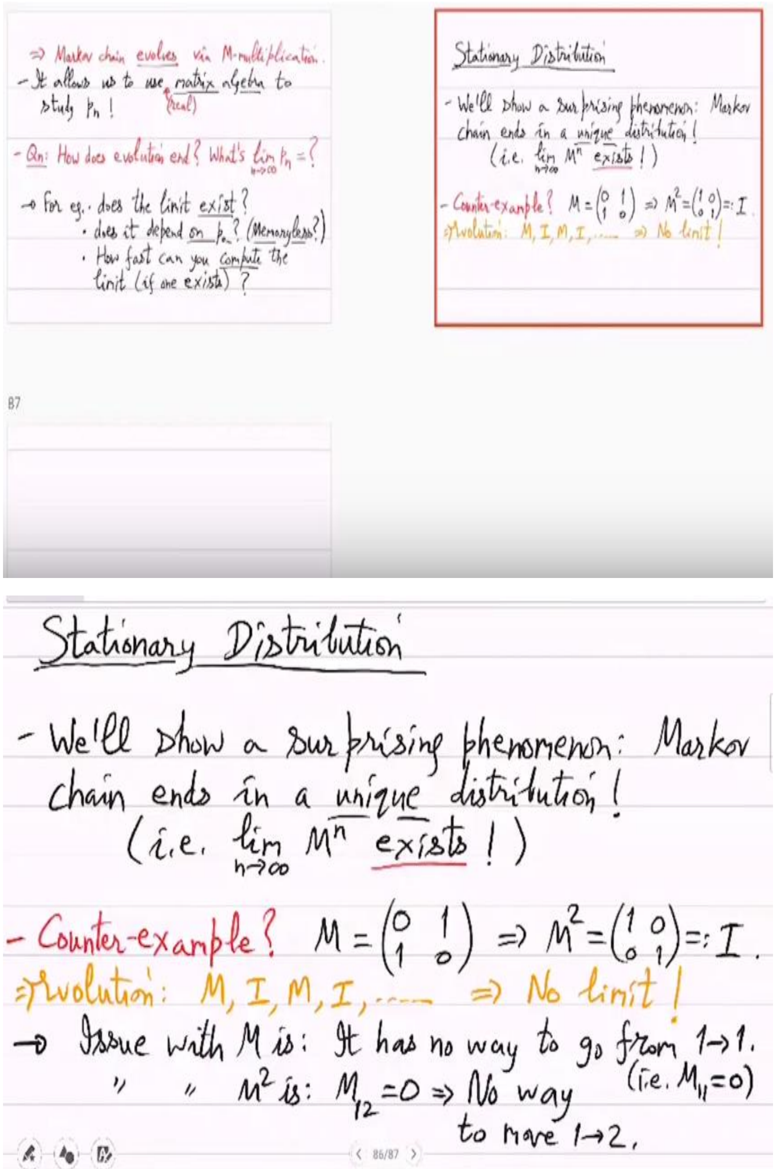
Probability for Computer Science
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology – Kanpur

Lecture - 20
Stationary Distribution

What do we do with these examples where the transition probabilities were 0 1 1 0?

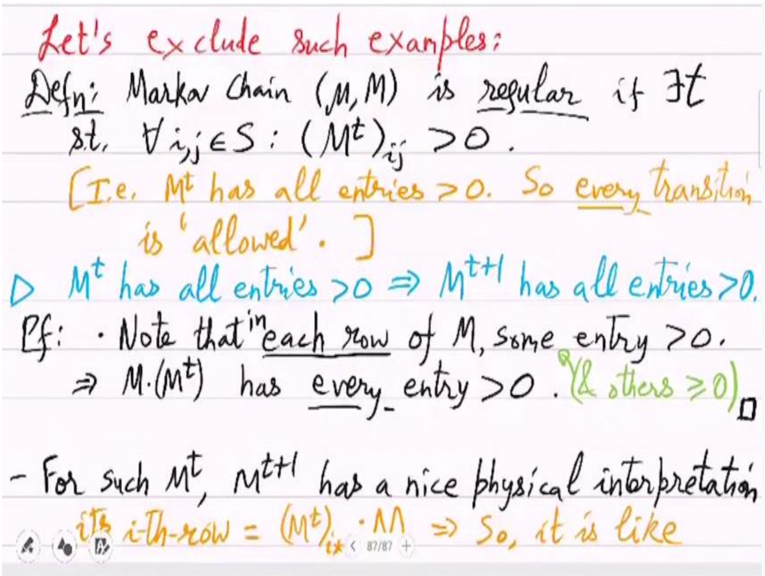
(Refer Slide Time: 00:23)



And then it keeps switching in the evolution. So, there seems to be no limit of the powers of this matrix. So, what is the problem here? So, issue with M is that it has no way to go from the state space element 1 state 1 to 2 itself. And similarly is from state 2 to 2. So, this probability is 0. This is what M 11 0 means. So, since there is it is not the case that from any i you can go to any j with positive probability.

So, because of that it is somehow not behaving in the most natural way. Similarly issue with M square is that M 21 is 0. So, 2 to 1 not possible. Note, this 1 2 is 0. So, no way to move 1 to 2. So, it gets actually the exploration in the evolution is not happening in the best way because of these problems because of these hurdles. So, the transition is not allowing you to go from every i to every j with some chance. It makes the chance actually 0 exactly 0. So, let us exclude these examples.

(Refer Slide Time: 02:33)



So, how do you exclude them? You look at processes where every i to j there is some chance of reaching. So, Markov chain given by mu and M is called regular. Regular or natural if there exists a time m. Let us call it just time. So, let there exist a time t at which point the transition probabilities are all positive such that for all i, j in S, M to the t ij is positive. That is M to the t has all entries positive. M to the t has only positive entries.

So, that is then a point after which the Markov chain will start behaving in a much better way. And you can understand the evolution better. Because there is a fair chance, well, it may not be fair but at least there is some chance that you can go from any i to any j. It is not that the evolution is excluding some option completely out. So, every transition is allowed. That is the good thing about regular Markov chain. After a point it becomes nice.

And let us then immediately prove a simple property of these. So, if M raised to t has all entries positive, then so does M raise to t plus 1. So, once you reach this point where all transitions are allowed then after that point of time that is always the case. How do you show this? So, first you note that each row of M has some entry positive. Why is that the case? Well, because the otherwise the row will be the 0 row and then the sum will not be 1.

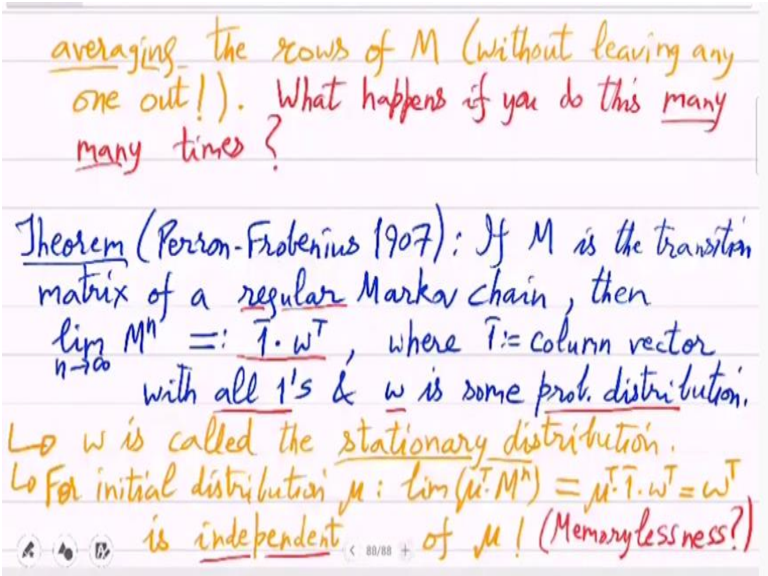
So, at least some entry is positive. It may be only 1 entry in which case it will be 1. So, each row in each row of M some entry is positive. So, now you look at M times M to the t. So, you are multiplying a row of M with columns of M to the t. Now, columns of M to the t every entry is positive. So, this has also every entry positive. Because you pick this and let us say look at the first row of M. Pick the position the entry which is positive.

And just multiply that with if you multiply it with any column corresponding location of M to the t, you will have a positive contribution. This here we are also using the fact that since negative entries are not there and others greater than equal to 0. That is also there. So, some entry in every row is positive and the others are non-negative. So, because of that that positive contribution will remain positive strictly positive. So, that is the proof.

So, this is why regularity is a very nice condition because once M to the t satisfies all positive entries it will remain true forever beyond that point. So, for regular M, M square has a nice physical interpretation which is that its ith row is the ith row of M times M. So, what is happening here is the ith row. Let me correct this. For such M to the t, M to the t plus 1 has a nice physical interpretation because if you look at the ith row then again it is just this.

Now, in fact, I have to do bit differently. So, now, M to the t ith row times M. So, M to the t has positive entry. So, this ith row is also positive. Every entry is positive. So, when you multiply with a column in M what is happening is you are taking an average in and you are not leaving anything out because M to the t has positive entries. So, the average actually gives some positive weight to every entry in the first column of M let us say.

(Refer Slide Time: 10:37)



So, it is like averaging the rows of M without leaving any out. So, every time you are multiplying by M, you are actually just keep taking the average. That is the intuition. So, what happens if you take this average infinitely for a long time? Again and again if you average out then you tend to get all things equal. So, you ask the question, what happens if you do this many times?

So, this is the guiding question and the guiding intuition which will lead you to a major theorem which describes the limit. So, this is a version of Perron-Frobenius theorem. It is very old matrix analysis. So, it says that if M is the transition matrix of a regular Markov chain then its limit of its powers is a matrix which is given by a product of the 1 column vector and some other row vector where 1 bar is all 1 column vector and w is some column vector.

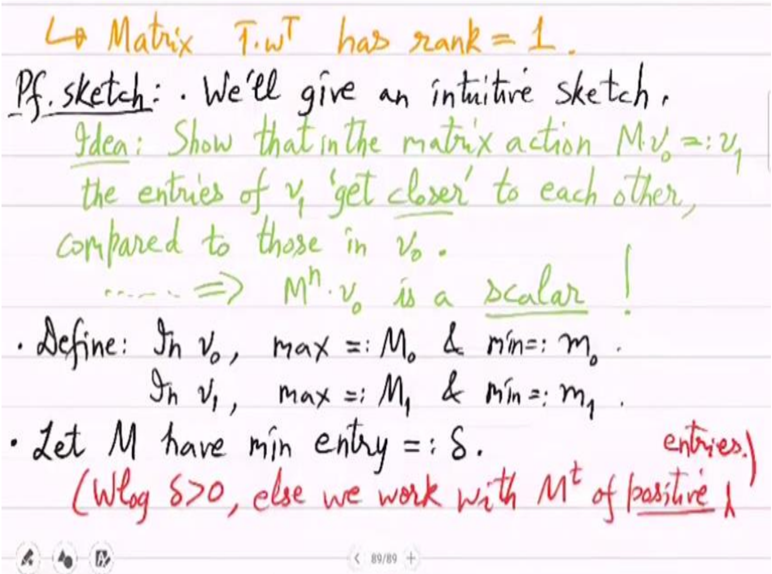
w is some probability distribution. So, we are actually this theorem is saying things but also defining things. So, assuming regularity it is saying that the limit actually exists. This is it is you are multiplying a column with a row. So, you will get a matrix square matrix which will have rank 1. So, 1 is basically it has all 1's and w is some probability distribution. So, let us interpret this.

So, this w is called the topic that we started studying stationary distribution. And this for any initial distribution mu what is now the limit of mu M to the n. So, that will be by the theorem it will be mu times 1 bar times w transpose now mu sorry this the row, so, mu transpose. So, mu transpose is a row multiplied with the column vector 1 but that you know is the value of mu which is 1. So, this is just w transpose.

So, what you see is that in the limit it does not matter what mu was. This is independent of mu. So, that is the amazing thing. That no matter what the starting point is the end point is the same if you wait long enough. So, this stationary distribution is very fundamental. It again tells you something about the memorylessness. So, this kind of evolution is actually memoryless.

It does not remember what it started from. It just remembers the transition probability. So, this w is only dependent on M not mu.

(Refer Slide Time: 16:43)



And another third thing is that this matrix 1 dot w transpose it has rank 1. So, in the end, you can start with any M of any rank stochastic matrix but ultimately the rank in the limit becomes 1. So, this is a very important and highly unexpected theorem. And this shows why Markov chains are structurally so important. So, how do you show this? What is the proof? So, we will take an intuitive route.

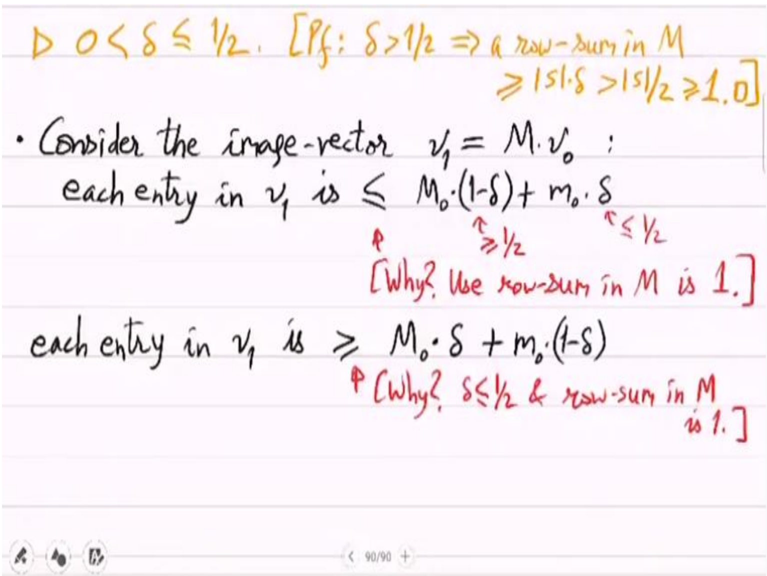
We will not go into the deep analysis of limit etcetera. We will just give you a more intuitive sketch so that you can read the proof somewhere else, formalize. Let us just first look at the idea of this. So, we will show that the action M on a vector. Let us call it. Let us call the result v 1. So, this matrix action the entries of v 1 get closer to each other compared to those in v 0. So, you start with the, I mean view matrix M as a linear transformation.

In other words, you multiply M with a column vector. You will get another column vector call it v 1. What we will show very soon is that whatever were the entries in v 0 what the difference between maximum and minimum that difference will shrink in this matrix action. So, in v 1 the difference will be smaller. So, if you continue doing this action infinitely many times ultimately the transformation M to the n will make all the entries equal.

So, v n will have equal entries. So, it will become a scalar. So, in the end, in the limit, it means that M to the n dot v 0 is a scalar. That is the bulk of the proof. And then based on this we will finish the theorem statement. So, let us attempt this. Let us implement this. So, define. So, inside v 0, max is let us say M 0 and min is small m 0. In the vector v 1, max is M 1, min is small m. And we are interested in the differences.

How has the difference changed? So, let the matrix m have minimum entry delta. Can delta be 0? Yes, it can be 0. But since we are assuming it to be a regular Markov chain, so, at some point M to the t will have only positive entries. So, let us assume that. So, without loss of generality, delta is positive, else we work with M to the t of positive entries. So, let us assume that delta is positive. That is the main entry in the matrix.

(Refer Slide Time: 22:38)



The other thing is here that delta cannot exceed half. Minimum cannot be too large. Why? Because after all this matrix M is stochastic, so, the row sum has to be 1 at all times. If delta is large then that will exceed. It will go beyond 1. If delta is greater than half then it would mean that sum row sum a row sum in M is actually more than S by I mean it is at least S times delta which is then more than S by 2. Now, state space is at least 2.

So, this is greater than 1. That is a contradiction. (()) (23:54) delta cannot be greater than half. It is between half and 0 but strictly positive. So, now, let us look at the entries in v 1. So, consider the image vector v 1 M times v 0. So, observe that each entry in v 1 is at most what. So, you have to look at entries that you get when you multiply a row of M with the vector v 0. So, the entries of the row let us say the first row of M it has entries at least delta.

And v 0 the entries are in the range small m 0 to big M 0. So, an upper bound is big M 0 times 1 minus delta plus small m 0 times delta. This is an upper bound. Why? Because this is greater than half and this is less than equal to half. So, the larger entries essentially the larger entries in v 0, they can contribute at most M 0 times 1 minus delta. And the smaller entries can contribute this much. So, this is the basic idea.

You have to prove this by using row sum to be 1 essentially. So, use that row sum in M is 1. So, since the row sum is 1 and you know the minimum is delta. So, the range is delta 2. I mean the delta part and then the 1 minus delta part. Now, the bigger part, the maximum contribution you can get from that is if you multiply it with M 0 and then the rest delta part is will come from m 0. So, that is an upper bound. What is a lower bound?

So, each entry in v 1 is at least now the reverse. So, combine M 0 with the smaller part and the smaller part with larger part. So, this will be the bound. This again you show by similar arguments. Delta is less than equal to half and row sum in M is 1. So, based on this, we will finish the proof next time.