

module_prob4_lecture1

Fri, 2/18 12:51PM 11:13

SUMMARY KEYWORDS

module, random variable, probability, number, continuous random variables, distribution, expectation, trials, lambda, values, called, squared, denoted, bernoulli, normal distribution, important, concept, occurrences, factorial, variance

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:06

So let's get started on a new module of probability. This is continuous random variables. So in the previous module, we covered what's called discrete random variables, those which could only take some discrete set of values, right? Values like 1, 2, 3, and so on. But nothing in between continuous random variables are those, which can also take all these in between values. Right, so let me quickly go over the module outline for this module. So we'll start with a quick review of the previous module on discrete random variables. Then I'll cover a topic, which is related to things which we did in the previous module, which is expectation and variance. So this time, we'll be covering the concept of covariance and correlation, which is the relationship between two sets. And then we'll come to the main feature of this module, which is continuous random variables. And I'll start by introducing you the important concepts of PDF, and density. And then this segment will be we'll be making heavy use of concepts from calculus, things like differentiation, and integration. So you may want to review the modules on differentiation and on integration, to better understand the material in this module. And then we'll be covering expectation and variance of continuous random variables. And then we'll be ending the module by looking at some common continuous distributions, in particular, the uniform and the normal distribution. And we'll be spending a fair bit of time on the normal distribution, because it's such an important distribution. So let's get started by reviewing some of the things that we looked at in the previous module on discrete random sets. So in that module, we focused on random variables, but mostly on discrete random variables. So random variables, which take discrete values, like 0, 1, 2, 3, or minus one, minus two, minus 10. Right, so they don't take continuous values in between those numbers. So I introduced two main concepts in terms of probability there. One is called the probability mass function of the PMF, which is the probability that the random variable X takes a particular value X . So that's called the probability mass function or the PMF of that. The second idea was the idea of the cumulative distribution function, which is what's the probability that the random variable takes values less than or equal to X , right? So if this is small X , what's the probability that the random variable takes values less than or equal to smallest. And this is going to be a super important concept, also in this module on continuous random variables. And the way we found it, is by adding up all the probabilities of all the values that the random variable can take below X , right? So technically, this is the summation of the Y 's - for all the Y values which are less than or equal to. So that

gave you the CDF. So then, we talked about two particular measures, one is expectation. So expectation, is denoted by μ of X . And this is given by X times the probability of X and we add it up across all the X 's that gives the expectation of that particular. So this is called the mean of the random variable. The second concept was variance. And variance gives you a measure of the dispersion or how the spread of the random variable, this is usually denoted by σ^2 of X . And this definition is the expectation of X minus the mean holding squared. And what we showed is that there is an easier version of this formula, which is given by expectation of X^2 minus the mean which is expectation of X whole thing squared. So this part is X^2 . And you subtract that from expectation of X^2 , that gives you the variance. And we ended the module by looking at some common discrete distributions. We started with the simplest possible, which is the Bernoulli. So that was a single trial. And what you're interested in is whether that trial is a success, in which case we denoted by one, or if it's a failure, in which case we denoted by C . So more generalized version of this was the binomial distribution, which is the number of successes in n trials. So what we're looking at is what's the probability that the number of trials can be 0, 1, 2, so on up to n . And in particular, the probability that this random variable takes a particular value x , that means what's the probability that there will be small x number of successes in n trials, that's given by this expression here, this is called n choose x p to the power x one minus p to the power n minus. So there are two things, two important parameters here. One is n , the number of trials and p , the probability of success in each trials. The other thing that we did is the points of distribution, which gives the number of occurrences of an event in a particular time period, or an area. So it's the number of occurrences in a particular time period, or in an area right, an example where things like the number of crimes committed in a city during a month, the number of emails arriving in an hour. So here, X can take values 0, 1, 2, so on up to infinity. And what's the probability that this random variable takes a specific value of X , that's given by this expression here, which is e to the power minus λ , λ to the power X over X factorial. And let me remind you, so X factorial means it's the so this is X factorial is X times X minus one times X minus two, so on all the way down to one. And for the poisson distribution, the main determining factor, the main parameter is λ . And this is also the mean of the points of distribution. So this is the average number of occurrences within that time period. So λ equal, so suppose there are five crimes, which occur in an hour on average, right? So if you're interested in this number of occurrences, number of crimes in an hour or in a day, in a particular city, that average would determine the particular λ for that district. So so these are the main things that we covered in the previous module. And so just to jog your memory back? So here's a clicker question based on something that we did in the previous module. So you may want to stop the video at this point, and try this clicker question. So hope you got a chance to look at that do the clicker question what it asks us, which one of the following would be appropriate to be modeled as a poisson distribution? So he is that whether a rule of the die comes up, even not. So this isn't a case where either there's just one trial, and I if it's even then you call it one, and if it's odd, then you call it zero? So this is an example of a Bernoulli. Because there's just one trial, one roll of the die, and either it's a success in the terms of it comes up even or not. Secondly, B is number of car owners in a sample of 100 adults. So this is a case where you have 100 trials, right? And what are you checking for? As the number of car owners, right, so this is an example of a binomial what we see is the number of customers arriving at a bank during an hour. So that's a particular time period. And this is the number of customers arriving. So it could be 0, 1, 2, 3, 4. And going on, let's this is exactly the setup of the poisson distribution, right? So the correct answer to this clicker questions see. And let me just finish by discussing D , D is number who get a in a class of 200 students. So again, so here there are as if there are 200 trials, and what you're interested is the number who get a so this is an example of a binomial distribution. So so that was a quick review of what we had done in the previous module. So let me stop this clip here.

