TEST #2
STA 5326 **Name:**__________________________________
November 3, 2021

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

Directions

- 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 ONLY what you need to write with (pens, pencils, erasers, etc) and (if you wish) an ordinary scientific calculator (TI-86 or below is fine).
- 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.
- Partial credit is available. (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.)
- You must show and explain your work (including your calculations) for all the problems (except those designated as "no work required"). **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.
- 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.
- **If your answer is valid only for a certain range of values, this should be stated as part of your answer. For example, if a density is zero outside of some interval, this interval should be stated explicitly.**
- 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. 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 **11** problems and **10** pages. There are a total of **100** points.

Problem 1. Let $X_1, X_2, X_3, \ldots, X_n$ be iid random variables with density $f(x) = 2(1-x)$ for $0 < x < 1$. Define $Y = \min_{1 \le i \le n} X_i$.

This problem is similar to exercise C3.

(a) (12%) Find the cumulative distribution function (cdf) of Y.

See page 7 of notes7.pdf.

Let F be the cdf of X_i. Then $F_Y(t) = 1 - (1 - F(t))^n$.

For $0 < t < 1$ we have $1 - F(t) = P(X_i > t) = \int_t^1 2(1-x)\,dx = (1-t)^2$ so that $F_Y(t) = 1-(1-t)^{2n}$ for $0 < t < 1$.

(b) (6%) Find EY.

Differentiating $F_Y(t)$ gives the pdf: $f_Y(t) = 2n(1-t)^{2n-1}$ for $0 < t < 1$.

$$EY = 2n \int_0^1 t(1-t)^{2n-1} = 2n\tfrac{\Gamma(2)\Gamma(2n)}{\Gamma(2n+2)} = \tfrac{2n(2n-1)!}{(2n+1)!} = \frac{1}{2n+1}\,.$$

Here we have recognized the integral as a Beta integral, which we can evaluate using the normalizing constant of the Beta distribution in the appendix. In the simplification we use the fact that $\Gamma(k) = (k-1)!$ for integers $k > 0$.

Problem 2. (12%) A hotel can accomodate up to 150 guests. The hotel has a restaurant in which it serves a complimentary (free) breakfast to its guests. From long experience it is known that each guest will choose with probability 0.4 to eat breakfast in the hotel (and with probability 0.6 they either skip breakfast or eat elsewhere). Assume that guests act independently.

What is the minimum capacity N the restaurant must have so that, when the hotel is full, the restaurant will have to turn away guests with probability at most 0.09? (Use a normal approximation.)

This is similar to exercise 3.8.

Let X be the number of guests who come to breakfast. We want the smallest integer N such that $P(X > N) \leq 0.09$. We know $X \sim$ Binomial$(150, 0.4)$. Let $X^* \sim N(\mu = 150 * 0.4 = 60, \sigma = \sqrt{150 * 0.4 * 0.6} = 6)$. Then using the continuity correction we have $P(X > N) \approx P(X^* > N + .5) = P(Z > (N + .5 - 60)/6)$ where $Z \sim N(0, 1)$. Using the normal table or calculator we find $P(Z > 1.34) \approx 0.09$ (a more accurate quantile is 1.3407550) and solving $(N+.5-60)/6 = 1.34$ gives $N = 67.54$ and rounding up gives $N = 68$. Note that rounding down does not give the correct answer since we desire $P(X > N) \leq 0.09$.

Students must use the continuity correction and do it correctly to receive full credit. It does not always make a difference in the answer, but they should do it anyway since you don't know in advance whether it will or not. Also, if you don't do the continuity correction, there is an ambiguity: Do you do $P(X > N) \approx P(X^* > N)$ or $P(X > N) = P(X \geq N + 1) \approx P(X^* \geq N + 1)$?

Problem 3. (12%) Use indicator random variables to prove the principle of inclusion-exclusion for the case of 3 events.

This is exercise C0. See the posted solution.

Problem 4. (12%) Let X have the cumulative distribution function (cdf)

$$F(x) = \frac{x}{1+x} \quad \text{for } x \geq 0.$$

For $t > 0$, find the value of

$$\lim_{\delta \downarrow 0} \frac{1}{\delta} P(t < X \leq t + \delta \,|\, X > t)\,.$$

(In your solution you may use the result of a homework exercise without proof.)

This limit is the definition of the hazard function $h(t)$. Differentiate F to find the density f, and then use $h(t) = f(t)/(1 - F(t))$.

Problem 5. (12%) An accident at a nuclear reactor released a small amount of radiation. Ten million (10^7) people live near this reactor. Each person has a small chance of getting cancer from this radiation, with the probability depending on their distance from the reactor at the time of the accident. Suppose that people are numbered according to their distance from the reactor and that person i has probability $p_i = 6 \times 10^{-14}\, i$ for $i = 1, 2, 3, \ldots, 10^7$ of getting cancer. (Person 1 is the farthest from the reactor and person 10^7 is the closest.) What is the probability that exactly 5 people get cancer as a result of this accident? Use an appropriate approximation. (Assume people are independent.)

This is similar to the example on page 31 of notes6.pdf.

Let X be the number who get cancer. The total number of people (10^7) is large, and the maximum probability of getting cancer is for person $i = 10^7$ whose probability of cancer is $p_i = (6 \times 10^{-14}) \times 10^7 = 6 \times 10^{-7}$ which is small. So we expect X to have approximately a Poisson distribution with mean $\lambda = \sum_{i=1}^{10^7} p_i = 6 \times 10^{-14} \sum_{i=1}^{10^7} i = 6 \times 10^{-14} \frac{10^7 \times (10^7+1)}{2} = 3.0000003 \approx 3$. So $P(X = 5) \approx \frac{3^5 e^{-3}}{5!}$. This solution should get full credit. A decimal answer is 0.100818813445, but this is not required.

Problem 6. (11%) Let us model the flow of traffic as a sequence of iid Bernoulli(p) random variables: time is divided into one second intervals, and the probability of a car passing during any given one second interval is p. Suppse a pedestrian can cross the street only if no car is to pass during the next **5** seconds. Find the probability that the pedestrian has to wait for exactly **6** seconds before starting to cross.

This is similar to exercise 3.3 and the solution is similar to the handwritten solution given on page 11 of solutions2_chapter3.pdf.

The answer is $(1-(1-p)^5)p(1-p)^5$.

Problem 7. (11%) Let X be a random variable with density (pdf)

$$f(x) = \frac{1}{2(1+|x|)^2} \quad \text{for } -\infty < x < \infty\,.$$

Does EX exist? Prove your answer.

This is similar to the Cauchy distribution example on page 5 of notes4.pdf.

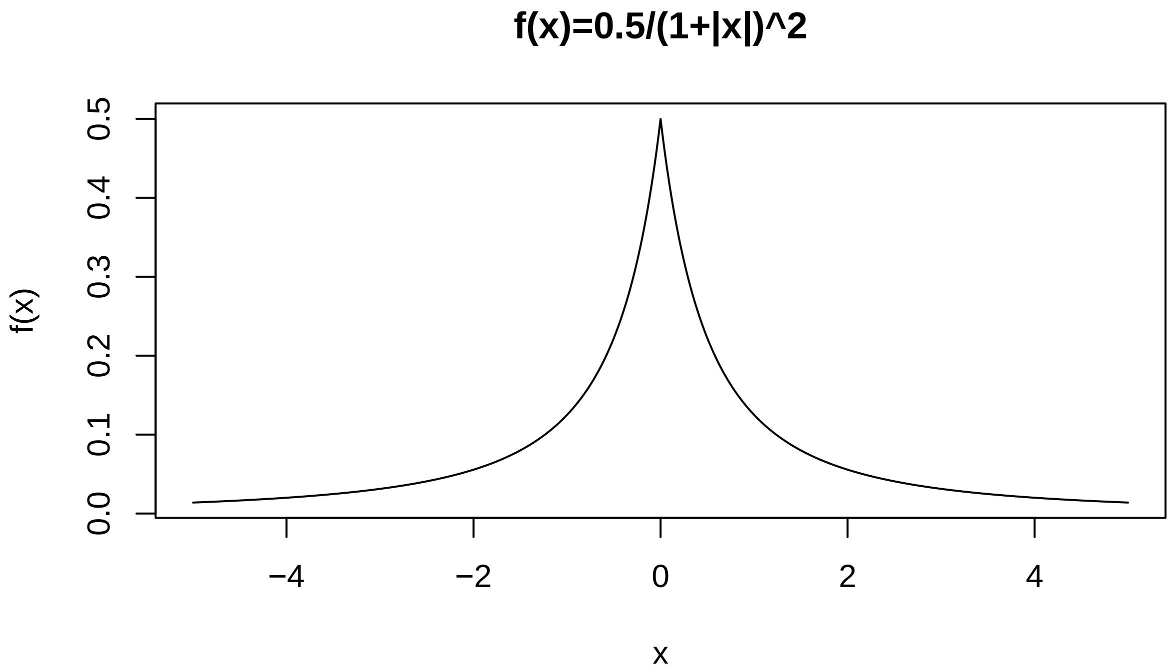
$$E|X| = \int_{-\infty}^{\infty} \frac{|x|}{2(1+|x|)^2}\,dx \quad (\dagger)$$

For large positive values of x*, the ratio* $\dfrac{|x|}{(1+|x|)^2} \approx \dfrac{1}{x}$ *and* $\displaystyle\int_c^\infty \frac{dx}{x} = \infty$ *for all* $c > 0$*. This implies that the integral* $(\dagger)$ *is* ∞ *so that* EX *does not exist.*

(A similar argument can be made using large negative values of x*, but it is only necessary to give one of these arguments to conclude that* $E|X| = \infty$ *so that* EX *does not exist.)*

The remaining problems require no work. You will receive full credit just for stating the correct answer. No partial credit is given.

Problem 8. (3%) A plot of the density $f(x)$ from the previous problem is given below.



If $\overline{X}$ is the sample mean of a random sample of size $n = 10^{100}$ from $f(x) = \dfrac{1}{2(1+|x|)^2}$, then $\overline{X}$ will be __________________.

Circle the **single** correct completion of the above sentence.

a) an extremely large positive value with probability close to one

b) an extremely large negative value with probability close to one

c) very close to zero with probability close to one

d) approximately normally distributed with mean zero

e)$\star$ none of the above

Since EX is undefined, the Law of Large Numbers and Central Limit Theorem will not be valid in this situation. So we do not expect the responses (c) or (d) to be true. It is intuitive (and not hard to prove) that, if a population is symmetric about zero, the distribution of the sample mean $\overline{X}$ will also be symmetric about zero. This rules out responses (a) and (b). This does not prove but strongly suggests that (e) is the correct answer.

Basically, the density $f(x)$ resembles the Cauchy density in some respects; for both distributions (1) the mean does not exist, (2) the density is symmetric about zero, and (3) both distributions have similar $1/x$ behavior in the tails. For the Cauchy density (e) is the correct answer, leading us to suspect (e) is also correct for $f(x)$.

This is not a rigorous proof. (You can empirically verify (e) is correct using simulations, but you will be forced to use sample sizes much smaller than 10^{100}.)

Problem 9. (3%) Suppose $X \sim$ Geometric($p = 1/n$) with $n = 10{,}000$. The cumulative distribution function (cdf) of $\dfrac{X}{n}$ is approximately equal to the cdf of what other distribution?

Give the name of the distribution and the value(s) of any parameter(s) which correctly answers the above question.

The answer is Exp(1)*, that is, Exponential with* $\beta = 1$.

See notes5.pdf, pages 19 and 20.

Problem 10. (3%) Suppose X and Y are independent, $X \sim$ Poisson($\lambda = 5$), and $Y \sim$ Poisson($\lambda = 2$). Define $Z = 3(X + Y)$. What is the moment generating function (mgf) of Z? (Simplify your answer).

The answer is $e^{7e^{3t}-1}$.

By the closure property for the Poisson distribution, $X + Y \sim$ *Poisson*(7) *with mgf* $e^{7e^{t}-1}$. *Now use the scaling property: for any rv* X*, the mgf satisfies* $M_{aX}(t) = M_X(at)$.

Problem 11. (3%) An urn contains R red balls and G green balls. The balls are randomly drawn from this urn one by one without replacement until the urn is empty. What is the probability that the third ball drawn from the urn is red?

The answer is $\dfrac{R}{R+G}$.

See page 22 of notes6.pdf.