Discrete Time Markov Chains and Poisson Processes
Professor Subhamay Saha
Department of Mathematics
Indian Institute of Technology, Guwahati
Exponential Distribution and Poisson Processes
Lecture 27
Order Statistics

Hello everyone, welcome to the 27th lecture of the course Discrete Time Markov chains and Poisson processes. So, in the last lecture, we solved a couple of problems on exponential distributions, where you we used all the properties of exponential distributions. So, today we will see another small thing, which we will need in for Poisson processes remember we are doing all these as preparation for Poisson processes. So, before we can start partial Poisson processes, we will need one more thing which we will learn in today's lecture.

And that thing is what is called order statistics. So, what is order statistics? So, let X so, it is a theorem. So, let $X_1, X_2, ..., X_n$ be independent and identically distributed continuous random variables with probability density function f. If we let X_i denote the ith of these smallest of these random variables. So, that means X_i is the smallest n is the maximum 1 X_2 is the second smallest X_3 is the third smallest and so, on.

So, if we let X_i denote the ith smallest of these random variables, so, you will have n iid continuous random variables iid means independent and identically distributed with probability density function f. Now, if we let $X_{(i)}$ denote the ith smallest of these random variables now, if I look at these collections $X_1, ..., X_n$, so, here it is now $X_{(1)}, X_{(2)}, ..., X_{(n)}$ these are called the order statistics.

So, you are looking at from smallest to largest. So, that is basically the definition of so given n iid continuous random variables, what is the order statistics corresponding to that, if you basically order those random variables from smallest to largest. So, now, the theorem says that the joint probability density function so, now, this is you can think of it as a random vector this $X_{(1)}, X_{(2)}, ..., X_{(n)}$.

So, the joint probability density function of this random vector. So, this again will be a continuous random vector is given by this formula. So, what is it? So, remember you are looking at order statistics, so, here this $X_{(1)}, X_{(2)}, ..., X_{(n)}$. So, these are the variables, so obviously because $X_{(1)}$. So, $X_{(1)} < X_{(2)} < ... < X_{(i)}$. So, these variables should also respect that ordering.

So, first of all this is non 0 only when this $X_{(1)} < X_{(2)} < ... < X_{(i)}$ because the random variables has this relationship between them. So, if this is not true, then the probability density function or the joint probability density function is 0 and if this is true, then it is given by this quantity. Now, we will not prove this theorem, but let me try to explain to you heuristically why you get this.

Now, suppose instead of a X_n let me look at say X_1, X_2, X_3. Now says for any 3 values, so, if I suppose it takes X_1, X_2, X_3. Now, you can permute each of these values but still when you order so for example, what I mean is suppose if I give you 1, 2, 3 or 3, 2, 1 or 2, 1, 3.

Now, if I so, when you look at So, in this case, this is X_1, this is X_2 and this is X_3 in this case, this is X_1, this is X_2, this is X_3, similarly, this is X_1, this is X_2 this is X_3. But if you look at $X_{(1)}$, then this is, so what I am saying is if I give you n numbers.

Now, all for all possible $n!$ permutations of those n numbers, if I just order them, they will give me the same ordering just like this 1, 2, 3, 3, 2, 1, 2, 1, 3. So, if I just look at it as a 3 topple then these are different, but if I just ordered the take the numbers and order them, they will give me just one thing 1, 2, 3. So, if you are given n numbers, if you are looking at it, so all if you are looking at doing permutations of them, and if you are looking at these as n topples, then they are different.

But if you are just ordering them, so all possible all n factorial possible permutations of n numbers will give you the same ordering that is important. That is why you get this n factorial because see, when will these X_1,X_2,...,X_n take these values x_1,x_2,...,x_n. You so it does not so for example, say if, as I said say if X_1 is 1 and X_2 is 2, or X_1 is 2, and X_2 is 1 in both these cases, you are X_1 is 1, and X_2 is 2. So, if you permute a set of numbers, its ordering does not change still.

So,1, 2 among 1, 2, 3 the lowest is 1 and the highest is 3 similarly, if I give you 2, 1, 3, still the lowest is 1 and the highest is 3, yes. Now, the lowest one is the second random variable, and the highest one is the third. So, then which one is the lowest and which one is the highest that might change. But just if you look at the ordering, that does not change, even if you permute the numbers, that is why all possible permutations. So, since there are n numbers, so all possible permutations, then all the number of possible permutations is n factorial.

So, you get this and why do you get now so, see that is the thing. So, your random variables to take the values X_1, X_2,...,X_2 but you know that they are independent. So, as well as identically distributed having this density function, so, the joint density becomes this product density. So, the $n!$ comes because you even if you permute the numbers the or still that will not change the order statistic just the example I gave if X_1 is 1 and X_2 is 2 or if X_1 is 2 and X_2 is 1 in both these cases the order statistic is X_1 is 1 and X_2 is 2.

So, corresponding to both these possibilities, you get the same order statistics, but here they are just to 2! is 2, but if you have n then you get and there are this $n!$ possibility, which will give rise to the same order statistic and from where this second term is coming, that is just independence and identically distributed, that is why you get this product of density of the marginal densities.

So, that is the heuristic explanation of this density again, if you have to prove it, the proof is slightly more complicated, but anyway the heuristic should be clear. Let me just clean this a bit. Now, when you have a joint density, then you can calculate the marginal densities, just by integrating out the other variables. So, doing that again, we will not see the proof of that because the proof will be some slightly tedious calculation again, the proof is not difficult like finding when you are given the marginal joint density finding the marginal density, you know what you have to do, you have to just integrate out the remaining variables.

But again, since here it is a but the integration can be complicated. So, we do not we will not do the integration here. But in principle, you know how to do how to find marginal density when you are given the joint density. So, here this corollary is saying. So, let x_1,x_2,...,x_n be iid continuous random variables with probability density function $f(.)$ and cumulative distribution function $F(.)$, then the marginal of the ith order statistic.

So, we are looking at $X_{(i)}$, then so the marginal of the ith order of the marginal density of the ith order statistic is given by this quantity. So, again although we will not see proof of this, but let us see. So, again this will be true for where all these things are positive. Now, again let us try to see if we can understand heuristically from where this density is coming. Now, if the ith minimum is X.

So, remember the ith minimum is X which means below that there are $i-1$ number and after that there are $n-i$ remember there are total n numbers you are told that X is the ith minimum that means, there are $i-1$ number below it and $n-i$ numbers after it. So, that you see. So, again this is f_X is coming because that because of this that the ith 1 is X. Now, this capital $F(x)^{i-1}$ comes from the fact that there are $i-1$ below X remember what is $F(X)$ if X or let me write it here.

So, $F_X(x)$ is $\mathbb{P}(X \leq x)$. So, this is $F(X)^{i-1}$ remember here these are all independent and identically distributed. So, it becomes just the product. So, basically, it tells you that I think should be less than X and that probability is given by that f sorry $i-1$ things should be less than X. So, that is $F(X)^{i-1}$ again it becomes this goes to the power because of independent and identically distributed and $n-i$ should be greater than that that is $(1-F(x))$.

Because $(1-F(x))$ is probability X greater than x. So, $n-i$ things should be greater than X that is when i will be the sorry X will be the ith minimum if $i-1$ is less than that and $n-i$ is greater than that. Now, so, basically what you are doing here, so, again. So, X will be the ith minimum when $i-1$ things are less than that and $n-i$ things are greater than that. But again remember in terms of the original random variables, you can permute things and but still you will get the same order statistics.

So, basically what you have to choose which $i-1$ will be less than which $n-i$ will be greater than and which one will be the ith minimum. So, basically from n people you are making three groups a group which will be this $i-1$ minimum that will contain $i-1$ things, then the which 1 will be the highest minimum that is a group of 1 and then a group of $i-1$ which will be greater than X.

So, the number of ways you can group n people in in these three groups one group containing $i-1$ one group containing 1 and the other group containing $n-i$. So, what you are doing you are group you are given say n things you need to group among them, you need to choose a group of $n-i$, $i-1$ and one which will be the ith minimum. So, first you choose $n-i-1$ give them values. So, for any possible values of them, which is less than X you give them values less than X.

Now, $n-i$ you give them values greater than X. So, now, you can give any values to this $n-i$, $i-1$ as long as $1 \leq X$ and the other is greater than X. So, basically you are given n numbers you have to make them divide them into three groups. The ith 1 should be X $i-1$ should be less than that and $n-i$ should be greater than that in how many ways you can

do such thing that thing you can do in these many ways $\frac{n!}{(i-1)!(n-i)!}$.
Because for all such groupings, you will get that the ith 1, ith minimum is X, $i-1$ is less than X and $n-i$ is greater than X. So, basically you have to choose which one will be less than among the first $i-1$ and which one will be in the letter $n-i$. So, and how many ways given n things, how many ways can you form such groups it is it again that is just permutation combination is just $\frac{n!}{(i-1)!(n-i)!}$ that gives you this density.
So, again, this is not a mathematical proof, but I have trying to give you a heuristic understanding of this density. So, if X has to be the ith minimum, then $i-1$ has to be less than that $n-i$ has to be greater than that and again, so, since order statistic does not change on permuting, so there are so you can get this in multiple ways in how many ways. So, given n things you have to choose, which i you want to give value $i-1$, you want to give values less than X and which $n-i$ you want to give values greater than X.
And that can be done in these many ways, that grouping can be done in $\frac{n!}{(i-1)!(n-i)!}$ ways. So, that gives rise to this density. But again, if you have to do it mathematically, you will have to just do it from this joint density by integrating out the other variables, but that integration is slightly complicated. So, I have not worked out that thing in this lecture.
But again, you heuristically, this density should make sense. So, this slide gives you what is the joint dense joint pdf of the order statistic given n iid continuous random variables with a common probability density function f and also it gives you the marginal density of the ith order statistics.
Now, let us see an example. So, let $X_{(1)}$,$X_{(2)}$ be two independent exponential random variables, each having parameter λ. Find expectation of $X_{(2)}$ and variance of $X_{(2)}$. So, we will use the previous thing. Now so basically, we get from this corollary, that the PDF of $X_{(2)}$ is given by this that is easy to see, just using this formula. Now once I have the density, now I can calculate the mean and the variance. So, to find the mean, we just integrate this from 0 to infinity and integrate the density from 0 to infinity multiplied by X.
Now, you just separated, so this and this, but now remember, if you take this 2 outside, this is just $\int_0^\infty x\lambda e^{-\lambda x}dx$, that is nothing but just mean of an exponential random variable, with parameter λ. So, that is $\frac{1}{\lambda}$. But since you have this 2, you get $\frac{2}{\lambda}$. And now if you look at this integral, this is just the so $2\lambda e^{-2\lambda x}$, that is just the density of the probability density function of an exponential distribution with parameter 2 λ.
And so here, you are just calculating mean of that. So, the mean of that will be just $\frac{1}{2\lambda}$. So, you get $\frac{2}{\lambda}-\frac{1}{2\lambda}=\frac{3}{2\lambda}$. So, again, this integration you can easily do, but what I am saying is that since you know what is the mean of exponential distribution with a given parameter, you can you do not even have to do this integration work out this integration explicitly.
If you remember that then you can straight away just write down these things. So, that is the mean. So, here $X_{(2)}$ since here is just two random variables. So, $X_{(2)}$ is basically $\max(X_1, X_2)$. So, you are trying to find the maximum expectation of the maximum of two independent exponential random variables, each having parameter λ. So, you already know what is minimum for minimum, we already know that the minimum of two exponential random variables is again exponential with parameter just some of the parameters.

So, here again, here we are not exactly finding the distribution, but again, so we actually have found the distribution because this is the density of the maximum. So, you can, so that gives you the distribution of the maximum and also we have calculated the mean of the maximum.

Now what about the variance? Now before finding the variance we first find the second moment. So, again, we need to calculate this because the density is what is given in the previous slide, now, again we will use so if X is $Exp(\lambda)$, then the expectation of X^2 is given by $\frac{2}{\lambda^2}$. Again, this you can easily check, so we will use it here. So, again we break the integral, now the first one again, so it 2, we can take out this 2 and then it is just $x^2\lambda e^{-\lambda x}dx$. So, this is just expectation of X^2 for where X is $Exp(\lambda)$. So, that will give you 2. So, this 2 is already there 2 times 2 over lambda square which gives you $\frac{4}{\lambda^2}$. And this is expectation of X^2 where X is $Exp(\lambda)$. Now, I look at this density $2\lambda e^{-2\lambda x}x^2dx$. So, now it is just the expectation of X^2 where X has this distribution.

So, now we get just 2 times. So, this is nothing so this is basically $\frac{2}{4\lambda^2}$ because it will be 2 over the parameter square. So, here the parameter is 2 λ, so, we get $\frac{1}{2\lambda^2}$ now, if you do this you get $\frac{7}{2\lambda^2}$. So, the variance is the expectation of X square minus expectation X the whole square. So, our expectation of this $X_{(2)}$ whole square minus expectation of $X_{(2)}$ the whole square we already know from the previous slide what expectation of $X_{(2)}$ is.

So, finally, we get that this is the variance. So, using this theorem in the first slide, we get this mean and variance of maximum of 2 independent and identically distributed exponential random variables with parameter λ. But now, we already have seen that exponential random variables are somewhat spatial.

Now, it turns out like, just again, this is the general method, the first method that we worked out is the general method. So, instead of exponential, if it was some other random variable, then we would have followed this method, which we did first. But now, we know that exponential random variables are somewhat special, it has some very special properties, one of them being this memoryless property.

Now, it turns out because of that, this particular problem is for exponential random variables can be done in a slightly easier way, because here you see, you need to do all these integrations. So, if here I use those facts mean of exponential, the second moment of exponential but if you do not remember all those things, you will have to do or work out all these integrals, but since this is since the random variables are exponentially distributed, the problem can be done in a simplified manner.

So, again, that is because it is exponential instead of exponential if it was something else, then the first method the way we calculate it, that is the way to do it, but since exponentially especial, so, the problem is for exponential random variables, they give a second method which you will see is a much simpler method to calculate mean and variance. Remember, because we are just interested in mean and variance.

So, the other method also gave us the distribution because it gave us the probability density function, but since we are just interested in mean and variance the problem asks you just about the mean and variance and the random variables are exponential you can do it in a simpler way what is that simpler way? So, now, by memoryless property $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent exponential random variables with parameters 2λ and λ respectively, why is that?

Because, what is this $X_{(2)} - X_{(1)}$ that is just simply like the remaining. So, X1 tells you $X_{(1)}$ is basically the minimum of 2 exponentials. So, that you know that has to be again an exponential random variable with parameter 2λ. But now, that is this $X_{(2)} - X_{(1)}$ and $X_{(1)}$ is again independent why because this $X_{(2)} - X_{(1)}$ is you can think of it as the excess life. So, how much bigger this $X_{(2)}$ will be for in comparison to $X_{(1)}$. So, what is the difference? So, it is again looking at the excess life or the remaining life. So, that we know by mmemorylessproperty that how much more it will be again does not depend on what is you what you are already given that, it is at least this how much more it can be. Now, for that, what it is at least is not important only thing is how much more it can be. So, that is $X_{(2)} - X_{(1)}$. So, but by the memory less property that simply exponential that is because, again we have seen this argument before that the remaining like we saw in the previous lecture that this remaining service time is just the original service time. Because of this memoryless property. So, here is the same thing. So, if you think of this $X_{(1)}$ that is the service time of the $X_{(1)}$ is the service time of the first customer and who to leave and this $X_{(2)} - X_{(1)}$ is just the remaining time service time to the customer who is still in service when the first customer the customer who leaves first leaves.

So, this again is exponential with the original parameter and also these are independent. So, $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent exponential random variables with parameters 2λ and λ respectively that $X_{(1)}$ the parameter has changed because now you know that $X_{(1)}$ this is basically minimum of 2 exponentials and the parameters are λ and λ. So, the when you add λ plus λ you get 2λ and this $X_{(2)} - X_{(1)}$ that it is independent of $X_{(1)}$ and it is again exponential lambda that is because of memory less property.

So, if you have to find the expectation then expectation of $X_{(2)}$ you can just simply write it in this way now expectation is just linear. So, expectation of $X_{(1)}$ since it is exponential with parameter 2λ. So, its mean is $\frac{1}{2\lambda}$ and expectation of $X_{(2)} - X_{(1)}$ since it is exponential with mean λ. So, it is $\frac{1}{\lambda}$ and you get this $\frac{3}{2\lambda}$ which is the same answer that you got here obviously, you should get the same answer because you are working out the same problem in two different methods.

So, here we basically did not need the independence of $X_{(1)}$ and $X_{(2)} - X_{(1)}$ because the expectation is linear, but when we want to calculate variance again we break it in this way. Now, we know that variance of if you have given two random variables X plus Y the formula is the variance of X plus variance of Y plus covariance of X Y, but in this case, since it is independent, so, this covariance term is not gone because you all you know that if two random variables are independent then covariance of X Y is 0 or in other words, they are what is called uncorrelated.

So, this covariance term will not be there, so, this will be 0. So, to calculate the variance we are using the fact that they are independent. So, variance of this plus this will be just variance of this $X_{(1)}$ plus variance of $X_{(2)} - X_{(1)}$. But again what is variance of $X_{(1)}$ we know

that if you have Exp(λ) then variance is $\frac{1}{\lambda^2}$ again you can calculate that. So, if X is has exponential λ distribution then variance of X is $\frac{1}{\lambda^2}$.
So, since $X_{(1)}$ is exponential with parameter 2λ, the is variance is $\frac{1}{4\lambda^2}$ and what is variance of $X_{(2)}$, $X_{(2)} - X_{(1)}$ that is just exponential λ. So, it is just $\frac{1}{\lambda^2}$ again if you do the summation, you get $\frac{5}{4\lambda^2}$ again the same thing that you got by the previous method.
So, you see since the given problem is for exponential random variables, you can do it in a much simpler way without using those theorems are corollary on order statistics that is, because these are exponential random variables, which has this spatial property of memoryless that exponential random variables have this memoryless property. So, and since you again remember the previous, the previous method also gave you the PDF.
Because from that corollary, you get the PDF of the maximum as well but this method you do not get the PDF, but what you again you can because you remember now, you know what that $X_{(1)}$ is what $X_{(1)}$, so, again although I say that you cannot get the PDF, I have not worked out the PDF here, but just let me tell you, even in this method also you can get PDF, why?
Because, you know, this $X_{(2)}$ you can write it as $X_{(1)}+X_{(2)}-X_{(1)}$. Now, $X_{(1)}$ $X_{(2)}$, $X_{(2)}-X_{(1)}$ these are independent random variables. Both are continuous random variables have certain density, you remember this, you know that if X and Y are two independent continuous random variables having density functions f and g then you have this convolution formula.
So, the density is actually given by $\int_0^\infty f(x-y)g(y)dy$. So, that is say the density of the sum provided X and Y are independent. So, here again since you can write $X_{(2)}$ as $X_{(1)}+X_{(2)}-X_{(1)}$ is exponential with parameter 2λ, $X_{(2)} - X_{(1)}$ is exponential with parameter λ and these are independent.
So, you can use the convolution formula. So, you can try it out as an exercise using the convolution formula find the density of $X_{(2)}$ and you should end up with this formula. So, the convolution formula should give you this density. So, although in this example, I have just worked out the mean and the variance you can use these properties of exponential you can also find out the density of maximum of two exponentials using this convolution formula.
But again, since the main thing was just the question asked about mean and variance, so, we can calculate it very simply for because the here the given random variables are exponential, but if it is any random variable, then the first method which we employed is the one to use.
But since exponential is special, there is an easier way to find out the answers that you are looking for. So, we have worked out the problem in two different methods.
Obviously, we have got the same answer we should get that but here there are two different methods because we are working with or the problem is about exponential random variables.
So, now we have finished all the prerequisites for Poisson processes. So, in the next lecture, we will start with the Poisson process. So, we will stop here today. Thank you.