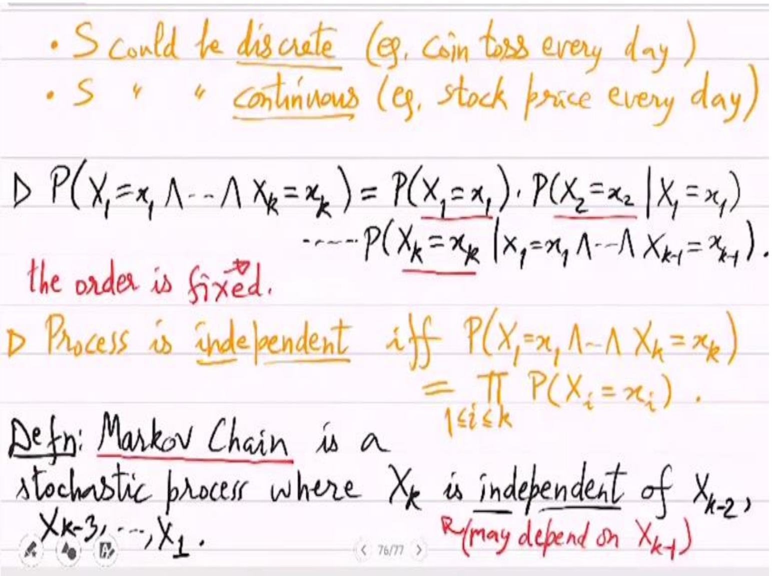
Probability for Computer Science
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology – Kanpur

Lecture - 18
Stochastic Process: Markov Chains

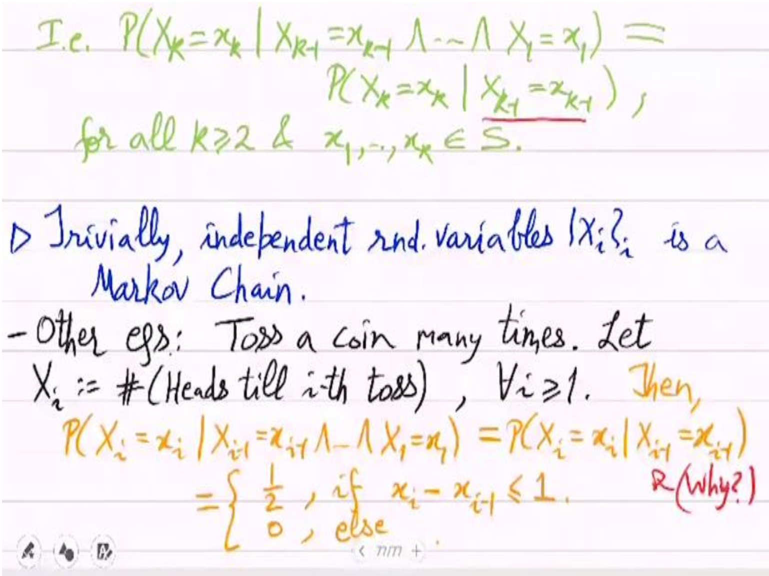
(Refer Slide Time: 00:13)



And the other special cases or the very first special cases if these experiments are independent, then what you can write. Then you can simplify this. So, if the stochastic process is independent then essentially this probability that you are looking at it should just be product of probabilities. That is what independence means. So, independent stochastic process has this product probability. General stochastic process has this ordered conditioning.

And then you take the product. So, based on this, we will study a very nice and specialized stochastic process which is called a Markov chain. So, Markov chain is a process where X k is independent of X k minus 2 X k minus 3 dot dot X 1. So, it may depend on the previous one immediate previous but not the history not anything before the previous.

(Refer Slide Time: 02:56)



So, that is the probability of X k being little x k. Given the previous ones, this probability is just the same as this one, X k minus 1. So, probability of small x k coming only may only depend on little x k minus 1 and nothing that precedes. And this is true for all k greater than equal to 2 and all x 1 little x 1 to little x k in the domain in the state space. This happens for all possibilities. Then you say there it is a Markov chain or a Markov process.

So, remember this means may depend on X k minus 1. That is what we really meant. So, it is clear that independent process is also a Markov process. So, independent random variables they do form a Markov chain because not only X k minus 2 but everything that comes before X k is independent of. So, the condition is satisfied. So, independent random variables do form a Markov chain but are there more interesting examples.

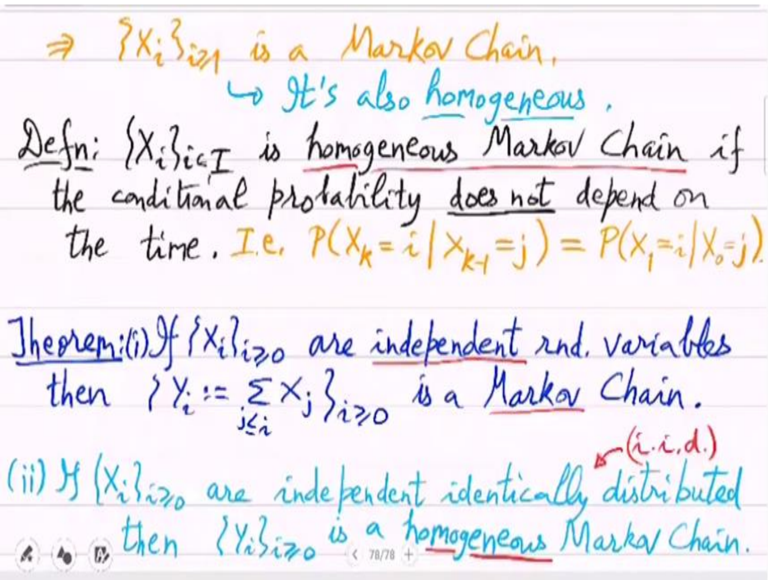
So, let us see one. So, toss a coin many times. So, obviously these will be independent events. But you are interested in counting, how many heads you saw till the ith toss? So, let X i be the number of heads till ith toss for all i at least 1. So, what you can show is probability that X i is small x i. Given the previous things, you can see that this all that matters is i minus 1 not the whole list. So, think about this.

So, the reason is that once you know the count if you know the count up to i minus 1 of heads then the ith count just depends on your current flip the current toss. It does not matter what happened in i minus 2 i minus 3 and so on so. Just from the previous number of heads you have to add the current experiment value whether it is 1 or 0. So, it just depends on the previous count which in any case is half.

So, no matter what the count was before. The count not changing means that you are tossing a tail the it is a you get a tail and the count increasing means that you got a head. So, this will actually be equal to half if x i is x i minus 1 or plus 1. So, let me write it like this. Either you are asking for one more or you are asking for no change. Else, it will be 0. So, if you are asking for a change of 2 heads then it is not possible.

And if you are asking for a change of minus 1 head then that is also not possible. So, it only the probability only depends on the previous which means that this is a Markov chain.

(Refer Slide Time: 09:04)



So, an interesting and very simple Markov chain is just counting the heads. Keep tossing coins count the heads. And these random variables they form a Markov chain. And there is some dependence on the previous value. That is a good example to keep in mind. In fact, it has one more property. It is homogeneous. So, what is that? So, let us define it. So, we call process homogeneous Markov chain if the conditional probability does not depend on time.

So, let me clarify this in terms of formula. What we are saying is probability of X k being a value given the previous value. So, after j getting an i, it does not depend where in the process you are asking the question. So, this is equal to the very first from the initial configuration this big X 0. If that was j, what was the probability next time you got i? This probability does not change if you look at X k on X k minus 1.

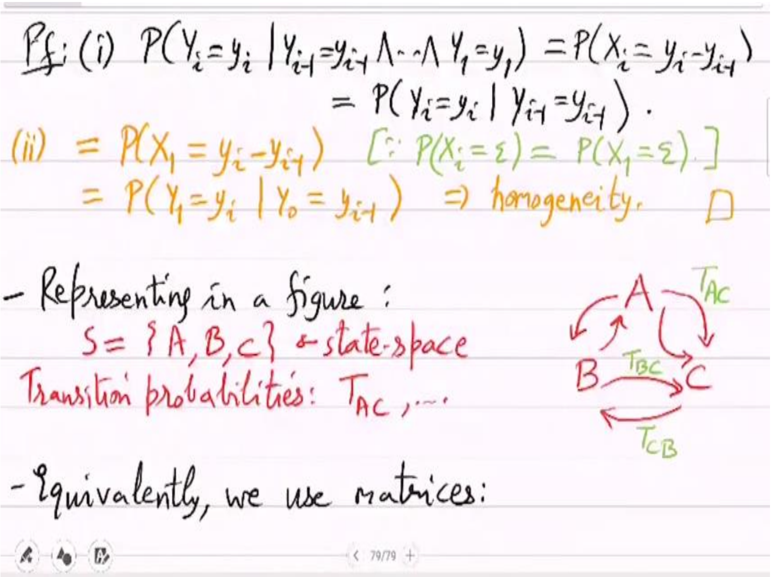
So, in this case, we called the Markov chain to be homogeneous. And this example of coin toss actually satisfies this because you look at the values this half and zero, they really do not depend on i. They just depend on what the difference of the values is. So, this is a very nice example. It gives you really the meaning of all these definitions that today you have seen. So, let us now prove a more general theorem which will justify or which will explain to you why the coin toss example is so nice.

So, this says that if you actually take independent random variables. So, if X i are independent random variables, then take their sum the initial sum. So, j going from 1 to i. So, these Y i’s are now just taking a sum of these of the first 0 to i ones. So, it is a new process and this is a Markov chain. This is why the coin toss example worked because you were getting heads and then you were taking actually some of the Bernoulli random variable.

And second is about homogeneity of this. So, if X i are identical. So, they are independent identically distributed. So, this is called iid, independent and identically distributed. So, if X i's are iid, then these Y i's this is actually a homogeneous Markov chain. So, this is a generic way of constructing Markov chains and how they are homogeneous versions. You either take independent random variables or you take identically distributed ones and independent.

That is why the coin toss example is a generic example. So, let us quickly prove this. Proof is now quite easy.

(Refer Slide Time: 15:43)



So, in the first case, what is the probability of Y i being small y i given the previous ones? So, remember, how we defined Y. That was a sum. So, given small y 1 to small y i minus 1 all that matters is what is the probability of this last random variable being this. It all matters. It, so, this is all that matters which is also what we did remember when we analyze this heads number of heads example. We looked at the difference.

So, that is all that matters and which is something equal to only the previous from Y i minus small y i minus 1 how do you go to small y i, so, which means that it is a Markov chain. So, that is done. Let us move on to the second one. So, in the second one, this probability, since we have assumed identical distribution, so, this will be equal to probability of in fact X 1 being this value. Why? Because, probability of X i being any value in fact.

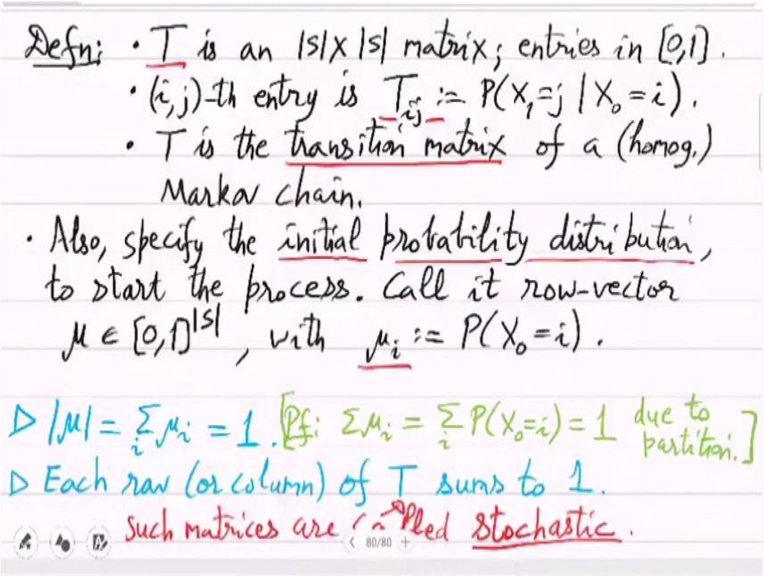
Probability of X i being value epsilon is the same as probability of X 1 being epsilon. This i does not matter, so, from that. So, this previous probability is actually equal to is the same as X 1 taking this which is equal to Y 1 being small y i and Y 0 being small y i minus 1 which means homogeneity. So, that was simple almost from the definition. So, we are studying homogeneous Markov chain for 2 reasons.

One is that many physical examples actually follow this and second is that it will be easy to represent and analyze. So, why? So, let me first give a alternate representation of it, so, representing in a figure. So, suppose you have a Markov chain where or homogeneous Markov chain where the state space has 3 entries A, B, C. And the transition probabilities like probability to go from value A to value C from X 0 to X 1.

So, then you can represent it like this. A, B, C are the elements in the state space. You go from A to B, B to A. These are all the ways in which you can in which the process can move A to B or B to A then A to C, C to A and B to C and C to B. And for example, this A to C transition, this happens with probability T AC and this C to B happens with probability T CB and B to C happens with T BC and so on.

So, these all these probabilities may be different. And this is how you can represent a homogeneous Markov chain. And given this picture, then you can see the evolution of the chain up to infinite time. So, with this picture in mind, it may not be easy to work with picture. So, we will instead work with matrices. So, alternatively in fact even equivalently we use matrices and let us define that.

(Refer Slide Time: 21:32)



So, T is an S cross S matrix. Entries in there will be a probability, so, 0, 1, this real interval. So, ijth entry is probability of going from value i to value j. So, T is the transition matrix of a homogeneous Markov chain. So, we have defined T. We have defined T ij transition matrix. And, how did the Markov chain start? That also you want to specify. Also, specify the initial distribution to start the process.

So, you can talk about a process only when there is a starting point. Only then can you talk about transition matrices or transition probabilities. So, that is the initial probability distribution. So, call it row vector mu. So, it is again these probabilities in the range 0 to 1. And they are as many as the state space the number of states. Moreover, you can see that the ith entry basically captures what is the probability of starting from value i.

That is the starting point. So, immediately you get certain properties from this definition. And this is the way homogenous Markov chains will be represented. So, first is that this if you sum up the coordinates of mu, then you get 1. Because somewhere you will be at least somewhere and these are disjoint events. Second is that each row or column of T sums to 1. So, what is the proof of this the first thing?

So, mu sigma mu i is the same as sum over the events X 0 equal to i. But these are this is a partition of the whole state space. So, this is this sum this probability is 1. That is a quick proof. And for row and columns, similar proof is there which we will see next time. And finally such a matrix T in general such matrices are called stochastic. So, the name comes actually from probability theory.

So, matrices which are non-negative real valued entries and the row and the sum columns are 1. They are called stochastic matrices. There they immediately define a Markov chain.