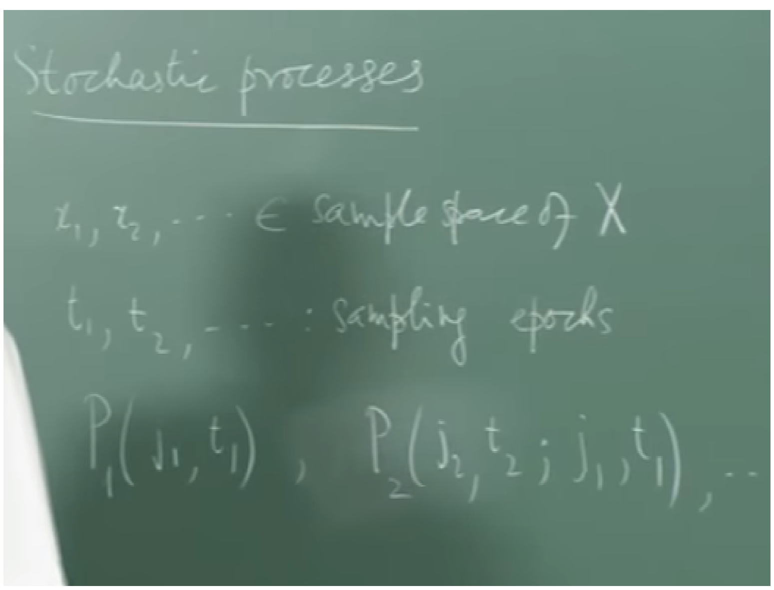
Physical Applications of Stochastic Processes
Prof. V. Balakrishnan
Department of Physics
Indian Institute of Technology-Madras

Lecture - 06
Stochastic Processes

Now, so far we talked about random variables both discrete and continuous random variables but we did not say anything about time dependence. We did not say anything about how these random variables could possibly change with time. From now on we will look at random variables which change with time, which evolve with time and then you have what is called a random process or a stochastic process.

So this is going to be our next topic which is concerned, this subject is concerned with the study of random variables with some rule for the evolution of certain probability distributions as a function of time okay. Now the first thing we have to appreciate is that a random process, if you sample this random process at discrete instance of time, you get a time series with values for the random variable drawn from the sample space of this random variable.

(Refer Slide Time: 01:38)



So if we for instance say that this random variable could have values x 1, x 2 dot, dot, dot, etc., then from this set of values if you sample this process at various instance of time say t 1, t 2 and so on; these are the sampling instance, instance of time or particular values of the time variable.

It is general technical term is epochs, sampling epochs and this is these are elements of the sample space of the random variable X okay.

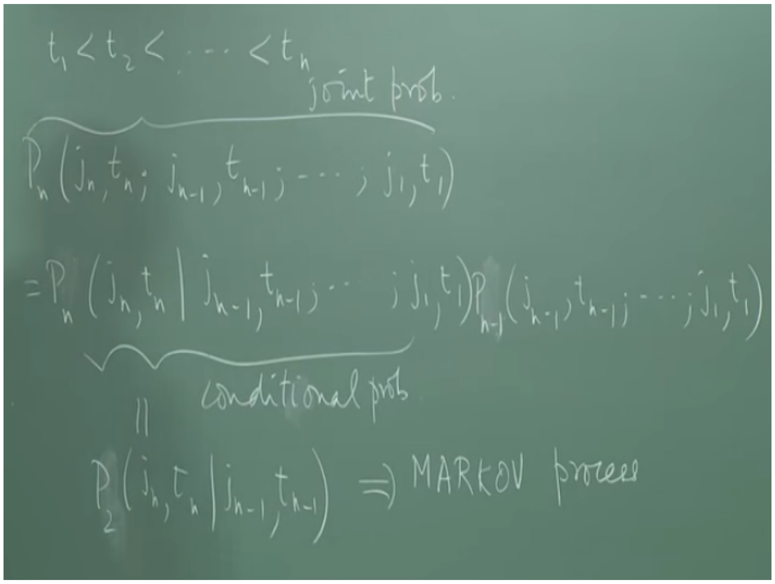
Then one ask for the probability that any of these variables values is attained at any given instant of time okay. Now we need, this is very cumbersome notation, so what I will do is to just take the index here and label this value here by that index okay and for this index I will call, use the symbol j, k, and so on and so forth. When I have too many of them I will call it j 1, j 2 etc., etc. So the question is what is the probability for a discrete random variable functions; what is the probability that at some instant of time t 1 the value happens to be some j 1 or x sub j 1.

So I will call that the one-time probability. Or if it is a continuous random variable then I use, I will interchangeably use this j 1 t 1, but I will be careful to indicate the fact that this thing is a continuous variable here. For the moment of course let us leave it at discrete and I have this. I could also ask what is the probability that you have the value x sub j 2 at time t 2 and the value j 1 x sub j 1 at time t 1. That is a different function. This is a joint probability.

It is a different function from this. There are two-time arguments here . To keep track of that, let me call this P 2 and let me call this P 1 and clearly this can go on. I look at the three-time probability, the four-time probability and so on. Now to specify this random variable completely, I need to tell you all these probability, joint probabilities. So the first thing we learn is that a stochastic process is described by an infinite hierarchy of probabilities or in the case of continuous variables, probability densities but it is an infinite hierarchy to start with.

Of course with this formidable problem there is not much one can do unless you start making certain simplifying assumptions. But there is one thing we can do which is not even an assumption and that is the following.

(Refer Slide Time: 04:47)



You can always take the n time probability. This n time probability can always be written as equal to the probability that we have j n, t n given that all these earlier things happen. So I am assuming here of course that t 1 less than t 2 less than dot, dot less than t n and I am writing the earliest times to the right and the latest times to the left; that is the standard notation and this n time probability can be written as the product of a conditional probability and a vertical bar will denote the probability of whatever is on this side given whatever is on the right hand side of the bar. That will be my notation. So we have j n - 1, t n - 1 all the way up to j 1 t 1.

This is again an n time argument probability, so it is still n. But this however is now a conditional probability. Whereas this one is just a joint probability multiplied by the probability that all these events have occurred that is j n - 1, t n minus 1 dot dot up to j 1 t 1 and that is a function of n minus 1 variable. So it is P n - 1 and in turn you can take this quantity and write it as a conditional probability of this last event occurring given that all the other events have occurred and so on.

So finally you can write it as a product of an n time conditional probability and n - 1 time, n - 2 time right up to the single time probability P 1 of j 1, t 1. So that is one simplification one can do right away immediately. But even that is not helpful because you still have this formidable task of specifying all these conditional probabilities if these things have happened okay and that is why the general theory is.

One can proceed further with this and so on but we are going to restrict ourselves to a very special instance, a very special kind of random process where the memory is a short-term memory in a very specific sense okay. Now this implies that the probability that this happens at time t n depends on all that happened earlier on earlier instance of time okay. But this is like saying you have a memory in this process.

Now experience tells us with random process of various kinds tells us that in nature very often if you use the right number of variables, if you take a complete set of variables in a very specific sense then it is short-term memory that occurs, never long-term memory. No history dependence in a certain specific sense. Just to give you an example, if you look at to give you a sort of trivial example, if you look at Newton's equation for a particle moving in space, this looks like a second order differential equation in time okay.

So not only to tell you what so if you want to plot a trajectory of a particle you have to know not only the position of the particle at a certain instance of time, but also the slope of the trajectory at that instant of time. This is like saying really to specify things completely, the fact that the force specifies the acceleration rather than the velocity, tells you that you need both the initial velocity and the initial position.

Which means that dynamics is really happening in a phased pace comprising the configuration space of coordinates as well as the velocity components or the momentum components right and once you put in in terms of those extra variables the full set of variables then the equations are motion of first order differential equations. So the initial state, any given, at any given instant of time will determine once you solve the equations of motion will determine the future state of the system right.

So that is an example where the dynamics is really first order in time so that the future is determined by the initial condition or the present and not on how you reach that present in exactly the same way as in quantum mechanics where the Schrodinger's equation is the first order differential equation in time for the state vector. So if you tell me the state vector at an initial instant of time and the Hamiltonian which gives you the rule of evolution, you can predict what the future state of the system is going to be in principle.

So this experience tells us that it may be worthwhile looking at those random processes or stochastic processes where this conditional end time probability is not dependent on the earlier variables other than the one immediately preceding here. So if this is equal to P n, now it is no longer P n but it is P j n, t n, j n - 1, t n - 1 and it is just a two-time probability; so it is P 2. If this is equal to this quantity here for all n so P 3, P 4, P 5 etc., it does not matter; every one of those things gets truncated to just this here.

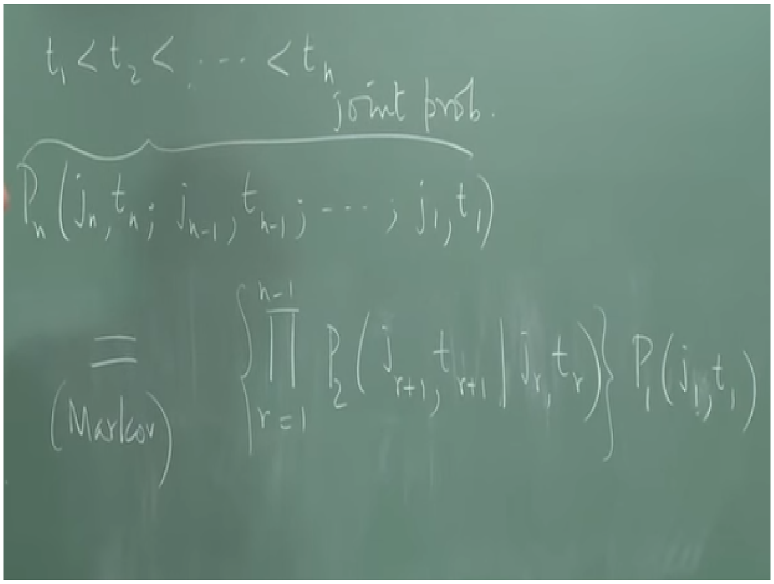
If that happens then it is called a Markov process. So again to repeat, a Markov process, it says nothing about the form of the probability distributions, it does not say anything about whether it is a Gaussian or whatever; those things come later. It says something about the level of memory in the process. Sometimes there are cases where you would like to have this dependent on the preceding 2 instance of time and then it is called a two-step Markov and so on.

But I am not going to get into that now. This is our straightforward definition of what a Markov process is, okay; does not always have to happen. But it turns out that if you model physical systems appropriately with the right number of variables, almost always you end up with a Markov process. Notable, there are notable exceptions. We will talk about a few of them. But the fact is that in most cases experience tells you how to model a random process and in general the most common one that you use always is a Markov process okay.

Now exactly as in the vector example I gave of a particle moving in space, it might so happen that the random variable is not a single random variable but a set of random variables, couple random variables. Then it would be a vector process of some kind maybe and then it is a Markov process still in terms of memory but there won't be a single index here but you need now several labels here for all the variables.

So that is a possibility we keep in mind okay and that is a matter of notation which we can sort out if the occasion arises but this is what I mean by a Markov process, this thing here. A similar thing for continuous processes, instead of probabilities the same thing is true for densities okay and then I will call it a conditional density in this case. But it is a two-time conditional density here. As soon as you have this, you immediately see that this joint probability simplifies enormously.

(Refer Slide Time: 13:15)



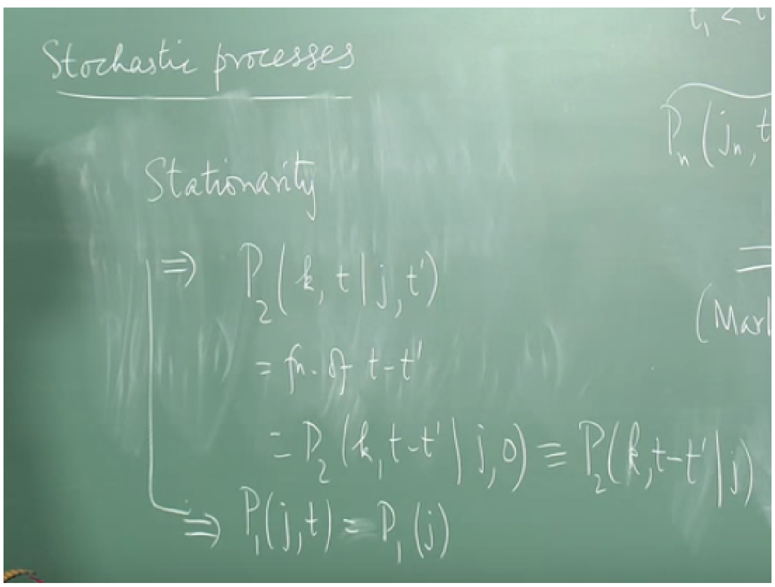
So if you make the Markov assumption, this becomes equal for a Markov process to a product of P 2 of j r + 1, t r + 1 given j r and t r and is a product from r = 1 to n - 1 out here so the last one is this guy here multiplied by a P 1 of j 1 t 1. So it at once simplifies okay into a product of two-time probabilities multiplied by a one-time probability P 1 okay. So the problem now reduces to specifying these 2 quantities and once you do that then we have all information we need for this infinite hierarchy of probabilities okay.

So it is a great simplifying assumption, the Markov assumption is a very it immediately changes the complexion of the whole problem and makes it a much more tractable problem to handle okay. As you will see this itself includes in it enormous amounts of complexity but it still makes the problem quite tractable. So we will focus on such cases here. We will look at many examples of Markov processes.

There is another further simplification that can happen and that has to do with the fact that the process that we are talking about may not change statistically speaking as time progresses. In other words it could be exactly the same process statistically no statistical properties change as a function of time. In other words the randomness is not ageing in some sense. There is no systematic drift or anything like that.

If that happens that would be the analog of an autonomous dynamical system where you do not have explicit time dependence in the way in the dynamical variables evolve in the dynamical rules. They will satisfy some kind of differential equations but then those differential equations do not explicitly involve the time okay. so the analog of that here would be a process where the origin of time does not matter and therefore this quantity here is a function only of the elapsed time t r plus 1 minus t sub r okay. And what would that imply and that is called a stationary random process.

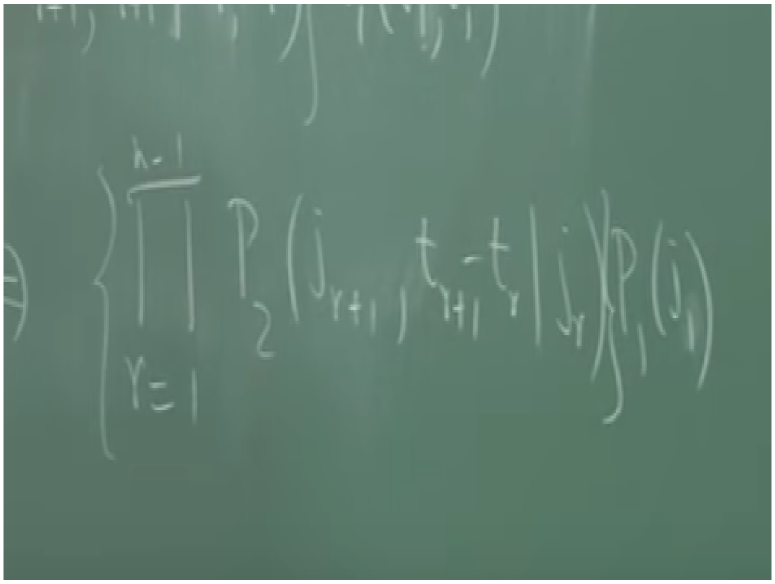
(Refer Slide Time: 16:02)



So stationarity implies statistical properties don't change with time at all. So it implies that P 2 of say k, t, give j at time t prime this quantity is a function of t minus t prime and not of t and t prime separately okay. So you could write this as equal to P 2 of k, t - t prime j, 0. In other words I can shift the origin of time and nothing happens. The probability do not change okay and very often I am going to make life easier and write this as P 2 of k, t, j where k and j are state labels or they stand for sample space elements.

I am going to use this kind of notation all the time. This is t - t prime j. I dropped the 0 here. It is understood that it is a function of difference of time arguments here. What would it imply also for this quantity P 1 of j, t. This should be independent of time. So all time dependents disappears in the one-time probability. So this is equal to P 1 of j okay. No t dependence at all and that together with the Markov assumption here.

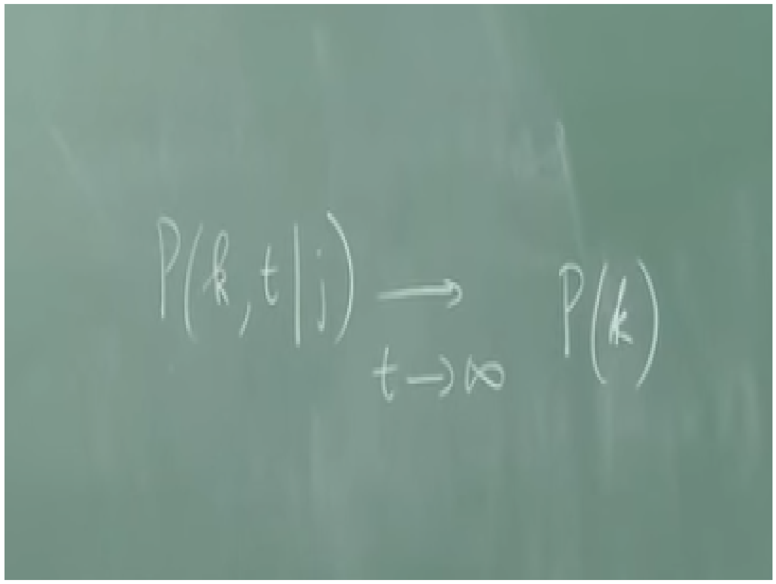
(Refer Slide Time: 18:09)



So for a stationary Markov process, for a stationary Markov process, this thing here implies this is equal to a product from r = 1 to n - 1 P 2 of j r + 1 t r + 1 - t r j r. So we now just have a two-time probability to handle, a one-time probability, time dependent probability, conditional probability to handle and an absolute probability here okay. So a stationary Markov process is completely defined if you tell me this quantity as a function of t minus t prime and this quantity out here okay.

Now all the models we talk about are going to specify these 2 quantities okay and if there is no confusion, once we reach that stage I will often drop this 1 and a 2. The moment there is a time argument and there are these arguments with this bar I know I am talking about a conditional density or probability and this for a probability itself in this case. You could put in one more bit of physical assumption or a physical input and that is the following although this is not absolutely essential. In general we won't need it.

(Refer Slide Time: 20:01)



But it will so turn out that you could ask what happens to this quantity k, t, j as t tends to infinity okay. Notice I've dropped this 2 here. It is supposed to be there but I just dropped it for convenience. What would you expect would happen to this quantity, this probability, conditional probability as t tends to infinity? Well you might expect, intuitively you might expect that this quantity should tend to something which depends on k but shouldn't depend on the initial condition j, initial state j.

As t becomes very long, memory is lost completely. So I would kind of expect in the same way I expect autocorrelations to die down and so on and so forth. I would expect that this tends to something which depends only on k n therefore it is just the probability k okay with a 1 here but this needs to be established. We need to make sure this really happens okay. On the other hand if the system has in the common example of some system in thermodynamic equilibrium for example I would expect the statistical properties are not changing.

Then if I choose a particular initial condition and ask what happens conditioned upon that initial state if some variable changes with time and I find some expression for the probability associated with it I could ask what happens if time elapses, a long time elapses and the system, nothing is happening to it statistically, I would expect it would this relation to hold good.

For instance if this was the velocity of a molecule and I start with a particular molecule whose velocity is some given number I specify and then I let it lose among all the other molecules and I ask what is the probability or probability density that it has a certain given velocity a long time after I started I would expect it to just attain the equilibrium density all over again okay. So I would expect it would tend to the Maxwellian distribution on this side independent of what initial velocity I started with okay.

Well, that is a physical expectation. If the system has enough junk in it and there are enough influences which are completely independent of each other randomizing the whole process then I would expect this to happen. In technical terms one says that if dynamical system has a sufficient degree or what is called mixing this will be true in general. So we will take a look at examples when this happens.

But remember that we have already assumed that it is a stationary process okay. If it is not stationary then of course this is even this is not true there is a time argument sitting here and it could well be that the initial state is remembered okay. So this poses an incredible amount of simplification once you have this. The moment you have a property like this, it means the entire process is determined completely by this one-time conditional probability because from that you get this the 0 time thing and you get all the other joint probabilities as well through this formula.

So a stationary Markov process with this property here of mixing actually is determined completely by determining this probability, this probability, conditional probability and then it reduces to a question of writing down equations for this probability in general okay. So the processes we will look at, a large number of them will fall into this category and we will write down specific equations for this quantity.

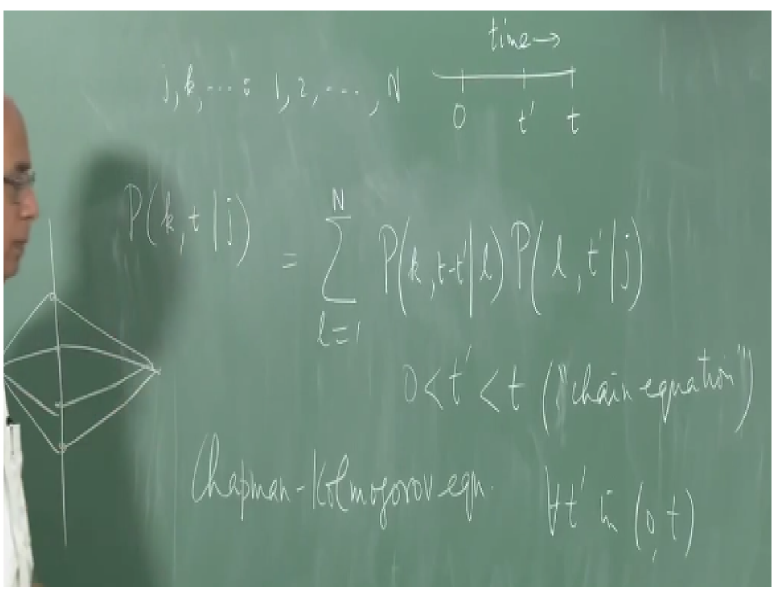
If you think a little bit you realize that any modeling that you do for physical systems of probabilities would be always to write down equations for conditional probabilities or probability densities. You need to know given something then what is the probability of something else happening and so on. You never say something about absolute probabilities itself. It is always conditional probabilities.

So conveniently for us joint probabilities reduced to conditional probabilities okay. So all we need to do is to model these conditional probabilities appropriately and then we are done okay. So it is important to distinguish between several assumptions here. First the Markov assumption has reduced things to one-step memory if you like and then the stationarity assumption reduces time arguments in this fashion here.

And it is important to remember that it does so for an arbitrary n. No matter how many time arguments you have out here this conditional probability depends only on the preceding instant of time okay. That instant is not specify this arbitrary, some earlier instant of time and that is it. That is all you need and then if it is true for every such earlier instant of time you have a Markov process okay.

So in a sense this process is kind of renewing itself at any instant of time it is forgotten the past and now it looks at what it does next in the future. So it is not surprising that there are going to be renewal equations and so on associated with this sort of process okay. For instance you could ask can I write down an equation for this p. And now let us use symbols like j, k, l etc., because we are not going to deal with these n time probabilities anymore but essentially just one-step memory.

(Refer Slide Time: 25:52)



So let us simplify notation and ask what is this likely to be k, t, j with 0 on this side okay. Now clearly, if it is a Markov process, which has this property of renewing itself all the time let us look at a case where j, k, etc., can take values 1, 2, up to some n. In other words, the sample space is discrete and you have capital N of these possible values. We could of course subsequently look at cases where n tends to infinity or becomes continuous and so on okay.

And this is equal to on this side the probability that you started with j and reached some intermediate state l at some intermediate time t prime. So on the time axis here is 0, and here is t prime, here is t and in the remaining time you move from l to k t minus so let us write it out properly. P of t prime, we started with j but reached an intermediate state l and then the probability that you went from that l in the remaining time t minus t prime to the state k.

But you could have done so through a variety of paths, all kinds of intermediate states l would have been allowed. So you have here a summation l equal to 1 to N in this fashion okay. So for a stationary Markov process, this tells you because it is not dependent on any earlier instance the memory is a one-step memory it says to go the probability of going from an initial state j to a final state k in time t is the probability of going from j to l at some till some intermediate time t prime and then in the remaining time going from l to the final state k, this desired state k okay.

And you must sum over all the intermediate possibilities (()) (28:30) and that is the summation over l out there okay. This is like a chain equation. It is got a technical name. It is called the Chapman–Kolmogorov equation. It should really be called the Chapman–Kolmogorov, Bachelier, Smoluchowski equation etc.; several people were associated with this equation. But it is popularly called the Chapman–Kolmogorov equation in this case okay.

Now if these were continuous random variables then you would have to integrate over this state, the intermediate state l rather than sum over it but that is a matter of notation in this case okay. What do you, what is the first thing that strikes you about this equation? Well, first let me say that this is not restricted to Markov processes. There are other processes which also obey the chain equation but Markov processes obey it. So it is not uniquely a property of Markov processes.

“Professor - student conversation starts” Ya. Pardon me. You are fixing t prime here. We are not fixing t prime. So this is true for any t prime in 0, t. I think from each n the t prime that we choose is the same right, when we make a sum. Yes, yes of course. Yes, certainly. You must sum over all intermediate states at some intermediate instant of time. **“Professor - student conversation ends”**.

So if you draw a picture, here is the initial state, here is the final state. Here are all the possible intermediate states. We are going from propagating from here to there, here to here in this fashion and there is a time slice here at this point at time t prime. So you are summing over all those possibilities and adding the probabilities appropriately to get this right here okay. So what is it that strikes you about this equation immediately?

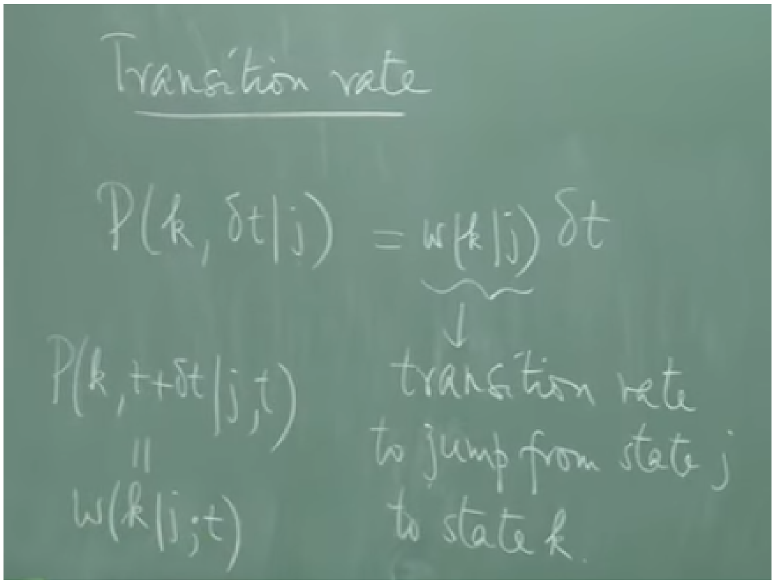
As a mathematical equation, this is not so tractable as it looks because it is a non-linear equation. This equation here is not linear in this P okay and therefore it is a fairly complicated equation. It is not immediately obvious what the solution will be okay.

“Professor - student conversation starts” Yes. In the first problem D, is it t + t prime or t - t prime? Well, the time interval left here is t - t prime. So that is all the time available for the system to go from the intermediate state to the final state. So it is this interval multiplied by that interval. Also this equation hold for one stationary processes but they need not be Markov, is that right. **“Professor - student conversation ends”**.

Well this chain equation yes. They are stationary processes, but there is a wider class of processes called renewal processes for which this equation would also hold good. It is called. It is an example of what is called a renewal equation right. But we are concerned here with Markov processes okay. So I am not going to get into the technicality of looking at processes other than that. If time permits we will talk about such renewal processes later on.

When we do Poisson processes and so on then I will mention what happens if you look at a more general case here. So this nonlinearity makes it intractable in some sense and if it is a continuous variable then for the probability densities you have an integral equation because there is an integral on the right hand side which is nonlinear and therefore fairly hard to solve. It would be convenient to write this in terms of a linear equation for this P. For this purpose one introduces the following idea. Does not always work, but when it does this is what happens.

(Refer Slide Time: 33:03)



So one introduces the idea of a transition rate and the idea is the following. Consider this probability here for extremely small values of t, very close to 0 or this probability for extremely small values of t minus t prime close to 0. So if you look at P of k delta t, j over here, this is state j at time 0 and this is state k at an infinitesimal time delta t. What would you expect this to be proportional to?

If delta t goes to 0 I would expect that it is going to remain at the initial state. I would expect a delta function there right. But if delta t is infinitesimal then I would expect that this quantity for all k not equal to j, for all k not equal to j this must be of the form some delta t multiplied by w k, j where this quantity is a transition probability per unit time that the system jumps from the state j to the state k okay.

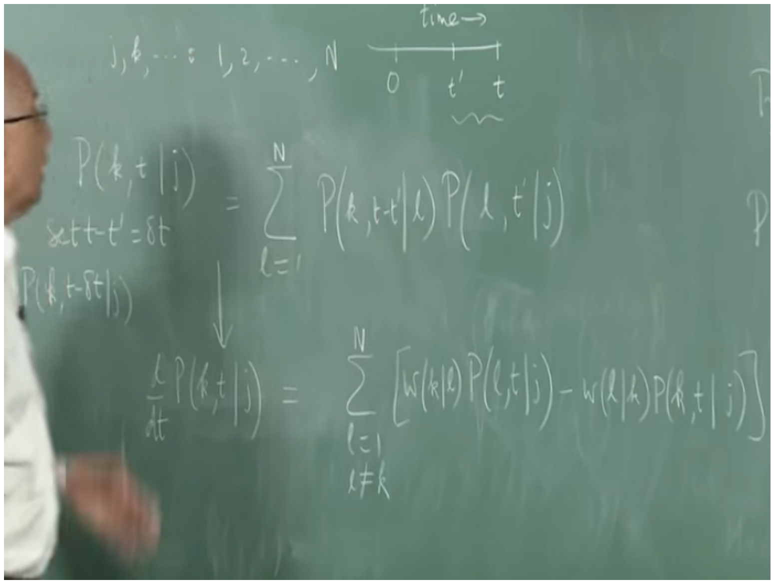
I would expect the answer to be proportional to delta t and a constant of proportionality is a per unit time. This is a probability. So this must have dimensions 1 over time okay and this is a transition probability or rate to jump. No guarantee that this exists. No guarantee at all this exists okay. But if it does then it has the physical connotation of a transition rate because when you multiply it by the time interval delta t you get the actual probability, conditional probability okay.

The same thing could well be true for even a non-stationary process. What would happen in that case if I had a t + delta t here? So if I have a non-stationary process of the form k, t + delta t, j at time t you could still assume that if delta t is sufficiently small and k is not equal to j, this should be proportional to delta t multiplied by a transition probability but that transition rate would depend on time right.

So the generalization of this idea of a transition rate to a non-stationary process is fairly straightforward. This would again become equal to w of k delta t k, t well k, j and then a t here to show that the transition rate itself could change as a function of time because the statistical properties is changing with time. So the great advantage of having made the stationarity assumption is that the transition rates are independent of time okay.

So this is a very physical thing that we are talking about. If I make that assumption, then what is the next step? What is going to happen here? Well the obvious thing to do is to say let us make t - t prime delta t and then for this quantity put that in, put that expression in and there would be answers, there would be things proportional to delta t. The obvious thing to do is to subtract from this k of t - delta t at time j from both sides and then divide out through delta t and convert it to a differential equation. So this is what one would do immediately right.

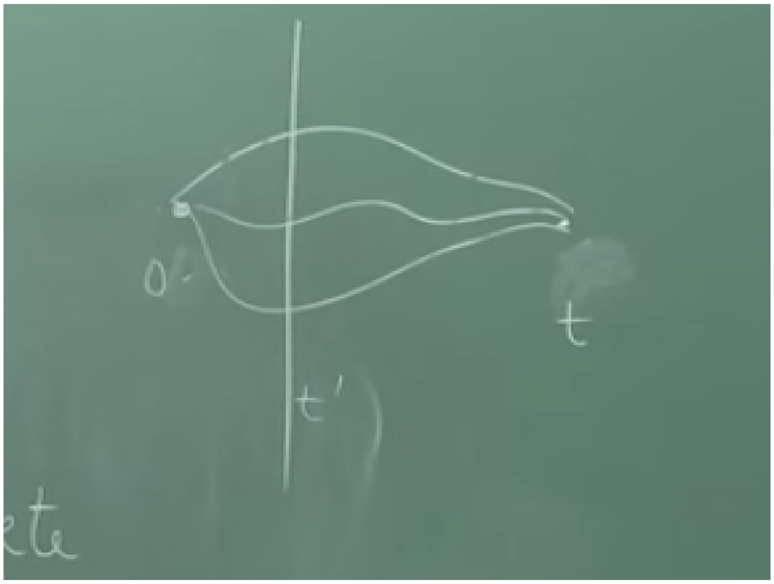
(Refer Slide Time: 37:29)



So I leave that to you as an exercise and it is not hard to show that with this assumption this equation translates to d over dt of P of k, t, j becomes equal to summation l equal to 1 to N and now we got to be a little careful . P of l, t, j; w of k, l; l not equal to k this side minus because you subtracted this quantity you end up with a minus sign here and now let us look at this equation carefully.

So the trick is to subtract from this both sides of this equation, subtract the following quantity minus P first set, set t - t prime equal to delta t and subtract P of k, t - delta t which is t prime by the way from both sides and put that in and maneuver. **"Professor - student conversation starts"** Sir. Ya. When we are considering the product of probability, yes, Chapman–Kolmogorov equation, why are we not considering all possible times? Ah, it is not necessary. Any time will be true. Okay. **"Professor - student conversation ends"**.

(Refer Slide Time: 39:18)



So look at it physically like the picture I drew. You want to start at t = 0 at this point and at time t you want to reach this point at time t you are starting in this state ending in this state and you have many routes to go through with different probabilities and now the statement is the probability to go from here to there, the total probability is the sum of all these individual probabilities such that you go from here to here at some time t prime and then you traverse the rest of the way and it does not matter where you take the time slice okay.

These quantities are mutually exclusive. They are different intermediate states which is the reason you sum over it okay. So when you sum over these probabilities, what is the meaning of the word and it means if you have several possibilities, you sum over their probabilities; this and this and this and this. If you have or then of course it is a different story. Sorry, if it is and you multiply the probabilities which is what I have done but if you have or you sum over them and that is what I have done because they are mutually exclusive.

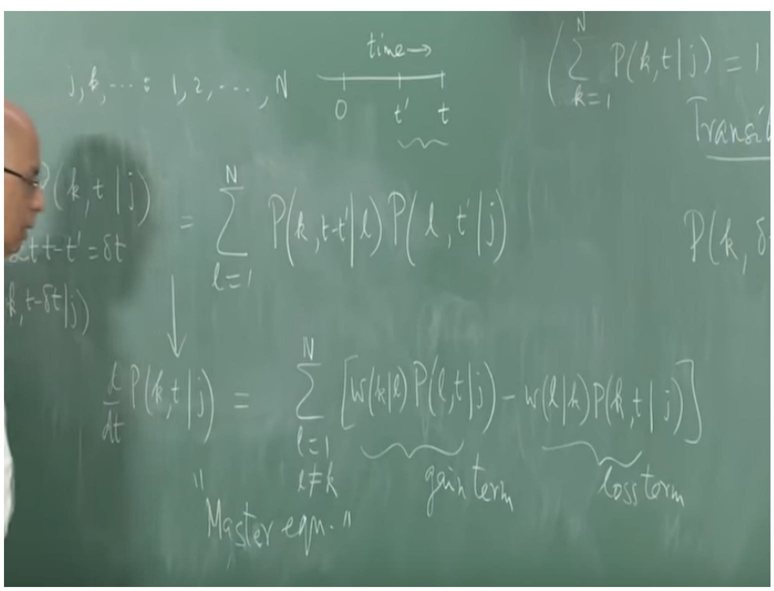
This is different from this is different from this. So it is only at the same instant of time that these are all mutually exclusive possibilities okay. So it is worth pointing this out. It is not an equation in time. It is not an integral in time. There are such renewal equations. We will talk about them subsequently. But this is a summation over intermediate states here at any given instant of time in between okay and therefore I can choose that interval, intermediate time as I please. I choose this to be infinite decimal.

“Professor - student conversation starts” Shouldn’t it be a double interval, double summation. No, no that will be over counting, that will be over counting. This is the physical way to look at it. This is over counting because these parts could intersect and so on. So you definitely have to do this at one instant of time. So you add over mutually excluded events okay. **“Professor - student conversation ends”**.

Now let us look at this equation a little bit. So this derivation is something I am going to leave to you, its straightforward enough. But what is the interpretation of this equation. It says the conditional probability to go from j to k, the rate of change of this probability has 2 contributions. One is a gain term out here where you go from j to l an intermediate state multiplied by the probability per unit time that you go from l to the final state desired k.

That is the gain term and this is the lost term exactly like in the rate equation because you have gone from j to k the state that you want to but then you jump out of it with this transition rate with this probability here okay. So the input into this is first you do this then you subtract this and then you use conservation of probability.

(Refer Slide Time: 42:34)



You use the fact that if you start with P k, t, j and you sum over all k out here, all k now from 1 to N, what should you get? You should get 1 because you start with a state and the system has not disappeared, it is in one of the states available to it including the initial state itself, so when you include that the sum should be equal to 1 for all t okay. This is equal to 1 for all t greater than equal to 0. That is input, that is put in.

You need to put that 1 in and that is how you get this minus term appropriately okay. So the interpretation is quite clear. The rate of change of this probability this increases when you have gain and it depletes when you have a loss and this is the precise equation for it okay. This is called the master equation. This word is used in many context, but this is the most common context okay. Now what is the great advantage of this master equation?

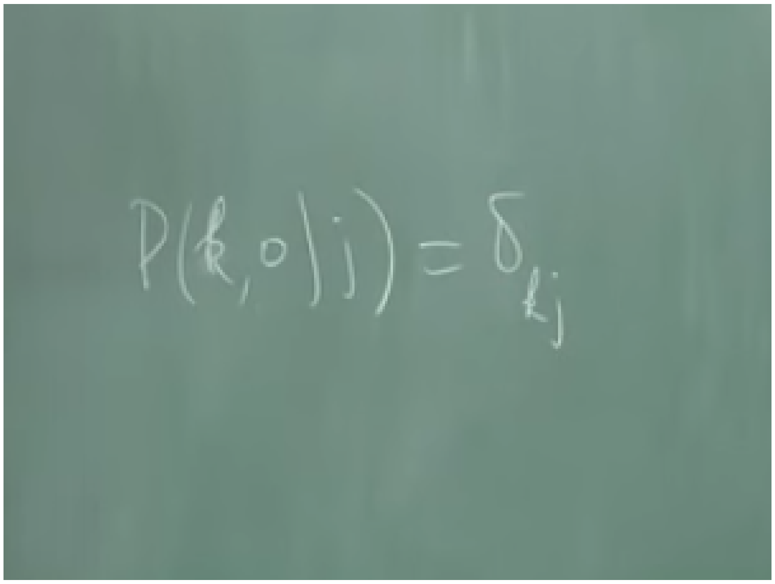
It is a linear equation. The price you pay for it of course is that it becomes a differential equation here in time, a first order differential equation okay. But it is a linear equation. The matter is not so simple even now because in general if it is a continuous variable then these would be probability, conditional probability densities and this would be an integral. So then you have a integrodifferential equation, linear but an integrodifferential equation and that is not so simple to solve either okay. In fact we are going to look at that.

What will happen in that case is that this side will get converted, there had been an integral here. We can get rid of that integral but we will get it converted to a partial differential equation in the variable itself but it will unfortunately be an infinite order partial differential equation in general okay at least formally and then we look at further cases, sub cases etc. But at the moment we are talking about discrete variables with discrete sample spaces.

Then this is what you have as the master equation okay. Now when you do chemical reactions you write down rate equations for the concentrations of various species. You have precisely the same sort of equation, set of equations. You have things which are gain terms and loss terms of this kind. So this is often called a rate equation or something like that but in this context these are equations for the conditional probability, probability itself okay. So the next task is to solve this.

By the way what is the initial condition? It is a first order differential equation and time, so we need an initial condition to solve it.

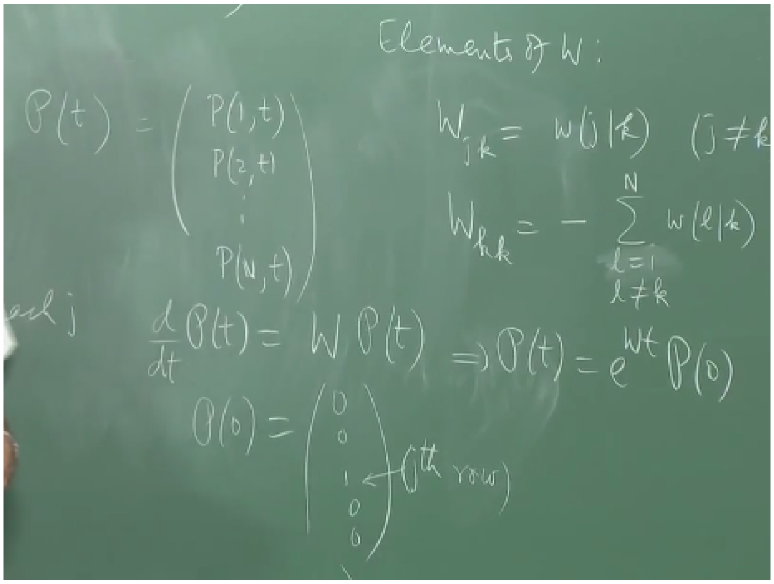
(Refer Slide Time: 45:48)



And of course P of k, 0, j = delta k j. Given that you are starting in the state j at t = 0, of course at t = 0 this becomes delta k j okay. Now in a slightly more general context you could look at this; j is sitting here as a dummy variable, as a sort of spectator throughout. You could write such an equation for the probabilities themselves without putting this j in and then specify an initial distribution of j's.

Then the initial condition would not be a delta function but some appropriate distribution. We will look at those cases as well. But this is the task one has to now attack, this quantity here. Now let us see what we can do about this. The first thing to do is to notice that if j and k run from 1 to N, these indices run from 1 to N then this equation has the following structure.

(Refer Slide Time: 46:57)



Let me write P of 1, t, j ; well let me let me write this, let us write this as a column vector P of t given j so let me suppress this j index for a moment because it is a spectator sitting out here. For every j this is true, for each and every j. You have such a master equation. So let me suppress that for a moment and write this P(t) to be a column vector which is P(1, t), P(2, t) up to P(N, t) with j's, understood, on the right hand side after the bar.

If I have defined a column vector of that kind then this equation here takes the form d over dt P(t) equal to some W P(t). W is a matrix of some kind okay and what are the elements of W? W j k is just w j k. This is for j not equal to k. On the other hand the diagonal elements of this matrix, remember here it is l that is getting summed over. So in that sense this term comes out into the summation and this fellow multiplies the sum over l okay.

So it is immediately clear the moments thought that W kk = minus the sum of all the other elements in that column okay. So this is equal to minus summation W l k l not equal to l equal to 1 to N, l not equal to k. So you can rewrite this this set of linear equations in the form of a matrix equation with a certain column vector P which determines all the probabilities that you want, conditional probabilities.

And then multiplied by on the right hand side you have this square matrix N by N matrix acting on this column vector where this matrix has off diagonal elements which are all the transition probabilities and the diagonal elements are minus the sum of the rest of the elements. That is a very special kind of matrix because it says the sum of the elements of every column of this matrix is 0 okay.

Now what does that tell us immediately about the eigenvalues of this matrix? Well the determinant is 0 because the sum of each column is 0 so the determinant is 0 right? The moment the determinant is 0, you know that 0 is a eigenvalue of this matrix right. So this means that this equation in general, this equation would have an eigenvector. You expect it to have a nontrivial eigenvector such that W on P is 0 which would imply that d over dt of that P is 0 which would imply that this is a stationary distribution.

It does not depend on time at all okay. So this is buried in it, this whole thing is buried in it and we will see what happens. Of course there are other eigenvalues as well. What is the formal solution to this equation? I have an equation of this kind, what is the formal solution. Well, it depends on the initial condition right? Now what the initial condition be? We know that at $t = 0$ this quantity here at $t = 0$ is a delta k j.

I have written this equation here. This is the k index and I have suppressed the j index. So at $t = 0$, what is $P(0)$? It is going to have 0's everywhere except at the j th element where you would have 1. So you got to solve this equation with the initial condition that $P(0) = 0, 0$ etc., till you hit a 1 and then 0's again and this will be the j th, it will be in the j th row okay. Now given that initial condition, what is the formal solution to this equation; the exponential.

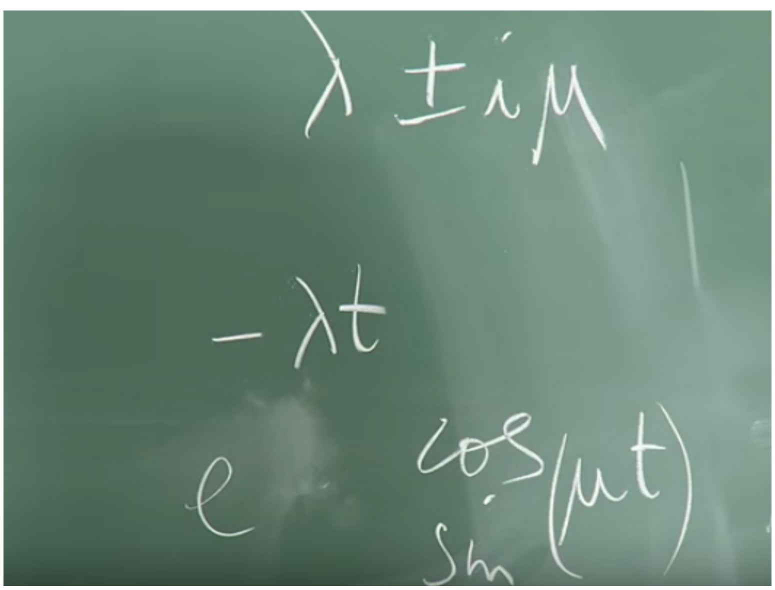
Because this W is independent of time and what is the physical assumption that made W independent of time, stationarity, stationarity. We assumed it was a stationary process. Otherwise is not true okay. You still have the formidable task of exponentiating this matrix. But we know in principle what is going to happen. If this matrix has eigenvalues lambda 1, lambda 2 to lambda n then in general generically barring repeated eigenvalues and so on we are going to have terms on the right hand side which go like e to the lambda 1 t, e to the lambda 2 t and so on.

So they are going to be exponentials of the eigenvalue multiplied by time. **"Professor - student conversation starts"** So this implies that the eigenvalues cannot be real and positive because if they are the probabilities keep on multiplying. Yes, absolutely. **"Professor - student conversation ends"**. Absolutely. This immediately tells you we know nothing about this matrix. At the moment we know nothing about it.

What we know is the following. We know that these elements, these fellows are all positive or maybe 0. There could be some states where there is no transition directly possible from k to l. So this could be 0 right, but certainly not negative. So we have a matrix whose elements are all real. All the off diagonal elements are either positive or 0. No negative elements and all the diagonal elements are negative because they are minus some positive numbers okay and the matrix is real, not necessarily symmetric.

Because there is nothing that says w j k must be w j k nothing at all. So given that we still see from this physically we would be very surprised if you got an eigenvalue which has got a positive real part because immediately it would imply that this probability is growing unboundedly with time. So you need to be sure that the eigenvalues cannot have positive real parts. They could be complex but there are current complex conjugate pairs.

(Refer Slide Time: 55:02)

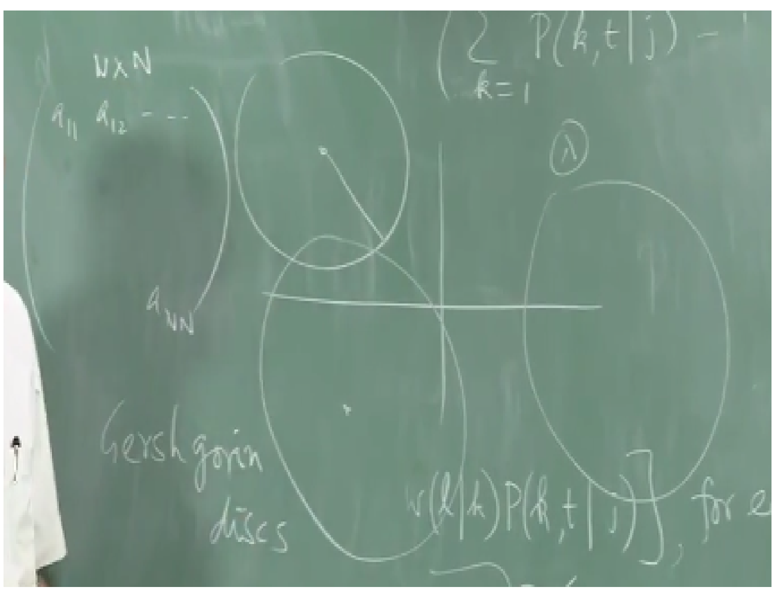


And once they do that what it would mean is if you have an eigenvalue of the form lambda plus or minus i mu then this would go like e to the lambda t cos or sin mu t. That is what the solution would look like but we must be sure that this lambda is in fact negative. So we would expect something like this e to the power minus lambda t where this is positive and possibly oscillatory behaviour etc.

So this is what we should make sure we have and we should expect. Now we would expect that as t becomes infinite, I would expect the t dependence have to disappear and thinks to go to where. Well, once I say that all the eigenvalues have negative real parts, all these fellows go to 0 but we know that 0 has to be an eigenvalue of this matrix. Therefore there'd be some constant which is sitting there and the P(t) will tend to that constant which will be the stationary probability right.

Now this is sort of formalized by a little theorem in matrix analysis called Gershgorin's theorem. I am not sure if you have heard of this but let me explain what this theorem is because it is simple enough.

(Refer Slide Time: 56:17)



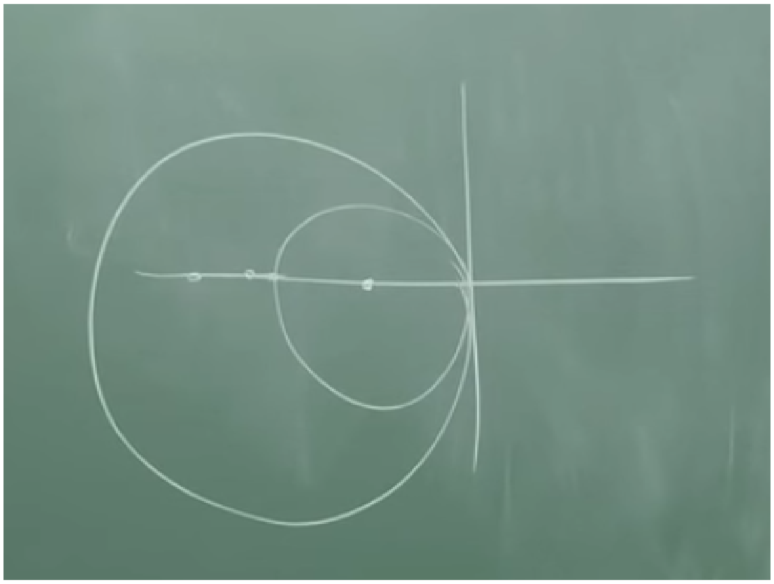
It says if you have a square matrix with various elements an N by N matrix then let us suppose the general element of this matrix is a 11, a 12 bla bla bla etc., a NN then it says the eigenvalues of this matrix, whatever be this matrix in the complex plane because eigenvalues are in general complex are located in certain circles or discs and these discs are found as follows. Take a 11 and mark it on the complex plane.

It could in general be a complex matrix with complex entries we do not care is sitting somewhere here and then you take the rest of these elements here at their moduli together and that gives you a positive number right and that positive number draw a circle of that radius about this point okay. So draw a circle whether you choose rows or columns it does not matter because the eigenvalues of a matrix are unchanged if you change the matrix to its transpose.

So the radius here would be the sums of these moduli. Similarly, take the next row take a 22, that is somewhere here and draw a similar circle etc. These things are called Gershgorin discs and the statement is all the eigenvalues will lie either in or around these circles, that is all and it is a very simple theorem to prove. You can prove it by elementary means okay. Now these are discs, could be disjoint. There could another disc here which is disjoint.

There could be things which overlap we don't care. What we do know, there is an extra theorem which says that if any of these discs is disjoint then you are guaranteed to have at least 1 eigenvalue there in the disc and this is a completely general theorem. It does not say anything about the nature of the matrix. It does not assume whether it is real elements, complex elements etc., we do not care, still true. Now if you apply this to this w what is going to happen?

(Refer Slide Time: 58:54)



In the case of w we know that all the diagonal elements are negative real numbers. So they are all sitting here or here or here etc. and in each case the elements add up. The rest of the elements are minus whatever was the diagonal element right so the radius is just this distance and this fellow here has a thing like this etc. and all the eigenvalues are in the intersection of these discs which means no eigenvalue can have a positive real part immediately.

And all the eigenvalues other than 0 will have negative real parts and therefore the system will the probabilities relax towards the equilibrium distribution. So w is called the relaxation matrix in the physical literature. So I stop here now and we take it from this point.