

MIT OpenCourseWare
<http://ocw.mit.edu>

14.30 Introduction to Statistical Methods in Economics
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Problem Set #3

14.30 - Intro. to Statistical Methods in Economics

Instructor: Konrad Menzel

Due: Tuesday, March 3, 2009

Question One

1. Write down the definition of a cumulative distribution function (CDF). Explain what it means in words, perhaps using an example.
 - Solution to (1): One definition of the CDF is $f(\cdot) : R \mapsto [0, 1]$ where $f(x) \equiv Pr(X \leq x)$. The CDF tells us the cumulative probability up to particular point of the ordered support of the random variable, X . What this means is that we can know what the chances are that something less than or equal to (or to the left, depending on how you wish to interpret the ordering) an outcome, x , occurs.
2. Verify whether the following function is a valid CDF. If yes, draw a graph of the corresponding PDF.

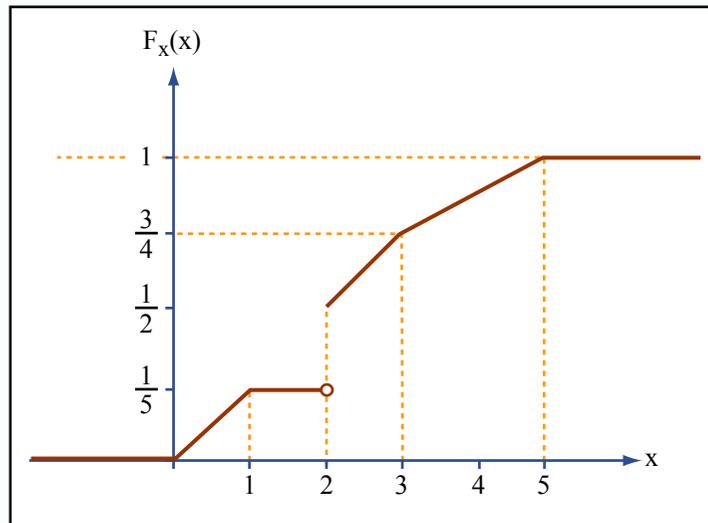


Image by MIT OpenCourseWare.

- Solution to (2): The function is in fact a valid CDF. It is bounded below by zero and above by one. It also satisfies the left and right limit conditions as $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow \infty} F_X(x) = 1$. However, it is a mixture random variable where it has a continuous distribution and a mass point at 2. The PDF

is the following equation:

$$f(x) = \begin{cases} 0 & x < 0 \cup 5 \leq x \\ \frac{1}{5} & 0 \leq x < 1 \\ 0 & 1 \leq x < 2 \\ \frac{3}{10} & x = 2 \\ \frac{1}{4} & 2 \leq x < 3 \\ \frac{1}{8} & 3 \leq x < 5 \end{cases}$$

3. Verify that the following function is a valid PDF and draw the corresponding CDF.

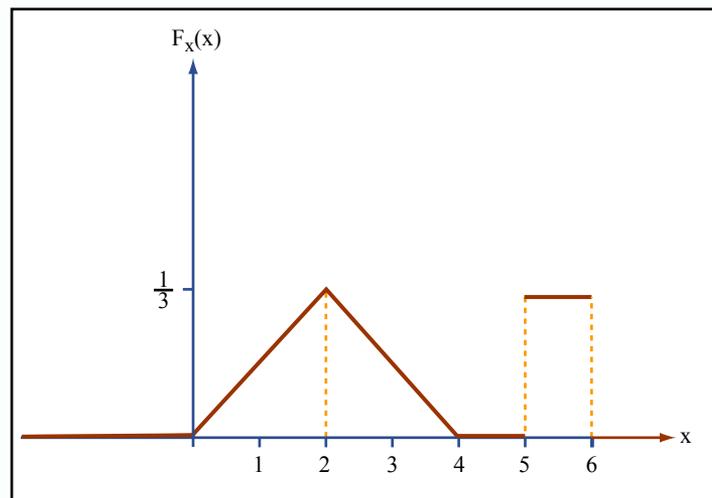


Image by MIT OpenCourseWare.

- Solution to (3): This function is, in fact, a valid PDF. It is positive everywhere and integrates to 1 (the triangle has area of $\frac{2}{3}$ and the interval from 5 to 6 has area of $\frac{1}{3}$ which together sum to 1). The CDF is straightforward. I will write the PDF and then CDF down analytically first, to make for easier integration:

$$f_X(x) = \begin{cases} 0 & x < 0 \cup 6 \leq x \\ \frac{1}{6}x & 0 \leq x < 2 \\ \frac{1}{6}(4-x) & 2 \leq x < 4 \\ 0 & 4 \leq x < 5 \\ \frac{1}{3} & 5 \leq x < 6 \end{cases} \quad F_X(x) = \begin{cases} 0 & x < 0 \cup 6 \leq x \\ \frac{1}{12}x^2 & 0 \leq x < 2 \\ \frac{2}{3} - \frac{1}{12}(4-x)^2 & 2 \leq x < 4 \\ \frac{2}{3} & 4 \leq x < 5 \\ \frac{2}{3} + \frac{1}{3}(x-5) & 5 \leq x < 6 \end{cases}$$

Drawing this curve is relatively straightforward, at least if you pay little attention to detail as I am not a graphic designer:

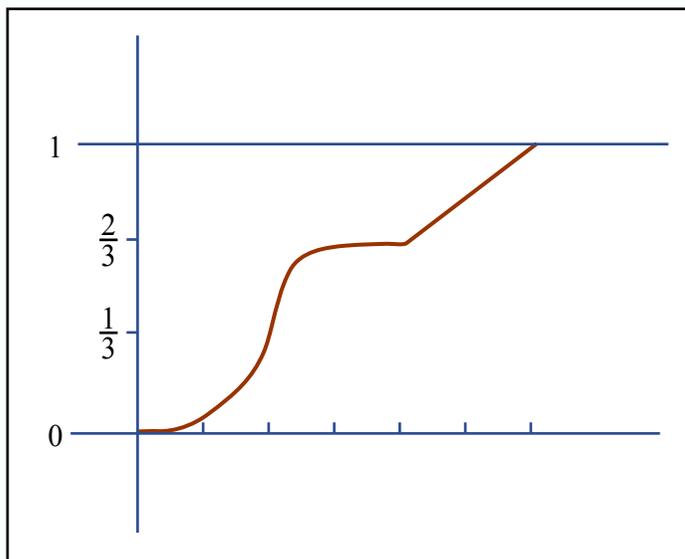


Image by MIT OpenCourseWare.

Question Two

1. Give a p.d.f. whose c.d.f. is not continuous but is right-continuous.

- Solution to (1): This will have to come from a distribution with at least one mass point (or it could be a completely discrete distribution). Konrad's lecture notes have an example of the CDF:

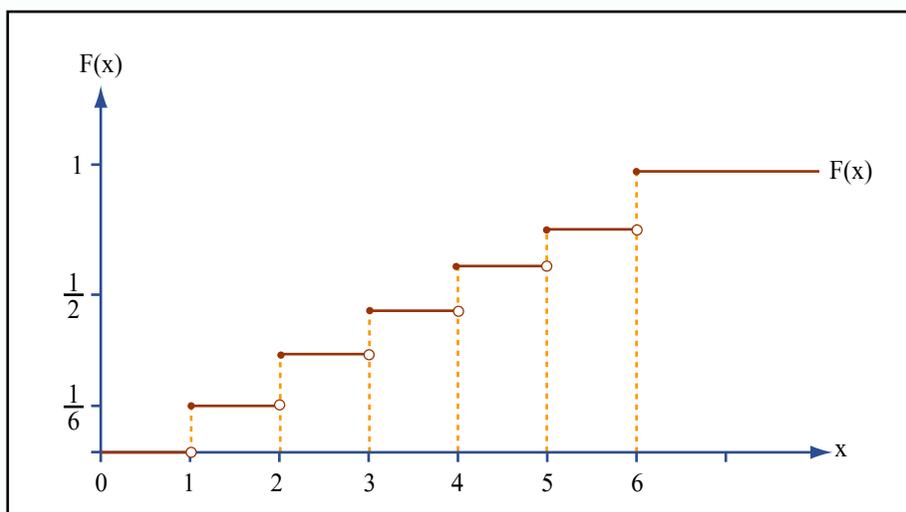


Image by MIT OpenCourseWare.

The PDF which corresponds to this graph is simply the PDF generated by the random variable X defined as the value on the face of a single roll of a die. The PDF follows:

$$f(x) = \begin{cases} \frac{1}{6} & \text{if } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}.$$

True/false/uncertain: Always give a brief explanation if the statement is true, or counter-examples and a short explanation of the counter-examples if the statement is false or uncertain.

1. If $P(A|B) > P(A)$ and $P(A|C) > P(A)$, then $P(A|B, C) > P(A)$.

- Solution to (1): False. Just because two conditional probabilities are large, does not mean that their joint probability will not be large. One example is the probability of getting sick (event A), given it is winter time (event B). You are more likely to get sick during winter than the average during the year. You are also more likely to get sick, conditional on taking medication that you don't need (randomly-event C). However, the probability of getting sick is likely lower if you take preventive medications only during the peak times of the winter when everyone else is getting sick. The same thing applies to flu vaccines (high probability of getting sick if you get a shot, high probability of getting sick during the winter, potentially lower interaction probability).

2. A continuous p.d.f. can never take a value greater than 1.

- Solution to (2): False. An example of a continuous PDF would be the uniform PDF defined as $f(x) = 6$ if $0 \leq x < \frac{1}{6}$ and zero elsewhere. This PDF clearly is properly defined and has a value greater than one over an uncountable set of outcomes. Further, this statement would be more properly applied to CDFs.

3. $P(A) = P(A|B)P(B)$ means that A and B are independent.

- Solution to (3): False. This means that A and B are actually correlated as the law of total probability gives us that

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

So, what this means is that either $P(A|B^c) = 0$ (which means that A only occurs if B does which would imply that they cannot be independent) or $P(B^c) = 0$ (which means that $B = S$, the sample space and then A and B would be independent). A simple example of A and B would be, in the space of a single roll of a die, $A = \{\text{Roll a 4}\}$ and $B = \{\text{Roll an even number}\}$. In this example, $P(A) = \frac{1}{6}$, $P(B) = \frac{1}{2}$, and $P(A|B) = \frac{1}{3}$. However, $P(A \cap B) = \frac{1}{6} \neq P(A)P(B) = \frac{1}{6} \cdot \frac{1}{2}$. However, if $B = S$, then, it turns out that since $P(S) = 1$, we actually do have that $P(A|S)P(S) = P(A)P(S) = P(A) \cdot 1 = P(A)$ which means both independence and the condition in the question both hold simultaneously. Still, this is not that interesting of a result. :)

Question Three

(Source: Bain/Engelhardt, Ch. 2, ex. 8)

A nonnegative integer-valued random variable X has a CDF of the form $F(x) = 1 - (1/2)^{x+1}$ for $x = 0, 1, 2, \dots$ and zero if $x < 0$.

1. Find the pdf of X .

- Solution to (1): The PDF is the difference in CDF at each point. Since this is an integer random variable, we need to compute the difference in probabilities of the adjacent discrete outcomes:

$$\begin{aligned}F(x) - F(x - 1) &= (1 - (1/2)^{x+1}) - (1 - (1/2)^{x-1+1}) \\ &= (1/2)^x(1 - (1/2)) \\ f(x) &= \frac{1}{2^{x+1}}\end{aligned}$$

Also, the PDF is such that $f(x) = 0$ for anything that is not a nonnegative integer.

2. Find $P[10 < X \leq 20]$.

- Solution to (2): For those of you who have seen the geometric series before, you quickly recognized that this is the just a finite difference in sums:

$$P[10 < X \leq 20] = \sum_{x=11}^{20} \frac{1}{2^{x+1}} = \frac{\frac{1}{2^{11+1}} - \frac{1}{2^{20+1+1}}}{1 - \frac{1}{2}} = \frac{1}{2^{11}} - \frac{1}{2^{21}}$$

Or, since we already have the CDF, you can just use that:

$$\begin{aligned}P[10 < X \leq 20] &= F(20) - F(10) \\ &= \left(1 - \frac{1}{2^{21}}\right) - \left(1 - \frac{1}{2^{11}}\right) \\ &= \frac{1}{2^{11}} - \frac{1}{2^{21}}\end{aligned}$$

3. Find $P[X \text{ is even}]$.

- Solution to (3): Since we just want even numbers, we just need to compute it for ever other number. We can do this easily by just multiplying each x by 2 which will allow us to skip all of the even numbers in the PDF. We then just need to compute the infinite sum:

$$\begin{aligned}P[X \text{ is even}] &= \sum_{x=0}^{\infty} \frac{1}{2^{2x+1}} \\ &= \frac{1}{2} \sum_{x=0}^{\infty} \frac{1}{4^x} \\ &= \frac{1}{2} \frac{1}{1 - \frac{1}{4}} \\ P[X \text{ is even}] &= \frac{2}{3}.\end{aligned}$$

Question Four

1. Suppose that a random variable has a PDF that is proportional to x on the interval $[0, 1]$. Write down a formula for this PDF. What is the corresponding CDF?

- Solution to (1): The PDF is proportional to x :

$$f(x) = kx \text{ if } x \in [0, 1]$$

and zero otherwise. By the integration property of PDFs, we know that $k = 2$. The CDF is just the integral of the PDF:

$$F(x) = x^2 \text{ if } x \in [0, 1].$$

For completeness we define the CDF as $F(x) = 1$ if $x > 1$ and $F(x) = 0$ if $x < 0$.

2. Now suppose that the random variable has a CDF that is proportional to x on the interval $[0, 1]$. Write down a formula for this CDF. What is the corresponding PDF?

- Solution to (2): The CDF is proportional to x :

$$F(x) = kx \text{ if } x \in [0, 1]$$

which, by the properties of CDFs means that $k = 1$ to ensure that the CDF is bounded by zero and one. We also define the CDF as $F(x) = 1$ if $x > 1$ and $F(x) = 0$ if $x < 0$ for completeness (don't forget to do this on exams and problem sets—you need to define PDFs and CDFs on the whole support of the real numbers). The corresponding PDF would be the derivative:

$$f(x) = 1 \text{ if } x \in [0, 1]$$

and zero elsewhere. This is the uniform PDF.

Question Five

Suppose that the joint PDF of X and Y is given by

$$f_{X,Y} = \begin{cases} kx^3y & \text{for } 0 < x < y < 1 \\ 0 & \text{elsewhere} \end{cases}.$$

1. What is the value of k ?

- Solution to (1): For $f_{X,Y}$ to be a PDF, the value of k must ensure that the density integrates to 1:

$$\begin{aligned}
 \int_{X \times Y} f_{X,Y} dy dx &= 1 \\
 \int_0^1 \int_x^1 kx^3 y dy dx &= \\
 \int_0^1 kx^3 \left[\frac{1}{2}y^2 \right]_x^1 dx &= \\
 \frac{1}{2}k \int_0^1 x^3 (1 - x^2) dx &= \\
 \frac{1}{2}k \int_0^1 (x^3 - x^5) dx &= \\
 \frac{1}{2}k \left(\frac{1}{4}x^4 - \frac{1}{6}x^6 \right) \Big|_0^1 &= \\
 k &= 24
 \end{aligned}$$

So, k must be equal to 24 for the density to integrate to 1 and, thus, be a proper probability density function.

2. What is the marginal PDF, $f_X(x)$, of x ?

- Solution to (2): The definition of the marginal PDF is

$$f_X(x) = \int_Y f(x, y) dy.$$

In words, to obtain the marginal of x , all we have to do is integrate out the other random variables in the joint density, which means we integrate out y in this case. We already performed this integral in the previous problem, so I will simply write down the solution from the work above:

$$\begin{aligned}
 f_X(x) &= \int_Y f(x, y) dy \\
 &= 12x^3 (1 - x^2).
 \end{aligned}$$

We could also solve for the marginal PDF of y , which would force us to do a different integral, but I leave that exercise undone here.

3. What is the value of the marginal cdf of x , $F_X(x)$, at $x = \frac{1}{2}$?

- Solution to (3): The marginal CDF of x is defined as the CDF of the marginal PDF or

$$F_X(x) = \int_{-\infty}^x f_X(\tilde{x}) d\tilde{x}.$$

So, by integrating this PDF, which we already did in part (1) of this question, we get:

$$\begin{aligned} F_X(x) &= 12 \left(\frac{1}{4} \tilde{x}^4 - \frac{1}{6} \tilde{x}^6 \right)_0^x \\ &= 12 \left(\frac{1}{4} x^4 - \frac{1}{6} x^6 \right) \\ F_X(x) &= 6x^4 \left(\frac{1}{2} - \frac{1}{3} x^2 \right). \end{aligned}$$

Which, when we evaluate this at $\frac{1}{2}$, we obtain $F_X\left(\frac{1}{2}\right) = 6\left(\frac{1}{2}\right)^4 \left(\frac{1}{2} - \frac{1}{3}\left(\frac{1}{2}\right)^2\right) = \frac{3}{16} \cdot \frac{5}{6} = \frac{5}{32}$.

4. What is the conditional PDF of y (conditional on x , i.e. $f(y|x)$)? Are X and Y independent? Explain.

- Solution to (4): We again write down the definition of the conditional distribution (which is just a simple extension of the definition of conditional probability):

$$f(y|x) = \frac{f(x, y)}{f_X(x)}.$$

We clearly have all parts of this equation. Interestingly enough, though, we didn't need the integration constant, k , to obtain the conditional distribution. This is relevant for many Bayesian applications and advanced statistical methods such as Metropolis-Hastings Markov Chain Monte Carlo. You'll learn more about its relevancy later, but here we will use the densities with k just to illustrate:

$$\begin{aligned} f(y|x) &= \frac{f(x, y)}{f_X(x)} \\ &= \frac{kx^3y}{\frac{1}{2}kx^3(1-x^2)} \\ f(y|x) &= \frac{2y}{1-x^2}. \end{aligned}$$

Again, for the interested reader, computing $f(x|y)$ is also possible upon obtaining the marginal of y , $f_Y(y)$. Then, we could either compute the conditional distribution directly or use Bayes Rule:

$$f(x|y) = \frac{f(y|x)f_X(x)}{f_Y(y)}.$$

We also must remember to answer the second part of the question: Are X and Y independent? It turns out that they are not independent since the conditional distribution of y is still a function of x . While at first glance it appears that we should be able to factor out the marginals from the joint density via the "factoring

rule,” we must remember that it only applies when the support of our joint distribution is rectangular. But, computing the marginals is the easiest way to see that it does, in fact, matter. While it is possible to have a carefully constructed non-rectangular support and still have two independent random variables composing the joint density, in our case it doesn't work out that way.

5. What is the probability that $X + Y < 1$?

- Solution to (5): We would like to integrate over the region where $X + Y < 1$. We can perform this many ways, but I'm going to just compute it directly. One thing to remember is that we need to be careful with our integration bounds. When integrating y out, we must remember that $x < y < 1 - x$ where the first inequality is from the definition of the PDF and the second inequality is from the condition that $X + Y < 1 \Rightarrow Y < 1 - X$. The combination of the two inequalities reminds us that $x < \frac{1}{2}$ is also a condition we must put on x when integrating. We now perform the integration:

$$\begin{aligned}
 P(X + Y < 1) &= 24 \int_0^{\frac{1}{2}} \int_x^{1-x} x^3 y dy dx \\
 &= 24 \int_0^{\frac{1}{2}} x^3 \left[\frac{1}{2} y^2 \right]_x^{1-x} dx \\
 &= 12 \int_0^{\frac{1}{2}} x^3 (1 - 2x + x^2 - x^2) dx \\
 &= 12 \int_0^{\frac{1}{2}} dx \\
 &= 12 \left[\frac{1}{4} x^4 - \frac{2}{5} x^5 \right]_0^{\frac{1}{2}} \\
 &= \frac{12}{2^4} \left[\frac{1}{4} - \frac{1}{5} \right] \\
 &= \frac{3}{80}
 \end{aligned}$$

which is fortunately less than 1. :) Also, an additional check would be to notice that the form of the density would put a lot more mass close to $(1, 1)$, and we only integrated out the smaller half of the triangular region bounded by $0 < x < y < 1$.

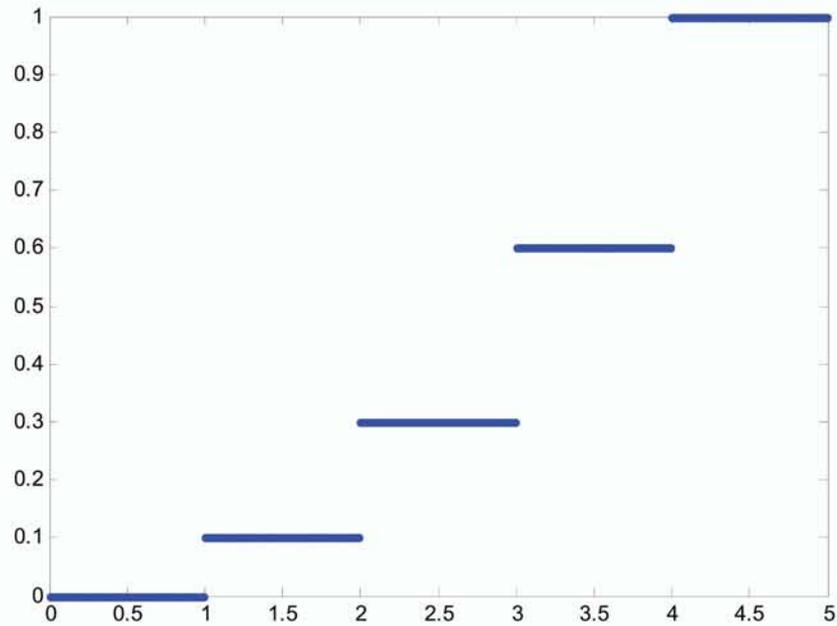
Question Six

(Bain/Engelhardt, Ch. 2, ex. 10)

Let X be a discrete random variable such that $P[X = x] = 0$ otherwise. Suppose the CDF is $F(x) = .05x(1 + x)$ at the values $x = 1, 2, 3,$ or 4 .

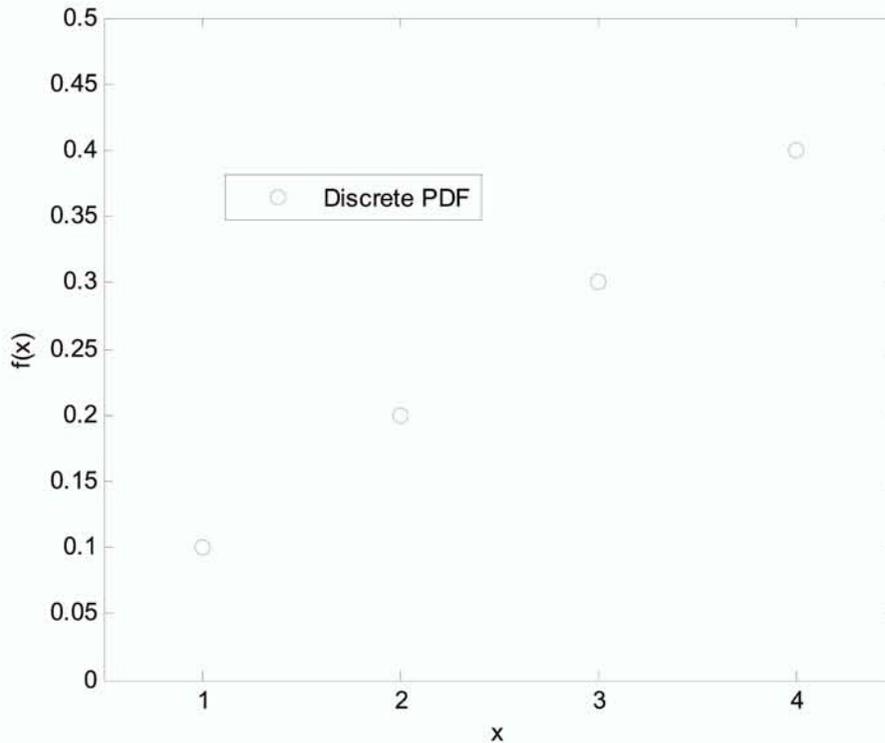
1. Sketch the graph of the CDF.

- Solution to (1): We can plot this discrete CDF in a program like Matlab by using the floor $\lfloor x \rfloor$ operator: $F(x) = .05 \lfloor x \rfloor (1 + \lfloor x \rfloor)$. Here is the plot of the CDF on the interval $[0, 5]$:



2. Sketch the graph of the discrete pdf, $f(x)$.

- Solution to (2): The PDF is simple to plot as well. We just want to use the difference where we put mass points at 1, 2, 3, and 4, and recognize that it is a discrete PDF:



3. Write down the definition of $E[X]$ and find $E[X]$.

- Solution to (3): The definition of $E[X]$ is different for discrete and continuous random variables, although they are related via a limiting argument. I write them both down for completeness, even though for this problem we will only be using the discrete one:

$$\text{Continuous: } E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$\text{Discrete: } E[X] = \sum_{x \in X} xf(x)$$

where $f(x)$ is the PDF of the random variable X . In this problem, we have a simple sum:

$$\begin{aligned} E[X] &= \sum_{x \in X} xf(x) \\ &= 1 \cdot .1 + 2 \cdot .2 + 3 \cdot .3 + 4 \cdot .4 \\ E[X] &= 3 \end{aligned}$$