5.114. We are given the marginal pdfs of Y_1 and Y_2. You should note that

$$\begin{aligned} Y_1 &\sim \text{gamma}(4,1) \\ Y_2 &\sim \text{exponential}(2). \end{aligned}$$

Therefore, $E(Y_1) = 4$, $V(Y_1) = 4$, $E(Y_2) = 2$, and $V(Y_2) = 4$.

(a) With $U = Y_1 - Y_2$, we have

$$E(U) = E(Y_1 - Y_2) = E(Y_1) - E(Y_2) = 4 - 2 = 2.$$

(b) Because Y_1 and Y_2 are independent by assumption, we have

$$V(U) = V(Y_1 - Y_2) = V(Y_1) + V(Y_2) - 2\underbrace{\text{Cov}(Y_1, Y_2)}_{=\,0} = 4 + 4 = 8.$$

Note that we could not calculate $V(U)$ if we did not make an independence assumption. We don't know the joint pdf of Y_1 and Y_2; we only know the marginal distributions.

(c) This part asks us to find $P(Y_1 - Y_2 < 0)$. This probability is found by using the joint distribution of Y_1 and Y_2. *Therefore, the only way we can find* $P(Y_1 - Y_2 < 0)$ *exactly is to assume* Y_1 *and* Y_2 *are independent.* Under this assumption, the joint pdf of Y_1 and Y_2, for $y_1 > 0$ and $y_2 > 0$, is given by

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{Y_1}(y_1) f_{Y_2}(y_2) \\ &= \frac{1}{6} y_1^3 e^{-y_1} \times \frac{1}{2} e^{-y_2/2} \;=\; \frac{1}{12} y_1^3 e^{-y_1 - y_2/2}. \end{aligned}$$

Summarizing, provided Y_1 and Y_2 are independent,

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \dfrac{1}{12} y_1^3 e^{-y_1 - y_2/2}, & y_1 > 0,\ y_2 > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The support of Y_1 and Y_2 is $R = \{(y_1, y_2) : y_1 \geq 0, y_2 \geq 0\}$, the entire first quadrant. This set is shown at the top of the next page (left). The joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ is a three-dimensional function which takes the value $\frac{1}{12} y_1^3 e^{-y_1 - y_2/2}$ over this region (i.e., the entire first quadrant) and is otherwise equal to zero.

We find $P(Y_1 - Y_2 < 0)$ by integrating the joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ over the set

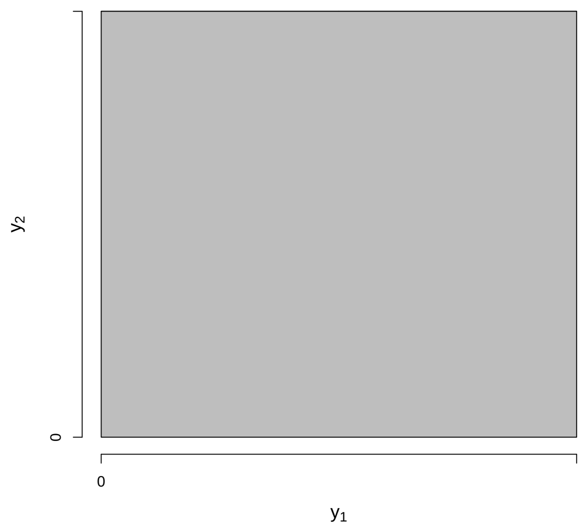
$$B = \{(y_1, y_2) : y_1 \geq 0,\ y_2 \geq 0,\ y_1 - y_2 < 0\}.$$

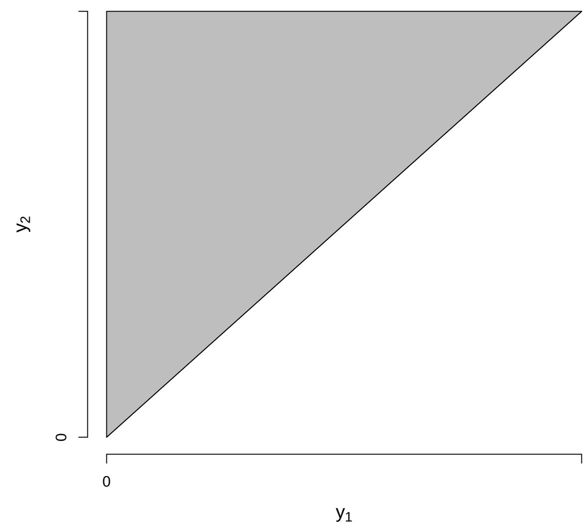
This set is shown at the top of the next page (right). Note that the boundary of this set is

$$y_1 - y_2 = 0 \quad \Longrightarrow \quad y_2 = y_1.$$

The limits to calculate $P(Y_1 - Y_2 < 0)$ come from this picture. We have

$$\begin{aligned} P(Y_1 - Y_2 < 0) &= \int_{y_2=0}^{\infty} \int_{y_1=0}^{y_2} \frac{1}{12} y_1^3 e^{-y_1 - y_2/2} \, dy_1 dy_2 \\ &= \frac{1}{12} \int_{y_2=0}^{\infty} e^{-y_2/2} \left(\int_{y_1=0}^{y_2} y_1^3 e^{-y_1} \, dy_1 \right) dy_2. \end{aligned}$$





Unfortunately, the only way to do the integral

$$\int_{y_1=0}^{y_2} y_1^3 e^{-y_1}\ dy_1$$

is to use integration by parts three times. Note that the integrand is the gamma$(4,1)$ kernel, but the integral is not over $(0,\infty)$, so this integral does not equal $\Gamma(4)=6$. There is no way I'm using integration by parts three times! Let's reverse the order of integration.

$$\begin{aligned}
P(Y_1 - Y_2 < 0) &= \int_{y_1=0}^{\infty}\int_{y_2=y_1}^{\infty} \frac{1}{12} y_1^3 e^{-y_1-y_2/2}\ dy_2 dy_1 \\
&= \frac{1}{12}\int_{y_1=0}^{\infty} y_1^3 e^{-y_1}\left(\int_{y_2=y_1}^{\infty} e^{-y_2/2}\ dy_2\right) dy_1 \\
&= \frac{1}{12}\int_{y_1=0}^{\infty} y_1^3 e^{-y_1}\left(-2e^{-y_2/2}\Big|_{y_2=y_1}^{\infty}\right) dy_1 \\
&= \frac{1}{12}\int_{y_1=0}^{\infty} y_1^3 e^{-y_1}\left(0+2e^{-y_1/2}\right) dy_1 \;=\; \frac{1}{6}\int_{y_1=0}^{\infty} y_1^3 e^{-3y_1/2} dy_1.
\end{aligned}$$

Note that the last integrand can be written as

$$y_1^3 e^{-3y_1/2} = y_1^{4-1} e^{-y_1/\frac{2}{3}},$$

which is the gamma$(4,2/3)$ kernel. Therefore, we have

$$P(Y_1 - Y_2 < 0) = \frac{1}{6}\int_{y_1=0}^{\infty} y_1^{4-1} e^{-y_1/\frac{2}{3}} dy_1 \;=\; \frac{1}{6}\Gamma(4)\left(\frac{2}{3}\right)^4 \approx 0.198.$$

Note: If we did not assume Y_1 and Y_2 are independent in this example, then we could not compute $P(U<0) = P(Y_1 - Y_2 < 0)$ exactly in part (c). The only thing we know about U is that it has mean $E(U) = \mu = 2$ and variance $V(U) = \sigma^2 = 8$, calculated in parts (a) and (b). In lieu of independence, we could use Tchebysheff's Theorem to calculate lower bounds on the probability

$$P(\mu - k\sigma < U < \mu + k\sigma)$$

for $k > 1$, but that's about all we could do. In other words, the independence assumption is critical in part (c).

5.126. We think of each item as a "trial" under the following framework:

			Probability	Count
	$\longrightarrow$	Category 1 ("exactly 1 defect")	$p_1 = 0.10$	Y_1
Trial outcome	$\longrightarrow$	Category 2 ("more than 1 defect")	$p_2 = 0.05$	Y_2
	$\longrightarrow$	Category 3 ("no defects")	$p_3 = 0.85$	Y_3

If $n = 10$ items are randomly selected, then $\mathbf{Y} = (Y_1, Y_2, Y_3)$ follows a trinomial distribution with $n = 10$ and the probabilities above; i.e., $\mathbf{Y} \sim \text{mult}\,(n = 10, \mathbf{p} = (0.10, 0.05, 0.85))$. The probability mass function of $\mathbf{Y}$ is

$$p_{\mathbf{Y}}(y_1, y_2, y_3) = \frac{10!}{y_1!y_2!y_3!}\,(0.10)^{y_1}(0.05)^{y_2}(0.85)^{y_3},$$

with support $R = \{(y_1, y_2, y_3) : y_j = 0, 1, 2, ..., 10;\ \sum_{j=1}^{3} y_j = 10\}$.

In this problem, we want to calculate $E(Y_1 + 3Y_2)$ and $V(Y_1 + 3Y_2)$. Recall that, marginally,

$$\begin{aligned} Y_1 &\sim b(n = 10, p_1 = 0.10) \\ Y_2 &\sim b(n = 10, p_2 = 0.05). \end{aligned}$$

Therefore, $E(Y_1) = np_1 = 1$, $E(Y_2) = np_2 = 0.5$, $V(Y_1) = np_1(1 - p_1) = 0.9$, and $V(Y_2) = np_2(1 - p_2) = 0.475$. We have

$$E(Y_1 + 3Y_2) = E(Y_1) + 3E(Y_2) = 1 + 3(0.5) = 2.5.$$

Also,

$$V(Y_1 + 3Y_2) = V(Y_1) + 3^2 V(Y_2) + 2(1)(3)\text{Cov}(Y_1, Y_2).$$

Recall that

$$\text{Cov}(Y_1, Y_2) = -np_1p_2 = -10(0.10)(0.05) = -0.05.$$

Therefore,

$$V(Y_1 + 3Y_2) = 0.9 + 9(0.475) + 6(-0.05) = 4.875.$$

5.130. We have $Y_1, Y_2, ..., Y_n$ are mutually independent random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$, for $i = 1, 2, ..., n$.

(a) The covariance of U_1 and U_2 is

$$\text{Cov}(U_1, U_2) = \text{Cov}\left(\sum_{i=1}^{n} a_iY_i, \sum_{j=1}^{n} b_jY_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ib_j\text{Cov}(Y_i, Y_j).$$

Now, we know that $\text{Cov}(Y_i, Y_j) = 0$ whenever $i \neq j$; i.e., whenever the subscripts don't match, because $Y_1, Y_2, ..., Y_n$ are mutually independent. Therefore, out of the n^2 terms in this sum, only n of them are nonzero−those that remain when the subscripts on each sum match; e.g., $i = j = 1$, $i = j = 2$, and so on. Therefore, we can write

$$\text{Cov}(U_1, U_2) = \sum_{i=1}^{n} a_i b_i \text{Cov}(Y_i, Y_i).$$

Recall that the covariance of any random variable with itself is the variance; i.e., $\text{Cov}(Y_i, Y_i) = V(Y_i) = \sigma^2$, for $i = 1, 2, ..., n$. Therefore,

$$\text{Cov}(U_1, U_2) = \sum_{i=1}^{n} a_i b_i \sigma^2 = \sigma^2 \sum_{i=1}^{n} a_i b_i.$$

Because $\sigma^2 > 0$, we have $\text{Cov}(U_1, U_2) = 0$ (and hence U_1 and U_2 are orthogonal) if and only if $\sum_{i=1}^{n} a_i b_i = 0$.

(b) We have $Y_1, Y_2, ..., Y_n$ are mutually independent **normal** random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$, for $i = 1, 2, ..., n$. We will learn in STAT 512 why U_1 and U_2 are bivariate normal (the authors just tell you this is true). Recall that in the bivariate normal model,

$$U_1 \text{ and } U_2 \text{ are independent } \iff \rho = 0.$$

The correlation of U_1 and U_2 is

$$\rho = \frac{\text{Cov}(U_1, U_2)}{\sqrt{V(U_1)V(U_2)}}.$$

We calculated $\text{Cov}(U_1, U_2)$ above. Note that

$$V(U_1) = V\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} a_i^2 V(Y_i) + \underbrace{\sum\sum_{i \neq j} a_i a_j \text{Cov}(Y_i, Y_j)}_{=\,0} = \sum_{i=1}^{n} a_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^{n} a_i^2.$$

The same argument shows

$$V(U_2) = \sigma^2 \sum_{i=1}^{n} b_i^2.$$

Therefore,

$$\rho = \frac{\text{Cov}(U_1, U_2)}{\sqrt{V(U_1)V(U_2)}} = \frac{\sigma^2 \sum_{i=1}^{n} a_i b_i}{\sqrt{\sigma^2 \sum_{i=1}^{n} a_i^2 \times \sigma^2 \sum_{i=1}^{n} b_i^2}} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2}}.$$

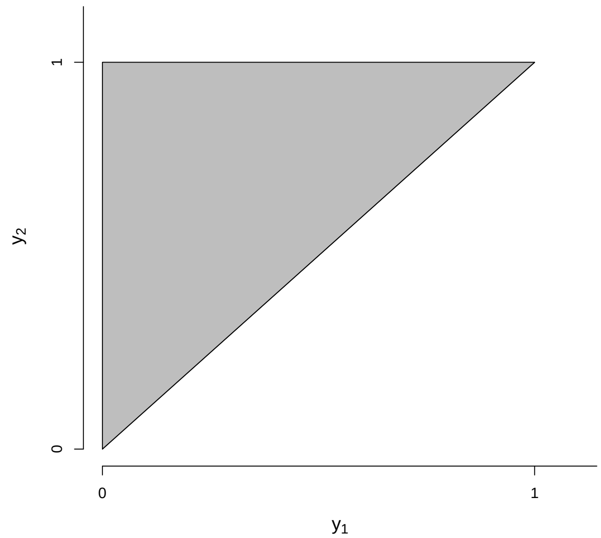
Suppose U_1 and U_2 are orthogonal. Then $\sum_{i=1}^{n} a_i b_i = 0$ from part (a). Therefore, $\rho = 0$. Therefore, U_1 and U_2 are independent.

Note: We know that $-1 \leq \rho \leq 1$, so $\rho^2 \leq 1$. In this last expression, this implies

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2,$$

which is the Cauchy-Schwarz Inequality for sums. Interesting!

5.133. The support is $R = \{(y_1, y_2) : 0 \le y_1 \le y_2 \le 1\}$, the upper triangle of the unit square. See below:



The joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ is a three-dimensional function which takes the value $6(1 - y_2)$ over this region and is otherwise equal to zero.

Note: We had this same joint distribution in Problem 5.77 from HW11. Recall that we calculated

$$f_{Y_1}(y_1) = \begin{cases} 3(1-y_1)^2, & 0 \le y_1 \le 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad f_{Y_2}(y_2) = \begin{cases} 6y_2(1-y_2), & 0 \le y_2 \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Because $Y_1 \sim \text{beta}(1, 3)$, we have

$$E(Y_1) = \frac{1}{1+3} = \frac{1}{4}.$$

(a) The conditional expectation $E(Y_1|Y_2 = y_2)$ is the mean of the conditional distribution described by $f_{Y_1|Y_2}(y_1|y_2)$. Let's find $f_{Y_1|Y_2}(y_1|y_2)$. Note that when Y_2 is fixed at y_2, then y_1 must fall between 0 and y_2; see the picture above. Therefore, $f_{Y_1|Y_2}(y_1|y_2) > 0$ when $0 \le y_1 \le y_2$, and otherwise $f_{Y_1|Y_2}(y_1|y_2) = 0$. Therefore, for $0 \le y_1 \le y_2$, we have

$$f_{Y_1|Y_2}(y_1|y_2) = \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_2}(y_2)} = \frac{6(1-y_2)}{6y_2(1-y_2)} = \frac{1}{y_2}.$$

Summarizing,

$$f_{Y_1|Y_2}(y_1|y_2) = \begin{cases} \dfrac{1}{y_2}, & 0 \le y_1 \le y_2 \\ 0, & \text{otherwise.} \end{cases}$$

We recognize this as the pdf of a uniform distribution which allows y_1 to range from 0 to y_2 (y_2 regarded as fixed). In other words, $Y_1|Y_2 = y_2 \sim \mathcal{U}(0, y_2)$. Borrowing what we know about the uniform distribution, the mean

$$E(Y_1|Y_2 = y_2) = \frac{0 + y_2}{2} = \frac{y_2}{2};$$

i.e., the midpoint of the support. If you wanted to calculate this using the formula

$$E(Y_1|Y_2 = y_2) = \int_{\mathbb{R}} y_1 f_{Y_1|Y_2}(y_1|y_2)dy_1,$$

simply calculate

$$E(Y_1|Y_2 = y_2) = \int_{y_1=0}^{y_2} y_1 \, \frac{1}{y_2} \, dy_1 = \frac{1}{y_2}\left(\frac{y_1^2}{2}\bigg|_{y_1=0}^{y_2}\right) = \frac{y_2^2}{2y_2} = \frac{y_2}{2}.$$

(b) In Problem 5.77 on HW11, we calculated

$$E(Y_1) = \frac{1}{4}.$$

Let's see if we get the same answer for $E(Y_1)$ when we use the iterated rule for expectations; i.e.,

$$E(Y_1) = E[E(Y_1|Y_2)].$$

We know from part (a) that $Y_1|Y_2 = y_2 \sim \mathcal{U}(0, y_2)$, so

$$E(Y_1|Y_2 = y_2) = \frac{y_2}{2} \quad \Longrightarrow \quad E(Y_1|Y_2) = \frac{Y_2}{2}.$$

Therefore,

$$E(Y_1) = E[E(Y_1|Y_2)] = E\left(\frac{Y_2}{2}\right) = \frac{1}{2}E(Y_2).$$

Recall that $Y_2 \sim \text{beta}(2, 2)$, so $E(Y_2) = \frac{1}{2}$. Therefore,

$$E(Y_1) = \frac{1}{2}E(Y_2) = \frac{1}{2}\left(\frac{1}{2}\right) = \frac{1}{4},$$

which is the same answer we obtained when we calculated $E(Y_1)$ from the marginal distribution of Y_1.

5.136. This problem is set up as a hierarchical model; i.e.,

$$\begin{aligned} Y|\lambda &\sim \text{Poisson}(\lambda) \\ \lambda &\sim \text{exponential}(1). \end{aligned}$$

(a) We want to find $E(Y)$. We don't have the marginal distribution for Y; instead, we only have the conditional distribution of Y, given λ (in the first level of the hierarchy above). Therefore, let's use the law of iterated expectation; i.e.,

$$E(Y) = E[E(Y|\lambda)].$$

We know $Y|\lambda \sim \text{Poisson}(\lambda)$, so the conditional expectation $E(Y|\lambda) = \lambda$, the mean of a Poisson(λ) distribution. Therefore,

$$E(Y) = E[E(Y|\lambda)] = E(\lambda) = 1,$$

because $\lambda \sim \text{exponential}(1)$.

(b) To find $V(Y)$, we use Adam's Rule. We have

$$V(Y) = E[V(Y|\lambda)] + V[E(Y|\lambda)].$$

Now, $Y|\lambda \sim \text{Poisson}(\lambda)$, so $E(Y|\lambda) = \lambda$ and $V(Y|\lambda) = \lambda$; recall the Poisson mean and variance are equal. Therefore,

$$V(Y) = E(\lambda) + V(\lambda) = 1 + 1 = 2.$$

Recall that if $\lambda \sim \text{exponential}(1)$, then $E(\lambda) = 1$ and $V(\lambda) = 1$.

(c) We want $P(Y > 9)$, which is calculated by using the marginal distribution of Y. Note that Y (a count of the number of defects) is a discrete random variable. We aren't given the marginal pmf of Y, but we can get it using some creativity. We are given

$$\begin{aligned} Y|\lambda &\sim \text{Poisson}(\lambda) \\ \lambda &\sim \text{exponential}(1). \end{aligned}$$

Therefore, the joint distribution of Y and λ can be obtained by

$$f_{Y,\lambda}(y, \lambda) = f_{Y|\lambda}(y|\lambda) f_\lambda(\lambda),$$

which is nonzero as long as $y = 0, 1, 2, ...,$ and $\lambda > 0$. For these values,

$$f_{Y,\lambda}(y, \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \times e^{-\lambda} = \frac{\lambda^y e^{-2\lambda}}{y!}.$$

Summarizing, the joint distribution of Y and λ is described by

$$f_{Y,\lambda}(y, \lambda) = \begin{cases} \dfrac{\lambda^y e^{-2\lambda}}{y!}, & y = 0, 1, 2, ..., \text{and } \lambda > 0 \\ 0, & \text{otherwise,} \end{cases}$$

which, interestingly, is a **mixture** of discrete and continuous components. To find the marginal pmf of Y, we integrate $f_{Y,\lambda}(y, \lambda)$ over $\lambda > 0$. This is easy; note that

$$\begin{aligned} \int_{\lambda=0}^{\infty} f_{Y,\lambda}(y, \lambda) d\lambda = \int_{\lambda=0}^{\infty} \frac{\lambda^y e^{-2\lambda}}{y!} d\lambda &= \frac{1}{y!} \int_{\lambda=0}^{\infty} \lambda^{(y+1)-1} e^{-\lambda/(\frac{1}{2})} d\lambda \\ &= \frac{1}{y!} \times \Gamma(y+1) \left(\frac{1}{2}\right)^{y+1} = \left(\frac{1}{2}\right)^{y+1}, \end{aligned}$$

because $\Gamma(y + 1) = y!$. Summarizing, the marginal probability mass function of Y is

$$p_Y(y) = \begin{cases} \left(\dfrac{1}{2}\right)^{y+1}, & y = 0, 1, 2, ..., \\ 0, & \text{otherwise.} \end{cases}$$

We can now calculate $P(Y > 9)$ by using $p_Y(y)$. We have

$$\begin{aligned} P(Y > 9) = 1 - P(Y \leq 9) &= 1 - \sum_{y=0}^{9} \left(\frac{1}{2}\right)^{y+1} \\ &= 1 - \frac{1}{2} \sum_{y=0}^{9} \left(\frac{1}{2}\right)^{y} = 1 - \frac{1}{2}\left[\frac{1 - \left(\frac{1}{2}\right)^{10}}{1 - \frac{1}{2}}\right] = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}. \end{aligned}$$

So, no it is not likely that $Y > 9$.

5.137. In this problem, Y_1 is the weight of the item stocked by the supplier and Y_2 is is the weight of the amount sold during the week. We have the hierarchical model:

$$\begin{aligned} Y_2|Y_1 = y_1 &\sim \mathcal{U}(0, y_1) \\ Y_1 &\sim \mathcal{U}(0, 1). \end{aligned}$$

We are given that $y_1 = 3/4$, so

$$\begin{aligned} Y_2|Y_1 = y_1 &\sim \mathcal{U}(0, 3/4) \\ Y_1 &\sim \mathcal{U}(0, 1). \end{aligned}$$

We want to find $E(Y_2)$, the expected amount sold during the week. We are going to find this using the law of iterated expectation; i.e.,

$$E(Y_2) = E[E(Y_2|Y_1)].$$

Note that $E(Y_2|Y_1 = 3/4) = 3/8$, the midpoint of the (conditional) uniform distribution from 0 to 3/4. Therefore, $E(Y_2|Y_1) = 3/8$ (a constant) and thus

$$E(Y_2) = E\left(\frac{3}{8}\right) = \frac{3}{8}.$$

5.141. It looks like we have another hierarchical model; here,

$$\begin{aligned} Y_2|Y_1 = y_1 &\sim \mathcal{U}(0, y_1) \\ Y_1 &\sim \text{exponential}(\lambda). \end{aligned}$$

We want to find $E(Y_2)$ and $V(Y_2)$. We are going to use the iterated laws again. The conditional distribution $Y_2|Y_1 = y_1 \sim \mathcal{U}(0, y_1)$, so the conditional mean

$$E(Y_2|Y_1 = y_1) = \frac{y_1}{2},$$

the midpoint of the (conditional) uniform distribution from 0 to y_1. Therefore,

$$E(Y_2) = E[E(Y_2|Y_1)] = E\left(\frac{Y_1}{2}\right) = \frac{1}{2}E(Y_1) = \frac{\lambda}{2},$$

because $Y_1 \sim \text{exponential}(\lambda)$. To get $V(Y_2)$, we will use Adam's Rule. The conditional distribution $Y_2|Y_1 = y_1 \sim \mathcal{U}(0, y_1)$, so the conditional variance

$$V(Y_2|Y_1 = y_1) = \frac{y_1^2}{12},$$

the variance of a uniform distribution from 0 to y_1. Therefore,

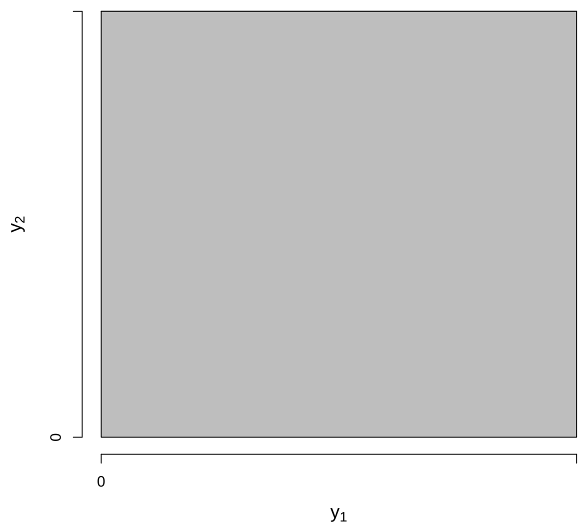
$$V(Y_2) = E[V(Y_2|Y_1)] + V[E(Y_2|Y_1)] = E\left(\frac{Y_1^2}{12}\right) + V\left(\frac{Y_1}{2}\right) = \frac{1}{12}E(Y_1^2) + \frac{1}{4}V(Y_1).$$

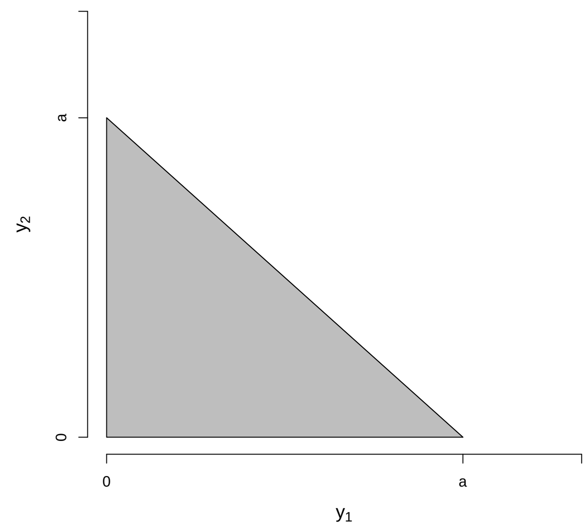
Now, $Y_1 \sim \text{exponential}(\lambda)$, so $V(Y_1) = \lambda^2$ and

$$E(Y_1^2) = V(Y_1) + [E(Y_1)]^2 = \lambda^2 + \lambda^2 = 2\lambda^2.$$

Therefore,

$$V(Y_2) = \frac{2\lambda^2}{12} + \frac{\lambda^2}{4} = \frac{5\lambda^2}{12}.$$





5.151. In this problem, we are given

$$\begin{aligned} Y_1 &\sim \text{exponential}(\beta) \\ Y_2 &\sim \text{exponential}(\beta) \end{aligned}$$

and Y_1 and Y_2 are independent. The joint pdf of $\mathbf{Y} = (Y_1, Y_2)$ is the product of the marginal pdfs (because of independence). Therefore,

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{Y_1}(y_1) f_{Y_2}(y_2) \\ &= \frac{1}{\beta} e^{-y_1/\beta} \times \frac{1}{\beta} e^{-y_2/\beta} \;=\; \frac{1}{\beta^2} e^{-(y_1+y_2)/\beta}. \end{aligned}$$

Summarizing,

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \dfrac{1}{\beta^2} e^{-(y_1+y_2)/\beta}, & y_1 \geq 0,\ y_2 \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

The support of Y_1 and Y_2 is $R = \{(y_1, y_2) : y_1 \geq 0, y_2 \geq 0\}$, the entire first quadrant. This set is shown at the top of this page (left). The joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ is a three-dimensional function which takes the value $(1/\beta^2)e^{-(y_1+y_2)/\beta}$ over this region (i.e., the entire first quadrant) and is otherwise equal to zero.

(b) We calculate $P(Y_1 + Y_2 \leq a)$ by integrating the joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ over the set

$$B = \{(y_1, y_2) : y_1 \geq 0, y_2 \geq 0,\ y_1 + y_2 \leq a\}.$$

This set is shown at the top of this page (right). The boundary of this set is

$$y_1 + y_2 = a \implies y_2 = a - y_1,$$

a linear function of y_1 with intercept $a > 0$ and slope -1. The limits on the double integral to calculate $P(Y_1 + Y_2 \leq a)$ come from this picture:

$$\begin{aligned}
P(Y_1 + Y_2 \leq a) &= \int_{y_1=0}^{a} \int_{y_2=0}^{a-y_1} \frac{1}{\beta^2} e^{-(y_1+y_2)/\beta} \, dy_2 dy_1 \\
&= \frac{1}{\beta^2} \int_{y_1=0}^{a} e^{-y_1/\beta} \left(\int_{y_2=0}^{a-y_1} e^{-y_2/\beta} \, dy_2 \right) dy_1 \\
&= \frac{1}{\beta^2} \int_{y_1=0}^{a} e^{-y_1/\beta} \left(-\beta e^{-y_2/\beta} \Big|_{y_2=0}^{a-y_1} \right) dy_1 \\
&= \frac{1}{\beta^2} \int_{y_1=0}^{a} e^{-y_1/\beta} \left(\beta - \beta e^{-(a-y_1)/\beta} \right) dy_1 \\
&= \frac{1}{\beta^2} \int_{y_1=0}^{a} \left(\beta e^{-y_1/\beta} - \beta e^{-a/\beta} \right) dy_1 \\
&= \frac{1}{\beta} \left(-\beta e^{-y_1/\beta} - y_1 e^{-a/\beta} \right) \Big|_{y_1=0}^{a} \\
&= \frac{1}{\beta} \left(-\beta e^{-a/\beta} - a e^{-a/\beta} + \beta + 0 \right) \; = \; 1 - e^{-a/\beta} - \frac{a}{\beta} e^{-a/\beta}.
\end{aligned}$$

Remark: Note that $P(Y_1 + Y_2 \leq a)$ is essentially the cdf of the "random variable" $Y_1 + Y_2$, evaluated at a. Interestingly, if you take a derivative of this expression above, you get

$$\frac{1}{\beta^2} a e^{-a/\beta} \quad (a > 0),$$

which, as a function of a, we recognize as a gamma$(2, \beta)$ pdf. Interesting!

5.164. Recall that for a univariate random variable X, the mgf of X is given by

$$m_X(t) = E(e^{tX}).$$

Finding $m_X(t)$ involves calculating a (single) sum or integral, depending on whether X is discrete or continuous, respectively. This question introduces you to joint moment generating functions. Suppose $\mathbf{X} = (X_1, X_2)$ is a bivariate random vector (discrete or continuous). The **joint moment generating function** (mgf) of X_1 and X_2 is

$$m_{X_1,X_2}(t_1, t_2) = E(e^{t_1 X_1 + t_2 X_2}).$$

For $m_{X_1,X_2}(t_1, t_2)$ to exist, this expectation must be finite in an open neighborhood about the origin $(0, 0)$. Look what happens when we put in $t_1 = 0$ or $t_2 = 0$ into the joint mgf:

$$\begin{aligned}
m_{X_1,X_2}(0, t_2) &= E(e^{t_2 X_2}) \; = \; m_{X_2}(t_2) \\
m_{X_1,X_2}(t_1, 0) &= E(e^{t_1 X_1}) \; = \; m_{X_1}(t_1).
\end{aligned}$$

Therefore, it is easy to get the marginal mgfs $m_{X_1}(t_1)$ and $m_{X_2}(t_2)$ from the joint mgf. Note that joint mgfs for random vectors in higher dimensions are defined in the same way. The joint mgf of $\mathbf{X} = (X_1, X_2, ..., X_n)$ is

$$m_{\mathbf{X}}(t_1, t_2, ..., t_n) = E\left(e^{t_1 X_1 + t_2 X_2 + \cdots + t_n X_n}\right).$$

Now, we get to the problem. For three random variables X_1, X_2, and X_3, the joint mgf of $\mathbf{X} = (X_1, X_2, X_3)$ is

$$m_{\mathbf{X}}(t_1, t_2, t_3) = E\left(e^{t_1X_1+t_2X_2+t_3X_3}\right).$$

(a) Put in $t_1 = t_2 = t_3 = t$ into $m_{\mathbf{X}}(t_1, t_2, t_3)$ and we get

$$m_{\mathbf{X}}(t, t, t) = E\left(e^{tX_1+tX_2+tX_3}\right) = E\left[e^{t(X_1+X_2+X_3)}\right],$$

which is the mgf of the "random variable" $X_1 + X_2 + X_3$.

(b) Put in $t_1 = t_2 = t$ and $t_3 = 0$ into $m_{\mathbf{X}}(t_1, t_2, t_3)$ and we get

$$m_{\mathbf{X}}(t, t, 0) = E\left(e^{tX_1+tX_2+0X_3}\right) = E\left[e^{t(X_1+X_2)}\right],$$

which is the mgf of the "random variable" $X_1 + X_2$.

(c) We don't have to do this problem in its full-blown generality (although it is not that hard). To get the main point, take $k_1 = k_2 = k_3 = 1$. That is, let's show

$$\left.\frac{\partial^3 m_{\mathbf{X}}(t_1, t_2, t_3)}{\partial t_1 \partial t_2 \partial t_3}\right|_{t_1=t_2=t_3=0} = E(X_1X_2X_3).$$

Proof. Note that

$$m_{\mathbf{X}}(t_1, t_2, t_3) = E\left(e^{t_1X_1+t_2X_2+t_3X_3}\right).$$

Therefore,

$$\frac{\partial^3 m_{\mathbf{X}}(t_1, t_2, t_3)}{\partial t_1 \partial t_2 \partial t_3} = \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} E\left(e^{t_1X_1+t_2X_2+t_3X_3}\right) = E\left(\frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} e^{t_1X_1+t_2X_2+t_3X_3}\right).$$

Interchanging the order of the derivative and expectation (i.e., triple sum or triple integral) is permitted as long as the mgf exists−recall that we had this same discussion with univariate mgfs; see pp 52 (notes). Now, note that

$$\begin{aligned}
\frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} e^{t_1X_1+t_2X_2+t_3X_3} &= \frac{\partial^2}{\partial t_2 \partial t_3} X_1 e^{t_1X_1} e^{t_2X_2+t_3X_3} \\
&= \frac{\partial}{\partial t_3} X_1 e^{t_1X_1} X_2 e^{t_2X_2} e^{t_3X_3} = X_1 e^{t_1X_1} X_2 e^{t_2X_2} X_3 e^{t_3X_3}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left.\frac{\partial^3 m_{\mathbf{X}}(t_1, t_2, t_3)}{\partial t_1 \partial t_2 \partial t_3}\right|_{t_1=t_2=t_3=0} &= \left.E\left(\frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} e^{t_1X_1+t_2X_2+t_3X_3}\right)\right|_{t_1=t_2=t_3=0} \\
&= \left.E\left(X_1 e^{t_1X_1} X_2 e^{t_2X_2} X_3 e^{t_3X_3}\right)\right|_{t_1=t_2=t_3=0} \\
&= E(X_1X_2X_3),
\end{aligned}$$

as claimed. Calculating higher-order mixed partial derivatives will establish the more general result for any k_1, k_2, and k_3.

5.165. This problem utilizes joint mgfs (discussed in Problem 5.164) to derive mathematical properties of the multinomial distribution. Suppose

$$\mathbf{X} = (X_1, X_2, X_3) \sim \text{mult}(n, \mathbf{p}; p_1 + p_2 + p_3 = 1),$$

where the category probabilities are in $\mathbf{p} = (p_1, p_2, p_3)$. Therefore, $\mathbf{X} = (X_1, X_2, X_3)$ has a **trinomial distribution** and the joint pmf is

$$p_{\mathbf{X}}(x_1, x_2, x_3) = \frac{n!}{x_1!x_2!x_3!}\; p_1^{x_1} x_2^{x_2} p_3^{x_3},$$

for values of x_1, x_2, x_3 in the support

$$R = \{(x_1, x_2, x_3) : x_j = 0, 1, 2, ..., n;\; x_1 + x_2 + x_3 = n\}.$$

(a) The joint mgf of $\mathbf{X} = (X_1, X_2, X_3)$ is

$$\begin{aligned} m_{\mathbf{X}}(t_1, t_2, t_3) &= E\left(e^{t_1 X_1 + t_2 X_2 + t_3 X_3}\right) \\ &= \sum_{(x_1,x_2,x_3)\in R} e^{t_1 x_1 + t_2 x_2 + t_3 x_3}\; \frac{n!}{x_1!x_2!\cdots x_n!}\; p_1^{x_1} p_2^{x_2} p_3^{x_3} \\ &= \sum_{(x_1,x_2,x_3)\in R} \frac{n!}{x_1!x_2!x_3!}\; (p_1 e^{t_1})^{x_1} (p_2 e^{t_2})^{x_2} (p_3 e^{t_3})^{x_3}, \end{aligned}$$

which we recognize as the multinomial expansion of $(p_1e^{t_1} + p_2e^{t_2} + p_3e^{t_3})^n$; see pp 20 (notes). Therefore, the mgf of $\mathbf{X} = (X_1, X_2, X_3)$ is

$$m_{\mathbf{X}}(t_1, t_2, t_3) = (p_1e^{t_1} + p_2e^{t_2} + p_3e^{t_3})^n.$$

(b) Take the joint mgf in part (a), and put in $t_1 = t$ and $t_2 = t_3 = 0$. The LHS is

$$m_{\mathbf{X}}(t, 0, 0) = E\left(e^{tX_1 + 0X_2 + 0X_3}\right) = E(e^{tX_1}), \quad \text{the marginal mgf of } X_1.$$

The RHS is

$$(p_1e^t + p_2e^0 + p_3e^0)^n = (p_1e^t + p_2 + p_3)^n = [(1 - p_1) + p_1e^t]^n,$$

because $p_1 + p_2 + p_3 = 1$. We recognize $[(1 - p_1) + p_1e^t]^n$ as the (marginal) mgf of a binomial distribution with number of trials n and probability of "success" p_1. Therefore, marginally, $X_1 \sim b(n, p_1)$.

(c) Recall the covariance computing formula

$$\text{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2).$$

We know $X_1 \sim b(n, p_1)$ and, analogously, $X_2 \sim b(n, p_2)$. Therefore, $E(X_1) = np_1$ and $E(X_2) = np_2$. From Problem 5.164 (c), we know we can find $E(X_1X_2)$ by calculating

$$\left.\frac{\partial^2 m_{\mathbf{X}}(t_1, t_2, t_3)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=t_3=0}.$$

Note that

$$\begin{aligned} \frac{\partial^2 m_{\mathbf{X}}(t_1, t_2, t_3)}{\partial t_1 \partial t_2} = \frac{\partial}{\partial t_2}\left[\frac{\partial m_{\mathbf{X}}(t_1, t_2, t_3)}{\partial t_1}\right] &= \frac{\partial}{\partial t_2}\left[\frac{\partial}{\partial t_1}(p_1e^{t_1} + p_2e^{t_2} + p_3e^{t_3})^n\right] \\ &= \frac{\partial}{\partial t_2}\left[n(p_1e^{t_1} + p_2e^{t_2} + p_3e^{t_3})^{n-1}\right] p_1e^{t_1} \\ &= n(n-1)(p_1e^{t_1} + p_2e^{t_2} + p_3e^{t_3})^{n-2} p_1e^{t_1} p_2e^{t_2}. \end{aligned}$$

Therefore,

$$\begin{aligned}
E(X_1X_2) = \left.\frac{\partial^2 m_{\mathbf{X}}(t_1,t_2,t_3)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=t_3=0} &= \left. n(n-1)(p_1e^{t_1}+p_2e^{t_2}+p_3e^{t_3})^{n-2}p_1e^{t_1}p_2e^{t_2}\right|_{t_1=t_2=t_3=0} \\
&= n(n-1)(p_1e^0+p_2e^0+p_3e^0)^{n-2}p_1e^0p_2e^0 \\
&= n(n-1)(p_1+p_2+p_3)^{n-2}p_1p_2 \\
&= n(n-1)p_1p_2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{Cov}(X_1,X_2) = E(X_1X_2) - E(X_1)E(X_2) &= n(n-1)p_1p_2 - np_1np_2 \\
&= n^2p_1p_2 - np_1p_2 - n^2p_1p_2 = -np_1p_2,
\end{aligned}$$

which agrees with how we derived this in class (i.e., by using our rules for covariances of linear combinations of random variables).