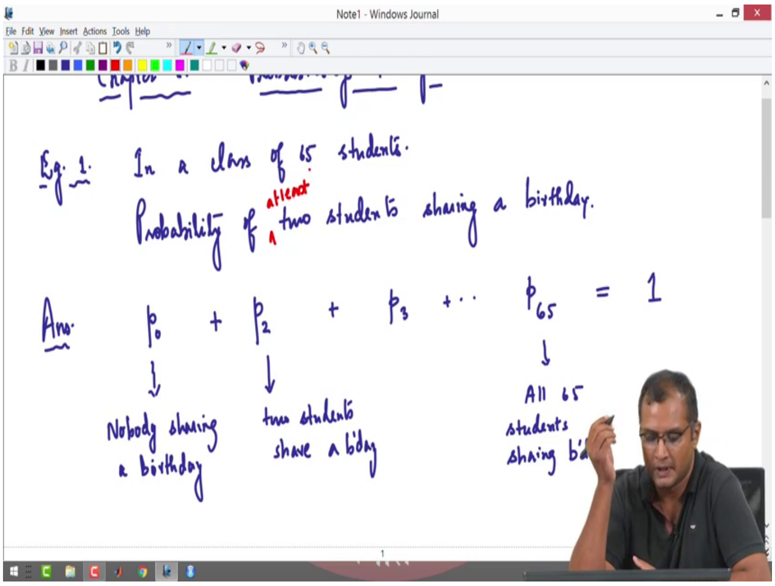
Statistical Mechanics
Prof. Ashwin Joy
Department of Physics
Indian Institute of Technology, Madras

Lecture – 01
Discrete Probability

Good afternoon to you all, today we will start the first chapter on Probability theory.

(Refer Slide Time: 00:19)



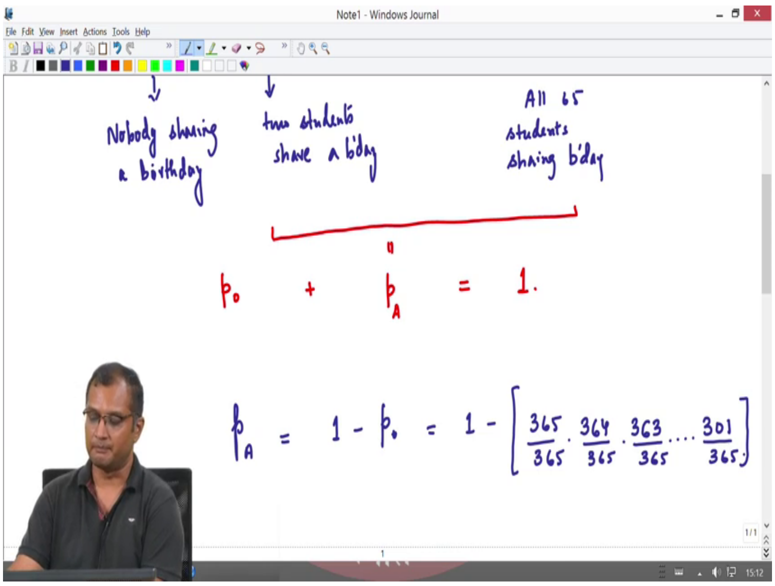
And, before we dive into the theoretical rigor of probability theory I think it is a good idea to start with some its very simple examples concerning discrete probability theory, where we will try to solve some examples using objective approach of probability theory. And the focus would be not to rely on formulas, but rather on the ability to solve the problem using what is given and what is required. And, to illustrate this I would take some examples very simple examples on probability theory. For example, the first example would be if you have a class of some students. Let us say you have a class of 65 students typically the strength of any master's class in IIT Madras.

And, you ask yourself a question what is the probability that any two of them will share a birthday, any two students in the class sharing a birthday ok. Now this is a very interesting problem. In fact, if you look it up in the web or if you look it up in basic textbooks on probability theory there is an analytical result which says that you require about 23 students in a class of, in a class such that the probability of any two students having a birthday which is same exceeds 50 percent, that is a result you can derive with the end of this in this problem.

But I am going to basically pose a simpler question that you have a class of n students n being 65 here, what is the probability that two students will share a birthday ok. So, this problem can be done objectively in the sense that you can say that let us say you have p 0 as the probability of nobody sharing a birthday. Which means all the 65 students have different birthdays plus you say two of them share a birthday and this way you take 3 of them, 4 of them so on all the way to let us say all 65 of them sharing their birthdays ok.

Now, I know that the sum of all these probabilities has to be 1, because all of them have an exact birthday. And, I have taken all the possible cases on the left hand side, its last case being all 65 students having birthday on the same day. So, this is the left hand side is exhaustive collectively exhaustive, the sum of all this has to be 1. What is asked is basically that what is the probability that at least two students, I should slightly correct my question and then if the question is at least two students sharing a birthday.

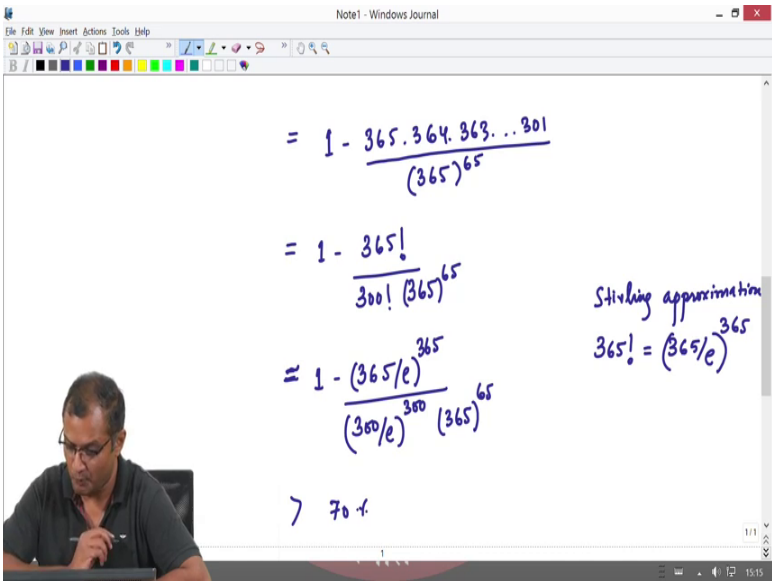
(Refer Slide Time: 05:07)



That is basically the sum of all these p 2 p 3 up to p 65 ok. So, this is basically my answer p of A and I have p 0 which is nobody sharing a birthday. So, p 0 plus p A should be 1 ok. So, which means my answer is now p A equals to 1 minus p 0 which is nothing, but 1 minus the probability that nobody shares a birthday ok. Now, a fact that nobody shares a birthday is a problem of simple permutation. Since, you have 65 students and you want to write down the probability that nobody shares the birthday, you can say that the first student has 365 days in the year to take as his birthday divide by 365 ok.

Now, the next student who comes has 364 days available because the first student is already taken one birthday in the calendar here. So, he is left with one less day ok, the third student who comes sees that the first two students have already taken 2 days in the year. So, his number of days assessable is restricted to 363 out of total of 365 and this way the last student has only 301 days available that is the probability of his birthday being one of them.

(Refer Slide Time: 07:13)

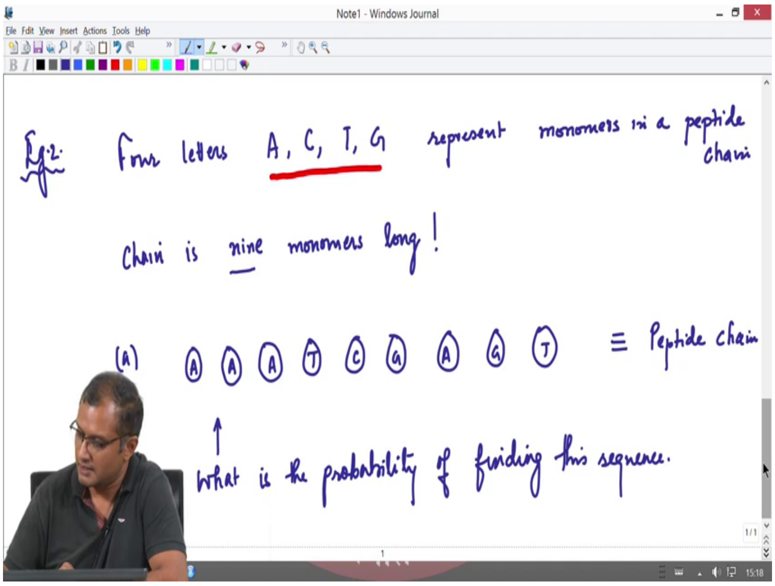


So, now you can basically write this as 1 minus 365 into 364 into 363 all the way to 301 divide by 365 to the power 65 and this can be written as 1 minus 365 factorial upon 300 factorial into 365 by to the power 65 ok. And, if you solve this the answer that you get you compute it in a calculator or you can use a simple Stirling's approximation that I can write down this, 365 to the power such because this is a large number, a large number factorial can be written as if n is large n factorial is 36 n n by e to the power n ok.

So, I can write this as 1 minus 365 by e to the power 365 upon my of course, I have to write down a nearly equal sign because I have used Stirling approximation over 300 by e to the power 300 into 365 to the power 65. If you solve this your answer would be I have not calculated this, but you can use a calculator to proceed from here ok. So, I leave that as an exercise, this answer should be greater than 70 percent ok. So, exact answer can be computed by using a calculator, but it would be definitely greater than 70 percent ok.

Now, so this is an example of doing this problem objectively without using any formulas. So, there is one more problem that I would like to immediately discuss which is basically a problem of letters.

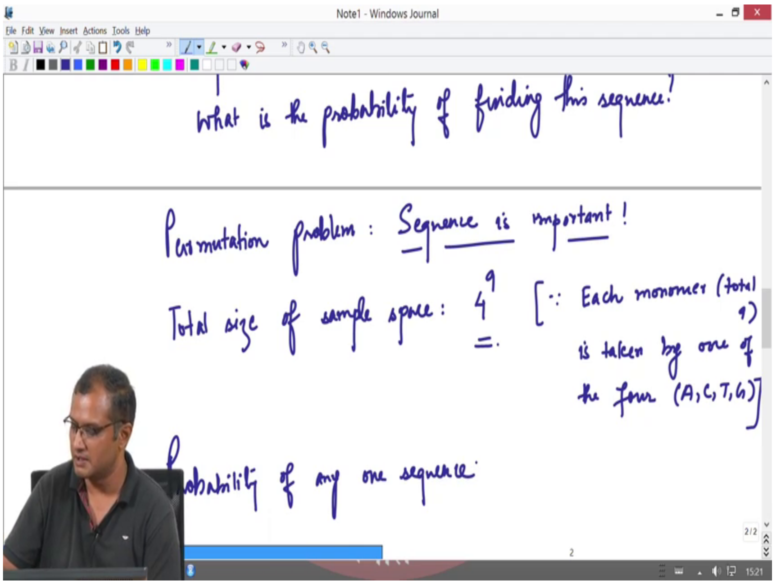
(Refer Slide Time: 10:11)



So, I will take the second example and you have basically let us say that you have four letters and the letters I spell them out as A, C, T and G and suppose these are the four letters that are basically represent monomers in a peptide chain. So, you have a chain of monomers which is basically composed of only these four letters ok. And you say that my peptide chain is nine monomers long and you ask yourself a question that. So, I have nine monomers 1 2 3 4 5 6 7 8 and 9.

So, this is your typical peptide chain and the empty slots have to be filled with these four letters, these four letters. Now, the question here is now, what is the probability of finding a sequence A A A T C G A G T ok. So, what is the probability of this sequence? Probability of finding this sequence; now you can construct very large number of sequences, of this large number of sequences one of the sequences in front of you A A A T C G A G T. What is the probability of finding this sequence, what of those very large number of sequences. So, this is already we have permutation problem.

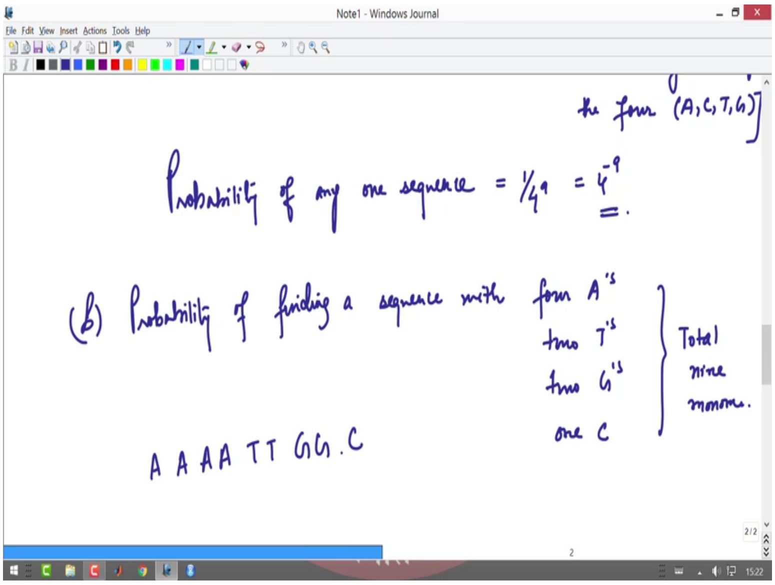
(Refer Slide Time: 12:59)



Because here the exact sequence of the monomers is important because, if I would change the relative positions of T and C then the peptide chain would be different. So, sequences important here. So, this is the permutation problem and then to answer what is the probability of finding this sequence I must first find out the total size of the sample space, which is total number of sequences that are possible ok.

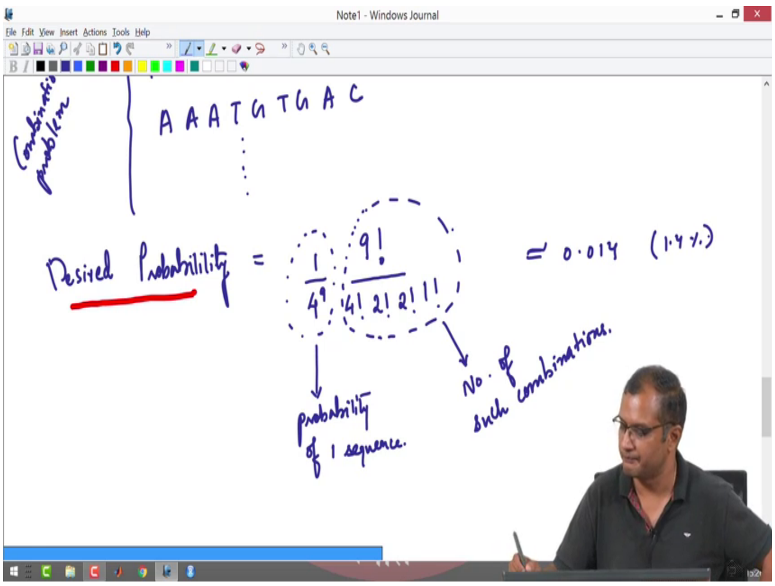
Now since there are nine monomer units each of them can be taken by these four monomers A C T and G which means the total number of realizations that I can construct is 4 to the power 9 ok. Since each monomer total 9 we have total 9 of them is taken by one of the 4 that is A C T and G ok. So, if we have tie 9 vacancies and each vacancy can be taken by one of these 4 the answer that how many, the answer to the question that how many total; you know what is the total number of peptide chains you can construct that is 4 to the power 9. And so the probability of any one sequence is of any one sequence.

(Refer Slide Time: 15:19)



In this case the sequence mentioned here is exactly 1 upon 4 to the power 9 or 4 raise to minus 9, that is the answer to your question. The second question is what is the probability that you will find a sequence with four A's, two T's, two G's, one C ok. So, we can see that four A's, two T's, two G's and one C's basically total 9 monomers. So, basically we can take four A together then you can take two T's together then you can take two G's together and one C, but you can also have you know three A's together.

(Refer Slide Time: 16:45)

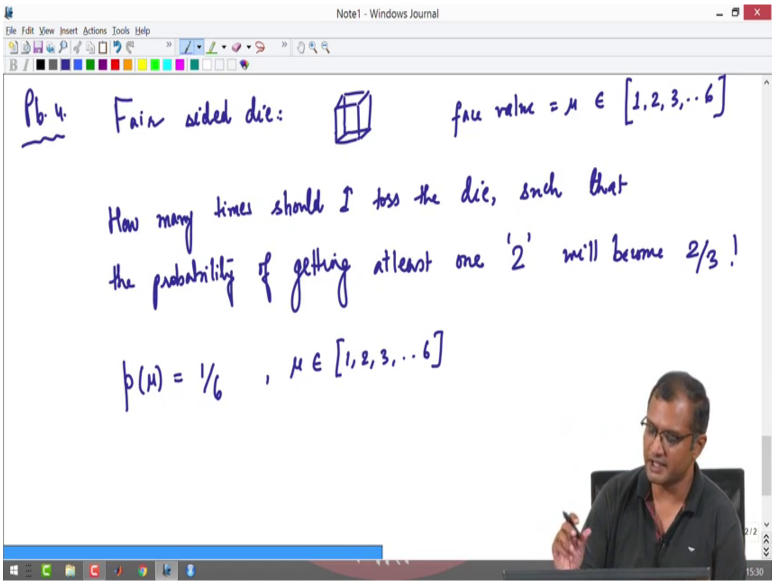


Then you can put one T, one G put another T another G one A and C and this where there are many sequences. So, the question as here is what is the probability of finding a sequence with four A's two T's two G one C basically a sum of all these you know permutations divided by the total number of permutations. So, this essentially a combination problem, where you do not say that I want exactly this permutation; what you have done is essentially asked a question that I am interested in a total number of ways in which I can get four A two T's two G's and one C ok. So, this is exactly it is different from the previous problem the sense that you are essentially saying that I do not care about the sequence.

All I want is: what is the probability of finding four A's two T's two G's and one C ok. So, this can be done in a very simpler way, you can find out the probability of you know you can find out the total number of. So, the probability of this particular case is given as the number of possible you know the combinations you can have which would be if you have 9 monomers. And, this has a multinomial combinatrics where you will have a, where we can say that this is equal to 9 factorial divide by 4 factorial into 2 factorial into 2 factorial into 1 factorial ok. So, we can see that 4 plus 2 plus 2 plus 1 is 9. So, this is like this is exactly the probability of this is the number of combinations that you can have. And, then if you divided by the total number of combinations which is 4 to the power 9, this will essentially give you the total the probability that we achieve.

One way to interpret this is that 1 upon 4 to the power 9 is the. So, this a decide probability 1 upon 4 to the power 9 is the probability of one sequence and this factor is combinator factor is basically number of such combinations. So, we have number of combinations multiplied by the probability of each sequence in this combination that gives us the desired probability ok, that is your answer. If you work it out, it would come out to be around 0.14 which is like saying 1.4 percent ok. So, you work it out.

(Refer Slide Time: 20:53)

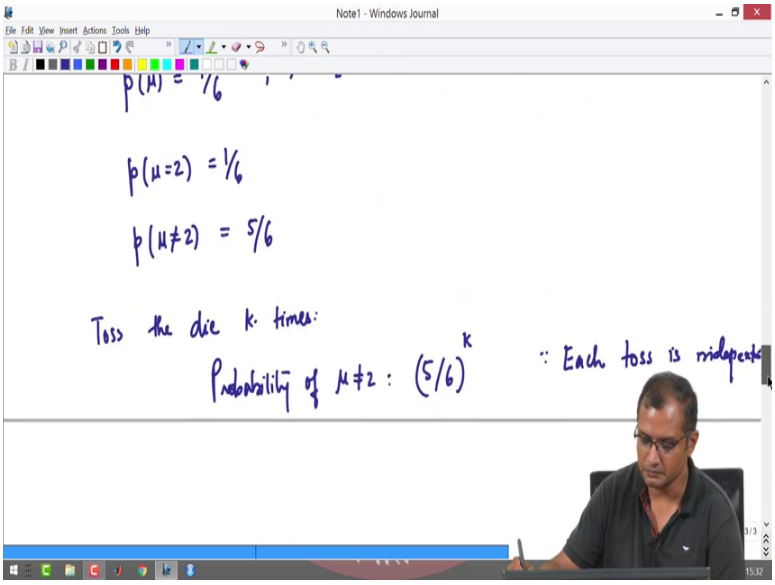


One problem that is important here that concerns the role of a die. So, if I take a fair sided die ok. So, this is another problem of a die which basically a cube shaped object and on each side of this cube you have a number written on it ok. So, you have you can say that the face value of a die is let us say mu; so the face value is mu and this mu belongs to one of these digits ok. And, the question is how many times I must toss this die such that the probability becomes two-third that at least 2 will appear at least one 2 will appear, which means the question is how many times should I toss the die such that the probability of getting at least one 2 will become two-third.

Suppose if I toss it to once, I know the probability of getting 1 2 3 all this digits is 1 by 6 ok. So, I know that I know apriori that p of mu is 1 by 6, where this mu belongs to the set of these 6 numbers. Now you may toss it is the first time and you will get 2. So, you got 2 straight away, but it is possible that if you toss it 3 times you may not get a 2, but you get a 2 the fourth (Refer Time: 23:31). So, basically how many times should I toss this die such that the theoretical probability of getting at least one 2 will be two-third, certainly you do not have to toss it 100 times because I know that all these faces are unbiased. So, there is a one sixth chance of getting a 2 on any trial. So, you do not require a very large number of tosses.

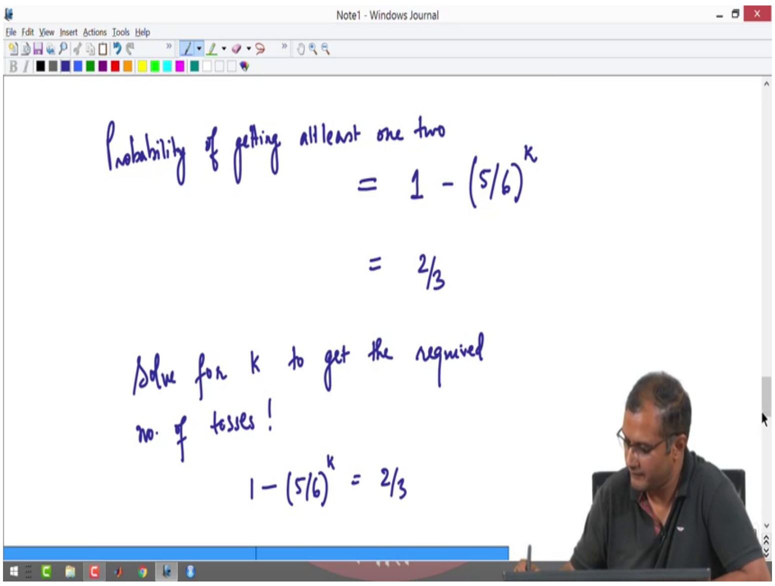
But just to compute the minimum number of tosses required statistically to say that the probability of getting 2 will become two-third ok. So, this question can be done in a certain, in a certain way that.

(Refer Slide Time: 24:21)



If you know that the probability of getting a 2 is 1 by 6 you can say that the probability of not getting a 2 is 5 by 6 ok. Now this is saying that if I toss it you know k term k times, suppose I toss the die k times ok. So, probability of not getting a 2 would just be 5 by 6 to the power k because each tosses independent of the previous toss, the probability is a multiplicative dependent. So, the probability is just keep on multiplying. Now, the question that was asked was: what is the probability that the occurrence of 2 becomes two-third.

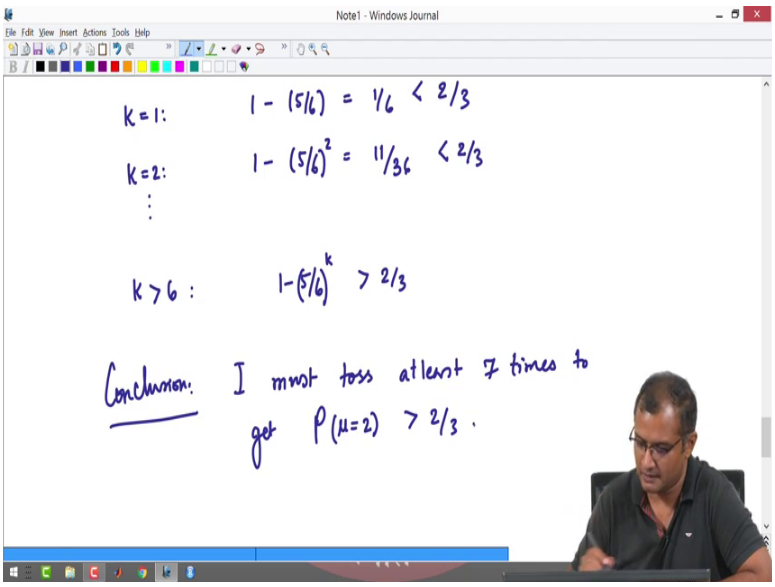
(Refer Slide Time: 25:51)



Which means if I just I know that the probability of you know I can say that after k tosses the probability of getting at least one 2 would be just 1 minus 5 by 6 to the power k right. Because, there were only 2 options either you get a 2 or you do not get a 2. 5 by 6 to the power k is the probability of getting no 2 at all. So, the answer to the question is simply 1 minus 5 by 6 to the power k.

So, what we have done is instead of tackling your problem directly you have basically divided your sample space into 2 2 2 sections. So, one is you get no 2's and the other section is basically you get at least one 2. The sum of these two regions is basically your entire sample space, which is denoted here as probability 1 ok. So, what is basically given to me is that 1 minus 5 by 6 to the power k is given as two-third. So, this is the probability desired probability of getting at least one 2, now I am must solve for k to get the required number of tosses because, k is the number of tosses. So, at the end of this problem I would get the total number of tosses required such that the probability of getting a 2 becomes one 2 by 3rd. So, that would mean you solve this equation for k and that would.

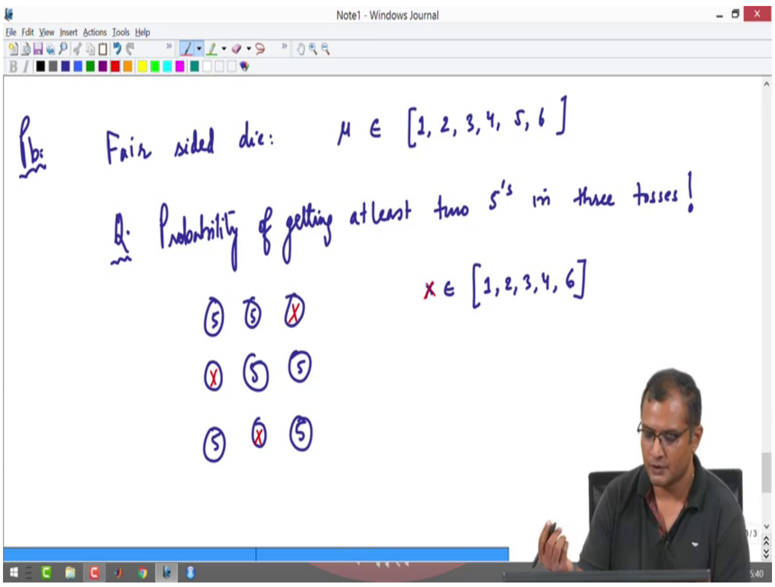
(Refer Slide Time: 28:13)



For example, if you take k equals to 1 this is the, if you just toss at once then you can see its 1 minus 5 by 6 equals to 1 by 6 and this is definitely less than 2 by 3 ok. So, I do not get the equality for 1 toss, now if you take 2 tosses you get 1 minus 5 by 6 to the power 2 which is basically 11 by 36 this is also less than 2 by 3. So, this way you keep on sort of increasingly eventually you will see that for k equals to for k greater than 6 and above for k greater than 6 actually, you will see that 1 minus 5 by 6 to the power k is basically greater than 2 by 3.

Because as k increases 5 by 6 to the power k will become smaller and smaller because; 5 by 6 is already less than 1. So, a larger power would only make 5 by 6 to the power k smaller which means 1 minus 5 by 6 to the power k will become larger. So, typically for k equals to 7 you will see that the probability becomes greater than two-third. What this means is that I must toss. So, the conclusion that take home conclusion is that you know, one must toss the die at least 7 times to get the probability of getting at least one 2 or greater than 2 by third that is the conclusion fine.

(Refer Slide Time: 30:43)

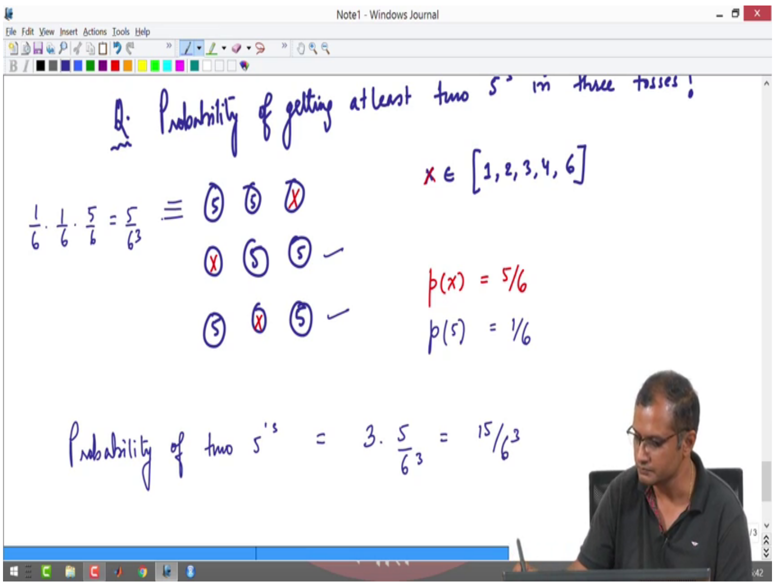


Next I have a allow a very simple problem at the end if it which will close this lecture. I will just give you one more problem where you can solve you know for probabilities objectively. So, let us take another just as before a fair sided die which means the face of a die can be 1, 2, 3, 4, 5, 6 that is the face of a die which is mu a variable and can take one of these values or one of the symbols 1 to 6.

So, suppose I ask a question what is the probability of getting at least two 5's in three tosses. So, basically your tossing it the first time, tossing it the second time, tossing it the third time; you have tossed this fair sided die of 6 faces 3 times. And what is being asked is that you should tell me the probability of at least getting two 5's which means I could get a 5 here and something else here ok. So, there is something else basically mean that you could be getting 1, 2, 3, 4 and 6 lot of 5, 5 is absent.

And you could also get something here and a 5 here and something here ok. So, this would be an in fact, you want two 5's. So, let me just a correct myself here you want to get two 5's that is the question right. So, this is obviously, not correct. So, let me just correct it. So, I am writing now all possible ways of getting two 5's is one more way of getting 2 5's that would be keeping something here and two 5's here ok. So, since we have two different types of symbols I am going to colour them with different colours ok, so the 3 ways in which you can basically get two 5's out of 3 tosses.

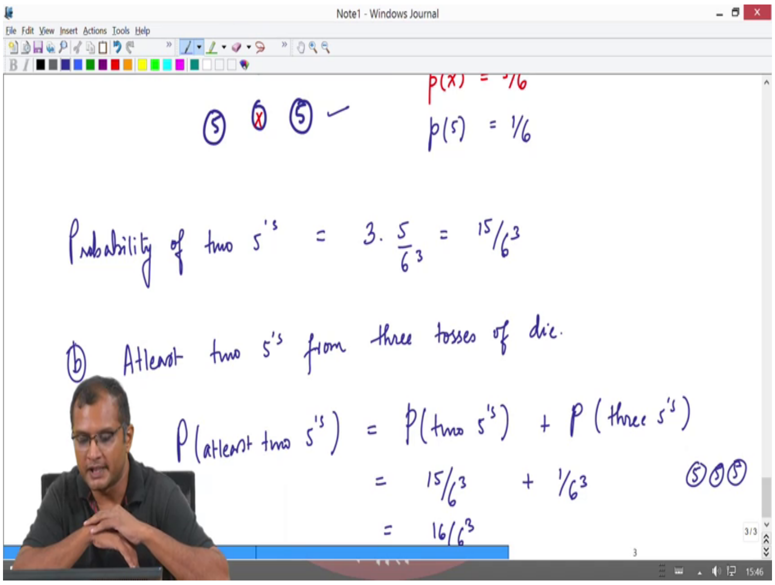
(Refer Slide Time: 33:53)



So, the probability of getting at least two 5's out of 3 tosses is see the probability of getting an x is basically 5 by 6 and the probability of getting of 5 is 1 by 6. So, if you look at this particular observation, probability of this observation is 5 by 6 cube, which is if I break it down the probability of this is. So, let us write it below, so let us write it below here. So, the probability of getting a 5 is 1 by 6 and probability of getting another 5 is 1 by 6, the probability of getting an x is 5 by 6 ok. So, which is basically 5 by 6 cube. Now there are 3 such ways in which is possible, you have one more way here and one more way here.

So, the total probability of getting two 5's in a 3 toss of two 5's is basically 3 into 5 by 6 cube which is 15 by 6 cube that is the answer to the problem.

(Refer Slide Time: 35:37)



You can also ask another interesting question is and that is as follows what is the probability of getting at least two 5's from 3 tosses of the die ok. So, what we solved previously was the probability of exactly two 5's, what I am asking now is the probability of at least two 5's; probability of getting at least two 5's is equal to probability of getting two 5's plus probability of getting three 5's ok

That is the only two possibilities and this would be already calculated, this is nothing but 15 by 6 cube as just calculated from the previous task. And the probability of getting three 5's is just 1 by 6 cube because there is only one way in which you can get three 5's. So, you can get 5 and a 5 and a 5 and that is it, is one combination of three 5's out of the 6 to the power 3 ways. So, the answer here is nothing, but 16 by 6 to the power 3 ok.

So, this is one way to solve it and that is the end of this particular lecture. In the next class will talk about random variables in general. So, this particular lecture was about discrete probabilities and discrete random variables, we will start the next lecture and discuss continuous random variables. And, this is the type of random variable that you will use throughout the course in statistical mechanics. And these random variables have a profound importance in all areas of equilibrium statistical mechanics for example, the velocity of gas particle in this room at some temperature is a random variable which a continuous random variable.

So, any theory that we develop in the context of continuous random variables their means, their higher moments, their cumulants would be of great importance in the theory of statistical mechanics that will developed in chapter 2. So, we will meet again in the next lecture and start off from that.