APMA 1650

Homework 7 - Solutions

Problem 1. (22 pts) Suppose that you are taking only two classes this semester (sounds nice). The professor for each class has three different dates to choose for their final exam (Day 1,Day 2 or Day 3). Each professor chooses one of the three days at random for their final. Let Y_1 be the number of exams you have on day 1, and Y_2 be the number of exams you have on day 2.

a. (5 pts) Find the joint PMF for Y_1 and Y_2. Write it as a table.

b. (5 pts) What is $F_{Y_1,Y_2}(1,0)$?

c. (5 pts) Find the marginal PMFs for Y_1 and Y_2 from the joint PMF.

d. (2 pts) Identify the name of the distribution for Y_1 and give its appropriate parameters. Explain why this is the case.

e. (5 pts) Are Y_1 and Y_2 independent? Explain.

Solution.

a. In general, this is given by the multinomial distribution (see the end of section 2.1.3 for a reminder)

$$P_{Y_1,Y_2}(y_1,y_2) = \binom{2}{y_1,\ y_2,\ 2-y_1-y_2}\left(\frac{1}{3}\right)^{y_1}\left(\frac{1}{3}\right)^{y_2}\left(\frac{1}{3}\right)^{2-y_1-y_2},$$

for $(y_1,y_2) \in R_{Y_1,Y_2}$, where the range is

$$R_{Y_1,Y_2} = \{(y_1,y_2) \in \mathbb{Z}\times\mathbb{Z} : y_1 \geq 0,\ y_2 \geq 0,\ y_1+y_2 \leq 2\}.$$

Here we think of each professor as a "trial" involving randomly picking from one of 3 possible outcomes with probability 1/3 each . Then, each professor corresponds to a trial (2 trials total). The multinomial probability expresses the probability that you see exactly y_1, y_2 and $2-y_1-y_2$ exams on day 1, day 2 and day 3 respectively.

However this is a bit overkill, we can simply work out each case separately in the table (this is totally fine to do for this problem). Indeed, if $Y_1 = 0$, and $Y_2 = 0$, then both professors picked the 3rd day, this has a probability of

$$(1/3)(1/3) = 1/9$$

of happening. Similarly, if $Y_1 = 0$ and $Y_2 = 1$, then either one professors picked day 2 and the other picked day 3, this has a probability of

$$(1/3)(1/3) + (1/3)(1/3) = 2/9$$

of happening. If $Y_1 = 0$ and $Y_2 = 2$, then both professors picked day 2 and therefore has a probability of

$$(1/3)(1/3) = 1/9$$

of happening. Finally if $Y_1 = 1$ and $Y_2 = 1$, then either one of the professors picked day 1 and the other picked day 2, this has a probability of

$$(1/3)(1/3) + (1/3)(1/3) = 2/9$$

of happening. All other probabilities are immediately determined, since the probability table is symmetric (i.e. $P(Y_1 = 0, Y_2 = 1) = P(Y_1 = 1, Y_2 = 0)$), and $P(Y_1 = y_1, Y_2 = y_2) = 0$ if $y_1 + y_2 > 2$ (since you are only taking two classes). Therefore the table we get is

$Y_1 \backslash Y_2$	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

.

b. Let find $F_{Y_1,Y_2}(1, 0)$. This is given by

$$F_{Y_1,Y_2}(1, 0) = P(Y_1 \leq 1, Y_2 \leq 0) = P(Y_1 \leq 1, Y_2 = 0) = 1/9 + 2/9 = 1/3.$$

c. We will work out the marginals via the augmented table

$Y_1 \backslash Y_2$	0	1	2	P_{Y_1}
0	1/9	2/9	1/9	4/9
1	2/9	2/9	0	4/9
2	1/9	0	0	1/9
P_{Y_2}	4/9	4/9	1/9	1

.

d. As we saw in the midterm, the distributions for Y_1 and Y_2 are just Binomial distributions, since we can think of the probability that a professor picks day 1 (or day 2) is just a Bernoulli trial with probability $p = 1/3$ of happening. Then the total number "successes" out of two "trials" is just

$$Y_1, Y_2 \sim \text{Binomial}(2, 1/3).$$

e. The are definitely not independent. For instance $P_{Y_1,Y_2}(2, 2) = 0$, but $P_{Y_1}(2)P_{Y_2}(2) = (1/9)(1/9) = 1/81$.

Problem 2. (20 pts) You have a bag with 9 balls in it 4 red, 3 green and 2 blue. You reach in an pull out 3 balls. Out of the three balls you pulled out, let X_1 be the number of red balls, and X_2 be the number of green balls.

a. (5 pts) What is the joint PMF of X_1 and X_2? State it as a function of x_1, x_2. Clearly state the range R_{X_1,X_2}. You may leave your answer in terms of combinatorial quantities.

b. (5 pts) Are X_1 and X_2 independent?

c. (5 pts) What is the conditional PMF of X_1 given $X_2 = x_2$? State it as a function of x_1 and x_2.

d. (5 pts) What is the probability $P(X_1 = 1 | X_2 \geq 1)$?

Solution.

a. We can calculate this probability using counting. It is unordered sampling without replacement. First, we consider the number of ways to to choose 3 balls our of 9 total. That is $\binom{9}{3}$. The number of ways to choose x_1 red balls out of 4 total is just $\binom{4}{x_1}$, the number of ways to choose x_2 green balls out of 3 total is just $\binom{3}{x_2}$, and the number of ways to choose the remaining $3 - x_1 - x_2$ balls as blue is then just $\binom{2}{3-x_1-x_2}$. This gives a PMF of

$$P_{X_1,X_2}(x_1, x_2) = \frac{\binom{4}{x_1}\binom{3}{x_2}\binom{2}{3-x_1-x_2}}{\binom{9}{3}},$$

for $(x_1, x_2) \in R_{X_1,X_2}$. The range R_{X_1,X_2} is determined by the fact that $3 - x_1 - x_2$ must be between 0 and 2 (namely the total number of red and green balls can't be more than 3, but you must always choose at least a red or a green since there are only two blues)

$$R_{X_1,X_2} = \{(x_1, x_2) \in \mathbb{Z} \times \mathbb{Z} : x_1 \geq 0, x_2 \geq 0, 1 \leq x_1 + x_2 \leq 3\}.$$

b. They are most certainly not independent. To determine this mathematically, we first need to know the marginals. In fact, you should know right off the bat that X_1 and X_2 are hypergeometric random variables, since it is just asking how many reds (or greens) are being chosen out of a bag of 4 reds and 5 non-reds. Note that $R_{X_1} = R_{X_2} = \{0, 1, 2, 3\}$ since the total number of red or greens that you draw out can always vary from 0 to 3. Therefore we have the hypergeometric PMFs

$$P_{X_1}(x_1) = \frac{\binom{4}{x_1}\binom{5}{3-x_1}}{\binom{9}{3}}$$

and

$$P_{X_2}(x_2) = \frac{\binom{3}{x_2}\binom{6}{3-x_2}}{\binom{9}{3}}.$$

It follows that X_1, X_2 are not independent since

$$P_{X_1,X_2}(3,3) = 0, \quad \text{while} \quad P_{X_1}(3)P_{X_2}(3) = \frac{\binom{4}{3}\binom{5}{0}\binom{3}{3}\binom{6}{0}}{\binom{9}{3}^2} \neq 0$$

Alternate Solution: Lets see how you could also compute the marginals directly from the joint distribution

$$\begin{aligned}
P_{X_1}(x_1) &= \sum_{x_2=0}^{3} P_{X_1,X_2}(x_1,x_2) \\
&= \sum_{x_2=1-x_1}^{3-x_1} \frac{\binom{4}{x_1}\binom{3}{x_2}\binom{2}{3-x_1-x_2}}{\binom{9}{3}} \quad \text{using that } 1 \leq x_1 + x_2 \leq 3 \\
&= \frac{\binom{4}{x_1}}{\binom{9}{3}} \sum_{x_2=\max\{1-x_1,0\}}^{3-x_1} \binom{3}{x_2}\binom{2}{3-x_1-x_2} \quad \text{pulling out constants} \\
&= \frac{\binom{4}{x_1}\binom{5}{3-x_1}}{\binom{9}{3}} \quad \text{using Vandermonde's identity} \quad \sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i} = \binom{m+n}{k}
\end{aligned}$$

Similarly we have for $P_{X_2}(x_2)$

$$\begin{aligned}
P_{X_1}(x_1) &= \sum_{x_1=0}^{3} P_{X_1,X_2}(x_1,x_2) \\
&= \sum_{x_1=0}^{3-x_1} \frac{\binom{4}{x_1}\binom{3}{x_2}\binom{2}{3-x_1-x_2}}{\binom{9}{3}} \quad \text{using that } 1 \leq x_1 + x_2 \leq 3 \\
&= \frac{\binom{3}{x_2}}{\binom{9}{3}} \sum_{x_2=0}^{3-x_1} \binom{4}{x_1}\binom{2}{3-x_1-x_2} \quad \text{pulling out constants} \\
&= \frac{\binom{3}{x_2}\binom{6}{3-x_2}}{\binom{9}{3}} \quad \text{using Vandermonde's identity.}
\end{aligned}$$

c. Using the joint PMF and the marginal for X_2, the conditional PMF of X_1 given $X_2 = x_2$ is for $1 \leq x_1 + x_2 \leq 3$, $x_1 \geq 0$, $x_2 \geq 0$

$$\begin{aligned}
P_{X_1|X_2}(x_1|x_2) &= \frac{P_{X_1,X_2}(x_1,x_2)}{P_{X_2}(x_2)} \\
&= \frac{\binom{4}{x_1}\binom{2}{3-x_1-x_2}}{\binom{6}{3-x_2}}
\end{aligned}$$

We can write this more carefully considering the range for each $x_2 = \{0,1,2,3\}$ and we have

$$P_{X_1|X_2}(x_1|x_2) = \frac{\binom{4}{x_1}\binom{2}{3-x_1-x_2}}{\binom{6}{3-x_2}}, \quad \text{for } \max\{1-x_2,0\} \leq x_1 \leq 3-x_2$$

Note that this means that $X_1|X_2 = x_2$ is Hypergeometric$(4, 2, 3 - x_2)$, which makes sense since if $X_2 = x_2$, you can just remove the x_2 green balls from the problem all together and think about drawing $3 - x_2$ balls out of a bag of 6 balls (4 red and 2 blue).

d. Note that by definition,

$$P(X_1 = 1|X_2 \geq 1) = \frac{P(X_1 = 1, X_2 \geq 1)}{P(X_2 \geq 1)}$$

We know from the formula for the PMF above that

$$\begin{aligned} P(X_1 = 1, X_2 \geq 1) &= P(X_1 = 1, X_2 = 1) + P(X_1 = 1, X_2 = 2) \\ &= \frac{\binom{4}{1}\binom{3}{1}\binom{2}{1}}{\binom{9}{3}} + \frac{\binom{4}{1}\binom{3}{2}\binom{2}{0}}{\binom{9}{3}} = \frac{2}{7} + \frac{1}{7} = \frac{3}{7}. \end{aligned}$$

and

$$\begin{aligned} P(X_2 \geq 1) &= 1 - P(X_2 = 0) \\ &= 1 - \frac{\binom{3}{0}\binom{6}{3}}{\binom{9}{3}} = 1 - \frac{5}{21} = \frac{16}{21}. \end{aligned}$$

Therefore

$$P(X_1 = 1|X_2 \geq 1) = \frac{\frac{3}{7}}{\frac{16}{21}} = \frac{9}{16}.$$

Problem 3. (25 pts) Let X and Y be two jointly continuous random variables with joint PDF

$$f_{XY}(x, y) = \begin{cases} cx & 0 \leq x \leq 1, \quad 0 \leq y \leq x^2 \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

a. (5 pts) Find the constant c that makes f_{XY} a valid joint PDF.

b. (5 pts) Find the marginal PDFs of X and Y. Be sure to state their range.

c. (5 pts) Find the conditional densities $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$. Be sure to state them as PDFs for each fixed value of the conditioned variable (e.g. state $f_{X|Y}(x|y)$ as a PDF in x for each fixed value $y \in R_Y$.) What is the "name" of the conditional distribution of Y given $X = x$?

d. (5 pts) Find the conditional expectations $E[X|Y = y]$ and $E[Y|X = x]$.

e. Let $Z = E[Y|X]$. What is the PDF of Z?

Solution:

a. We need

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}y\mathrm{d}x = \int_0^1 \int_0^{x^2} cx\mathrm{d}y\mathrm{d}x = 1$$

This reduces to

$$c\int_0^1 \int_0^{x^2} x^3\mathrm{d}x = c\left[\frac{1}{4}x^4\right]_0^1 = \frac{c}{4} = 1$$

Therefore we need $c = 4$.

b. To compute the marginals we find for $x \in [0,1]$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y = \int_0^{x^2} 4x\mathrm{d}y = 4x^3.$$

Therefore

$$f_X(x) = \begin{cases} 4x^3 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

Similarly we note that if $x \geq 0$ then

$$0 \le y \le x^2 \le \quad \Leftrightarrow \quad 0 \le \sqrt{y} \le x \le 1.$$

So for $y \in [0,1]$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x = \int_{\sqrt{y}}^1 4x\mathrm{d}x = \left[2x^2\right]_{\sqrt{y}}^1 = 2(1-y).$$

Therefore

$$f_Y(y) = \begin{cases} 2(1-y) & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

c. To find the conditional densities, we compute for $(x,y) \in R_{XY}$

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{4x}{2(1-y)} = \frac{2x}{1-y}$$

and therefore for $y \in [0,1)$

$$f_{X|Y}(x|y) = \begin{cases} \frac{2x}{1-y} & \sqrt{y} \le x \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

Note that for $y = 1$ we actually get a delta function at $x = 1$

$$f_{X|Y}(x|1) = \delta(x-1).$$

Similarly for $(x, y) \in R_{XY}$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{4x}{4x^3} = \frac{1}{x^2}$$

and therefore for $x \in [0,1]$

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{x^2} & 0 \le y \le x^2 \\ 0 & \text{otherwise} \end{cases}.$$

We see that the random variable $Y|X = x$ is just a Uniform$(0, x^2)$ random variable

d. The conditional expectations are just given by

$$\begin{aligned} E[X|Y=y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\mathrm{d}x \\ &= \int_{\sqrt{y}}^{1} \frac{2x^2}{1-y}\mathrm{d}x = \frac{2}{3(1-y)}\left[x^3\right]_{\sqrt{y}}^{1} = \frac{2(1-y^{3/2})}{3(1-y)}. \end{aligned}$$

and

$$\begin{aligned} E[Y|X=x] &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\mathrm{d}y \\ &= \int_{0}^{x^2} \frac{y}{x^2}\mathrm{d}y = \left[\frac{y^2}{2x^2}\right]_0^{x^2} = \frac{1}{2}x^2. \end{aligned}$$

e. From part d we see that

$$Z = E[Y|X] = \frac{1}{2}X^2.$$

Therefore we are looking for the PDF of $\frac{1}{2}X^2$. Computing the CDF, we find (since $R_X = [0,1]$)

$$F_Z(z) = P(Z \le z) = P(X \le \sqrt{2z}) = F_X(\sqrt{2z}).$$

Taking the derivative of both sides with respect to z, when $0 \le \sqrt{2z} \le 1$

$$\begin{aligned} f_Z(z) &= f_X(\sqrt{2z})\frac{\mathrm{d}}{\mathrm{d}z}\left(\sqrt{2z}\right) \\ &= 4(2z)^{3/2}\frac{1}{\sqrt{2z}} \\ &= 8z \end{aligned}.$$

Therefore, since for $z \ge 0$,

$$0 \le \sqrt{2z} \le 1 \quad \Leftrightarrow \quad 0 \le z \le \frac{1}{2},$$

we have

$$f_Z(z) = \begin{cases} 8z & 0 \le z \le 1/2 \\ 0 & \text{otherwise} \end{cases}.$$

Problem 4. (15 pts) Let (Y_1, Y_2) be uniformly distributed in the triangle $T \subset \mathbb{R}^2$ with corners

$$(0,0), (3,1), (3,3).$$

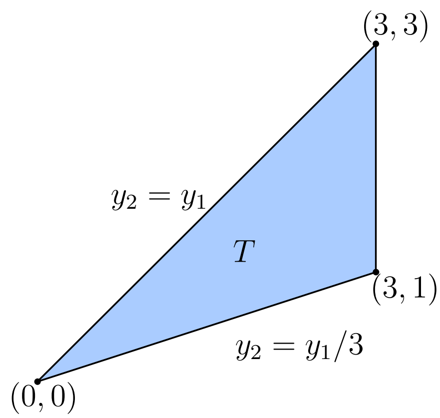
a. (5 pts) What is the joint PDF of Y_1, Y_2? Be sure to specify it's domain and make sure that it is a valid density function.

b. (5 pts) What is $P(Y_1 \geq 1, Y_2 \leq 1)$? (It helps to draw a picture!)

c. (5 pts) What is $P(3Y_2 - 2Y_1 \geq 0)$? (Again, draw a picture!)

Solution:

a. The triangular domain T is pictured below in the (y_1, y_2) plane



It can be describes a the region bounded by the two curves $y = x/3$ and $y = x$.

$$T = \{(x, y) \,:\, 0 \leq x \leq 1,\, x/3 \leq y \leq x\}$$

Being a uniform RV the PDF is given by

$$f_{XY}(x, y) = \begin{cases} c & (x, y) \in T \\ 0 & \text{otherwise} \end{cases}$$

To ensure that it is a valid PDF we find need

$$1 = \int_0^3 \int_{x/3}^{x} c\mathrm{d}y\mathrm{d}x = \int_0^3 \frac{2c}{3} x\mathrm{d}x = \frac{c}{3}x^2\Big|_0^3 = 3c.$$

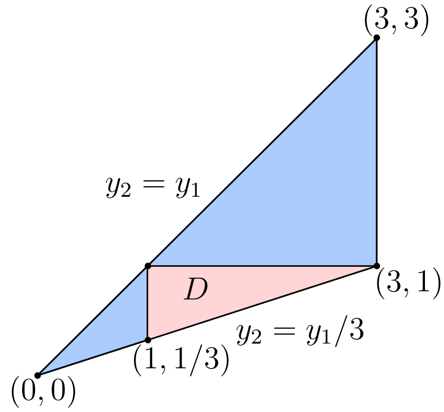
Therefore $c = 1/3$. It follows that

$$f_{XY}(x, y) = \begin{cases} \frac{1}{3} & (x, y) \in T \\ 0 & \text{otherwise} \end{cases}.$$

Alternate Solution Another way of seeing this without integrating is that for a uniform distribution on T

$$c = \frac{1}{\text{Area}(T)} = 1/3.$$

b. To compute this probability, we look at the intersection of the rectangle $\{y_1 \geq 1, y_2 \leq 1\}$ with the triangle T. This gives another triangle D shown in the figure below



the triangle D can be described by

$$D = \{(y_1, y_2) : 1 \leq y_1 \leq 3, \ , y_1/3 \leq y_2 \leq 1\}$$

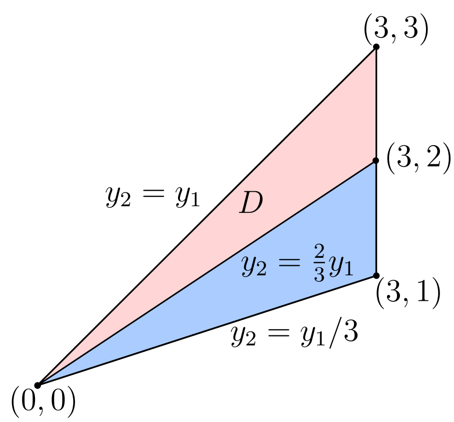
The probability is therefore

$$\begin{aligned} P(Y_1 \geq 1, Y_2 \leq 1) &= \iint_D \frac{1}{3} \mathrm{d}A = \int_1^3 \int_{y_1/3}^1 \frac{1}{3} \mathrm{d}y_2 \mathrm{d}y_1 \\ &= \int_1^3 \frac{1}{3}(1 - y_1/3)\mathrm{d}y_1 = \frac{1}{3}\left[y_1 - \frac{1}{6}y_1^2\right]_1^3 = \frac{2}{9} \end{aligned}$$

Alternate Solution Another way of seeing this is that

$$P(Y_1 \geq 1, Y_2 \leq 1) = \frac{\text{Area}(D)}{\text{Area}(T)} = \frac{\frac{2}{3}}{3} = \frac{2}{9}$$

c. Finally, we calculate $P(3Y_2 - 2Y_1 \geq 0)$. To do this, we look at the intersection of the region $\{3y_2 - 2y_1 \geq 0\}$ (the region above the line $y_2 = \frac{2}{3}y_1$) with the triangle T. This gives another triangle D illustrated below



The triangle can be described by

$$D = \left\{ (y_1, y_2) \,:\, 0 \le y_1 \le 3,\, \frac{2}{3}y_1 \le y_2 \le y_1 \right\}$$

and therefore the probability is just

$$\begin{aligned} P(3Y_2 - 2Y_1 \ge 0) &= \iint_D \frac{1}{3}\mathrm{d}A = \int_0^3 \int_{\frac{2}{3}y_1}^{y_1} \frac{1}{3}\mathrm{d}y_2\mathrm{d}y_1 \\ &= \frac{1}{3}\int_0^3 [y_2]_{\frac{2}{3}y_1}^{y_1}\,\mathrm{d}y_1 = \frac{1}{9}\int_0^3 y_1\mathrm{d}y_1 \\ &= \frac{1}{9}\left[\frac{y_1^2}{2}\right]_0^3 = \frac{1}{2}. \end{aligned}$$

Alternate Solution.You can see from the picture that D is exactly half of the area of T. Specifically we have

$$P(3Y_2 - 2Y_1 \ge 0) = \frac{\text{Area}(D)}{\text{Area}(T)} = \frac{\frac{3}{2}}{3} = \frac{1}{2}.$$

Problem 5. (18 pts) Suppose that the number of defects Y in a given chip fabrication process is known to follow a Poisson distribution with rate Λ. However, the rate Λ is itself an Exponential(2) random variable.

a. (5 pts) Use the law of iterated expectation find the expected number of defects per chip by first finding the expected number of defects for a given $\Lambda = \lambda$.

b. (5 pts) Use the law of total variance to find the variance of Y.

c. (8 pts) Show that $Y + 1$ is just a Geometric random variable $Y + 1 \sim \text{Geometric}(p)$. What is the value of p? (Hint: Use the law of total probability and the definition of the Gamma function $\Gamma(k+1) = \int_0^\infty x^k e^{-x}\mathrm{d}x = k!$ to find the PMF of Y).

Solution:

a. Using the law of iterated expectation and the fact that $Y|\Lambda = \lambda$ is Poisson(λ)

$$E[Y|\Lambda = \lambda] = \lambda.$$

We find using the mean of the exponential random variable

$$EY = E\left[E[Y|\Lambda]\right] = E[\Lambda] = \frac{1}{2}.$$

b. Similarly for the law of total variance, we have using the variance of Poisson(λ) that

$$\text{Var}(Y|\Lambda = \lambda) = \lambda.$$

Therefore using the properties of exponential random variables

$$\begin{aligned}
\text{Var}(Y) &= E[\text{Var}(Y|\Lambda)] + \text{Var}(E[Y|X]) \\
&= E[\Lambda] + \text{Var}(\Lambda) \\
&= \frac{1}{2} + \left(\frac{1}{2}\right)^2 = \frac{3}{4}.
\end{aligned}$$

c. To show that $Y + 1$ is a geometric random variable, we use the law of total probability

$$\begin{aligned}
P(Y+1=k) &= \int_0^\infty P(Y=k-1|\Lambda=\lambda)P_\Lambda(\lambda)\mathrm{d}\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!}2e^{-2\lambda}\mathrm{d}\lambda \\
&= \frac{2}{(k-1)!}\int_0^\infty \lambda^{k-1}e^{-3\lambda}\mathrm{d}\lambda \\
&= \frac{2}{(k-1)!}\left(3^{-k}\int_0^\infty u^{k-1}e^{-u}\mathrm{d}u\right) \quad \text{substitution } u = 3\lambda \\
&= (2)(3^{-k})\frac{\Gamma(k)}{(k-1)!} \quad \text{definition of } \Gamma(k) \\
&= \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^k \quad \Gamma(k) = (k-1)!
\end{aligned}$$

Therefore $Y + 1 \sim \text{Geometric}(2/3)$.