5.3.10. (a)

\[ M(t) = \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} \]
\[ = \frac{p}{1-p} \sum_{k=1}^{\infty} ((1-p)e^t)^k \]
\[ = \frac{p}{1-p} \left( \frac{1}{1 - (1-p)e^t} - 1 \right) \]
\[ = \frac{1}{1 - (1-p)e^t}, \text{ if } (1-p)e^t < 1. \]

(b)

\[ M(t) = \sum_{n=2}^{\infty} \left( \frac{n-1}{1} \right) e^{nt} p^2(1-p)^{n-2} \]
\[ = p^2e^{2t} \sum_{n=2}^{\infty} ((1-p)e^t)^{n-2} \]
\[ = p^2e^{2t} \cdot \frac{d}{du} \left[ \sum_{u=1}^{\infty} u^{n-1} \right]_{u=(1-p)e^t} \]
\[ = p^2e^{2t} \cdot \frac{d}{du} \left[ \frac{1}{1-u} \right]_{u=(1-p)e^t} \]
\[ = p^2e^{2t} \cdot \left[ \frac{1}{(1-u)^2} \right]_{u=(1-p)e^t} \]
\[ = \frac{p^2e^{2t}}{(1-(1-p)e^t)^2}, \text{ if } (1-p)e^t < 1. \]

(c) The time \( T \) until the second success is the sum of the time \( T_1 \) until the first success and the time \( T_2 \) strictly after the first success until the second success, and these two times are independent. Thus, \( M_T(t) = M_{T_1}(t) \cdot M_{T_2}(t) \). Both of the two m.g.f.'s on the right are as in part (a). Their product agrees with the result in part (b), thus \( T \) must have the desired negative binomial distribution.

5.4.4. The sample mean has the value \( \bar{X} = 288.1 \) for the \( n = 22 \) students in this data set. If the chance of observing a mean score this high is very small under the assumption that \( \mu = 256 \), then we conclude that \( \mu \) is probably higher than 256. We have

\[ P[\bar{X} \geq 288.1] = P[Z \geq \frac{288.1 - 256}{29/\sqrt{22}} = 5.19]. \]

This probability is nearly zero, so we can be very sure that \( \mu \) exceeds 256.

5.4.5. We require

\[ .95 = P[\bar{X} - c < \mu < \bar{X} + c] = P[-c < \bar{X} - \mu < c] = P\left[ \frac{-c}{\sigma/\sqrt{n}} < Z < \frac{c}{\sigma/\sqrt{n}} \right], \]

where \( Z \) is standard normal. From the normal table,

\[ 1.96 = \frac{c}{\sigma/\sqrt{n}} = \frac{c}{2/6} \Rightarrow c = \frac{1}{3}, 1.96 = .653. \]
(b) By the same reasoning as in part (a),

\[ M_Y(t) = \frac{1}{6} e^t + \frac{1}{6} e^{2t} + \frac{1}{6} e^{3t} + \frac{1}{3} e^{4t} + \frac{1}{12} e^{5t} + \frac{1}{12} e^{6t}. \]

(c) Again by the reasoning of (a),

\[ M(t) = \sum_{i=1}^{n} p_i e^{z_i t}. \]

(d) The m.g.f. of the sum is the product of the individual m.g.f.'s. Hence,

\[ M_{X+Y}(t) = M_X(t)M_Y(t) \]

\[ = \left( \frac{1}{12} e^{-2t} + \frac{1}{6} e^{-t} + \frac{1}{4} + \frac{1}{6} e^{t} + \frac{1}{6} e^{2t} + \frac{1}{12} e^{3t} \right) \]

\[ \times \left( \frac{1}{6} e^{4t} + \frac{1}{6} e^{3t} + \frac{1}{3} e^{2t} + \frac{1}{12} e^{t} + \frac{1}{12} e^{-t} \right) \]

\[ = \frac{1}{72} e^{-t} + \frac{1}{6} e^{0} + \frac{1}{12} e^{t} + \frac{1}{12} e^{2t} + \frac{1}{24} e^{3t} + \frac{25}{36} e^{4t} + \frac{23}{144} e^{5t} + \frac{1}{144} e^{6t} + \frac{9}{144} e^{7t} + \frac{3}{144} e^{8t} + \frac{1}{144} e^{9t}. \]

By part (c), the coefficients are the probabilities, and the coefficients of \( t \) in the exponents are the states they are associated with. Reducing the fractions, states \(-1, 0, 1, \ldots, 9\) respectively have probability masses \(1/72, 1/24, 1/12, 5/36, 25/144, 3/16, 23/144, 1/9, 1/16, 1/48, 1/144\).

5.4.2. If \( R \) denotes the length of the radial vector, then

\[ P[|R| \geq 2.14] = P[R^2 \geq 4.6] = P[X^2 + Y^2 \geq 4.6]. \]

But the random variable \( X^2 + Y^2 \) has the \( \chi^2(2) \) distribution, hence this probability is about \( .1 \).

5.4.8. The sample standard deviation is \( s = 5.17 \). The observed value of the random variable \( (n-1)s^2/\sigma^2 \) is \( 19 \cdot (5.17)^2/16 = 31.74 \) which is an observed value from the \( \chi^2(19) \) distribution if the true variance is \( 16 \). Since this number is so large (less than 5% likely), we have fairly strong evidence that the true variance is more than \( 16 \).

5.4.13. As in Exercise 12, since we can write the transformation as \( Y = AX \), where

\[ Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \]

we can conclude that \( Y \) is bivariate normal with the following mean vector and covariance matrix:

\[ \mu_Y = A\mu_X = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix} = \begin{bmatrix} -5 \\ -1 \end{bmatrix}, \]

\[ \Sigma_Y = A\Sigma_X A' = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 3 & 3 \end{bmatrix}. \]

Note that \( \rho = 3/(\sqrt{7} \cdot \sqrt{3}) = 3/\sqrt{21} \).