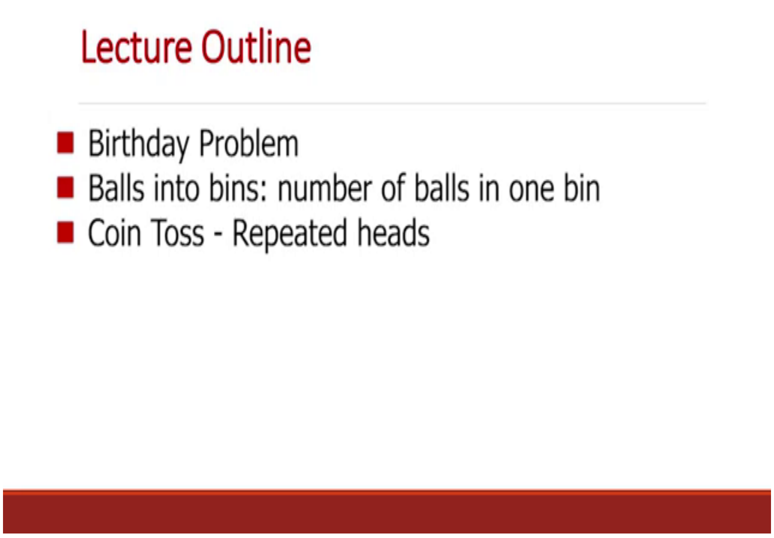
Probability Foundations for Electrical Engineers
Prof. Andrew Thangaraj
Department of Electrical Engineering
Indian Institute of Technology, Madras

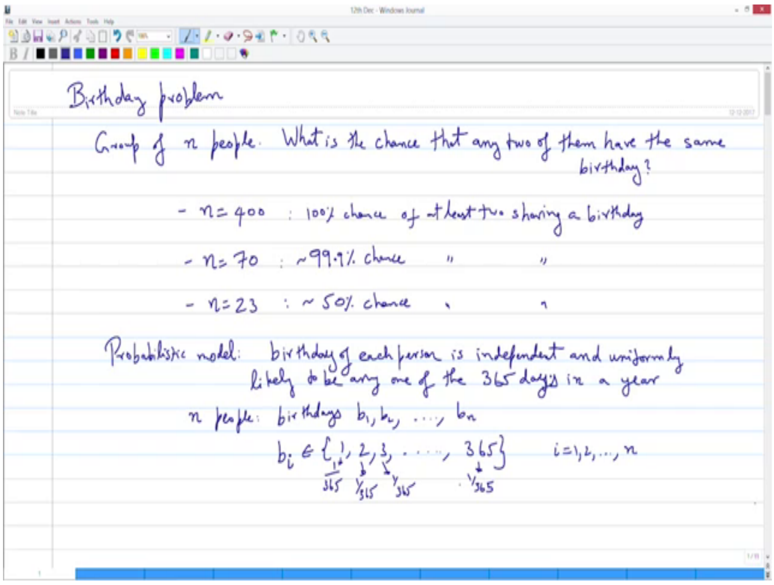
Lecture – 20
Examples: Binomial and Geometric Model

(Refer Slide Time: 00:13)



In this lecture we are going to see maybe ine example of a application of content conditional probability to a very interesting problem, and then maybe a couple of a simple examples following that for binomial model and geometric model and things like that. So, let us gets started the example I want to talk about is called the birthday problem ok.

(Refer Slide Time: 00:36)



So, very popular problem in the in this area, you might want to search it on Google or something and you will find a Wikipedia page, which talks about it and in my presentations very standard it is based on this is very standard ideas. So, the question that asked here is let us say you have a group of n people.

So, the basic question they want to ask is; what is the chance that, any two of them have the same birthday. So, that is the question. So, it is the sort of interesting question how a group of n people n could be any number 10, 20, 32, 100, 400 whatever what is the chance excuse me what is the chance that any two of them have the same birthday. Now this is an interesting phenomena maybe some of you have experience observing, it in a as you grow up in your own classroom, you would have let us say 30 people 40 people in a classroom and you will see invariably that might be 1 or 2 people who two people.

Who are the same birthday ends up happening quite often, and that is the very interesting answer to this question you can put a probabilistic model to model the birthdays of people and then come up with the calculation which will show you this number and there very surprising answers to this number.

So, for instance one way to start this argument is to first say, suppose n is say 400. So, if you think about it and the number of different birthday is possible is 365 right there are 365 days in a year may be if you include the leap year and February 29th you are going to get 366 right. So, there are only 365 possibilities and there are 400 people. So, if you start sending birthdays to each of these guys, you will have two people at least two people who share the same birthday ok.

So, in fact, this is very large. So, you imagine the number of people you need before two people will start sharing the same birthdays of this order, maybe 3 65 30 100 350 it is like some something of that order what you naturally think, but it turns out surprisingly. So, let me complete this if n is 400, then with you know 100 percent chance of two people sharing same birthdates; of at least two sharing a birthday right, because there just too many people and this would work. Now I will show you a probabilistic model some sort of a uniform probabilistic model, where you will have some surprising results. So, it turns out if n is something like 70, then roughly the chance of having two people having the same birthday is ninety 99.9.

So, even for 70 people it is almost certain the two people will have the same birthday apparently according to some probabilistic model that will put down, a chance is the same chances is 99.9 percent. So, you do not really need of the order of like 350 400 people before two people sharing a birthday. Even 70 people are enough when the probabilistic model is basically uniform and you will see, it is its very interesting model and it definitely works in practice. So, quite often this is true. So, if you have a class or group of people which are 70 and you expect their birthdays to be very random, it should not you know badly pick that group of people. Then almost surely to 99.9 percent I mean it is a very high percentage very very very high chance, that two people will have the same birthday. In fact, the much more interesting, suppose if n is 23, it turns out it is roughly 50 percent chance.

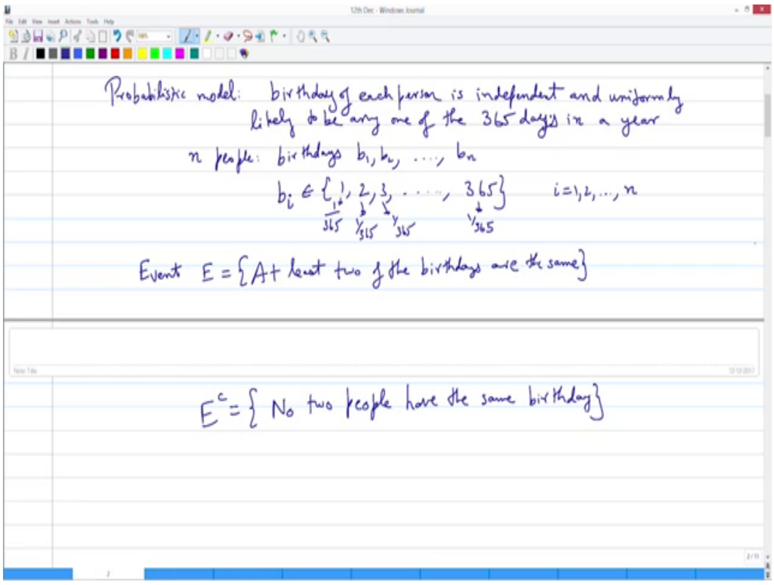
So, even if you have only 23 people in a group with probability half which is quiet high you know you think about 50 percent chance that two people have the same birthday. So, this is also something you can try and of course, to 23 people it may happen that you know two people have a birthday is only 50 percent it is not very high if it is 70 it becomes almost 100 percent. So, what is this model what is this interesting model that we have when, who does it work and it turns out like I said in the beginning conditional probability place and interesting role here.

So, let us define these probabilistic models. So, I am going to have a. So, in my model the birthday of each person is independent and uniform uniformly likely to be any one of the 365 days in a year. So, this is the model. So, what is; that means? So, let us say we have n people and their birthdays are b 1, b 2 so on till b n. Each b 1 can belong to I will simply number the days of the year from 1 to 365 of course; you can argue approximating here, I am not looking a leap years and all that.

So, we will just throw way people who born on February 29. So, we are not consider that we will just take a 365 days of the year and we will say birthdays are equally likely to be anywhere there so. In fact, you can bring in more complicated models; it only makes your life more difficult. So, this is easy enough. So, this is uniform. So, what does that mean? So, each birthday, let me put i here to indicate this i could be from 1 to n, every b i. So, every one of these birthdays can be either 1 or 2 or 3 all the way till 365 remember 365 would be December 31st one would be January first. So, that is just the way of way in which I am counting that days of the year from 1 to 365. So, each of these is 1 by 365 will probability, this is 1 by 365, this is 1 by 365 all the way till here. So, now, it turns out this may not be a very very accurate model if you go by a may be biological or other considerations.

This is panatela not true, but you know for our model it is quite reasonable and it works. So, this is the model that we are going to have this each bi's like this like this and there are there all independence in b 1 and b 2 are an independent events. So, this is also something important to consider is that. So, each birthday could be equally likely to be any one of the 365 days and they are all independent. So, in this setting I want to calculate that there is at least two of these birthdays are equal that is the idea is that hopefully you can see by the calculation come from. So, let us see how to go about doing a problem like this ok.

(Refer Slide Time: 08:53)

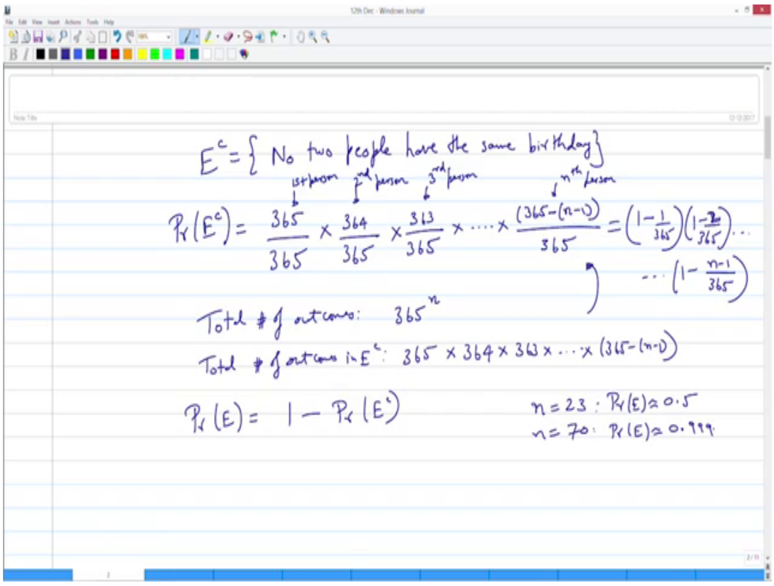


So, hopefully the model is clear. So, event what is the event I am interested in? The event I am interested in I the event E which is at least two of the birthdays are the same they fall on the same day. So, it turns out this event if you directly try to calculate the probability it is not impossible, but it is going to be lot more painful. So, quite often when you try to compute probabilities for a particular event, you should always think of the compliment of the event, because in many cases the compliment of the event might be easier to calculate. So, in this case also that is true let us look at E compliment what is E complement? No two people have the same birthday. So, this is what I am going to look at. So, no two people have the same birthday. So, I am each birthday is independent and uniformly likely to be any 1 of the 365 days that is my model.

And I want to find the probability that no two of them have the same birthday they should all have different birthdays. So, that is the little calculation that we wanted to here, and I mean you might see that this is this can be done using conditional probability. So, we first look at the first person's birthday. The first person's birthday could fall on any one of the 365 days. Now given that first persons of the birthday has fallen somewhere where should the second persons birthday fall, it should not be on the same day as the first person's birthday. It can be any one of the remaining 3 164 possibilities what about the third person given the first two persons birthday where different it cannot be any one of the two possibilities which should already there. So, it should be there 363 possibilities. So, that is the way in which you can conditionally argue given each of the birthdays ok.

So, let me try to write that down in a reasonable fashion ok.

(Refer Slide Time: 11:06)



So, I am going to write the probability of E compliment by conditioning on the first persons birthday like, I mentioned the first persons birthday can be any one of the 365 possibilities. It could be anything and min my event of E compliment no two people having the same birthday, it is not going to be false given any day for the first person.

Now given that the first person as a particular birthday and given that the second persons birthday is independent of the first persons birthday, and conditioned on that; what are the favorable outcomes for me for no two people having the same birthday the second person should not have the same birthday as the first person right if you has the same birthday as a first person my event is violated. So, this is not true, second person should not should not have the same birthday as the first person. So, that becomes 364 by 365 possibilities. So, notice I have used both the fact that you know I have use the fact that the two birthdays are independent.

So, this is for the first person this is for the second person. So, I have used the fact that the birthdays of the two people are actually independent; and I also used conditioning a first return the event has a combined event of birthdays of all these people then I am looking at the first persons birthday condition that the first persons birthday was at a particular date, what are the favorable outcomes for the second one then I can keep on going. So, what about the third person? Third person can be born on any one of the remaining 365 63 possibilities and so on. So, you can go on till 365 minus. So, so this is for the third person. So, for the n-th person you will have n minus 1 here by 365.

So, how did I get n minus 1 here you can think about it for the second person I have 365 minus 1, for the third person I have 365 minus 2. So, for the n-th person I should have 365 minus n minus 1. So, this is the probability that no two people have the same birthday. So, you think about how conditional probability and independence together are helping you to build up the probability for this event, in a very simple manner. So, you can also do it in so many other ways. So, for instance you can look at the total number of outcomes. The total number of possible birthdays for n people is 365 power n. So, that is the total number of outcomes. What is the total number of favorable outcomes? Outcomes in E compliment that will be no two of them should repeat.

The first one can be any one of the 365 possibilities, next one should be 365 next one should be 363. So, until 365 minus n minus 1 right, that is the idea is that. So, the same thing happened here, but you use conditional probability and independence like the way I showed you, you can directly write down. So, this is the probability of E compliment that no two people have the same birthday. So, what is the probability of E? That at least two people have a birthday it is 1 minus probability of E compliment and now you can do a calculations for different n you can go ahead and calculate this, you could be a little bit care full I would suggest the way you calculate this it is best to write it like this, the first one is just one the second one is 1 minus 1 by 365 times 1 minus 1 by 2 by 365 sorry, on till 1 minus n minus 1 by 365.

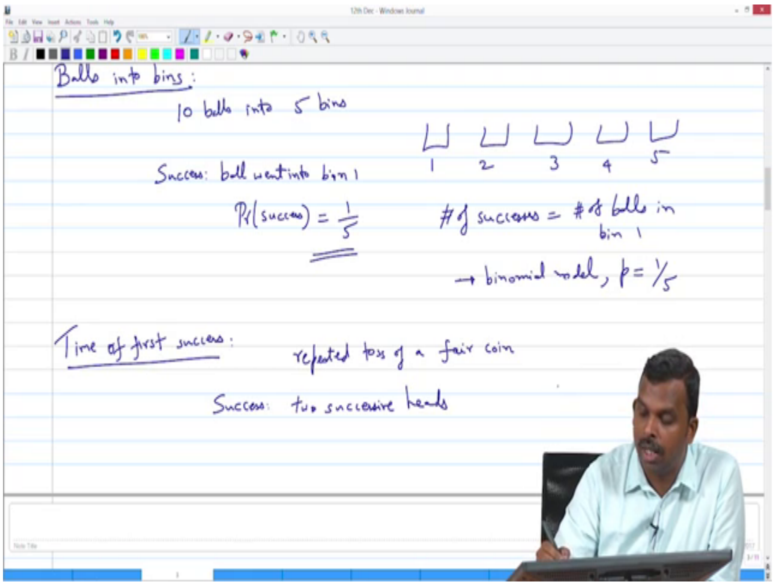
You write it like this and then do the multiplication and find the probability and you will see surprisingly for n equals 23 probability of E is roughly about 0.5 and for n equals 70 probability of E is roughly about 0.999. So, this is the result for the birthday problem quite a simple calculation of the end, but it is little bit surprising in the way you interpret so. In fact, even the birthday problem the way it is framed is the model that I build here is actually the same as balls and bins, where are the bins and where are the balls there are 365 bins each representing 1 day of the year, and you have n people these are the n balls that I am throwing uniformly and independently at random into this bins. And the event

E at least two of the birthdays are same asks are there at least two balls in any 1 of the bins what is E compliment? No two people have the same birthday all the bins have only either 1 ball or 0 balls that is the E compliment and you can do it in this fashion.

So, if you have just 23 balls and you are throwing them into 365 bins uniformly and independently, you have a 50 percent chance that two of them have the say 1 bin has two balls two balls went into the same bin. So, bit unreasonable and some sense, but it is the calculation then this model uniform independent model that is the calculation.

So, once again it is nice illustration of conditional conditioning on the sequence of events, conditioning on the previous event next event and all that and using the independence of different things to come up with this come up with this (Refer Time : 16:38). So, you can see so many nice results on can drive in this fashion. So, hopefully this was an illustrative example for you. One more keeping up with these balls in bins example I also want to show you one more nice way of thinking about binomial model.

(Refer Slide Time: 16:58)



So, if you look at balls into bins.

So, if you are throwing a let us say 10 balls into 5 bins. 5 bins 1 2 3 4 5 you throwing 10 balls into 5 bins, I am interested in a success event which is. In fact, ball went into bin one. So, I am throwing this ball, I am going to count a success if the ball went into bin 1 and failure if it went into anything else. So, the probability of success is actually 1 by 5.

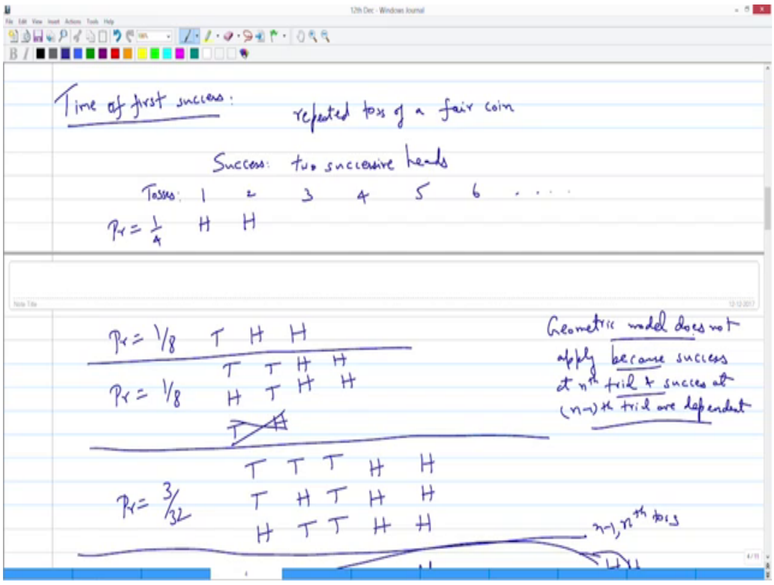
So, number of successes this is what? Is the number of successes it is actually the number of balls in bin 1, right.

So, this experiment is actually if you if you had only worry about the number of balls that got into a particular bin or say for instance bin 1 here, this is nothing different from some sort of a coin toss and counting the number of heads accept that instead of probability of heads and probably of tails being hafiz, probability of successes 1 by 5. So, if you want to count the number of balls in bin 1 you once again have a binomial model with p which is probability of success being 1 by 5. So, this binomial model this way arises in many many interesting situations anytime u want to count whether you had a success or not in a particular attempt and you want to count the total number of successes by the binomial model naturally arises. So, I want to leave you with 1 intriguing question about the geometric model, we will come back and look at it little bit later on the geometric model usually talks about the time of your first success.

So, turns out one can define for success in a slightly more complicated way to make the problem more interesting. So, for instance supposing you have a repeated coin toss repeated toss of a fair coin. So, you might define success as the first head. So, that person example is very classic example for geometric model, the probability that the time at which you have the first head or you can also define in a slightly more complicated way you can talk think of a probability of having the second head. So, these are all very popular models. So, I am going to talk about a model which is slightly more intriguing and more difficult. So, success I am going to define as two successive heads.

That this model for success, you can immediately see is a little bit more complicated. So, what do I mean by two successive heads I keep tossing a coin if I get two heads in a row I can stop now have succeeded. So, the kind of question that would ask is, what is the probability, that you will have the first time two successive heads come is at the particular position. So, for instance, you are tossing a coin. So, let us say you have. So, you have your sequence of tosses.

(Refer Slide Time: 20:34)

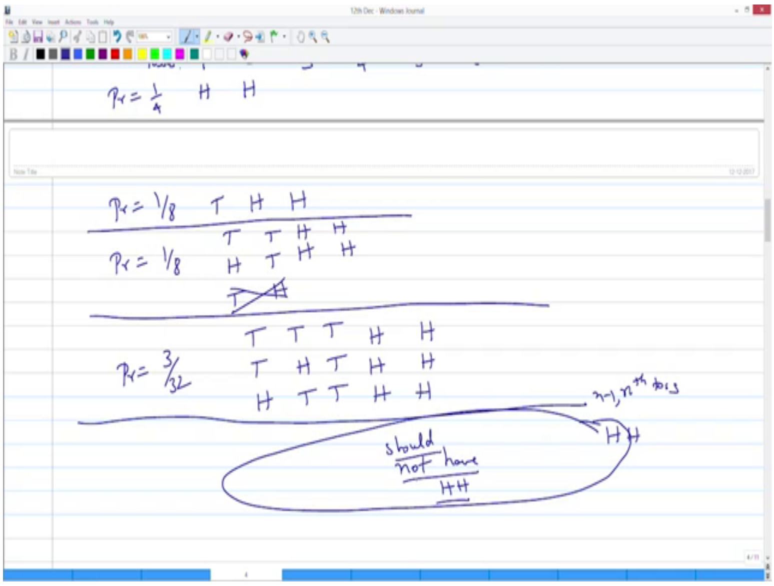


So, the first toss, second toss, third toss, fourth toss, fifth toss and so on. So, you will have a sequence of heads or tails. So, supposing if you could have heads heads here in which case you will stop. So, what is the probability of this? Probability is 1 by 4 what is the probability that the first time you have to consecutive two heads, is in the second and third position? So, you should have consecutive heads in the second and third position, you should not have consecutive heads before that which means it should be a tail and the probability is 1 by 8.

So, please can see well it is going. So, what about first time two consecutive heads occurred is 3 and 4. So, so hopefully you can see. So, I am tossing a coin repeatedly when I get two successive heads I stop. So, first question is; what is the probability that you will stop at the second toss. So, first two should be heads. So, you will stop at the second toss what is the probability will stop at the third toss? Second and third must be heads and you should not have the first also as heads because then you do stop that the second one right. So, this is THH and that would be 1 by8. So, what about the probability that will stop at the fourth place this is a little bit more complicated now right. So, you can have in the first two tosses right you could have for instance TT right you can have TH.

you can have can you have TH I am sorry I am sorry you can have TT you can have HT, you cannot have any other possibility you cannot have for instance TH, if you have TH what will happen? You had HH here hear you had stop. So, this is actually not possible. So, the probability here of these two things happening, you may be draw line here to indicate that this is ended the probability of these two things happening it could be TTHH or HTHH. So, that is 1 by 16 plus 1 by 16. So, that probability is 1 by 8. So, by the fourth toss you already have half the probability. So, what happens if you go to the fifth toss? So, maybe I should draw this little bit blow.

(Refer Slide Time: 22:55)



So, if you want to look at two consecutive heads by the fifth toss, you would have H H here and what are the various things that are possible here in the first 3 tosses.

you have to look at all those possibilities you could have TTT or you could have THT right and you cannot have THH, you can have the HTT that is it this is 3 possibilities right. So, these are the possibilities under which you can have two consecutive heads appearing only in the fifth place. So, that probability is already you know this each one of these is 1 by 32 in probability. So, you have 3 by 32. So, you see the probability is dropping quiet drastically and it is not really very easy to compute so. In fact, what is the thing for that you will have a success at the n-th toss and you will have to stop at the n-th toss, n minus 1 and n have to be HH and in all these guys including the n minus 1 th toss you should not have HH anywhere.

So, turns out computing this kind of things it is quite tricky it is not very easy. So, this is not really a pure geometric model in the sense why is that happening, why does this happen because this is not a pure geometric model right; your success in one position a second success in the next position are not independent. So, geometric model applies when you have a sequence of events they are all independent and you define success based on 1 event. Now what happened here is you repeatedly tossing a coin and your defining a success event based on two events and that complicates everything. So, once you have two events successive events because the probability the event that you succeed is the second toss in the event that you succeed in the third toss now end up being related. And you have to count this successive heads very very carefully, and computation becomes a bit more difficult.

So, it is not very difficult or impossible to calculate this exact probability, but much more challenge geometric model. Does not apply in this case why does geometric model does not apply because success at n-th trail and success at n minus 1 th trail are dependent. So, I needs to be careful and also in any example that you come up with you should be aware that the what conditions the model applies, and in what condition the model does not apply. So, dependents is always a situation that you have to take care of a very carefully. So, that is the end of this lecture then.

In the ensuing lectures we will start looking at random variables we will see processor Aravind's lectures explaining what a random variables, and different types of random variables and I will come back and do some examples of those random variables as well.

Thank you very much.