TEST #3
STA 5326 **Name:**____________________________________
December 8, 2022

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 ONLY what you need to write with (pens, pencils, erasers, etc).
- 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.
- You must show and explain your work (including your calculations) for all the problems except those on the last page and the problems labeled **NWR** (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.
- Partial credit is available (except on the last page). 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.
- 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 or joint density is zero outside of some set, this set should be stated as part of the answer.**
- 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 **10** problems and **10** pages. There are a total of **100** points.

Problem 1. Suppose

$$Y|X \sim \text{NegativeBinomial}(r, X) \quad \text{and} \quad X \sim \text{Beta}(\alpha, \beta)\,.$$

(This problem uses the textbook's definition of the NegativeBinomial distribution, which is the main definition given in the Table of Common Distributions.)

This is exercise 4.34(b).

Students should make the more obvious simplifications, but it is OK if students leave things in terms of Gamma functions in their answers. Also, algebra does not have to be in the most simple form. (This applies to all the problems.)

(a) (10%) Find the marginal mass function (pmf) of Y.

[**Problem 1 continued**]

(b) (10%) Find EY.

[**Problem 1 continued**]

See 4.34(b)_correct_variance.pdf in the hw3_solutions folder for the answer.

(c) (10%) Find $\mathrm{Var}(Y)$.

Problem 2. Suppose (X, Y) has joint density given by

$$f(x,y) = \frac{4x}{x+y} \quad \text{for } (x,y) \in D$$

where D is the triangular region $D = \{(x,y) : x > 0, y > 0, x + y < 1\}$.

(a) (10%) For $0 < x < 1$, find $E(Y \mid X = x)$.

This resembles the lecture example in notes9 on pages 12, 13, and 17.

Useful calculus facts for this part: $\quad \dfrac{d}{dy} \log(x+y) = \dfrac{1}{x+y} \quad$ and $\quad \dfrac{d}{dy}[\, y - x\log(x+y)\,] = \dfrac{y}{x+y}$

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy = \frac{\displaystyle\int y f_{X,Y}(x,y)\, dy}{\displaystyle\int f_{X,Y}(x,y)\, dy} = \frac{\displaystyle\int_0^{1-x} y \frac{4x}{x+y}\, dy}{\displaystyle\int_0^{1-x} \frac{4x}{x+y}\, dy} = \frac{\displaystyle\int_0^{1-x} \frac{y}{x+y}\, dy}{\displaystyle\int_0^{1-x} \frac{1}{x+y}\, dy}$$

$$= \frac{y - x\log(x+y)\Big|_0^{1-x}}{\log(x+y)\Big|_0^{1-x}} = \frac{1 - x + x\log(x)}{-\log(x)}$$

[**Problem 2 continued**]

Recall that (X, Y) has joint density given by

$$f(x,y) = \frac{4x}{x+y} \quad \text{for } (x,y) \in D$$

where D is the triangular region $D = \{(x,y) : x > 0, y > 0, x + y < 1\}$.

(b) (10%) Let $U = X/(X+Y)$ and $W = X+Y$. Find the joint density of (U, W). (Make sure to specify the support.)

This is similar to exercise 4.19(b). See the homework solution.

The joint density of (U, W) *is* $f_{U,W}(u,w) = 4uw$ *for* $0 < u < 1, 0 < w < 1$.

Problem 3. (8%) An urn contains 10 balls; 2 are red and the other 8 are green. Balls are randomly selected, one at a time and withOUT replacement, until both of the red balls have been drawn. Let X be the number of draws needed to get the first red ball, and Y be the number of **additional** draws needed to get the second red ball. (For example, if the draws are $GGRGGGGGR$, then $X = 3$ and $Y = 6$.) Are X and Y independent? Answer YES or NO and prove your answer.

No, X and Y are NOT independent. One way to prove this is to observe that $P(X = 6) > 0$ and $P(Y = 6) > 0$ but $P(X = 6, Y = 6) = 0 \neq P(X = 6)P(Y = 6)$, which violates the definition of independence. (Of course 7, 8, or 9 could be used in place of 6 here, and other similar arguments can be given.)

Another way to state the argument is to note that the support of the joint pmf of (X, Y) is **not** *the product of the support of X times the support of Y, which it must be if the rv's are independent. The support of (X, Y) is*

$$\{(x, y) : x \text{ and } y \text{ are integers satisfying } x \geq 1,\ y \geq 1 \text{ and } x + y \leq 10\}.$$

and the supports of X and Y are both equal to $\{1, 2, \ldots, 9\}$.

If students give the support of the joint distribution and note that it is not a product set, that should receive full credit.

Another approach is to explicitly evaluate $f_{X,Y}(x, y)$, $f_X(x)$, and $f_Y(y)$, and show that $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ is false for at least one point (x, y). But that is more work.

Problem 4. (7%) A random point (X, Y) is distributed uniformly on the **triangle** with vertices $(2, 0)$, $(0, 2)$, and $(-2, 0)$. Find the probability that $X^2 + Y^2 < 1$.

This is similar to exercise 4.1. Note that only half the circle lies inside the triangle.

If students draw a picture and give the correct ratio of areas $= \dfrac{0.5 \cdot \pi}{0.5 \cdot 4 \cdot 2} = \dfrac{\pi}{8}$, then give them full credit.

Problem 5. Let X and Y be independent random variables with $X \sim \text{Binomial}(M, p)$ and $Y \sim \text{Binomial}(N, p)$.

This problem is similar to 4.15 but using the Binomial distribution instead of the Poisson, and relying on the closure property for the Binomial distribution instead of the closure property for the Poisson.

(a) (10%) Find the distribution of $X \,|\, X+Y$. That is, evaluate $P(X = j \,|\, X+Y = k)$ for all j and k for which this conditional probability is positive.

The conditional probability is

$$P(X = j \,|\, X+Y = k) = \frac{\binom{M}{j}\binom{N}{k-j}}{\binom{M+N}{k}}$$

which we recognize as a Hypergeometric probability, i.e., it is the probability that, in a sample of size k from an urn with M red balls and N green balls, there will be exactly j red balls.

(b) (**NWR**) (4%) What is the value of the conditional expectation $E(X \,|\, X+Y = k)$ for $0 \le k \le M+N$.

Using the formula for the mean of a Hypergeometric distribution, The expected number of red balls in a sample of size k from an urn with M red balls and N green balls is

$$E(X \,|\, X+Y = k) = k\,\frac{M}{M+N}\,.$$

Problem 6. (**NWR**) (6%) Suppose that X and Y have a joint density $f_{X,Y}(x, y)$. State a general formula for the density of $Z = Y/X$.

(If you can remember the formula, you get full credit for stating it. If not, you have room below to work it out.)

See notes10.pdf, page 16.

No work is required for the remaining problems.

Problem 7. (4%) Suppose P is the transition probability matrix of a finite-state Markov chain which is irreducible and aperiodic. If the matrix P is raised to a very large power, then ________________.

Circle the letter(s) of **all** the choices which correctly complete the sentence above.

a) The entries along the diagonal will all be equal

b) The entries in the last column will all equal one (when rounded to 7 places)

c) All the entries **except** those in the last column will be zero (when rounded to 7 places)

d)$\star$ The entries in each row sum to one

e) The entries in each column sum to one

f)$\star$ Each row of the matrix will be the same (when rounded to 7 places)

g) Each column of the matrix will be the same (when rounded to 7 places)

Deduct 2 points for each error (i.e., a wrong choiced circled or a right choice not circled) up to a maximum of four points.

Problem 8. (4%) Let P be the transition probability matrix of a Markov chain and $p_{ij}^{(n)}$ be the ij entry of the matrix P^n. Write the probability below as a product of values of the form $p_{ij}^{(n)}$.

See pages 13 and 17 of notes13_markov_chains.pdf. Page 13 states the basic result, and page 17 gives an example.

The answer is $p_{32}^{(5)} p_{25}^{(3)} p_{54}^{(1)} p_{46}^{(2)}$.

One possibility is to give one point for each correct term above, but there are other possibilities which are reasonable.

$P_3(X_5 = 2, X_8 = 5, X_9 = 4, X_{11} = 6) =$ ________________________________

Problem 9. (4%) If states i and j communicate ($i \leftrightarrow j$), then ________________.

Circle the letter of the **single** response which correctly completes the above statement.

a) $p_{ij} > 0$ and $p_{ji} > 0$

b) $p_{ij}^{(n)} > 0$ for all $n \geq 0$

c) the gcd of the periods $d(i)$ and $d(j)$ equals 1

d) i and j are aperiodic

e)$\star$ i and j have the same period ($d(i) = d(j)$)

f) $\pi_i = \pi_j$

Problem 10. (3%) Suppose you have a finite-state Markov chain which is irreducible and aperiodic. If you wanted to report a number which indicated how rapidly this Markov chain approached its stationary distribution, what number could you report? Describe this number as precisely as you can.

The answer: The second largest eigenvalue of the transition probability matrix P (sometimes denoted λ_2).