A Brief Note on CDFs

"So, should I integrate or differentiate here?"

by Andy Palan[1]

[1] Thank you to Dibya Ghosh and Chin Chang for reading a draft of this note.

On the most recent quiz, a surprising number of students seemed to struggle with the question on the CDF.

The CDF is a very important tool in your Stat 134 toolbox and it is fundamental that you are comfortable in dealing with it. Unfortunately, in future sections, we will not have the time to return to this topic, hence why I am resorting to writing this brief note instead.

In this note, I will attempt to discuss a general approach for questions involving the CDF by using the exact question from the quiz as a guiding example.

The Question

Find the c.d.f. of X with density function

$$f_X(x) = \frac{1}{2}e^{-|x|},\ -\infty < x < \infty$$

The Approach

First, recall that the CDF of a random variable X, which we will henceforth denote F_X, is defined[2] as

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f_X(t)dt \end{aligned}$$

[2] It is fundamental you understand the distinction and relationship between a density and a CDF. Too many students confused the two in this question.

In most questions, there isn't much more to do to compute the CDF other than simply plug in the density into the formula above. Unfortunately, this question isn't quite as simple.

Let us consider the density function as outlined in the question

$$f_X(x) = \frac{1}{2}e^{-|x|},\ -\infty < x < \infty$$

The argument, x, is contained within absolute values and hence, we cannot directly evaluate the integral in its current form. We have to massage this density into a form that we can integrate easily; in other words, we want to remove the absolute value sign from the density[3].

[3] Most CDF questions will have some similar component, requiring you to massage an otherwise complicated function into a simpler form, which you can then use to compute the CDF. Without such a portion added in these questions, these questions would be much too trivial.

To do so, simply note that

$$-|x| = \begin{cases} x & x \leq 0 \\ -x & x \geq 0 \end{cases}$$

Hence, we can rewrite the density function as a piecewise function using the same principle as above

$$f_X(x) = \frac{1}{2}e^{-|x|} = \begin{cases} \frac{1}{2}e^{x} & x \leq 0 \\ \frac{1}{2}e^{-x} & x \geq 0 \end{cases}$$

As desired, our density function is now separably integrable over each of the two domains presented above.

HOWEVER, WE STILL cannot directly plug this piecewise function into the formula for the CDF. The eagle-eyed among you would have noticed that the CDF must also be similarly piecewise. Why?[4] (Try to answer this question for yourself before moving on.)

[4] Not recognizing this fact was the single biggest cause of error on this question among students.

NOTE THAT FOR values of x less than or equal to 0, we only need integrate the top expression in the piecewise function above, because for these values of x, the CDF can be simply expressed as

$$\begin{aligned} F_X(x) &= P(-\infty < X < x),\ x \leq 0 \\ &= \int_{-\infty}^{x} f(t)dt \\ &= \int_{-\infty}^{x} \frac{1}{2}e^{t}dt \end{aligned}$$

But, for values of x greater than 0, we need to integrate across the entire domain of the top expression and up to the required value of x in the lower expression. To see why this is true, note that for these values of x, the CDF is expressed as

$$\begin{aligned} F_X(x) &= P(-\infty < X < x),\ x \geq 0 \\ &= P(-\infty < X < 0) + P(0 < X < x) \\ &= \int_{-\infty}^{0} \frac{1}{2}e^{t}dt + \int_{0}^{x} \frac{1}{2}e^{-t}dt \end{aligned}$$

Now, note that there is no need to integrate to compute the first term in the expression above. Assuming that you've already computed the CDF for values of $x \leq 0$, we simply have that $P(-\infty < X \leq 0) = \int_{-\infty}^{0} \frac{1}{2}e^{t}dt = F_X(0)$.[5]

[5] Alternatively, you could have noticed that the density function was symmetric about the origin, hence we must have had that $P(-\infty < X \leq 0) = \frac{1}{2}$.

NOW THAT WE have a set approach and a few basic formulas to compute the CDF for this density, let's do that. Below is the exact solution I would have written had I been taking this quiz instead of you.

The Solution

We first write

$$f_X(x) = \begin{cases} \frac{1}{2}e^x & x \leq 0 \\ \frac{1}{2}e^{-x} & x \geq 0 \end{cases}$$

For $x \in (-\infty, 0]$:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} \frac{1}{2}e^t dt \\ &= \frac{1}{2}\left[e^t\right]_{-\infty}^{x} \\ &= \frac{1}{2}e^x \end{aligned}$$

For $x \in (0, \infty)$:

$$\begin{aligned} F_X(x) &= F_X(0) + \int_{0}^{x} \frac{1}{2}e^{-t} dt \\ &= \frac{1}{2} + \frac{1}{2}\left[e^{-t}\right]_{0}^{x} \\ &= \frac{1}{2} + \frac{1}{2}\left(1 - e^{-x}\right) \\ &= 1 - \frac{1}{2}e^{-x} \end{aligned}$$

Hence, we have that

$$F_X(x) = \begin{cases} \frac{1}{2}e^x & x \leq 0 \\ 1 - \frac{1}{2}e^{-x} & x \geq 0 \end{cases}$$

A General Approach

This is a general approach that almost always works for similar questions. Given a piecewise density, to compute the piecewise CDF, you simply have to

1. **Compute the lowest portion of the piecewise CDF as you usually would for any regular density;**

2. **For higher portions, first add the value of the CDF evaluated at the lower bound of this portion's domain;**

3. **Then integrate the density of the distribution corresponding to this domain as you usually would, but only over its own domain**.

In our example above, note that we followed this exact general approach.

1. We first computed the lowest portion of the CDF, corresponding the domain $-\infty < x \leq 0$;
2. In computing the second portion of the CDF, over the domain $0 < x < \infty$, we added $F_X(0)$ first (note that 0 is the lower bound of this portion's domain), to account for the possibility that $X \leq 0$;
3. Then, in the integration that followed, we only integrated the function over its own domain (i.e. from 0 to x). We don't have to worry about what happens below its domain (i.e. below 0), since we already added $F_X(0)$ to the expression.