \frac{5\cdot 4\cdot 3!}{3!\cdot 2!}(0.09)(0.343) = 10(0.03087) = 0.3087$$

Moment about the origin:

1) **First moment about the origin:**

$$\mu_1' = \sum_{r=0}^{n} r.(nCr)p^r q^{n-r}$$
$$= np$$

2) Second moment about the origin:

$$\mu_2' = \sum_{r=0}^{n} r^2.(nCr)p^r q^{n-r}$$
$$= npq + n^2p^2$$

Moment about the Mean:

1) First moment about the mean is 0.
2) Second moment about the mean or variance is given by=npq

   Standard deviation=$\sqrt{npq}$

Examples:

(1) Six dice are thrown 729 times. How many times do you expect at least three dice to show a five or six?

Solution:) We know that when a die is thrown , the probability to show a 5 or 6 = 2/6 = 2/3 = p (say)

q= 1-p = 1- (1/3) = 2/3

The probability to show a 5 or 6 in at least 3 dice

= $\sum_{x=3}^{6} p(x)$ = p(3) + p(4) + p(5) +p(6), where p(x) is the probability to show 5 or 6

$= {}^6C_3 q^3 p^3 + {}^6C_4 p^4 q^2 + {}^6C_5 p^5 q + {}^6C_6 p^6 = \frac{233}{729}$ = p (say)

SO the required no. = np = 233

(2) The mean and variance of a binomial variate are 16 & 8. Find i) P (X= 0)

ii) P(X ≥ 2)

Mean = ? = np = 16

Variance = $?^2$ = npq = 8

npq / np = 8/16 = 1/2

ie, q = ½

p = 1 – q = ½

np = 16 ie, n = 32

i) $P(X = 0) = nC_0\, p^0 q^{n-0}$

$= (½)^0 (1/2)^{32}$

$= (1/2)^{32}$

ii)P(X ≥ 2) = 1 – P (X < 2)

= 1 – P(X = 0 ,1)

= 1 – P(X = 0) – P(X = 1)

$= 1 - 33\ (1/2)^{32}$

3) Six dice are thrown 729 times.Howmany times do you expect atleast 3

dice to show a 5 or 6 ?

Solution : Here n = 6 ,N = 729

$P(x \geq 3) = 6C_x\, p^x q^{\,n-x}$

Let p be the probability of getting 5 or 6 with 1 dice

ie, $p = 2/6 = 1/3$

$q = 1 – 1/3 = 2/3$

$P(x \geq 3) = P\ (x = 2,3,4,5,6)$

$= p(x=3)+p(x=4)+p(x=5)+p(x=6)$
$=0 .3196$

number of times = 729*0.3196 = 233

4)A basket contains 20 good oranges and 80 bad oranges . 3 oranges are drawn at random from this basket . Find the probability that out of 3 i) exactly 2 ii)atleast 2 iii)atmost 2 are good oranges.

Solution: Let p be the probability of getting a good orange

ie, $p = \dfrac{80C_1}{100C_1}$
$= 0.8$

$q = 1- 0.8 = 0.2$

i) $p\ (x=2) = 3C2\ (0.8)^2(0.2)^1 = 0.384$

ii) $p(x\geq2) = P(2) +p(3)= 0.896$

iii) $p(x\leq 2) = p(0) +p(1) + p(2) = 0.488$

5) In a sampling a large number of parts manufactured by a machine , the mean number of defective in a sample of 20 is 2. Out of 1000 such samples howmany would expected to contain atleast 3 defective parts.

$n=20 \quad n\,p =2$

ie , $p=1/10 \qquad q = 1-p = 9/10$

$p(x \geq 3) = 1 – p\ (x < 3)$

= 1 – p(x = 0,1,2) = 0.323

Number of samples having atleast 3 defective parts = 0.323 * 1000

= 323

The process of determining the most appropriate values of the parameters from the given observations and writing down the probability distribution function is known as fitting of the binomial distribution.

Problems

1) Fit an appropriate binomial distribution and calculate the theoretical distribution

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

Here n = 5 , N = 100

$$\text{Mean} = \frac{\sum x_i f_i}{\sum f_i} = 2.84$$

np = 2.84

p = 2.84/5 = 0.568

q = 0.432

$p(r) = 5C_r\,(0.568)^r\,(0.432)^{5-r}$, r = 0,1,2,3,4,5

Theoretical distributions are

r	p(r)	N* p(r)
0	0.0147	1.47 = 1
1	0.097	9.7 =10
2	0.258	25.8 =26
3	0.342	34.2 =34
4	0.226	22.6 =23
5	0.060	6 =6
		Total = 100

Poisson Distribution :.

The **Poisson distribution** is a discrete distribution. It is often used as a model for the number of events (such as the number of telephone calls at a business, number of customers in waiting lines, number of defects in a given surface area, airplane arrivals, or the number of accidents at an intersection) in a specific time period.

The mean is λ. The variance is λ.

Therefore the P.D. is given by

$P(r) = \frac{e^{-m}m^r}{r!}$ where r=0,1,2,3…

m is the parameter which indicates the average number of events in the given time interval.

Poisson distribution examples

1. The number of road construction projects that take place at any one time in a certain city follows a Poisson distribution with a mean of 3. Find the probability that exactly five road construction projects are currently taking place in this city. (0.100819)

2. The number of road construction projects that take place at any one time in a certain city follows a Poisson distribution with a mean of 7. Find the probability that more than four road construction projects are currently taking place in the city. (0.827008)

3. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 7.6. Find the probability that less than three accidents will occur next month on this stretch of road. (0.018757)

4. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 7. Find the probability of observing exactly three accidents on this stretch of road next month. (0.052129)

5. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 6.8. Find the probability that the next two months will both result in four accidents each occurring on this stretch of road. (0.009846)

Examples:In a certain factory turning razor blades, there is a small chance (1/500) for any blade to be defective. The blades are in packets of 10. Use Poisson's distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10,000 packets.
Solution: Here p = 1/500, n = 10 , N = 10,000 so m = np = 0.02

Now e^{-m} = $e^{-0.02}$ = 0.9802

The respective frequencies containing no defective, 1 defective & 2 defective blades are given

As follows

Ne^{-m}, Ne^{-m}.m, N$e^{-m}.\frac{1}{2}m^2$

i.e. 9802 ; 196; 2

Normal Distribution:

The normal (or Gaussian) distribution is a continuous probability distribution that frequently occurs in nature and has many practical applications in statistics.

Characteristics of a normal distribution

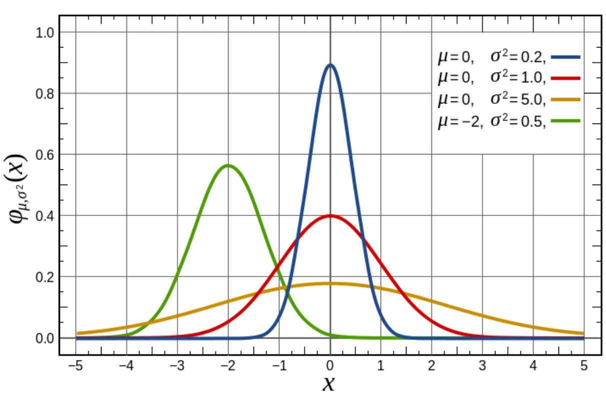
- Bell-shaped appearance
- Symmetrical
- Unimodal
- Mean = Median = Mode
- Described by two parameters: mean (μ_x) and standard deviation (σ_x)
- Theoretically infinite range of *x*: ($-\infty < x < +\infty$)

- The normal distribution is described by the following formula:

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

where the function f(x) defines the probability density associated with X = x. That is, the above formula is a probability density function

Because μ_x and σ_x can have infinitely many values, it follows there are infinitely many normal distributions:



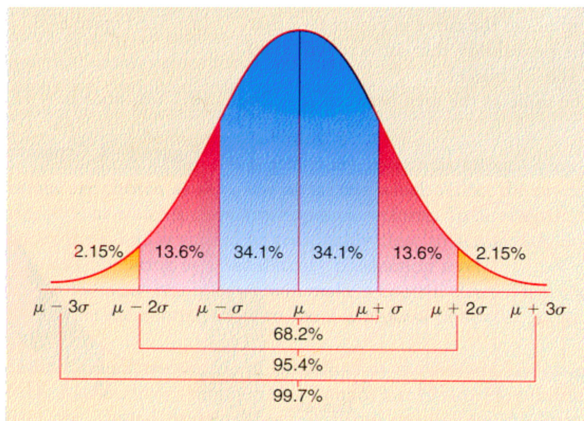
.

A standard normal distribution is a normal distribution rescaled to have $\mu_x = 0$ and $\sigma_x = 1$. The *pdf* is:

$$f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}} \quad \infty < x < \infty$$

The ordinate of the standard normal curve is no longer called *x*, but *z*.

For a normal curve, approximately 68.2%, 95.4%, and -99.7% of the observations fall within 1, 2, and 3 standard deviations of the mean, respectively.



Areas Under the Normal Curve

By standardizing a normal distribution, we eliminate the need to consider μ_x and σ_x; we have a standard frame of reference.

Areas Under the Standard Normal Curve

X (x values) of a normal distribution map into Z (z-values) of a standard normal distribution with a 1-to-1 correspondence.

If X is a normal random variable with mean μ_x and σ_x, then the standard normal variable (normal deviate) is obtained by:

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

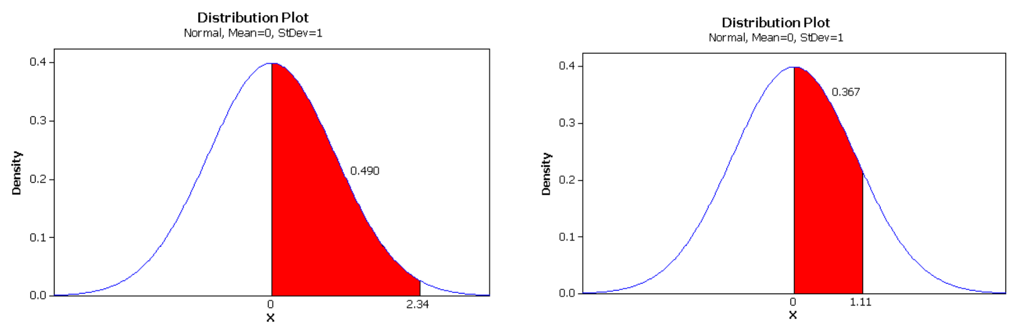
Example 1: What is the probability that Z falls z = 1.11 and z = 2.34?

$\Pr(1.11 < z < 2.34)$

= area from $z = 2.34$ to $z = 1.11$

= area from ($-\infty$ to $z = 2.34$) minus area from ($-\infty$ to $z = 1.11$)

= .9904 – .8665 = .1239



Note: figures above should also shade region from –inf to 0.

Practice Question:

Q.1 (a): From a pack of 52 cards ,6 cards are drawn at random. Find the probability of the following events:

(a) Three are red and three are black cards (b) three are king and three are queen

Q.1 (b): Out of 800 families with 4 children each, how many families would be expected to have?

(a) 2 boys & 2 girls (b) at least one boy (c) no girl (d) at most 2 girls?

Assume equal probabilities for boys and girls.

Q.2(a):One bag contain 4 white,6 red & 15 black balls and a second bag contains 11 white, 5 red& 9 black balls. One ball from each bag is drawn. Find the probability of the following events:(a) both balls are white (b) both balls are red (c) both balls are black (d)both balls are of the same colour.

Q.2(b): Find mean and variance of Binomial distribution .

Q.3(a): In a bolt factory, machines A,B & C manufacture respectively 25%,35% & 40% of the total. Of their output 5,4,2 percent of defective bolts. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by machine A, B or C.

Q.3(b): In a normal distribution 31% of the items are under 45and 8% are over 64. Find the mean and S.D. of the distribution.

Q.4(a): A sample of 100 dry battery cells tested to find the length of life produced the following results:

μ=12hours , σ=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 & 14 hours

Q.4(b): Define the following:

(a) Random Variable (b) Mathematical Expectation (c) Moment generating function

Curve Fitting using Least Square Method:

Curve fitting is a general problem to find the equation of an approximate curve fitting to the given data.

Fitting a Straight Line:

Suppose we require to fit a straight line Y=a+bx

So, Its Normal equation are given by

$\sum y = n a + b \sum x$

$\sum xy = a\sum x + b\sum x^2$

Now, Solve this equation for a & b & put the value of a& b in standard equation of Straight line.

Fitting A Second Degree Parabola:

Standard equation is

$Y = a + bx + cx^2$

And normal equations are given by

$\sum y = na + b\sum x + c\sum x^2$

$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$

$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$

Now, Solve this equation for a , b&c. & put the value of a,b,c in standard equation of Second degree parabola.

Example(1) Fit a straight line to the following data **:**

x	0	1	2	3	4
y	1	1.8	3.3	4.5	6.3

Solution(1) Suppose a st. line to be fitted to the given data is as follows , y = a + bx ………… (1)

Then the normal eqns. Are as : $\sum y$ = ma + b $\sum x$ …………(2)

$\sum xy$ = a$\sum x$ + b $\sum x^2$ ………..(3)

We have the data

X	y	xy	x^2
0	1	0	0
1	1.8	1.8	1

2	3.3	6.6	4
3	4.5	13.5	9
4	6.3	25.2	16

Total

$\sum x$ =10 $\sum y$= 16.9 $\sum xy$= 47.1 $\sum x^2$= 30

Here m = 5, sub. These values in the normal eqns., we get

16.9 = 5a + 10b ………..(4), 47.1 = 10a +30b ………..(5)

Solving (4) & (5), we get a = 0.72 & b = 1.33.

Thus the required eqn. of the st. line is y = 0.72 + 1.33x

Practice Question :

1) Fit a second degree parabola to the following data:

x	0	1	2	3	4
y	1	5	10	22	38

****