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

Name: CJ Kennedy

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 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 is a non-linear order, it's a good idea to do $\text{Kernel} \rightarrow \text{Restart \& Run All}$ 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** $\text{write a summary of the results}$ 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.pylab

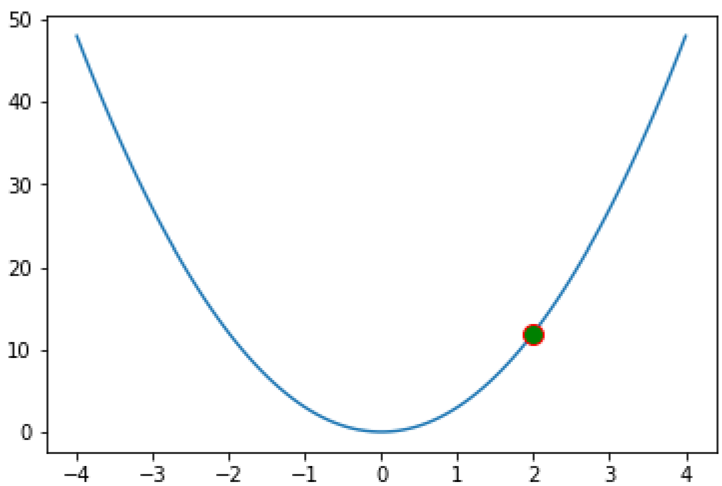
In [1]:
```
# Per the standard import pandas as 'pd', numpy as 'np', and matplotlib as 'plt'
import pandas as pd
import numpy as np
import matplotlib.pylab as plt
%matplotlib inline
    # 'inline' puts your graph in the cell versus a new popup window
```

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(-4,4,100)
y = 3*x**2

plt.plot(x,y)

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

print(3*np.sqrt(11/3)**2)
```

```
11.0
```



Part A

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

solution:

A valid PDF would entail:

$\int_a^2 (3x^2)dx = 1$

$\Rightarrow |_a^2 (x^3) = 1$

$(8) - (a)^3 = 1$

$7 = a^3$

$a = 7^{\frac{1}{3}}$

Part B

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

solution:

$P(X = 1.95) = 3(1.95^2) = 11.4075$

Part C

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

solution:

$P(X < 2) = \int_a^2 f(x)dx$

$= \int_a^2 3x^2 dx = |_a^2 x^3$

$= (2^3) - (7^{\frac{1}{3}})^3$

$= (8 - 7) = 1$

Part D

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

solution:

$P(1.95 < X < 2) = \int_{1.95}^2 f(x)dx$

$= \int_{1.95}^2 3x^2 dx = |_{1.95}^2 x^3$

$= (2^3) - (1.95)^3$

$= (8 - 7.414875)$

$\therefore P(1.95 < X < 2) = 0.585125$

Problem 2

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

title

Part A

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

solution: Put your solution to Part A here:

First, the PDF over intervals (-1, 0) is $1+x$ while over (0,1) it is $1-x$. Furthermore, before -1 and after 1, the values are 0. Explicitly define $f(x)$:

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

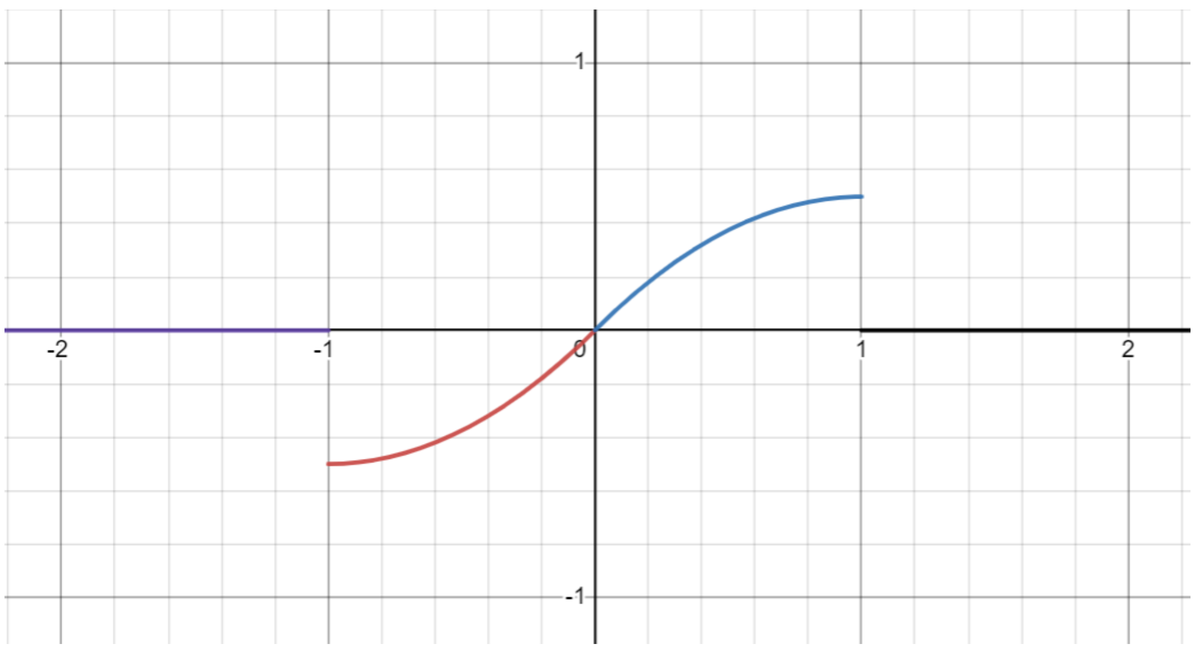
Note, the "otherwise" applies to $x > 1$ or $x < -1$. Then, integration to find $F(x)$:

$$F(x) = \begin{cases} x+\frac{x^2}{2} & \text{if } -1 \leq x \leq 0 \\ x-\frac{x^2}{2} & \text{if } 0 < x \leq 1 \\ 0 & otherwise \end{cases} \tag{2}$$

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: Put your solution to Part C here:

Noting the bounds:

$F(x) = x + \frac{x^2}{2}$

Evaluation:

$\Rightarrow F(x)|_{-\frac{1}{2}}^{0}$

$= (0) - (-\frac{1}{2} + \frac{1}{2}(-\frac{1}{2})^2)$

$= \frac{3}{8}$

$\therefore P(-\frac{1}{2} < X < 0) = \frac{3}{8}$

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: Put your solution to Part A here:

$F(x) = \int_5^x f(x)du = \int_5^x \frac{1}{4} du$

$F(x) = |_5^x \frac{u}{4} = \frac{x}{4} - \frac{5}{4}$

The median, $\pi_{.50}$, can be solved for with:

$P(X \leq m) = F(x) == .5$

$\Rightarrow \frac{x}{4} - \frac{5}{4} = \frac{1}{2}$

$\Rightarrow \frac{x}{4} = \frac{7}{4}$

$\therefore x = 7$

The median is 7.

Part B

(2 points) What is the 3rd quartile?

solution:

We know $F(x) = |_5^x \frac{u}{4} = \frac{x}{4} - \frac{5}{4}$.

The 3rd quartile can be represented as $\pi_{.75}$:

$P(X \leq m) = F(x) == .75$

$\Rightarrow \frac{x}{4} - \frac{5}{4} = \frac{3}{4}$

$\Rightarrow \frac{x}{4} = \frac{8}{4}$

$\therefore x = 8$

The 3rd quartile is 8.

Part C

(3 points) What is the 80th percentile?

solution:

$F(x) = \frac{x}{4} - \frac{5}{4}$.

The 80th percentile, $\pi_{.80}$, can be solved for with:

$P(X \leq m) = F(x) == .80$

$\Rightarrow \frac{x}{4} - \frac{5}{4} = \frac{8}{10}$

$\Rightarrow \frac{x}{4} = \frac{41}{20}$

$\therefore x = \frac{41}{5}$

The 80th percentile is 8.2.

Problem 4

Let X be a continuous random variable with the following probability density function:
$f(x) = \frac{1}{2}(x+1)$ for $-1 < x < 1$.

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

solution: Put your answer to Problem 4 here:

$f(x) = \frac{1}{2}(x+1)$

$F(x) = \int_{-1}^{x} f(x)du$

$F(x) = \frac{1}{2}(u+1)du$

$F(x) = \frac{u^2}{4} + \frac{u}{2}|_{-1}^{x}$

$F(x) = (\frac{x^2}{4} + \frac{x}{2}) - (\frac{(-1)^2}{4} + \frac{-1}{2})$

$F(x) = (\frac{x^2}{4} + \frac{x}{2}) - (\frac{1}{4} - \frac{2}{4})$

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

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

The 64th percentile, $\pi_{.64}$, can be solved for with:

$\Rightarrow F(x) = .64$

$\frac{1}{4}(x^2 + 2x + 1) = .64$

$(x^2 + 2x + 1) = 2.56$

Solving for x:

$x^2 + 2x - 1.56 = 0$

$x = -2.6, 0.6$

Obviously, the negative number is ignored.

So, the 64^{th} percentile of X is 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 [3]:
```
# code your function for Part A here (Allowed to use python libraries):
def PDF(x,mu,sigma):
    P = 1/(sigma*np.sqrt(2*np.pi))*np.exp((-1/2)*((x-mu)/sigma)**2)
    return P
```

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 [4]:
```
# Solution to Part B, #1 here:
print("P(𝑥=1)=",PDF(1,0,1))
```

```
P(𝑥=1)= 0.24197072451914337
```

In [5]:
```
# Solution to Part B, #2 here:
print("P(𝑥=1.1)=",PDF(1.1,0,1))
```

```
P(𝑥=1.1)= 0.21785217703255053
```

In [6]:
```
# Solution to Part B #3 here:
print("P(𝑥=1.2)=",PDF(1.2,0,1))
```

```
P(𝑥=1.2)= 0.19418605498321295
```

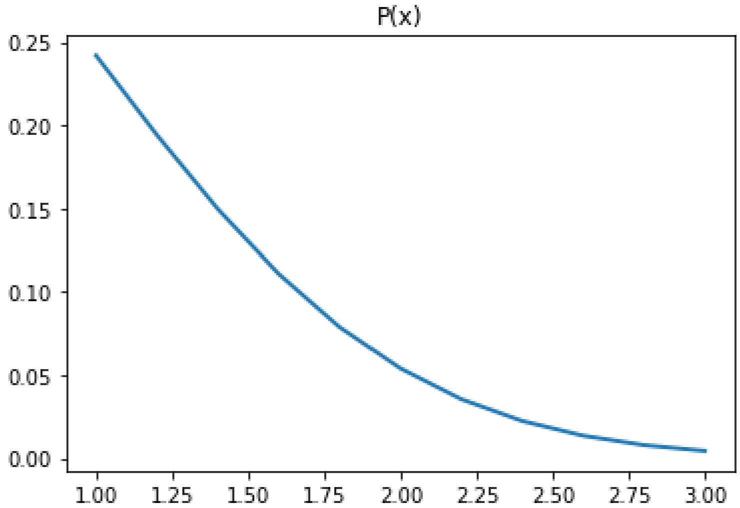
Part C

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 [7]:
```
# Solution to Part B here
x = np.linspace(1,3,11)
partC = np.zeros(11)
for i in range(0,11):
    partC[i] =PDF(x[i],0,1);
print("\n Values of P(1:3) are:")
print(partC)
print("\n Plot of Section:")
# plot
fig, ax = plt.subplots();

ax.plot(x, partC, linewidth=2.0);
ax.set_title("P(x)");
```

```
 Values of P(1:3) are:
[0.24197072 0.19418605 0.14972747 0.11092083 0.07895016 0.05399097
 0.03547459 0.02239453 0.01358297 0.00791545 0.00443185]

 Plot of Section:
```



Part D

The graphs below are probability density functions produced by P:

Different input values create the different graphs.

title

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

solution:

The normal distribution shown by the graphs above are also described by the equation listed before part A.

It is clear that decreasing σ or the standard deviation would increase the maximum value of x. Furthermore, increasing the standard deviation would spread out the function to make it more evenly distributed.

Let's assign σ_i to graph i for 1:3.

Then: $\sigma_3 > \sigma_2 > \sigma_1$

Part E

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

solution:

Increasing μ or the mean has the effect of "shifting" the graphs to the right. This is because the mean describes the centerpoint of the PDF graph. However, each of the graphs are centered at the dashed line. So, all the means are the same.

Thus, assigning μ_i to each graph i: $\mu_1 = \mu_2 = \mu_3$

Rubric Check

(5 points) Makesure 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.