

module_prob4_lecture9

Fri, 2/18 12:51PM 15:49

SUMMARY KEYWORDS

integral, pdf, squared, expectation, continuous random variable, cdf, probabilities, variance, function, negative infinity, equal, random variable, limit, uniform distribution, probability density function, slope, denoted, dx, positive infinity, positive

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So so far, I have introduced you to the idea of PDF, the probability density function, and the CDF, the cumulative distribution function for continuous random variables, and showed you the connection between the two. So in this clip, I'll be talking about a couple of properties of the PDF, the density function, and also introducing you to finding expectation and variance for continuous random variables. To the PDF, which is given by small f of x is obtained by the derivative of the slope of the CDF, which is the capital F function. And vice versa. If you want to get the CDF at a particular point, you look at the integral of the area under the under the PDF up to that. So the first property of a PDF is that the PDF is always greater than or equal to zero. And why is that? Remember, the PDF is the slope of the CDF. So here is my cdf, the CDF is always increasing, because the cumulative distribution function, and so you're adding more and more probabilities to it. So it's derivative or its slope is always going to be positive. So that means the PDF must always be it cannot be negative, right? It has to be greater than or equal to zero, it can be zero, for example, here, the slope is zero, so the PDF is zero, but it cannot be named. The second idea is the following. Remember, I looked at one way of getting the probability density function was looking at the probabilities of smaller and smaller intervals, right. And in the limit, this curve gives you the PDF. Now, if we add up all of these probabilities, so they cover up all the possible values of x , so if you add up all the probabilities of all the possibilities, they must add up to one. So what it means is that if you look at the PDF, the density function and look at the total area under it, that must be one. Because what you're doing is that you're summing up the probabilities of all possible values of x . So technically, how do you write it? So if you look at the integral of f of x from the lowest possible value, which is minus infinity, up to the highest possible value, which is positive infinity, if we integrate f of x over that limit, that must add up that must be equal to one. So let me show you an example of this. So remember, the uniform distribution pdf that we have been doing so far, right? So it's, it's $f(x)$ is $1/10$, for x between zero and 10. And it's zero elsewhere. So, so here's the plot of the uniform distribution pdf of zero to 10, which you've seen before. So it's $1/10$, between zero and 10. And it's, it's zero everywhere else. So now if you're trying to look at the total area under this, it's the total area will be the will be this area here. And this here is a is a rectangle whose height is $1/10$. And what is its base its basis from zero to 10. So the $1/10$ times standards equal to one. But this property is not only true for uniform, but it is true for any any probability

density function, if you look at the total area under the PDF, it will always be so now, let me talk about the mean and the variance of a continuous random variable number for a discrete random variable the mean just given by the summation of x times the probability, you add it up over all the x 's. For a continuous random variable, the range of x is continuous. It's not that it's a discrete number of x 's, right? So you cannot do this summation. Because you will be summing up over an infinite and uncountably many links, small, small, small, small, all the x 's. So the equivalent thing to do here is take an integral. And what is the integral off? It's x times f of x , dx . This is the mean or the expected value of a continuous right? See the analogy here, right here, it was x times the probabilities, right? Summing those up here, instead of summing, you're doing the integral. And you're doing X , and instead of the probability, you're replacing it by the PDF. So it's the integral of x times f of x from negative infinity to positive infinity, over all the possible values of x . So that's the expectation of X tilde. And as you know, it's denoted by μ affects the similar idea if you want to, if you want to find the mean, not just of x , but have some some function of x , right? So for example, suppose you want to do X squared, or you want to find the expectation of square root of x . What do you do the same idea, instead of x , you're going to put an h of x and multiply it by the PDF h of x times f of x and take the integral from minus infinity to plus. So this is how you figure out the expected value of a function and what it would variance. So remember, our formula for variance. So formula for variance, the definition of variance is that it's the expectation of X minus the mean, holding squared. So how do you find the expectation of that, so you take x minus μ of X whole thing squared, and multiply it by the PDF, and take the integral from negative infinity to positive that's the variance. And, and we also remember, we have the easier formula for variance, which is, this is the equivalent, and we derived it for the discrete random variable, the same formula holds here as well. So the variance of x is given by expectation of X squared minus expectation of X halting squared. So that holds here as well. And variance is denoted by by this letter sigma squared. So in the case of continuous random variable, see, the additional thing here is that you need to do the integrals. So it requires to actually compute this, it requires not some knowledge of calculus. So depending on your level of comfort with calculus, right, these may be easy or hard. So let me do a relatively simple example. So let's do the uniform distribution, which has the PDF f of x is equal to $1/10$ for x between zero and 10, and f of x is zero. So how would you compute the expectation of this? So its expectation is x times f affects between negative infinity to positive. So now see, this f of x is, is positive only between zero and 10. Everywhere else, this is zero. So that means if you look at this integral, right, it's going to be zero from negative infinity up to zero. And it's also going to be zero for values of x beyond. So the only place where it really We will have some positive value is when x is between zero and 10. And there, it's going to be x times f of x , which is one term, dx of x . Now $1/10$ is a constant, so I can pull it outside. And this is going to be an integral of x dx from zero to. Now, if we look at this, this is possibly the easiest possible integral, right? It's just the integral of x . So the integral of x , if you go back and review from the calculus module of the course, is the integral of x is given by x squared over two. So this is going to be $1/10$ times x squared over two. And see this is a definite integrals, it has a lower limit of zero, upper limit of 10. So this has to be evaluated from zero. And how do we do that, so that's $1/10$, this is the constant bit, right, and this is the part that goes from zero to 10. So let's first do it a term. So 10 is 10 squared, that's 100 over two minus at zero, this is going to be zero. So this is one term, times 50. That's equal to five. So this is the expectation of the this particular random variable, which has the uniform distribution between zero. So what you've done is you've looked at the integral of x times f of x . What if we were to try and figure out its variance. So for its variance, we need to know more the expectation of x squared. So the expectation of x squared is going to be integral of x squared times f of x . So again, f of x is only positive between zero and 10. So we just put it to $10x$ squared times x . Again, if I build the one over term, the constant outside is the integral of x squared dx between zero to 10. So again, we need to know the integral of x squared. So just as

an aside, if you do integral of x to the power and dx , this is going to be x to the power and plus one over plus one. So this is again, a fact from integration. So in this case, we are looking at x squared, so n is equal to two. So that means the integral of x squared is going to be x cubed. So two plus one divided by two plus one, that's three. And again, we need to evaluate this from zero to 10. And we're nearly there. So this is $1/10$ times x cubed, so that's 10 cubed. So that's 1000 divided by three minus zero. So this is going to be 100 divided by three. So that's expectation of x squared. And now, if we want to find the variance, so we figured out so the formula, the easy formula for variance is expectation of x squared, this we already figured out is 100 over three minus expectation of X squared expectation of x we figured out is five. So this will be 100 over three minus five squared. That's 25. So this is 32.33 minus a minus 25 so this will be eight. So, this is how you would find expectation and variance for this record. Now, you can follow the same procedure for any random variable. But of course, in that case you will have to deal with you have to deal with integral in figuring this out. And how easy or difficult is to find the expectation and variance depends on how easy or difficult your PDF the f of x functions