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. 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.

2. 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.

3. Let $X_1, X_2, \ldots$ be iid positive random variables with finite mean μ and finite variance. Define $W_n = \frac{X_1}{X_1+\ldots+X_n}$.
   (a) Find $\mathbb{E}[W_n]$. *Hint: consider* $\frac{X_1}{X_1+\ldots+X_n} + \frac{X_2}{X_1+\ldots+X_n} + \ldots + \frac{X_n}{X_1+\ldots+X_n}$.
   (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!

4. A group of n people play "Secret Santa": each person writes their name on a slip of paper. The names are placed in a hat, and each person picks a name randomly from the hat (without replacement) 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. Find $\mathbb{E}[Y]$. Simplify.

5. 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 $P(W) < 1/2$.

6. Suppose $X \sim \text{Beta}(a, b)$. Find the distribution of $1 - X$.

7. Seven balls are randomly and independently thrown into seven boxes, with each box equally likely. Each ball lands in exactly one box, but some boxes could be empty. Define X as the number of boxes that contain exactly 3 balls. Derive the PMF of X.

8. On any given day, a certain stock has low volatility with probability p and high volatility with probability $1-p$. When the stock is low volatility, the percent change in the stock price is $X_1 \sim N(0, \sigma_1^2)$. When the stock is high volatility, the percent change in the stock price is $X_2 \sim N(0, \sigma_2^2)$, with $\sigma_1 < \sigma_2$. Define X as the percent change of the stock on a certain day, where we represent X as:

$$X = \mathbf{1}X_1 + (1 - \mathbf{1})X_2$$

where $\mathbf{1}$ is the indicator that the day was low volatility. We will assume that $\mathbf{1}, X_1, X_2$ are all mutually independent.

   (a) Find $\text{Var}(X)$ using Eve's law.

   (b) Find $\text{Var}(X)$ using the property that $\text{Var}(X) = \text{Cov}(X, X)$.

9. A closet has n unique pairs of shoes. If $2k$ shoes are chosen at random (all shoes equally likely) with $2k < n$, what is the probability that there will be no matching pair in the sample?

10. 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.

11. 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?

12. A physicist makes 25 independent measurements of the specific gravity of a certain body. She knows that her equipment isn't perfect, so the measurements have a standard deviation of $\sigma < \infty$ units.

   (a) Find a lower bound for the probability that the average of her 25 measurements will differ from the actual gravity of the body by less than $\sigma/4$ units.

   (b) Using a relevant theorem, find an approximate value of the probability in part (a).

13. Suppose the town of Middlebury has a population of 9000 people, and that 225 of of these people own an electric vehicle (EV). If I randomly contact 1000 in Middlebury (without replacement), what is the probability (exact or approximate) that 20 of them have an EV?

14. We will work with a *hybrid* joint distribution: a joint where some elements are continuous and some elements are discrete. Remember the CDF uniquely determines a distribution, so if we wanted to find the joint distribution of Z and W, we could do so by finding $F_{Z,W}(z,w) = P(Z \leq z, W \leq w)$. As we've seen, when Z and W are discrete it's usually easier to work with their joint PMF $P(Z = z, W = w)$. Now suppose Z is continuous and W is discrete. To specify their joint distribution, we could specify joint probabilities of the form $P(Z \leq z, W = w)$ for all z, w in their joint support.

    Define $X \sim \text{Exp}(\lambda_1)$ independent of $Y \sim \text{Exp}(\lambda_2)$. Suppose it is impossible to obtain direct observations of X and Y, and we only observe the following versions which contain less information:

    $$Z = \min\{X, Y\} \qquad \text{and} \qquad W = \begin{cases} 1 & \text{if } Z = X \\ 0 & \text{if } Z = Y \end{cases}$$

    (a) We will find the joint distribution of (Z, W) by obtaining formulas for the probability $P(Z \leq z, W = w)$. In particular, because the support of $W = \{0, 1\}$, find formulas for $P(Z \leq z, W = 0)$ and $P(Z \leq z, W = 1)$. This tells us all the ways Z and W and behave together.

    (b) Determine if Z and W are independent.

15. 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 $\mathbb{E}[N^2|X]$.

16. The same group of n friends go out for dinner every week. Every dinner, one person is chosen uniformly at random to pay the entire bill, independent of previous dinners.

    (a) Find the probability that in k dinners, no one will have to pay the bill more than once. Make sure to specify this for all k.

    (b) Find the expected number of dinners it takes in order for everyone to have paid at least once.

    (c) Becky and Christian are two of the friends. For a fixed number of k dinners, find the covariance between how many times Becky pays and how many times Christian pays.

17. Let X_1, X_2, X_3 be independent random variable, where $X_i \sim \text{Exp}(\lambda_i)$. Remember the useful theorem you've derived multiple times: 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 $M = \max\{X_1, X_2, X_3\}$.

18. Let the joint PDF of (X, Y) be $f_{X,Y}(x, y) = 1$ for $0 < x < 1, x < y < x + 1$, and 0 otherwise.

    (a) Draw a picture indicating the support of the joint PDF.

    (b) Find the marginal distributions of X and Y. Use your picture to help you!

    (c) Find $\text{Cov}(X, Y)$.

19. 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:

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

    (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 $\text{Var}(\bar{X}^*) > \text{Var}(\bar{X})$.

20. 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, $P(|X - Y| > 2)$ ___ $\frac{\text{Var}(X)}{2}$

21. Let $X_1, X_2, \dots$ be iid variables, each with CDF F_X. For every x, define a function $R_n(x)$ to be the number of $X_1, \dots, X_n$ 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) What does the WLLN tell us about $\frac{1}{n}R_n(x)$?

22. A certain hereditary disease can be passed from a mother to her children. Given that the mother has the disease, her children will independently have the disease with probability 1/2. If the mother doesn't have the disease, then the children will not either. A certain mother has two children, and it is known that she has the disease with probability 1/3.
    (a) Find the probability that neither child has the disease.
    (b) Is whether the older child has the independent of whether the young child has the disease? Are they conditionally independent?
    (c) Suppose we know that both children do not have the disease. What is the probability that the mother has the disease?

23. Let (X_1, X_2, X_3) have joint PDF
$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = 6e^{-x_1-x_2-x_3} \qquad 0 < x_1 < x_2 < x_3 < \infty$$
    (a) Are the X_i independent? How can you tell why/why not?
    (b) Consider the transformation $Y_1 = X_1$, $Y_2 = X_2 - X_1$, and $Y_3 = X_3 - X_2$. Find the joint distribution of (Y_1, Y_2, Y_3).
    (c) Based on (b), are the Y_i independent? What are their marginal distributions? *Note: this result should feel "familiar".*

24. We have the following joint PDF for some $c > 0$:
$$f_{X,Y}(x, y) = \begin{cases} c(x + 2y) & 0 < y < 1, 0 < x < 2 \\ 0 & o.w. \end{cases}$$
    (a) Find the value of c.
    (b) Find the marginal distribution of X.
    (c) Tricky: find the joint CDF of X and Y, $F_{X,Y}$. i.e., find $P(X \le x, Y \le y)$ for all $(x, y) \in \mathbb{R}^2$.