GROUND RULES:

- This exam contains 7 questions; each question is worth 10 points. The maximum number of points on this exam is 70.
- This is a closed-book and closed-notes exam. I will provide the summary formulae for the discrete and continuous distributions.
- You may use a calculator if you wish, but show all of your work (on paper) and explain all of your reasoning to receive full credit. Correct answers with no work provided will receive only a small fraction of the total points possible.
- Any discussion or otherwise inappropriate communication between examinees, as well as the appearance of any unnecessary material, will be dealt with severely.
- Print your name at the top of this page in the upper right hand corner.
- You have 3 hours to complete this exam. GOOD LUCK!!

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. For $-1 < t < 1$, the random variable Y has the moment-generating function

$$m_Y(t) = \frac{1}{1-t^2}.$$

(a) Find $E(Y) = \mu$ and $V(Y) = \sigma^2$.
(b) Recall that the skewness of Y is given by

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

Show that $\xi = 0$. What does this imply about the distribution of Y?

2. The manager of a local pizzeria obtains the following information from its records:

- All customers order at least one pizza.
- 40 percent of the customers order more than one pizza.
- 30 percent of the customers order a Pepsi product.
- Of those customers who order more than one pizza, 10 percent order a Pepsi product.

(a) Calculate the probability that a randomly selected customer orders exactly one pizza **and** does not order a Pepsi product.
(b) Calculate the probability that a randomly selected customer orders more than one pizza, **given** that the customer does not order a Pepsi product.

3. A medical investigation is undertaken to learn about the relationship between brain lesion frequency for patients with advanced multiple sclerosis. A Poisson model is assumed for Y, the number of brain lesions per subject. Specifically, it is assumed that $Y \sim \text{Poisson}(\lambda)$, where $\lambda = 2.6$.

(a) Discuss the three Poisson postulates in the context of this example (i.e., talk about lesions and brains); that is, what three conditions have to be true for the Poisson model to be appropriate here?

(b) Find the probability that a subject has no more than two brain lesions.

(c) The function

$$g(Y) = 0.1e^{Y/2} + 2.95Y^2$$

is used to describe the cost of treatment (in \$1000s) for a subject with Y lesions. Find the expected cost of treatment for a given subject. *Hint:* Note that $E(e^{Y/2}) = m_Y(1/2)$.

4. Suppose that Y has a beta distribution with parameters α and β so that the pdf of Y is

$$f_Y(y) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Prove that the mean of Y is

$$E(Y) = \frac{\alpha}{\alpha+\beta}.$$

(b) When $\alpha = \beta = 2$, show that the pdf of Y reduces to $6y(1-y)$, for $0 < y < 1$. Graph this particular pdf.
(c) For the model in part (b), compute $F_Y(0.75)$, where $F_Y(y)$ denotes the cumulative distribution function (cdf) of Y. It is not necessary to derive $F_Y(y)$, but you can if you want.

5. Suppose that a medical investigation is underway to study patients who are in the advanced stages of AIDS and are refusing treatment. To clinicians, it is important to measure the following times:

$$\begin{aligned} Y_1 &= \text{time until CD4 cell count is less than 150} \\ Y_2 &= \text{time until CD4 cell count is less than 500.} \end{aligned}$$

The typical behaviour of CD4 cell counts in AIDS patients is that CD4 cell counts decrease as the disease progresses. Because of this, it is reasonable to assume that Y_1 is larger than Y_2. Both times are measured in years using "time to onset of AIDS" as a baseline measure. Suppose that the joint distribution of (Y_1, Y_2) is given by

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{1}{4}e^{-y_1/2}, & 0 < y_2 < y_1 < \infty \\ 0, & \text{otherwise,} \end{cases}$$

and suppose that this is the correct model in answering the questions below.
(a) Sketch a graph of the support set in two dimensional space. Put y_1 on the horizontal axis and y_2 on the vertical axis. Describe in words what $f_{Y_1,Y_2}(y_1, y_2)$ represents.
(b) Find the probability that a single patient has CD4 count still above 150 after one year in the study; that is, compute $P(Y_1 > 1)$.
(c) Find the variance of Y_2.
(d) True or False: Y_1 and Y_2 are independent. Explain why your answer is correct.

This is an extra page for Problem 5. Use it if you wish.

6. On August 29, 2005, Hurricane Katrina blasted the Gulf Coast as a powerful Category 3 hurricane. The destruction was severe, and the onset of disease was a profound public health emergency. Shortly after the storm, a team of researchers from USC traveled to coastal Mississippi to study the transmission of West Nile Virus (WNV) among mosquitos. WNV is a rare infection, but it can have deleterious effects in the human population if transmitted. For this problem, we will assume that all mosquitos under investigation test positive for WNV with probability 0.001 and that the statuses of all mosquitos are **independent** random variables.

Testing mosquitos individually for WNV is very expensive. In order to reduce testing costs, mosquitos were caught in traps, frozen in a liquid nitrogen solution, and "combined" into pools of size 50. Then, the pool of 50 mosquitos was tested (that is, all 50 mosquitos were tested simultaneously using one test). If all mosquitos in the pool are negative, then the pool will be negative too. If **at least one** mosquito in the pool is positive, then the pool will test positive.

(a) The probability that a pool of size 50 mosquitos tests positive is about $p = 0.049$ (to three decimal places). Show this, providing explanation to support your calculations. Even if you can not show this fact, you may still use this fact in the parts below.
(b) What is the probability that, out of 10 pools tested, exactly 1 of the pools is positive?
(c) What is the probability that the first positive pool is found on the 6th pool tested?
(d) Suppose that Z denotes the number of pools to find the 3rd positive pool. What is the distribution of Z? Be precise.

7. At any one period of time, an insurance company classifies its customers as one of two types: nonstandard or standard. Define the following variables:

$$\begin{aligned} Y_1 &= \text{proportion of nonstandard customers} \\ Y_2 &= \text{proportion of standard customers.} \end{aligned}$$

Actuaries have posited that the joint distribution of (Y_1, Y_2) is

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} 60y_1y_2^2, & 0 < y_1 < 1,\ 0 < y_2 < 1,\ y_1 + y_2 < 1 \\ 0, & \text{otherwise} \end{cases}$$

Suppose that this is the correct model in answering the questions below.
(a) Sketch a graph of the support set in two dimensional space. Put y_1 on the horizontal axis and y_2 on the vertical axis.
(b) Find both conditional distributions. Make sure to note the support in each. Remember that a conditional distribution regards the conditioning variable as "fixed."
(c) Compute $P(Y_1 > 0.75|Y_2 = 0.10)$.
(d) Compute $E(Y_2|Y_1 = y_1)$. Your answer should be a function of y_1.

This is an extra page for Problem 7. Use it if you wish.