TEST #2
STA 5326 **Name:**__________________________________
November 6, 2023

Please read the following directions.
DO NOT TURN THE PAGE UNTIL INSTRUCTED TO DO SO

- The exam is closed book and closed notes. You will be supplied with scratch paper, and a copy of the Table of Common Distributions from the back of our textbook.
- During the exam, you may use what you need to write with (pens, pencils, erasers, etc) and an ordinary scientific calculator.
- All other items (INCLUDING CELL PHONES) must be left at the front of the classroom during the exam. This includes backpacks, purses, books, notes, etc. You may keep small items (keys, coins, wallets, etc., but NOT CELL PHONEs) so long as they remain in your pockets at all times.
- Your calculator should NOT be able to do algebra or calculus. It should have a small screen and limited memory. It should NOT have any internet or phone capability. Computers or tablets are NOT allowed.
- You must show and explain your work for all the problems (except for the last problem labeled **NWR**). **No credit is given without work or explanation!**. But don't get carried away! Give enough explanation and work so that what you have done is clearly understandable.
- Partial credit is available (except for problem labeled **NWR**). If you know part of a solution, write it down. If you know an approach to a problem, but cannot carry it out – write down this approach. If you know a useful result, write it down.
- **No work is required** for the very last problem which is marked **NWR**. For this problem, you will receive full credit just for stating the correct answer.
- Make sure that the grader can easily see how you get from one step to the next. If you needed scratch paper to work something out, make sure to transfer your work to the exam.
- You should give only one answer to each problem. **Circle your answer** if there is any chance for confusion.
- Simplify your answers when it is easy to do so. But more difficult arithmetic does **not** have to be done completely unless specifically requested. Answers can be left as fractions or products. You do not have to evaluate large binomial coefficients, factorials or powers. Answers can be left as summations (unless there is a simple closed form such as when summing a geometric or exponential series).
- All algebra and calculus must be done completely. (Only arithmetic can be left incomplete.)
- Do **not** quote homework results. If you wish to use a result from homework in a solution, you must prove this result.
- All the work on the exam should be your own. No "cooperation" is allowed.
- The exam has **8** problems and **9** pages. There are a total of **100** points.

Problem 1. Suppose X has the probability mass function (pmf)

$$f(x) = (1-p)^2(x+1)p^x \quad \text{for } x = 0, 1, 2, 3, \ldots.$$

where $0 < p < 1$.

In the parts (a) and (b) below, **do NOT use any results from the Appendix** (the Table of Distributions). You may use without proof the fact that the formula above is a valid pmf for all p satisfying $0 < p < 1$.

(a) (10%) Show that the moment generating function (mgf) of X is

$$M(t) = \frac{(1-p)^2}{(1-pe^t)^2}.$$

This part is similar to exercise 2.30(d) and also 2.38(a).

$$\begin{aligned}
M(t) = Ee^{tX} &= \sum_{x=0}^{\infty} e^{tx}(1-p)^2(x+1)p^x = (1-p)^2 \sum_{x=0}^{\infty}(x+1)\left(pe^t\right)^x \\
&= \frac{(1-p)^2}{(1-p')^2} \sum_{x=0}^{\infty}(1-p')^2(x+1)\left(p'\right)^x \quad \textit{where } p' = pe^t \\
&\quad \textit{(If } p' < 1\textit{, then the sum equals one since it is summing a pmf over all possible values.)} \\
&= \frac{(1-p)^2}{(1-p')^2} = \frac{(1-p)^2}{(1-pe^t)^2}
\end{aligned}$$

Note: the pmf $f(x)$ is actually just a Negative Binomial pmf with $r = 2$ in which the roles of p and $1-p$ are interchanged. Students who notice this can get the mgf from the mgf of the Negative Binomial given in the Appendix just by replacing p everywhere by $1-p$. That is why students are not allowed to quote from the Appendix in parts (a) and (b).

(b) (3%) What is the range of values of t for which the mgf of X is finite (well-defined) and given by the above formula?

The mgf is finite for $t < -\log p$.

The summation for the mgf in part (a) converges only when $pe^t < 1$ so that the terms $(pe^t)^x$ go to zero as $x \to \infty$. Taking logs on both sides, $pe^t < 1$ is equivalent to $t < -\log p$.

[**Problem 1 continued**]

In the next part you are allowed to use the Appendix.

(c) (10%) Define a new random variable by $Y = (1-p)X$. Use mgf's to show that, as $p \uparrow 1$, the random variable Y converges in distribution to a random variable with a Gamma distribution. Specify the values of the parameters α and β for this Gamma distribution.

Note: $p \uparrow 1$ means the same as $p \to 1-$.

This part is similar to exercise 2.38(b).

As $p \uparrow 1$, the mgf of Y converges to $\frac{1}{(1-t)^2}$ for $t < 1$ which is the mgf of a Gamma($\alpha = 2, \beta = 1$) random variable. Since convergence of mgf's implies convergence of distributions, we know that Y converges in distribution to a Gamma($\alpha = 2, \beta = 1$) distribution as $p \uparrow 1$.

$$M_Y(t) = M_X((1-p)t) = \left(\frac{1-p}{1-pe^{(1-p)t}} \right)^2 \quad \textit{for } (1-p)t < -\log p \textit{ or } t < \frac{-\log p}{1-p}$$

so that

$$\begin{aligned}
\lim_{p\uparrow 1} M_Y(t) &= \lim_{p\uparrow 1} \left(\frac{1-p}{1-pe^{(1-p)t}} \right)^2 = \left(\lim_{p\uparrow 1} \frac{1-p}{1-pe^{(1-p)t}} \right)^2 = \left(\lim_{p\uparrow 1} \frac{1}{e^{(1-p)t}} \cdot \frac{1-p}{e^{-(1-p)t}-p} \right)^2 \\
&= \left(\lim_{p\uparrow 1} \frac{1}{e^{(1-p)t}} \cdot \lim_{p\uparrow 1} \frac{1-p}{e^{-(1-p)t}-p} \right)^2 = \left(1 \cdot \lim_{p\uparrow 1} \frac{-1}{te^{-(1-p)t}-1} \right)^2 = \frac{1}{(1-t)^2}\,.
\end{aligned}$$

Basically, we are using L'Hospital's rule to evaluate the limit, but first pulling out a factor to make this easier. (If you want, you could apply L'Hospital's immediately without pulling out the factor.)

Problem 2. Suppose that X has density given by

$$f(x) = \frac{7^8}{x^8} \quad \text{for } x > 7\,.$$

(a) (10%) Find EX^k for all positive integers k. (More precisely, carefully state the values of k for which EX^k is well-defined and finite, and give the value of EX^k for these values of k.)

Similar to notes4.pdf, page 23.

The distribution of X is a special case of the Pareto distribution.

The moments are well-defined and finite for $k \leq 6$, and are $+\infty$ or undefined for $k \geq 7$.

$EX^k = \dfrac{7^{k+1}}{7-k}$ *for* $k \leq 6$.

(b) (6%) What is the value of $P(X > 11 \,|\, X > 9)$?

See notes6.pdf, pages 11-12.

$P(X > x) = (7/x)^7$ *for* $x \geq 7$ *so that* $P(X > 11 \,|\, X > 9) = \dfrac{P(X > 11)}{P(X > 9)} = \left(\dfrac{9}{11}\right)^7.$

Problem 3. In the fictional slightly red state of Arozida, 50.5% of the registered voters plan to vote Republican in the next election, 49.0% plan to vote Democrat, and 0.5% plan to vote Libertarian. A random sample of 400 registered voters is selected from this state.

In **both** parts (a) and (b) below:
use an appropriate approximation and **give a decimal answer**.

(a) (12%) What is the approximate probability the sample contains at least 201 people who plan to vote **Democrat** in the next election?

Use a normal approximation with a continuity correction.

Let X be the number of voters in the sample who plan to Democrat. Then $X \sim$ Binomial$(n = 400, p = 0.49)$. Let $X^* \sim$ Normal$(\mu = np = 196, \sigma^2 = np(1-p) = 99.96)$.

$$P(X \geq 201) \approx P(X^* \geq 201 - 0.5) = 1 - \Phi\left(\frac{200.5 - 196}{\sqrt{99.96}}\right) = 1 - \Phi(.45009)$$
$$\approx 1 - \Phi(.45) = 1 - .67364 = 0.32636$$

The continuity correction does make a definite improvement in this case, and students should lose some credit if they don't use it or use it incorrectly. If students use the 'normal' button on their calculator, they will get 0.3263228. If students use the 'binomial' button on their calculator, they wil get the exact answer of 0.326253. But if they have only done this and have not also done the normal approximation, they should lose a lot of credit because they were told to use an appropriate approximation.

(b) (6%) What is the approximate probability the sample contains exactly 3 people who plan to vote **Libertarian**?

Use a Poisson approximation to the Binomial distribution.

Let Y be the number in the sample who plan to vote Libertarian.

$Y \sim$ Binomial$(n = 400, p = 0.005) \approx$ Poisson$(\lambda = np = 2)$ so that $P(Y = 3) \approx \dfrac{2^3 e^{-2}}{3!} = 0.180447$. (Actually, I don't care too much about the final decimal answer in this part and students should receive full credit even if they stopped with $\dfrac{2^3 e^{-2}}{3!}$. However, students should use the Poisson approximation. If they only compute the exact Binomial probability and do **not** *do the Poisson approximation, they should lose lots of credit; maybe half the points.)*

Problem 4. (11%) Suppose X has density

$$f(x) = \frac{3x^5}{\beta^2} e^{-x^3/\beta} \quad \text{for } x > 0.$$

Find EX^2.

The calculations involved are similar to those in exercise 3.24(a).

Express EX^2 *as an integral and then make the substitution* $u = x^3/\beta$. *This converts the integral into the form of a gamma integral so that we get* $EX^2 = \beta^{2/3}\Gamma(8/3)$ *which can be re-expressed as* $10\beta^{2/3}\Gamma(2/3)/9$ *(and in other ways too) using* $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ *for* $\alpha > 0$.

Problem 5. (11%) Use an argument involving indicator random variables to give an expression for

$$P(A^c \cap B \cap C^c) + P(A \cap B^c \cap C)\,.$$

This expression should only contain terms chosen from the list

$$1, P(A), P(B), P(C), P(A \cap B), P(A \cap C), P(B \cap C), P(A \cap B \cap C)\,.$$

This resembles the lecture example on page 12 of notes4.pdf and also problem C0.

Here is a very detailed solution.

$$
\begin{aligned}
& P(A^c \cap B \cap C^c) + P(A \cap B^c \cap C) && (1)\\
&= EI_{A^c \cap B \cap C^c} + EI_{A \cap B^c \cap C}\\
&= E\left(I_{A^c \cap B \cap C^c} + I_{A \cap B^c \cap C}\right)\\
&= E\left(I_{A^c} I_B I_{C^c} + I_A I_{B^c} I_C\right)\\
&= E\left((1 - I_A) I_B (1 - I_C) + I_A (1 - I_B) I_C\right) && (2)\\
&= E\left([I_B - I_B I_C - I_A I_B + I_A I_B I_C] + [I_A I_C - I_A I_B I_C]\right) && (3)\\
&= E\left(I_B - I_B I_C - I_A I_B + I_A I_C\right)\\
&= E\left(I_B - I_{B \cap C} - I_{A \cap B} + I_{A \cap C}\right)\\
&= EI_B - EI_{B \cap C} - EI_{A \cap B} + EI_{A \cap C}\\
&= P(B) - P(B \cap C) - P(A \cap B) + P(A \cap C) && (4)
\end{aligned}
$$

One can vary the details and the order of the steps. It is OK if students show less detail than this so long as there is still enough to see what they are doing. Even if students show **only** *the numbered steps above, that is enough for full credit.*

Many students will probably tackle $P(A^c \cap B \cap C^c)$ and $P(A \cap B^c \cap C)$ separately, develop an expression for each probability, and then add these together to get the same final answer. This is also correct.

Problem 6. (8%) A nuclear power plant had a small radiation leak which escaped into the atmosphere and spread around the world, very slightly elevating the risk of death from cancer for everyone on Earth. Suppose that all of the 8.1 billion people on earth are placed in order 1, 2, 3, ..., 8.1×10^9 according to their degree of exposure to this radiation, and that person i in this ordering has probability $p_i = 10^{-6}\exp(-10^{-6}\, i)$ of dying from cancer due to this radiation leak.

What is the approximate probability that at least 2 people die of cancer due to this radiation?

See notes6.pdf, pages 30-31 and the solution to problem C2.

Let X be the number of people who die from cancer due to this radiation. The probabilities p_i are all very small. So, if the events $A_i = \{$person i gets cancer from this radiation$\}$ are mutually independent (which is an unstated assumption of this problem), then X should be approximately Poisson. The series $\sum_i p_i$ is a geometric series, and so easily summed; we obtain

$$X \sim \mathit{Poisson}\left(\lambda \approx \sum_{i=1}^{\infty} 10^{-6} e^{-10^{-6} i} = \frac{10^{-6}\exp(-10^{-6})}{1-\exp(-10^{-6})} = 0.9999995 \approx 1\right)$$

so that $P(X \geq 2) \approx 1 - f(0) - f(1) = 1 - 2\exp(-1) \approx 0.2642$ *where f is the Poisson*(1) *pmf* $f(x) = \exp(-1)/x!$.

Problem 7. (8%) A lightbulb has an exponentially distributed lifetime with a mean of 10,000 hours. If it has been in use for 13,268 hours and is still working, what is the approximate probability it will fail during the next hour of use? (Give a decimal answer.)

See page 4 of notes7.pdf.

The probability that a bulb which is still working at time t will fail in the next small period of time δ can be found using the hazard function $h(\cdot)$ and is approximately $h(t)\delta = (1/\beta)\delta = 10^{-4}\times 1 = 10^{-4}$.

Students who don't remember this result about the hazard function can still figure out the answer in a couple of ways.

First: The memoryless property of the exponential distribution says that the remaining lifetime of the lightbulb after 13,268 hours (call this Y) is still exponential($\beta = 10000$) so that

$$P(\textit{fail in next hour}) = P(Y < 1) = \int_0^1 \frac{1}{\beta} e^{-y/\beta}\, dy = 1 - e^{-1/\beta} = 1 - e^{-10^{-4}} \approx 10^{-4}.$$

Second: One can also get the answer right from the definition of conditional probability, essentially re-deriving the memoryless property. Let X denote the lifetime of the lightbulb.

$$\begin{aligned}
&P(13268 < X < 13268 + 1 \,|\, X > 13268) \\
&= \frac{P(13268 < X < 13268 + 1)}{P(X > 13268)} \\
&= \frac{F(x_0 + 1) - F(x_0)}{1 - F(x_0)} \quad \textit{where } x_0 = 13268 \textit{ and } F(t) = 1 - \exp(-t/\beta) \\
&= \frac{\exp(-x_0/\beta) - \exp(-(x_0+1)/\beta)}{\exp(-x_0/\beta)} = \frac{\exp(-x_0/\beta) - \exp(-(x_0+1)/\beta)}{\exp(-x_0/\beta)} \cdot \frac{\exp(x_0/\beta)}{\exp(x_0/\beta)} \\
&= 1 - \exp(-1/\beta) = 1 - \exp(-10^{-4}) \approx 10^{-4}
\end{aligned}$$

where F above denotes the cdf of the exponential(β) distribution.

Problem 8. (5%) **(NWR)** An urn contains 7 red balls and 13 green balls. The balls are randomly drawn from the urn one by one until the urn is empty. What is the probability that the last two balls drawn from the urn are both red?

See notes6.pdf, page 24.

The answer is $\frac{7}{20} \cdot \frac{6}{19}$. The probability the last two balls are red is the same as the probability the first two balls are red.