

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

CSCI 3022
Midterm Exam
Spring 2019

Name:

Student ID:

Section number:

Read the following:

- **RIGHT NOW!** Write your name on the top of your exam.
- You are allowed **one** 3×5 in 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.

Page	Points	Score
2	6	
3	9	
4	6	
5	9	
6	9	
7	20	
9	20	
11	20	
For luck!	1	
Total	100	



Potentially Useful Facts

Bayes' theorem	$p(A B) = \frac{p(B A)p(A)}{p(B)}$	Law of total probability	$p(E) = \sum_{i=1}^N p(E F_i)p(F_i)$
Union of sets	$p(A \cup B) = p(A) + p(B) - p(A \cap B)$	Conditional probability	$p(A B) = \frac{p(A \cap B)}{p(B)}$
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: $[0, 1, 5, 7, x]$, where $x \in \mathbb{R}$ is an unknown quantity. What is the set of possible values that the median of this data set must belong to?

- A. $(-\infty, \infty)$
- B. $[1, 5]$
- C. $[0, 7]$
- D. $\{5\}$
- E. \emptyset

B

2. (3 points) Suppose Courtney has a Python list `emails` consisting of all email addresses for students in the Computer Science Department. She conducted a survey of how many hours per week Computer Science students spend studying, by drawing a sample from her email list using the Python command below, and everyone she contacted responded. What type of sample did Courtney collect?

```
sample = np.random.choice(emails, replace=False, size=500)
```

- A. Simple random sample
- B. Systematic sample
- C. Census sample
- D. Stratified sample
- E. Free samples at Trader Joe's on Saturdays

A

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

Arugula, Butterhead, and Cress are three hedgehogs that live in a cabin in Rocky Mountain National Park. They have several hobbies, including eating lettuce and enjoying recreational activities in the park. These hedgehogs have a complicated group dynamic. They have a limited menu in the park, with the only two options being lettuce or insects, and they always eat one of those two things at mealtimes (but the hedgehogs do not necessarily all eat the same thing). At mealtimes, Arugula eats lettuce with a probability of 0.8, and with probability 0.6 Butterhead and Arugula are both eating lettuce.

3. (3 points) The probability that Arugula and Butterhead both eat insects at mealtimes is 0.1. What is the probability that Butterhead eats lettuce?

A. 0.1
B. 0.3
C. 0.5
D. 0.7
E. 0.9

D

4. (3 points) Arugula is eating lettuce. What's the probability that Butterhead is eating lettuce?

A. 0.25
B. 0.4
C. 0.5
D. 0.6
E. 0.75

E

5. (3 points) Cress is a strong, independent hedgehog who don't need no man, so she doesn't care what the other two eat. Thus, Cress's choice of meal is statistically independent of the other two hedgehogs' choices. She will eat insects with probability 0.9. What is the chance that all three hedgehogs eat the same thing at mealtime?

A. 0.06
B. 0.09
C. 0.15
D. 0.65
E. 0.85

C

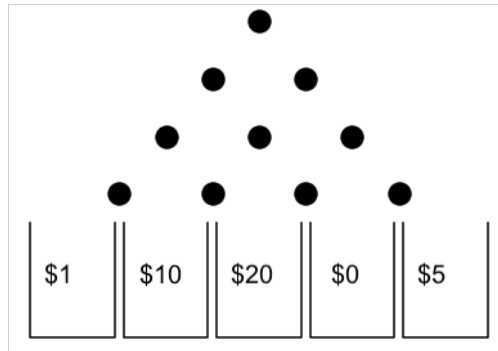
6. (3 points) Suppose Courtney, Jared and Nick open a lemonade stand because that is how Computer Science Mafia works. On average, they get 10 customers per hour. What is the probability that they will have 8 or more customers in an hour?

- A. $\sum_{k=0}^7 \frac{10^k}{k!} e^{-10}$
 B. $\sum_{k=0}^8 \frac{10^k}{k!} e^{-10}$
 C. $1 - \sum_{k=0}^7 \frac{10^k}{k!} e^{-10}$
 D. $1 - \sum_{k=0}^8 \frac{10^k}{k!} e^{-10}$
 E. $\int_{k=7}^{\infty} \frac{10^k}{k!} e^{-10} dk$
 F. $\int_{k=8}^{\infty} \frac{10^k}{k!} e^{-10} dk$

C

7. (3 points) A game of **Plinko** is to be played on the board shown below. The pegs are unbiased, meaning that the disc has equal probability of moving left or right at each peg. Furthermore, the disc can only be dropped from directly above the top-most peg. Which one of the following expressions gives the expected value of your winnings (in dollars) with a single disc?

Reminder/hint: the binomial PMF is $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, where the binomial coefficients are $\binom{n}{k} = \frac{n!}{k! (n-k)!}$.



- A. $42/16$
 B. $166/16$
 C. $130/16$
 D. $36/5$
 E. $28/5$

B

8. (3 points) Nick is playing a game of Flappy Birds, where the goal of the game is to get your bird to flap its wings as many times as possible per minute. He notices that the number of bird wing flaps per minute follows a Poisson distribution with rate 10 flaps per minute. Which distribution is the most appropriate for a random variable representing the amount of time between bird wing flaps?

A. Bernoulli
B. Uniform
C. Normal
D. Poisson
E. Exponential
F. Negative binomial

E

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 bachelor Jarrod is waiting for the contestants to arrive at his mansion. Consider the following function simulating this scenario. What distribution does the return value of the function belong to? Note that the `np.random.exponential` function's input $1/q$ corresponds to an exponentially-distributed random variable with rate $\lambda = q$.

```
def gonna_find_a_wife_on_this_reality_show(q):  
    j = 0  
    t = np.random.exponential(1/q)  
    while t <= 1:  
        t += np.random.exponential(1/q)  
        j += 1  
    return j
```

A. Bernoulli
B. Binomial
C. Geometric
D. Negative binomial
E. Poisson
F. Exponential

E

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(flips1=='H')
```

- A. $P(\text{both flips are heads} \mid \text{first flip is heads})$
- B. $P(\text{first flip is heads} \mid \text{both flips are heads})$
- C. $P(\text{both flips are heads} \cup \text{first flip is heads})$
- D. $P(\text{neither flip is heads})$
- E. $P(\text{both flips are heads})$

A

12. (3 points) Let $f(x) = kx^3$ 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

F

13. (3 points) Consider the probability density function given by $f(x) = \frac{1}{4}(1+x)$ for $0 \leq x \leq 2$ and 0 otherwise. What is the associated cumulative distribution function $F(x)$ for $0 \leq x \leq 2$?

- A. $F(x) = \frac{1}{4}$
- B. $F(x) = \frac{3}{4}$
- C. $F(x) = 1 - \frac{1}{4}(1+x)$
- D. $F(x) = \frac{1}{4} \left(x + \frac{1}{2}x^2 \right)$
- E. $F(x) = 1 - \frac{1}{4} \left(x + \frac{1}{2}x^2 \right)$

D

14. (20 points) No justification is necessary for this problem. Consider the following six data sets, each with 9 elements.

$$X_1 = \{1, 3, 4, 5, 6, 6, 7, 7, 9\}$$

$$X_2 = \{1, 4, 5, 5, 6, 6, 7, 7, 9\}$$

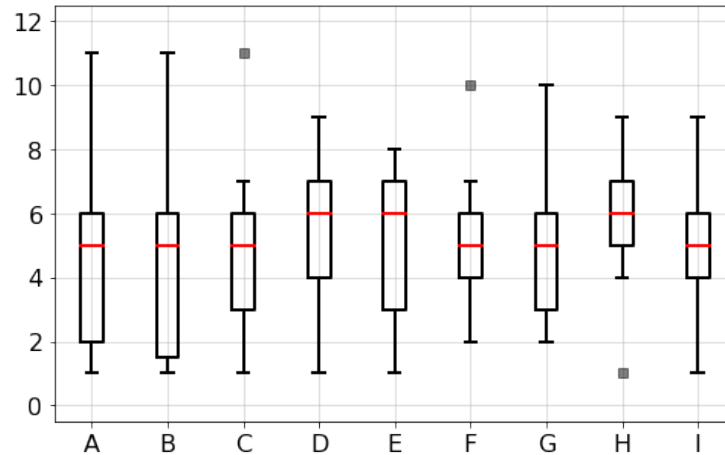
$$X_3 = \{1, 2, 2, 5, 5, 6, 6, 6, 11\}$$

$$X_4 = \{2, 2, 4, 5, 5, 5, 6, 7, 10\}$$

$$X_5 = \{1, 2, 3, 5, 6, 6, 7, 7, 8\}$$

$$X_6 = \{1, 2, 3, 5, 5, 6, 6, 7, 11\}$$

- (a) (12 points) Match each of the six data sets to their boxplot (box-whisker plot) below. Use the conventions from lecture, and clearly mark the letter 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.



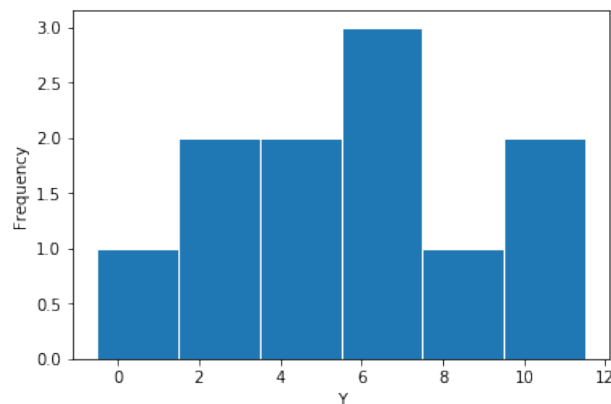
X_1 : D	X_2 : H	X_3 : A	X_4 : F	X_5 : E	X_6 : C
----------------	----------------	----------------	----------------	----------------	----------------

- (b) (4 points) For Parts (b) and (c), consider the data set $Y = \{1, 3, 3, 5, 5, 6, 6, 7, 8, 10, 11\}$. Y has 11 values in it. Write your answers in the boxes provided. No justification is necessary.

Solution: What is the median of Y ? **Answer:** 6

What is the 3rd (upper) quartile, Q_3 , of Y ? **Answer:** 7.5 (there is an odd number of data points, so median is included in the calculations for Q_1 and Q_3)

- (c) (4 points) Draw the **frequency** histogram of Y . Use bins of width 2, with the left-most bin edge at -0.5 . Label your axes.



15. (20 points) Eloise the hedgehog is the queen of the hipsters. While she wears her calculator watch, she churns butter, because that store-bought stuff is too mainstream. She also flips coins by hand and records the results, because you just can't trust that digital junk. Eloise has a box with 6 coins in it. Each coin belongs to one of three categories, and all coins are equally likely to be drawn from the box.

- 1 of the coins is **fair** (F) such that heads (H) and tails (T) are equally likely.
- 2 of the coins are **biased towards heads** (BH), such that heads comes up three times as often as tails.
- 3 of the coins are **biased towards tails** (BT), such that tails comes up three times as often as heads.

Please answer the following questions, and be sure to show your work—for sufficient space, a blank page follows this one.

- (a) (3 points) Eloise chooses a coin at random. What is the probability that it is a fair coin?
- (b) (6 points) Eloise chooses a coin at random and flips it. What is the probability that the coin comes up heads?
- (c) (7 points) Eloise chooses a coin at random and flips it three times. It comes up tails two times and heads one time. Given this information, what is the probability that the coin that she chose was a fair coin? You may express your answer as a fraction or as a decimal rounded to 3 decimal places.
- (d) (4 points) Suppose Eloise picks a coin at random and flips it. Are the events “flip comes up heads” and “Eloise picked a fair coin” independent? Justify your answer **with math**.

Solution:

(a) $P(F) = 1/6 = 0.167$

(b) We are looking for $P(H)$. From the law of total probability, we have:

$$\begin{aligned} P(H) &= P(H | F)P(F) + P(H | BH)P(BH) + P(H | BT)P(BT) \\ &= \frac{1}{2} \cdot \frac{1}{6} + \frac{3}{4} \cdot \frac{2}{6} + \frac{1}{4} \cdot \frac{3}{6} \\ &= \frac{2}{24} + \frac{6}{24} + \frac{3}{24} \\ &= \boxed{\frac{11}{24} \approx 0.458} \end{aligned}$$

(c) We are looking for $P(F | 2T, 1H)$. From Bayes' rule, we have:

$$\begin{aligned} P(F | 2T, 1H) &= \frac{P(2T, 1H | F)P(F)}{P(2T, 1H)} && \text{(Bayes)} \\ &= \frac{P(2T, 1H | F)P(F)}{P(2T, 1H | F)P(F) + P(2T, 1H | BH)P(BH) + P(2T, 1H | BT)P(BT)} && \text{(LTP)} \\ &= \frac{3 \cdot (\frac{1}{2})^3 \cdot \frac{1}{6}}{3 \cdot (\frac{1}{2})^3 \cdot \frac{1}{6} + 3 \cdot (\frac{1}{4})^2 \cdot \frac{3}{4} \cdot \frac{2}{6} + 3 \cdot (\frac{3}{4})^2 \cdot \frac{1}{4} \cdot \frac{3}{6}} \\ &= \frac{\frac{1}{8}}{\frac{1}{8} + \frac{6}{64} + \frac{27}{64}} = \frac{\frac{8}{64}}{\frac{8}{64} + \frac{33}{64}} = \frac{8}{64} \cdot \frac{64}{41} \\ &= \boxed{\frac{8}{41} \approx 0.195} \end{aligned}$$

(d) **No, not independent.** We see this by checking (for example) whether $P(H | F) \stackrel{?}{=} P(H)$:

$$P(H) = 0.458 \neq 0.5 = P(H | F)$$

We could also check $P(H \cap F) \stackrel{?}{=} P(H)P(F)$ or $P(F | H) \stackrel{?}{=} P(F)$.

16. (20 points) The probability distribution of a discrete random variable X is given by

$$P(X = -1) = \frac{1}{10}, \quad P(X = 0) = \frac{3}{10}, \quad P(X = 1) = \frac{4}{10}, \quad P(X = 2) = \frac{2}{10}$$

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

Solution:

$$\begin{aligned} E[X] &= \sum_x x \cdot P(X = x) = -1 \cdot P(X = -1) + 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) \\ &= -1 \cdot \frac{1}{10} + 1 \cdot \frac{4}{10} + 2 \cdot \frac{2}{10} \\ &= -\frac{1}{10} + \frac{8}{10} = \boxed{\frac{7}{10} = 0.7} \end{aligned}$$

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

Solution: The possible values of Y are -1, 0 and 3, as shown below:

$$(-1)^2 - 1 = 0, \quad (0)^2 - 1 = -1, \quad (1)^2 - 1 = 0, \quad (2)^2 - 1 = 3$$

The probabilities are:

$$P(Y = -1) = P(X = 0) = \frac{3}{10} = 0.3$$

$$P(Y = 0) = P(X = 1) + P(X = -1) = \frac{1}{10} + \frac{4}{10} = \frac{5}{10} = 0.5$$

$$P(Y = 3) = P(X = 2) = \frac{2}{10} = 0.2$$

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

Solution:

$$\begin{aligned} E[Y] &= \sum_y y \cdot P(Y = y) = -1 \cdot P(Y = -1) + 0 \cdot P(Y = 0) + 3 \cdot P(Y = 3) \\ &= -1 \cdot \frac{3}{10} + 3 \cdot \frac{2}{10} \\ &= \frac{6 - 3}{10} = \boxed{\frac{3}{10} = 0.3} \end{aligned}$$

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

Solution: There is no way to use the definition of Variance to calculate this, because we do not actually have any sense of what the data are. So we use the formula:

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

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

$$E[Y] = E[X^2 - 1] = E[X^2] - 1$$

$$E[X^2] = E[Y] + 1 = 0.3 + 1 = 1.3$$

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:

$$\begin{aligned} \text{Var}(X) &= E[X^2] - E[X]^2 \\ &= 1.3 - (0.7)^2 \\ &= 1.3 - 0.49 \\ &= \boxed{0.81} \end{aligned}$$

(e) (4 points) Let F represent the cumulative distribution function (CDF) for X , and compute $F(0.25)$

$$\textbf{Solution: } F(0.25) = P(X \leq 0.25) = P(X = -1) + P(X = 0) = \boxed{4/10 = 0.4}$$