

Lsn5

Clark

Admin

Review

We have been talking, up to this point, about Random variables. Recall that a random variable is?

Picture

If our random variable is *discrete* we say that the number of points, x where $P(X = x) > 0$ is, at most, countably infinite. In other words, the domain of $X(s)$ is countable.

Perhaps this assumption isn't appropriate for the situation we are interested in. If the domain of our random variables is \mathbb{R} we say that X is a continuous random variable. Note that we no longer can talk about $P(X = x)$ as it is always 0, but what we can say is that for any set, $A \subset \mathbb{R}$ it makes sense to talk about $P(X \in A) \geq 0$.

Regardless of whether our random variable is discrete or continuous we can define the *distribution* of X , $F(X) = P(X \leq x)$ for $x \in \mathbb{R}$. **Important notational aside, our book refers to this as the distribution function, this is different than our books definition of a probability distribution, 3.3 vs 4.1.** From the definition, we can see that regardless of whether our random variable is discrete or continuous if we have any set $A = (-\infty, a]$ where $a \in \mathbb{R}$ we can find $P(X \in A)$. If X is discrete the fact that this is a half-closed set matters if X is continuous it does not (why?)

If X is continuous, we define the distribution function of X as $f(x) = \frac{dF(x)}{dx}$. In order for our distribution function to be valid we must have

$$\int_{-\infty}^{\infty} f(x)dx = 1$$
$$f(x) \geq 0$$

A box contains five keys, only one of which will open a lock. Keys are randomly selected and tried, one at a time, until the lock is opened. Let Y be the number of the trial on which the lock is opened.

Is Y discrete or continuous?

Write out the cumulative distribution function of Y .

Draw a picture of the cumulative distribution function.

Draw a picture of the probability distribution.

Suppose that $f(y) = ky(1 - y)$ if $0 \leq y \leq 1$. What value of k makes $f(y)$ a probability density function?

Does $P(.4 \leq Y < 1) = P(.4 \leq Y \leq 1)$? Why or why not?

What is the CDF of Y ?

Extra Credit 2 pt.- Due Next Lesson Find $P(Y \leq .4 | Y \leq .8)$

Functions of Random Variables

Similar to discrete random variables, we can calculate functions of continuous random variables. For any function of Y , say $g(Y)$ then the expected value is

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$$

With a very important caveat that this integral must exist. If we recall from 205/255 that an integral is just a fancy way to do addition, we can see the connection with what we did using discrete random variables. $E[.]$ is just a function of our random variables where we either integrate or sum over the values of Y .

The temperature Y at which a thermostatically controlled switch turns on has probability density function given by $f(y) = .5\mathbb{1}_{(59,61)}$. What, in words, is this problem saying about the temperature of the thermostat?

Find $E[Y]$, $E[Y^2]$, and $E[e^Y]$

Let's sketch the CDF of Y in this instance.

Couple of things to think about. When we integrate, we are taking the limit of the Reiman sum

$$\int f(y)d(y) = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(y_i)\Delta y_i$$

If we consider a picture of the distribution of a discrete random variable we see that taking this limit on the right hand side doesn't really make sense. We still have *stuff* under our curve but to find the total amount of *stuff* we can't really use this definition of an integral. For the curious student, we can define a new integral, called the Lebesgue integral that allows us to treat both discrete random variables and continuous random variables in the same manner.

When we think about probability we need to think about measuring stuff, while the probability *stuff* that we are measuring is abstract, keep in mind that it ties back to a real, tangible problem on a sample space associated with our experiment. We don't find $E[Y]$ because we love to do integrals, we find $E[Y]$ because it tells us, if our probability model is correct, what value on average should we observe if we do an experiment. As we eventually move into inferential statistics we can use this to determine if our probability model is correct, or, in other words, does our experiment show the results that we think our experiment should.