

Homework 05: CRV's, PDF's, CDF's


Name: Julia Troni

This assignment is due on Canvas by **6:00PM on Friday September 30**. Your solutions to theoretical questions should be done in Markdown directly below the associated question. Your solutions to computational questions should include any specified Python code and results as well as written commentary on your conclusions. Remember that you are encouraged to discuss the problems with your classmates, but **you must write all code and solutions on your own**.

NOTES:

- Any relevant data sets should be available in the Homework 01 assignment write-up on Canvas. To make life easier on the grader if they need to run your code, do not change the relative path names here. Instead, move the files around on your computer.
 - If you're not familiar with typesetting math directly into Markdown then by all means, do your work on paper first and then typeset it later. Remember that there is a [reference guide \(https://math.meta.stackexchange.com/questions/5020/mathjax-basic-tutorial-and-quick-reference\)](https://math.meta.stackexchange.com/questions/5020/mathjax-basic-tutorial-and-quick-reference) linked on Canvas on writing math in Markdown. **All** of your written commentary, justifications and mathematical work should be in Markdown.
 - Because you can technically evaluate notebook cells in a non-linear order, it's a good idea to do **Kernel → Restart & Run All** as a check before submitting your solutions. That way if we need to run your code you will know that it will work as expected.
 - It is **bad form** to make your reader interpret numerical output from your code. If a question asks you to compute some value from the data you should show your code output **AND write a summary of the results** in Markdown directly below your code.
 - This probably goes without saying, but... For any question that asks you to calculate something, you **must show all work and justify your answers to receive credit**. Sparse or nonexistent work will receive sparse or nonexistent credit.
 - **All the solutions should be solved by hand instead of using python in-built functions or external libraries unless mentioned explicitly.**
-

Import Pandas, NumPy, and matplotlib.pyplot

```
In [1]:  # Per the standard import pandas as 'pd', numpy as 'np', and matplotlib as 'p  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
%matplotlib inline  
import math  
    # 'inline' puts your graph in the cell versus a new popup window
```

Problem 1

Suppose X is a continuous random variable with a probability density function given by $f(x) = 3x^2$ over $(a, 2)$.

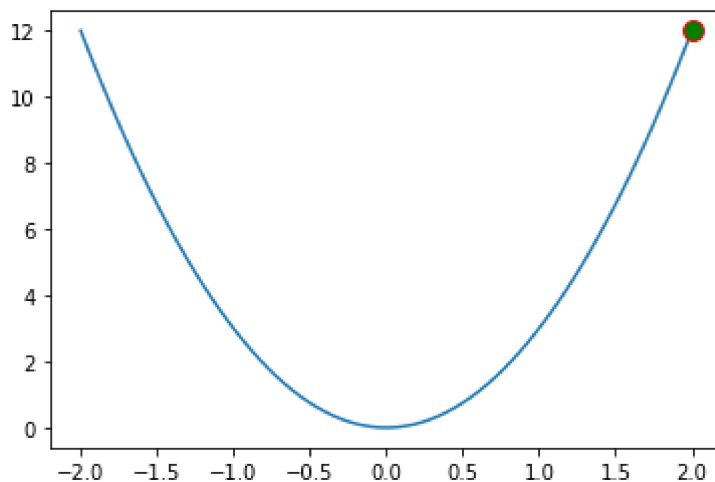
```
In [2]: # This is just a visual for the PDF

x = np.linspace(-2,2,100)
y = 3*x**2

plt.plot(x,y)

plt.plot(2,12, marker="o", markersize=10, markeredgecolor="red", markerfacecolor="green")
```

Out[2]: [



Part A

(4 points) What is a such that $f(x)$ is a valid PDF?

solution: Put your solution to Part A here:

- First check: $f(x) = 3x^2 \geq 0$ for all x
- Now check $\int_a^2 f(x)dx = 1$

$$\int_a^2 3x^2 dx = x^3 \Big|_a^2 = 2^3 - a^3 = 1$$

$$\text{So } a = \sqrt[3]{7} = 1.9129$$

$$\text{Thus } PDF = f(x) = \begin{cases} 3x^2 & 1.913 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Part B

(2 points) What is $P(X = 1.95)$?

solution: Put your solution to Part B here:

For any exact value, c , the $P(X = c) = 0$ so $P(X = 1.95) = 0$

Part C

(2 points) What is $P(X < 2)$?

solution: Put your solution to Part C here:

Since the PDF in part A satisfies $\int_a^b f(x)dx = 1$, then we know that $P(X < 2) = 1$

Part D

(3 points) What is $P(1.95 < X < 2)$?

solution: Put your solution to Part D here:

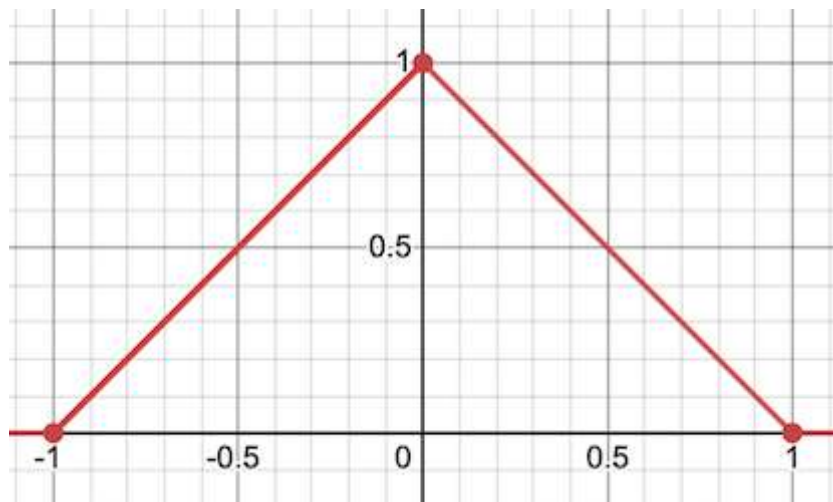
$$P(1.95 < X < 2) = \int_{1.95}^2 3x^2 dx = x^3 \Big|_{1.95}^2 = 2^3 - 1.95^3 = 0.585$$

```
In [3]:  a= 2**(3)
        b= 1.95**(3)
        a-b
```

```
Out[3]: 0.5851250000000006
```

Problem 2

Suppose the PDF of a continuous random variable, X , appeared as follows:



Part A

(4 points) Write the CDF as an explicit function.

solution: Put your solution to Part A here:

$$\text{Since } PDF = f(x) = \begin{cases} x + 1 & -1 \leq x \leq 0 \\ 1 - x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Then } CDF = F(x) = \int_a^b f(x)dx = 1$$

must be defined over four intervals

1. $x \leq -1$ Since no probability accumulates over that interval, $\int_{-\infty}^{-1} 0dx = 0$

2. For $-1 < x \leq 0$ $F(x) = \int_{-1}^x (t + 1)dt = \frac{1}{2}t^2 + t \Big|_{-1}^x = \frac{1}{2}x^2 + x + \frac{1}{2}$

3. For $0 < x \leq 1$ since we know that at $F(0) = \frac{1}{2}x^2 + x + \frac{1}{2} \Big|_{x=0} = \frac{1}{2}$, Then over

$$0 < x < 1, F(x) = F(0) + \int_0^x (1 - t)dt = \frac{1}{2} + \left(t - \frac{1}{2}t^2\right) \Big|_0^x = \frac{1}{2} + x - \frac{1}{2}x^2$$

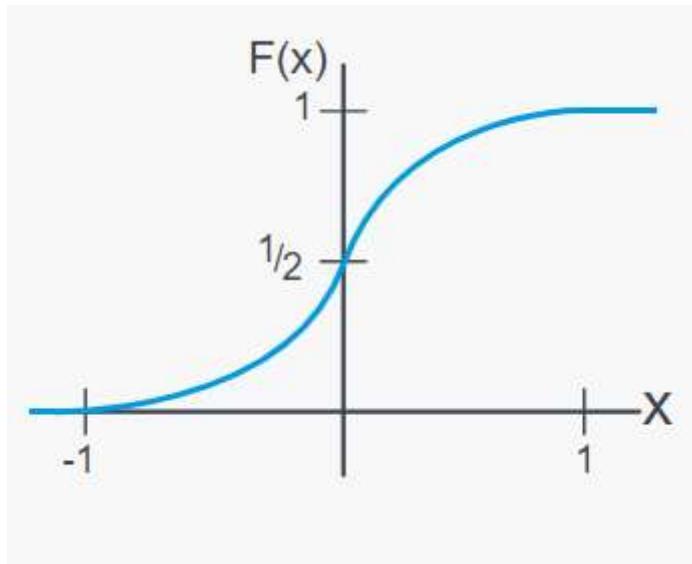
4. For $x \geq 1$ Since all of the probability has been accumulated for beyond 1,
 $F(1) = P(x \geq 1) = 0$

$$\text{Thus } CDF = F(x) = \begin{cases} 0 & -\infty < x \leq -1 \\ \frac{1}{2}x^2 + x + \frac{1}{2} & -1 < x \leq 0 \\ \frac{1}{2} + x - \frac{1}{2}x^2 & 0 < x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

Part B

(3 points) Graph $F(x)$.

solution: Put your solution (drawn, attached, or coded) to Part B here:



Part C

(2 points) What is $P(-\frac{1}{2} < X < 0)$?

solution:

$$P(-\frac{1}{2} < X < 0) = F(0) - F(-\frac{1}{2}) = \left(\frac{1}{2} + x - \frac{1}{2}x^2\right)\Big|_{x=0} - \left(\frac{1}{2}x^2 + x + \frac{1}{2}\right)\Big|_{x=-\frac{1}{2}}$$

$$= \frac{1}{2} - \frac{1}{8} = 0.375$$

In [4]: `1/2-1/8`

Out[4]: 0.375

Problem 3

Suppose X is a continuous random variable with probability density function given by $f(x) = \frac{1}{4}$ for $5 < x < 9$.

Part A

(2 points) What is the median? Justify your answer using the density functions.

solution:

$$X \sim U(5, 9) \text{ So the CDF is } F(x) = \int_5^x \left(\frac{1}{4}\right) dt = \left[\frac{1}{4}t\right]_{t=5}^{t=x} = \frac{1}{4}(x - 5)$$

The median is the 50th percentile so we want to find $\pi_{0.5}$

$$0.5 = F(\pi_{0.5}) = \frac{1}{4}(\pi_{0.5} - 5)$$

$$\pi_{0.5} = 7$$

Part B

(2 points) What is the 3rd quartile?

solution:

The 3rd quartile is 75% so we want to find $\pi_{0.75}$

$$0.75 = F(\pi_{0.75}) = \frac{1}{4}(\pi_{0.75} - 5)$$

$$\pi_{0.75} = 8$$

Part C

(3 points) What is the 80th percentile?

solution:

The 8th percentile is 80% so we want to find $\pi_{0.8}$

$$0.8 = F(\pi_{0.8}) = \frac{1}{4}(\pi_{0.8} - 5)^2$$

$$\pi_{0.8} = 8.2$$

Problem 4

Let X be a continuous random variable with the following probability density function:

$$f(x) = \frac{1}{2}(x + 1) \text{ for } -1 < x < 1.$$

(4 points) What is the 64th percentile of X ?

solution: Put your answer to Problem 4 here:

$$F(x) = \frac{1}{2} \int_{-1}^x (t + 1) dt = \frac{1}{2} \left[\frac{(t + 1)^2}{2} \right]_{t=-1}^{t=x} = \frac{1}{4}(x + 1)^2$$

$$0.64 = F(\pi_{0.64}) = \frac{1}{4}(\pi_{0.64} + 1)^2$$

$$2.56 = (\pi_{0.64} + 1)^2$$

$$\pi_{0.64} = -2.6 \text{ or } 0.60$$

So because $-1 < x < 1$ we know the 64th percentile must be 0.6

Problem 5

Consider the PDF named P below:

$$P(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Part A

(3 points) Write a function that takes x , σ and μ as inputs.

The function should return the value of P .

```
In [5]: ▶ def pdf_norm(x,mu,sig):
        piroot= (2*np.pi)
        denom= sig*(piroot**.5)
        exp= -0.5*((x-mu)/sig)**2
        num= (math.e)**(exp)
        ans= (num/denom)
        print("The probability of {} with mean of {} and std {} is : {}".
              format(x,mu,sig,ans))
        return ans
```

Part B

Suppose $\mu = 0$ and $\sigma = 1$

- 1] (1 point) Use your function to find the height of the graph at $x = 1$.
- 2] (1 point) Use your function to find the height of the graph at $x = 1.1$.
- 3] (1 point) Use your function to find the height of the graph at $x = 1.2$.

Note : Do not round off your answers.

```
In [6]: ▶ # Solution to Part B, #1 here:
        pdf_norm(1,0,1)
```

The probability of 1 with mean of 0 and std 1 is : 0.24197072451914337

Out[6]: 0.24197072451914337

```
In [7]: ▶ # Solution to Part B, #2 here:
pdf_norm(1.1,0,1)
```

The probability of 1.1 with mean of 0 and std 1 is : 0.21785217703255055

Out[7]: 0.21785217703255055

```
In [8]: ▶ # Solution to Part B #3 here:
pdf_norm(1.2,0,1)
```

The probability of 1.2 with mean of 0 and std 1 is : 0.19418605498321298

Out[8]: 0.19418605498321298

Part C

Still using $\mu = 0$ and $\sigma = 1$.

(4 points) Use your function to find the values of $P(x)$ for $x \in [1, 3]$ in steps of 0.2. Hint: use an array; do NOT enter all values by hand.

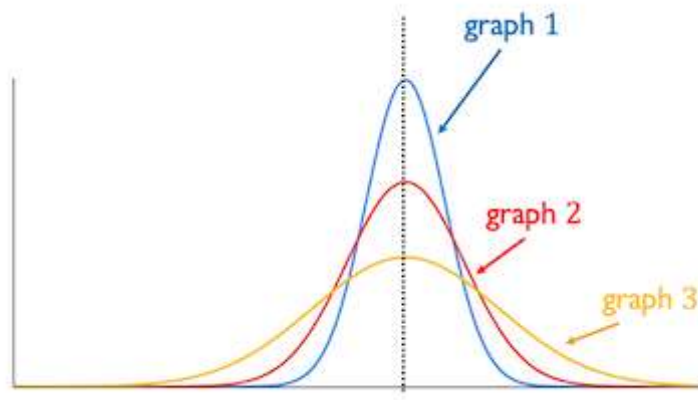
```
In [9]: ▶ # Solution to Part B here
vals= np.arange(1,3.2,0.2)
for i in vals:
    pdf_norm(i,0,1)
```

The probability of 1.0 with mean of 0 and std 1 is : 0.24197072451914337
The probability of 1.2 with mean of 0 and std 1 is : 0.19418605498321298
The probability of 1.4 with mean of 0 and std 1 is : 0.1497274656357449
The probability of 1.5999999999999999 with mean of 0 and std 1 is : 0.1109208346794556
The probability of 1.7999999999999998 with mean of 0 and std 1 is : 0.0789501583008942
The probability of 1.9999999999999998 with mean of 0 and std 1 is : 0.053990966513188084
The probability of 2.1999999999999997 with mean of 0 and std 1 is : 0.03547459284623146
The probability of 2.3999999999999995 with mean of 0 and std 1 is : 0.022394530294842934
The probability of 2.5999999999999996 with mean of 0 and std 1 is : 0.013582969233685634
The probability of 2.8 with mean of 0 and std 1 is : 0.00791545158297997
The probability of 2.9999999999999996 with mean of 0 and std 1 is : 0.004431848411938013

Part D

The graphs below are probability density functions produced by P :

Different input values create the different graphs.



(2 points) How do the sizes of σ compare? i.e. which graph has the smallest σ and which has the largest?

solution:

When the data points are more spread out, the standard deviation, σ , will be larger. Thus the curve with the larger standard deviation will have a wider (more spread out) and shorter graph. Thus $\sigma_3 > \sigma_2 > \sigma_1$

Part E

(2 points) Referring to the same graphs mentioned/pictured in Part C, how do the sizes of μ compare?

solution:

Since normal curves are symmetric about their mean, μ , and since all 3 graphs are symmetric about the same position then $\mu_1 = \mu_2 = \mu_3$

Rubric Check

(5 points) Make sure your answers are thorough but not redundant. Explain your answers, don't just put a number. Make sure you have matched your questions on Gradescope. Make sure your PDF is correct and your LaTeX is correct. etc. etc. BE NEAT.