1. I have two boxes. Each box has four balls in it numbered 1, 2, 3, and 4, respectively. I conduct the following experiment: I randomly select two balls, one from the first box and one from the second box, and record the number on each ball. Define the random variables

$$\begin{aligned} X &= \text{minimum ball value,} \\ Y &= \text{maximum ball value, and} \\ R &= Y - X. \end{aligned}$$

For example, if I selected a 2 and a 3, then $y = 3$, $x = 2$, and $r = y - x = 3 - 2 = 1$. If I selected a 4 and a 4, then $y = 4$, $x = 4$, and $r = y - x = 4 - 4 = 0$.
(a) Find the probability mass function (pmf) of R, and sketch a graph of it. Label everything on your graph.
(b) Compute $E(R)$ and $V(R)$.

2. I am preparing an itinerary to visit 5 cities: Birmingham, Raleigh, Iowa City, Seattle, and Dallas. The order in which I visit the cities will determine the cost of the entire trip.
(a) How many different itineraries are possible?
(b) Let Y denote the number of cities visited before Iowa City. Find the probability mass function (pmf) of Y. Assume that each itinerary is equally likely.
(c) Use your pmf in part (b) to compute $E(Y)$ and $V(Y)$.

3. Let Y be a discrete random variable with probability mass function (pmf)

$$p_Y(y) = \begin{cases} y/8, & y = 1, 2, 5 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Provide a graph for the pmf of Y.
(b) Find $E(Y)$ and $V(Y)$.
(c) Find $E(2Y + 3)$ and $V(2Y + 3)$.
(d) Find the moment-generating function (mgf) of Y.

4. Two fair dice are rolled. Let Y equal the product of the two dice.
(a) Find the pmf of Y.
(b) Find the mean and variance of Y.

5. Suppose that Y is a random variable with $E(Y) = 3$ and $V(Y) = 1$.
(a) Compute $E[(Y - 1)^2]$.
(b) Compute $V(2 - 3Y)$.

6. Suppose that Y is a discrete random variable with pmf given by

y	0	1	4	9
$p_Y(y)$	0.5	0.2	0.2	0.1

(a) Compute $E(Y)$ and $V(Y)$ without using the mgf of Y.
(b) Find the mgf of Y. Compute $E(Y)$ and $V(Y)$ using the mgf of Y.

7. Let Y be a discrete random variable with probability mass function (pmf)

$$p_Y(y) = \begin{cases} ce^{-y}, & y = 0, 1, 2, ... \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show that $c = (e - 1)/e$.
(b) Compute $P(Y \geq 2)$ and $P(Y \geq 2 | Y \leq 4)$.
(c) Show that the moment generating function of Y is given by

$$m_Y(t) = c(1 - e^{t-1})^{-1},$$

for values of $t < 1$. Make sure you argue why $t < 1$.

8. Suppose that the random variable Y has pmf

$$p_Y(y) = \begin{cases} \theta(1-\theta)^y, & y = 0, 1, 2, ..., \\ 0, & \text{otherwise,} \end{cases}$$

where $0 < \theta < 1$.
(a) Show that $p_Y(y)$ is a valid pmf.
(b) Derive the mgf of Y, and use it to find expressions for $E(Y)$ and $V(Y)$.

9. Suppose that the random variable X has the pmf

$$p_X(x) = \begin{cases} \frac{1}{12}(c - 2x), & x = 0, 1, 2 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the value of c which makes this a valid pmf.
(b) Compute $E(X)$.

10. Suppose $a, b \in \mathbb{R}$. For any random variable Y, show that

$$V(a + bY) = b^2 V(Y).$$

11. A discrete random variable Y has pmf

$$p_Y(y) = \begin{cases} c(1/2)^y, & y = 1, 2, 3, ... \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the value of c.
(b) Find $E(Y)$ and $V(Y)$.

12. Suppose that Y is a discrete random that assigns positive probabilities to only the positive integers; i.e., the support of Y is $R = \{y : y = 1, 2, 3, ..., \}$. Prove that

$$E(Y) = \sum_{y=1}^{\infty} P(Y \geq y).$$

13. A discrete random variable Y has probability mass function

$$p_Y(y) = \begin{cases} \dfrac{c}{y2^y}, & y = 1, 2, 3, ... \\ 0, & \text{otherwise,} \end{cases}$$

where the constant $c = (\ln 2)^{-1}$.
(b) Show that the moment generating function of Y is

$$m_Y(t) = -c \ln\left(1 - \frac{e^t}{2}\right), \quad \text{for } t < \ln 2.$$

Hint: Write out $h(a) = -\ln(1-a)$ in its Maclaurin series expansion. Where does the condition $t < \ln 2$ come from?
(c) Find $E(Y)$ and $V(Y)$.

14. A truncated discrete distribution arises when a particular value cannot be observed and is eliminated from the support. In particular, if Y has the pmf $p_Y(y)$ with support $0, 1, 2, ...,$ and the "0" value cannot be observed, the 0-truncated random variable Y_T has pmf $p_{Y_T}(y)$, where

$$p_{Y_T}(y) = \begin{cases} \dfrac{p_Y(y)}{1 - p_Y(0)}, & y = 1, 2, 3, ..., \\ 0, & \text{otherwise.} \end{cases}$$

For $0 < \theta < 1$, suppose that the random variable Y has pmf

$$p_Y(y) = \begin{cases} \theta(1-\theta)^y, & y = 0, 1, 2, ..., \\ 0, & \text{otherwise,} \end{cases}$$

(a) Show that the pmf of the 0−truncated version of Y, Y_T, is given by

$$p_{Y_T}(y) = \begin{cases} \theta(1-\theta)^{y-1}, & y = 1, 2, 3, ..., \\ 0, & \text{otherwise.} \end{cases}$$

(b) Find the moment generating function of Y_T in part (a) and use it to compute $E(Y_T)$.

The following problems have been added to this handout to prepare for the second midterm.

15. Let Y be a random variable with mean $\mu = E(Y)$ and variance $\sigma^2 = V(Y)$. The **skewness** of Y, denoted by ξ, is given by

$$\xi = \frac{E[(Y-\mu)^3]}{\sigma^3}.$$

The skewness ξ quantifies how much a probability distribution departs from symmetry. Thus, if $\xi = 0$, then the distribution associated with Y is **symmetric** in appearance. Values of ξ that are larger (smaller) than zero correspond to distributions that are skewed right (left).
(a) Show that the skewness of $Y \sim b(n, p)$ is given by

$$\xi = \frac{1-2p}{\sqrt{np(1-p)}}.$$

Hint: First show $E[(Y-\mu)^3] = E(Y^3) - 3\mu E(Y^2) + 2\mu^3$. For $Y \sim b(n,p)$, we know $E(Y) = \mu = np$ and $V(Y) = \sigma^2 = np(1-p)$. Use the mgf to find $E(Y^3)$. You can find $E(Y^2)$ using either the mgf or by noting that $E(Y^2) = \sigma^2 + \mu^2$.
(b) For what value of p is $\xi = 0$? Interpret.
(c) For the $b(n,p)$ model, what does ξ converge to as $n \to \infty$? Interpret.
(d) Derive ξ for $Y \sim \text{Poisson}(\lambda)$.

16. The zinc-phosphate coating on the threads of steel tubes used in oil and gas wells is critical to their performance. However, 12 percent of all tubes receive an improper amount of coating (either much too low or much too high). That is, about 12 percent of the tubes are defective. Assume that the tubes are independent.
(a) If we take a sample of 10 tubes, what is the probability at least 2 are defective?
(b) If we continually observe tubes until we find the first defective tube, what is the probability we will observe no more than 3 tubes?
(c) If we continually observe tubes until we find the second defective tube, what is the probability we will observe more than 4 tubes?

17. Screening for infectious diseases in a blood-bank setting is a major part of ensuring blood safety. At a local clinic, subjects' blood donations are tested for infection. Suppose that 10 percent of all blood donations are infected. Assume that all subjects are independent.
(a) Let W denote the number of subjects tested to find the second infected blood donation. Find $P(W > 3)$.
(b) Suppose that, during a given day, there are 30 donations. Find the probability that no more than two of these donations are infected.

18. Past studies have shown that 1 out of every 10 cars on the road has a speedometer that is miscalibrated. For this problem, assume that different cars are independent and that each has the same 1/10 probability of being miscalibrated.
(a) In a sample of 10 cars, let X denote the number of cars which are miscalibrated. Write the pmf for X and compute $P(X \geq 3)$. Interpret this probability in words.
(b) Suppose that we continually observe cars until we find the first car with a miscalibrated speedometer. Let Y denote the number of cars that we will observe. Write the pmf for Y and compute $P(Y \leq 6)$. Interpret this probability in words.
(c) Suppose that we continually observe cars until we find the second car with a miscalibrated speedometer. Let Z denote the number of cars that we will observe. Write the pmf for Z and compute $P(Z = 4)$. Interpret this probability in words.

19. Sowbugs are primarily nocturnal, thrive in a moist environment, and they eat decaying leaf litter and vegetable matter. Suppose that Y, the number of sowbugs on a square-foot plot, follows a Poisson distribution with $\lambda = 15$.
(a) Find the probability that a square-foot plot contains exactly 10 sowbugs.
(b) In terms of the Poisson pmf, write an expression for $P(Y \geq 100)$.
(c) Suppose the cost (measured in dollars) incurred from sowbug damage, per square-acre plot, is $C = 0.03Y^2 + 0.05Y$. What is the expected cost of damage?

20. Suppose that Y has a Poisson distribution with mean $\lambda > 0$.
(a) Derive the mgf of Y. Make sure you explain all the steps.
(b) For a support point y, use the Poisson pmf to show that

$$\frac{P(Y = y + 1)}{P(Y = y)} = \frac{\lambda}{y + 1}.$$

(c) What value of y maximizes $P(Y = y)$? This value y is called the **mode** of Y. *Hint*: Consider the result in (b).
(d) If $3P(Y = 1) = P(Y = 2)$, find $P(Y = 4)$.

21. Suppose that Y has a geometric distribution with success probability p. Let $F_Y(y) = P(Y \leq y)$ denote the cumulative distribution function of Y.
(a) Show that

$$F_Y(a) = 1 - (1 - p)^a,$$

for any positive integer a.
(b) Using the result in (a), argue that the geometric distribution possesses the **memoryless property**; i.e., show that, for $b > 0$,

$$P(Y > a + b|Y > a) = P(Y > b).$$

22. When circuit boards used in the manufacture of blue ray players are tested, the percentage of defectives is 5 percent. A sample of 20 circuit boards is available for analysis. Let Y denote the number of defectives (out of 20) and assume a binomial model.
(a) Briefly, discuss the Bernoulli assumptions in the context of this example. That is, tell me what would have to be true for the binomial model to hold.
(b) If the sample of 20 circuit boards contains 2 or more defectives, the entire sample is rejected. Find the probability that the sample is rejected.
(c) In lots of size 20, the cost (in dollars) associated with reworking defective circuit boards is $C = 200Y^2 - 10Y + 100$. Find the expected value of C.

23. Suppose that Y, the number of infected trees distributed on a square-acre plot, follows (approximately) a Poisson distribution with mean $\lambda = 3$.
(a) What is the probability a single square-acre plot will contain 2 or fewer infected trees?
(b) Suppose that we continually observe square-acre plots (in a very large forest) until we observe the fourth plot with 2 or fewer infected trees. Let X denote the number of plots we will need to observe. Write down the pdf for X and compute $P(X = 6)$. You may assume that the square-acre plots are independent.
(c) The time, T, (measured in hours) needed to locate and treat all infected trees is thought to be a linear function of Y. One researcher assumes that $T = 2Y + 3$. Find the mean and variance of T.

24. An animal biologist observes the number of sparrow nests on a 5000 m^2 plot. Let Y denote the number of sparrow of nests per plot, and assume that Y has a Poisson distribution with mean 2.

(a) What is the probability that a randomly selected plot contains no sparrow nests?
(b) Suppose that the biologist observes 10 plots, each of the same size, and let W_1 denote the number of plots that have no sparrow nests on them. Give the distribution of W_1 and find $P(W_1 \leq 1)$. You may assume that the 10 plots are independent.
(c) Suppose that the biologist continues to observe plots until she finds the first plot that contains no sparrow nests, and let W_2 denote the number of plots she observes. Give the distribution of W_2 and find $P(W_2 = 3)$. You may assume that the plots are independent.
(d) Suppose that the biologist continues to observe plots until she finds the second plot that contains no sparrow nests, and let W_3 denote the number of plots she observes. What is the (named) distribution of W_3? (Just give me the distribution name). You may assume that the plots are independent.

25. A phase II clinical trial, currently being conducted in Bethesda, MD, is examining how well ixabepilone works in treating patients with renal cell carcinoma (kidney cancer). Patients are currently being enrolled in the trial. For this problem, we will assume that 35 percent of all patients will respond positively to ixabepilone.
(a) Let Y denote the number of patients in the trial who respond positively to ixabepilone (e.g., the drug reduces tumor size). Clearly explain, in the context of the problem, what assumptions must be true for Y to have a binomial distribution.
(b) Suppose now that patients will be enrolled continually in the trial until the 30th patient responds positively to ixabepilone. What is the probability that no more than 100 patients will be needed? Write an expression that gives this probability. Do not evaluate this expression numerically.
(c) At the end of the recruitment period, suppose that there are 90 patients. Ten of these patients are HIV positive. If we select a random sample of 5 patients from the 90 (without replacement), what is the probability that exactly one is HIV positive? Here, I am looking for a numerical answer.

26. Let Y denote the number of calls received per day by the USC Campus Police. Suppose that Y has a Poisson distribution with mean 6.
(a) What is the probability that, on a given day, there are at most 2 calls?
(b) Calculate the median of Y.
(c) Suppose that the daily cost (in dollars) to respond to Y calls is given by

$$C = 150 + 100Y$$

Find the expected value and variance of C.