Introductory Mathematical Methods for Biologists

Prof. Ranjith Padinhateeri
Department of Biosciences & Bioengineering
Indian Institute of Technology, Bombay

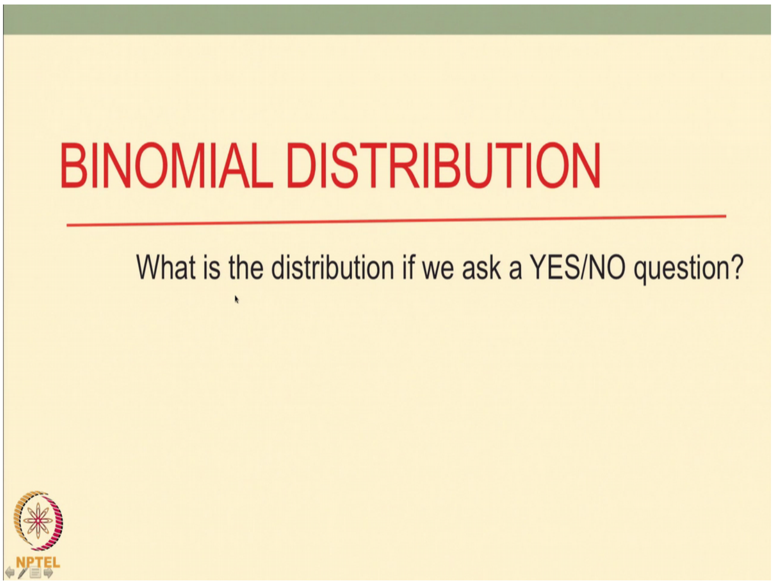
Lecture – 38
Binomial Distribution

Hi, welcome to this lecture on Mathematical Methods for Biologist. We have been discussing statistics and we discuss probability distribution. Today we will discuss some examples of probability distribution. The aim of mathematics is to quantify things as we discussed and when we want to quantify things around us the experiments that we do there are many ways one can do and then very often, you have some data or you have planning to do some experiment, you could ask the experiment in the following way where you could ask a yes or no question for example, if you are doing a survey you could of course, ask a question which has an answer yes or no survey is also an experiment.

But even if you do experiments with cells for example, you add some cells at a particular concentration then you could ask the following question is the cell is alive or dead right is it viable. So, there is a dead or not dead alive. So, this is yes or no question there is only 2 answers either dead or not dead right, there is this is one way of asking a question in experiment and the answer is yes or no is the cell dead the answer is yes or no. This is the question that we can ask in an experiment, in other this could be in many other case then somewhere if you want to study some protein binding you could ask the question whether the protein is bound or not bound the answer is yes or no.

So, there are many examples many situations where you would want to design your experiment such a way that the answer is yes or no if you do that our statistics gives us some help. So, if we have a situation where the answer you ask a question and the answer is yes or no using the ideas from statistics we can analyze, it in a very interesting manner and that is of this lecture how to do that is what we will discuss in this lecture.

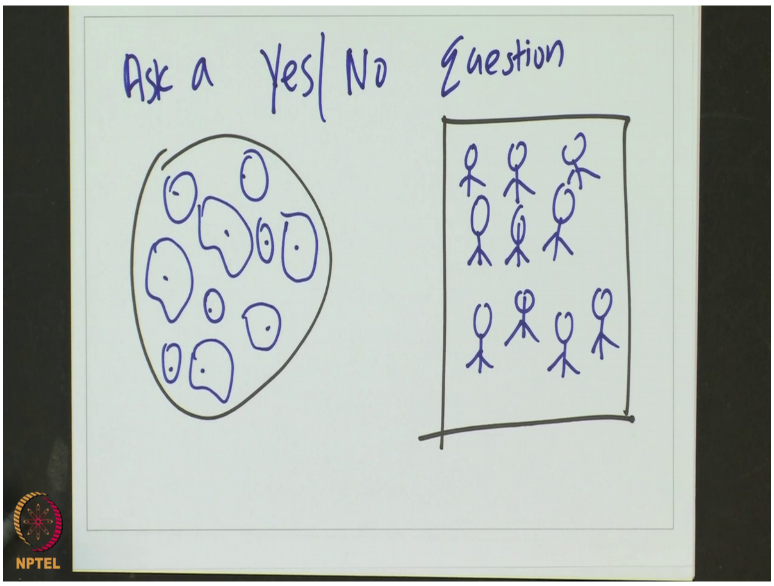
(Refer Slide Time: 02:31)



So, the topic of this lecture is binomial distribution this is a name of a particular distribution and this the question will answer is what is the distribution if we ask a yes or no question.

So, if we ask a yes or no question the resulting distribution is going to be a binomial distribution and that is what we would discuss today to do that let us think of some situation. So, you have a population again. So, we have either a population of cells.

(Refer Slide Time: 03:04)



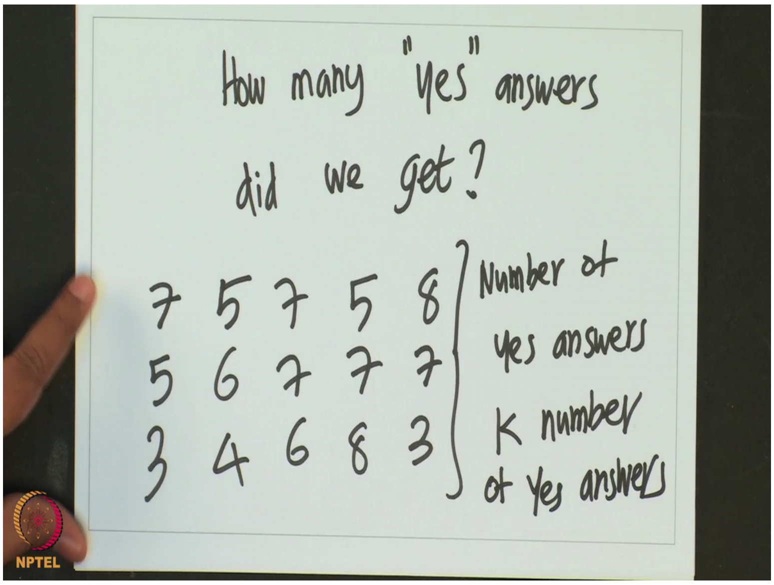
So, you have you have population of cells, which you would see under microscope let us say there are 10 cells. So, 1 2 3 4 5 6 7 8 9 10 there are 10 cells or you would go to a sample or you have some number of individuals.

So, let us say you have a room or you have 10 individuals. So, let us say you have 10 individuals 3 5 6 7 8 9. So, there are 10 individuals whom you would ask this question and this could be this question could be like for example, if I have a drug and you give this drug and this 10 individuals have a disease is it cured or not after taking this much dose of the dose of the drug is the cure is a disease cured or not and the answer will be yes or no. After adding this much drug is the cell, alive or not answer is yes or no by making a particular mutation is a particular function happening or not.

So, whenever you have a you ask a yes question for which the answer is yes or no. So, ask a yes or no question you ask a question for which the answer is yes or no, now out of this 10 people when you do experiment once maybe 7 of them for 7 individuals or 7 cells the answer is yes maybe some other time if you repeat this experiment maybe 5 cells were alive. So, the answer is yes for 5 cells if the cells alive if the question 5 of them would say 5 of the cells would be alive.

So, you would get for example, at the end of it. So, how many yes answers we got.

(Refer Slide Time: 05:32)

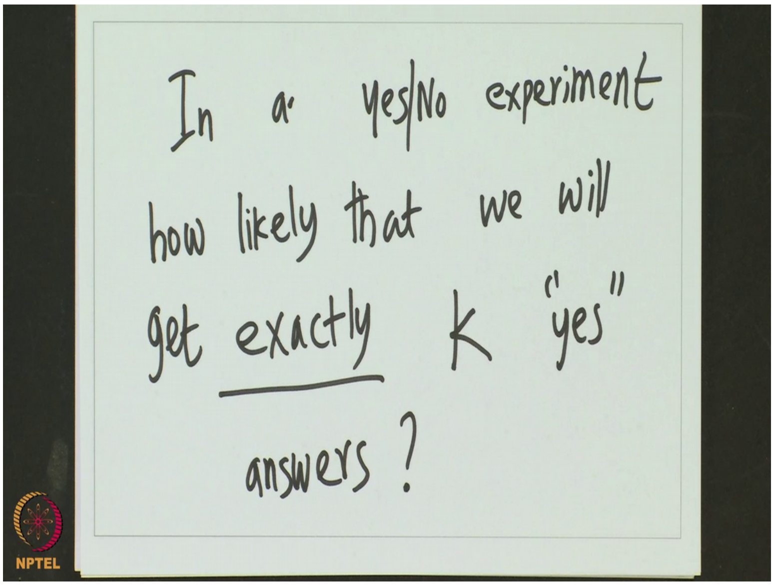


So, this is after doing this experiment many times we would tabulate how many "yes" answers did we get this is a how many yes and answers did we get? So, when we first time when you do experiment out of 10 7 out of 10 we got yes answers, next time maybe 5 another time 3 you repeat a experiment you got 5 you again repeat a experiment 6 4 7 7 6 5 7 8 8 7 3 and so on and so forth you could get many values.

So, let us these values. So, this is how many yes answer. So, this is number of yes answers. So, we have 7 number of yes answers, 5 number of yes answers, 3 number of yes answers. So, K number of yes answers where K is any of this number K number of yes answers now how many yes answers will we get. So, the question is if you take a population and add a particular drug how many cells will be alive right.

So, how likely that you would get exactly 3 cells alive how likely that you would get exactly 7 cells alive right. So, if you do an yes or no experiments how likely that you would get exactly K yes answers this is if you ask this question.

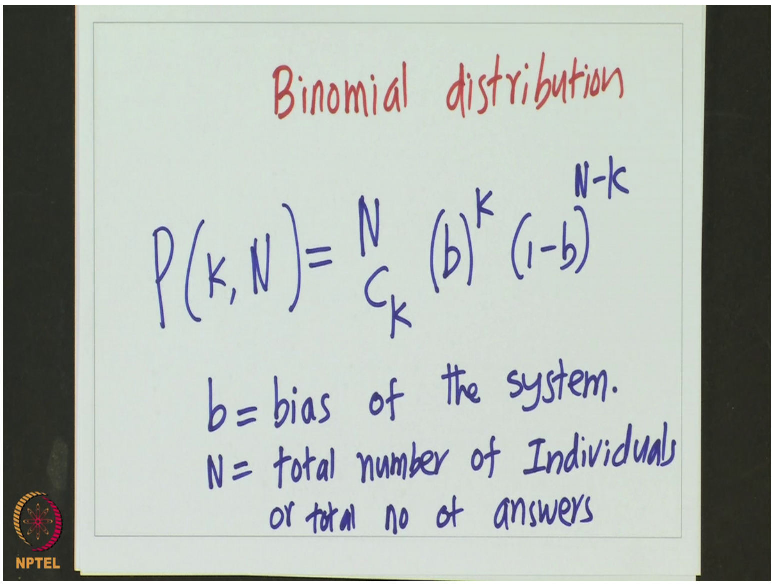
(Refer Slide Time: 07:32)



So, in a yes or no experiment in a yes or no experiment where the question is framed such a way that the answer is yes or no, how likely that how likely that we will get exactly this is the keyword and exactly K "yes" answers in a yes or no experiment how likely that we will get exactly K "yes" answers.

Now, the this question can be answered mathematically and this is called a binomial distribution.

(Refer Slide Time: 08:37)

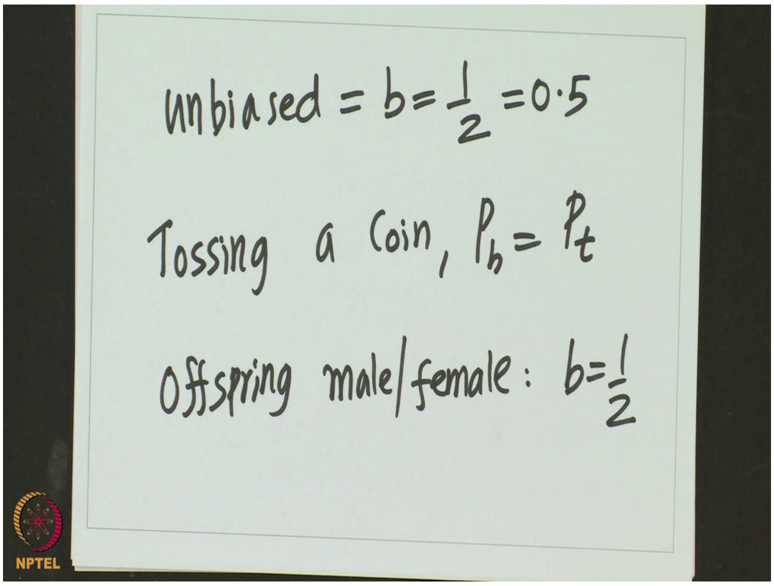


So, the Binomial distribution answers this question. So, this binomial distribution so, the question is how likely that you will get exactly K “yes” answers. So, the probability of getting K “yes” answers that is what the probability of getting K “yes” answers given there are N individuals answering this or other N cells. So, N is the total number of cells we have this is given by NCK, NCK is combination N choose K and I will tell you what this is some of you have heard this in school NCK N choose K there is a parameter which is called a b and I will tell you what b is b power K 1 minus b power N minus K b is the bias of the system and what is b, I will explain what b is N is total number of individuals or cells individuals or total number of answers.

So, there are total number of answers or total number of answers there would be some yes answers and no answers yes plus no total number is N. So, if there are total number of answers how likely that you will get K of them, yes is given by this or b is something which is called a bias of the system now what is bias. So, if things are completely unbiased there is equal likelihood of getting yes or no answer. For example, if you toss a coin there is an equal chance of getting head or tail. So, there the bias is half like if in a if in a in a unbiased in a typical situation if your offspring’s male or female there is a equal probability of offspring being male or female. So, the probability is half. So, the bias is half.

So, b is the bias typically unbiased would mean b is equal to half.

(Refer Slide Time: 11:20)



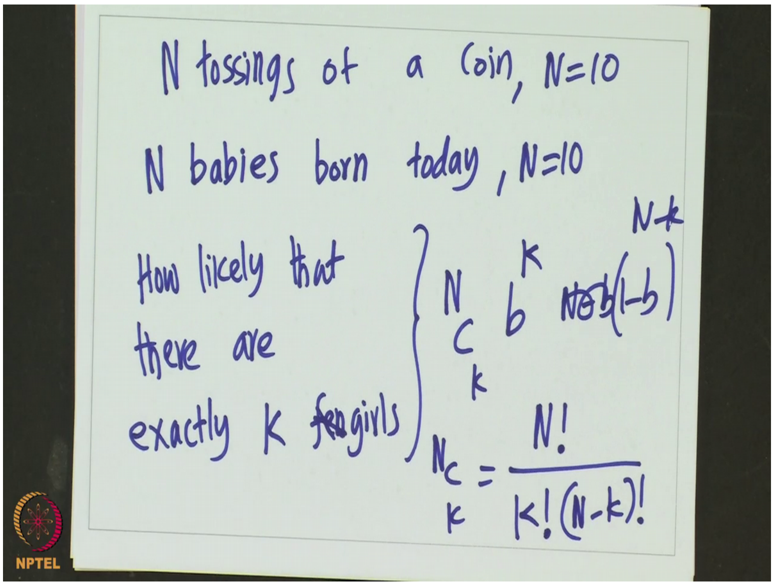
So, unbiased situation b is equal to half examples are tossing a coin probability of head probability of tail a probability of head is equal to probability of tail therefore, this is completely unbiased therefore, b is equal to half which is 0.5 a same offspring the likelihood of offspring being a male of or offspring is like male or female this is also biases unbias. So, probability of either of this is half.

So, in such situation the probability would be half, but many a situation if you do not have a fair coin, if you have an unfair coin you might get more heads than tail if you have an asymmetric, if you do not have a nice if you take some object which is not symmetric you would fall 1 side more than the other in that case the b is not half it is biased towards 1 side. So, the system that you would do experiments you may not know which way it is biased if the drug that you apply is very efficient then it will be bias towards curing it. So, then the answer will bias towards.

So, by doing this we can also know the bias of the system how much it is bias this way. So, if we know the probability we can calculate the bias or if we know the bias we can calculate the probability of this. So, let us think of this question in the context of unbias thing that we know let us say if you have 10 times if you toss a coin, how many times what likelihood that you how likely that you will get exactly 3 head like, if many couples you take and if you take 10 couples and you ask the question how likely that. So, each of this 10 couples have 1 exactly 1 offspring 1 kid 1 baby and this baby could be a male or a female and out of this N how many babies are female right what is the probability of getting exactly K female.

So, this is are the questions that let us first answer. So, I would write here you have N tossings of a coin.

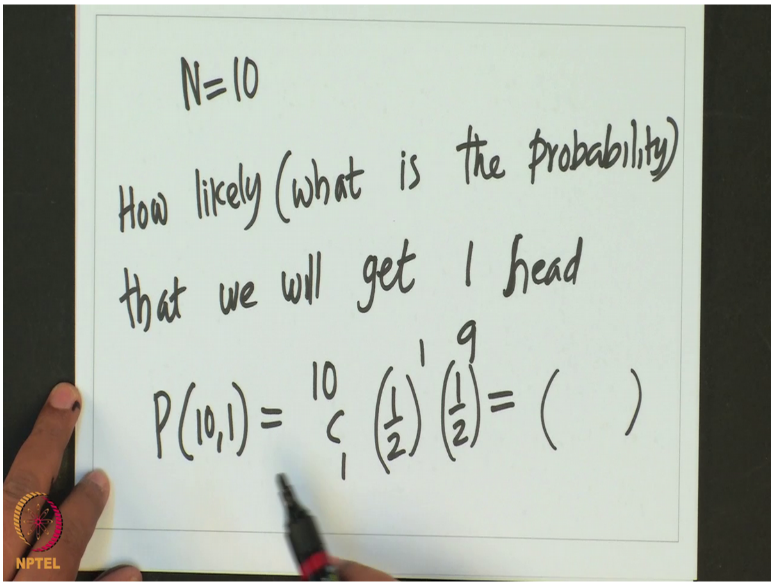
(Refer Slide Time: 14:11)



So, here there is N is 10 let us say 10 tossing or N babies if you take right for to N parents. So, there N couples and there are N babies born today and you ask the question how many of them are. So, here also let us say N is equal to let us take 10 how likely. So, the question we would answer here is how likely what is the probability how likely, how likely that there are exactly K female babies K girls or K heads in the case of a coin and this is given by this formula N choose K b power K where b is half here N minus b sorry 1 minus b power N minus K this is the formula, where NCK is N choose K which can be computed as N factorial divided by K factorial times N minus K factorial. So, this is the if you this is NCK is a number and this can be computed like this N factorial divided by K factorial times N minus K factorial.

But it is very important to understand what is the question we are asking we are asking how likely that there will be exactly this many male or this many head right. So, let us take some simple examples right. So, let us ask let us take the case of N is equal to 10.

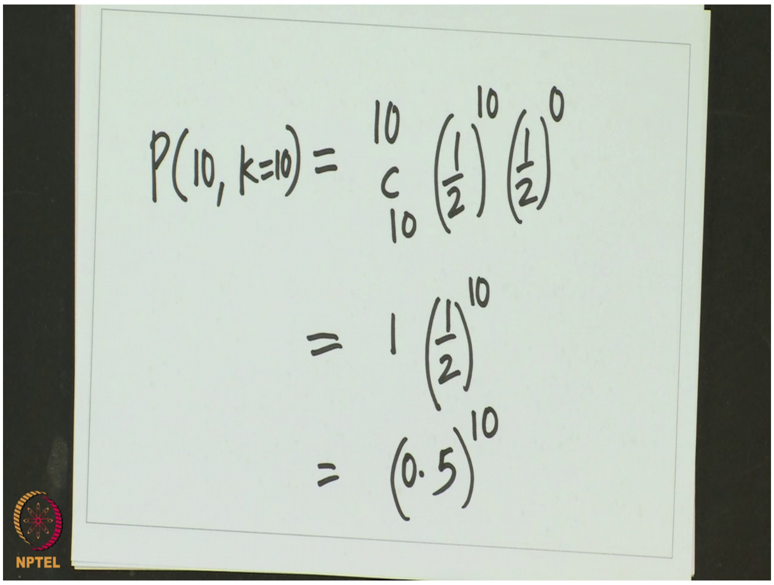
(Refer Slide Time: 16:23)



There are 10 individuals or 10 10 couples or 10 tossing’s and we would ask the question; how likely which is same as what is the probability what is the probability, how likely that we will get 1 head or 1 1 head or it could be 1 girl. So, exactly 1 and 9 is boys or in the other case 9 is tail or so this would of course, be P 10, 1 and the bias is half here.

So, 10 C 1 half power 1 1 minus half is also half half power N minus K is 9 9 plus 1 this should be this is the answer that we can get. So, 10 C 1 is 10 factorial divided by 1 factorial divided by 10 minus 1 which is 9 factorial. So, we can calculate this approximately and this number would give us the probability to have exactly 1 head or 1 girl though opposite we could ask how likely that all the 10 girls born all the 10 babies born in a particular day are girls right.

(Refer Slide Time: 18:15)

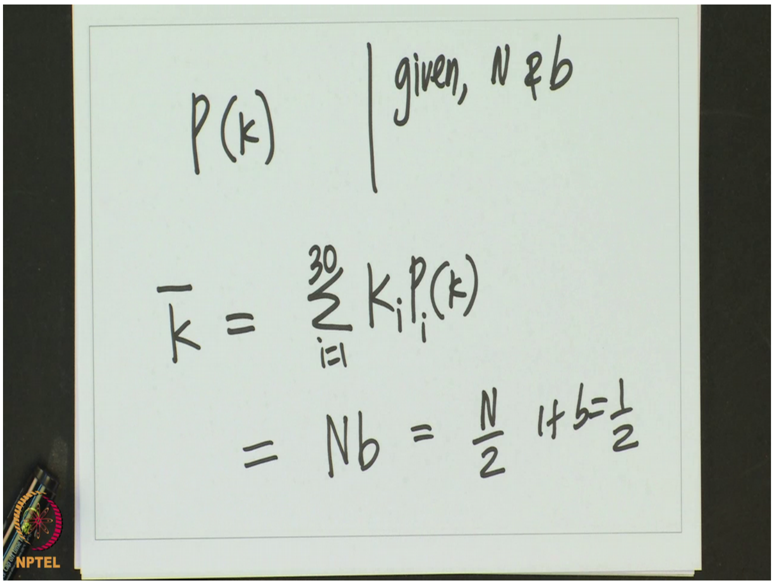


Probability that out of 10 births on a particular day all the 10 there are exactly 10 this of course, can be writ10 10 c 10 half power 10 when half power 0. So, this is basically 10 C 10 is 1 and half power 10 is this. So, this is the answer.

So, this will be half 0.5 power 10 would be the answer that we know. So, this would give us answers to this question that for example, this gives the probability that all 10 are girls the probability is 0.5 power 10, you can ask how many of them how what is probability that 1 of them is a girl what is probability that 2 of them would be a girl exactly 2 of them what is the probability that there will be exactly 3 girls out of this 10 births happened on a particular day.

So, there are many such examples where we know things are unbiased there is no bias then the b is half very often we would know what b is. So, we would be able to back calculate b if you if we know the P value. So, this is the use of binomial distribution to answer this question now we would want to know what is the mean number for example, if. So, we have what we have we have P of K for a given N given N and b, so, given N and b we have P of k.

(Refer Slide Time: 20:08)



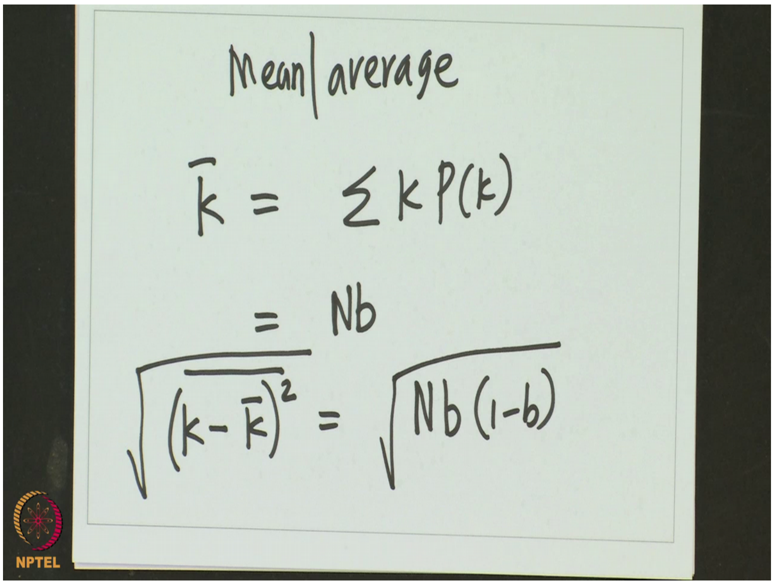
Now, what is the mean value of K. So, you do this experiment let us say you do 10 tossing's many many times and you would ask how many what is the average number of heads I got equivalently, I could go to different hospitals for example, and look at 10 a group of 10 births and you ask out of this 10 births in different hospital what is average number of girls born on a particular day.

So, this is basically can be written as K bar and this is nothing, but sum over K P of K. So, if I just go to 100 or let us say I go to 30 hospitals or 30 times I repeat. So, this is like this would go from 1 to 30. So, this is if I want to write K i I could write here i is equal to 1 to or if I go to 30 hospitals this K i P K i.

So, this i could write this way and this would be the mean value and it turns out that this is N times b. So, the number of girls if the bias is half this would be exactly N by 2 if b is equal to half. So, that we know that if you have hundred coins or if you have 100 if you have 10 tossings and you go to many many places and do 10 tossings or you take and then your answer will be N by 2 like there is a probability average value will be half 5 and N by 2 is 5 0 10 tossing the average a head number of heads will be N by 2 which is 5.

So, this is if the bias is half is N by 2 if the bias we do not know the mean value will be N times b. So, this is something that we have to remember the mean of a binomial distribution.

(Refer Slide Time: 22:43)

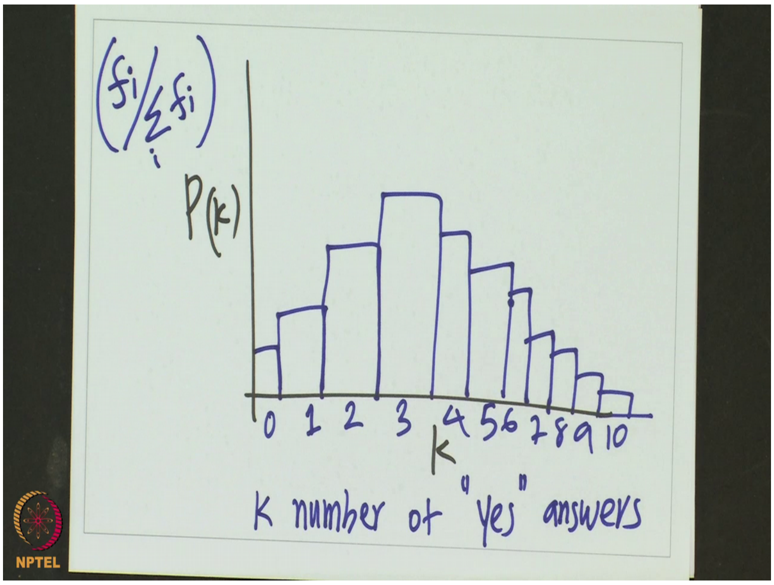


So, the mean or average of a binomial distribution would be K bar would be equal to as we wrote here sum over of K P of K and this would be N times b.

Now, you could also ask what is K minus K bar whole square average root of this right you could ask what is K minus K bar whole square whole square average root of this right. So, this is going to be N b times 1 minus b and of course, this is very root of this. So, the mean and standard deviation can be obtained like this and now given this we would also think about plotting this distribution.

So, let us think of plotting P of K versus K.

(Refer Slide Time: 24:04)

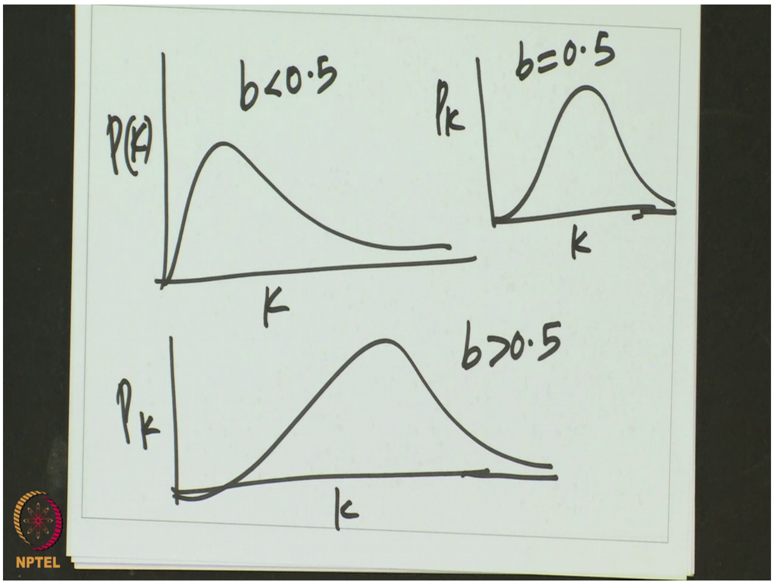


So, I just quickly wanted to plot P of K what does this mean this means if I do this experiment many times how many times I got no head at all or how many times I got no answer right or here is a how many times I got 0 yes answer. So, may be many often you approach 10 people you would get no no from always.

So, how many times you got 0 so, this is what this is how many times you got 1 yes answer, how many times you got 2 yes answers, how many times you got 3 yes answers, 4 yes answers, 5 6 7 8 9 10. So, this is 0 1 2 3 4 5 6 7 8 9 10. So, these are different number of yes answers. So, this is K is a K number of yes answers K number of yes answers and this P of K is frequency.

So, fi divided by sum over I fi this is that is what is P of K is. So, the frequency how many times I got divided by sum over f normalized frequency there is a probability as we said. So, this kind of plot we can plot and you would get a plot like this how many times we got K yes answers in from a population if you repeat this many times. So, that gives us P of K and the plot of P of K would depend would depend on how biased it is.

(Refer Slide Time: 26:15)



If it is you could think of different ways and you could think of a P of K like this or you could think of a P of K like this or you could think of a P of K like this.

So, depending on the bias if b is very small if b is much smaller than 1 is much is if smaller than half, you would think of this way if b is larger than half, you would think of some answer and if b is close to half you would get a distribution like this. So, P of K versus K this is P of K versus K P of K versus K if I plot it I would get different 1 can get different kind of distributions depending on the value of b that we have.

So, to summarize we studied binomial distribution and this distribution helps us and tells us how many yes answers you would get, if you ask a question which has a yes or no answers or the same thing could be applied to no answer then the b would be the bias for the no answer. So, you could equivalently ask how many no answers you would get is just that the b the meaning of b would be the bias towards no the b is the bias towards yes if K is number of yes answers b is the bias towards no if K is a number of no answers.

So, the binomial distribution answers the question what is the probability of getting exactly K yes answers, if there are total N answers and the formula is given known to us and this will help us in designing experiments as a yes or no question and quantifying and understanding it appropriately. So, I urge you to design your experiment as yes or no question as much if possible and think about this and apply this idea and use it for your quantifying your data with this I will stop this lecture and continue in the next class.