MITx Video | 6.041x | Spring 2014 | Video transcript | L23-6: Random incidence in the Poisson process

In this segment, we discuss the so-called "random incidence" paradox for the Poisson process. It's a paradox because it involves a somewhat counterintuitive phenomenon. However, we will understand exactly what's going on, and in the end, it will cease to be a paradox and we will have an intuitive understanding of what exactly is happening.

So consider a Poisson process that has been running forever, or think of it as a Poisson process that started a very long time back in the past. To make things concrete, suppose that the arrival rate is 4 arrivals per hour so that the expected interarrival time is one fourth, in hours, or that would be the same as 15 minutes. For example, suppose that the bus company in your town claims that buses arrive to your stop according to a Poisson process with this particular rate. But you don't really believe that your bus company is telling the truth and you decide to investigate.

So what you do is the following. You show up at some time at your bus stop and wait until the next arrival comes and also ask someone who lives near the bus stop, what time was the last arrival? And they tell you the last arrival happened at that time instant. And you measure this amount of time, which is the interarrival time, record what it is, repeat this experiment on many days, and calculate an average. What you're likely to see turns out to be something around 30 minutes.

At this point, you could go to the bus company and challenge them. You claim an arrival rate of 4 arrivals per hour, which would translate into interarrivals of 15 minutes, but every day I go and check the interarrival time and I find that they are close to 30 minutes. What's the explanation? What's going on? Is it that the belief or the claim of the bus company is incorrect, or is there something more complicated?

So let us try to understand what's going on by being very precise and careful. You show up at the bus station at some time-- let's call that time t star. You ask someone who has been at the station, when was the last arrival time, and they tell you, and it is some number U. You wait until the next bus, and the next bus arrives at some future time capital V. You are interested in the interarrival time that you're observing, which is the difference between these two random variables V minus U.

Now this difference-- let us split it into two pieces. There's one piece from t star until V, which is V minus t star. And there's another piece, which is the first interval, and this is t star minus U. Now t star, the time at which you arrive, is just a constant. Suppose that you arrive at the bus station at exactly 12 noon. There's nothing random about it. However, V and U are random variables. What kind of random variable is this? You show up at 12 noon and you wait until the first arrival. Because a Poisson process starts fresh at any given time-- so after 12 noon it starts fresh-- this is the time until the first arrival in a Poisson process with rate lambda, so this is a random variable which is exponential with parameter lambda.

Now let us understand what this random variable is. One way of thinking about it is to think of the Poisson process running backwards in time, so you live time backwards. You show up at 12 noon, and then time runs backwards, and you wait until you see the first arrival coming in this backwards universe. So we're dealing here with the time until an arrival in a Poisson process that runs backwards in time.

What kind of process is a backwards Poisson process? If you take a Poisson process in reverse time, the independence assumption is not affected. Disjoint time intervals are independent. Even if you reverse time, disjoints time intervals still remain independent. Any given time interval of small length delta will have certain probabilities of an arrival or of two arrivals, and these will be the same whether time goes forward or time goes backward. So the conclusion from this discussion is that the backwards running Poisson process is also a Poisson process, and so this time until the first arrival in the backwards process is just like the time until the first arrival in a Poisson process. So this also is an exponential random variable with parameter lambda.

Even more than that, these two random variables are independent of each other. Why are they independent? The length of this time interval has to do with the history of the Poisson process after time t star. The length of this time interval has to do with the history of the Poisson process before time t star, but in the Poisson process because of the independence property, the past and the future are independent, and therefore, this random variable is independent from that random variable.

In any case, the expected value of the interarrival interval that you see, the expected value of this random variable, is going to be the expected value of one exponential, which is 1 over lambda, plus the expected value of another exponential, which is 1 over lambda, and we get a result of 2 over lambda. And that's why when you actually carried out the experiment, you saw interarrival intervals that had a length of 30 minutes as opposed to the 15 minutes that you were expecting in the first place.

Now how can this be? Since the interarrival times in a Poisson process have expected value 1 over lambda, how can it be that the expected length of the interarrival times that you see have an expected value of 2 over lambda? Well, the resolution of this paradox has to do [with] what exactly we mean when we use the words an interarrival time.

There's one interpretation which is the first interarrival time, the second one, the hundredth interarrival time-- each one of these actually has an expected value of 1 over lambda. But this is a different kind of interarrival time. It's not the first or the second or some specific k-th interarrival time. It's the interarrival time that you selected to watch. When you show up at a certain time, like 12 noon, you're more likely to fall inside a large interarrival interval rather than a smaller interarrival interval. So just the fact that you're showing up at a certain time that's uncoordinated with the rest of the process makes you more likely to be biased towards longer rather than shorter intervals. And this bias is what causes this factor of 2.

So it's an issue really about how you sample or how you choose the interarrival time that you're going to watch, and this particular sampling method has a bias towards longer intervals. As we will see, this is not something that's specific to the Poisson process. In general, in many occasions there are different ways of sampling which give you different answers, and we will go through a number of examples that will give you some intuition about the source of the discrepancy between these two answers.