4.17 (a) \( P(Y = i + 1) = \int_{i}^{i+1} e^{-x} \, dx = (-e^{-x})_{i}^{i+1} = e^{-i}(1 - e^{-1}) \), which is geometric with \( p = 1 - e^{-1} \).

(b) Since \( Y = i + 1 \) if and only if \( i \leq X < i + 1 \), \( Y \geq 5 \iff X \geq 4 \).

\[
P(X - 4 \leq x | Y \geq 5) = P(X - 4 \leq x | X \geq 4)
\]

By the “memoryless” property (on page 101 of Casella & Berger) of the exponential distribution:

\[
P(X > s | X > t) = P(X > s - t),
\]

\[
P(X - 4 \leq x | X \geq 4) = P(X \leq x + 4 | X \geq 4) = 1 - P(X > x | X \geq 4)
\]

\[
= 1 - P(X > x + 4 - 4) = 1 - P(X > x) = P(X \leq x) = 1 - e^{-x}, \quad x \in \{0, 1, 2, \ldots\}
\]

4.20 (a) This transformation is not one-to-one because you cannot determine the sign of \( X_2 \) from \( Y_1 \) and \( Y_2 \). So partition the support of \((X_1, X_2)\) into \( S_X^0 = (-\infty < x_1 < \infty, x_2 = 0) \), \( S_X^1 = (-\infty < x_1 < \infty, x_2 > 0) \) and \( S_X^2 = (-\infty < x_1 < \infty, x_2 < 0) \). The support of \((Y_1, Y_2)\) is \( S_Y = (0 < y_1 < \infty, -1 < y_2 < 1) \). It is easy to see that \( P\{X \in S_X^0\} = 0 \). The inverse transformation from \( S_Y \) to \( S_X^1 \) is \( x_1 = y_2\sqrt{y_1} \) and \( x_2 = \sqrt{y_1 - y_1y_2^2} \) with Jacobian

\[
J_1(y_1, y_2) = \begin{vmatrix}
\frac{1}{2} \sqrt{1 - y_2^2} & \frac{\sqrt{y_1}}{2} \\
\frac{1}{2} \frac{y_2}{\sqrt{y_1}} & \frac{y_2\sqrt{y_1}}{2\sqrt{1 - y_2^2}}
\end{vmatrix} = \frac{1}{2\sqrt{1 - y_2^2}}
\]

The inverse transformation from \( S_Y \) to \( S_X^2 \) is \( x_1 = y_2\sqrt{y_1} \) and \( x_2 = -\sqrt{y_1 - y_1y_2^2} \) with \( J_2 = -J_1 \). Then

\[
f_{Y_1, Y_2}(y_1, y_2) = \frac{1}{2\pi\sigma^2} e^{-y_1/(2\sigma^2)} \left| \frac{1}{2\sqrt{1 - y_2^2}} \right| + \frac{1}{2\pi\sigma^2} e^{-y_1/(2\sigma^2)} \left| \frac{1}{2\sqrt{1 - y_2^2}} \right| = \frac{1}{2\pi\sigma^2} e^{-y_1/(2\sigma^2)} \frac{1}{\sqrt{1 - y_2^2}}
\]

(b) We can see in the above expression that the joint pdf factors into a function of \( y_1 \) and a function of \( y_2 \). So \( Y_1 \) and \( Y_2 \) are independent. \( Y_1 \) is the square of the distance from \((X_1, X_2)\) to the origin. \( Y_2 \) is the cosine of the angle between the positive \( x_1 \)-axis and the line from \((X_1, X_2)\) to the origin. So
Let independence says that the distance from the origin is independent of the orientation (as measured by the angle).

4.23 (a) \( X \sim \text{beta}(\alpha, \beta), Y \sim \text{beta}(\alpha + \beta, \gamma). \) \( X \) and \( Y \) are independent.

The inverse transformation is \( y = v, x = u/v \) with Jacobian

\[
J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{v} - \frac{u}{v^2} \\ 0 \end{vmatrix} = \frac{1}{v}
\]

\[
f_{U,V}(u, v) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta)\Gamma(\gamma)} \cdot \left(\frac{u}{v}\right)^{\alpha-1} (1 - \frac{u}{v})^{\beta-1} v^{\alpha+\beta-1} (1-v)^{\gamma-1} \frac{1}{v}, 0 < u < v < 1
\]

So

\[
f_U(u) = \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1} \int_u^1 v^{\beta-1} (1 - v)^{\gamma-1} (\frac{v-u}{v})^{\beta-1} dv
\]

\[
= \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} u^{\alpha-1} (1 - u)^{\beta+\gamma-1} \int_0^1 y^{\beta-1} (1 - y)^{\gamma-1} dy \left( y = \frac{v-u}{1-u}, dy = \frac{dv}{1-u} \right)
\]

\[
= \frac{\Gamma(\alpha + \beta + \gamma)}{\Gamma(\alpha)\Gamma(\beta + \gamma)} u^{\alpha-1} (1 - u)^{\beta+\gamma-1}
\]

Thus, \( U \sim \text{beta}(\alpha, \beta + \gamma). \)

4.24 Let \( z_1 = x + y, \ z_2 = \frac{x}{x+y}, \) then \( x = z_1 z_2, \ y = z_1(1 - z_2) \) and

\[
|J(z_1, z_2)| = \begin{vmatrix} \frac{\partial x}{\partial z_1} & \frac{\partial x}{\partial z_2} \\ \frac{\partial y}{\partial z_1} & \frac{\partial y}{\partial z_2} \end{vmatrix} = \begin{vmatrix} z_2 & z_1 \\ 1 - z_2 & -z_1 \end{vmatrix} = z_1
\]

The support \( S_{XY} = \{x > 0, y > 0\} \) is mapped onto the set \( S_Z = \{z_1 > 0, 0 < z_2 < 1\}. \) So

\[
f_{Z_1,Z_2}(z_1, z_2) = \frac{1}{\Gamma(r)} (z_1 z_2)^{r-1} e^{-z_1 z_2} \cdot \frac{1}{\Gamma(s)} (z_1 - z_1 z_2)^{s-1} e^{-z_1 z_1 z_2} \cdot z_1
\]

\[
= \frac{1}{\Gamma(r + s)} z_1^{r+s-1} e^{-z_1} \cdot \frac{\Gamma(r + s)}{\Gamma(r)\Gamma(s)} z_2^{r-1}(1 - z_2)^{s-1}, \quad 0 < z_1 < +\infty, 0 < z_2 < 1.
\]
Z₁ and Z₂ are independent because \( f_{Z₁,Z₂}(z₁, z₂) \) can be factored into two terms, one depending only on \( z₁ \) and the other depending only on \( z₂ \). Inspection of these kernels shows \( Z₁ \sim \text{gamma}(r + s, 1) \), \( Z₂ \sim \text{beta}(r, s) \).

4.26 (a) The only way to attack this problem is to find the joint cdf of \((Z, W)\). Now \( W = 0 \) or \( 1 \). For \( 0 \leq w < 1 \),

\[
P(Z \leq z, W \leq w) = P(Z \leq z, W = 0)
\]

\[
= P(\min(X, Y) \leq z, Y \leq X) = P(Y \leq z, Y \leq X)
\]

\[
= \int_{0}^{z} \int_{y}^{\infty} \frac{1}{\lambda} e^{-x/\lambda} \frac{1}{\mu} e^{-y/\mu} \, dx \, dy
\]

\[
= \frac{\lambda}{\lambda + \mu} \left( 1 - \exp\left(-\left(\frac{1}{\mu} + \frac{1}{\lambda}\right)z\right) \right)
\]

and

\[
P(Z \leq z, W = 1) = P(\min(X, Y) \leq z, X \leq Y)
\]

\[
= P(X \leq z, X \leq Y) = \frac{\mu}{\lambda + \mu} \left( 1 - \exp\left(-\left(\frac{1}{\mu} + \frac{1}{\lambda}\right)z\right) \right)
\]

can be used to give the joint cdf.

(b) \( P(W = 0) = P(Y \leq X) = \int_{0}^{\infty} \int_{y}^{\infty} \frac{1}{\lambda} e^{-x/\lambda} \frac{1}{\mu} e^{-y/\mu} \, dx \, dy = \frac{\lambda}{\mu + \lambda} \)

\[
P(W = 1) = 1 - P(W = 0) = \frac{\mu}{\mu + \lambda}
\]

\[
P(Z \leq z) = P(Z \leq z, W = 0) + P(Z \leq z, W = 1) = 1 - \exp\left(-\left(\frac{1}{\mu} + \frac{1}{\lambda}\right)z\right)
\]

Therefore, \( P(Z \leq z, W = i) = P(Z \leq z)P(W = i) \), for \( i = 0, 1, z > 0 \) which implies that \( Z \) and \( W \) are independent.

4.36 (a) \( EY = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} E(E(X_i|P_i)) = \sum_{i=1}^{n} E(P_i) = \sum_{i=1}^{n} \frac{\alpha}{\alpha + \beta} = \frac{n\alpha}{\alpha + \beta} \)
(b) Since the trials are independent,

\[ V_Y = \sum_{i=1}^{n} V X_i = \sum_{i=1}^{n} [V(E(X_i|P_i)) + E(V(X_i|P_i))] = \sum_{i=1}^{n} [V(P_i) + E(P_i(1 - P_i))] \]

\[ = \sum_{i=1}^{n} [V(P_i) - E((P_i)^2) + E(P_i)] = \sum_{i=1}^{n} [-(E(P_i))^2 + E(P_i)] \]

\[ = \sum_{i=1}^{n} \left[ -\left( \frac{\alpha}{\alpha + \beta} \right)^2 + \frac{\alpha}{\alpha + \beta} \right] = \sum_{i=1}^{n} \frac{\alpha \beta}{(\alpha + \beta)^2} = \frac{n \alpha \beta}{(\alpha + \beta)^2} \]

(c) We show that \( X_i \sim \text{Bernoulli}(\frac{\alpha}{\alpha + \beta}) \). Then, because \( X_i \)'s are independent, \( Y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, \frac{\alpha}{\alpha + \beta}) \). Let \( f_{P_i}(p) \) denote the beta(\( \alpha, \beta \)) density for the \( P_i \)'s and conditioning on \( P_i \) gives

\[
P(X_i = x) = \int_{0}^{1} P(X_i = x|P_i = p)f_{P_i}(p) \, dp = \int_{0}^{1} p^x(1 - p)^{1-x} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha} (1 - p)^{\beta - 1} \, dp
\]

\[ = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{0}^{1} p^{\alpha + x - 1} (1 - p)^{\beta - x + 1} \, dp \]

\[ = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + x)\Gamma(\beta - x + 1)}{\Gamma(\alpha + \beta + 1)} = \begin{cases} 
\frac{\beta}{\alpha + \beta} & x = 0 \\
\frac{\alpha}{\alpha + \beta} & x = 1 
\end{cases} \]

4.47 Because the sample space can be decomposed into the subsets \([XY > 0]\) and \([XY < 0]\) (and a set of \((X, Y)\) probability zero), we have for \( z < 0 \), by definition of \( Z \),

\[ P(Z \leq z) = P(X \leq z \text{ and } XY > 0) + P(-X \leq z \text{ and } XY < 0) \]

\[ = P(X \leq z \text{ and } Y < 0) + P(X \geq -z \text{ and } Y < 0) \text{ (because } z < 0) \]

\[ = P(X \leq z)P(Y < 0) + P(X \geq -z)P(Y < 0) \text{ (because } X \text{ and } Y \text{ are independent)} \]

\[ = P(X \leq z)P(Y < 0) + P(X \leq z)P(Y < 0) \text{ (symmetry of the distributions of } X \text{ and } Y \text{ )} \]

\[ = P(X \leq z) \]

By a similar argument, for \( z > 0 \), we get \( P(Z > z) = P(X > z) \), and hence, \( Z \sim X \sim N(0,1) \).

(b) By definition of \( Z \), \( Z > 0 \) \( \Leftrightarrow \) either (i) \( X < 0 \) and \( Y > 0 \) or (ii) \( X > 0 \) and \( Y > 0 \). So \( Z \) and \( Y \) always have the same sign, hence they cannot be bivariate normal.
These parts are exercises in the use of conditioning.

(a) 

\[ \text{Cov}(X,Y) = E(XY) - E(X)E(Y) \]
\[ = E[E(XY|X)] - E(X)E[E(Y|X)] \]
\[ = E[XE(Y|X)] - E(X)E[E(Y|X)] \]
\[ = \text{Cov}(X,E(Y|X)) \]

(b) 

\[ \text{Cov}(X,Y - E(Y|X)) = \text{Cov}(X,Y) - \text{Cov}(X,E(Y|X)) = 0 \quad \text{by (a)} \]

So, \( X \) and \( Y - E(Y|X) \) are uncorrelated.

(c) 

\[ \text{Var}(Y - E\{Y|X\}) = \text{Var}(E\{Y - E\{Y|X\}|X\}) + E\{\text{Var}[Y - E\{Y|X\}|X]\} \]
\[ = \text{Var}(E\{Y|X\} - E\{Y|X\}) + E\{\text{Var}(Y - E\{Y|X\}|X)\} \]
\[ = \text{Var}(0) + E\{\text{Var}(Y|X)\} \quad \text{E\{Y|X\} is constant wrt the [Y|X] distribution} \]
\[ = E\{\text{Var}(Y|X)\} \]