Discrete-time Markov Chains and Poission Processes
Professor Subhamay Saha
Department of Mathematics
Indian Institute of Technology, Guwahati
Lecture 25
Properties of Exponential Distribution

Hello, everyone, welcome to the 25th lecture of the course Discrete-time Markov Chains and Poisson Processes. So in the last lecture, we finished the Discrete Time Markov Chains part of the course. So today, we will move on to the Poisson processes part. But actually, we will not start with Poisson processes today, before starting Poisson processes, we need to learn something more about exponential distributions. That is what we will be looking at in today's lecture.

So, you already know what is an exponential random variable. So, we will just recall that. So, a random variable X is said to have exponential distribution with a parameter $\lambda > 0$ if the probability density function of X is given by this quantity. So, it is $\lambda e^{-\lambda x}$ if $x > 0$, and it is 0 if $\lambda \leq 0$. So, random variable X is said to have exponential distribution with parameter $\lambda > 0$ if it is a continuous random variable with probability distribution defined in this way. So, again, this kind of thing, so, you already know what is an exponential distribution from your basic knowledge of probability, so, it is just a recall. And what is the cumulative distribution function of X, so, it is given by this. So, you know that if it is a continuous random variable then its distribution function is given by $\int_{-\infty}^{x} f_X(x)dx$. So, you know, $f_X(x)$ is this, so, if you just do the integration, you can easily check that the cumulative distribution function of an exponential random variable is given by this. So, it is $1 - e^{-\lambda x}$ if $x > 0$, and zero otherwise. So, this is a random variable that takes only positive real values, so, it does not take nonnegative real or nonpositive real values. So, that is an exponential distribution or a random variable having exponential distribution and we will use this notation. So, $x\ Exp(\lambda)$. So, we will use this notation to mean that X is a random variable with exponential distribution having parameter $\lambda > 0$. So, that is the exponential distribution. Like you also know many properties of an exponential random variable. Like for example, if you calculate you can easily check what is the expectation of X. So, first of all expectation exists and it is equal to $\frac{1}{\lambda}$. Similarly, you can calculate what is the variance of X and so on. So, that is the exponential distribution or exponential random variable.

Now, these are some plots of PDFs and CDFs of Exp lambda for several values of λ. So, you see, we have plotted here for three different values of λ, $\lambda = 2$, $\lambda = 1$ and $\lambda = 0.5$. So, you see the smaller the λ for PDF, it decays much slower or the higher the λ the decay is much faster. And similarly, if you look at the cumulative distribution function, the bigger the λ the higher it goes towards 1. So, you see the blue curve goes towards 1 at a much faster rate or it goes much, much quicker to 1 as compared to the green or red curve. And similarly, here, you see the blue curve decays much faster as compared to the green or red curve. So again, these are just some observations from the PDFs and CDFs of $Exp(\lambda)$ for several values of λ. Again, you must have seen all these things in your basic course on probability. So again, as I said, this part is just simply a recall.

Now, we will start with some properties of the exponential distribution, which we will be needing for Poisson processes. So, the first property is what is called the memoryless property so we state as a theorem. So, what is the theorem? Suppose X has an exponential distribution with parameter λ. So, everywhere we will not say $\lambda > 0$, but whenever we say that X has an exponential distribution with parameter λ, we mean that $\lambda > 0$. Then for any $t, s \geq 0$ if you look at this quantity $\mathbb{P}(X > t + s | X > t)$. That means, what is this quantity? So, you are told that $X > t$ and you are being asked what is the probability that $X > t + s$. So, you have been given, so, it is a conditional probability. You are given the information that the value of $X > t$ and you are being asked okay what is the chance or what is the conditional probability given this information that $X > t + s$. Now, what this property is telling you or what this theorem is telling you is that, that is just simply probability $X > s$. Now, this is what is called the memoryless property, why? Now, what is this saying? It is saying that this information that $X > t$, so, you know that okay $X > t$ and you are asking what is the probability it will be greater than $t + s$. What this is saying is that this information that it is greater than t is actually not like it is being used in the sense that this is the same as just probability $X > s$.

Now, why am I calling this memoryless property? Now, let me explain it by giving an example. Now, think of suppose lifetime of some electronic gadget, say lifetime of a bulb, suppose, it follows an exponential distribution. Suppose the lifetime of an electric bulb follows an exponential distribution with parameter λ then what this probability $X > t + s$ given $X > t$ means, that means, you are told that the bulb has already lasted or has already survived for t units of time and what is the chance that it will survive and s additional units of time. So, you are told that okay, it has already survived four t units of time and you are asking what is the probability that it will survive for $t + s$ units of time or in other words, you are asking what is the probability that it will survive for additional s units of time. Now, what this theorem is telling you that is same as the probability that the bulb will just survive s units of time. So, this information that it has already survived for t units of time is not being used or is not useful. So, when you are looking at above, if you know that the bulb has an, again the lifetime of the bulb follows an exponential distribution, then suppose you just come to inspect a bulb, you do not need to know how long it has been working, in order to answer the question that how much further it will work. So, that is what this property is saying. That is why it is called memoryless because it does not have any memory. It does not keep in memory that it has already survived for t units of time. So, when you are asking, okay, what is the probability that it will survive for s additional units of time, it is as if, it is assuming that the bulb, so it is new bulb, it is forgetting that okay, it has worked for t units of time. That is why it is called memoryless property. So, that is what this probability $X > t + s$ says, given $X > t$ equal to probability $X > s$, if you have to think of it in a real-life situation, this is what it is saying.

Further exponential distribution is the only continuous distribution or it is the only continuous random variable with this property. So, it is kind of an again if and only if statement. So, if you, if x is an exponential random variable with parameter λ, then it satisfies this property and if you are or it is what is called this memoryless property. Similarly, if you are told that okay, X is a continuous random variable having this property then straightaway you can say it is an exponential random variable with some parameter $\lambda > 0$. So if it is an exponential random variable then it has the memoryless property. Similarly, if you are told that okay it is a random variable that has the memoryless property, then it is for sure the exponential random variable. That is what is an, so this is what is an if and only statement. So, it is memoryless if and only if it is an exponential random variable. Now, we will see proof, but again we will just prove the first part that if it is exponential, then it has this memoryless property will not prove this other part that if it is a continuous random variable with the memoryless property, then it is exponential distribution, that proof is slightly complicated we will not prove it in this course, but we will see one side of the proof that if it is an exponential distribution with some parameter $\lambda > 0$, then it satisfies this memoryless property or it has the memoryless property. So, we will look at this $\mathbb{P}(X > t + s | X > t)$. So, we are using the formula for conditional probability, we know that probability A given B is probability of A intersection B divided by probability of B. So, this is nothing but $\frac{\mathbb{P}(X>t+s \cap X>t)}{\mathbb{P}(X>t)}$. Now, $\mathbb{P}(X > t + s \cap X > t)$ is just simply this, why?

Because, if $(X > t + s)$, then for sure it is greater than t. So, this set where $(X > t + s)$, it is a smaller set, because if $(X > t + s)$, then for sure it is greater than t, but not other way if it is greater than t you cannot say it is greater than $t + s$. So, this set, set of all Ω for which, if you look at this, set of all Ω for which X of Ω is greater than $t + s$, that is a subset of all Ω such that X of Ω is greater than t. So, this intersection this is this, so, if you have a smaller set and a bigger set, so, if $A \subset B$, then $A \cap B = A$.

So, that is why we get that this is basically, so, $\mathbb{P}(X > t + s | X > t)$ is just simply $\frac{\mathbb{P}(X>t+s)}{\mathbb{P}(X>t)}$. Now, you know it is an exponential random variable, so, you know what its CDF is. So, now, $\mathbb{P}(X > t + s)$ so, this will be just $1 - F(t + s)$, so that you get as $e^{-\lambda(t+s)}$ and the denominator is probability X greater than t. So, it is $e^{\lambda t}$. Now, if you just do the cancellation you get $e^{\lambda s}$ which is simply probability X greater than s. So, we have proved that if X is an exponential distribution with parameter λ then it has the memoryless property which is $\mathbb{P}(X > t + s | X > t)$ is just simply probability X greater than s. So, this distribution does not have memory.

So, if given that an electric bulb has survived for t units of time and if you are being asked that, what is the probability that it will serve for s additional units of time that is same as asking what is the probability that a new bulb will survive s units of time. That is the memoryless property in real-life situations. Provided you are told that the lifetime of the electric bulb follows an exponential distribution with some parameter $\lambda > 0$.

So, that was the first property. Now, second important property if $X_1, X_2, ..., X_n$ are i.i.d. That means, independent and identically distributed random variables with having distribution exponential λ. So, each X_i is $Exp(\lambda)$ and X_i's are independent of one another. So, that is the meaning of $X_1, X_2, ..., X_n$ are i.i.d. $Exp(\lambda)$. Then if I look at this some random variable. So, $S_n = \sum_{i=0}^{n} X_i$. Now, what this theorem is telling that the sum is again a continuous random variable having this probability density function. Now, if you have seen this density before then you should know that this is the density of $\Gamma(n, \lambda)$ distribution. So, this is the density of a random variable that has $\Gamma(n, \lambda)$ distribution. So, what this theorem is telling you is that if $X_1, X_2, ..., X_n$ are i.i.d. $Exp(\lambda)$, then if you look at the$\sum_{i=0}^{n} X_i$ then that has $\Gamma(n, \lambda)$ distribution. What is $\Gamma(n, \lambda)$ distribution? It is a continuous random variable having probability density function equal to this. Again, you must have seen Γ distribution or Γ distributed random variable in your basic course in probability. So, then this density should be familiar to you, but anyway if you have not seen this before then if $X_1, X_2, ..., X_n$ are independent and identically distributed $Exp(\lambda)$ random variables then if you look at the sum then it has $\Gamma(n, \lambda)$ distribution, which is that it has, it is a continuous random variable having this probability density function. Now, this is true for any n. So, we will prove it by mathematical induction. Now, it is easy to see that this is true for n equal to 1, n equal to 1 means you are just looking at X_1, but what is X_1, X_1 has exponential distribution. And what is the density? The density is $\lambda e^{-\lambda s}$ for $s > 0$, 0 otherwise. So now, if you compare this and this. Now, remember what is $\Gamma(n)$? $\Gamma(n)$ is $(n-1)!$.
So, that is the formula. So, if n is, then this is true for any $n \geq 1$. So, and also 0! is equal to 1. So, if you plug in $n = 1$ here, so you get $\frac{\lambda}{\Gamma(1)}$ which is 0!, which is 1. Now, s^{1-1}, so s^0, that is 1, $e^{-\lambda s}$. So, you get precisely this thing. That the density is $\lambda e^{-\lambda s}$ for $s > 0$.
So, if I just look at for $n = 1$, that means I am just looking at a single random variable, since this is exponential, since this has exponential λ distribution, so, its PDF is given by this and if you plug in $n = 1$ in this formula, then you just get this using the fact that $\Gamma(n)$ is $(n-1)!$. So, how does mathematical induction work? You first prove it for $n = 1$ or $n = 0$ depending on the situation.
So, here we prove it for $n = 1$. Now, we assume it for some $k-1$ and then you prove it for k. That this assumption that it is true for $k-1$ that is what is called the induction hypothesis. So, we assume it, assume that it is true for $n = k - 1$. So, this is what is called induction hypothesis. So, we assume that it is true for $n = k - 1$. Now, we will show that this is true for $n = k$. That is how that is the process or method of mathematical induction.
Now, if X and Y are two independent continuous random variables. So, before going to the final step, which is the inductive step, we need one more fact. So, the fact is if X and Y are two independent continuous random variables with probability density functions f and g respectively. Remember, independence is important. So, if it is, it is important that X and Y are independent.
So, if they are two independent continuous random variables having probability density functions f and g respectively, then if you look at $X + Y$ the sum that is again a continuous random variable with probability density function given by this formula. This is what is

called the convolution formula. Again, the name does not matter, but this is the formula. So if X and Y are two independent continuous random variables.
Independent is very, very important. If X and Y have some dependency, then this formula is not true. So, if X and Y are two independent random, continuous random variables having probability density function f and g respectively, then if you look at the random variable X plus Y, that is again a continuous random variable having probability density function given by this formula $\int_{-\infty}^{\infty} f(s-t)g(t)dt$.
So, this formula is what is called the convolution formula, but anyway, the terminology does not matter. Now, we will use this fact and the induction hypothesis. The induction hypothesis is, that we have assumed that, so we have assumed that this thing is true for $n = k$ that if I look at $i = 1, ..., k$ X_i, that is a continuous random variable having this PDF for $n = k$. So, now, we want to prove it for $n = k$. So, we have assumed it for $n = k$ sorry, so, we have assumed this for $n = k - 1$.
So, $i = 1, ..., k - 1$ X_i has this density with $n = k - 1$. Now, we will prove it for $n = k$. So, we want to find the density of S_k which is k running, $\sum_{i=1}^{k} X_i$. So, again, so, now remember, since this each X_i's are independent, now, if I look at S_{k-1}. Remember what is S_{k-1}, that is $i = 1, ..., k - 1$ X_i and you have X_k. Now, since X_is are i.i.d., S_{k-1} and X_k are independent. So, these two are independent S_{k-1} and X_k are independent are independent. Again, both these are continuous random variables having some probability density function. So using this fact, the density function of $f(S_k)$, so, which is $S_{k-1} + X_k$ by using these fact the should be equal to this. But now, see, first of all, $f(X_k)$ remember, this is an exponential random variable, so, this is only positive for when t is, sorry, so, this should not be t, so, this should be S, this is S, this is S. So, $f(X_k)$ remember X_k is exponential.
So, the exponential random variable, its density is positive only for $t > 0$. So, the part minus infinity to 0 vanishes. Again, if you look at this part, remember, this density is positive only when this argument is positive. So, this thing is positive only when $t < s$. So, finally, this integral limit becomes 0 to s because outside that either this is equal to 0 or this is equal to 0.
So, this first part is equal to 0 for $t < 0$ and the second part is equal to 0 for $t > s$. So, finally, you get this integral from 0 to s. Remember, so, this is a typo it should not be t should be s. So, integration 0 to s, this thing. Now, if you just do the cancellation. So, here this is one $e^{-\lambda t}$, this is $e^{-\lambda t}$ all those things will can get canceled. The integrating variable is t, so $e^{-\lambda t}$ will come out. There is one $(\lambda)^{k-1}$.
Here, there is a λ. So, you get $(\lambda)^k$ and again this $\Gamma(k-1)e^{-\lambda s}$. Now, if you do this integration you will see, you just get this is equal to, so just this integration is equal to $\frac{S_{k-1}}{k-1}$. Now, remember what is $\Gamma(k-1)$? $\Gamma(k-1)$ is $(k-2)!$. Now, with $(k-2)!$, if you multiply $k - 1$ then you end up with, so if you multiply $(k-1)(k-2)!$, you get $(k-1)!$, which is nothing but $\Gamma(k)$.
So, doing all this you finally end up with this. So, you have shown that, that S_k has precisely this density. So, it has this density for $(n = k)$. So, by mathematical induction, we have proved the result for all n. So, if X_1,X_2,...,X_n are independent and identically distributed

exponential λ random variables then if you look at the sum which is $X_1 + X_2 +, ... + X_n$ then it has what is called gamma distribution.
And the meaning $\Gamma(n, \lambda)$ distribution, the meaning of that is it is a continuous random variable having this particular density. So, again this is 0 for $s < 0$ and it is equal to this for $s > 0$, and we have proved it by mathematical induction. So, we have shown first that it is true for $n = 1$ because when it is $n = 1$ it is just X_1 and it has exponential distribution. Then we assume it for $n = k - 1$ and then we prove it for $n = k$, but in order to prove it for $n = k$, we need this important fact that if X and Y are two independent continuous random variables having probability density functions f and g respectively, then if you look at the some $X + Y$ that is again a continuous random variable and which has density given by this formula. So, remember it is very, very important that you have independence of X and Y. This if X and Y has some dependent structure then this formula is not true.
Now, using that fact and induction hypothesis, we show that $f(S_k)$ or we show that S_k also has the correct density. So, the fact we use, using the fact we get this first formula and then we just write down the densities do the integration and finally, we end up with the correct result. So, by process or by method of mathematical induction, we have proved our result.
Now, moving on to the last property that we will be seeing in today's lecture. So, again we have a theorem if X_1,X_2,...,X_n are independent exponential lambda i. Now, (())(27:25) each X_is are independent, but they are not identically distributed each X_i has exponential λi distribution, so, the parameter, so, each of them is exponential, but the parameter can be different for different random variables, but independence is required.
So, if X_1,X_2,...,X_n are independent exponential λi then if you look at this random variable, which is $min_i X_i$, then that again has an exponential distribution where the parameter is sum of all these parameters. So, for example, if I just look at two things say X_1 and X_2, then and suppose this as parameter λ_1 and this is parameter λ_2, then if I look at minimum of X_1 and X_2 that will have exponential $\lambda_1 + \lambda_2$ distribution.
So, if you look at independent exponential random variables minimum of that, then that is again an exponential random variable, where the parameter is some of the parameters of the random variables whose minimum you are taking. Remember, again independence is crucial, you do not have this if there is some dependence structure, even if the marginal distributions are exponential, but if there is some dependent structure, this is not true, independence is very, very crucial.
Now, the proof is very simple. So, for $t > 0$. Again, so, if since each X_i takes values only greater than 0, because each of them is exponential, so, the minimum will also take value greater than 0. So, we only need to check for $t > 0$, because the probability that minimum will be less than or equal to 0 is 0. So, probability $min_i(X_i > t)$. That means, so, if minimum is strictly greater than t.
That means, each of the random variables is strictly greater than t. So, you get $(X_1 > t)$, $(X_2 > t)$ and so on, $(X_n > t)$. But now remembers X_is are independent and if you are independent and probability of intersection is product of the probabilities. So, you get probability of, this probability is basically product of this $(X_i > t)$ i running from 1 to n. So,

these follows from independence.
Now, since each X_i is an exponential random variable, we know what probability $(X_i > t)$ is that is just $e^{\lambda_i t}$, product i running from 1 to n, but if you do product of exponential, the exponential just gets added, the thing in the exponent just gets added. So, this product is equal to this. But this is precisely nothing but the, so this is equal to $1 - F_X(t)$ where $F_X(t)$ is $1 - e^{\sum_i \lambda_i t}$. And this you know is precisely the CDF of exponential lambda i.
So, we get that, since like distributed, there is this 1 to 1, so you can identify a random variable from its distribution function. So, as soon as we know that. So, here we are not exactly looking at the distribution function, but 1 minus the CDF, but if you have 1 minus CDF from there, you can easily find the CDF. So, we have seen that the CDF of this minimum random variable is equal to $(1 - e^{\sum_i \lambda_i t})t$.
So, which is precisely the CDF of an exponential random variable with parameter i running from 1 to n, λ_i. Hence, we have shown the result. So, again you see we have crucially used independence in our proof.
So, actually, there is one more theorem that we are going to see in today's lecture, so, that the previous one was not the last. So, but this is going to be the last property that we are going to see in today's class. So, again we have another theorem, which says that again if X_1,X_2,...,X_n are independent exponential λ_i. So, again independence is required, but we are allowing that each X_i can have its own parameter λ_i.
Then the $\mathbb{P}(X_i = \min_j X_j)$ is equal to this. That means, what is the probability that, so, in the previous theorem was how is the minimum, what is the distribution of the minimum if X_1s are independent exponentials? Now, this property says like what is the probability, so suppose you are given n independent exponential random variables. What is the probability that say X_1 will be a minimum of X_1,X_2,...,X_n, what is that probability?
What is the probability that the i'th random variable will be the minimum of among these n random varia?les. So, the what this theorem is telling that probability is equal to $\frac{\lambda_i}{\sum_j \lambda_j}$. So, this λ_i corresponds to this i. So, the probability that X_i will be the minimum is given by the parameter of X_i divided by the sum of the parameters. So, it is kind of an intuitive result.
So, the probability that the i'th random variable will be the minimum among the n random variables is equal to the parameter of i divided by the sum of the parameters. Now, the proof, so it says for any n, so we will just prove it for $n = 2$. So, using this idea, you can prove it for any n. So, but we will see just for $n = 2$, but tthe idea should be clear from this proof. So, we are interested in what is the probability that X_1 is equal to minimum of X_1 and X_2.
But that is same as $\mathbb{P}(X_1 < X_2)$, in that case only X_1 is equal to minimum of X_1 and X_2. But what is $\mathbb{P}(X_1 < X_2)$? So, here we are using this conditional formula which you have seen in the preliminary. So, $\mathbb{P}(X_1 < X_2)$ is $\int_0^\infty \mathbb{P}(X_1 < X_2|X_1)$ is equal to x multiplied by that density. So, you have seen this formula in the just the first few lectures that you saw, which was the preliminaries.
So, $\mathbb{P}(X_1 < X_2)$ is integration 0 to infinity, $\int_0^\infty \mathbb{P}(X_1 < X_2|X_1)$ is equal to x times the density

dx. So, you have seen this formula. But now, you know that X_1 and X_2 are independent. So, now, if I, if you are told that $X_1 = x$, so, you are told that $X_1 = x$, but then this just becomes $\mathbb{P}(x < X_2|X_1 = x)$. But now X_1 and X_2 are independent, so the conditional probability will be same as the unconditional probability.

So, this is basically now becoming probability that $\mathbb{P}(x < X_2|X_1 = x)$. But now, if you look at this, since X_2 and X_1 are independent, now, this conditional probability will be just equal to the unconditional probability. That is, because X_2 and X_1 are independent. So, that is where we are using independence, this is where we are using indie independence. So, this becomes equal to this, but now, again you know X_2 is exponential with parameter λ_2.

So, $\mathbb{P}(x < X_2)$ is just simply e to the power minus lambda 2x times this density. But now, so, this λ_1 will just come out so, this will inside integration it will be just $-(\lambda_1 + \lambda_2)x$. Now, the integration of that and then if you plug in the limits, you will just get $\frac{\lambda_1}{\lambda_1+\lambda_2}$. So, this integration you can easily do.

So, we have got that probability that X_1 will be minimum among X_1 and X_2 is given by $\frac{\lambda_1}{\lambda_1+\lambda_2}$. Similarly, if you would have tried to check that what is the $\mathbb{P}(X_1 = \min(X_1, X_2))$ then you would have got the answer as $\frac{\lambda_2}{\lambda_1+\lambda_2}$. So, you can just check this just following the same method.

In this case, you will have to find out what is the $\mathbb{P}(X_1 < X_2)$. Again, you can follow the same process and you will end up with this answer. So, given n independent exponential random variables, the probability that the i'th random variable will be the minimum among the n random variables is equal to the parameter of the i'th random variable divided by the sum of all the parameters. So, we have seen today four properties.

First, the memoryless property. Second is if you have n independent identically distributed exponential random variables with parameter λ, then their sum has gamma n lambda distribution, and I have explained what Γ(n,λ) distribution is. So, it is a continuous random variable having a certain probability density function. So, that was the second property.

The third property was that if you take n independent exponential random variables each having their own parameter λ_i, then the minimum of them has again exponential distribution with parameter the sum of the parameters. And the last property that we saw is that again if you are given n independent exponential random variables having like possibly different parameters, then the probability that the i'th random variable will be the minimum among the n random variables that is given by the parameter of the i'th random variable divided by the sum of that parameters.

So, you will see that all these properties we will be needing, when we look at Poisson processes, when we try to prove various properties of Poisson processes, we will need this property. So, this lecture and we will be having one more lecture at least on exponential distribution. So, these are all preparations for the Poisson processes. So, we will stop here today. Thank you all.