

module_prob4_lecture7

Fri, 2/18 12:51PM 15:14

SUMMARY KEYWORDS

cdf, derivative, probabilities, probability density function, pdf, finer, intervals, equal, temperature, likelihood, point, suppose, relative, occurring, degree, clicker question, greater, rounding, derive, random variable

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

Since the last clip, I introduced this concept of the probability density function on a PDF, which gives us an idea about what are the relative likelihood of the occurrence of particular points. So, the definition of the PDF, which was given by small f of x is that it's the derivative of the CDF. And wherever f of x , if it's high, that means the relative likelihood of the occurrence of x , this is high. And if f of x is low, the relative likelihood of the occurrence effects is low. So let's do an example where we derive a PDF from a given CD. So here, I'll, I'll use some calculus, because when I have to look at derivative, I need to use calculus. So let's consider what's called a uniform distribution over zero to 10. So this, the CDF is given by f of x is equal to zero. For x less than zero. It's equal to one for x greater than 10. And in between zero and 10, f of x is given by x over 10. So here, I've graphed it. So see, for x less than zero, f of x 's. Here, the f of x is zero. So between zero and 10, it's given by x over 10. And f of 10, it's given by one. So you can use it to figure it out. So suppose you have a point four, right? So what's f of four? That's going to be equal to x over 10. So that's four over 10, which is equal to point. So now, if you want to know what's the relative likelihood of four occurring versus five, occurring versus six occurring, and so on. So how do we know the relative likelihoods of different values? So for this, we need to figure out the PDF, the probability density function, right? So the PDF is just given by the derivative of the CDF. So now, let's do that. So C for x less than zero, the PDF has given by f of x equal to zero. Now if you take the derivative of something, which is a constant, that derivative, if you want to take dy/dx of zero, that's going to be zero. So that means if I'm looking at the f of x , that means wherever x is less than zero, this density is going to be zero. Similarly, if you look at x greater than 10, there, again that the CDF is always equal to f of x equal to one, it's not change, right? So if you take the derivative of something, which does not change, so that's great to see. So that means when x is greater than 10, right? The CDF is correct. So the most interesting cases here, where for x between x between 10 and zero so they're f of x is equal to x over 10. So from if I want to find the density, so I need to take the derivative of x . And if you recall from calculus, right, so this 10 is a constant, so that can come out and we just need to take derivative of dy/dx x . So the derivative of x is just going to be equal to one. So this is just screen to view. So that means the density for x between zero and 10, that's equal to one. Right, so this is zero for x less than zero, this is one over 10, for x between zero and 10. And it's again zero for x . So for this particular random variable, the

density function looks like this. For less than zero, it's zero. For greater than 10, it's again, zero, and in between, it's one. So what it means is that if you are looking at the relative likelihood of these points occurring. So here, that these points are impossible to occur, because their relative likelihood is zero, here, the relative likelihood of points greater than 10, that's also zero. But in between, right in between zero to 10, all of these points, they have the same relative likelihood of occur, it's one over. So that's why this is called a uniform distribution. So all of these points between zero to 10, they all have the same relative likelihood. So this, this is how you can derive the probability density function of the PDF from a given c. Basically, just by taking the derivative of the CDF. So let me stop here and ask you to do a clicker question. It's an easy clicker question. So please do attempt it. So the clicker question as to what does CDF and PDF stand for CDF is the cumulative distribution function. PDF is the probability density function. So the correct answer to that is so now, now that I've introduced you to the concept of probability density function, which is essentially the slope, or the derivative of the CDF, let me explain it to you in an alternative way, right, just to give you an alternate way of understanding this concept, because this is a super important concept for continuous. So let's think of an example that suppose the random variable is the temperature in a city. And as I say, right, like although we often measure it in whole numbers, right, one degree two degree three degree temperature is actually a continuous random variable, because if you could measure it carefully, you could measure whether the temperature is 1.4235 versus or whether it is 1.5162, and so on. But let's start with that. Suppose you just measured the temperature in whole degrees. Because and because your temperature, the thermometer could only give you hold degrees, right? So in this case, you would have temperatures like 0, 1, 2, 3, 4 minus one, minus two so on, and you could just look at the probabilities of these different ones. So this is called the histogram and statistics, right? But on this axis here, I have the probabilities, right? So this is the probability of the temperature being zero degrees, the probability of temperature be one degree mean two degree three degrees. So what do you have done is that your thermometer cannot quite distinguish between point six and point seven and point eight it calls all of the more similarly, it cannot distinguish between 1.2 1.3 It calls all of them as one. Similarly, it rounds off to the nearest. What if you could have a better thermometer right now which could measure in point five degrees? So now you round off to the nearest point five degrees, so now your temperatures are in? Zero, 0.5, one, 1.5, negative 1.5, negative one, negative 1.5. So now suppose again, you plotted the probabilities of all of these. And these are your various probabilities, right? So what we have done is we have made this intervals finer right? Earlier, it was just one, two, so on. Now we've made it finer, we've made it zero 2.5 to one to 1.5 minutes. All right, so now that as a result, we have these more bars. Now, suppose we got an even finer thermometer, and broad, and it was measuring in point two, five degrees, right. So in this case, again, if we plot the probabilities, this will be finer, a lot more bars, right, representing the many finer points that we have. Now, suppose we continue doing spreads, suppose we continue making these intervals finer, and finer, and finer and finer. So now we have more and more of these bars. And these bars are very close to each other. And if we keep continuing doing this, right, what happens is that these bars, like these intervals now become infinite as Miss Lee small, and we have probabilities corresponding to each of these tiny intervals. And if we push this to the limit, that's the sense of calculus, right? If we push this to the limit, what we ultimately get, if we join the top points of all of these probabilities, is that we get a curve. And this curve here is exactly the probability. Right? So what the probability density function is the limit of all of these probabilities of these finer and finer and finer points. So that's another way of viewing the probability density function as giving you the relative likelihood of of each of these points, right? So if we look at a point like here, right, so what we have done is we've made like we've started with a broad interval, like, what's the probability of, of, of temperatures occurring around this rounding off to the nearest whole number, then rounding off to the nearest point five degree, then rounding off to the nearest point to five

degree point to one five degree and so on, and so on. And in the limit, we get the relative likelihood of this point, or this temperature occurring, and that is exactly your probability density function. So to sum it up, write the CDF, the cumulative distribution function, which is represented by capital F of x . This is the probability of the random variable taking values less than or equal to x . And the PDF, which gives you the relative likelihood of this point x occurring is just the derivative of the CDF. So let me end this clip with a clicker question, again, asking you to derive the PDF from occupancy. So hope you are able to do this clicker question. So what it says is that there is a CDF, which is given by $F(x) = x^2$ for x between zero and one is zero for x less than zero, and it's one for x greater than one. So which of the following is its PDF for values of x between zero and one. So for values of x between zero and one, $f(x)$ is given by x^2 . So the PDF is just going to be the derivative of this. And if you recall from calculus, the derivative of x^2 is just going to be $2x$. So if you recall the formula, the derivative of x^n is just going to be $n x^{n-1}$. So here, n is equal to two. So the derivative of x^2 is going to be $2x^{2-1}$, which is $2x$. So the correct answer here is this. So let's stop the tip here and in the next clip I'm going to talk a little bit more about the connection between PDF and CDF.