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

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

(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?

(a) $$m_Y'(t) = (-1)\left(1-t^2\right)^{-2} \times (-2t) = 2t(1-t^2)^{-2}$$

$$m_Y''(t) = 2(1-t^2)^{-2} + (2t) \times (-2) \times (1-t^2)^{-3} \times (-2t)$$
$$= 2(1-t^2)^{-2} + 8t^2(1-t^2)^{-3}$$

$$\mu = E(Y) = m_Y'(0) = 0$$

$$\sigma^2 = V(Y) = E(Y^2) - (E(Y))^2 = m_Y''(0) - 0^2 = 2 - 0 = 2$$

(b) $$\xi = \frac{E((Y-\mu)^3)}{\sigma^3} = \frac{E(Y^3)}{2^3}$$

$$m_Y'''(t) = 2 \times (-2) \times (1-t^2)^{-3} \times (-2t) + 16t \times (1-t^2)$$
$$+ 8t^2 \times (-3) \times (1-t^2)^{-4} \times (-2t)$$

$$E(Y^3) = m_Y'''(0) = 0 \quad \text{So} \quad \xi = \frac{0}{8} = 0$$

it means the distribution of Y

is symmetric w.r.t $\mu = 0$

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. $\to A$ $\quad P(A)=0.4$
- 30 percent of the customers order a Pepsi product. $\to B$ $\quad P(B)=0.3$
- Of those customers who order more than one pizza, 10 percent order a Pepsi product.

$P(B|A)=0.1$

(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.

(a) $$P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B)$$
$$= 1 - (P(A) + P(B) - P(A \cap B))$$
$$= 1 - (P(A) + P(B) - P(B|A)P(A))$$
$$= 1 - (0.4 + 0.3 - 0.1 \times 0.4) = \ldots$$

(b) $$P(A|\bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{P(A) - P(A \cap B)}{P(\bar{B})}$$
$$= \frac{P(A) - P(B|A)P(A)}{1 - P(B)}$$
$$= \frac{0.4 - 0.1 \times 0.4}{1 - 0.3} = \ldots$$

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? ← Page 55 of Notes

(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 $E(e^{Y/2}) = m_Y(1/2)$.

(b) $P(Y \le 2) = \text{poissoncdf}(\,- - -\,)$

(c) $E[g(Y)] = E[0.1e^{Y/2} + 2.95Y^2]$

$= 0.1\,E[e^{Y/2}] + 2.95\,E(Y^2)$

$= 0.1\,m_Y(\frac{1}{2}) + 2.95 \times (V(Y) + E(Y)^2)$

$Y \sim \text{Poisson}(\lambda = 2.6)$

$m_Y(t) = \exp(\lambda(e^t - 1))$

$V(Y) = E(Y) = \lambda$

$= 0.1 \times \exp(2.6 \times (e^{0.5} - 1)) + 2.95 \times (2.6 + 2.6^2)$

$= \ldots$

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.

(a) $$E(Y) = \int_0^1 y f_Y(y)\,dy = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha}(1-y)^{\beta-1}\,dy$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 y^{\alpha}(1-y)^{\beta-1}\,dy$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}$$

$$= \frac{\cancel{\Gamma(\alpha+\beta)}}{\cancel{\Gamma(\alpha)\Gamma(\beta)}} \times \frac{\cancel{\Gamma(\alpha)} \times \alpha \times \cancel{\Gamma(\beta)}}{\cancel{\Gamma(\alpha+\beta)} \times (\alpha+\beta)}$$

$$= \frac{\alpha}{\alpha+\beta}$$

(b) $\alpha=\beta=2$ $$f_Y(y) = \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} y(1-y) = \frac{3!}{1!\,1!} y(1-y)$$

$$= 6y(1-y) \quad 0<y<1$$

(c) $$F_Y(0.75) = \int_0^{0.75} 6y(1-y)\,dy = 6\int_0^{0.75} (y-y^2)\,dy$$

$$= (3y^2 - 2y^3)\Big|_0^{0.75} = \dots$$

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:

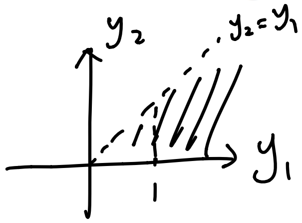
$$\begin{aligned} Y_1 &= \text{time until CD4 cell count is less than 150} \\ Y_2 &= \text{time until CD4 cell count is less than 500.} \end{aligned}$$

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.

$0 < y_2 < y_1 < +\infty$



(b) $P(Y_1 > 1) = \int_1^{+\infty} \int_0^{y_1} \frac{1}{4} e^{-y_1/2} \, dy_2 \, dy_1$

$= \int_1^{+\infty} \frac{1}{4} y_1 e^{-y_1/2} \, dy_1$

or. $P(Y_1 > 1) = 1 - P(Y_1 \le 1) = 1 - \int_0^1 \int_0^{y_1} \frac{1}{4} e^{-y_1/2} \, dy_2 \, dy_1$

$= 1 - \underbrace{\int_0^1 \frac{1}{4} y_1 e^{-y_1/2} \, dy_1}_{\text{TI-84}}$

(c) $V(Y_2) = E(Y_2^2) - [E(Y_2)]^2$

$$f_{Y_2}(y_2) = \int_{y_2}^{+\infty} \frac{1}{4} e^{-y/2} \, dy_1$$

$$= \frac{1}{2}\left(1 - e^{-y_1/2}\right)\Big|_{y_2}^{+\infty}$$

$$= \frac{1}{2} - \frac{1}{2}\left(1 - e^{-y_2/2}\right)$$

$$= \frac{1}{2} e^{-y_2/2}, \quad y_2 > 0$$

$Y_2 \sim exp(2)$

$V(Y_2) = 4$

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.

$P(\text{WNV Positive}) = 0.001$

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.

(a) $P(\text{a pool of size 50 tests positive})$

$= 1 - P(\text{a pool of size 50 tests negative})$

$= 1 - P(\text{all 50 mosquitos test negative})$

$= 1 - [P(\text{one tests negative})]^{50}$

$= 1 - [1 - 0.001]^{50} = 0.049$

(b) X: # of pools, out of the 10 pools, test positive

$X \sim \text{Binomial}(n=10, p=0.049)$

$$P(X=1) = \binom{10}{1} p (1-p)^{10-1}$$

$$= 10 \times 0.049 \times (1-0.049)^9$$

$$= \ldots$$

(c) Y: # of pools to find the 1st positive pool

$Y \sim \text{Geometric}(p=0.049)$

$$P(Y=6) = (1-p)^5 p = (1-0.049)^5 \times 0.049$$

(d) $Z \sim \text{nib}(r=3, p=0.049)$

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:

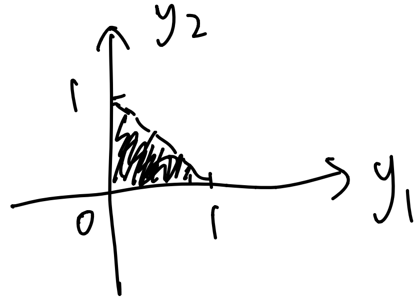
$$\begin{aligned} Y_1 &= \text{proportion of nonstandard customers} \\ Y_2 &= \text{proportion of standard customers.} \end{aligned}$$

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.

(a)



(b) $f_{Y_1}(y_1) = \int_0^{1-y_1} 60y_1y_2^2\,dy_2 = 20y_1 \times \left(y_2^3 \Big|_0^{1-y_1}\right)$

$= 20y_1(1-y_1)^3 \qquad 0<y_1<1$

$f_{Y_2}(y_2) = \int_0^{1-y_2} 60y_1y_2^2\,dy_1 = 30y_2^2\left(y_1^2 \Big|_0^{1-y_2}\right)$

$= 30y_2^2(1-y_2)^2 \qquad 0<y_2<1$

$$f_{Y_1|Y_2}(y_1|y_2) = \frac{f_{Y_1Y_2}(y_1,y_2)}{f_{Y_2}(y_2)} = \frac{60y_1y_2^2}{30y_2^2(1-y_2)^2} = \frac{2y_1}{(1-y_2)^2}, \quad 0<y_1<1-y_2$$

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_1Y_2}(y_1,y_2)}{f_{Y_1}(y_1)} = \frac{60y_1y_2^2}{20y_1(1-y_1)^3}$$

$$= \frac{3y_2^2}{(1-y_1)^3}, \quad 0<y_2<1-y_1$$

(d) The density of Y_1 given $Y_2=0.1$ is $f_{Y_1|Y_2=0.1}(y_1|0.1)$

$$= \frac{2y_1}{(1-0.1)^2}, \quad 0<y_1<1-0.1$$

So: $P(Y_1>0.75 \mid Y_2=0.1)$

$$= \int_{0.75}^{1-0.1} \frac{2y_1}{0.9^2}\, dy_1$$

$$= \frac{1}{0.9^2} \times y_1^2 \Big|_{0.75}^{0.9}$$

$$= \frac{1}{0.9^2} \times (0.9^2 - 0.75^2) = \ldots$$

(e) $$E(Y_2 \mid Y_1=y_1) = \int y_2 f_{Y_2|Y_1}(y_2|y_1)\, dy_2$$

$$= \int_0^{1-y_1} \frac{3y_2^3}{(1-y_1)^3}\, dy_2$$

$$= \frac{1}{(1-y_1)^3} \times \frac{3}{4}\left(y_2^4 \Big|_0^{1-y_1}\right)$$

$$= \frac{3}{4} \times \frac{(1-y_1)^4}{(1-y_1)^3} = \frac{3}{4}(1-y_1)$$