Final Exam: Review Problems

The majority of these problems focus on material from the last third of the course (though all of probability builds on what we learned starting from the first day). Many (not all) of these problems are more challenging than what I might expect you to do during the final. But mastering more difficult problems will prepare you well for the final.

1. Consider the Binomial distribution. Name at least 4 other distributions that are related to the Binomial and explain how they are related to the Binomial.

2. For events A and B such that $P(A) > 0$ and $P(B) > 0$ prove:

   (a) If A and B are mutually exclusive, then they cannot be independent.

   (b) If A and B are independent, then they cannot be mutually exclusive.

3. If n balls are placed randomly into different n bins where each bin is equally likely, find the probability that exactly one bin remains empty.

4. Suppose a test consists of 20 multiple-choice questions, each with four possible answers. I have not come to class the entire semester, so I must guess on each question, independent of the other questions. Find the probability (no need to necessarily evaluate explicitly) that I get at least 10 questions correct.

5. It's even closer to being Christmas! A group of n people play "Secret Santa": each person writes their name on a slip of paper in a hate, picks a name randomly from the hat (without replacement), and then buys a gift for that person. Unfortunately, they overlook the possibility of drawing one's own name, so some people may be buying a gift for themselves. Assume $n \geq 2$.

   (a) Let X be the number of people who pick their own names. Find the expected value of X. Simplify.

   (b) With the same X as above, obtain an approximate distribution of X if n is large.

   (c) Let Y be the number of pairs of people who pick each other's name. That is, pairs of people A and B such that A picks B and B picks A. This assumes that $A \neq B$ and order doesn't matter. Simplify.

6. A fair 20-sided die is rolled repeatedly until a number greater than or equal to 11 is shown. Let W be the event that the first such number appears on an even numbered roll (i.e if the first three rolls are 3, 2, 9, and the fourth roll is 12, then W occurs, since 4 is even).

   (a) Give at least two different ways to calculate $P(W)$, using material from our course.

   (b) Explain why it makes sense that $P(W) < 1/2$.

7. An urn contains red, green, and blue balls. Balls are chosen randomly with replacement (each time, the color is noted and then the ball is put back.) Let r, g, b represent the probabilities of drawing a red, green, or blue ball, respectively. $(r + g + b = 1)$.

(a) Find the expected number of balls chosen before obtaining the first red ball, not including the red ball itself.

(b) Find the expected number of different *colors* of balls obtained before getting the first red ball.

(c) Suppose we draw n balls (still randomly with replacement). What is the probability that at least 2 of the balls are red, given at least one is red? (Avoid $\sum$ notation).

8. Suppose we have a continuous random variable Y whose CDF is

$$F_Y(y) = \begin{cases} 1 - \frac{1}{y^2} & y \geq 1 \\ 0 & y < 1 \end{cases}$$

   (a) Verify that F_Y is a valid CDF.

   (b) Obtain the pdf of Y.

   (c) Let $Z = 10(Y - 1)$. Find the PDF of Z in two different ways.

9. Suppose $X \sim \text{Beta}(a, b)$ and $Y \sim \text{Beta}(a + b, c)$ are independent random variables. Define $U = XY$ and $V = X$. Find the joint PDF of U and V (don't forget the support). Are U and V independent?

10. Suppose a lake contains just two species of fish, where 20% of the fish are species 1, and 80% of the fish are species 2. Let N be the number of fish in the lake, and assume $N \sim \text{Poisson}(\lambda)$. Let X be the number of fish of species 1 in the lake, and let Y be the number of species 2.

   (a) Find the joint PMF of X and Y.

   (b) Find $\mathbb{E}[N|X]$ and $\mathbb{E}[N^2|X]$.

11. Let X_1, X_2, X_3 be independent random variable, where $X_i \sim \text{Exp}(\lambda_i)$. A useful fact (which you may use) is that if $X \sim \text{Exp}(\lambda_X)$ is independent of $Y \sim \text{Exp}(\lambda_Y)$, then $P(X < Y) = \frac{\lambda_X}{\lambda_X + \lambda_Y}$.

   (a) Find $\mathbb{E}[X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3]$ in terms of $\lambda_1, \lambda_2.\lambda_3$.

   (b) Find $P(X_1 = \min\{X_1, X_2, X_3\})$ (i.e. the probability that X_1 is the smallest of the three). *Hint: re-state this probability as one in terms of* X_1 *and* $\min\{X_2, X_3\}$.

   (c) For the case of $\lambda_1 = \lambda_2 = \lambda_3 = 1$, find the PDF of $M = \max\{X_1, X_2, X_3\}$.

12. A researcher is studying the length of hospital stays for patients with a certain disease. Let $X \sim \text{Pois}(\lambda)$ denote the number of days a randomly selected person with the disease stays at a hospital. Data however, is only collected for patients with the disease who were admitted to the hospital (i.e. no data is available for patients with the disease who did not stay at a hospital). Therefore, averaging the length of hospital stays among this population does not estimate $E[X]$. Instead, we can assess the conditional distribution of X, given that the person stayed at least 1 day in the hospital.

(a) Find a formula for the conditional PMF of X given $X \geq 1$.

(b) Use the previous result to calculate $\mathbb{E}[X|X \geq 1]$.

(c) Use part (a) to calculate $\text{Var}(X|X \geq 1)$.

13. The number of accidents on a highway is modeled as a Poisson Process with rate parameter λ. Suppose exactly one accident has occurred by time t (i.e. in the time interval $(0, t]$). For s such that $0 < s < t$, find the probability that the accident occurred by time s.

14. Let $X_1, \ldots, X_n$ ($n \geq 2$) be iid random variables with mean μ and variance σ^2. A bootstrap sample is a random sample with replacement of size n from these $X_1, \ldots, X_n$, where all variables are equally likely to be sampled. The bootstrap sample takes the form $X_1^*, \ldots, X_n^*$, where X_j^* represents the j-th randomly sampled variable. Let $\bar{X}^*$ denote the sample mean of the bootstrap sample:

$$\bar{X}^* = \frac{1}{n}\left(X_1^* + \cdots + \ldots, X_n^*\right)$$

(a) Obtain $\mathbb{E}[X_j^*]$ and $\text{Var}(X_j^*)$ for each j. *Hint*: think about what the distribution of each X_j^* might be.

(b) Calculate $\mathbb{E}[\bar{X}^*|X_1, \ldots, X_n]$ and $\text{Var}(\bar{X}^*|X_1, \ldots, X_n)$. *Hint*: conditional on $X_1, \ldots, X_n$, the X_j^* are independent with a PMF that puts probability $\frac{1}{n}$ at each of the points $X_1, \ldots, X_n$.

(c) Calculate $\mathbb{E}[\bar{X}^*]$ and $\text{Var}(\bar{X}^*)$.

(d) Explain intuitively why $\text{Var}(\bar{X}^*) > \text{Var}(\bar{X})$.

15. Let X and Y be positive random variables, not necessarily independent unless otherwise stated. Assume that the various expected values below exist. For each of the following, fill the blank in with the most appropriate choice of $\leq, \geq, =$ or ? (where ? means that no relation holds in general).

(a) $\mathbb{E}[\mathbb{E}[X|Y] + \mathbb{E}[Y|X]]$ ___ $\mathbb{E}[X]$

(b) $P(X + Y = 2)$ ___ $P(\{X \geq 1\} \cup \{Y \geq 1\})$

(c) $P(|X + Y| > 2)$ ___ $\frac{1}{16}\mathbb{E}[(X + Y)^4]$

(d) Assuming X and Y are iid, $P(|X - Y| > 2)$ ___ $\frac{\text{Var}(X)}{2}$

16. Let $X_1, X_2, \ldots$ be iid variables, each with CDF F_X. For every x, define a function $R_n(x)$ to be the number of $X_1, \ldots, X_n$ that are less or equal to x.

(a) Find the mean and variance of $R_n(x)$ (in terms of n and $F_X(x)$).

(b) Show that with probability 1, $\lim_{n\to\infty} \frac{R_n(x)}{n} = F_X(x)$.

17. Let $X_1, X_2, \ldots$ be iid positive random variables with mean μ. Define $W_n = \frac{X_1}{X_1 + \ldots + X_n}$.

(a) Find $\mathbb{E}[W_n]$. *Hint: consider* $\frac{X_1}{X_1+...+X_n} + \frac{X_2}{X_1+...+X_n} + \ldots + \frac{X_n}{X_1+...+X_n}$.

(b) What random variable does the nW_n converge to as $n \to \infty$?

(c) Suppose in this part that the $X_i \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. Name the exact distribution of W_n without using calculus!