

Total Quality Management-II
Prof. Raghunandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 03
Distribution of a Random Variable-I

A very good morning, good afternoon, good evening my dear friends I am Raghunandan Sengupta, from the IME department IIT Kanpur, and welcome to this TQM 2 lecture three. And as you know that we have been discussing over the basic backgrounds, I will try to finish with within 2 classes or 2 lectures and hopefully from the fifth one; we will start about the concepts of design of experiments, which is the main bulk portion of that and once you see, the I did not say these words in the first 2 lectures.

But when you see the introductory video you will understand what actuality you came to waste because I did not want to go into again the nitty gritty of the coverage of that you came, and once you see the syllabus see the video lectures what the three formats one which is the introductive 1 for the TQM 2 you will understand. So, consider considering our discussion which was there in the concept of p d f p m f c d f and the corresponding concepts so, let us see this example.

(Refer Slide Time: 01:26)

Example 6

In an examination each question has four alternatives, answer of which only one is correct. If a student knows the correct alternative then he/she is definitely able to identify it. Otherwise he/she picks one of the alternatives at random. Given that a student has identified the correct alternative what is the conditional probability that he/she knew it, assuming 70% of the student know the correct alternative to the question under consideration.

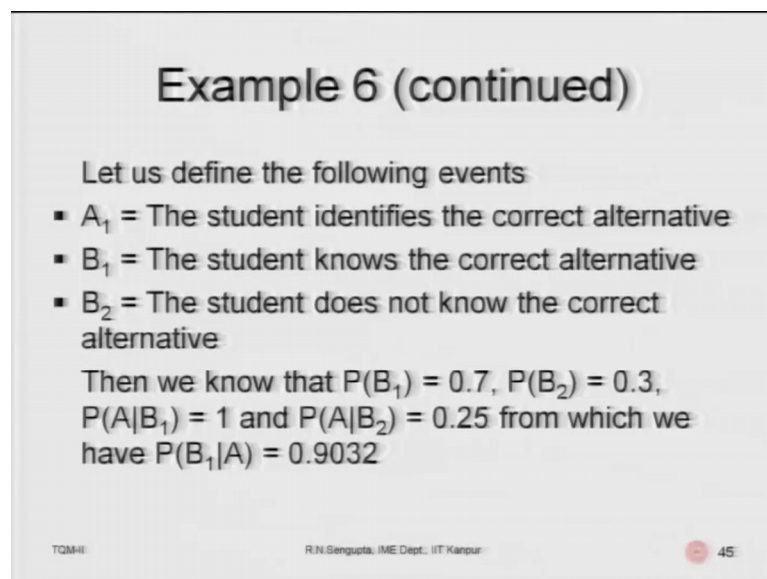
TQM-IIR.N. Sengupta, IME Dept., IIT Kanpur44

In an example each question has four alternatives of answers of which is one is correct. So, if you want if you are reading this problem we will understand is something to do

with Baye's theorem, and the concepts of Baye's theorem which I did discuss. If a student knows the correct alternative then he she is definitely able to identified, otherwise he she picks one of those alternatives at random given that; a students are identified the correct alternative what is the correct conditional probability that he she knew it assuming 70 percent of the student know the correct alternative the question under consideration, now 1 thing is absolutely clear from this problem.

So, if you know the problem you answer it correctly which is the probability one, and if you do not know the answer questions answer; obviously, you will make a guess and the probability of; obviously, would be a 1 by 4 because there are 4 alternative. So, this 2 bullet points which is very intuitive and very simple is easy to understand. So, with this we will proceed, so let us define the following events A 1 as the student identifies the correct alternative.

(Refer Slide Time: 02:26)



Example 6 (continued)

Let us define the following events

- A_1 = The student identifies the correct alternative
- B_1 = The student knows the correct alternative
- B_2 = The student does not know the correct alternative

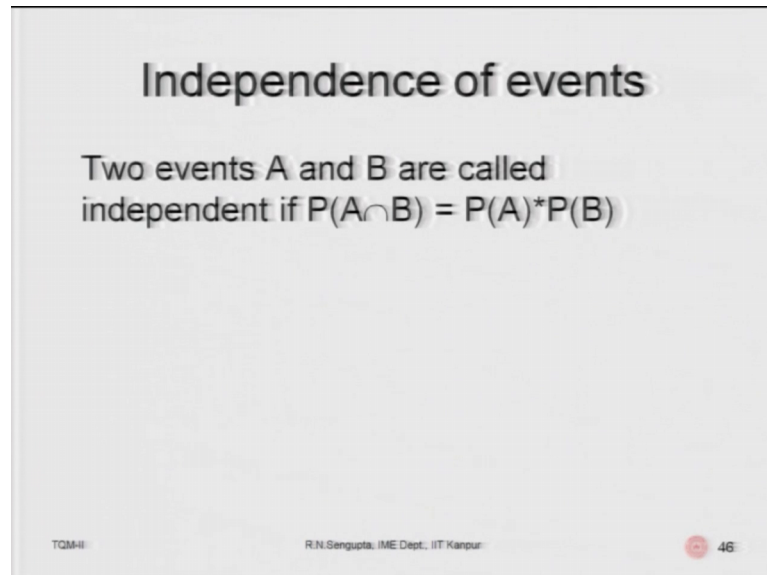
Then we know that $P(B_1) = 0.7$, $P(B_2) = 0.3$,
 $P(A|B_1) = 1$ and $P(A|B_2) = 0.25$ from which we
have $P(B_1|A) = 0.9032$

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 45

B 1 student knows the correct alternatives, B 2 student does not know so; obviously, B 1 and B 2 would be complementary then we know that the probability of that B 1; that means, 70 percent know is 0.7, B 2 which is the percentage who do not know is 0.3 given you know the question any of answer you will answer it; obviously, 1 as I said and given that you do not know the answer and you mark it correctly is one-fourth, from which you will have that given the student has identified the correct alternatives what is

the probability that the student knew the answer is given by the Baye's theorem is about 0.9032.

(Refer Slide Time: 03:11)



Now, we will cover two other concepts very simple one they would be used later on the 1 is the concept of 2 we will say the 2 events are called independent if the intersection of them is a multiplication of the probabilities of A and B.

So, correspondingly we can extend it to 3 and more events also, but for the timing we will only stick to the 2 the concept of for 2 events now we will go into the concept of distribution.

(Refer Slide Time: 03:39)

Distribution

Depending what are the outcomes of an experiment a random variable (r.v) is used to denote the outcome of the experiment and we usually denote the r.v using X , Y or Z and the corresponding probability distribution is denoted by $f(x)$, $f(y)$ or $f(z)$

- Discrete: probability mass function (pmf)
- Continuous: probability density function (pdf)

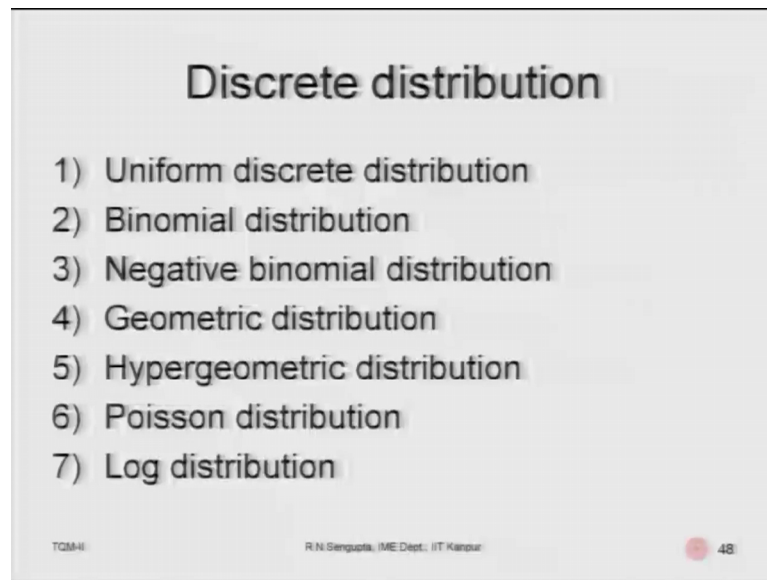
TQM/H
R.N.Singupta, IIM Dept., IIT Kanpur 47

If you remember we did mention about probability mass function probability density function, and you are discussing our CDFs and the corresponding overall view that how less than type greater than type can be used to we will consider those in a little bit more details because you will see later on some few of the distributions would be very important for the design of experiments and hypothesis testing.

So, depending what are the outcomes of an experiment a random variable is used to denote the outcome the of the experiment. And we usually denote the random variable using X or Y or Z , and the corresponding probability distribution or probability mass functions are denoted by small f of x small f of y small f of z and so on and so forth. So, they can be if you remember I did mention it in the second class they and the second lecture.

So, if the random variable is discrete you have the probability mass function if the random variables are continuous you have the probability distribution functions. So, this what is mentioned the last two bullet points discrete means probability mass function continuous means probability density function.

(Refer Slide Time: 04:48)



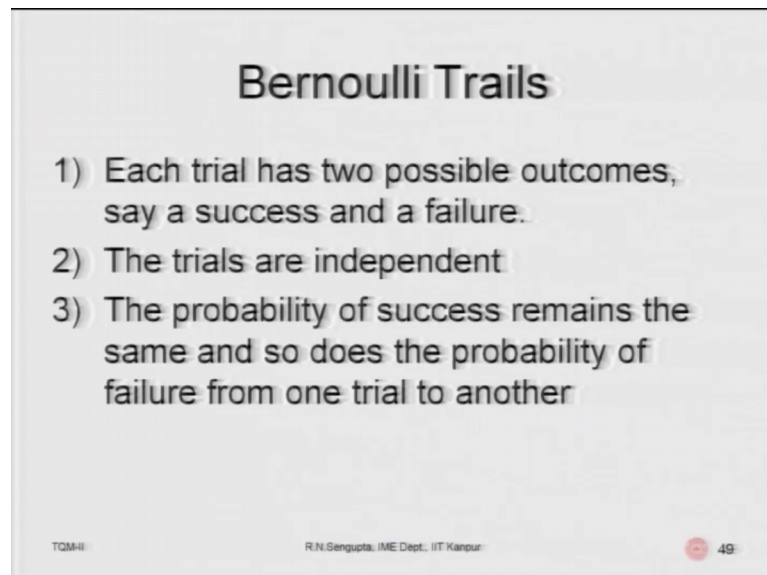
Discrete distribution

- 1) Uniform discrete distribution
- 2) Binomial distribution
- 3) Negative binomial distribution
- 4) Geometric distribution
- 5) Hypergeometric distribution
- 6) Poisson distribution
- 7) Log distribution

TQM-II R.N.Singupta, IIM Dept., IIT Kanpur 48

So, under the discrete we would have basically uniform, we will leave more emphasis on few of them will come to that we have uniform you have binomial, negative binomial, geometric, hyper geometric, poisson log distributions. So, these are the few important one; obviously, there are other we are not going to discuss them in this course.

(Refer Slide Time: 05:10)



Bernoulli Trials

- 1) Each trial has two possible outcomes, say a success and a failure.
- 2) The trials are independent
- 3) The probability of success remains the same and so does the probability of failure from one trial to another

TQM-II R.N.Singupta, IIM Dept., IIT Kanpur 49

So, now, let us see what is Bernoulli Trials? So, each trial has basically when you are basically rolling a dying. Now, this concept of Bernoulli trial is only applicable for the no only I would not see the use the word only applicable. We are basically discussing that

for two trials for any outcome there are two when you throw a die or toss a coin they would be how would the throw of the die would be important I will come to that when. So, there are two outcomes positive negative plus minus yes, no, red, blue, black, white, so on and so forth.

So, when you toss the coin it is head tail, so there are two when now the question would definitely be coming in your mind that when you roll the die how they are to. So, consider we are only interested in finding out the even and the odd. So, the; obviously, there would be background concept will be they are complementary or say for example, we consider the numbers as less than equal to 4 and greater than 4 or the numbers may be say for example, equal to 1 and not equal to 1.

So, when you are rolling the die, so in this way we basically will have. So, called two outcomes and considering that that trial has trial whatever the trial you are doing. So, we will basically consider that is a Bernoulli trial provided these other concepts are also true; that means, for any trial outcomes would be would be as I said yes, no, black, white, red, blue, a good item bad item success failure.

So, these are the outcomes the trials would be independent, it means that if we keep rolling the die the in the probability of the first trial would not affect the second, second would not affect the third and so on and so forth, neither would be the second being affected by third and all this thing. So, we will consider that and we will also consider the probability of the success or the failure or you if you consider red or blue or if you consider black or white. They remain the same from trial to trial and; that means, they do not affect.

So, let us consider a very simple notion you are playing a game of rolling the die, and your head and your win and a loss in the initial stage and then you first start the play start playing the game is say for example, half and half now consider you start winning or you start losing whatever it is, then if the probability of you winning or losing in the next trial depending on what state you are whether you are in a winning position or a losing position. Then if that changes; obviously, this concept of Bernoulli trials would not hold true. So, it is basically independent from trial to trial and it will continue remaining that.

Now, before I discuss the uniform discrete distribution I did mention where I started the Bernoulli trial, it is only 2 trials say for example, if I have 3 trial trials or 4 trials;

obviously, they would not be Bernoulli trials, but they can be basically considered that in a very simple way where the sum of the concepts of the sum the intuitive field of Bernoulli trial can be brought into the picture in those cases also, so coming back to the distribution.

(Refer Slide Time: 08:32)

Uniform discrete distribution
 $[X \sim UD(a, b)]$

$f(x) = 1/n \quad x = a, a+k, a+2k, \dots, b$

- a and b are the parameters where $a, b \in \mathbb{R}$
- $E[X] = a + \frac{k}{2}(n-1)$
- $V[X] = k^2 \left(\frac{n^2 - 1}{12} \right)$
- Example: Generating the random numbers 1, 2, 3, ..., 10. Hence $X \sim UD(1, 10)$ where $a=1, k=1, b=10$. Hence $n=10$.

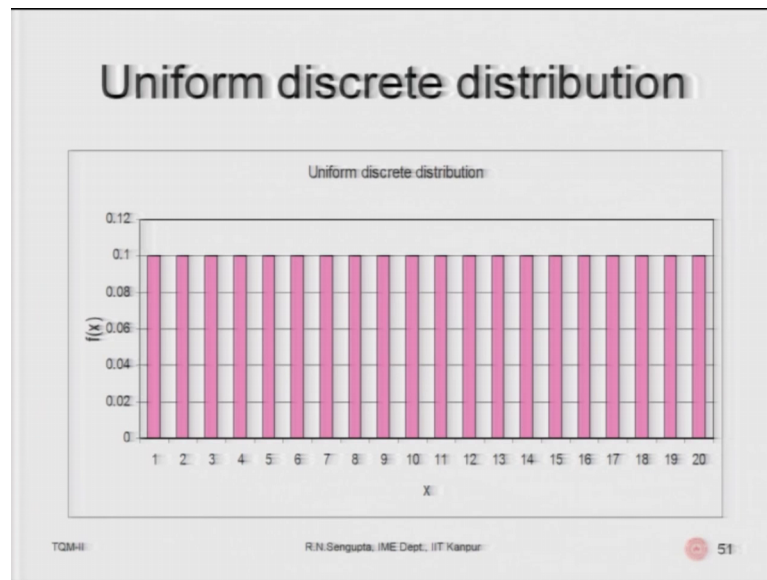
TQM-II
R.N.Sengupta, IIM Dept., IIT Kanpur
50

So, we will consider the uniform discrete distribution. So, that f of x is given by $1/n$ and a and b are the parameters such that a and b along the real line. So, a can be negative or positive 0, b can be negative or positive 0, but; obviously, b would be greater than a . The expected value is given by and by the way. So, this k would be the amount of increment which is happening, like say for example, you pick up chits and chits are marked as 1, 2, 3, 4 till 100.

So, in that case the k value would be 1 in case say for example, chits are marked as 1, 3, 5, 7, in that well case the k value is 2. So, expected value is given the variance is given, so example. So, I am just discussing a variance and expected value they would be if at all used in the cases we will just take these formulas and do our problems accordingly and they would not definitely not be any proofs.

So, example mean generating the random numbers 1 to 10 hence it is a uniform distributing 1 to 10 where the value of a as I said is 1 and I said means based on the example which I gave the value of B is 10 and the increments are 1. So, hence k is 1, so and; obviously, n is 10.

(Refer Slide Time: 09:49)



So, this is a uniform discrete distribution how it look like, starting from 1 to 20 for this case and the f of x which is the p d p m f probability mass function is plotted along the y axis.

(Refer Slide Time: 10:02)

Binomial distribution

$[X \sim B(p, n)]$

$$f(x) = {}^nC_x p^x q^{n-x} \quad x = 0, 1, 2, \dots, n$$

- n and p are the parameters where $p \in [0, 1]$ and $n \in \mathbb{Z}^+$
- $E[X] = np$
- $V[X] = npq$
- Example: Consider you are checking the quality of the product coming out of the shop floor. A product can either pass (with probability $p = 0.8$) or fail (with probability $q = 0.2$) and for checking you take such 50 products ($n = 50$). Then if X is the random variable denoting the number of success in these 50 inspections, we have:

$$X \sim {}^{50}C_x (0.8)^x (0.2)^{50-x}$$

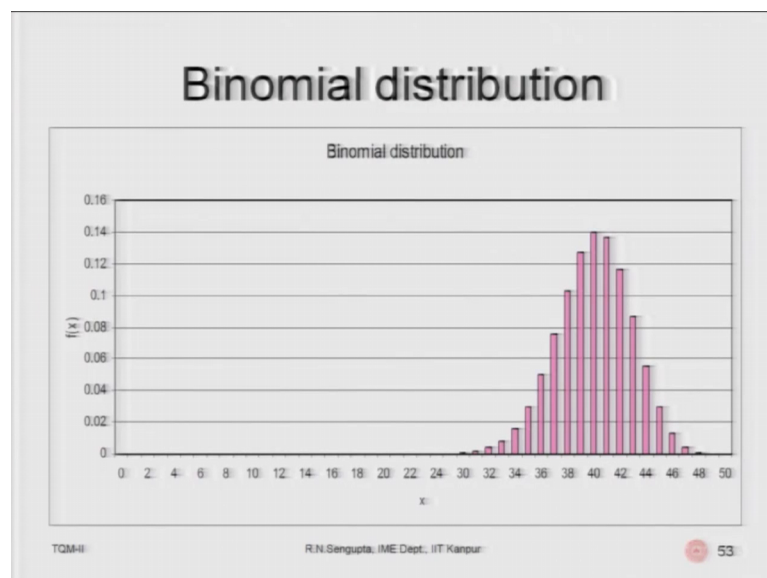
TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 52

Now, consider the binomial distribution, so it is basically comes from the binomial distribution which you know in general mathematics we must have done in class 10 or class 11. So, the f of x is given by $n C x p$ to the power $x q$ to the power n minus x . So, n and p are the parameters where p is basically between 0 and 1 and n is a number which is

integer; obviously, positive integers because n can be 3 4 5 6, whatever it is the expected value is given by np variance is given by npq .

So, consider you are checking the quality of the product coming out of shop floor or product and either pass or fail with probability point eight and 0.2, and for checking you take such 50 products. So, n is 50 then if x is the random variable denoting the number of successes in this 50 inspection. So, you will basically have ${}^{50}C_x p^x q^{n-x}$ to the power n minus x and here n is 50 as we know and the values of p and q are given as 0.8 and 0.2.

(Refer Slide Time: 11:09)



So, this is the binomial distribution just a sample which we have just brought it on excel and thus given it for our own benefit to understand. So, f of x which is the p m f along the y axis x values along x axis they start from 0 and can go on depending on whatever the n value is.

(Refer Slide Time: 11:31)

Solved example (Binomial distribution)

Question: 20% of the BSNL shares are owned by Mr. Murali Lal. A random sample of 5 shares is chosen. What is the probability that at most 2 of them will be found to be owned by Mr. Murali Lal?

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 54

So, a very simple, so we will consider very simple examples. So, question is 20 percent of BSNL shares are owned by Mister. Murali Lal. A random sample of 5 shares is chosen. What is the probability that at most 2 of them would be found to be owned by Murali Lal. So, you can basically solve it using the simple binomial distribution.

(Refer Slide Time: 11:49)

Solved example (Binomial distribution)

Answer: Here $p=0.2$, $q=0.8$. Hence the required probability is:

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$
$$P(X \leq 2) = {}^5C_0 0.2^0 0.8^5 + {}^5C_1 0.2^1 0.8^4 + {}^5C_2 (0.2)^2 (0.8)^3$$

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 55

Where p value is 0.2, q value is 0.8 and n value is given as 5. So, we want to find out the probability of x being less than equal to 2 so; obviously, the x will value take the value of 0 1 2 or other ways can you do is that basically 1 minus the complementary of that. So,

the probability comes out to be as given in the last line starting the first term being 5C_0 . 2 to the power 0.8 to the power 5 where now; obviously, and the x value is 0 and n minus x is 5 till the third value is given with 5C_2 to the power 2. So, negative binomial distribution is given and by the formula $r + x - 1$ c $r - 1$ or.

(Refer Slide Time: 12:39)

Negative binomial distribution

$[X \sim \text{NB}(p, r)]$

$$f(x) = {}^{r+x-1}C_{r-1} p^r q^x \quad x = r, r+1, \dots$$

- p and r are the parameters where $p \in [0, 1]$ and $r \in \mathbb{Z}^+$
- $E[X] = r/p$
- $V[X] = r/p^2$
- Example: Consider the example above where you are still inspecting items from the production line. But now you are interested in finding the probability distribution of the number of failures preceding the 5th success of getting the right product. Then, we have, considering $p=0.8, q=0.2$

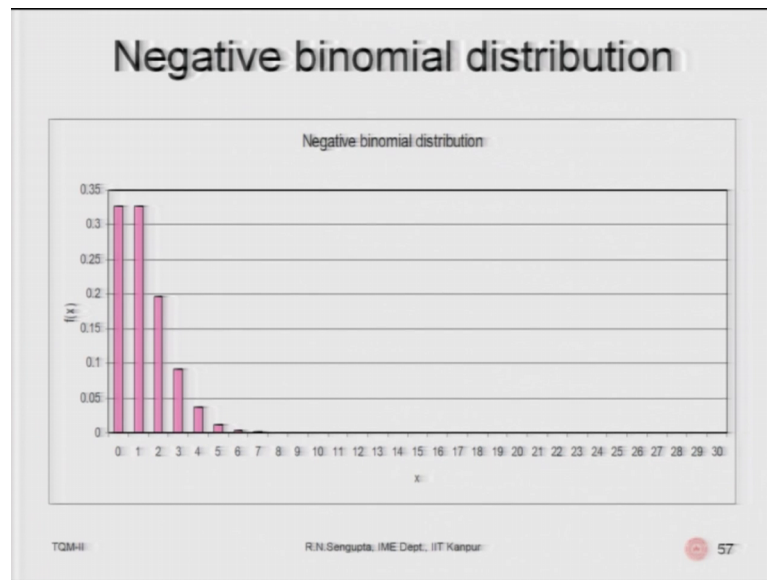
$$X \sim {}^{5+x-1}C_{5-1} (0.8)^5 (0.2)^x$$

TQM-II
R.N.Sengupta, IIM Dept., IIT Kanpur
56

${}^{r+x-1}C_{r-1}$ whatever way you are looking at p to the power r q to the power x ; so, x values basically take from r r plus 1 and it continuously in infinity. So, p and n are the parameters expected value is given by r/p and variance is given by r/p^2 .

So, consider example that the where you are still inspecting items from the production line, but now you are interested in finding the probability distribution on the number of failures predicting the fixed success getting the right product, then we will have considering p is equal to 0.8 q is equal to 0.2, the distribution of x is given as shown in the last line, so this is the negative binomial distribution.

(Refer Slide Time: 13:23)



With some values again we are plotting p m f p m f along the y axis x basically along the x axis you have the values for x .

(Refer Slide Time: 13:36)

Solved example (Negative binomial distribution)

Question: Suppose that the probability of a manufacturing process producing a defective item is 0.05. Suppose further that the quality of any one item is independent of the quality of any other item produced. If a Mr. Rao the quality control officer selects items at random from the production line and stops his inspection when he gets 10 good items. Then what is the probability of getting 2 defective items before he stops inspection?

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 58

So, suppose that the probability of manufacturing process producing a defective item is 0.05, suppose further that the quality of any 1 item is independent of the quality of any other item produced. So, consider Mister. Rao the quality control officer selects the items at random, from the production line and stops his inspection when he gets 10 good items then what is the probability of getting 2 defective items before he stops production.

(Refer Slide Time: 14:02)

Solved example (Negative binomial distribution)

Answer: Here $p=0.95$, $q=0.05$. Hence the required probability is

$$P(X=2) = {}^{10+2-1}C_{10-1} (0.95)^{10} (0.05)^2$$

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 59

So, you need to find it, so given p as 0.95, q is 0.05 hence the required probability is given as when x is equal to 2 and you have those values in front of you.

(Refer Slide Time: 14:16)

Geometric distribution

$[X \sim G(p)]$

$f(x) = pq^x \quad x = 0, 1, 2, \dots$

- p is the parameter where $p \in [0, 1]$
- $E[X] = q/p$ ($r = 1$ in the Negative Binomial distribution case)
- $V[X] = q/p^2$ ($r = 1$ in the Negative Binomial distribution case)
- Example: Consider the example above. But now you are interested in finding the probability distribution of the number of failures preceding the 1st success of getting the right product. Then, we have considering $p=0.8$, $q=0.2$

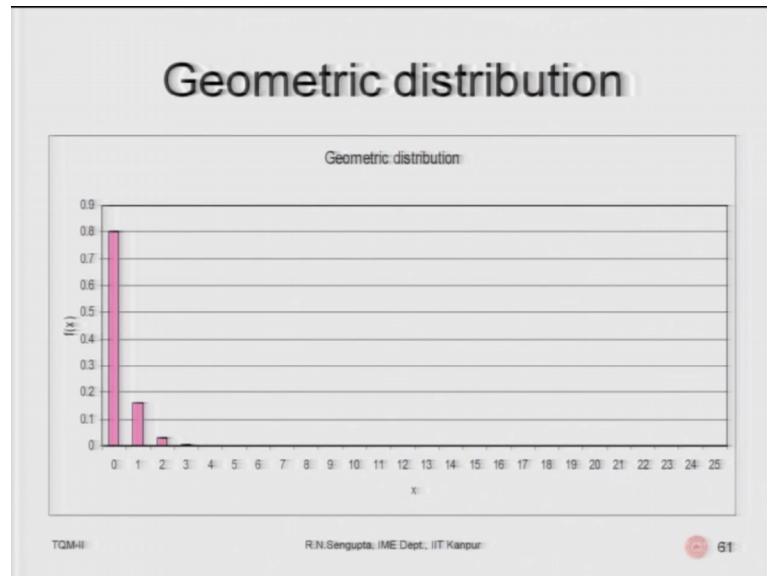
$X \sim (0.8)(0.2)^x$

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 60

Consider the geometric distribution, so geometric distribution f of x which is the p m f is given by p x to the power q . So, p to the power 1 and q to the power x , so technically it can be consider the case where you get the head or a tail for the first time very simply, p is the parameter and expected value is given by q by p .

So, if you remember in the earlier example it was r q by p . So, it is now q by p because r is 1 and variance is given by q by p square. So, consider the example above, but now you are interested in finding the probability distribution and the number of failures predicting the first success and getting the right products. So, the probabilities are found out given the values of p and q as 0.8 and 0.2.

(Refer Slide Time: 15:01)



So, this is the geometric distribution how it would look like.

(Refer Slide Time: 15:04)

Solved example (Geometric distribution)

Question: A recent study indicates that Colgate toothpaste has a market share of 45% (versus 55% of Pepsodent). Miss. Dabawallah the marketing research executive firm wants to conduct a new taste test for which she wants users of Colgate. Potential participants for the test are selected by random screening of users of toothpaste to find Colgate users. What is the probability that 5 users will have to be interviewed by Miss. Dabawallah to find the 1st Colgate toothpaste user?

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 62

So, a recent study indicates that Colgate toothpaste have a market share of 45 percent versus 55 percent of Pepsodent, Miss. Dabawallah the marketing research executive firm wants to conduct a new taste test, for which he wants users of Colgate potential participants from the testers selected by random screening of users of toothpaste to find the Colgate users. So, now, you want to find out that given if the probability the that there are 5 users then, you want to find out that probability that 5 users will have to be interviewed which is not 5 users my mistake.

So, they have to be interviewed, so if the probability is to be found out if there 5 users will have to be interviewed by Miss. Dabawallah to find out the 1st Colgate toothpaste. So, it can be the 1st trial; 2nd, 3rd, 4th and so on and so forth.

(Refer Slide Time: 15:59)

Solved example (Geometric distribution)

Answer: Here $p=0.45$, $q=0.55$. Hence the required probability is

$$P(X=4) = (0.45)^1(0.55)^4$$

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 63

So, p is given as 45 p q is given as 55, hence the probability that x equal to 4 we can find out from the formula as p to the power 1 and q to the power 4.

(Refer Slide Time: 16:14)

Hypergeometric distribution

$[X \sim \text{HG}(N, n, p)]$

$$f(x) = \frac{{}^Np_x {}^Nq_{n-x}}{{}^NC_n} \quad 0 \leq x \leq Np \text{ and } 0 \leq (n-x) \leq Nq$$

- N , n and p are the parameters
- $E[X] = np$
- $V[X] = npq\{(N-n)/(N-1)\}$
- Example: Consider the example above. But now you are interested in finding the probability distribution of the number of failures(success) of getting the wrong(right) product when we choose n number of products for inspection out of the total population N . If the population is 100 and we choose 10 out of those, then the probability distribution of getting the right product, denoted by X is given by

$$X \sim {}^{.85}C_x {}^{.15}C_{10-x} / {}^{100}C_{10}$$

- Remember:
 - p (0.85) and q (0.15) are the proportions of getting a good item and bad item respectively.
 - In this distribution we are considering the choosing is done without replacement

TQM/II R.N. Sengupta, IIM Dept., IIT Kanpur 64

In the high, but hyper geometric distribution, so f of x is given. So, no here I won't like to mention 1 very important things, for in all the distributions we have considered technically the; obviously, this is right the population is infinite here also in hyper geometric the population infinite, but will consider it with such very large value and it is denoted by capital N , because N would be coming into the calculations for the hyper geometric distribution.

So, this is the unique case and remember that, so f of x is given and you remember that here that capital N is coming into the formula. So, that will make sense in the whole calculation. So, parameters are capital N small n p expected value is given by n into p . So, this n into p for the other distribution binomial distribution would make sense, so in the limiting case. So, a variance is given by $n p q$ and in the bracket capital N minus small n divided by capital N minus 1. Now, again it will make sense because as capital n increases to infinity this reach this ratio of capital N minus small n by capital N minus 1 becomes 1, hence the variance also becomes equal to the binomial distribution.

So, example consider the example above now we are interested in finding in the probability distribution on the number of failures of success for getting the wrong or the right product when we choose n number of products, and inspection is being done for a total population of n consider n is very large here it is 100. So, given that we can find out the distribution x , so remember here the the p and q values we are the actually the

proportions which we can consider if you remember in the last second class. We did mentioned that corresponding to relative frequency chance it becomes actually equal to the probability in the long run as the sample size increases.

So, p and q are the proportions of getting a good item and a bad item respectively, in this distribution we are considering choosing is done without replacement if replacement is done so; obviously, the formed about change is the hyper geometric distribution how it looks like.

(Refer Slide Time: 18:25)

Solved example (Hypergeometric distribution)

Question: Suppose that automobiles arrive at a Mr. Ghosh's garage in lots of 10 and that for time and resource considerations, he can inspect only 5 out of each 10 for safety. The 5 cars are randomly chosen from the 10 on the lot. If 2 out of the 10 cars on the lot are below standards for safety, what is the probability that at most 1 out of the 5 cars to be inspected by Mr. Ghosh will be found not meeting the safety standards?

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 66

Now, suppose that the automobiles arrived at misses this is an example at misses good Mister. Ghosh's garage in lots of 10 and that for a time being the resource consideration he can inspect 5 out of this 10 for safety, the 5 cars are randomly chosen from 10 off the lot if 2 out of 10 no cars on the lot below standard for safety.

Then what is the probability that most 1 out of 5 cars is to be inspected by Mister. Ghosh will be found not meeting the safety standards.

(Refer Slide Time: 18:58)

Solved example (Hypergeometric distribution)

Answer: Here $N=10$, $n=5$, $Np=2$, $Nq=8$
Hence the required probability is
 $P(X \leq 1) = P(X=0) + P(X=1)$
 $P(X \leq 1) = {}^2C_0 {}^8C_5 / {}^{10}C_5 + {}^2C_1 {}^8C_4 / {}^{10}C_5 =$

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 67

So, here capital N 10 small n 5 n into p, now going back to the concept of p is given. So, you multiply p into capital N to get the numbers in the values of the frequencies and given q and capital N into q comes out to be 8 hence the required probability is given for when x is less than 1. So, that would be 0 and 1 and you find out the probabilities accordingly.

(Refer Slide Time: 19:26)

Poisson distribution
 $[X \sim P(\lambda)]$

$f(x) = e^{-\lambda} \lambda^x / x!$ $x = 0, 1, 2, \dots$

- λ is the parameter where $\lambda > 0$
- $E[X] = \lambda$
- $V[X] = \lambda$
- Example: Consider the arrival of the number of customers at the bank teller counter. If we are interested in finding the probability distribution of the number of customers arriving at the counter in specific intervals of time and we know that the average number of customers arriving is 5, then, we have
 $X \sim e^{-5} 5^x / x!$

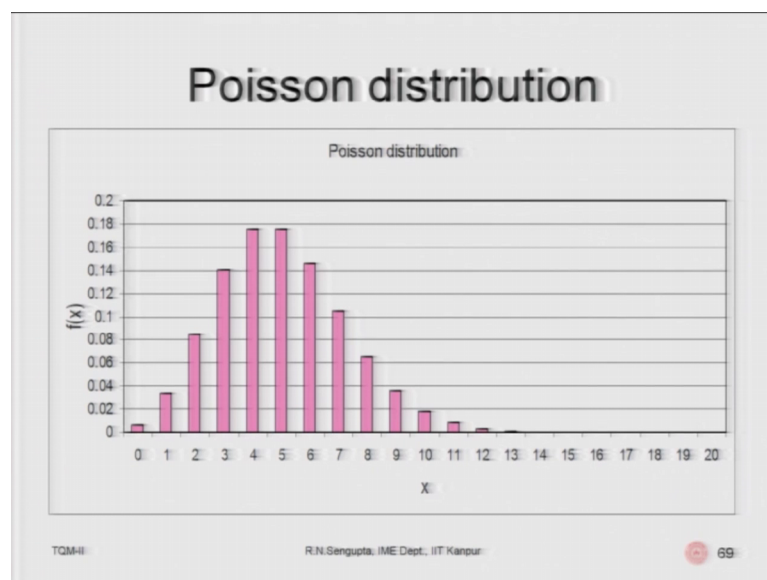
TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 68

Now, the Poisson distribution is given by the formula where lambda is the parameter expected value in variance both are lambda. So, now, Poisson distribution would become

slowly very important. So, basic it is something due to it with the discrete distribution number of people entering the queue is or basically being served for a server or. So, called people standing in front the ticketing counter and all these things can be cited.

So, we would not go into the details of the derivation or this thing as I said. So, we will come back to the example directly considering the level of number of customers at the bank teller counter if you are interested in finding the probability distribution of the number of customers arriving at the counter. In a specific interval of time we know that the average number of customers is 5 and based on that you can find out the distribution which is position distribution.

(Refer Slide Time: 20:23)



So, this is the Poisson distribution how it looks like again along y axis f of x is there that is vertical 1 and along the horizontal 1 is x.

(Refer Slide Time: 20:39)

Solved example (Poisson distribution)

Question: Mr. Gurneek Singh who is the cashier at the departmental store cash counter notices that the average number of customer arriving at the cash counter per 5 minutes is 10. Then what is the probability that more than 5 customers arrive at the cash counter with the interval of 5 minutes?

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 70

So, consider Mister. Gurneek Singh who is the cashier at the department store cash counter notices that the average number of customers arriving at the cash counter per 5 minutes per unit interval is 10. Then what is the probability that more than 5 customers arrive the cash counter with within the interval of 5 minutes?

(Refer Slide Time: 21:01)

Solved example (Poisson distribution)

Answer: Here $\lambda=10$. Hence the required probability is

$$P(X \geq 6) = 1 - \{P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)\}$$
$$P(X \geq 6) = 1 - \{\exp(-10)10^0/0! + \exp(-10)10^1/1! + \exp(-10)10^2/2! + \exp(-10)10^3/3! + \exp(-10)10^4/4! + \exp(-10)10^5/5!\}$$

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 71

So, basically we consider lambda has given within that 5 minutes which is 10 and procedure along with that, so here lambda is 10 hence the required probability is given and that can be found out very easily.

(Refer Slide Time: 21:08)

Log distribution

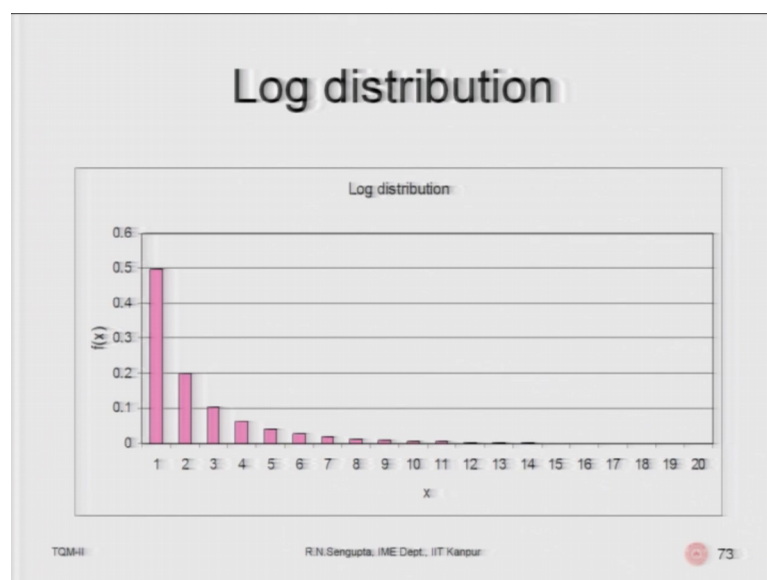
$$[X \sim L(p)]$$
$$f(x) = -(\log_e p)^{-1} x^{-1} (1-p)^x \quad x = 1, 2, 3, \dots$$

- p is the parameter where $p \in (0, 1)$
- $E[X] = -(1-p)/(\log_e p)$
- $V[X] = -(1-p)[1 + (1-p)/\log_e p]/(p^2 \log_e p)$
- Example
 - 1) Emission of gases from engines against fuel type
 - 2) Used to represent the distribution of the number of items of a product purchased by a buyer in a specified period of time

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 72

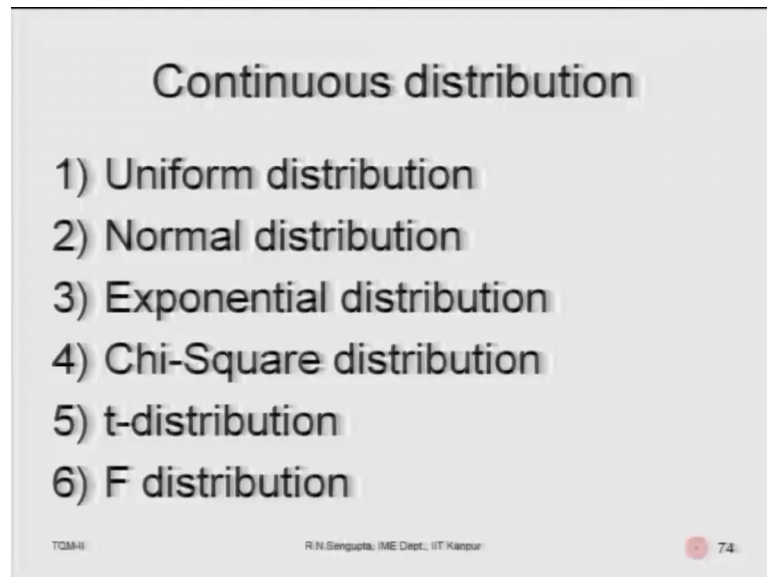
Log distribution is given by this formula this should be remembered x 's are all positive p is the parameter between 0 and 1 expected variance is given. So, we will consider examples being emission of gases from engines against fuel type used to represent the distribution on the number of items of product purchased by a buyer in a specific period of time and so on and so forth.

(Refer Slide Time: 21:37)



So, this is the log distribution putting some values and. So, it looks similarly the for looks like the exponential distribution or the Poisson distribution conceptually like shape wise no exact not the exact way, but the shape how it looks like.

(Refer Slide Time: 21:59)



Now we will consider the continuous distribution which are you which are very specific for our studies there are a lot of continuous lot of discrete distribution, but we won't consider that. So, they are the uniform distribution the normal distribution, the exponential the chi square the t and the F distribution.

So, mainly which we will consider once the normal case is over we will consider the chi, t and F distribution would be which would be very important for us.

(Refer Slide Time: 22:27)

Uniform distribution

$[X \sim U(a, b)]$

$f(x) = 1/(b - a) \quad a \leq x \leq b$

- a and b are the parameters where $a, b \in \mathbb{R}$ and $a < b$
- $E[X] = (a+b)/2$
- $V[X] = (b-a)^2/12$
- Example: Choosing any number between 1 and 10, both inclusive, from the real line

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 75

So, the uniform distribution is almost similar to the discrete case, but only the values between a and b are now continuous. So, a and b are the parameters where a and b along the real line and b is greater than a . The expected value is given as the midpoint of centre gravity which is $a + b$ by 2. Variance is given accordingly.

So, example choosing any number between 1 and 10 both inclusive from the real line; would give you some hint about the concept of uniform distribution. This is the uniform distribution.

(Refer Slide Time: 23:00)



(Refer Slide Time: 23:02)

Normal distribution

$$[X \sim N(\mu, \sigma^2)]$$
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \quad -\infty < X < \infty$$

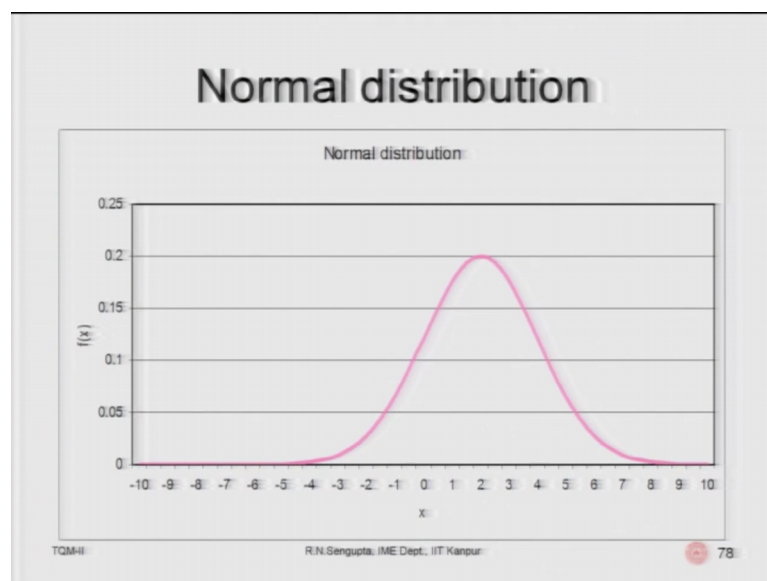
- μ_X, σ_X^2 are the parameters where $\mu_X \in \mathbb{R}$ and $\sigma_X^2 > 0$
- $E[X] = \mu_X$
- $V[X] = \sigma_X^2$
- Example: Consider the average age of a student between class VII and VIII selected at random from all the schools in the city of Kanpur

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 77

So, now, we will consider the normal distribution with the parameters μ and σ square. So, distribution f of x given as is it looks like and μ is an, is along the real line and σ square is greater than 0 expected value is given by μ and variance is given by σ square.

So, example consider the average age of a student between class 7 and eight selected at random the city of Kanpur.

(Refer Slide Time: 23:32)



So, this is the normal distribution, but; obviously, normal distribution would be very important. So, we will discuss something more about the normal distribution as we go.

(Refer Slide Time: 23:43)

Exponential distribution
 $[X \sim E(a, \theta)]$

$$f(x) = \frac{1}{\theta} e^{-\frac{(x-a)}{\theta}} \quad a < x < \infty$$

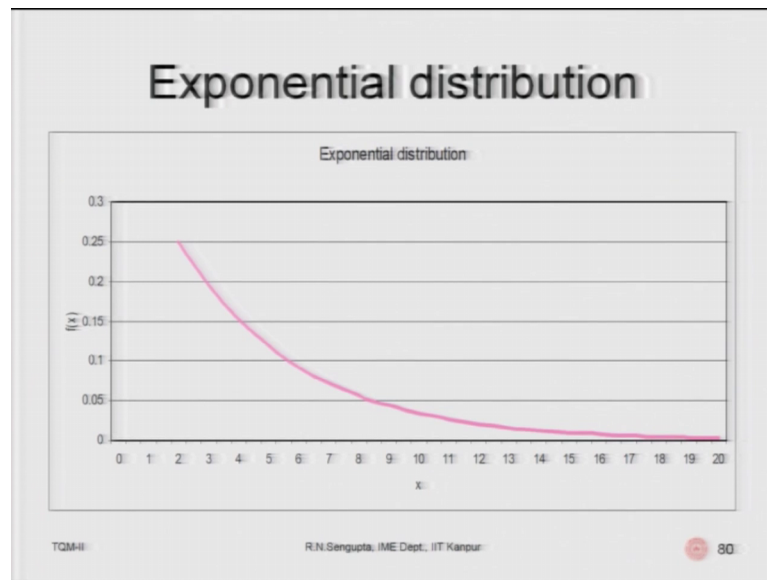
- a and θ are the parameters where $a \in \mathbb{R}$ and $\theta > 0$
- $E[X] = a + \theta$
- $V[X] = \theta^2$
- Example: The life distribution of the number of hours a electric bulb survives.

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 79

Now, consider the exponential distribution which if you see the shape of that curve, would almost look exactly like the Poisson distribution just for your interest and; obviously, there is some relationship between the Poisson and the exponential we will come to that later.

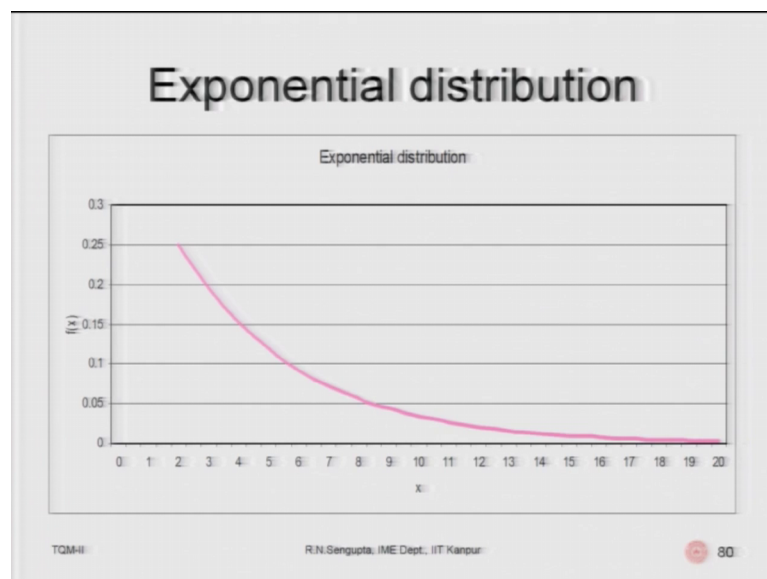
So, a and θ the parameters where a is along the real line θ is greater than 0 expected value is given by $a + \theta$ variance is given by θ^2 . So, it is basically to do with the light distribution or number of us and of an electric bulb till it survives.

(Refer Slide Time: 24:14)



So, with the exponential distribution falls exponentially and only one thing that value in case if say for example, the starting value is 0. So, the whole distribution will shift on to the left.

(Refer Slide Time: 24:29)



So, I will come to that, but just for interest.

(Refer Slide Time: 24:32)

Solved example (Exponential distribution)

Question: Mr. K. Bharadwaj an electrician knows that the time a bulb operates before it gets fused is exponentially distributed with a mean life of 100 hours. What is the probability that any bulb will work continuously for at least 200 hours?

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 81

Mister. Bharadwaj an electrician knows that the time a bulb operates before it gets fused is exponential distribution with the mean life of 100 hours. So, what is the probability that any bulb will stop will work continuously for at least 200 hours?

(Refer Slide Time: 24:44)

Solved example (Exponential distribution)

Answer: Here $\theta = 100$. hence the required probability is

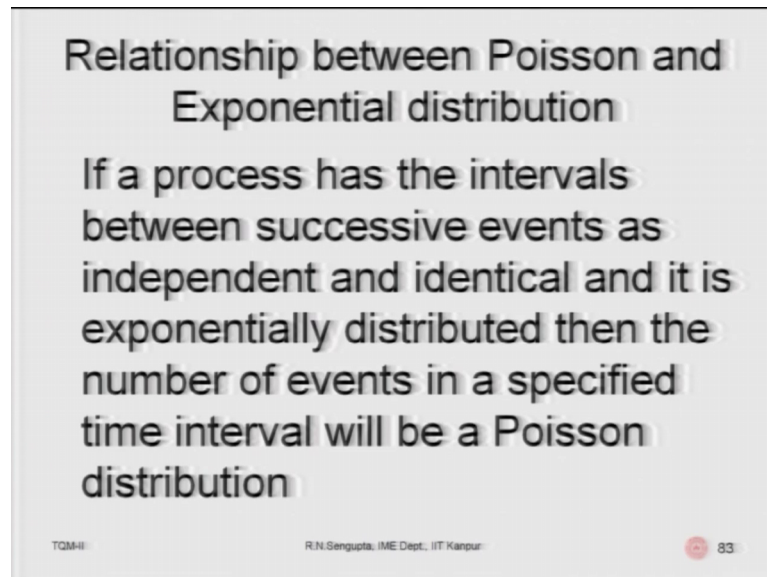
$$P(X \geq 200) = 1 - P(X \leq 200)$$
$$P(X \geq 200) = 1 - \int_{x=0}^{200} \frac{1}{100} \exp\left(-\frac{x}{100}\right) dx$$

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 82

So, given theta is 100 we and we replace that in the formula, but only remember that finding of the c d f. Now, rather than summing it up we have to basically if we integrate it because it is continuous on that you will find out the value. So, greater than 200 would also mean the complementary part being on the left hand side which is less than to 100,

so total probability is 1 1 minus any of them would give you the other. Now, had I did mention 3 slides before there is a relationship between Poisson and exponential this is the case and this is very important conceptually.

(Refer Slide Time: 25:25)



Obviously theoretically it is very interesting to note that, but we would not go into theory, if a process has the intervals between successive events as independent and identical and it is exponentially distributed, the number of events in a specified time interval will be given by Poisson distribution.

So, there are intervals between them the arrivals are given then; obviously, the processing time would be another distribution, so a discrete being Poisson continuous being exponential.

(Refer Slide Time: 25:53)

Cumulative distribution function (cdf) or the distribution function

We denote the distribution function by $F(x)$

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$
$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx = \int_{-\infty}^x dF(x)$$

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 84

So, cumulative distribution; we know that is the sum of all the values starting from 0 or minus infinity to that x and for the discrete case it is summation for the continuous case it is basically in the integration.

(Refer Slide Time: 26:08)

Properties of distribution function

- 1) $F(x)$ is non-decreasing in x , i.e., if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$
- 2) $\lim_{x \rightarrow -\infty} F(x) = 0$
- 3) $\lim_{x \rightarrow +\infty} F(x) = 1$
- 4) $F(x)$ is right continuous

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 85

The properties of the distribution function c d f values it is a non-decreasing so; obviously, if x_1 is less than x_2 or less than equal to x_2 . So, capital F of x_1 would always be less than equal to x_2 because for any incremental value from x_1 to x_2 the value cannot be negative it can be either 0 or positive because this probability. So, if x

tends to minus infinity capital F of x which is the c d f a is 0 if x tends to positive infinity can the f of x of x is 1 and f of x is right continuous.

(Refer Slide Time: 26:44)

Standard normal distribution

Putting $Z=(X-\mu_X)/\sigma_X$ in the normal distribution we have the standard normal distribution

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

Where: $\mu_Z = 0$ and $\sigma_Z = 1$

Remember

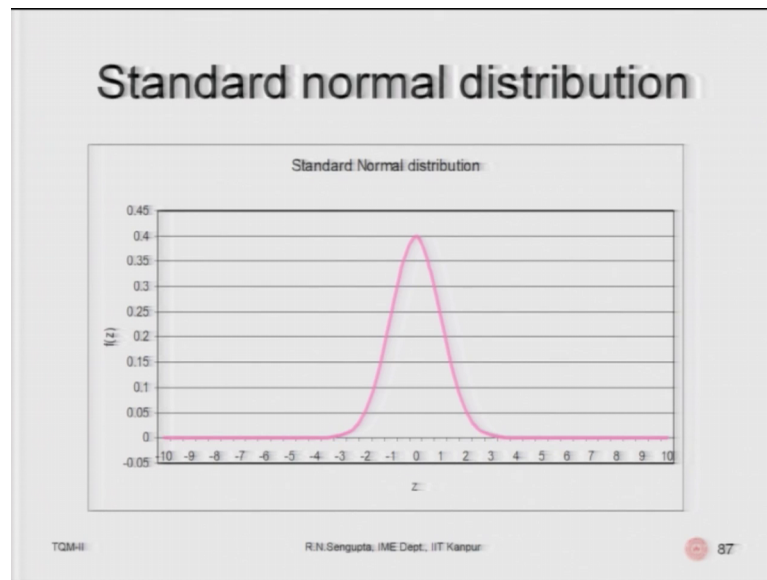
- $F(x) = P(X \leq x) = F(z) = F(Z \leq z)$
- $f(z) = \phi(z)$
- $F(z) = P(Z \leq z) = \int_{-\infty}^z f(z) dz = \int_{-\infty}^z dF(x) = \Phi(z)$

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 86

So, if you put values of the do a simple transformation put the value of z as x minus mu by sigma square we get the standard normal and the standard normal tables are there we will use that accordingly.

So, remember now capital F of x is equal to capital F of z because the transformation we gives you the values small f of x will be given by small f phi of z, which is the actually technically the p d f values and similarly the value of capital F of z is also given by capital phi of z, so these are very simple to understand.

(Refer Slide Time: 27:23)



This is the standard normal distribution with the mean value of 0. So, the variance is 1 and for them, tables are available.

(Refer Slide Time: 23:33)

Normal distribution results

- $\phi(-z) = \phi(z)$, i.e., $f(-z) = f(z)$
- $\Phi(-z) = 1 - \Phi(z)$, i.e., $P(Z \leq z) = 1 - P(Z \geq z)$
- $P(a \leq X \leq b) = \Phi[(b - \mu_X)/\sigma_X] - \Phi[(a - \mu_X)/\sigma_X]$
- $P(X \leq b) = \Phi[(b - \mu_X)/\sigma_X]$
- $P(a \leq X) = 1 - \Phi[(a - \mu_X)/\sigma_X]$

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 88

Now, as it is symmetric is very simple to understand small of small phi of minus z will be equal to small phi of plus z, because its symmetric yeah; capital of that minus Z would basically be the complementary part of the of say for example, capital Z and probabilities of a being less than x and x being less than b; you can find out using the cumulative distribution function and when the left limit tends to minus infinity, and the

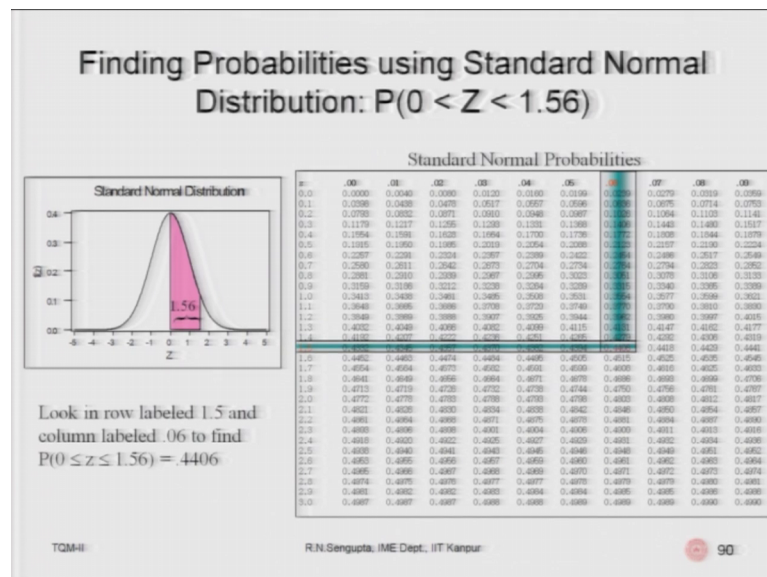
right limit tends to the positive infinity in that two cases we can solve the problems according.

(Refer Slide Time: 28:16)



So, this is the normal distribution these are the a and b values just for example. So, we can find out the c d f values capital phi and all those values accordingly here, so we will come to that later.

(Refer Slide Time: 28:22)

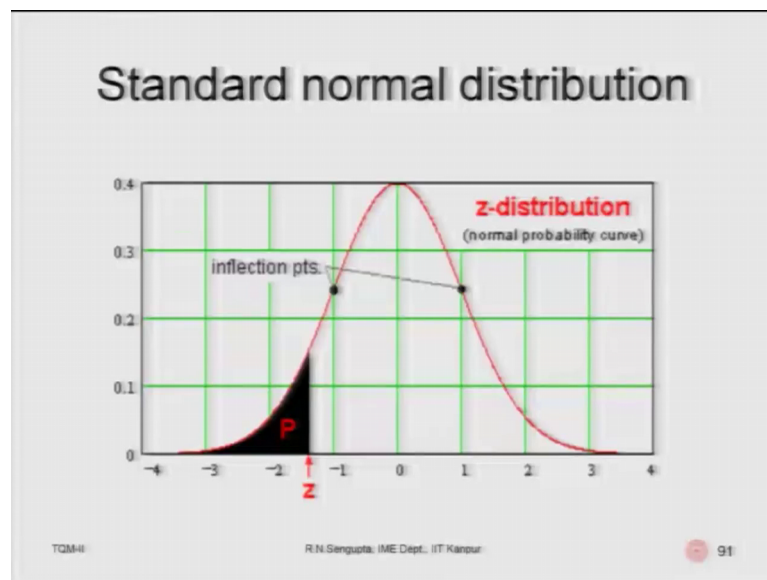


Now, this how we will come to that later on we have the charts here, so the standard normal distribution is given. So, if you basically find out the first column z values are

given from 0 to 3 it can continue to infinity, but; obviously, the values would, would be very, very small and on the top most row the values of z point something of values are given from 0 to 0.09, and the values which are inside this sheet all these are the.

So, called cumulative distribution functions c d f value starting the integration from 0 to that being that that particular value if you see the pink curve so; obviously, there would be complementary part on the left hand side it can use that accordingly.

(Refer Slide Time: 29:08)



So, only one thing, so this is z z distribution which is therefore, various standard deviation is 1. So, at plus 1 and minus 1 these are the inflection points they can be found out very easily any cancel the problems for many of the cases.

(Refer Slide Time: 29:26)

Solved example (Normal distribution)

In an examination 20% of the students failed (i.e., obtained a score which is less than or equal to 40 marks out of 100) and 10% of the students obtained a grade A (score of 70 marks or above out of 100). Assuming normal distribution of marks find the mean and the standard deviation of the distribution

TQM/HI R.N.Singupta, IIM Dept., IIT Kanpur 92

So, consider this in an example 20 percent of the students failed that is obtained a score.

(Refer Slide Time: 29:50)

Solved example (Normal distribution)

Steps

- 1) $P(X \leq 40) = 0.2 = P[(X - \mu_X)/\sigma_X \leq (40 - \mu_X)/\sigma_X] = P(Z \leq z_1) = \Phi(z_1) = -0.84$
- 2) $P(X \geq 70) = 0.1 = P[(X - \mu_X)/\sigma_X \geq (70 - \mu_X)/\sigma_X] = P(Z \geq z_2) = 1 - P(Z \leq z_2) = 1 - \Phi(z_2)$. Hence $\Phi(z_2) = 0.9$

Hence we have from the above two equations:

- $z_1 = (40 - \mu_X)/\sigma_X = -0.84$
- $z_2 = (70 - \mu_X)/\sigma_X = +0.90$
- $\mu_X = 54.12; \sigma_X = 17.64$

TQM/HI R.N.Singupta, IIM Dept., IIT Kanpur 93

Which is less than equal to forty 10 percent of students obtain a grade of seventy marks and above assuming normal distribution of marks find the mean and the standard deviation on the distribution and you can solve the problems accordingly.

So, here steps which will do solve the problem probability of x being less than point forty marks is equal to for the corresponding value of z would be given, and say similarly when you want to find out the value of probability of x greater than 70, you can find out

the other z putting those values to there are two equations two unknowns mean and standard deviation you can find them and the value comes out to be 50 4.12, and 17.64.

(Refer Slide Time: 30:15)

Solved example (Normal distribution)

Question: In Prof. Ram Pal's mathematics examination 20% of the students failed (i.e., obtained a score which is less than or equal to 40 marks out of 100) and 10% of the students obtained a grade A (score of 70 marks of above out of 100). Assuming normal distribution of marks find the mean and the standard deviation of the distribution of marks in mathematics?

TQM-II R.N.Sengupta, IME Dept., IIT Kanpur 94

So, Professor Ram Pal's mathematic is summation 20 percent the students failed and 10 percent the same example a students continued got a score of 70 and based on that we calculate the values as given. So, this is the standard normal distribution.

(Refer Slide Time: 30:36)

Solved example (Normal distribution)

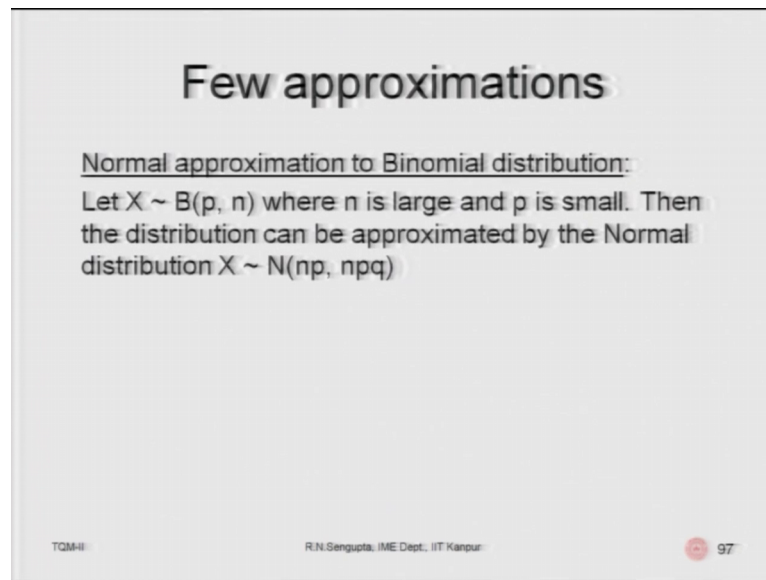
Finding Probabilities using Standard Normal

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2089 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4266 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |
| 2.5 | 0.4938 | 0.4940 | 0.4941 | 0.4943 | 0.4945 | 0.4946 | 0.4948 | 0.4949 | 0.4951 | 0.4952 |
| 2.6 | 0.4953 | 0.4955 | 0.4956 | 0.4957 | 0.4959 | 0.4960 | 0.4961 | 0.4962 | 0.4963 | 0.4964 |
| 2.7 | 0.4965 | 0.4966 | 0.4967 | 0.4968 | 0.4969 | 0.4970 | 0.4971 | 0.4972 | 0.4973 | 0.4974 |
| 2.8 | 0.4974 | 0.4975 | 0.4976 | 0.4977 | 0.4977 | 0.4978 | 0.4979 | 0.4979 | 0.4980 | 0.4981 |
| 2.9 | 0.4981 | 0.4982 | 0.4982 | 0.4983 | 0.4984 | 0.4984 | 0.4985 | 0.4985 | 0.4986 | 0.4986 |
| 3.0 | 0.4987 | 0.4987 | 0.4987 | 0.4988 | 0.4988 | 0.4989 | 0.4989 | 0.4989 | 0.4990 | 0.4990 |

TQM-II R.N.Sengupta, IME Dept., IIT Kanpur 96

And we solve the problems accordingly.

(Refer Slide Time: 30:39)



Few approximations

Normal approximation to Binomial distribution:
Let $X \sim B(p, n)$ where n is large and p is small. Then the distribution can be approximated by the Normal distribution $X \sim N(np, npq)$

TQM-II R.N.Sengupta, IME Dept., IIT Kanpur 97

So, now few approximations are there, so the normal approximation which would be important for the binomial distribution would be let x be equal to binomial p n n p n n and the parameters. So, where n is large and p is small then the distribution can be approximately with the normal with the mean equal to the distribution in the binomial distribution which $n p$ and the variance equal to $n p q$.

So, we can solve according to. So, if it is say for example, Poisson distribution it can be converted to the using this concept were normally distribution if it is say for example, the exponential distribution it can be converted to a normal distribution, but remember the mean value and the variance of the corresponding normal distribution will be coming from where the original distribution is so; obviously, they will be approximation, but it gives us good solution.

(Refer Slide Time: 31:33)

De Moivre Laplace Limit Theorems

- 1) $P(a \leq X \leq b) \cong \Phi\left[\frac{b - np}{\sqrt{npq}}\right] - \Phi\left[\frac{a - np}{\sqrt{npq}}\right]$
- 2) $P(a \leq X) \cong 1 - \Phi\left[\frac{a - np}{\sqrt{npq}}\right]$
- 3) $P(X \leq b) \cong \Phi\left[\frac{b - np}{\sqrt{npq}}\right]$

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 98

De Moivre's Laplace Limit Theorems, so these are very easy to understand in case say for example, you want to consider the binomial distribution main mean value in place of say for example, μ becomes np and variance in place of σ^2 or square root of σ^2 visits and division it becomes square root of npq and we can solve the problems accordingly.

(Refer Slide Time: 31:58)

Markov Inequality

Let Z be a non-negative r.v such that $E[Z]$ exists. Then for every positive t we have

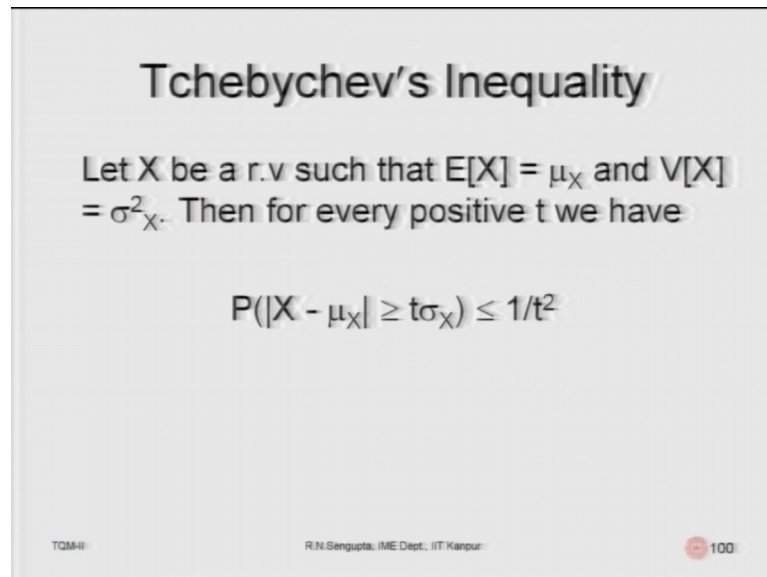
$$P(Z \geq t) \leq E[Z]/t$$

TQM-II R.N. Sengupta, IIM Dept., IIT Kanpur 99

So, these would not be actually used, but they may be interesting to discuss later on when we consider TQM 2. So, in Markov Inequality; let Z be a non-negative random

variable such that expected value exists then for every positive t will have probability of Z greater and then t is given by expected value of Z divided by t , so t is basically a positive integer.

(Refer Slide Time: 32:24)



Tchebychev's Inequality

Let X be a r.v such that $E[X] = \mu_X$ and $V[X] = \sigma_X^2$. Then for every positive t we have

$$P(|X - \mu_X| \geq t\sigma_X) \leq 1/t^2$$

TQM-II R.N.Sengupta, IIM Dept., IIT Kanpur 100

In the case it is a Tchebychev's inequality, so it does not mention anything about the negative or positive values of the random variable, but if its first moment and the second moment exists; then the mod of the distance between x and its mean value within a certain bound is if given by even the value of t then there is a relationship that when the mod would be the bond would be less than equal to or greater than equal to t sigma is given my inverse of 1 by t square.

So, higher or lower the value of t will basically you mean the answer.

(Refer Slide Time: 31:00)

Bernoulli's Theorem

Let X_n be the number of success in ' n ' number of Bernoulli trials, each with success probability ' p '. Then for arbitrary positive ε we have

$$\lim_{n \rightarrow \infty} P[|X_n/n - p| \leq \varepsilon] = 1$$

TQM/H R.N.Sengupta, IIM Dept., IIT Kanpur 101

So, let X be the number of successes in, so let X be the successes of n number of trials in a Bernoulli trials. So, each with the success p then for arbitrary positive values of ε ; we have limiting in case it basically its X is X/n by n minus p being greater than less than equal to ε is given by 1 as n tends to infinity have a good day and.

Thank you very much.