CSCI 3022
Midterm Exam
Fall 2018

Name & Section: $\underbrace{\text{Solutions - Wong \& Larremore - October 2018}}_{\text{By writing my name I promise to abide by the Honor Code}}$

Read the following:

- **RIGHT NOW**! Write your name and section on the top of your exam.
- **Then**! Flip your exam over and write your name on the back page.
- You are allowed one 8 1/2 × 11in sheet of **handwritten** 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 on the provided answer line.
- 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.
- If you need more space for free-response questions, there are blank pages at the end of the exam. If you choose to use the extra pages, make sure to **clearly** indicate which problem you are continuing.

Problem	Points	Score
1	2	
2	2	
3	3	
4	3	
5	3	
6	3	
7	3	
8	3	
9	3	
10	3	
11	3	
12	3	
13	3	
14	3	

Problem	Points	Score
Mult. Choice Total	40	
15	20	
16	20	
17	20	
Total	100	

1. (2 points) Jimmy the Octopus is recovering from a ski accident, and is regaining his dexterity by tossing around a football with his buddies. But his suckers are just not gripping the ball like they used to. He goes to the doctor, Dr. Cøraline Reefsdottír. The doctor sees that Jimmy has over 320 suckers across his 8 tentacles, and decides to number them all and evaluate the suction on every 16th sucker. What kind of sample is this?

   A. Systematic

   B. Thorough

   C. Stratified

   D. Unbiased

   E. Census

1. **A**

2. (2 points) Katie, arriving at a bus stop, just misses the bus. Suppose that she decides to walk if the (next) bus takes longer than 8 minutes to arrive. Suppose also that the time in minutes between the arrivals of buses at the bus stop is a continuous random variable with a $U(4, 9)$ distribution. What is the probability that Katie will end up walking.

   A. 1

   B. 4/5

   C. 3/5

   D. 1/5

   E. 1/2

2. **D**

3. (3 points) If y can be any real number, what values can the median of this dataset take?

$$[1, 2, \pi, y, 5]$$

   A. $[1, 5]$

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

   C. $[2, \pi]$

   D. $(2, \pi)$

   E. In this case, the median is undefined since π is irrational.

   F. In this case, the median is undefined since π is unreasonable.

3. **C**

4. (3 points) Rebecca has finally purchased a new TV that has way more graphics, and so, at last, she sets about beating the hot new video game, Fortnite. Let A be the event that Rebecca beats Fortnite on her first try. Let B be the event that Rebecca gets a new high score. Suppose that the probability that A occurs is 0.4, the probability that B occurs is 0.5, and the probability that neither A nor B occurs is 0.3. What is $P(A \cap B)$?

   A. 0.7
   B. 0.6
   C. 0.2
   D. 0.1
   E. 0.01
   F. None of the Above

4. **C**

5. (3 points) You are stuck in a YouTube clickhole, late on a Thursday night, and cannot stop yourself from clicking Next...Next...Next...on all the videos. Homework is due for CSCI 3022 the next day, but you simply cannot get enough of these amazing clips of Violet the Octopus playing Plinko. What a legend! A new notification appears on your mobile, and it says:

$$f(x) = kx - x^2 \text{ for } 0 \leq x \leq 1 \text{ and } f(x) = 0 \text{ for any other value of } x$$

What value of k makes $f(x)$ a valid probability density function?

   A. 8/3
   B. 4/3
   C. 2/3
   D. 1/4
   E. None of the above.

5. **A**

6. (3 points) You are taking your 3022 midterm when you start thinking about octopuses for some reason. Consequently, consider the following:

   The probability that a problem on a midterm exam is related to octopuses is 3/4 if Dan and Tony are tired, and is 1/2 if they are not tired. Dan and Tony are tired with probability 1/4. Given that this problem is about an octopus named Frederique, what is the probability that Dan and Tony are le tired?

   A. 1/4
   B. 2/4
   C. 3/4
   D. 1/3
   E. 2/3
   F. 2/5
   G. 3/5
   H. 4/5

6. **D**

7. (3 points) Claire is in CSEL. It is basically Antarctica in there. Everyone is freezing. On average, the people studying in CSEL get hypothermia at a rate of 3 people per hour. What a mess. Michael tries to adjust the thermostat. Nothing happens. Other Michael also tries to adjust the thermostat - still nothing. Which expression, below, computes the probability that 2 or more people get hypothermia in CSEL in any particular hour?

   A. $\frac{3^2}{2!}e^{-3}$
   B. $1 - \sum_{k=0}^{1} \frac{3^k}{k!}e^{-3}$
   C. $1 - \sum_{k=0}^{2} \frac{3^k}{k!}e^{-3}$
   D. $\frac{2^3}{3!}e^{-2}$
   E. $\int_{k=2}^{\infty} \frac{3^k}{k!}e^{-3}dk$

7. ______**B**______

8. (3 points) You are hanging out with your friends, flipping through memes on the socials. Everyone there agrees that the crew should get off the couch and do something with the night. You attended a seminar on leadership, so you say, "Let us keep browsing these memes until we lol at 4 more good ones. Since each meme produces a lol independently with probability p, then the probability that we browse through M memes total by the time we finally observe the fourth good one will be well modeled by . . ."

   A. a geometric distribution
   B. a poisson distribution
   C. a uniform distribution
   D. an exponential distribution
   E. a binomial distribution
   F. a negative binomial distribution
   G. a hypergeometric distribution
   H. Impossible; you'd never stop at 4.

8. ______**F**______

9. (3 points) The background radiation rate as measured in a particular room with a particular Geiger counter is known to be 800 counts per hour on average. You run the Geiger counter in the room for a five second interval and write down the count. You run the Geiger counter again for another five second interval and write down that count. You repeat, and write down a third count. And a fourth. And a fifth ... . As you store your data in a convenient Pandas dataframe, you begin to notice that your measurements appear to be drawn from which of the following distributions?

   A. Binomial
   B. Poisson
   C. Geometric
   D. Uniform
   E. Exponential
   F. None of the Above

9. ______**B**______

10. (3 points) Some engineers use the Rankine temperature scale. To convert from a temperature in Rankine T_R to a temperature in Celsius T_C, the following conversion is used.

$$T_C = \frac{T_R - 491.67}{1.8}$$

Suppose that a temperature is represented by the random variable X, whose units are in Rankine, with standard deviation 6 Rankine. Let Y be the equivalent random variable with units converted to degrees Celsius. What is the variance of Y, rounded to one decimal point?

A. 36.0

B. 1.9

C. -269.8

D. 269.8

E. 11.1

F. The null hypothesis should be rejected.

10. **E**

11. (3 points) Consider the following function related to incoming texts to a cellphone during class. What distribution does the return value of the function belong to?

```
def text_me_im_so_stoked_about_3022(q):
    j = 0
    t = np.random.exponential(1/q)
    while t <= 1:
        j += 1
        t += np.random.exponential(1/q)
    return j
```

A. Binomial

B. Geometric

C. Poisson

D. Uniform

E. Exponential

F. None of the Above

11. **C**

12. (3 points) Consider simulating the roll of a fair, six-sided die. Which of the following quantities does the following function estimate?

```
def rolling_rollers(num_samples):
    rolls = np.random.choice([1,2,3,4,5,6], size=num_samples)
    return np.sum(np.logical_and((rolls % 2)==1),(rolls > 2)) / np.sum((rolls % 2)==1)
```

A. $P(X > 2 \mid X$ **is Odd**)

B. $P(X$ is Odd $\mid X > 2)$

C. $P(X > 2 \cap X$ is Odd)

D. $P(X > 2 \cup X$ is Odd)

12. **A**

13. (3 points) Consider the following function related to finding an open parking spot in a large parking lot where the probability of an individual spot being open is given by p. What distribution does the return value of the function belong to?

```
import numpy as np
def shoulda_taken_the_bus(p):
    total_spots = 1
    successes = 0
    while np.random.choice([0,1], p=[1-p, p]) == 0:
        total_spots += 1
    successes += 1
    return total_spots
```

A. Binomial

B. Geometric

C. Poisson

D. Uniform

E. Exponential

F. The function above will produce an error if executed

G. None of the above

13. **B**

14. (3 points) It's Saturday night and you are at a party on the rooftop of Norlin Library. Everybody is there. The temperature is uniformly distributed between ideal and perfect. The music is loud, if you're into loud music, but also it is more on the quiet side, if you're more into that. Suddenly the music of your preferred volume stops and a voice comes over the speakers:

"Suppose there exists a continuous random variable X with probability density function $f(x)$!"

People scream, but you do not; you have been training for this moment.

"The variance of X is 12 and $\int_{-\infty}^{\infty} x^2 f(x)\, dx = 16$"

You smile. You're pretty sure you know where this is going.

"Compute $\int_{-\infty}^{\infty} x f(x)\, dx$ or the party is over! Answer me!"

What do you answer to get the party restarted? (Assume that you are trying to answer correctly because you want the rooftop party to continue.)

A. 2

B. 4

C. -4

D. Does not exist.

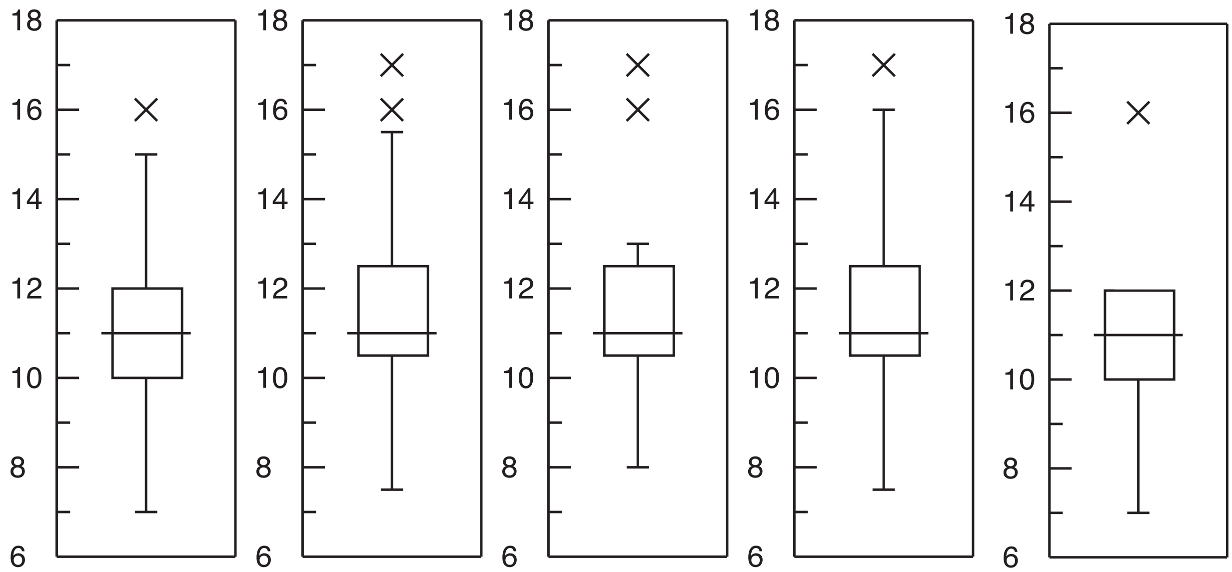
E. Cannot compute using the information given.

F. None of the above.

14. **A**

15. (20 points) Consider the following two datasets:

$$X = \{8,\ 8,\ 10,\ 11,\ 11,\ 11,\ 11,\ 12,\ 13,\ 16,\ 17\}$$
$$Y = \{7,\ 8,\ 10,\ 10,\ 11,\ 11,\ 12,\ 12,\ 12,\ 12,\ 16\}$$



(a) There are 4 boxplots and one empty plot above. For each dataset, X and Y, identify which boxplot, if any, represents that dataset. If the correct boxplot for one of the datasets is not provided, please correctly draw it in the blank axes. Use the conventions for boxplots introduced in class, and fully justify your choices using the space provided. Be sure to address all aspects of the boxplot.

Solution:

Let's start with X, my dudes!

- The values above are in order. There are $n = 11$ items. The Q_2 (also called the median) is the 6th item. So $Q_2 = 11$.
- To find the upper and lower quantiles, we split the dataset in two, and duplicate the median, placing it in both halves:

$$\{8,\ 8,\ 10,\ 11,\ 11,\ 11\} \quad \text{and} \quad \{\ 11\ 11,\ 12,\ 13,\ 16,\ 17\}$$

- Q_1 is the median of the left dataset, so $Q_1 = \frac{10+11}{2} = 10.5$
- Q_3 is the median of the right dataset, $Q_3 = \frac{12+13}{2} = 12.5$
- $\Rightarrow$ So the 2nd, 3rd and 4th plots are all candidates for X
- $IQR = 12.5 - 10.5 = 2$, so whiskers extend *at most* $1.5 \cdot 2 = 3$ above Q_3 and below Q_1
- Whiskers extend to the outermost datapoints within $[Q_1 - 3, Q_3 + 3] = [7.5, 15.5]$. That's 8 on the low side, and 13 on the high side.
- There are no data points below the lower whisker, so no low fliers. On the high side, 16 and 17 are both above the whisker, so they are fliers.
- The **THIRD** panel corresponds to X

And now, it is Y time!

- The values above are in order. There are $n = 11$ items. The Q_2 (also called the median) is the 6th item. So $Q_2 = 11$.
- To find the upper and lower quantiles, we split the dataset in two, and duplicate the median, placing it in both halves:

$$\{7,\ 8,\ 10,\ 10,\ 11,\ 11\} \quad \text{and} \quad \{\ 11\ 12,\ 12,\ 12,\ 12,\ 16\}$$

- Q_1 is the median of the left dataset, so $Q_1 = 10$
- Q_3 is the median of the right dataset, $Q_3 = 12$
- $\Rightarrow$ So the first plot is the only candidate at this point for Y
- $IQR = 12 - 10 = 2$, so whiskers extend *at most* $1.5 \cdot 2 = 3$ above Q_3 and below Q_1
- Whiskers extend to the outermost datapoints within $[Q_1 - 3, Q_3 + 3] = [7, 15]$. That's 7 on the low side, and 12 on the high side. (Yes, the whisker coincides with the box edge.)
- There are no data points below the lower whisker, so no low fliers. On the high side, 16 is above the whisker, so there is one flier at 16.
- Since the first plot fails at drawing the whisker properly, we must draw Y ourselves! See above.

(b) Classify each dataset, X and Y, as symmetric, positively skewed, or negatively skewed. Clearly justify your response.

Solution:

Let's start with X again. The median is $\tilde{x} = 11$. The mean is $\bar{x} = 11.64$. Because the mean is greater than the median, we call this distribution positively (or right) skewed.

For Y, the mean and median are both 11. So we would say this distribution is symmetric, or not skewed.

16. (20 points) Researchers are conducting an experiment to test for Extra-Sensory Perception (ESP). Their experiment involves showing a test-subject the back of a card with one of four shapes on the front side and asking them to correctly guess the shape. They perform this test on each test-subject 3 times. The study's designers know from experience that a normal person without ESP will correctly guess the shape on the card with probability 1/4. On the other hand, a person with ESP will guess the correct shape with probability 7/8. The researchers also know that 1 out of every 10 people has ESP.

(a) If a test-subject in fact has ESP, what is the probability that they correctly guess all 3 cards that they are shown?

Solution: Let $5C$ be the event that the test-subject gets 5 cards correct. If we condition on the fact that we know the ESP/non-ESP status of the test subject then the events that they get a particular card correct are independent. Thus, we have

$$p(5C \mid ESP) = \frac{7}{8} \cdot \frac{7}{8} \cdot \frac{7}{8} = \left(\frac{7}{8}\right)^3 \approx 0.6699$$

(b) What is the probability that any test-subject (whose ESP-status is unknown) correctly guesses all 3 cards that they are shown?

Solution: Let $3C$ again be the event that a test-subject gets all 3 cards correct. Using the Law of Total Probability, we have

$$p(3C) = p(3C \mid ESP)p(ESP) + p(3C \mid ESP^C)p(ESP^C)$$

$$= \left(\frac{7}{8}\right)^3 \cdot \frac{1}{10} + \left(\frac{1}{4}\right)^3 \cdot \frac{9}{10} \approx 0.08105$$

(c) Suppose that a test-subject (whose ESP-status is unknown) correctly guesses all 3 cards that they are shown. What is the probability that the test-subject has ESP?

Solution: We want to compute $p(ESP \mid 3C)$. By Bayes' Theorem, we have

$$p(ESP \mid 3C) = \frac{p(3C \mid ESP)p(ESP)}{p(3C)} = \frac{\left(\frac{7}{8}\right)^3 \cdot \frac{1}{10}}{\left(\frac{7}{8}\right)^3 \cdot \frac{1}{10} + \left(\frac{1}{4}\right)^3 \cdot \frac{9}{10}} \approx 0.8265$$

(d) Let C_1 be the event that the test subject correctly guesses the first card and C_2 be the event that the test-subject correctly guesses the second card. First, state an appropriate mathematical definition of independence. Then, use that definition to determine whether C_1 and C_2 are independent events.

Solution: To check if C_1 and C_2 are independent, we will check $p(C_1 \cap C_2) \stackrel{?}{=} p(C_1)p(C_2)$.
We have, by the Law of Total Probability, for C_i with $i = 1, 2$:

$$p(C_i) = p(C_i \mid ESP)p(ESP) + p(C_i \mid ESP^C)p(ESP^C) = \frac{7}{8} \cdot \frac{1}{10} + \frac{1}{4} \cdot \frac{9}{10} = 0.3125$$

Thus $p(C_1)p(C_2) = (0.3125)^2 = 0.09765625$

Next, we have

$$p(C_1 \cap C_2) = p(C_1 \cap C_2 \mid ESP)p(ESP) + p(C_1 \cap C_2 \mid ESP^C)p(ESP^C)$$

$$= \left(\frac{7}{8}\right)^2 \cdot \frac{1}{10} + \left(\frac{1}{4}\right)^2 \cdot \frac{9}{10} = 0.1328125$$

Since $p(C_1 \cap C_2) = 0.1328125 \neq 0.09765625 = p(C_1)p(C_2)$ we conclude that the events C_1 and C_2 are **NOT** independent.

17. (20 points) The probability distribution of a discrete random variable X is given by

$$P(X=-1)=\frac{1}{10},\quad P(X=0)=\frac{1}{10},\quad P(X=1)=\frac{3}{10},\quad P(X=2)=\frac{5}{10}$$

(a) Compute $E[X]$

Solution:

$$\begin{aligned}
E[X] &= \sum_x x\cdot P(X=x) = -1\cdot P(X=-1)+0\cdot P(X=0)+1\cdot P(X=1)+2\cdot P(X=2)\\
&= -1\cdot\frac{1}{10}+1\cdot\frac{3}{10}+2\cdot\frac{5}{10}\\
&= \frac{2}{10}+\frac{10}{10}=\frac{12}{10}=\boxed{\frac{6}{5}=1.2}
\end{aligned}$$

(b) Let Y be the random variable $Y=X^2+1$. Write down the probability distribution of Y.

Solution: The possible values of Y are 1, 2 and 5, as shown below:

$$(-1)^2+1=2,\quad (0)^2+1=1,\quad (1)^2+1=2,\quad (2)^2+1=5$$

The probabilities are:

$P(Y=1)=P(X=0)=\frac{1}{10}=0.1$

$P(Y=2)=P(X=1)+P(X=-1)=\frac{1}{10}+\frac{3}{10}=\frac{4}{10}=0.4$

$P(Y=5)=P(X=2)=\frac{5}{10}=0.5$

(c) Compute $E[Y]$

Solution:

$$\begin{aligned}
E[Y] &= \sum_y y\cdot P(Y=y) = 1\cdot P(Y=1)+2\cdot P(Y=2)+5\cdot P(Y=5)\\
&= 1\cdot\frac{1}{10}+2\cdot\frac{4}{10}+5\cdot\frac{5}{10}\\
&= \frac{1+8+25}{10}=\frac{34}{10}=\boxed{\frac{17}{5}=3.4}
\end{aligned}$$

(d) Compute 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 $E[Y]$:

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

$$E[X^2] = E[Y] - 1 = 3.4 - 1 = 2.4$$

$$\begin{aligned}\text{Var}(X) &= E[X^2] - E[X]^2 \\ &= 2.4 - (1.2)^2 \\ &= 2.4 - 1.44 \\ &= \boxed{0.96}\end{aligned}$$

(e) Let F represent the cumulative distribution function (CDF) for X, and compute $F(0.5)$

Solution: $F(0.5) = P(X \leq 0.5) = P(X = -1) + P(X = 0) = \boxed{2/10}$