GROUND RULES:

- This exam contains 5 questions. Each question is worth 20 points. This exam is worth 100 points.
- This is a open-book and open-notes exam. You cannot use anything else other than the book or the notes; no internet; no discussion with anyone else except me, etc.
- Each question contains subparts. On each part, there is opportunity for partial credit, so show all of your work and explain all of your reasoning. **Translation:** No work/no explanation means no credit.
- If you use R to answer any part or to check your work, please include all code and output as attachments. Do not just write out the code you used.
- **My preference:** (1) Print out exam if you are attending the class online. (2) Print your name legibly at the top of this page. (3) Write your solution to each problem on its own page. You can use the back of the page if needed; you can append extra pages if needed. (4) Collect your pages when done and staple in upper left corner. Keep all pages/solutions in order.
- I prefer you write out your solutions "by hand." In the past, I have noticed students who type their solutions (in LaTeX or Word) often do not provide enough detail and/or do not explain their reasoning very well.
- Your solutions should be turned in to me no later than Wednesday, October 14 at 5:00pm. I would prefer your solutions be delivered to me in person. If this is not possible, let me know via email.

HONOR PLEDGE FOR THIS EXAM:

After you have finished the exam, please read the following statement and sign your name below it.

> *I promise that I did not discuss any aspect of this exam with anyone other than the instructor, that I neither gave nor received any unauthorized assistance on this exam, and that the work presented herein is entirely my own.*

1. Two litters of Labrador Retriever puppies have been born. Here are the litters and the colors of the puppies in each litter:

Litter 1: 1 chocolate, 1 black, 2 yellows
Litter 2: 2 blacks, 2 yellows

I will randomly select 2 puppies from each litter. The order of selection is not important. Let Y denote the total number of yellow puppies I select from the two litters combined.

(a) Determine the probability mass function of Y and graph it (like we do in the notes). Neatness counts.
(b) Find the moment generating function of Y.

2. The discrete random variable Y has a probability mass function (pmf) which is summarized in the table below:

y	-2	0	2
$p_Y(y)$	p	$1-2p$	p

(a) Describe which values of p are allowed in this pmf.
(b) Show $E(Y) = 0$ and $V(Y) = 8p$.

The *skewness* of a random variable Y is defined as

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

where $\mu = E(Y)$ is the mean of Y and $\sigma^2 = V(Y)$ is the variance of Y. The *kurtosis* of Y is defined as

$$\kappa = \frac{E[(Y-\mu)^4]}{\sigma^4}.$$

The skewness ξ measures how a distribution "departs" from symmetry. If $\xi = 0$, then the distribution is symmetric about its mean. If $\xi > 0$ ($\xi < 0$), then the distribution is skewed to the right side (left side). The kurtosis κ measures how "peaked" the distribution is in the center. The larger the value of κ, the more peaked the distribution is.

(c) Calculate ξ and κ for the distribution described in the table above. What values of p will make the kurtosis large?
(d) Instead of observing Y, suppose I observe $X = Y + 3$ instead. Describe how the values of μ, σ^2, ξ, and κ associated with X would change (if at all).

3. The South Carolina Department of Motor Vehicles reports that among all customers who arrive to renew their driver's license, 30 percent have one that is already expired.

(a) Thinking of each customer as a "trial," state carefully the assumptions needed for the Bernoulli trial assumptions to hold. Don't give me a generic description. Talk in terms of customers and driver's licenses.
(b) Find the probability that among the first 8 customers seeking renewal, at most 2 will have expired licences.
(c) Let W denote the number of customers an employee will help until he finds the first customer with an expired license. Find $P(W \leq 5|W > 2)$.
(d) Another employee will take a break after she helps 3 customers with an expired license. Give a formula for the probability mass function (pmf) of the total number of customers she will help. Make sure to state the support in your pmf and define all of your notation.

4. Let Y denote the number of positive coronavirus tests each day among faculty, students, and staff at the University of South Carolina. Suppose Y follows a Poisson distribution with $\lambda = 10$.

(a) A warning flag is raised if, on any day, Y exceeds $\mu + 2\sigma$, where $\mu = E(Y)$ and σ is the standard deviation of Y. Calculate $P(Y > \mu + 2\sigma)$.
(b) The daily cost of quarantining positive cases (in dollars) is a random variable Q given by

$$Q = 3Y^4 + (2Y - 1)^2 + 10000.$$

Calculate $E(Q)$.
(c) The number of positive tests each day from September 26 through October 5 (a total of 10 days) at the University of South Carolina are given below:

```
5   1   3   12   4   11   6   4   1   69
```

Based on these observations, what do you think about the assumption that the number of positive tests each day follows a Poisson distribution with $\lambda = 10$?

5. The Poisson distribution is commonly used to model "counts;" e.g., the number of coronavirus cases per day as in Problem 4. Another discrete distribution that is useful in modeling counts is the *logarithmic series distribution.* A random variable Y has this distribution if its probability mass function is

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

The value p satisfies $0 < p < 1$. Note that observing $y = 0$ is not allowed under this model; i.e., the value $y = 0$ is not in the support of Y.

Remark: The logarithmic series distribution probably would not be a good distribution to model the number of positive coronavirus tests each day among faculty, students, and staff at the University of South Carolina. First of all, this model does not allow for a day to have no positive cases. In addition, the *mode* of Y, that is, the value of y that maximizes $p_Y(y)$, is equal to 1. This model would imply having exactly one positive case on any given day is the most likely outcome (which is probably not true right now).

(a) Provide a mathematical argument verifying the mode of Y is 1.
(b) The moment generating function (mgf) of Y is

$$m_Y(t) = \frac{\ln(1-qe^t)}{\ln p}, \quad \text{for } t < -\ln q,$$

where $q = 1 - p$. Derive formulas for $E(Y)$ and $E(Y^2)$ by using this mgf.
(c) Derive $E(Y)$ *without* using the mgf.