CSCI 3022
Midterm Exam
Fall 2020

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

Name: Jonathan Phumirat
Student ID: 106054641
Section number: 001 1:50pm

Read the following:

- **RIGHT NOW!** Write your name, student ID and section number on the top of your exam. If you're handwriting your exam, include this information at the top of the first page!
- You may use the textbook, your notes, lecture materials, and Piazza as recourses. Piazza posts should not be about exact exam questions, but you may ask for technical clarifications and ask for help on review/past exam questions that might help you. You may not use external sources from the internet or collaborate with your peers.
- You may use a calculator.
- If you print a copy of the exam, clearly mark answers to multiple choice questions in the provided answer box. If you type or hand-write your exam answers, write each problem on their own line, clearly indicating both the problem number and answer letter.
- 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. For handwriting multiple choice answers, clearly mark both the number of the problem and your answer for each and every problem.
- 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.
- The Exam is due to Gradescope by midnight on Monday, October 26.
- When submitting your exam to Gradescope, use their submission tool to mark on which pages you answered specific questions. Submitting your exam properly is worth 1/100 points. The other problems sum to 99.

Multiple choice problems: Write your answers in the boxes if using a printed version of the exam.

1. (3 points) Under what standard conditions for *unimodal* data is the median value $\tilde{x}$ larger than the mean value $\bar{x}$?

   A. This always happens.

   B. This happens when the data is right-skew.

   C. This happens when the data is left-skew.

   D. This happens when the interquartile range is small.

   E. This can not happen.

   C

Not sure

Q1 Q2 Q3

2. (3 points) Consider the data set: $[0, 1, 4, 4, 6, 8, 9, 11, 12, 15, x]$, where $x \in \mathbb{R}$ is an unknown quantity. What is the (minimal) set of possible values to which the lower quartile of this data set **must** belong? Use Tukey's method, *including* the median in each half of the data.

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

   B. $[1, 4]$

   C. $[2.5, 4]$

   D. $\{4\}$

   E. $[4, 5]$

   F. $[4, 7]$

   G. $\emptyset$

   C

3. (3 points) Suppose Zach has a list consisting of all the first generation Pokémon. He is conducting a study of how many of them are actually stronger than Mudkip - the cutest Pokémon ever - by drawing a sample from his Pokédex, which he has sorted alphabetically.

   He checks the stats of every 7th Pokémon on his list and compares them to Mudkip's.

   What type of sample did Zach collect?

   A. Simple random sample

   B. Systematic sample

   C. Census sample

   D. Stratified sample

   E. Free samples, all you can eat!

   B

Use the following information for Problems 4 – 6, which may build off of each other.

It's time for a game of "Data Scientists Among Us," where two players compete. One player is the **"Data Scientist" (D)**, who treats their data with dignity and respect, and the other player is the **"Imposter" (I)**, a devilish rogue that breaks mathematics regularly.

Suppose the Data Scientist performs valid calculations **(V)** 80% of the time. On the other hand, the Imposter only does so 60% of the time.

4. (3 points) Half the time, *both* players make valid calculations. What is the exact probability that *neither* player makes valid calculations?

A. 0.1
B. 0.2
C. .25
D. 0.48
E. 0.5
F. 0.52
G. .75
H. 0.8
I. 0.9

F

5. (3 points) The Imposter is sneaky and tries to mimic the Data Scientist. *Given* the Data Scientist is *not* making valid calculations, the Imposter is equally likely to make valid or invalid calculations. What is the probability that the Imposter is making valid calculations *and* the Data Scientist is making invalid ones?

A. 0.1
B. 0.2
C. .25
D. 0.48
E. 0.5
~~F.~~ 0.52
G. .75
H. 0.8
I. 0.9

A

6. (3 points) What is the probability that the Data Scientist is making valid calculations *given* the Imposter is making invalid ones?

Not sure

~~A.~~ 0.1
B. 0.2
C. .25
D. 0.48
E. 0.5
~~F.~~ 0.52
G. .75
H. 0.8
I. 0.9

H

independent events

7. (3 points) Medical records show that, among patients suffering from a given disease, 75% will die of it within 5 years. Out of 10 people suffering from D, let X be the random variable counting the number of people who survive more than 5 years. What is an appropriate random variable for X?

? A. Binomial

~~B.~~ Negative binomial

? C. Uniform → all equally likely

~~D.~~ Normal

? E. Poisson

~~F.~~ Exponential

$\boxed{A}$

8. (3 points) Suppose when you get to the cafeteria, you and your best friend have a competition for pizza slices. For each and every slice, you each type NP.RANDOM.RAND() and the person who gets the higher number wins. A pizza with 16 slices is put before you. What is the probability that you get exactly 9 slices?

A. $\binom{16}{2}(0.5)^2(0.5)^16$

B. $\binom{16}{8}(0.5)^8(0.5)^8$

C. $\binom{16}{9}(0.5)^7(0.5)^9$

D. $1 - \sum_{i=9}^{16}\binom{16}{i}(0.5)^i(0.5)^{16-i}$

E. $1 - \sum_{i=0}^{8}\binom{16}{i}(0.5)^{16-i}(0.5)^i$

F. $1 - \sum_{i=9}^{16}\binom{i}{16}(0.5)^i(0.5)^{16-i}$

$\boxed{C}$

9. (3 points) Suppose we know that the general antiderivative of a function $g(x)$ is $\int g(x)\,dx = (x-1)e^{x-2} + C$. Then from $[1, 2]$ the cumulative density function of a random variable with pdf $f(x)$,

$$f(x) = \begin{cases} g(x) & 1 < x \leq 2 \\ 0 & else \end{cases}$$

is given by:

A. $F(x) = (x-1)e^{x-2} + C$

B. $F(x) = (x-2)e^{x-2} + C$

C. $F(x) = (x-1)e^{x-2}$

D. $F(x) = (x-2)e^{x-2}$

E. $F(x) = xe^{-\lambda x}$

F. None of the Above.

$\boxed{C}$

10. (3 points) Suppose we have a random variable X satisfying $E[X^2] = a$ and $Var[X^2] = b$. What is $E[X^4]$?

A. 0

B. a^2

C. b^2

D. $b + a^2$

E. $\sqrt{b}$

F. $b^2 - a^2$

G. $a^2 + b^2$

D

Come back

11. (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 temperature in Fahrenheit F by $C = \frac{5}{9}(F - 32)$, what is the *variance* 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. $11 \cdot \left(\frac{5}{9}\right)^2$

B

12. (3 points) Consider the following function, where the probability p is some constant. What is the *average* return value for this function?

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

~~A.~~ p

B. $5p$

C. $1/p$

D. $5/p$

~~E.~~ $\frac{5p}{1-p}$

F. p^2

B

13. (3 points) You are sampling the weights of various puppies from a population with a known mean of 15 pounds and variance of 16 pounds^2. You obtain a measurement from an adorable Beagle of $X = 19$ pounds. What is the corresponding value of the standardized normal random variable, Z?

A. 0.25

B. 0.5

C. 1

D. $\frac{19}{16}$

E. 2

F. $\frac{19}{4}$

G. 15

$\bar{x} = 15 \text{lbs}$

$\text{var} = 16 \text{lbs}^2$

$X = 19 \text{lbs}$

$Z = \frac{X - \mu}{\sigma}$

$\boxed{C}$

Free Response problems: Write your answers in the spaces following each prompt if possible. Make note if your work continues elsewhere!

14. (10 points) You are in awe of your desk plant Fernoulli Jr.'s grandeur. It's growing so successfully that you're considering renaming it Fernomial! It's now time for *quantifying* Fernoulli Jr.'s majesty. Suppose your plant has 324 leaves which each have length independently and identically distributed from the normal distribution with mean 5 cm and variance 1.5 cm.

(a) (5 points) You admire a branch that contains 10 leaves. What is the exact distribution of the *average* leaf length of the 10 leaves along that branch? Cite any relevant theorems.

(b) (5 points) In your research, a botanist tells you that the distribution of colors of leaves (via spectral analysis) is *also* independently distributed, and that you should expect a mean wavelength of 515nm with standard deviation of 10*nm*. The exact distribution is unknown, however. What conclusions can you draw about the Fernoulli's *average* spectral leaf color? Cite any relevant theorems.

a) Being asked to find distribution of average leaf length of 10 leaves so find the standard deviation of between the 10 leaves of the branch. Will utilize the central limit theorm $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

mean $\rightarrow$ $\mu = 5\text{cm}$
variance $\rightarrow$ $\sigma^2 = 1.5\text{cm}$
amounts $\rightarrow$ ~~x~~ $= 10$ leaves

We know $\mu = 5$ because our sample follows the normal distribution and $\frac{\sigma^2}{n} = \frac{(1.5)}{10\text{cm}}$ since its given. So we can say $\bar{x}_{10} \sim N(5, \frac{1.5}{10})$ and now that we know that this is normal can find the standard deviation by finding the standard error for 10 leaves. $S.E.(\bar{x}) = \frac{\sigma}{\sqrt{n}} \rightarrow \frac{\sqrt{1.5}}{\sqrt{10}}$. $\boxed{\text{Exact distribution is } \frac{\sqrt{1.5}}{\sqrt{10}}}$

b) Based on the information we should be able to conclude that on average, the leaf colors should be fairly similar in color. We can find the actual value of average by applying the central limit theorm and calculate the $S.E(\bar{x})$ since this is a normal distribution.

$S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}} \Rightarrow \sigma = 10\text{nm}, \; n = 324 \; = \boxed{S.E(\bar{x}) = \frac{10}{\sqrt{324}}}$

15. (15 points) You are an analyst charged with the task of gauging support for a new ballot measure. You find that the probability of a Democrat supporting this ballot measure is 0.2, the probability of a Republican supporting this ballot measure is 0.8, and the probability of an Independent supporting the measure is 0.4. Furthermore, you know that in your area, 60% of voters are registered Democrats, 30% are registered Republicans, and 10% are registered Independents.

(a) (5 points) You interview a voter at random. What is the probability that they support this ballot measure?

(b) (5 points) You interview someone at random and find out that they support this ballot measure. Given this information, what is the probability that they are a Republican?

(c) (5 points) Are the events "voter is a Republican" and "voter supports this ballot measure" independent? Justify your answer using math.

a) $P(\text{support}) = \left(\frac{6}{10}\right)\left(\frac{2}{10}\right) + \left(\frac{3}{10}\right)\left(\frac{8}{10}\right) + \left(\frac{1}{10}\right)\left(\frac{4}{10}\right) \Rightarrow P(\text{support}) = \boxed{40\%}$

b) $P(\text{Rep}|\text{support}) = \frac{P(\text{support}|\text{Rep}) \cdot P(\text{Rep})}{P(\text{support})} = \frac{\left(\frac{8}{10}\right)\cdot\left(\frac{3}{10}\right)}{\left(\frac{4}{10}\right)} = \boxed{60\%}$

If $P(A|B) \neq P(A)$ then we know the events are dependent. A = voter is rep
B = supports the ballot

$P(A) = \frac{3}{10}$ $P(A|B) = \frac{8}{10}$

$P(A) \neq P(A|B)$, these are $\boxed{\text{DEPENDENT}}$.

16. (15 points) You've discovered that games which involve rolling dice are often pretty boring, and you could spice them up a bit with a die that is weighted. Suppose we have a six-sided die that satisfies the following claim: "The probability of rolling each number is proportionate to the *square* of the face." In other words, a 2 is 4 times more likely than a 1, a 3 is 9 times more likely than a 1, and so forth.

(a) (7 points) What is the probability mass function for the face of the die?

(b) (4 points) You roll the die twice and sum the faces. What is the probability you have rolled a 3 or less?

(c) (4 points) You decide to roll the die until you observe two consecutive 1's. What is the probability that it takes you exactly 4 rolls for this to happen (so the third and fourth rolls were each "1").

a) $P(X=1) = \frac{1}{91}$

$P(X=2) = \frac{4}{91}$

$P(X=3) = \frac{9}{91}$

$P(X=4) = \frac{16}{91}$

$P(X=5) = \frac{25}{91}$

$P(X=6) = \frac{36}{91}$

$\boxed{\text{pmf} = P(x) = \frac{x^2}{91}}$

b) trials, $P(\text{sum} \leq 3)$? Law of Total probability

2 dice are rolled and sum

$P(\text{sum} \leq 3) = 2(P(X=1)\cdot P(X=2)) + P(X=1)^2$

$\boxed{P(\text{sum} \leq 3) = 2\left(\frac{1}{91}\cdot\frac{4}{91}\right) + \left(\frac{1}{91}\right)^2}$

c) $P(\text{trials}=4)$ Another Total Law of probability

$P(\neg\text{one}) \cdot P(\neg\text{ones}) \cdot P(\text{one}) \cdot P(\text{one})$

$\left(\sum_{i=2}^{6} f(i)\right)^2 \cdot \left(\frac{1}{91}\right)^2$

+

$P(\text{one}) \cdot P(\neg\text{one}) \cdot P(\text{one}) \cdot P(\text{one})$

$\sum_{i=2}^{6} f(i) \cdot \left(\frac{1}{91}\right)^3$

$P(4 \text{ trials}) =$

$\left(\sum_{i=2}^{6} f(i)\right)^2 \cdot \left(\frac{1}{91}\right)^2$

+

$\sum_{i=2}^{6} f(i) \cdot \left(\frac{1}{91}\right)^3$

17. (20 points) Mathematically justify all answers. Suppose you have a probability distribution of the form

$$f(x) = \begin{cases} a + bx & \text{for } x \in [0, 2] \\ 0 & \text{else} \end{cases}$$

where a and b are some unknown constants (real numbers).

(a) (4 points) Leaving a and b constant, integrate the pdf over the entire range of X.

(b) (4 points) Compute the expected value of X as a function of a and b.

(c) (4 points) Suppose you are only now told that $E[X] = \frac{10}{9}$. Use this and the information in parts (a) and (b) to to solve for the constants a and b.

(d) (4 points) Compute the median of X.

(e) (4 points) Compute Var(X).

a) $\int_0^2 a + bx\,dx \Rightarrow \int_0^2 f(x) \Rightarrow ax + \frac{bx^2}{2}\Big|_0^2$

$\Rightarrow \left[2a + \frac{4b}{2}\right] - \left[a(0) + \frac{b(0)^2}{2}\right] \Rightarrow 2a + \frac{4b}{2}$

$\Rightarrow \boxed{\int_0^2 f(x) = 2a + 2b}$

b) $f(x) : a + bx$

eq. $\int_0^2 x f(x)\,dx$ solve. $\int_0^2 x(a+bx)\,dx \Rightarrow \int_0^2 ax + bx^2\,dx$

$\Rightarrow \frac{ax^2}{2} + \frac{bx^3}{3}\Big|_0^2 = \frac{4a}{2} + \frac{8b}{3} \Rightarrow \boxed{2a + \frac{8b}{3}}$

c) from a & b we know (systems of equations)

$2a + 2b = 1$

$2a + \frac{8b}{3} = \frac{10}{9}$

$\frac{8b}{3} - 2b$

$\frac{8b}{3} - \frac{6b}{3} = \frac{2b}{3}$

solve for b first

$-(2a + 2b = 1)$

$2a + \frac{8b}{3} = \frac{10}{9}$

$\frac{2b}{3} = \frac{1}{9}$

$2b = \frac{3}{9}$

$b = \frac{3}{18}$

$\boxed{b = 1/6}$

now solve for a

$2a + \frac{2}{6} = 1$

$2a = 1 - \frac{2}{6}$

$2a = \frac{4}{6}$

$a = \frac{4}{12}$

$\boxed{a = \frac{1}{3}}$

Problem continued on separate page

d) Compute the median

eq. $a=\frac{1}{3}$ $f(x) = a+bx$ $\qquad \int \frac{1}{3} + \frac{x}{6}\,dx = \frac{1}{3}x + \frac{x^2}{12}$

$b=\frac{1}{6}$ $f(x) = \frac{1}{3}+\frac{1}{6}x$

set equal to .5

$$\frac{x}{3} + \frac{x^2}{12} = \frac{1}{2}$$

$$\frac{4x}{12} + \frac{x^2}{12} = \frac{1}{2} \Rightarrow \frac{x^2+4x}{12} = \frac{1}{2} \Rightarrow x^2+4x = 6$$

use quadratic equation

$$\frac{-b \pm \sqrt{b^2-4ac}}{2a}$$

$\rightarrow x^2+4x-6=0$ $\quad a=1$, $b=4$, $c=-6$ * solve for roots

$$\frac{-4 \pm \sqrt{4^2-4(1)(-6)}}{2(1)} \Rightarrow \frac{-4 \pm \sqrt{16+24}}{2} \Rightarrow \frac{-4 \pm \sqrt{40}}{2}$$

, so $\tilde{x}$ at $\frac{-4-\sqrt{40}}{2}$ and $\frac{-4+\sqrt{40}}{2}$

Since we are bounded at positive numbers $\boxed{\tilde{x} = \frac{-4+\sqrt{40}}{2}}$

e) Compute Var[X]

eq. $Var(x) = E[X^2] - [E[X]]^2$

We know $E[X] = \frac{10}{9}$ so $[E[X]]^2 = \left(\frac{10}{9}\right)^2$

need to compute $E[X^2]$

$$\int_0^2 x^2 f(x) \rightarrow \int_0^2 x^2(a+bx)\,dx \rightarrow \int_0^2 ax^2+bx^3\,dx$$

*plug in a & b

$$\rightarrow \frac{ax^3}{3} + \frac{bx^4}{4}\Big|_0^2 \rightarrow \frac{2^3 a}{3} + \frac{2^4 b}{4} \rightarrow \frac{2^3(\frac{1}{3})}{3} + \frac{2^4(\frac{1}{6})}{4}$$

$$\boxed{Var(x) = \frac{2^3(\frac{1}{3})}{3} + \frac{2^4(\frac{1}{6})}{4}}$$