Write **clearly** and **in the box**:

CSCI 3022
Midterm Exam
Fall 2019

Name:
Student ID:
Section number:

Read the following:

- **RIGHT NOW**! Write your name on the top of your exam.
- You are allowed **one** 3×5in notecard of notes (both sides). No magnifying glasses!
- You may use a calculator provided that it cannot access the internet or store large amounts of data.
- You may **NOT** use a smartphone as a calculator.
- Clearly mark answers to multiple choice questions **in the provided answer box**.
- Mark only one answer for multiple choice questions. If you think two answers are correct, mark the answer that **best** answers the question. No justification is required for multiple choice questions.
- If you do not know the answer to a question, skip it and come back to it later.
- For free response questions you must clearly justify all conclusions to receive full credit. A correct answer with no supporting work will receive no credit.
- You have **90 minutes** for this exam.

Potentially Useful Facts

Bayes' theorem	$p(A \mid B) = \dfrac{p(B \mid A)p(A)}{p(B)}$	Law of total probability	$p(E) = \sum_{i=1}^{N} p(E \mid F_i)p(F_i)$
Union of sets	$p(A \cup B) = p(A) + p(B) - p(A \cap B)$	Conditional probability	$p(A \mid B) = \dfrac{p(A \cap B)}{p(B)}$
Expectation & variance	$Var(X) = E[X^2] - E[X]^2$	You can do it!	

Multiple choice problems: Write your answers in the boxes, or they will not be graded!

1. (3 points) Consider the data set: $[x, 2, 4, 10, 11, 13]$, where $x \in \mathbb{R}$ is an unknown quantity. What is the set of possible values that the mean of this data set must belong to?

   A. $[2, 13]$

   B. $[2, 8]$

   C. $\{8\}$

   D. $\emptyset$

   E. $(-\infty, \infty)$

   $\boxed{\text{E}}$

2. (3 points) Suppose Kyle has a roster consisting of all students in the Computer Science Department. He is conducting a survey of how many hours per week Computer Science students spend studying, by drawing a sample from the roster. The students on the roster are divided into class; freshman, sophomore, junior, and senior. Kyle randomly chooses 10% of the students from each class to contact.

   What type of sample did Kyle collect?

   A. Simple random sample

   B. Systematic sample

   C. Census sample

   D. Stratified sample

   E. Free samples at Trader Joe's on Saturdays

   $\boxed{\text{D}}$

Use the following information for Problems 3 – 5, which may build off of each other.

Beowolf and Hrothgar are defending their kingdom from a monster. As Beowolf and Hrothgar fight the terrible monster, they use two tactics: they either throw boiling oil or they launch a catapult. These are their only two fighting tactics. Suppose the probability that Beowolf is throwing boiling oil at the monster is 0.4 while the probability that Hrothgar is throwing boiling oil at the monster is 0.7.

3. (3 points) Suppose that both Beowolf and Hrothgar throw boiling oil at the monster with probability 0.3. What is the probability that Beowulf or Hrothgar is throwing boiling oil at the monster?

   A. 0.3

   B. 0.4

   C. 0.7

   D. 0.8

   E. 0.9

   D

4. (3 points) Beowulf is launching a catapult. What is the probability that Hrothgar is launching a catapult?

   A. 0.25

   B. 0.33

   C. 0.5

   D. 0.67

   E. 0.75

   B

5. (3 points) What is the probability that Beowulf and Hrathgar are using the same tactic at the same time?

   A. 0.1

   B. 0.2

   C. 0.3

   D. 0.4

   E. 0.5

   E

6. (3 points) Suppose you are playing pool with your friends. Instead of simply keeping score, your friends are tracking the number of turns it takes you to sink 5 balls. Suppose that you can sink a pool ball into its pocket with probability 0.4. What is the probability that it takes you exactly 8 turns to sink 5 pool balls.

   A. $\binom{7}{4}(0.4)^5(0.6)^3$

   B. $\binom{8}{5}(0.4)^5(0.6)^3$

   C. $1 - \sum_{i=5}^{8} \binom{i-1}{4}(0.4)^5(0.6)^i$

   D. $1 - \sum_{i=0}^{8} \binom{i-1}{4}(0.4)^5(0.6)^{i-5}$

   E. $1 - \sum_{i=0}^{8} \binom{i}{5}(0.4)^5(0.6)^{i-5}$

A

7. (3 points) You decide to run a coin flipping experiment for your probability class. You flip a biased coin repeatedly until you achieve a Heads and you record the number of coin flips that it took for you to achieve this. The coin is biased so that the probability of Heads is 0.1. How many times on average do you expect to flip the coin before observing a Heads?

   A. 0

   B. 1

   C. 5

   D. 10

   E. 11

   F. ∞

D

8. (3 points) Suppose you want to count the number of birds flying over a certain area every given hour. You do this and find that on average, 47 birds fly by each hour. You'd like to model the time interval between bird flyovers. What distribution would you use to characterize this phenomenon?

   A. Binomial

   B. Negative binomial

   C. Uniform

   D. Normal

   E. Poisson

   F. Exponential

   F

9. (3 points) Consider the following function, where the probability p is some constant. What distribution does the return value of the function belong to?

```
def what_the_function(p):
x = 0
y = 0
while y < 5:
   draw = np.random.choice([0,1], p=[1-p, p])
   x += 1
   if draw == 1:
      y += 1
return x
```

   A. Bernoulli

   B. Binomial

   C. Geometric

   D. Negative binomial

   E. Poisson

   D

10. (3 points) The average high temperature for Boulder, CO on October 31 is 58° Fahrenheit with a standard deviation of 11 degrees. If the temperature C in Celsius is calculated from the temperate in Fahrenheit F by $C = \frac{5}{9}(F - 32)$, what is the *mean* of the temperature in Boulder on October 31 in degrees Celsius?

   A. $\frac{5}{9} \cdot (58 - 32)$

   B. $11^2 \cdot \frac{5^2}{9^2}$

   C. $11 \cdot \frac{5}{9}$

   D. $\left(\frac{5}{9}\right)^2 26^2$

   E. $11^2 \cdot \frac{5}{9}$

   F. $11 \cdot \left(\frac{5}{9}\right)^2$

   A

11. (3 points) Consider simulating two flips of a potentially biased coin, where 'H' represents Heads and 'T' represents Tails. Which of the following quantities does the following function estimate?

```
def flippyFloppy(num_samples):
   flips1 = np.random.choice(['H','T'], size=num_samples)
   flips2 = np.random.choice(['H','T'], size=num_samples)
   return np.sum(np.logical_and((flips1=='H'),(flips2=='H'))) / np.sum(flips2=='H')
```

A. P(both flips are heads | second flip is heads)

B. P(second flip is heads | both flips are heads)

C. P(both flips are heads $\cup$ second flip is heads)

D. P(neither flip is heads)

E. P(both flips are heads)

F. 1.

A

12. (3 points) Let $f(x) = kx$ for $0 \leq x \leq 1$, and $f(x) = 0$ for $x \notin [0, 1]$, where k is some unknown constant. What value of k will make f a valid probability density function?

A. 1/4

B. 1/2

C. 3/4

D. 1

E. 2

F. 4

G. No such value of k exists

E

13. (3 points) Suppose we know that the general antiderivative of a function $g(x)$ is $\int g(x)\,dx = \frac{x^2}{8} + C$. Then from $[1, 3]$ the cumulative density function of a random variable with pdf $f(x)$,

$$f(x) = \begin{cases} g(x) & 1 < x \leq 3 \\ 0 & else \end{cases}$$

is given by:

A. $F(x) = \dfrac{x^3}{24}$

B. $F(x) = \dfrac{x}{4}$

C. $F(x) = 1 - \dfrac{x^2}{8}$

D. $F(x) = \dfrac{x^2 - 1}{8}$

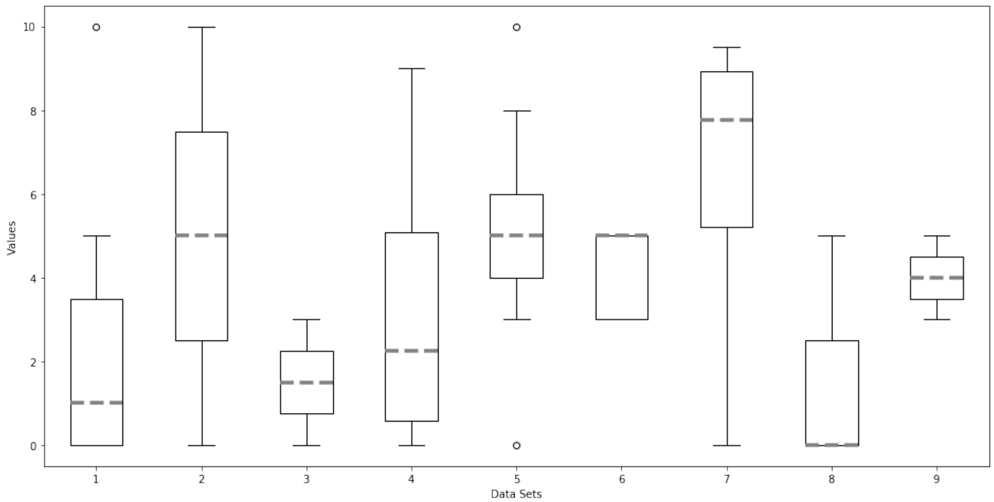
E. $F(x) = 1 - \dfrac{x^3}{24}$

D

14. (20 points) No justification is necessary for this problem. Consider the following four data sets, each with 11 elements.

```
A = [0, .3, .6, .9,1.2,1.5,1.8,2.1,2.4,2.7, 3]
B = [10-10*np.exp(-x) for x in B]
C = [0, 0, 0, 0, 0, 1, 2, 3, 4, 5, 10]
D = [5, 5, 5, 5, 5, 5, 5, 3, 3, 3, 3]
```

(a) (12 points) Match each of the four data sets to their boxplot (box-whisker plot) below. Use the conventions from lecture, and clearly mark the number corresponding to your choice of boxplot in the boxes below for each data set. No boxplot is used more than once, and some are not used.



A: 3	B: 7	C: 1	D: 6

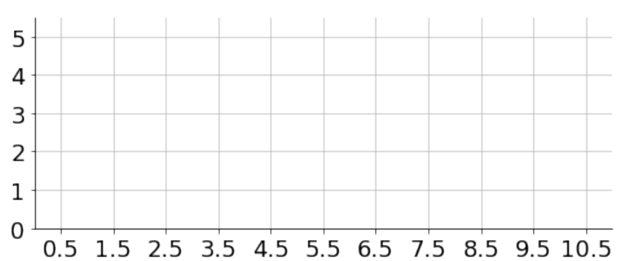
(b) (3 points) For each of A, B, and C, classify the distribution of the data as symmetric, positively skewed, or negatively skewed.

A: Symmetric

B: Negative Skew

C: Positive Skew

(c) (5 points) Consider the data set $C = \{0, 0, 0, 0, 0, 1, 2, 3, 4, 5, 10\}$. Draw the **frequency** histogram of C. Use bins of width 2, with the left-most bin edge at -0.5. Label your axes.



15. (20 points) It is well-known that the mood of the legendary headmaster of the CU College of Witchcraft and Wizardry, Rhondolf, is greatly affected by the weather. The weather each day is exactly one of three states: nice (N), cloudy (C) or rainy (R). In Rhondolf's location, 2/10 of the days are nice, 3/10 of the days are cloudy, and 5/10 of the days are rainy. On days when it is nice, Rhondolf is happy (H) with probability 9/10; on days when it is cloudy, she is happy with probability 5/10; and on days when it is rainy, she is happy with probability 1/10.

(a) (8 points) Without knowing the weather, what is the probability that Rhondolf is **not happy** on any given day?

Solution: Note that $p(\bar{H}) = 1 - p(H)$. Using the Law of Total Probability:

$$\begin{aligned} p(\bar{H}) = 1 - p(H) &= 1 - (p(H \mid N)p(N) + p(H \mid C)p(C) + p(H \mid R)p(R)) \\ &= 1 - \left(\frac{9}{10} \cdot \frac{2}{10} + \frac{5}{10} \cdot \frac{3}{10} + \frac{1}{10} \cdot \frac{5}{10}\right) \\ &= 1 - \frac{18 + 15 + 5}{100} \\ &= 1 - \frac{38}{100} \\ &= \frac{62}{100} \end{aligned}$$

(b) (8 points) Suppose you see Rhondolf and find that she is happy. Given that Rhondolf is happy, what is the probability that it is rainy?

Solution: Using Bayes' Theorem:

$$\begin{aligned} p(R \mid H) &= \frac{p(H \mid R)p(R)}{p(H)} \\ &= \frac{1/10 \cdot 5/10}{38/100} \\ &= \frac{5/100}{38/100} \\ &= \frac{5}{38} \approx 0.132 \end{aligned}$$

(c) (4 points) Are the events "Rhondolf is happy" (H) and "it is nice" (N) independent? Fully justify your answer **using math**.

Solution: They are **not independent**. We check by comparing $p(H)$ and $p(H \mid N)$:
From part (a), we know $p(H) = 38/100 = 0.38$
And from the problem set-up, we know that $p(H \mid N) = 9/10 = 0.9$
These are not equal, so the events are not independent.

Alternatively, you could check one of:

- $p(N)$ against $p(N \mid H)$
- $p(H \cap N)$ against $p(H)\ p(N)$

16. (20 points) Mathematically justify all answers. Suppose you have a biased zero-indexed pyramidal 4-sided die, which you roll and then record the face down side as X. X has probability distribution

$$P(X=0) = \frac{1}{10}, \quad P(X=1) = \frac{2}{10}, \quad P(X=2) = \frac{6}{10}, \quad P(X=3) = \frac{1}{10}$$

(a) (4 points) Compute $E[X]$

Solution:

$$\begin{aligned}
E[X] &= \sum_x x \cdot P(X=x) = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) \\
&= 0 \cdot \frac{1}{10} + 1 \cdot \frac{2}{10} + 2 \cdot \frac{6}{10} + 3 \cdot \frac{1}{10} \\
&= \frac{17}{10} = 1.7
\end{aligned}$$

(b) (4 points) Let Y be the random variable $Y = (X-1)^2$. Write down the probability distribution (probability mass function) of Y.

Solution: The possible values of Y are 0, 1 and 4.
The probabilities are:
$P(Y=0) = P(X=1) = \frac{2}{10} = 0.2$
$P(Y=1) = P(X=0) + P(X=2) = \frac{1}{10} + \frac{6}{10} = \frac{4}{10} = 0.7$
$P(Y=4) = P(X=3) = \frac{1}{10} = 0.1$

(c) (4 points) Compute $E[Y]$

Solution:

$$\begin{aligned}
E[Y] &= \sum_y y \cdot P(Y=y) = 0 \cdot P(Y=0) + 1 \cdot P(Y=1) + 4 \cdot P(Y=4) \\
&= \frac{7+4}{10} = \boxed{1.1}
\end{aligned}$$

(d) (4 points) Compute $\text{Var}(X)$

Solution: We use the formula:

$$\text{Var}(X) = E[X^2] - E[X]^2$$

And we can obtain $E[X^2]$ from the linearity of $E[Y]$:

$$E[Y] = E[(X-1)^2] = E[X^2] - 2E[X] + 1$$

$$E[X^2] = E[Y] + 2E[X] + 1 = 1.1 + 2 \cdot 1.7 - 1 = 3.5$$

Note that you could also have gone straight from the change-of-variables formula:

$$E[X^2] = \sum_i a_i^2 \cdot P(X = a_i)$$

Anyway, we plug into our relationship between expected value and variance to find:

$$\begin{aligned} \text{Var}(X) &= E[X^2] - E[X]^2 \\ &= 3.5 - (1.7)^2 \\ &= 3.5 - 2.89 \\ &= \boxed{0.61} \end{aligned}$$

(e) (4 points) Suppose we independently roll this die twice. What is the probability that the sum of the two rolls is 5 or greater?

Solution: This is 3 outcomes: $\{23, 32, 33\}$. These outcomes have probabilities of $\frac{6}{10}\frac{1}{10}, \frac{1}{10}\frac{6}{10}, \frac{1}{10}\frac{1}{10}$ respectively (multiplicate rule), and since each is a disjoint event, the total probability is their sum of $\boxed{0.13}$.