Homework 4

PSTAT 5A: Spring 2023, with Ethan P. Marzban

Instructions

- Please submit your work to Gradescope by no later than **11:59pm on Tuesday, May 2**. As a reminder, late homework will not be accepted.
- Recall that you will be asked to upload a **single** PDF containing your work for *both* the programming and non-programming questions to Gradescope.
  - You can merge PDF files using either Adobe Acrobat, or using adobe's online PDF merger at this link.

Problem 1: Classifying Random Variables

Classify each of the following random variables as either discrete or continuous. Justify your answers.

a. X = the height (in ft) of a randomly selected skyscraper in Los Angeles

> **Solution:** Continuous: the state space is $S_X = [0, \infty)$ which has no "jumps".

b. Y = the number of car crashes we observe at a particular intersection

> **Solution:** Discrete: the state space is $S_X = \{0, 1, 2, \cdots\}$ which has "jumps".

c. Z = the number of blueberries we find in a randomly-selected cubic-inch of muffin

> **Solution:** Discrete: the state space is $S_X = \{0, 1, 2, \cdots\}$ which has "jumps".

d. W = the amount of time (in minutes) a person has to spend waiting at a randomly-selected red light

> **Solution:** Continuous: the state space is $S_X = [0, \infty)$ which has no "jumps".

Problem 2: A Discrete Random Variable

Let X be a random variable with the following probability mass function (p.m.f.):

$\boldsymbol{k}$	-1	-0.5	0	1.4	3
$\mathbb{P}(\boldsymbol{X = k})$	0.12	0.23	0.34	0.07	$\boldsymbol{a}$

a. What must the value of $\boldsymbol{a}$ be?

Solution: We know that the sum of the probabilities in a p.m.f. must be 1; that is,

$$0.12 + 0.23 + 0.34 + 0.07 = a$$

which means

$$a = 1 - (0.12 + 0.23 + 0.34 + 0.07) = 0.24$$

b. What is the probability that X is either 0 or 1?

Solution: $\mathbb{P}(\{X = 0\} \cup \{X = 1\}) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = 0.34 + 0 = 0.34$

c. Compute $\mathbb{P}(X \leq 0)$.

Solution: $\mathbb{P}(X \leq 0) = \mathbb{P}(X = -1) + \mathbb{P}(X = -0.5) + \mathbb{P}(X = 0) = 0.12 + 0.23 + 0.34 = 0.69$

d. Compute $\mathbb{E}[X]$, the expected value of X

Solution:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_k k \cdot \mathbb{P}(X = k) \\
&= (-1) \cdot \mathbb{P}(X = -1) + (-0.5) \cdot \mathbb{P}(X = 0.5) + (0) \cdot \mathbb{P}(X = 0) \\
&\qquad + (1.4) \cdot \mathbb{P}(X = 0.07) + (3) \cdot \mathbb{P}(X = 3) \\
&= (-1) \cdot (0.12) + (-0.5) \cdot (0.23) + (0) \cdot (0.34) + (1.4) \cdot (0.07) + (3) \cdot (0.24) = 0.583
\end{aligned}$$

e. Compute $\mathrm{Var}(X)$, the variance of X.

Solution: We first compute

$$\begin{aligned}
\sum_k k^2 \cdot \mathbb{P}(X = k) &= (-1)^2 \cdot \mathbb{P}(X = -1) + (-0.5)^2 \cdot \mathbb{P}(X = 0.5) + (0)^2 \cdot \mathbb{P}(X = 0) \\
&\qquad + (1.4)^2 \cdot \mathbb{P}(X = 0.07) + (3)^2 \cdot \mathbb{P}(X = 3) \\
&= (-1)^2 \cdot (0.12) + (-0.5)^2 \cdot (0.23) + (0)^2 \cdot (0.34) + (1.4)^2 \cdot (0.07) \\
&\qquad + (3)^2 \cdot (0.24) = 2.4747
\end{aligned}$$

Therefore, by the second variance formula,

$$\mathrm{Var}(X) = \left(\sum_k k^2 \cdot \mathbb{P}(X = k) \right) - (\mathbb{E}[X])^2 = 2.4747 - (0.583)^2 \approx 2.1348$$

Problem 3: Binomial, or Not?

For each of the following random variables, determine whether they are Binomially distributed or not. If they are binomially distributed, provide the parameters of the Binomial distribution they follow.

Solution: For each part, we check the four Binomial criteria.

a. X = the number of heads we observe in 120 independent tosses of a *biased* coin that lands 'heads' with probability 0.8.

Solution:
(1) Trials are independent: yes.
(2) Number of trials is fixed: yes ($n = 120$)
(3) Each trial outcome can be classified as "success" or "failure": yes ("success" = "landing heads" and "failure" = "landing tails")
(4) Fixed probability of success across trials: yes ($p = 0.8$)

Hence, X does follow the Binomial distribution; specifically, $X \sim \text{Bin}(120, 0.8)$.

b. Y = the number of people who own a car in a pool of 100 UCSB students that is selected *without* replacement, assuming 63% of the total UCSB student population owns a car.

Solution:
(1) Trials are independent: yes.
(2) Number of trials is fixed: **no** (sampling done *without* replacement)
(3) Each trial outcome can be classified as "success" or "failure": yes ("success" = "owning a car" and "failure" = "not owning a car")
(4) Fixed probability of success across trials: **no** (sampling done *without* replacement)

Hence, Y does not follow the Binomial distribution.

c. Z = the number of people who own a car in a pool of 100 UCSB students that is selected *with* replacement, assuming 63% of the total UCSB student population owns a car.

Solution:
(1) Trials are independent: yes.
(2) Number of trials is fixed: yes ($n = 100$)
(3) Each trial outcome can be classified as "success" or "failure": yes ("success" = "owning a car" and "failure" = "not owning a car")
(4) Fixed probability of success across trials: yes ($p = 0.63$)

Hence, Z does follow the Binomial distribution; specifically, $Z \sim \text{Bin}(100, 0.63)$.

d. You roll a fair 6−sided die, and record the number showing; then, you toss as many fair coins as there are spots showing on the die (e.g. if the die shows the number 3, you then toss 3 coins). Assume the coin tosses are independent of each other. Let W = the number of heads in these coin tosses.

Solution:

(1) Trials are independent: yes.
(2) Number of trials is fixed: **no**
(3) Each trial outcome can be classified as "success" or "failure": yes ("success" = "landing heads" and "failure" = "landing tails")
(4) Fixed probability of success across trials: yes ($p = 0.5$, since the coin is stated to be fair)

Hence, W does not follow the Binomial distribution.

Problem 4: Tracking Defects

It is found that 2% of *GauchoCompute*-brand laptops contain a defect. An inspector selects 52 *GauchoCompute*-brand laptops with replacement, and records the number of defective laptops.

a. Define the random variable of interest, and call it X.

Solution: Let X denote the number of defective *GauchoCompute*-brand laptops in the sample of 52.

b. Does X follow the Binomial Distribution? (Be sure to check the conditions!) If so, what are the values of n and p?

Solution:

(1) Trials are independent: yes.

(2) Number of trials is fixed: yes ($n = 52$)

(3) Each trial outcome can be classified as "success" or "failure": yes ("success" = "laptop is defective" and "failure" = "laptop is not defective")

(4) Fixed probability of success across trials: yes ($p = 0.02$)

Hence, X does follow the Binomial distribution; specifically, $X \sim \text{Bin}(52, 0.02)$.

c. What is the probability that the inspector observes exactly 2 defective laptops in her sample of 52?

Solution: Recall that if $Y \sim \text{Bin}(n, p)$, then the p.m.f. of Y is given by

$$\mathbb{P}(Y = y) = \binom{n}{y} \cdot p^y \cdot (1-p)^{n-y}$$

Hence, since $X \sim \text{Bin}(n = 52,\ p = 0.02)$ we have that the p.m.f. of X is given by

$$\mathbb{P}(X = x) = \binom{52}{x} \cdot (0.02)^x \cdot (0.98)^{52-x}$$

and so

$$\mathbb{P}(X = 2) = \binom{52}{2} \cdot (0.02)^2 \cdot (0.98)^{52-2} \approx \boxed{0.1932 = 19.32\%}$$

d. What is the probability that the inspector observes at least 3 defective laptops in her sample of 52?

Solution:

$$\begin{aligned}
\mathbb{P}(X \geq 3) &= 1 - \mathbb{P}(X \leq 2) \\
&= 1 - [\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2)] \\
&= 1 - \left[\binom{52}{0} \cdot (0.02)^0 \cdot (0.98)^{52-0} + \binom{52}{1} \cdot (0.02)^1 \cdot (0.98)^{52-1} + \binom{52}{2} \cdot (0.02)^2 \cdot (0.98)^{52-2}\right] \\
&\approx 1 - 0.9140 = \boxed{0.086 = 8.6\%}
\end{aligned}$$

e. What is the expected number of defective laptops the inspector will observe in her sample of 52?

Solution: $\mathbb{E}[X] = n \cdot p = (52) \cdot (0.02) = \boxed{1.04 \text{ laptops}}$

Problem 5: I've Got a Golden Ticket!

At the State Fair, you encounter the following game: a gamemaster has a box containing 4 blue tickets, 3 red tickets, and 1 gold ticket and selects a single ticket at random from this box. If the ticket selected is blue you win a dollar, if it is red you *lose* a dollar, and if it is gold you win 10 dollars. Let W denote your winnings after playing this game once.

a. What is S_W, the state space of W?

Solution: The only three possibilities are: we gain a dollar, we gain 10 dollars, or we lose a dollar. Hence,

$$\boxed{S_W = \{-1,\ 1,\ 10\}}$$

b. Find the p.m.f. (probability mass function) of W.

Solution: We consider each point in the state space separately:

- $\underline{\mathbb{P}(W=-1)}$: The only time we lose a dollar is if the gamemaster drew a red ticket. The probability that a red ticket was drawn is, by the Classical Approach to probability, simply the number of red tickets divided by the total number of tickets: hence,

$$\mathbb{P}(W=-1)=\mathbb{P}(\text{red was drawn})=\frac{3}{8}$$

- $\underline{\mathbb{P}(W=1)}$: The only time we win a dollar is if the gamemaster drew a blue ticket. By a similar reasoning as with the previous bullet point:

$$\mathbb{P}(W=1)=\mathbb{P}(\text{blue was drawn})=\frac{4}{8}=\frac{1}{2}$$

- $\underline{\mathbb{P}(W=10)}$: The only time we win ten dollars is if the gamemaster drew the gold ticket. By a similar reasoning as with the previous bullet points:

$$\mathbb{P}(W=10)=\mathbb{P}(\text{gold was drawn})=\frac{1}{8}$$

Hence, putting everything together, our p.m.f. for W looks like:

$\boldsymbol{k}$	-1	1	10
$\mathbb{P}(\boldsymbol{X=k})$	$3/8$	$1/2$	$1/8$

As a quick sanity check, note that the probabilities do indeed sum to 1.

c. What is the expected amount you will win each time you play the game? (I.e. what is $\mathbb{E}[W]$?)

Solution:

$$\begin{aligned}\mathbb{E}[W]&=\sum_k k\cdot\mathbb{P}(W=k)\\&=(-1)\cdot\mathbb{P}(W=-1)+(1)\cdot\mathbb{P}(W=1)+(10)\cdot\mathbb{P}(W=10)\\&=(-1)\cdot\left(\frac{3}{8}\right)+(1)\cdot\left(\frac{1}{2}\right)+(10)\cdot\left(\frac{1}{8}\right)=\frac{11}{8}=\$1.375\end{aligned}$$

Problem 6: Programming

Part (a): Basic Plotting

! Task 1

Plot the function $f(x) = x\sin(x) + xe^{-x^2}$ between $x = -4\pi$ and $x = 4\pi$. Add appropriate axis labels, and add a title.

Solutions

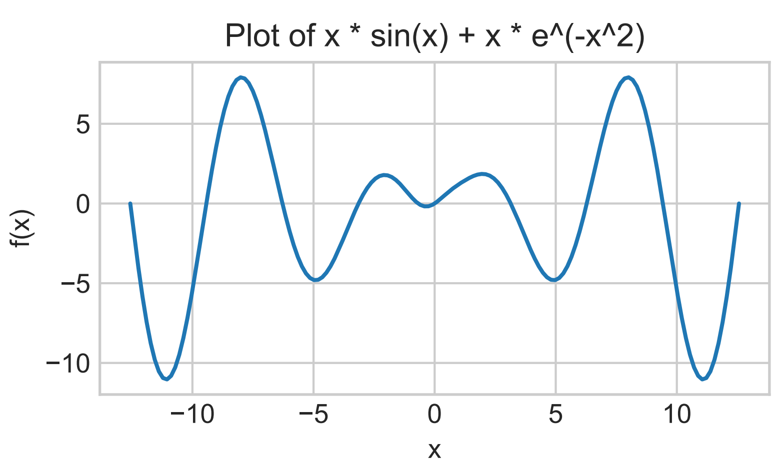
```python
import numpy as np

%matplotlib inline
import matplotlib
import matplotlib.pyplot as plt
plt.style.use('seaborn-v0_8-whitegrid')

def f(x):
  """return x * sin(x) + x * e^(-x^2)"""
  return x * np.sin(x) + x * np.exp(- x ** 2)

x = np.linspace(-4 * np.pi, 4 * np.pi, 150)

plt.plot(x, f(x));
plt.xlabel("x");
plt.ylabel("f(x)");
plt.title("Plot of x * sin(x) + x * e^(-x^2)");
```



❗ Task 2

Copy-paste the following code into a cell, and run the cell:

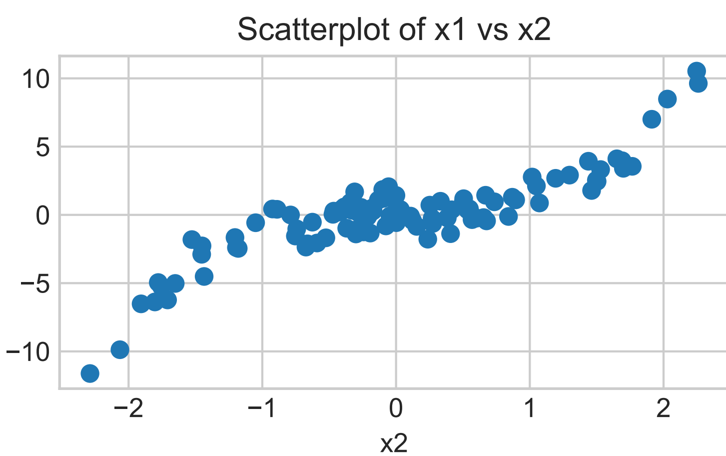
```
import numpy as np
np.random.seed(7)

x1 = np.random.normal(0, 1, 100)
x2 = (x1 ** 3) + np.random.normal(0, 1, 100)
```

Generate a scatterplot of `x1` vs `x2` (i.e. `x1` should be on the horizontal axis and `x2` should be on the vertical axis), and comment on whether there appears to be an association between `x1` and `x2`. What type of association is there?

Solutions

```
plt.scatter(x1, x2);
plt.xlabel("x1");
plt.xlabel("x2");
plt.title("Scatterplot of x1 vs x2");
```



Part (b): Superimposing Plots

It will sometimes be necessary to superimpose two or more plots on top of each other. The goal of this problem is to walk through how to do this.

Recall the following: given a function `f()` and a variable `x` that has been assigned a value resulting from a call to `numpy.linspace()`, we generate a graph of `f()` using (assuming `matplotlib.pyplot` has been imported as `plt`):

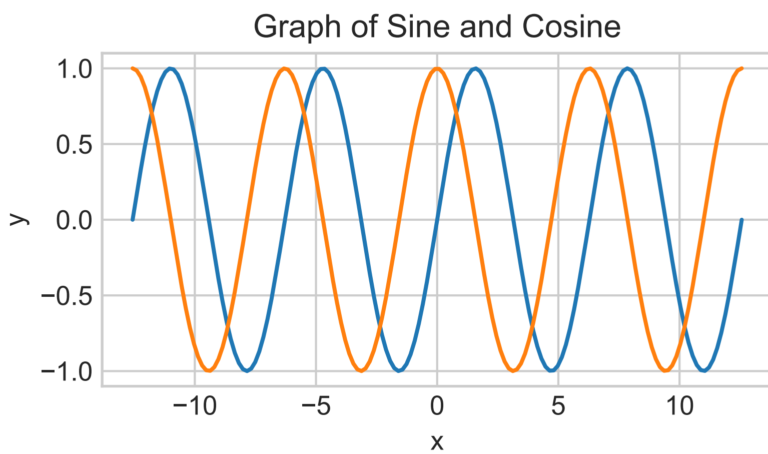
```
plt.plot(x, f(x));
```

It stands to reason, then that given another function `g()` we should be able to superimpose the graph of `g()` onto the graph of `f()` by simply adding another call to `plt.plot()`:

```
plt.plot(x, f(x));
plt.plot(x, g(x));
```

❗ Task 3

Generate a graph of `sin()`; on top of this graph, superimpose the graph of `cos()`. Restrict the x values on the graph to be between -4π and 4π. Your final graph should look like the following (**pay attention to the axis labels and title!**):



Solutions

```
plt.plot(x, np.sin(x));
plt.plot(x, np.cos(x));
plt.xlabel("x");
plt.ylabel("y");
plt.title("Graph of Sine and Cosine");
```

Now, as it stands, it's a bit difficult to determine which curve corresponds to the sine curve and which corresponds to the cosine curve. As such, we should add some labels!

❗ Task 4

Copy your code from Task 3 above into a new code cell, and

- add `label = "sine"` to your call to `plt.plot()` containing the sine curve
- add `label = "cosine"` to your call to `plt.plot()` containing the cosine curve.

Does this new plot look any different than the plot you generated in Task 3?

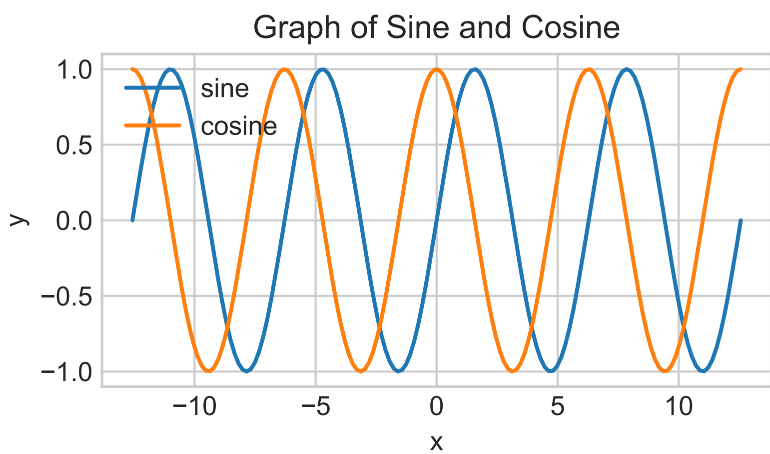
Solutions

```
plt.plot(x, np.sin(x), label = "sine");
plt.plot(x, np.cos(x), label = "cosine");
plt.xlabel("x");
plt.ylabel("y");
plt.title("Graph of Sine and Cosine");
```

Hm, doesn't look like anything changed... That's because we didn't add a **legend** to our plot! To add a legend, we simply tack on a call to `plt.legend()` after our code from above.

Task 5

Copy your code from Task 4 above into a new code cell, and add a line underneath it containing a call to `plt.legend()`. Look up the help file to figure out what arguments you need to pass in to obtain the following graph (note the **position** of the legend):



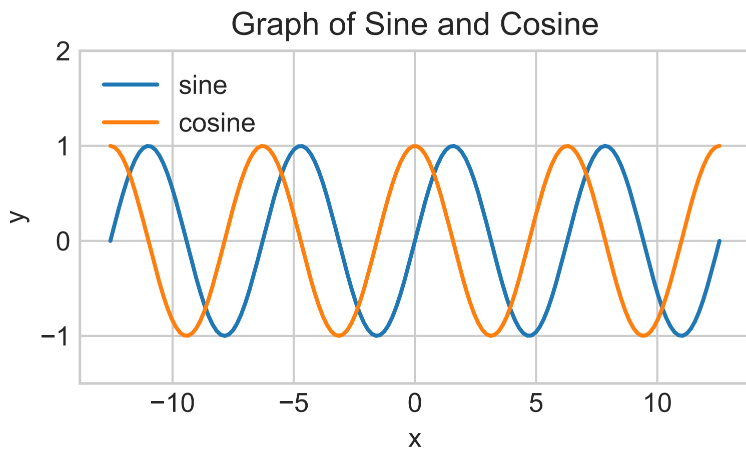
Solutions

```
plt.plot(x, np.sin(x), label = "sine");
plt.plot(x, np.cos(x), label = "cosine");
plt.xlabel("x");
plt.ylabel("y");
plt.title("Graph of Sine and Cosine");
plt.legend(loc="upper left");
```

Okay, we're almost there! The only issue is that now the legend is covered up by the actual graphs. One way we can fix this is by extending the y−axis further, using the function `plt.ylim()`:

❗ Task 6

Copy your code from Task 5 above into a new code cell, and add a line underneath it containing a call to `plt.ylim()`. Look up the help file to figure out what arguments you need to pass in to obtain a lower y−limit of -1.5 and an upper y−limit of 2.0. Your final graph should look like this:



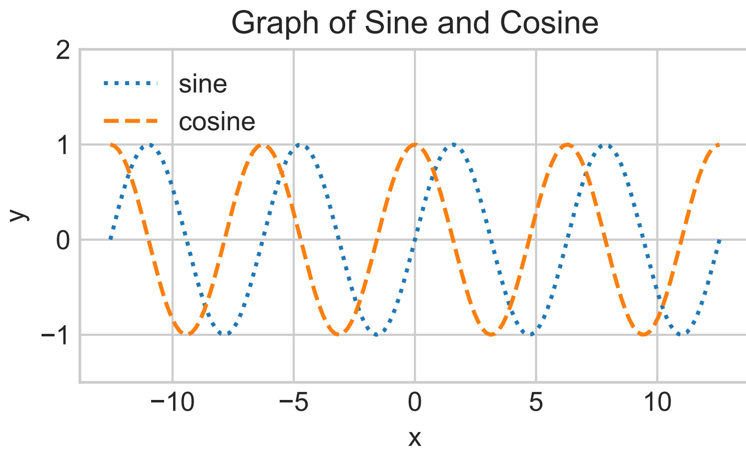
Solutions

```
plt.plot(x, np.sin(x), label = "sine");
plt.plot(x, np.cos(x), label = "cosine");
plt.xlabel("x");
plt.ylabel("y");
plt.title("Graph of Sine and Cosine");
plt.legend(loc="upper left");
plt.ylim(-1.5, 2.0);
```

Finally, it is sometimes considered bad form to rely too heavily on colors in plots. This is because doing so alienates readers who are colorblind. One way around this is to rely on different *line types*; e.g. used dashed lines for one graph and dotted lines for another.

❗ Task 7

Copy your code from Task 6 above into a new code cell. Read the following help file and figure out how to pass in a value to the `linestyle` argument to your two calls to `plt.plot()` to generate the following plot:



Note that the sine curve is now dotted, and the cosine curve is now dashed.

Solutions

```
plt.plot(x, np.sin(x), label = "sine", linestyle = ":");
plt.plot(x, np.cos(x), label = "cosine", linestyle = "--");
plt.xlabel("x");
plt.ylabel("y");
plt.title("Graph of Sine and Cosine");
plt.legend(loc="upper left");
plt.ylim(-1.5, 2.0);
```