Exploring Data
In Engineering, the Sciences, and Medicine
Solutions Manual

Ronald K. Pearson

December 31, 2011

This manual presents detailed solutions to all of the problems in Exploring Data in Engineering, the Sciences and Medicine except those designated as computer exercises.

Chapter 3

Exercise 1:

Show that the first moment of the binomial distribution is \( m_1 = np \).

[Hint—the following result may be useful:

\[
k \binom{n}{k} = \frac{kn!}{k!(n-k)!} = \frac{n(n-1)!}{(k-1)!(n-1-[k-1])!} = n \binom{n-1}{k-1}.
\]

It is easily derived from the definition of the binomial coefficient and the properties of factorials.]

Solution:

1. First, write out the defining expression for \( m_1 \) and note that:

\[
m_1 = \sum_{k=0}^{n} kp_k = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}
\]

\[
= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k},
\]

since the term in the sum for \( k = 0 \) vanishes.

2. Next, use the hint to rewrite this result as:

\[
m_1 = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k}.
\]
3. To simplify, define \( j = k - 1 \) and \( m = n - 1 \) and rewrite the above expression as:

\[
m_1 = \sum_{j=0}^{m} n \binom{m}{j} p^{j+1}(1-p)^{m-j} = np \sum_{j=0}^{m} \binom{m}{j} p^{j}(1-p)^{m-j}
\]

4. It follows immediately from this result that \( m_1 = np \) since the sum in this expression is simply the normalization sum for the binomial distribution.

**Exercise 2:**

Show that the Poisson distribution satisfies the normalization requirement that \( \sum_{k=0}^{\infty} p_k = 1 \).

**Solution:**

1. First, substitute the definition for the Poisson probability \( p_k \) into the sum:

\[
\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}
\]

2. Second, note that the sum appearing above is the Taylor series representation for the exponential function, yielding:

\[
\sum_{k=0}^{\infty} p_k = e^{-\lambda} \cdot e^\lambda = 1
\]

**Exercise 3:**

Derive the moments \( m_1 \) and \( \mu_2 \) for the Poisson distribution.

**Solution:**

1. From the definition of \( m_1 \), for the Poisson distribution we have:

\[
m_1 = \sum_{k=0}^{\infty} kp_k = \sum_{k=0}^{\infty} ke^{-\lambda} \frac{\lambda^k}{k!}
\]

\[
= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}
\]

\[
= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^{j}}{j!}
\]

\[
= \lambda e^{-\lambda} \lambda
\]

\[
= \lambda
\]
2. The easiest way to compute the central moment $\mu_2$ is to recall that $\mu_2 = m_2 - m_1^2$ where:

$$m_2 = \sum_{k=0}^{\infty} k^2 p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{k^2 \lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{k \lambda^k}{(k-1)!}$$

3. As before, write $j = k - 1$, converting this sum to:

$$m_2 = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(j+1) \lambda^{j+1}}{j!}$$

$$= \lambda e^{-\lambda} \left[ \sum_{j=0}^{\infty} \frac{j \lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \right]$$

$$= \lambda e^{-\lambda} \left[ \lambda \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} + e^\lambda \right]$$

$$= \lambda e^{-\lambda} \left[ \lambda \sum_{\ell=0}^{\infty} \frac{\lambda^{\ell}}{\ell!} + e^\lambda \right]$$

$$= \lambda e^{-\lambda} [\lambda e^\lambda + e^\lambda]$$

$$= \lambda^2 + \lambda$$

4. Combining these results yields:

$$\mu_2 = m_2 - m_1^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

**Exercise 4:**

Derive the mean and variance for the discrete uniform distribution.

**Solution:**

1. For the discrete uniform distribution with $M > 1$ levels, $p_k = 1/M$ for all $k$, and the mean $m_1$ is given by:

$$m_1 = \sum_{k=1}^{M} k p_k = \frac{1}{M} \sum_{k=1}^{M} k = \frac{1}{M} \left[ \frac{M(M+1)}{2} \right] = \frac{M+1}{2}$$

2. To compute the variance $\mu_2$, again use the fact that $\mu_2 = m_2 - m_1^2$. It follows from Eq. (3.21) that $m_2$ is given by:

$$m_2 = \sum_{k=1}^{M} k^2 p_k = \frac{1}{M} \sum_{k=1}^{M} k^2 = \frac{1}{M} \left[ \frac{M(M+1)(2M+1)}{6} \right] = \frac{(M+1)(2M+1)}{6}$$
3. Combining these results and simplifying the algebra yields:

\[
\mu_2 = \frac{(M+1)(2M+1)}{6} - \left[ \frac{M+1}{2} \right]^2
\]
\[
= \frac{2M^2 + 3M + 1}{6} - \frac{M^2 + 2M + 1}{4}
\]
\[
= \frac{1}{2} \left[ \frac{2(2M^2 + 3M + 1) - 3(M^2 + 2M + 1)}{6} \right]
\]
\[
= \frac{1}{2} \left[ \frac{4M^2 + 6M + 2 - 3M^2 - 6M - 3}{6} \right]
\]
\[
= \frac{M^2 - 1}{12}
\]

Exercise 5: NOTE that there is an error in the original problem statement. This error has been corrected here.

The standard estimator for the variance \( \sigma_x^2 \) of a sequence \( \{x_k\} \) of \( N \) real numbers is given in Chapter 1 and discussed further in Chapter 6. It is:

\[
\hat{\sigma}^2 = \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \bar{x})^2,
\]

where \( \bar{x} \) is the mean of the sequence. Show that the Simpson heterogeneity measure for a normalized sequence \( \{p_i\} \) of \( M \) probabilities is exactly equal to \( M\hat{\sigma}_p^2 \).

Solution:

1. First, note that for a normalized sequence of probabilities, we have:

\[
\sum_{i=1}^{M} p_i = 1 \Rightarrow \bar{p} = \frac{1}{M} \sum_{i=1}^{M} p_i = \frac{1}{M}
\]

2. Thus, the estimated variance is:

\[
\sigma_p^2 = \frac{1}{M-1} \sum_{i=1}^{M} (p_i - \bar{p})^2
\]
\[
= \frac{1}{M-1} \sum_{i=1}^{M} (p_i - 1/M)^2
\]
\[
= \frac{1}{M-1} \sum_{i=1}^{M} (p_i^2 - 2p_i/M + 1/M^2)
\]
\[
= \frac{1}{M-1} \left[ \sum_{i=1}^{M} p_i^2 - \frac{2}{M} \sum_{i=1}^{M} p_i + M \cdot \frac{1}{M^2} \right]
\]
\[ I_{\text{Simpson}} = \frac{M \sum_{i=1}^{M} p_i^2 - 1}{M - 1} = M \sigma_p^2 \]

Chapter 4

Exercise 1—NOTE: there is an error in the original problem statement. The term \( r \) in the original statement should be \( n \).

Prove the relationship between the central moments \( \mu_n \) and the non-central moments \( m_r \) given in Eq. (4.15).

Solution:

1. From the definition of central moments, \( \mu_n = E\{(x - m_1)^n\} \). By the binomial expansion—Eq. (3.12) on page 118—we have:
   \[
   (x - m_1)^n = (x + [-m_1])^n = \sum_{j=0}^{n} \binom{n}{j} x^j [-m_1]^{n-j}.
   \]

2. Since the expectation operator is linear, taking the expectation with respect to the random variable \( x \) leads to the final result:
   \[
   \mu_n = E \left\{ \sum_{j=0}^{n} \binom{n}{j} x^j [-m_1]^{n-j} \right\} = \sum_{j=0}^{n} \binom{n}{j} E\{x^j\} [-m_1]^{n-j}
   \]
   \[
   = \sum_{j=0}^{n} \binom{n}{j} m_j [-m_1]^{n-j}
   \]

Exercise 2:

Suppose the density \( p(x) \) is symmetric about \( x_0 \). Prove that the mean satisfies \( \bar{x} = x_0 \).

Solution:

1. The mean is given by:
   \[
   \bar{x} = E\{x\} = \int_{-\infty}^{\infty} xp(x)dx
   \]
2. Let \( z = x - x_0 \) so that \( x = x_0 + z \), giving:

\[
\bar{x} = \int_{-\infty}^{\infty} (x_0 + z)p(x_0 + z)dz
\]

\[
= x_0\int_{-\infty}^{\infty} p(x_0 + z)dz + \int_{-\infty}^{\infty} zp(x_0 + z)dz
\]

\[
\equiv x_0\int_{-\infty}^{\infty} g(z)dz + \int_{-\infty}^{\infty} zg(z)dz,
\]

where \( g(z) = p(x_0 + z) \) is simply a shifted version of \( p(x) \).

3. Note that the first of these integrals is equal to 1 by the normalization condition for \( g(z) \), while the second integral is the mean of the shifted distribution \( g(z) \). For notational convenience, denote this second integral by \( I \).

4. Since \( p(x) \) is symmetric about \( x_0 \), it follows that \( g(z) \) is symmetric about zero:

\[
g(z) = p(x_0 + z) = p(x_0 - [-z]) = p(-z - x_0) = p(-[x_0 + z]) = g(-z).
\]

5. Thus, the integral \( I \) is:

\[
I = \int_{-\infty}^{\infty} zg(z)dz = \int_{-\infty}^{\infty} zg(-z)dz
\]

6. Let \( w = -z \) and rewrite this integral as:

\[
I = \int_{-\infty}^{\infty} (-w)g(w)d(-w)
\]

\[
= \int_{-\infty}^{\infty} wg(w)dw
\]

\[
= -\int_{-\infty}^{\infty} wg(w)dw = -I.
\]

7. Thus, it follows that \( I = 0 \) and that \( \bar{x} = x_0 \) as claimed.

Exercise 3:

Suppose the density \( p(x) \) is symmetric about \( x_0 \). Prove that the median satisfies \( x^\dagger = x_0 \).

Solution:
1. As in Exercise 2, define \( z = x - x_0 \) and \( g(z) = p(x_0 + z) \), and consider the integral:

\[
I = \int_{-\infty}^{x_0} p(x) \, dx \\
= \int_{-\infty}^{0} p(x_0 + z) \, dz \\
= \int_{-\infty}^{0} g(z) \, dz
\]

2. Note that

\[
1 - I = \int_{-\infty}^{\infty} g(z) \, dz - \int_{-\infty}^{0} g(z) \, dz \\
= \int_{0}^{\infty} g(z) \, dz \\
= \int_{0}^{\infty} g(-z) \, dz \quad \text{(by the symmetry of } p(x) \text{ about } x_0) \\
= \int_{-\infty}^{0} g(w) \, d(-w) \quad \text{(substituting } w = -z) \\
= -\int_{-\infty}^{0} g(w) \, dw \\
= \int_{-\infty}^{0} g(w) \, dw = I
\]

3. It follows from this result that \( 1 - I = I \), implying \( I = 1/2 \). Consequently, we have:

\[
\int_{-\infty}^{x_0} p(x) \, dx = \frac{1}{2},
\]

which is the defining condition for the median \( x^\dagger \).

**Exercise 4:**

Suppose the density \( p(x) \) is unimodal and symmetric about \( x_0 \). Prove that the mode satisfies \( x^* = x_0 \).

**Solution:**

1. As in the previous two problems, define \( z = x - x_0 \) and \( g(z) = p(x_0 + z) \), from which it follows that \( g(-z) = g(z) \) (see the solution to Exercise 2). If \( p(x) \) is unimodal with a mode at \( x^* \), it follows that \( g(z) \) is also unimodal, with a mode at \( z^* = x^* - x_0 \).

2. It follows from the above definitions that \( g(-z^*) = g(z^*) \). Since \( g(z) \) is unimodal with a mode at \( z^* \), this implies \( -z^* = z^* = 0 \), which further implies \( x^* = x_0 \).
Exercise 5:

Consider the Pareto distribution:

\[ p(x) = \frac{ak^a}{x^{a+1}}, \quad x \geq k > 0, a > 0. \]

For this distribution, compute:

a. the mean \( \bar{x} \),

b. the median \( x^\dagger \),

c. the mode \( x^* \).

Solution:

1. The mean of this distribution is:

\[
\bar{x} = E\{x\} = \int_k^\infty xp(x)dx \\
= \int_k^\infty x \left(\frac{ak^a}{x^{a+1}}\right) dx \\
= ak^a \int_k^\infty x^{-a}dx \\
= ak^a \left[\frac{x^{-a+1}}{-a+1}\right]_k^\infty \\
= 0 - \left(\frac{ak^a}{-a+1}\right) (\text{provided } a > 1) \\
= \frac{ak}{a-1}
\]

2. The median is the value \( x^\dagger \) for which:

\[
\int_k^{x^\dagger} p(x)dx = \frac{1}{2}
\]

Integrating the density yields:

\[
\int_k^{x^\dagger} p(x)dx = \int_k^{x^\dagger} \frac{ak^a}{x^{a+1}} dx \\
= (ak^a) \int_k^{x^\dagger} x^{-a-1}dx \\
= \left[\frac{ak^a x^{-a}}{-a}\right]_k^{x^\dagger} \\
= -\left(\frac{k}{x^\dagger}\right)^{a} k^{a} + \left(\frac{k}{x^\dagger}\right)^{a} \\
= 1 - \left(\frac{k}{x^\dagger}\right)^{a}
\]
Solving for the median yields:

\[
1 - \left( \frac{k}{x^a} \right)^a = \frac{1}{2} \quad \Rightarrow \quad \left( \frac{k}{x^a} \right)^a = \frac{1}{2}
\]

\[
\Rightarrow \quad \left( \frac{x^a}{k} \right)^a = 2
\]

\[
\Rightarrow \quad \frac{x^a}{k} = 2^{\frac{1}{a}}
\]

\[
\Rightarrow \quad x = 2^{\frac{1}{a}} k.
\]

3. The mode is the value for which the density is a maximum. Plotting the Pareto distribution for any values of \(a\) and \(k\) shows that it is J-shaped, decaying monotonically from its mode at \(x^* = k\). To prove this result, note that the derivative of the density with respect to \(x\) is:

\[
\frac{dp(x)}{dx} = \frac{d}{dx} \left( \frac{ak^a}{x^{a+1}} \right)
\]

\[
= ak^a \left[ -(a + 1)x^{-a-2} \right]
\]

\[
= -a(a + 1)k^a \frac{x^{-a-2}}{x^{a+2}}.
\]

Since \(a\), \(k\), and \(x\) are all positive numbers, it follows that this derivative is negative for all \(x\), implying that the function is nonincreasing. Thus, the mode must occur at the minimum possible value for \(x\), at \(x^* = k\).

Exercise 6:

Consider the Laplace distribution:

\[
p(x) = \frac{1}{2\phi} e^{-|x-\mu|/\phi}, \quad \phi > 0.
\]

a. For this distribution, compute the exact probability \(P\{|x-\mu| > c\phi\}\):

b. Construct either a plot or a small table, comparing these exact values with those given by Chebyshev’s inequality.

Solution:

1. Since this distribution is symmetric about \(\mu\), it follows that:

\[
P\{|x-\mu| > c\phi\} = 2P\{x-\mu > c\phi\} = 2P\{x > \mu + c\phi\}
\]

\[
= 2 \int_{\mu+c\phi}^{\infty} p(x)dx
\]

\[
= \frac{1}{\phi} \int_{\mu+c\phi}^{\infty} e^{-(x-\mu)/\phi}dx
\]

\[
= \frac{1}{\phi} \int_{\mu+c\phi}^{\infty} e^{-(x-\mu)/\phi}dx
\]
Figure 1: Comparison of exact probability that $|x - \mu| > c\phi$ (solid line) with
the corresponding Chebyshev bound (dashed line).

2. To simplify the subsequent algebra, define $z = (x - \mu)/\phi$ and note that:

$$
\mathcal{P}\{|x - \mu| > c\phi\} = \int_{c}^{\infty} e^{-z}dz
= -e^{-z} \bigg|_{c}^{\infty}
= -[0 - e^{-c}] = e^{-c}.
$$

3. The plot in Fig. 1 compares these exact values (solid line) with the Chebyshev bounds (dashed line).
Chapter 5

Exercise 1:

Derive the expressions given in Eq. (5.79) for the maximum likelihood estimates \( \hat{\sigma}^2 \) and \( \hat{\phi} \) of the Gaussian and Laplace scale parameters.

Solution:

1. For the Gaussian case, the likelihood as a function of \( a \), \( b \), and \( \sigma^2 \) may be written as the following slight modification of Eq. (5.71):

\[
\ell(a, b, \sigma^2) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^{N} (y_k - ax_k - b)^2
\]

2. Necessary conditions for minimizing this function with respect to these three parameters are:

\[
\frac{\partial \ell(a, b, \sigma^2)}{\partial a} = \frac{1}{\sigma^2} \sum_{k=1}^{N} (y_k - ax_k - b)x_k = 0
\]

\[
\frac{\partial \ell(a, b, \sigma^2)}{\partial b} = \frac{1}{\sigma^2} \sum_{k=1}^{N} (y_k - ax_k - b) = 0
\]

\[
\frac{\partial \ell(a, b, \sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{k=1}^{N} (y_k - ax_k - b)^2 = 0.
\]

3. The first two of these conditions lead to the minimization problem for fixed \( \sigma \) given in Eq. (5.72). Since these ML estimates do not depend on \( \sigma^2 \), we can fix these coefficients at their estimated values \( \hat{a} \) and \( \hat{b} \). Substituting these into the third equation above and solving yields the desired result:

\[
-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{k=1}^{N} (y_k - \hat{ax}_k - \hat{b})^2 = 0
\]

\[
\Rightarrow \frac{N}{2\sigma^2} = \frac{1}{2\sigma^4} \sum_{k=1}^{N} (y_k - \hat{ax}_k - \hat{b})^2
\]

\[
\Rightarrow \hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^{N} (y_k - \hat{ax}_k - \hat{b})^2
\]

4. For the Laplace distribution, the likelihood is given by Eq. (5.74) as:

\[
\ell(a, b, \phi) = -N \ln 2 - N \ln \phi - \frac{1}{\phi} \sum_{k=1}^{N} |y_k - ax_k - b|
\]
Here, because $\ell(a, b, \phi)$ is not a smooth function of $a$ and $b$, the ML estimate is not obtained by setting partial derivatives to zero. As in the Gaussian case just considered, however, the ML estimates $\hat{a}$ and $\hat{b}$ do not depend on the scale parameter $\phi$ (i.e., they correspond to the solution of the optimization problem defined by Eq. (5.75)), so we can fix these parameters at their ML estimates and solve for $\hat{\phi}$. Since the dependence of $\ell(a, b, \phi)$ on $\phi$ is smooth, we can differentiate $\ell(\hat{a}, \hat{b}, \phi)$ with respect to $\phi$ to obtain Eq. (5.79) for the ML estimate $\hat{\phi}$:

$$\frac{\partial \ell(\hat{a}, \hat{b}, \phi)}{\partial \phi} = -\frac{N}{\phi} + \frac{1}{\phi^2} \sum_{k=1}^{N} |y_k - \hat{a}x_k - \hat{b}| = 0$$

$$\Rightarrow \frac{N}{\phi} = \frac{1}{\phi^2} \sum_{k=1}^{N} |y_k - \hat{a}x_k - \hat{b}|$$

$$\Rightarrow \hat{\phi} = \frac{1}{N} \sum_{k=1}^{N} |y_k - \hat{a}x_k - \hat{b}|$$

**Exercise 2:**

Derive Eqs. (5.53), (5.54), and (5.55) for the TLS line-fitting solution.

**Solution:**

1. For the TLS case, the distance to be minimized is:

$$J_{TLS}(a, b, \hat{x}(k)) = \sum_{k=1}^{N} \{|y(k) - a\hat{x}(k) - b|^2 + |\hat{x}(k) - x(k)|^2\}$$

To determine $\hat{x}(k)$, set the corresponding partial derivative to zero:

$$\frac{\partial J_{TLS}(a, b, \hat{x}(k))}{\partial \hat{x}(k)} = -2a[y(k) - a\hat{x}(k) - b] + 2[\hat{x}(k) - x(k)] = 0$$

This leads to the desired result (Eq. (5.53)):

$$2a[y(k) - b] + 2x(k) = 2(a^2 + 1)\hat{x}(k)$$

$$\Rightarrow \hat{x}(k) = \frac{x(k) + a[y(k) - b]}{1 + a^2}$$

2. From Eq. (5.30) we obtain Eq. (5.54) as:

$$\hat{y}(k) = a\hat{x}(k) + b$$

$$= a \left[ \frac{x(k) + a[y(k) - b]}{1 + a^2} \right] + b$$

$$= \frac{ax(k) + a^2y(k) - a^2b + (1 + a^2)b}{1 + a^2}$$

$$= \frac{a^2y(k) + ax(k) + b}{1 + a^2}$$
3. Combining these results yields Eq. (5.55):

\[
\rho(v_k, v_k) = \left[ \hat{y}(k) - y(k) \right]^2 + \left[ \hat{z}(k) - x(k) \right]^2
\]

\[
= \left[ y(k) - \frac{a^2 y(k) + ax(k) + b}{1 + a^2} \right]^2 + \left[ x(k) - \frac{x(k) - a[y(k) - b]}{1 + a^2} \right]^2
\]

\[
= (1 + a^2)^{-2} \left\{ [(1 + a^2)y(k) - a^2y(k) - ax(k) - b]^2
\right. \\
+ \left. [(1 + a^2)x(k) - x(k) - a(y(k) - b)]^2 \right\}
\]

\[
= (1 + a^2)^{-2} \left\{ [y(k) - ax(k) - b]^2 + [a^2 x(k) - ay(k) + ab]^2 \right\}
\]

\[
= (1 + a^2)^{-2} \left\{ [y(k) - ax(k) - b]^2 + [-a(y(k) - ax(k) - b)]^2 \right\}
\]

\[
= (1 + a^2)^{-2} \left\{ [y(k) - ax(k) - b]^2 + a^2[y(k) - ax(k) - b]^2 \right\}
\]

\[
= (y(k) - ax(k) - b)^2
\]

Exercise 3:

Prove that the second derivative of \( J_{TLS}(a, b) \) with respect to \( a \) at \( a = a_{TLS} \) is given by Eq. (5.63).

Solution:

1. From Eq. (5.60), we have:

\[
J_{TLS}(a, b_{TLS}) = N \left[ \hat{\sigma}_y^2 - 2a\hat{c}_{xy} + a^2\hat{\sigma}_x^2 \right] \frac{1}{1 + a^2}
\]

Defining \( \lambda = (\hat{\sigma}_y^2 - \hat{\sigma}_x^2)/\hat{c}_{xy} \), note that \( \hat{\sigma}_y^2 = \hat{\sigma}_x^2 - \lambda\hat{c}_{xy} \). Substituting this result into the above expression yields:

\[
J_{TLS}(a, b_{TLS}) = N \left[ \hat{\sigma}_x^2(1 + a^2) - \hat{c}_{xy}(\lambda + 2a) \right] \frac{1}{1 + a^2}
\]

\[
= N\hat{\sigma}_x^2 \frac{N\hat{c}_{xy}(\lambda + 2a)}{1 + a^2}
\]

2. The first derivative of this expression with respect to \( a \) is:

\[
\frac{\partial J_{TLS}(a, b_{TLS})}{\partial a} = -N\hat{c}_{xy} \frac{\partial}{\partial a} \left[ \frac{\lambda + 2a}{1 + a^2} \right]
\]

\[
= -N\hat{c}_{xy} \left[ \frac{(1 + a^2) \cdot 2 - (\lambda + 2a) \cdot 2a}{(1 + a^2)^2} \right]
\]

\[
= \frac{2N\hat{c}_{xy}}{(1 + a^2)^2} \left[ a^2 + a\lambda - 1 \right]
\]

13
Note that setting this derivative to zero yields Eq. (5.61) for the TLS estimator $a_{TLS}$.

3. Differentiating again yields:

\[
\frac{\partial^2 J_{TLS}(a, b_{TLS})}{\partial a^2} = 2N\hat{c}_{xy} \frac{\partial}{\partial a} \left[ \frac{a^2 + a\lambda - 1}{(1 + a^2)^2} \right]
\]

\[= \frac{2N\hat{c}_{xy}}{(1 + a^2)^3} \left[ (1 + a^2)(2a + \lambda) - 4a(2a^2 + a\lambda - 1) \right] \]

\[= \frac{2N\hat{c}_{xy}}{(1 + a^2)^3} \left[ 2a + \lambda + 2a^3 + a^2\lambda - 4a^3 - 4a^2\lambda + 4a \right] \]

\[= \frac{2N\hat{c}_{xy}}{(1 + a^2)^3} \left[ 6a + \lambda - 2a^3 - 3a^2\lambda \right] \]

4. To evaluate this result for $a = a_{TLS}$, note that this quantity is defined by Eq. (5.61):

\[a_{TLS}^2 + \lambda a_{TLS} - 1 = 0 \Rightarrow a_{TLS}^2 = 1 - \lambda a_{TLS} \]

Let $T$ denote the term in brackets in the second derivative expression, evaluated at $a_{TLS}$:

\[T = 6a_{TLS} + \lambda - 2a_{TLS}^3 - 3a_{TLS}^2\lambda \]

\[= 6a_{TLS} + \lambda - 2a_{TLS}(1 - \lambda a_{TLS}) - 3(1 - \lambda a_{TLS})\lambda \]

\[= 6a_{TLS} + \lambda - 2a_{TLS} + 2\lambda a_{TLS}^2 - 3\lambda + 3\lambda a_{TLS} \]

\[= 4a_{TLS} - 2\lambda + 2\lambda(1 - \lambda a_{TLS}) + 3\lambda^2 a_{TLS} \]

\[= 4a_{TLS} - 2\lambda + 2\lambda - 2\lambda^2 a_{TLS} + 3\lambda^2 a_{TLS} \]

\[= 4a_{TLS} + \lambda^2 a_{TLS} \]

\[= (\lambda^2 + 4)a_{TLS} \]

5. From Eq. (5.64), we have $\lambda^2 + 4 = (\lambda + 2a_{TLS})^2$. Substituting this expression into the second derivative result yields:

\[
\frac{\partial^2 J_{TLS}(a, b_{TLS})}{\partial a^2} \bigg|_{a=a_{TLS}} = \frac{2N\hat{c}_{xy}}{(1 + a_{TLS}^2)^3} (\lambda + 2a_{TLS})^2 a_{TLS} \]

To simplify this result to the desired one, note that:

\[(\lambda + 2a_{TLS})a_{TLS} = \lambda a_{TLS} + 2a_{TLS}^2 \]

\[= \lambda a_{TLS} + 2(1 - \lambda a_{TLS}) \]

\[= 2 - \lambda a_{TLS} \]

\[= 1 + a_{TLS}^2 \]
Thus, the above result reduces to Eq. (5.63):

\[
\frac{\partial^2 J_{TLS}(a, b_{TLS})}{\partial a^2} \bigg|_{a=a_{TLS}} = \frac{2N\hat{c}_{xy}(\lambda + 2a_{TLS})(1 + a_{TLS}^2)}{(1 + a_{TLS}^2)^3} = \frac{2N\hat{c}_{xy}(\lambda + 2a_{TLS})}{(1 + a_{TLS}^2)^2}
\]

Exercise 4:

In 1923, Subotin proposed the following family of probability distributions:

\[
p(x) = \frac{1}{C\phi} \exp \left[ -\frac{1}{2} \left| \frac{x - \theta}{\phi} \right|^{2/\delta} \right],
\]

\[
C = 2^{(\delta/2)+1}\Gamma\left(\frac{\delta}{2} + 1\right),
\]

where \(\theta\) is an arbitrary real constant and \(\phi\) and \(\delta\) are positive constants. Suppose the line fit errors \(\{e_k\}\) form a statistically independent, identically distributed, zero-mean sequence corresponding to this distribution with \(\theta = 0\) and \(\delta > 0\) fixed and known. Reduce the maximum likelihood estimation problem to an optimization problem. What norm of \(\{e_k\}\) is being minimized?

Solution:

1. The likelihood function for this problem is defined by Eq. (5.69):

\[
\ell(a, b) = \sum_{k=1}^{N} \ln p(e_k)
\]

where \(e_k = ax_k + b - y_k\).

2. For the distribution considered here—with \(\theta = 0\) and \(\delta\) known and fixed—we have:

\[
\ln p(e_k) = -\ln C(\delta) - \ln \phi - \frac{1}{2} \left| \frac{e_k}{\phi} \right|^{2/\delta}
\]

Thus, since \(\phi > 0\), the log likelihood can be expressed as:

\[
\ell(a, b) = -N \ln C(\delta) - N \ln \phi - \frac{1}{2\phi^{2/\delta}} \sum_{k=1}^{N} |y_k - ax_k - b|^{2/\delta}
\]

3. Since only the last term in this expression depends on the parameters \(a\) and \(b\), the maximum likelihood estimation problem is equivalent to the minimization of:

\[
J(a, b) = \sum_{k=1}^{N} |y_k - ax_k - b|^{2/\delta}
\]

Note that this is a minimization of the \(\ell_p\) norm of the model errors for \(p = 2/\delta\).
Chapter 6
Exercise 1:

Suppose \( \{ x_k \} \) is a sequence of \( N \) independent, identically distributed random variables with mean \( \mu \) and variance \( \sigma^2 \). If \( \mu \) and \( \sigma^2 \) are finite, show that the following estimator of \( \sigma^2 \) is unbiased:

\[
\hat{\sigma}^2 = \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \bar{x}_N)^2.
\]

Solution:

1. To show that the estimator is unbiased, we need to show that \( E\{\hat{\sigma}^2\} = \sigma^2 \).

To do this, first note that:

\[
E\{\hat{\sigma}^2\} = E \left\{ \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \bar{x})^2 \right\}
= \frac{1}{N-1} \sum_{k=1}^{N} E\{(x_k - \bar{x})^2\}
= \left( \frac{N}{N-1} \right) E\{(x_k - \bar{x})^2\}
\]

2. Next, note that:

\[
(x_k - \bar{x})^2 = [(x_k - \mu) - (\bar{x} - \mu)]^2
= (x_k - \mu)^2 - 2(x_k - \mu)(\bar{x} - \mu) + (\bar{x} - \mu)^2
\Rightarrow E\{(x_k - \bar{x})^2\} = \sigma^2 - 2E\{(x_k - \mu)(\bar{x} - \mu)\} + E\{(\bar{x} - \mu)^2\}
\]

3. Expanding the first of these expectations gives:

\[
E\{(x_k - \mu)(\bar{x} - \mu)\} = E \left\{ (x_k - \mu) \left( \frac{1}{N} \sum_{j=1}^{N} x_j - \mu \right) \right\}
= E \left\{ (x_k - \mu) \left( \frac{1}{N} \sum_{j=1}^{N} (x_j - \mu) \right) \right\}
= \frac{1}{N} \sum_{j=1}^{N} E\{(x_k - \mu)(x_j - \mu)\}
= \frac{1}{N} \left[ \sum_{j \neq k} E\{(x_k - \mu)(x_j - \mu)\} + E\{(x_k - \mu)^2\} \right]
\]
\[
\frac{1}{N} \left[ \sum_{j \neq k} E[(x_k - \mu)E[(x_j - \mu)] + \sigma^2 \right]
\]

\[
= \frac{\sigma^2}{N}
\]

4. The second expectation yields:
\[
E[(\bar{x} - \mu)^2] = E \left\{ \left( \frac{1}{N} \sum_{j=1}^{N} x_j - \mu \right)^2 \right\}
\]
\[
= E \left\{ \left( \frac{1}{N} \sum_{j=1}^{N} [x_j - \mu] \right)^2 \right\}
\]
\[
= \frac{1}{N^2} E \left\{ \sum_{j=1}^{N} (x_j - \mu) \sum_{k=1}^{N} (x_k - \mu) \right\}
\]
\[
= \frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} E[(x_j - \mu)(x_k - \mu)]
\]

As in the previous result, the only nonzero term in the sum over \(k\) is that \(j = k\), which is equal to \(\sigma^2\). Thus, the second expectation is:
\[
E[(x_k - \mu)^2] = \frac{1}{N^2} \sum_{j=1}^{N} \sigma^2 = \frac{\sigma^2}{N}
\]

5. Combining these results yields:
\[
E[(\bar{x} - \mu)^2] = \sigma^2 - \frac{2\sigma^2}{N} + \frac{\sigma^2}{N}
\]
\[
= \sigma^2 \left[ \frac{1}{N} - 1 \right]
\]
\[
= \left( \frac{N - 1}{N} \right) \sigma^2
\]

Thus, the estimator is unbiased as claimed since:
\[
E[\sigma^2] = \left( \frac{N}{N-1} \right) E[(x_k - \bar{x})^2] = \left( \frac{N}{N-1} \right) \left( \frac{N-1}{N} \right) \sigma^2 = \sigma^2
\]

Exercise 2:

For the unbiased estimator considered in Exercise 1, show that the estimator variance is given by:
\[
\text{var} \{\hat{\sigma}^2\} = \frac{(\kappa + 2)\sigma^4}{N} + \frac{2\sigma^4}{N(N - 1)}
\]
where $\kappa$ is the kurtosis of the common distribution of $\{x_k\}$.

Solution:

1. To simplify subsequent algebra, define $z_k = x_k - \mu$ and note that:

   \[
   E\{z_k\} = E\{x_k\} - \mu = 0
   \]

   \[
   \Rightarrow \operatorname{var}\{z_k\} = E\{z_k^2\} = \sigma^2
   \]

   and $E\{z_k z_\ell\} = E\{z_k\}E\{z_\ell\} = 0$ if $k \neq \ell$

2. Also note that:

   \[
   x_k - \bar{x} = (x_k - \mu) - (\bar{x} - \mu)
   \]

   \[
   = z_k - \left( \frac{1}{N} \sum_{j=1}^{N} x_j - \mu \right)
   \]

   \[
   = z_k - \left( \frac{1}{N} \sum_{j=1}^{N} [x_j - \mu] \right)
   \]

   \[
   = z_k - \frac{1}{N} \sum_{j=1}^{N} z_j
   \]

   \[
   = z_k - \bar{z}
   \]

3. Thus, the variance estimator is:

   \[
   \hat{\sigma}^2 = \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \bar{x})^2
   \]

   \[
   = \frac{1}{N-1} \sum_{k=1}^{N} (z_k - \bar{z})^2
   \]

   \[
   = \frac{1}{N-1} \sum_{k=1}^{N} (z_k^2 - 2z_k \bar{z} + \bar{z}^2)
   \]

   \[
   = \frac{1}{N-1} \left[ \sum_{k=1}^{N} z_k^2 - 2\bar{z} \sum_{k=1}^{N} z_k + \sum_{k=1}^{N} \bar{z}^2 \right]
   \]

   \[
   = \frac{1}{N-1} \left[ \sum_{k=1}^{N} z_k^2 - 2N\bar{z}^2 + N\bar{z}^2 \right]
   \]

   \[
   = \frac{1}{N-1} \left[ \sum_{k=1}^{N} z_k^2 - N\bar{z}^2 \right]
   \]

   \[
   = \frac{1}{N-1} \sum_{k=1}^{N} (z_k^2 - \bar{z}^2)
   \]
4. The variance of $\hat{\sigma}^2$ is:

$$\text{var} \left\{ \hat{\sigma}^2 \right\} = E[\hat{\sigma}^4] - [E[\hat{\sigma}^2]]^2 = E[\hat{\sigma}^4] - \sigma^4 \quad \text{(from Exercise 1)}$$

5. The first expectation is:

$$E[\hat{\sigma}^4] = E \left\{ \left( \frac{1}{N-1} \sum_{k=1}^{N} [z_k^2 - \bar{z}^2] \right)^2 \right\}$$

$$= E \left\{ \frac{1}{(N-1)^2} \sum_{k=1}^{N} [z_k^2 - \bar{z}^2] \sum_{\ell=1}^{N} [z_\ell^2 - \bar{z}^2] \right\}$$

$$= \frac{1}{(N-1)^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{z_k^2 z_\ell^2 - z_k^2 \bar{z}^2 - z_\ell^2 \bar{z}^2 + \bar{z}^4\}$$

$$= \frac{1}{(N-1)^2} \left[ \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{z_k^2 z_\ell^2\} - \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{z_k^2 z_\ell^2\} \right]$$

$$- \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{z_\ell^2 \bar{z}^2\} + \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{\bar{z}^4\}$$

$$= \frac{1}{(N-1)^2} \left[ \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{z_k^2 z_\ell^2\} - N \sum_{k=1}^{N} E\{z_k^2 \bar{z}^2\} \right]$$

$$- N \sum_{\ell=1}^{N} E\{z_\ell^2 \bar{z}^2\} + N^2 E\{\bar{z}^4\}$$

$$= \frac{1}{(N-1)^2} \left[ \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{z_k^2 z_\ell^2\} - 2N \sum_{k=1}^{N} E\{z_k^2 \bar{z}^2\} + N^2 E\{\bar{z}^4\} \right]$$

6. Define the double sum as:

$$S_1 = \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{z_k^2 z_\ell^2\}$$

$$= \sum_{k=1}^{N} \left[ E\{z_k^4\} + \sum_{\ell \neq k} E\{z_k^2 z_\ell^2\} \right]$$

$$= \sum_{k=1}^{N} \left[ (\kappa + 3)\sigma^4 + \sum_{\ell \neq k} E\{z_k^2 E\{z_\ell^2\}\} \right]$$

$$= \sum_{k=1}^{N} \left[ (\kappa + 3)\sigma^4 + \sum_{\ell \neq k} \sigma^4 \right]$$

19
\[ \sum_{k=1}^{N} [(\kappa + 3)\sigma^4 + (N - 1)\sigma^4] \]
\[ = N[(\kappa + 3)\sigma^4 + (N - 1)\sigma^4] \]
\[ = N(N + \kappa + 2)\sigma^4 \]

7. Next, define:

\[ S_2 = 2N \sum_{k=1}^{N} E\{z_k^2 z^2\} \]
\[ = 2N \sum_{k=1}^{N} E \left\{ z_k^2 \left( \frac{1}{N} \sum_{\ell=1}^{N} z_\ell \right) \left( \frac{1}{N} \sum_{m=1}^{N} z_m \right) \right\} \]
\[ = \frac{2}{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \sum_{m=1}^{N} E\{z_k^2 z_\ell z_m\} \]

Note that if \( m \neq \ell \), then we must consider three cases:

\[
  \begin{align*}
  k = \ell \neq m & \Rightarrow E\{z_k^2 z_\ell z_m\} = E\{z_k^3 z_m\} = E\{z_k^3\} E\{z_m\} = 0 \\
  k = m \neq \ell & \Rightarrow E\{z_k^2 z_\ell z_m\} = E\{z_k^3 z_\ell\} = E\{z_k^3\} E\{z_\ell\} = 0 \\
  k \neq m, \ell & \Rightarrow E\{z_k^2 z_\ell z_m\} = E\{z_k^3\} E\{z_\ell\} E\{z_m\} = 0 
  \end{align*}
\]

Thus, the only nonzero contribution from the sum over \( m \) in \( S_2 \) is the term \( m = \ell \), which implies:

\[ S_2 = \frac{2}{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{z_k^2 z_\ell^2\} \]
\[ = \frac{2}{N} S_1 \]
\[ = 2(N + \kappa + 2)\sigma^4 \]

8. Finally, define:

\[ S_3 = N^2 E\{z^4\} \]
\[ = N^2 \left\{ \left( \frac{1}{N} \sum_{k=1}^{N} z_k \right)^4 \right\} \]
\[ = \frac{1}{N^2} \left\{ \sum_{k=1}^{N} \sum_{\ell=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} z_k z_\ell z_m z_n \right\} \]
\[ = \frac{1}{N^2} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} E\{z_k z_\ell z_m z_n\} \]
\[
\begin{align*}
  &= \frac{1}{N^2} \left[ \sum_{k=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} E\{z_k^2 z_m z_n\} + \sum_{k=1}^{N} \sum_{\ell \neq k} \sum_{m=1}^{N} \sum_{n=1}^{N} E\{z_k z_\ell z_m z_n\} \right] \\
  &= \frac{1}{N^2} \left[ \frac{N S_2}{2} + \sum_{k=1}^{N} \sum_{\ell \neq k} \sum_{m=1}^{N} \sum_{n=1}^{N} E\{z_k z_\ell z_m z_n\} \right]
\end{align*}
\]

Note that for \( k \neq \ell \), it follows that \( E\{z_k z_\ell z_m z_n\} = 0 \) unless either \( m = \ell \) and \( n = k \) or \( m = k \) and \( n = \ell \). In either of these cases:

\[
E\{z_k z_\ell z_m z_n\} = E\{z_k^2 z_\ell^2\} = E\{z_k^2\}E\{z_\ell^2\} = \sigma^4
\]

Thus, the quadruple sum above reduces to:

\[
\sum_{k=1}^{N} \sum_{\ell \neq k} \sum_{m=1}^{N} \sum_{n=1}^{N} E\{z_k z_\ell z_m z_n\} = \sum_{k=1}^{N} \sum_{\ell \neq k} 2\sigma^4 = 2N(N-1)\sigma^4
\]

This implies that:

\[
S_3 = \frac{1}{N^2} \left[ \frac{N S_2}{2} + 2N(N-1)\sigma^4 \right] \\
= \frac{1}{N^2} \left[ \frac{N}{2} \cdot 2(N + \kappa + 2)\sigma^4 + 2N(N-1)\sigma^4 \right] \\
= \frac{\sigma^4}{N^2} \cdot [N(N + \kappa + 2)\sigma^4 + 2N(N-1)\sigma^4] \\
= \frac{\sigma^4}{N} \cdot [N + \kappa + 2 + 2(N-1)] \\
= \frac{(3N + \kappa)\sigma^4}{N}
\]

9. Combining these results yields:

\[
E\{[\sigma^2]\} = \frac{1}{(N-1)^2} [S_1 - S_2 + S_3] \\
= \frac{1}{(N-1)^2} \left[ N(N + \kappa + 2)\sigma^4 - 2(N + \kappa + 2)\sigma^4 + \frac{3N + \kappa + \sigma^4}{N} \right] \\
= \frac{\sigma^4}{N(N-1)^3} [N^2(N + \kappa + 2) - 2N(N + \kappa + 2) + 3N + \kappa] \\
= \frac{\sigma^4}{N(N-1)^2} [N^3 + 2N^2 - 2N^2 - 4N + 3N + \kappa(N^2 - 2N + 1)] \\
= \frac{\sigma^4}{N(N-1)^2} [N^3 - N + \kappa(N-1)^2] \\
= \frac{\sigma^4}{N(N-1)^2} \left[ N^2 - 1 \right] N + \kappa(N-1)^2 \sigma^4 \\
= \frac{\sigma^4}{N(N-1)^2} \left( N^2 - 1 \right) N + \kappa(N-1)^2 \sigma^4
\]
\[
\begin{align*}
\frac{(N+1)\sigma^4}{N-1} + \frac{\kappa \sigma^4}{N}
\end{align*}
\]
Thus, the variance is given by:
\[
\text{var}\{\hat{\sigma}^2\} = \mathbb{E}\{\hat{\sigma}^4\} - \sigma^4
\]
\[
= \sigma^4 \left[ \frac{N+1}{N-1} + \frac{\kappa}{N} \right]
\]
\[
= \sigma^4 \left[ \frac{N+1-(N-1)}{N-1} + \frac{\kappa}{N} \right]
\]
\[
= \frac{2\sigma^4}{N-1} + \frac{\kappa \sigma^4}{N}
\]
10. Note that this result can be re-written in the desired form:
\[
\frac{2\sigma^4}{N-1} + \frac{\kappa \sigma^4}{N} = \frac{2\sigma^4}{N-1} - \frac{2\sigma^4}{N} + \frac{(\kappa+2)\sigma^4}{N}
\]
\[
= \frac{2\sigma^4}{N(N-1)} + \frac{(\kappa+2)\sigma^4}{N}
\]

**Exercise 3:**

Using the convolution representation for the density of the sum of two independent random variables, determine the distribution of the sum of \(N\) independent random variables, each uniformly distributed on the interval \([0,1]\). (Hint: start with the simplest case \(N=2\) and then proceed by induction to obtain the Irwin-Hall distribution discussed in Sec. 6.3.1.)

**Solution:**

1. To prove the result by induction, first show that the form of the Irwin-Hall distribution is correct for \(N=2\), and then show that if it is true for \(N\), this implies it is also true for \(N+1\). For \(N=2\), the Irwin-Hall density is:
\[
p_2(x) = \frac{1}{1!} \sum_{j=0}^{k} (-1)^j \binom{2}{j} (x-j) \text{ for } k \leq x \leq k+1,
\]
where \(k = 0 \text{ or } 1\). More explicitly:
\[
p_2(x) = \begin{cases} 
\binom{2}{0} x & \text{for } 0 \leq x \leq 1 \\
\binom{2}{0} x - \binom{2}{1} (x-1) & \text{for } 1 \leq x \leq 2
\end{cases}
\]
\[
= \begin{cases}
  x & \text{for } 0 \leq x \leq 1 \\
  x - 2(x - 1) & \text{for } 1 \leq x \leq 2
\end{cases}
\]

2. From the convolution result Eq. (6.21), the density for \( z = x + y \) where \( x \) has density \( p(x) \) and \( y \) is statistically independent and uniformly distributed on \([0, 1]\) is given by:

\[
g(z) = \int_{-\infty}^{\infty} p(x) u(z - x) dx
\]

where the uniform density is:

\[
u(y) = \begin{cases}
  1 & 0 \leq y \leq 1 \\
  0 & \text{otherwise}
\end{cases}
\]

Thus, the shifted uniform distribution in the convolution integral is:

\[
u(z - x) = \begin{cases}
  1 & 0 \leq z - x \leq 1 \\
  0 & \text{otherwise}
\end{cases}
\]

\[
u(z - x) = \begin{cases}
  1 & -z \leq -x \leq 1 - z \\
  0 & \text{otherwise}
\end{cases}
\]

\[
u(z - 1) \leq x \leq z
\]

Combining these results yields:

\[
g(z) = \int_{z-1}^{z} p(x) dx
\]

3. Applying this result first to the case \( p(x) = u(x) \) shows that the distribution of the sum of two independent uniformly distributed random variables on \([0, 1]\) is the Irwin-Hall distribution \( p_2(x) \):

\[
g(z) = \int_{z-1}^{z} u(x) dx
\]

\[
= \begin{cases}
  0 & z < 0 \\
  \int_{0}^{z} dx & 0 \leq z \leq 1 \\
  \int_{z-1}^{1} dx & 0 \leq z - 1 \leq 1 \\
  0 & z - 1 > 1
\end{cases}
\]

\[
= \begin{cases}
  0 & z < 0 \\
  z & 0 \leq z \leq 1 \\
  1 - (z - 1) & 1 \leq z \leq 2 \\
  0 & z > 2
\end{cases}
\]
4. Now, suppose the Irwin-Hall distribution is correct for the sum of \( N \) independent random variables, each uniformly distributed on \([0, 1]\), and consider the distribution of the sum of \( N + 1 \) such variables. From the convolution results just discussed, this sum has the distribution:

\[
g(z) = \int_{-\infty}^{\infty} p_N(x)u(z-x)\,dx
\]

\[
= \int_{z-1}^{z} p_N(x)\,dx
\]

5. For \( 0 \leq z \leq 1 \), since \( p_N(x) = 0 \) for \( x < 0 \), this integral becomes:

\[
g(z) = \int_{0}^{z} p_N(x)\,dx
\]

Note that this range corresponds to \( k = 0 \) in the definition of \( p_N(x) \), so that:

\[
p_N(x) = \frac{1}{(N-1)!} \sum_{j=0}^{0} (-1)^j \binom{N}{j} (x-j)^{N-1}
\]

\[
= \frac{1}{(N-1)!} \binom{N}{0} x^{N-1}
\]

\[
= \frac{x^{N-1}}{(N-1)!}
\]

Thus, for \( 0 \leq z \leq 1 \), we have:

\[
g(z) = \int_{0}^{z} \frac{x^{N-1}}{(N-1)!}\,dx
\]

\[
= \frac{z^N}{N(N-1)!}\bigg|_{0}^{z}
\]

\[
= \frac{z^N}{N!}
\]

By the same reasoning as in Step 1 above, note that:

\[
p_{N+1}(z) = \frac{z^N}{N!} \quad \text{for } 0 \leq z \leq 1
\]

\[
\Rightarrow g(z) = p_{N+1}(z) \quad \text{for } 0 \leq z \leq 1
\]
6. More generally, note that the sum of \(N + 1\) uniformly distributed random variables on \([0, 1]\) can take values on the interval \([0, N + 1]\). For \(1 \leq z \leq N + 1\), there exists a unique integer \(k\) between 1 and \(N\) such that:

\[
k \leq z \leq k + 1 \Rightarrow z - 1 \leq k \leq z \leq k + 1
\]

This implies that:

\[
g(z) = \int_{z-1}^{z} p_N(x) \, dx = \int_{z-1}^{k} p_N(x \mid k - 1 \leq x \leq k) \, dx + \int_{k}^{z} p_N(x \mid k \leq x \leq k + 1) \, dx
\]

7. The first of these integrals is:

\[
I_1 = \int_{z-1}^{k} p_N(x \mid k - 1 \leq x \leq k) \, dx
\]

\[
= \int_{z-1}^{k} \left[ \frac{1}{(N - 1)!} \sum_{j=0}^{k-1} (-1)^j \binom{N}{j} (x - j)^{N-1} \right] \, dx
\]

\[
= \frac{1}{(N - 1)!} \sum_{j=0}^{k-1} (-1)^j \binom{N}{j} \int_{z-1}^{k} (x - j)^{N-1} \, dx
\]

\[
= \frac{1}{(N - 1)!} \sum_{j=0}^{k-1} (-1)^j \binom{N}{j} \left[ \frac{(x - j)^N}{N} \right]_{z-1}^{k}
\]

\[
= \frac{1}{N(N - 1)!} \sum_{j=0}^{k-1} (-1)^j \binom{N}{j} (k - j)^N
\]

\[
- \frac{1}{N(N - 1)!} \sum_{j=0}^{k-1} (-1)^j \binom{N}{j} (z - [j + 1])^N
\]

\[
= \frac{1}{N} \sum_{j=0}^{k-1} (-1)^j \binom{N}{j} (k - j)^N - \frac{1}{N!} \sum_{\ell=1}^{k} (-1)^{\ell-1} \binom{N}{\ell - 1} (z - \ell)^N
\]

\[
= \frac{1}{N} \sum_{j=0}^{k-1} (-1)^j \binom{N}{j} (k - j)^N + \frac{1}{N!} \sum_{j=1}^{k} (-1)^j \binom{N}{j - 1} (z - j)^N
\]

where the last line follows from the fact that \(\ell\) and \(j\) are dummy summation indices.

8. The second integral is:

\[
I_2 = \int_{k}^{z} p_N(x \mid k \leq x \leq k + 1) \, dx
\]
\[ \int_k^z \left[ \frac{1}{(N-1)!} \sum_{j=0}^{k} (-1)^j \binom{N}{j} (x-j)^{N-1} \right] dx \]

\[ = \frac{1}{(N-1)!} \sum_{j=0}^{k} (-1)^j \binom{N}{j} \left[ \frac{(x-j)^N}{N} \right]_k \]

\[ = \frac{1}{N!} \sum_{j=0}^{k} (-1)^j \binom{N}{j} [(z-j)^N - (k-j)^N] \]

\[ = \frac{1}{N!} \sum_{j=0}^{k} (-1)^j \binom{N}{j} (z-j)^N - \frac{1}{N!} \sum_{j=0}^{k} (-1)^j \binom{N}{j} (k-j)^N \]

9. Thus, the density of the sum is:

\[ g(z) = I_1 + I_2 \]

\[ = \frac{1}{N!} \sum_{j=0}^{k-1} (-1)^j \binom{N}{j} (k-j)^N + \frac{1}{N!} \sum_{j=1}^{k} (-1)^j \binom{N}{j-1} (z-j)^N \]

\[ + \frac{1}{N!} \sum_{j=0}^{k} (-1)^j \binom{N}{j} (z-j)^N - \frac{1}{N!} \sum_{j=0}^{k} (-1)^j \binom{N}{j} (k-j)^N \]

\[ = \frac{1}{N!} \left[ \sum_{j=1}^{k} (-1)^j \binom{N}{j-1} (z-j)^N + \sum_{j=0}^{k} (-1)^j \binom{N}{j} (z-j)^N \right] \]

\[ = \frac{1}{N!} \left\{ (-1)^k \binom{N}{k} (k-k)^N + (-1)^0 \binom{N}{0} (z-0)^N \right\} \]

\[ + \sum_{j=1}^{k} (-1)^j \left[ \binom{N}{j-1} + \binom{N}{j} \right] (z-j)^N \]

\[ = \frac{1}{N!} \left\{ z^N + \sum_{j=1}^{k} (-1)^j \left[ \binom{N}{j-1} + \binom{N}{j} \right] (z-j)^N \right\} \]

10. By result 3.1.4 from Abramowitz and Stegun, p. 10 [1]:

\[ \binom{N}{j-1} + \binom{N}{j} = \binom{N+1}{j} \]

\[ \Rightarrow g(z) = \frac{1}{N!} \left\{ z^N + \sum_{j=1}^{k} (-1)^j \binom{N+1}{j} (z-j)^N \right\} \]

26
\[
\frac{1}{N!} \sum_{j=0}^{k} (-1)^j \binom{N + 1}{j} (z - j)^N
\]

\[= p_{N+1}(z)\]

**Exercise 4:**

The rescaled Irwin-Hall density for \(N = 3\) was considered in Sec. 6.3.1:

\[p(x) = \begin{cases} 
(3 - x^2)/8 & |x| \leq 1, \\
(3 - |x|)^2/16 & 1 \leq |x| \leq 3, \\
0 & |x| \geq 3,
\end{cases}\]

where it was noted that this density provides a reasonable basis for approximating the Gaussian cumulative distribution function (CDF). Integrate this density to obtain this approximation to the Gaussian CDF.

**Solution:**

1. The cumulative distribution function is given by:

\[P(x) = \int_{-\infty}^{x} p(z) \, dz = \int_{-3}^{x} p(z) \, dz\]

2. For \(-3 \leq x \leq -1\), we have:

\[
P(x) = \int_{-3}^{x} \frac{(3 - |z|)^2}{16} \, dz
= \frac{1}{16} \int_{-3}^{x} (3 + z)^2 \, dz
= \frac{1}{16} \int_{0}^{x+3} w^2 \, dw \quad \text{(substitute } w = z + 3) \\
= \frac{w^3}{16} \bigg|_{0}^{x+3}
= \frac{(x + 3)^3}{48}
\]

Note that at the upper limit of this range, \(P(-1) = 2^3/48 = 1/6\).

3. For \(-1 \leq x \leq 0\), we have:

\[
P(x) = \int_{-3}^{x} p(z) \, dz
= \int_{-3}^{-1} p(z) \, dz + \int_{-1}^{x} p(z) \, dz
\]

27
\[
\begin{align*}
&= \frac{1}{6} + \int_{-1}^{x} \left( \frac{3 - z^2}{8} \right) dz \\
&= \frac{1}{6} + \frac{1}{8} \left[ 3z - \frac{z^3}{8} \right]_{-1}^{x} \\
&= \frac{1}{6} + \frac{1}{8} \left[ 3x - \frac{x^3}{3} + 3 - \frac{1}{3} \right] \\
&= \frac{1}{2} + \frac{9x - x^3}{24}
\end{align*}
\]

Note that at the upper limit of this range, we have \( P(0) = 1/2 \).

4. For \( 0 \leq x \leq 1 \), we have:

\[
\begin{align*}
P(x) &= \int_{-3}^{0} p(z)dz + \int_{0}^{x} p(z)dz \\
&= \frac{1}{2} + \int_{0}^{x} \left( \frac{3 - z^2}{8} \right) dz \\
&= \frac{1}{2} + \frac{1}{8} \left[ 3z - \frac{z^3}{8} \right]_{0}^{x} \\
&= \frac{1}{2} + \frac{9x - x^3}{24}
\end{align*}
\]

For the upper limit of this range, we have \( P(1) = 5/6 \).

5. For \( 1 \leq x \leq 3 \), the result is:

\[
\begin{align*}
P(x) &= \int_{-3}^{1} p(z)dz + \int_{1}^{x} p(z)dz \\
&= \frac{5}{6} + \int_{1}^{x} \frac{(3 - z)^2}{16} dz \\
&= \frac{5}{6} + \frac{1}{16} \int_{-2}^{x-3} w^2 bw (\text{substitute } w = z - 3) \\
&= \frac{5}{6} + \frac{1}{16} \left[ \frac{w^3}{3} \right]_{-2}^{x-3} \\
&= \frac{5}{6} + \frac{1}{48} [(x - 3)^3 + 8] \\
&= 1 + \frac{(x - 3)^3}{48} \\
&= 1 - \frac{(3 - x)^3}{48}
\end{align*}
\]

Note that the upper limit of this distribution is \( P(3) = 1 \).
6. Overall, the cumulative distribution function is:

\[
P(x) = \begin{cases} 
0 & x < -3 \\
\frac{(x+3)^3}{48} & -3 \leq x < -1 \\
\frac{1}{2} + \frac{9x-x^3}{24} & -1 \leq x < 1 \\
1 - \frac{(3-x)^3}{48} & 1 \leq x < 3 \\
1 & x \geq 3
\end{cases}
\]

Exercise 5:

Consider the following weighted sum of a sequence \(\{x_k\}\) of independent, identically distributed random variables:

\[
z_N = \sum_{k=1}^{N} \zeta_k x_k.
\]

Show that the kurtosis \(\kappa(z_N)\) of this sum is related to the kurtosis \(\kappa(x)\) of the data sequence via:

\[
\kappa(z_N) = \kappa(x) \left[ \frac{\sum_{k=1}^{N} \zeta_k^4}{(\sum_{k=1}^{N} \zeta_k^2)^2} \right].
\]

Solution:

1. Since kurtosis is independent of both location and scale, there is no loss of generality in assuming the variables have zero mean and unit variance, implying:

\[
E\{x_i\} = 0 \\
E\{x_i^2\} = 1 \\
E\{x_i x_j\} = 0 \text{ if } i \neq j, \text{ and} \\
E\{z_N\} = E\{ \sum_{i=1}^{N} \zeta_i x_i \} = \sum_{i=1}^{N} \zeta_i E\{x_i\} = 0
\]

2. Thus, the desired kurtosis is:

\[
\kappa(z_N) = \frac{E((z_N - E\{z_N\})^4)}{[E((z_N - E\{z_N\})^2)]^2} - 3 = \frac{E\{z_N^4\}}{[E(z_N^2)]^2} - 3
\]

3. For the denominator, we have:

\[
E\{z_N^2\} = E\left\{ \left( \sum_{i=1}^{N} \zeta_i x_i \right)^2 \right\}
\]
6. Thus, the quadruple sum above reduces to the following terms:

\[
E\{z_N^4\} = \sum_{i=1}^{N} \zeta_i x_i \sum_{j=1}^{N} \zeta_j x_j
\]

\[
= \sum_{i=1}^{N} \sum_{j=1}^{N} \zeta_i \zeta_j E\{x_i x_j\}
\]

\[
= \sum_{i=1}^{N} \zeta_i^2 E\{x_i^2\}
\]

\[
= \sum_{i=1}^{N} \zeta_i^2
\]

4. For the numerator, by analogous reasoning:

\[
E\{z_N^4\} = E\left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \zeta_i \zeta_j \zeta_k \zeta_\ell x_i x_j x_k x_\ell \right\}
\]

\[
= \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \zeta_i \zeta_j \zeta_k \zeta_\ell E\{x_i x_j x_k x_\ell\}
\]

5. Note that if any of the four indices \(i, j, k, \ell\) are distinct from all of the others, the expectation vanishes. Specifically, if \(\ell\) is distinct from the other three, we have:

\[
E\{x_i x_j x_k x_\ell\} = E\{x_i x_j x_k\} E\{x_\ell\} = 0
\]

The same reasoning applies to the other three indices. Thus, the only nonzero contributions are the following four cases:

- \(i = j = k = \ell\) \(\Rightarrow\) \(E\{x_i x_j x_k x_\ell\} = E\{x_i^4\} = \kappa(x) + 3\)
- \(i = j \neq k = \ell\) \(\Rightarrow\) \(E\{x_i x_j x_k x_\ell\} = E\{x_i^2 x_j x_k\} = 1\)
- \(i = k \neq j = \ell\) \(\Rightarrow\) \(E\{x_i x_j x_k x_\ell\} = E\{x_i^2 x_k x_\ell\} = 1\)
- \(i = \ell \neq j = k\) \(\Rightarrow\) \(E\{x_i x_j x_k x_\ell\} = E\{x_i^2 x_j x_k\} = 1\)

6. Thus, the quadruple sum above reduces to the following terms:

\[
E\{z_N^4\} = \sum_{i=1}^{N} \zeta_i^4 [\kappa(x) + 3] + \sum_{i=1}^{N} \sum_{\ell \neq i}^{N} \zeta_i^2 \zeta_\ell^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \zeta_i^2 \zeta_j^2 + \sum_{i=1}^{N} \sum_{k \neq i}^{N} \zeta_i^2 \zeta_k^2
\]

\[
= \kappa(x) \sum_{i=1}^{N} \zeta_i^4 + \sum_{i=1}^{N} \sum_{\ell \neq i}^{N} \zeta_i^2 \zeta_\ell^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \zeta_i^2 \zeta_j^2 + \sum_{i=1}^{N} \sum_{k \neq i}^{N} \zeta_i^2 \zeta_k^2
\]

\[
= \kappa(x) \sum_{i=1}^{N} \zeta_i^4 + 3 \left( \sum_{i=1}^{N} \zeta_i^2 \right)^2
\]
7. Combining these results yields:

\[ \kappa(z_N) = \kappa(x) \frac{\sum_{i=1}^{N} \zeta_i^4 + 3 \left( \sum_{i=1}^{N} \zeta_i^2 \right)^2}{\left( \sum_{i=1}^{N} \zeta_i^2 \right)^2} - 3 \]

Exercise 6:

Derive Eq. (6.25): suppose \( x_1 \) and \( x_2 \) are independent, exponentially distributed random variables with the common distribution:

\[ p(x) = \begin{cases} 
  e^{-x} & x \geq 0, \\
  0 & x < 0,
\end{cases} \]

and let \( y = x_1 - x_2 \). Show that \( y \) has the Laplace distribution:

\[ \phi(y) = \frac{1}{2} e^{-|y|}. \]

Solution:

1. Write the difference \( y = x_1 - x_2 \) as the sum of \( x_1 \) and \( z = -x_2 \), which has the density:

\[ g(z) = \begin{cases} 
  e^z & z \leq 0 \\
  0 & z > 0
\end{cases} \]

2. Since \( x_1 \) and \( x_2 \) are statistically independent, so are \( x_1 \) and \( z \), and the distribution for their sum follows from the convolution relationship:

\[ f(y) = \int_{-\infty}^{\infty} p(z) g(y - z) dz = \int_{0}^{\infty} e^{-z} g(y - z) dz \]

3. Note that:

\[ g(y - z) = \begin{cases} 
  e^{y-z} & y - z \leq 0 \\
  0 & \text{otherwise}
\end{cases} \]

\[ = \begin{cases} 
  e^{y}e^{-z} & z \geq 0 \\
  0 & \text{otherwise}
\end{cases} \]

31
4. Thus, for \( y \geq 0 \), the density for the sum is:

\[
f(y) = \int_{y}^{\infty} e^{-z} e^{y} e^{-z} dz
\]

\[
= e^{y} \int_{y}^{\infty} e^{-2z} dz
\]

\[
= e^{y} \left[ \frac{-e^{-2z}}{-2} \right]_{y}^{\infty}
\]

\[
= e^{y} \left[ 0 - \frac{e^{-2y}}{-2} \right]
\]

\[
= \frac{1}{2} e^{-y}
\]

\[
= \frac{1}{2} e^{-|y|}
\]

5. Note that if \( y < 0 \), the lower limit on the integral defining \( f(y) \) only extends to zero, implying:

\[
f(y) = \int_{0}^{\infty} e^{y-2z} dz
\]

\[
= e^{y} \left[ \frac{-e^{-2z}}{-2} \right]_{0}^{\infty}
\]

\[
= \frac{1}{2} e^{y}
\]

\[
= \frac{1}{2} e^{-|y|}
\]

6. Note that the same expression holds for both \( y > 0 \) and \( y < 0 \), corresponding to the density for the Laplace distribution, as claimed.

**Chapter 7**

**Exercise 1:**

For the Laplace distribution defined by:

\[
p(x) = \frac{1}{2\phi} e^{-|x-\mu|/\phi},
\]

compute the normalization factor \( \gamma \) required so that the MADM scale estimate \( S = \gamma S_0 \) is an unbiased estimator of \( \phi \).

**Solution:**

32
1. By the results presented in Sec. 7.4.2, the expected value of the un-normalized MADM scale estimator $S_0$ is:

$$E\{S_0\} = F^{-1} \left( \frac{3}{4} \right) - \mu$$

2. For the Laplace distribution, the inverse cumulative distribution function is given by Eq. (4.80):

$$F^{-1}(q) = \mu + \phi \ln \left[ \frac{1}{2(1-q)} \right]$$

3. Combining these results gives:

$$E\{S_0\} = \phi \ln 2$$

Thus, the required normalization factor is:

$$\gamma = \frac{1}{\ln 2} \approx 1.443$$

Exercise 2:

For Galton’s skewness measure $\gamma_G$ discussed in Sec. 7.6.1, prove that $|\gamma_G| \leq 1$ provided $x_{0.75} \neq x_{0.25}$. (Hint: define $\alpha = x_{0.75} - x_{0.25}$ and $\beta = x_{0.50} - x_{0.25}$. What do we know about $\alpha$ and $\beta$?)

Solution:

1. The Galton skewness measure is:

$$\gamma_G = x_{0.75} + x_{0.25} - 2x_{0.50}$$

$$\gamma_G = \frac{x_{0.75} - x_{0.25}}{x_{0.75} - x_{0.25}} \left( x_{0.75} + x_{0.25} - 2x_{0.50} \right)$$

2. Defining $\alpha = x_{0.75} - x_{0.25}$ and $\beta = x_{0.50} - x_{0.25}$, note that $\gamma_G$ can be written as:

$$\gamma_G = \frac{\alpha - 2\beta}{\alpha}$$

3. Since $x_{0.25} \leq x_{0.50} \leq x_{0.75}$, it follows that:

$$0 \leq \beta \leq \alpha$$
Further, since $x_{0.75} \neq x_{0.25}$, it follows that $\alpha > 0$, which implies:

$$0 \leq \frac{\beta}{\alpha} \leq 1 \implies -2 \leq -2(\beta/\alpha) \leq 0$$

$$\implies -1 \leq 1 - 2(\beta/\alpha) \leq 1$$

$$\implies |\gamma| \leq 1$$

Exercise 3:

Consider the following data sequence:

$$x_k = \begin{cases} 
\mu & k = 1, 2, \ldots, N - p, \\
\lambda \mu & k = N - p + 1, \ldots, N.
\end{cases}$$

This sequence may be viewed as a collection of $N - p$ nominal data points with value $\mu$ and $p$ outlying data points with value $\lambda \mu$.

a. Suppose $\lambda > 1$ and define $\epsilon = p/N$ as the contamination level. What is the maximum contamination permitted if the error in estimating the nominal mean $\mu$ is less than 10%?

b. What is the limiting value of this contamination as $\lambda \to \infty$?

c. Under what conditions will the sample median fail to give the desired result (i.e., $\mu$)? Do these conditions depend on $\lambda$?

Solution:

1. For part a, note that the mean of the sequence $\{x_k\}$ is:

$$\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x_k = \frac{1}{N} \left[ \sum_{k=1}^{N-p} \mu + \sum_{k=N-p+1}^{N} \lambda \mu \right]$$

$$= \frac{1}{N} [(N - p)\mu + p\lambda \mu]$$

$$= \mu - (p/N)\mu + (p/N)\lambda \mu$$

$$= \mu [1 - \epsilon + \lambda \epsilon]$$

$$= \mu [1 + (\lambda - 1)\epsilon]$$

2. Note that since $\lambda > 1$ and $\epsilon > 0$, it follows that $\bar{x} > \mu$. Thus, we want to know, for fixed $\lambda$, the largest value of $\epsilon$ such that $\bar{x}$ does not exceed $1.10\mu$, i.e.:

$$\bar{x} \leq 1.10\mu \implies 1 + (\lambda - 1)\epsilon \leq 1.10$$

$$\implies (\lambda - 1)\epsilon \leq 0.10$$

$$\implies \epsilon \leq \frac{0.1}{\lambda - 1}$$

34
3. For part b, note that the limiting value of $\epsilon$ as $\lambda \to \infty$ is $\epsilon = 0$. This is why the average is a “zero-breakdown estimator.”

4. To compute the median, note that the sample consists of $N - p$ values equal to $\mu$ and $p$ values equal to $\lambda \mu > \mu$. Thus, the median can assume only the following three values:

   a. $x^\dagger = \mu$ if $N - p > p$, implying $p < N/2$ and $\epsilon < 1/2$;
   b. $x^\dagger = \lambda \mu$ if $N - p < p$, implying $p > N/2$ and $\epsilon > 1/2$;
   c. $x^\dagger = \frac{\mu + \lambda \mu}{2} = \frac{1 + \lambda}{2} \mu$ if $N - p = p$, implying $p = N/2$ and $\epsilon = 1/2$.

Thus, the correct result is obtained so long as $\epsilon < 1/2$, which is why the median is called a “50% breakdown estimator.” Also, note that this result does not depend on the magnitude of the contamination $\lambda$, in contrast to the case of the mean considered above.

Exercise 4:

Prove that $\hat{\sigma} = 0$ implies $x_k = \bar{x}_N$ for all $k$, where $\hat{\sigma}$ is the traditional standard deviation estimate and $\bar{x}_N$ is the arithmetic mean of the sample.

Solution:

1. Note that the standard estimator $\hat{\sigma}$ is:

   $$\hat{\sigma} = \left[ \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \bar{x}_N)^2 \right]^{1/2}$$

2. Thus, $\hat{\sigma} = 0$ implies:

   $$\frac{1}{N-1} \sum_{k=1}^{N} (x_k - \bar{x}_N)^2 = 0 \Rightarrow (x_k - \bar{x}_N)^2 = 0 \text{ for all } k$$
   $$\Rightarrow x_k = \bar{x}_N \text{ for all } k$$

Chapter 8

Exercise 1:

Show that for the negative binomial distribution, the following relation holds:

$$\frac{k p_k}{p k-1} = (1 - p)k + (1 - p)(r - 1)$$

Solution:
1. For the negative binomial distribution, the probability \( p_k \) is given by:
\[
p_k = \binom{r + k - 1}{r - 1} (1 - p)^k p^r = \frac{\Gamma(r + k)}{\Gamma(r) \Gamma(k + 1)} \cdot (1 - p)^k p^r
\]

2. The ratio \( \frac{p_k}{p_{k-1}} \) is therefore given by:
\[
\frac{p_k}{p_{k-1}} = \frac{\Gamma(r + k)}{\Gamma(r) \Gamma(k + 1)} \cdot \frac{\Gamma(r) \Gamma(k)}{\Gamma(r + k - 1) \Gamma(k + 1)} \cdot \frac{(1 - p)^k p^r}{(1 - p)^{k-1} p^r}
\]

3. A defining characteristic of the gamma function is that \( \Gamma(z + 1) = z \Gamma(z) \), which implies that:
\[
\frac{p_k}{p_{k-1}} = \frac{r + k - 1}{k} \cdot (1 - p)
\]

4. Substituting this result into the ratio requested in the problem statement and rearranging yields the desired result:
\[
\frac{k p_k}{p_{k-1}} = (r + k - 1)(1 - p) = (1 - p)k + (1 - p)(r - 1)
\]

**Exercise 2:**

**NOTE:**

There is an error in Eq. (8.19), which was carried over into this exercise. Eq. (8.20) is correct, resulting in the correct expression for the negative binomial count metamer. This statement of the exercise and its solution correct the error in Eq. (8.19).

**Corrected statement:**

The negative binomialness plot is based on the following result for the negative binomial distribution:
\[
\ln p_k = r \ln p + k \ln(1 - p) + \ln \Gamma(r + k) - \ln \Gamma(r) - \ln \Gamma(k + 1).
\]
Prove this result.

**Solution:**

1. The solution follows immediately on applying the logarithm to the expression for the negative binomial probability \( p_k \) given above:
\[
p_k = \frac{\Gamma(r + k)}{\Gamma(r) \Gamma(k + 1)} \cdot (1 - p)^k p^r
\]
\[
\Rightarrow \ln p_k = \ln \Gamma(r + k) - \ln \Gamma(r) - \ln \Gamma(k + 1) + k \ln(1 - p) + r \ln p
\]
Exercise 3:

Consider the box and triangle kernel density estimates discussed in Sec. 8.6. For a fixed data distribution \( p(x) \), a fixed sample size \( N \), and a fixed bandwidth \( h \):

a. which kernel has smaller bias?

b. which kernel has smaller variance?

Solution:

1. The bias for the different kernels is determined by the factor defined in Eq. (8.51):
   \[
   \mu_2(K) = \int_{-\infty}^{\infty} z^2 K(z) \, dz
   \]

2. For the box kernel:
   \[
   K(z) = \begin{cases} 
   1 & |z| \leq 1/2 \\
   0 & |z| > 1/2 
   \end{cases}
   \Rightarrow \mu_2(K) = \int_{-1/2}^{1/2} z^2 \, dz
   \]
   \[
   = \frac{z^3}{3} \bigg|_{-1/2}^{1/2}
   = \frac{1}{8} - \left(-\frac{1}{8}\right)
   = \frac{1}{12}
   \]

3. For the triangle kernel:
   \[
   K(z) = \begin{cases} 
   1 - |z| & |z| \leq 1 \\
   0 & |z| > 1 
   \end{cases}
   \Rightarrow \mu_2(K) = \int_{-1}^{1} z^2 (1 - |z|) \, dz
   \]
   \[
   = 2\int_{0}^{1} z^2 (1 - z) \, dz
   = 2 \left[ \frac{z^3}{3} - \frac{z^4}{4} \right]_0^1
   = 2 \left( \frac{1}{3} - \frac{1}{4} \right)
   = \frac{1}{6}
   \]
4. Comparing the results from 2 and 3 above, the box kernel exhibits the smaller bias, consistent with the visual appearance of the plots for these two kernels shown in Figs. 8.17 through 8.20.

5. By Eq. (8.59), the variance for different estimators is proportional to:

\[ V = \int_{-\infty}^{\infty} K^2(z)dz \]

6. For the box kernel, this factor is:

\[ V = \int_{-1/2}^{1/2} dz = 1 \]

while for the triangle kernel, it is:

\[
\begin{align*}
V &= \int_{-1}^{1} [1 - |z|]z^2dz \\
&= 2 \int_{0}^{1} [1 - z]z^2dz \\
&= 2 \int_{0}^{1} [1 - 2z + z^2]dz \\
&= 2 \left[ z - z^2 + \frac{z^3}{3} \right]_0^1 \\
&= \frac{2}{3}
\end{align*}
\]

7. Thus, the triangle kernel has smaller variance, again consistent with the appearance of the plots for the box and triangle kernel results shown in Figs. (8.17) through (8.20).

Exercise 4:

Show that the Nadaraya-Watson kernel estimator does not depend on the kernel \( K_y(\cdot) \), provided it is a symmetric, proper density function; that is, derive Eq. (8.68).

Solution:

1. This estimator is obtained by substituting \( \hat{p}(x,y) \) from Eq. (8.67) into Eq. (8.66):

\[
\hat{m}(x) = \frac{\int_{-\infty}^{\infty} y\hat{p}(x,y)dy}{\int_{-\infty}^{\infty} \hat{p}(x,y)dy}
\]
2. Note that both the numerator and denominator of this expression may be written in the following form, for a suitable choice of the function $f(y)$:

\[
\int_{-\infty}^{\infty} f(y) \hat{p}(x, y) dy = \int_{-\infty}^{\infty} f(y) \left[ \frac{1}{Nh_x h_y} \sum_{k=1}^{N} K_x \left( \frac{x - x_k}{h_x} \right) K_y \left( \frac{y - y_k}{h_y} \right) \right] dy
\]

\[
= \frac{1}{Nh_x} \sum_{k=1}^{N} K_x \left( \frac{x - x_k}{h_x} \right) \int_{-\infty}^{\infty} f(y) K_y \left( \frac{y - y_k}{h_y} \right) dy
\]

3. For the denominator, the corresponding function is $f(y) = 1$ and the resulting integral over $y$ simply yields Eq. (8.46):

\[
\int_{-\infty}^{\infty} K \left( \frac{y - y_k}{h_y} \right) dy = 1
\]

Thus, the denominator of $\hat{m}(x)$ is:

\[
D(x) = \frac{1}{Nh_x} \sum_{k=1}^{N} K_x \left( \frac{x - x_k}{h_x} \right)
\]

4. The numerator corresponds to $f(y) = y$, and the integral over $y$ becomes:

\[
\int_{-\infty}^{\infty} y K \left( \frac{y - y_k}{h_y} \right) dy = y_k
\]

Thus, the numerator of $\hat{m}(x)$ is:

\[
N(x) = \frac{1}{Nh_x} \sum_{k=1}^{N} y_k K_x \left( \frac{x - x_k}{h_x} \right)
\]

5. Combining these results yields Eq. (8.68):

\[
\hat{m}(x) = \frac{N(x)}{D(x)} \frac{\sum_{k=1}^{N} y_k K_x \left( \frac{x - x_k}{h_x} \right)}{\sum_{k=1}^{N} K_x \left( \frac{x - x_k}{h_x} \right)}
\]

Chapter 9

Exercise 1:

It was noted in Sec. 9.4.4 that the Yuen-Welch test is a generalization of the Welch test of the hypothesis that two means are equal in the case where the variances are both unknown and unequal, obtained by replacing ordinary means with trimmed means. The original Welch method computes the same test statistic as the standard $t$-test, but evaluates it relative to a $t$-distribution with a modified number of degrees of freedom $\nu$. From the results given for the Yuen-Welch test, derive an expression for $\nu$ for the classical Welch test.
Solution:

1. The degrees of freedom for the Yuen-Welch test is given by:

\[ \nu = \frac{(d_x^2 + d_y^2)^2}{\frac{d_x^2}{n_x-1} + \frac{d_y^2}{n_y-1}} \]

where \( d_x = \frac{(n_x - 1)s_{Wx}^2}{(n_x - 2g_x)(n_x - 2g_x - 1)} \)

with \( d_y \) defined analogously.

2. The classical Welch test corresponds to the case of no trimming, implying \( g_x = 0 \) and \( s_{Wx}^2 = s_x^2 \), the classical variance estimator. In this case:

\[ d_x = \frac{s_x^2}{n_x} \text{ and } d_y = \frac{s_y^2}{n_y} \]

Note: there is an error in the text on page 404, where \( d_x \) is incorrectly given as \( \frac{s_x^2}{n_x-1} \).

3. Substituting the (correct) un-trimmed result into the Yuen-Welch degrees of freedom expression yields the classical Welch test degrees of freedom:

\[ \nu = \frac{(s_x^2 + s_y^2)^2}{\frac{s_x^2}{n_x^2(n_x-1)} + \frac{s_y^2}{n_y^2(n_y-1)}} \]

Exercise 2:

Prove that the Benjamani-Hochberg FDR control procedure can be expressed as a sequence of tests of hypotheses at a constant target level \( \alpha \) using the modified individual rank-ordered \( p \)-values defined in Eq. (9.81):

\[ \tilde{p}(m) = p(m), \quad \tilde{p}(i) = \min\left\{ \tilde{p}(i+1), \frac{m}{i} p(i) \right\} \quad \text{for } i = m-1, m-2, \ldots, 1. \]

Solution:

1. Note that the definition of the modified probabilities \( \tilde{p}(i) \) implies three things:

\[ (1) \quad \tilde{p}(i) = \min\{\tilde{p}(i+1), \frac{m}{i} p(i)\} \leq \tilde{p}(i+1) \]

\[ (2) \quad \tilde{p}(i) \geq \alpha \Rightarrow \frac{m}{i} p(i) \geq \tilde{p}(i) \geq \alpha \]

\[ \Rightarrow p(i) \geq \frac{i \alpha}{m} \]

\[ (3) \quad \tilde{p}(i) < \alpha \leq \tilde{p}(i+1) \Rightarrow \tilde{p}(i) = \frac{m}{i} p(i) < \alpha \]

\[ \Rightarrow p(i) < \frac{i \alpha}{m} \]

40
2. From (1), it follows that if \( \tilde{p}(m) < \alpha \), then \( \tilde{p}(i) < \alpha \) for \( i = 1, 2, \ldots, m \), and the modified procedure rejects all \( m \) null hypotheses. Conversely, if \( \tilde{p}(m-1) < \alpha \leq \tilde{p}(m) \), it follows from (2) that \( p(m) \geq \alpha \), and from (3) that \( p(m-1) < (m-1)\alpha/m \). Hence, the modified procedure accepts the null hypothesis \( H(m) \) but rejects null hypotheses 1 through \( m-1 \). Note that these results correspond exactly to Step \( m \) of the original Benjamani-Hochberg procedure.

3. More generally, if \( \tilde{p}(j) < \alpha \) for some \( j \), the condition also holds for all \( i = 1, 2, \ldots, j-1 \) and the results correspond to Step \( j \) of the original procedure.

4. Finally, note that if \( \tilde{p}(1) \geq \alpha \), it follows from (1) that \( p(i) \geq \alpha \) for all \( i \), and from (2) that \( p(1) \geq \alpha/m \). Hence, all \( m \) null hypotheses are accepted under exactly the same conditions as in Step 1 of the original procedure.

Exercise 3:

The Holm stepdown procedure for controlling FWER in multiple comparison problems can also be written in terms of modified ranked \( p \)-values, like those given in Eq. (9.81) for the Benjamani-Hochberg procedure. Derive these modified \( p \)-values.

Solution:

1. As with the Benjamani-Hochberg procedure, the form of the modified Holm stepdown procedure is to first define:

   \[
   \tilde{p}(m) = p(m), \quad \tilde{p}(i) = \min\{\tilde{p}(i+1), \gamma(i)p(i)\}, \quad \text{for } i = 1, 2, \ldots, m-1
   \]

   This structure leads to the following three conditions, analogous to those for the modified Benjamani-Hochberg procedure:

   \[
   \begin{align*}
   (1') \quad &\tilde{p}(i) = \min\{\tilde{p}(i+1), \gamma(i)p(i)\} \leq \tilde{p}(i+1) \\
   (2') \quad &\tilde{p}(i) \geq \alpha \Rightarrow \gamma(i)p(i) \geq \tilde{p}(i) \geq \alpha \\
   &\Rightarrow p(i) \geq \alpha/\gamma(i) \\
   (3') \quad &\tilde{p}(i) < \alpha \leq \tilde{p}(i+1) \Rightarrow \tilde{p}(i) = \gamma(i)p(i) < \alpha \\
   &\Rightarrow p(i) < \alpha/\gamma(i)
   \end{align*}
   \]

2. Since the general structure of the Holm stepdown procedure is identical to that of the Benjamani-Hochberg procedure, the key to obtaining the modified Holm procedure of interest here is to choose the function \( \gamma(i) \) to yield the test conditions of the original Holm procedure. Since the threshold at Step \( i \) in the Holm procedure is \( \alpha/(m+1-i) \), this requirement yields \( \gamma(i) = m + 1 - i \).
3. Thus, the modified Holm stepdown procedure tests each null hypothesis using the following modified ranked \( p \)-values:

\[
\tilde{p}_{(m)} = p_{(m)} \\
\tilde{p}_{(i)} = \min\{\tilde{p}_{(i+1)}, (m + 1 - i)p_{(i)}\}
\]

**Chapter 10**

**Exercise 1:**

Using the transformation defined by Eq. (10.53), show that \( E\{x\} = \mu \) for the multivariate normal random vector with the distribution \( N(\mu, \Sigma) \).

**Solution:**

1. Using the representation (10.53), we have:

\[
E\{x\} = E\{\mu + Uv\} = \mu + E\{Uv\} = \mu + UE\{v\} = \mu
\]

This last result follows from the fact that \( v \) is the \( N \)-vector defined in Eq. (10.54) with statistically independent, zero-mean components.

**Exercise 2:**

Again using the transformation defined by Eq. (10.53), show that the covariance of \( x \) is given by Eq. (10.55):

\[
\text{cov } x \equiv E\{(x - \mu)(x - \mu)^T\} = \Sigma.
\]

**Solution:**

1. Again by Eq. (10.53), the covariance is:

\[
E\{(x - \mu)(x - \mu)^T\} = E\{Uv(Uv)^T\} = E\{Uvv^TU^T\} = UE\{vv^T\}U^T
\]

2. Since \( v \) has \( N \) statistically independent components, each with variance \( \sigma_i^2 \), \( E\{vv^T\} = D \), the diagonal matrix defined in Eq. (10.52). Thus, \( E\{(x - \mu)(x - \mu)^T\} = UD^TU^T = \Sigma \), as claimed.

**Exercise 3:**
Once more using this transformation, show that if \( x \sim N(\mu, \Sigma) \), then \( Ax + b \) is normally distributed with mean \( A\mu + b \) and covariance \( A^T\Sigma A \).

Solution:

1. First, note that by Eq. (10.53):
\[
Ax + b = A[\mu + Uv] + b = A\mu + b + AUv
\]

2. Taking the expectation of this result yields:
\[
E\{Ax + b\} = A\mu + b + E\{AUv\} = A\mu + b + AU\mathbb{E}\{v\} = A\mu + b
\]

3. The covariance matrix \( C \) is given by:
\[
C = E\{(Ax + b - [A\mu + b])(Ax + b - [A\mu + b])^T\}
= E\{(Ax - A\mu)(Ax - A\mu)^T\}
= E\{A(x - \mu)(x - \mu)^T A^T\}
= A\Sigma A^T
\]

4. Since \( Ax + b \) is a linear combination of the jointly Gaussian components of \( x \), it follows that this combination is also Gaussian and thus completely specified by the mean and covariance given here.

Exercise 4:

Consider the Laplace distribution with mean 0 and scale parameter \( \phi \), defined by the density:
\[
p(x) = \frac{e^{-|x|/\phi}}{2\phi}.
\]

Further, consider the truncated version of this distribution, restricted to the interval \(-L \leq x \leq L\).

a. Derive an expression for the truncated density \( p(x) - L \leq x \leq L \).

b. Show that all odd moments of this distribution vanish.

c. Derive an expression for the variance of this distribution, as a function of the truncation limit \( L \).
d. Derive an expression for the kurtosis $\kappa(L)$. Is this distribution platykurtic, leptokurtic, or mesokurtic?

Solution:

1. The general form for a truncated distribution is given by Eq. (10.31) as:
   \[ p(x| -L \leq x \leq L) = \frac{p(x)}{F(L) - F(-L)} \]

2. For the Laplace distribution, it follows from Eq. (4.79) that:
   \[ F(L) = 1 - \frac{1}{2} e^{-L/\phi} \]
   \[ F(-L) = \frac{1}{2} e^{-L/\phi} \]
   Thus, the denominator in the above expression is:
   \[ F(L) - F(-L) = 1 - e^{-L/\phi} \]
   The truncated Laplace density is therefore given explicitly by:
   \[ p(x| -L \leq x \leq L) = \frac{e^{-|x|/\phi}}{2\phi(1 - e^{-L/\phi})} \]

3. Since the truncated distribution is symmetric, it follows from the discussion on page 147 that all odd-order moments—including the mean—vanish.

4. Since the distribution has mean zero, the variance is simply equal to the second moment:
   \[ \sigma^2 = \int_{-L}^{L} x^2 p(x| -L \leq x \leq L) dx \]
   \[ = \frac{1}{2\phi(1 - e^{-L/\phi})} \int_{-L}^{L} x^2 e^{-|x|/\phi} dx \]
   \[ = \frac{1}{\phi(1 - e^{-L/\phi})} \int_{0}^{L} x^2 e^{-x/\phi} dx \]
   \[ = \frac{1}{1 - e^{-L/\phi}} \int_{0}^{L} (\phi z)^2 e^{-z} dz \quad \text{(substituting } z = x/\phi) \]
   \[ = \frac{\phi^2}{1 - e^{-c}} \int_{0}^{c} z^2 e^{-z} dz \quad \text{(defining } c = L/\phi) \]

5. From Gradshteyn and Ryzhik [122], integral 2.322, no. 2:
   \[ \int_{0}^{c} z^2 e^{-z} dz = \left[ e^{-z} (z^2 - 2z - 2) \right]_{0}^{c} \]
   \[ = [-e^{-c}(c^2 + 2c + 2) - (-2)] \]
   \[ = 2 - e^{-c}[2 + 2c + c^2] \]
6. Thus, combining these results yields:

$$\sigma^2 = \left\{ \frac{2 - e^{-c}[2 + 2c + c^2]}{1 - e^{-c}} \right\}\phi^2$$

Note that as $L \to \infty$ we also have $c \to \infty$ and $\sigma^2 \to 2\phi^2$, corresponding to the correct limiting value for the un-truncated Laplace distribution.

7. Again since the distribution is zero-mean, the kurtosis is:

$$\kappa = \frac{E(x^4)}{[E(x^2)]^2} - 3$$

In terms of the truncation parameter $c$, the fourth moment is:

$$E(x^4) = \frac{1}{2\phi(1 - e^{-c})} \int_{-L}^{L} x^4 e^{-|x|/\phi} dx$$

$$= \frac{1}{2(1 - e^{-c})} \int_{-c}^{c} (\phi z)^4 e^{-|z|} dz \quad \text{(substitute } z = x/\phi)$$

$$= \frac{\phi^4}{2(1 - e^{-c})} \int_{-c}^{c} z^4 e^{-|z|} dz$$

$$= \frac{\phi^4}{(1 - e^{-c})} \int_{0}^{c} z^4 e^{-z} dz$$

8. From integral 2.321, no. 2 in Gradshteyn and Rhyzhik [122]:

$$\int x^n e^{ax} dz = e^{ax} \left( \frac{x^n}{a} + \sum_{k=1}^{n} \frac{(-1)^k n(n - 1) \cdots (n - k + 1)}{a^{k+1}} x^{n-k} \right)$$

For $n = 4$, this result specializes to:

$$\int x^4 e^{ax} dx = e^{ax} \left( \frac{x^4}{a} - \frac{4x^3}{a^2} + \frac{12x^2}{a^3} - \frac{24x}{a^4} + \frac{24}{a^5} \right)$$

9. The desired fourth moment is thus:

$$E(x^4) = \left( \frac{\phi^4}{1 - e^{-c}} \right) [e^{-z}(-z^4 - 4z^3 - 12z^2 - 24z - 24)]_0^c$$

$$= \frac{\phi^4}{1 - e^{-c}} [-e^{-c}(c^4 + 4c^3 + 12c^2 + 24c + 24) + \frac{24}{a^5}]$$

$$= \frac{\phi^4[24 - (24 + 24c + 12c^2 + 4c^3 + c^4)e^{-c}]}{1 - e^{-c}}$$

Combining this with the variance result given above, the kurtosis is:

$$\kappa = \frac{E(x^4)}{[E(x^2)]^2} - 3$$

$$= \frac{\phi^4[24 - (24 + 24c + 12c^2 + 4c^3 + c^4)e^{-c}]}{1 - e^{-c}} \cdot \frac{(1 - e^{-c})^2}{\phi^4[2 - (2 + 2c + c^2)e^{-c}]^2} - 3$$

$$= \frac{(1 - e^{-c})[24 - (24 + 24c + 12c^2 + 4c^3 + c^4)e^{-c}]}{[2 - (2 + 2c + c^2)e^{-c}]^2} - 3$$

45
10. This expression is messy enough to yield little insight by itself, but note that as $c \to \infty$, we should recover the un-truncated result. Note that in this limit $e^{-c} \to 0$ rapidly enough that $e^{n}e^{-c} \to 0$ for all finite $n$. Thus:

$$c \to \infty \Rightarrow \kappa \to \frac{24}{2^2} - 3 = 6 - 3 = 3$$

which is the correct limiting kurtosis for the un-truncated Laplace distribution. It can also be shown by a limiting argument that $\kappa \to -6/5$ (the uniform distribution limit) as $c \to 0$. It is, however, more informative to plot $\kappa$ over a reasonable range of $c$ values to see the general behavior, which can be platykurtic, mesokurtic, or leptokurtic, depending on the particular value of $c$. This plot is shown in Fig. 2.
Exercise 5—NOTE: there is an error in the original problem statement, which has been corrected here

For the bivariate Gaussian distribution defined by Eq. (10.65), show that $E((x - \bar{x})(y - \bar{y})) = \rho \sigma_x \sigma_y$.

Solution:

1. To simplify subsequent algebra, first define the following quantities:

   \[
   A = 1 - \rho^2
   \]

   \[
   u = \frac{x - \bar{x}}{\sigma_x} \Rightarrow x = \bar{x} + \sigma_x u
   \]

   \[
   v = \frac{y - \bar{y}}{\sigma_y} \Rightarrow y = \bar{y} + \sigma_y v
   \]

2. The required expectation is:

   \[
   E((x - \bar{x})(y - \bar{y})) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})(y - \bar{y})p(x, y)dx dy
   \]

   where $p(x, y)$ is defined in Eq. (10.65):

   \[
   p(x, y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left[ \frac{(x - \bar{x})^2}{\sigma_x^2} - \frac{2\rho(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y} + \frac{(y - \bar{y})^2}{\sigma_y^2} \right] \right\}
   \]

   \[
   = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{A}} \exp\left\{ -\frac{u^2 - 2\rho uv + v^2}{2A} \right\}
   \]

3. Thus, the expectation may be written as:

   \[
   E((x - \bar{x})(y - \bar{y})) = \frac{\sigma_x \sigma_y}{2\pi \sqrt{A}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u v \exp\left\{ -\frac{u^2 - 2\rho uv + v^2}{2A} \right\} dudv
   \]

4. For simplicity, denote the expectation incorporating this integral by $I$ and note that:

   \[
   u^2 - 2\rho uv + v^2 = (u - \rho v)^2 + Av^2
   \]

   \[
   = (u - \rho v)^2 + Av^2
   \]

   Thus, the expectation becomes:

   \[
   I = \frac{\sigma_x \sigma_y}{2\pi \sqrt{A}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv \exp\left\{ -\frac{(u - \rho v)^2 + Av^2}{2A} \right\} dudv
   \]

   \[
   = \frac{\sigma_x \sigma_y}{2\pi \sqrt{A}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv \exp\left\{ -\frac{(u - \rho v)^2}{2A} \right\} \exp\left\{ -\frac{v^2}{2} \right\} dudv
   \]
Define \( w = u - \rho v \) so that \( u = w + \rho v \) and rewrite the integral as:

\[
I = \frac{\sigma_x \sigma_y}{2\pi \sqrt{A}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (w + \rho v) e^{-w^2/2A} e^{-v^2/2} dw dv
\]

\[
= \frac{\sigma_x \sigma_y}{2\pi \sqrt{A}} \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} wve^{-w^2/2A} e^{-v^2/2} dw dv + \rho \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v^2 e^{-w^2/2A} e^{-v^2/2} dw dv \right]
\]

\[
= \frac{\sigma_x \sigma_y}{2\pi \sqrt{A}} \left[ \int_{-\infty}^{\infty} we^{-w^2/2A} dw \int_{-\infty}^{\infty} ve^{-v^2/2} dv + \rho \int_{-\infty}^{\infty} e^{-w^2/2A} dw \int_{-\infty}^{\infty} v^2 e^{-v^2/2} dv \right]
\]

5. Note that the first pair of integrals both vanish, corresponding to the expressions for the means of the zero-mean Gaussian random variables \( w \) and \( v \). Thus, the expectation reduces to:

\[
I = \frac{\rho \sigma_x \sigma_y}{2\pi \sqrt{A}} \int_{-\infty}^{\infty} e^{-w^2/2A} dw \int_{-\infty}^{\infty} v^2 e^{-v^2/2} dv
\]

\[
= \rho \sigma_x \sigma_y \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-w^2/2A} dw \right] \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} v^2 e^{-v^2/2} dv \right]
\]

Note that the first term in brackets is the normalization integral for \( w \), equal to 1, while the second integral is the variance of \( v \), which is also equal to 1. Thus, we have the final result:

\[
E\{ (x - \bar{x})(y - \bar{y}) \} = \rho \sigma_x \sigma_y
\]

Exercise 6:

In what follows, suppose \( P \) is any \( N \times N \) permutation matrix.

a. Define \( e = [1, 1, \ldots, 1]^T \in \mathbb{R}^N \) and show that \( Pe = e \).

b. Define \( E \) as the \( N \times N \) matrix, all of whose elements are 1. Show that \( PE = E \) and \( EP^T = E \).

c. Consider the \( N \)-dimensional random vector \( x \sim N(\mu, \Sigma) \) where:

\[
\mu = [\bar{x}, \bar{x}, \ldots, \bar{x}]^T
\]

\[
\Sigma = \begin{pmatrix}
\sigma^2 & \rho \sigma^2 & \cdots & \rho \sigma^2 \\
\rho \sigma^2 & \sigma^2 & \cdots & \rho \sigma^2 \\
\vdots & \vdots & \ddots & \vdots \\
\rho \sigma^2 & \rho \sigma^2 & \cdots & \sigma^2
\end{pmatrix}
\]

Show that \( Px \sim N(\mu, \Sigma) \), implying that the components of \( x \) are exchangeable, even though they are not independent unless \( \rho = 0 \).
Solution:

1. A permutation matrix is an $N \times N$ matrix with one element equal to 1 in each row and each column, with all other elements equal to zero. Thus, if $P$ is a permutation matrix and $e = [1, 1, \ldots, 1]^T$, the $i^{th}$ element of the product $Pe$ is:

$$[Pe]_i = \sum_{j=1}^{N} P_{ij} e_j = \sum_{j=1}^{N} P_{ij} = 1$$

for $i = 1, 2, \ldots, N$. Thus, it follows that $Pe = e$.

2. If $E$ is the $N \times N$ matrix with all 1’s, it follows that:

$$[PE]_{ij} = \sum_{k=1}^{N} P_{ik} E_{kj} = \sum_{k=1}^{N} P_{ik} = 1$$

for $i, j = 1, 2, \ldots, N$. Thus, it follows that $PE = E$. Similarly:

$$[EP^T]_{ij} = \sum_{k=1}^{N} E_{ik} P^T_{kj} = \sum_{k=1}^{N} E_{ik} P_{jk} = \sum_{k=1}^{N} P_{jk} = 1$$

for $i, j = 1, 2, \ldots, N$, implying $EP^T = E$.

3. If $x \sim N(\mu, \Sigma)$ and $P$ is a permutation matrix, it follows from Eq. (10.56) that $Px \sim N(P\mu, P^T\Sigma P)$. Here:

$$\mu = [\bar{x}, \bar{x}, \ldots, \bar{x}]^T = xe$$

$$\Rightarrow P\mu = \bar{x}Pe = \bar{x}e = \mu$$

4. For the covariance result, note that $\Sigma$ may be written as:

$$\Sigma = \begin{bmatrix} \rho \sigma^2 & \rho \sigma^2 & \cdots & \rho \sigma^2 \\ \cdots & \cdots & \cdots & \cdots \\ \rho \sigma^2 & \rho \sigma^2 & \cdots & \rho \sigma^2 \\ \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

Thus, the transformed covariance matrix is:

$$P^T\Sigma P = P^T[\rho \sigma^2 E - (\rho - 1)\sigma^2 I]P$$

$$= \rho \sigma^2 P^T EP - (\rho - 1)\sigma^2 P^T IP$$

For any permutation matrix $P$, the transpose $P^T$ is also a permutation matrix and corresponds to the inverse of $P$. These results imply that:

$$P^T EP = EP = E$$

$$P^T IP = P^T P = I$$
Combining these results yields:

\[ \mathbf{P}^T \Sigma \mathbf{P} = \rho \sigma^2 \mathbf{I} - (\rho - 1)\sigma^2 \mathbf{I} = \Sigma \]

Thus, it follows that \( \mathbf{P} \mathbf{x} \sim N(\mu, \Sigma) \) for any permutation matrix \( \mathbf{P} \).

Exercise 7:

Let \( R_x(i) \) denote the ranks for the \( N \) elements of a data sequence \( \{x_k\} \).

a. Show that the average of the sequence \( \{R_x(i)\} \) is \( (N + 1)/2 \).

b. Show that the variance of this sequence is \( (N^2 - 1)/12 \).

Solution:

1. The ranks \( R_x(i) \) for the \( N \) element sequence \( \{x_k\} \) consist of the numbers from 1 to \( N \) in some specified order. Thus:

\[
\bar{R} = \frac{1}{N} \sum_{i=1}^{N} R_x(i) = \frac{1}{N} \sum_{j=1}^{N} j = \frac{1}{N} \left( \frac{N(N + 1)}{2} \right) = \frac{N + 1}{2}
\]

2. The variance is most easily computed as:

\[
\text{var}\{R_x(i)\} = E\{R_x(i)^2\} - \bar{R}^2
\]

The first of these expectations is:

\[
E\{R_x(i)^2\} = \frac{1}{N} \sum_{i=1}^{N} R_x(i)^2 = \frac{1}{N} \sum_{j=1}^{N} j^2 = \frac{1}{N} \left( \frac{N(N + 1)(2N + 1)}{6} \right) = \frac{N(2N + 1)}{6}
\]

Thus, the variance is given by:

\[
\text{var}\{R_x(i)\} = \frac{N + 1}{2} \left[ \frac{4N + 2 - 3N - 3}{6} \right] = \frac{N + 1}{2} \left[ \frac{N - 1}{6} \right] = \frac{N^2 - 1}{12}
\]

Exercise 8:
Show that the ranks \( R_x(i) \) for a data sequence are invariant under arbitrary increasing transformations. That is, if \( y_k = f(x_k) \) for any increasing function \( f(\cdot) \), then \( R_y(i) = R_x(i) \).

**Solution:**

1. The ranks \( R_x(i) \) are the integers from 1 through \( N \) that define the position of each individual element \( x_k \) in the re-ordered sequence:

\[
x(1) \leq x(2) \leq \cdots \leq x(N)
\]

That is, if \( x_k = x(j) \), then \( R_x(k) = j \). Now, suppose \( y_k = f(x_k) \) where \( f(\cdot) \) is an increasing function. This means that:

\[
x_k \leq x_\ell \Rightarrow y_k = f(x_k) \leq f(x_\ell) = y_\ell
\]

Thus, the order in the transformed sequence:

\[
y(1) \leq y(2) \leq \cdots \leq y(N)
\]

is the same for \( \{y_k\} \) as it is for the original sequence \( \{x_k\} \). Thus:

\[
R_x(k) = j \Rightarrow x_k = x(j) \Rightarrow y_k = y(j) \Rightarrow R_y(k) = j
\]

**Exercise 9:**

Consider the Spearman rank correlation coefficient \( \rho_S \) between two sequences \( \{x_k\} \) and \( \{y_k\} \) introduced in Sec. 10.5.

- Show that \( \rho_S = +1 \) if and only if \( y_k = f(x_k) \) for some increasing function \( f(\cdot) \).
- Show that \( \rho_S = -1 \) if and only if \( y_k = f(x_k) \) for some decreasing function \( f(\cdot) \).

(Hint: remember that \( R_x(i) \) and \( R_y(i) \) must assume every value between 1 and \( N \) once and only once. Also, note Exercise 8.)

**Solution:**

1. The Spearman rank correlation \( \rho_S \) is the product-moment correlation coefficient between the ranks \( R_x(i) \) and \( R_y(i) \). Thus, \( \rho_S = +1 \) if and only if:

\[
R_y(i) = aR_x(i) + b
\]

for some \( a > 0 \). Since \( R_x(i) \) and \( R_y(i) \) each take on all of the integer values between 1 and \( N \), this condition can only be met for \( a = 1 \) and \( b = 0 \), implying that \( R_x(i) = R_y(i) \) for all \( i \). In other words, \( \rho_S = +1 \) if and only if the ranks of \( \{x_k\} \) and \( \{y_k\} \) are identical, implying that these sequences are related by an increasing transformation.
2. Similarly, \( \rho_S = -1 \) implies \( R_y(i) = aR_x(i) + b \) for some \( a < 0 \). This is only possible for \( a = -1 \) and \( b = N + 1 \), giving:

\[
R_y(i) = N + 1 - R_x(i)
\]

Note that this linear transformation reverses the rank order, requiring that \( x(1) \) be mapped into \( y(N) \), that \( x(2) \) be mapped into \( y(N-1) \), and so forth. This implies that these two sequences are related via a decreasing transformation \( f(\cdot) \) such that:

\[
x_i < y_j \Rightarrow y_i = f(x_i) > f(x_j) = y_j
\]

**Exercise 10:**

Show that the variance of the mixture density defined by Eq. (10.105) is given by Eq. (10.110).

**Hint:** recall that \( \text{var}\{x\} = E\{x^2\} - [E\{x\}]^2 \).

**Solution:**

1. Consider the mixture density defined in Eq. (10.105):

\[
p(x) = \sum_{j=1}^{p} \pi_j f_j(x)
\]

As noted in the hint, the variance of this distribution can be written in terms of the non-central first and second moments as:

\[
\text{var}\{x\} = E\{x^2\} - [E\{x\}]^2
\]

These moments are given by:

\[
E\{x\} = \sum_{j=1}^{p} \pi_j E\{x_j\} = \sum_{j=1}^{p} \pi_j \bar{x}_j
\]

\[
E\{x^2\} = \sum_{j=1}^{p} \pi_j E\{x_j^2\} = \sum_{j=1}^{p} \pi_j [\bar{x}_j^2 + \sigma_j^2]
\]

2. The square of the first moment may be written as:

\[
[E\{x\}]^2 = \left(\sum_{j=1}^{p} \pi_j \bar{x}_j\right)^2 = \sum_{j=1}^{N} \sum_{k=1}^{N} \pi_j \pi_k \bar{x}_j \bar{x}_k
\]

Combining these results yields Eq. (10.110):

\[
\text{var}\{x\} = \sum_{j=1}^{p} \pi_j \sigma_j^2 + \sum_{j=1}^{p} \pi_j \bar{x}_j^2 - \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k \bar{x}_j \bar{x}_k
\]

\[
= \sum_{j=1}^{p} \pi_j \sigma_j^2 + \sum_{j=1}^{p} \pi_j \bar{x}_j \left[ \bar{x}_j - \sum_{k=1}^{p} \pi_k \bar{x}_k \right]
\]
Exercise 11:

Show that the following expressions are equivalent, provided the coefficients $\pi_j$ are nonnegative and sum to 1:

$$\sum_{j=1}^{p} \pi_j z_j \left[ z_j - \sum_{k=1}^{p} \pi_k z_k \right] = \frac{1}{2} \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (z_j - z_k)^2 \geq 0.$$ 

Further, show that this lower bound is achievable if and only if $z_j = z_k$ for all $1 \leq j, k \leq p$.

Solution:

1. First, note that since the coefficients $\pi_i$ sum to 1, we have:

$$S = \sum_{j=1}^{p} \pi_j z_j \left[ z_j - \sum_{k=1}^{p} \pi_k z_k \right] = \sum_{j=1}^{p} \pi_j z_j \left[ z_j - \sum_{k=1}^{p} \pi_k - \sum_{k=1}^{p} \pi_k z_k \right] = \sum_{j=1}^{p} \pi_j z_j \left[ \sum_{k=1}^{p} \pi_k (z_j - z_k) \right] = \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (z_j - z_k).$$

2. Note that:

$$z_j(z_j - z_k) = [(z_j - z_k) + z_k](z_j - z_k) = (z_j - z_k)^2 + z_k(z_j - z_k)$$

Thus, we have:

$$S = \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k [(z_j - z_k)^2 + z_k(z_j - z_k)]$$

$$= \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (z_j - z_k)^2 + \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k z_k(z_j - z_k)$$

$$= \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (z_j - z_k)^2 - \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (z_k - z_j)$$

$$= \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (z_j - z_k)^2 - S$$

$$\Rightarrow 2S = \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (z_j - z_k)^2$$

$$\Rightarrow S = \frac{1}{2} \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (z_j - z_k)^2.$$
3. Since all of the terms in this double sum are nonnegative, the lower bound of zero is achievable if and only if each term is separately zero, implying $z_j = z_k$ for all $1 \leq j, k \leq p$.

Exercise 12:

Show that, if all component densities $f_j(x)$ in the mixture density defined by Eq. (10.109) are zero-mean and symmetric, the kurtosis $\kappa(x)$ for the mixture density is given by Eq. (10.112).

Solution:

1. If all component densities are symmetric about zero, it follows that $\bar{x}_j = 0$ for all $j$ and the variance expression reduces to:

$$\text{var}\{x\} = E\{x^2\} = \sum_{j=1}^{p} \pi_j \sigma_j^2$$

2. Similarly, the fourth moment expression becomes:

$$E\{x^4\} = \sum_{j=1}^{p} \pi_j E\{x_j^4\}$$

$$= \sum_{j=1}^{p} \pi_j (\kappa_j + 3) \sigma_j^4$$

$$= \sum_{j=1}^{p} \pi_j \kappa_j \sigma_j^4 + 3 \sum_{j=1}^{p} \pi_j \sigma_j^4$$

where $\kappa_j$ is the kurtosis of component $j$.

2. Combining these results, it follows that the kurtosis for the mixture distribution is given by Eq. (10.112):

$$\kappa(x) = \frac{E\{x^4\}}{[E\{x^2\}]^2} - 3$$

$$= \frac{E\{x^4\} - 3[E\{x^2\}]^2}{[E\{x^2\}]^2}$$

$$= \left[ \sum_{j=1}^{p} \pi_j \sigma_j^2 \right]^{-2} \left\{ \sum_{j=1}^{p} \pi_j \kappa_j \sigma_j^4 + 3 \sum_{j=1}^{p} \pi_j \sigma_j^4 - 3 \sum_{j=1}^{p} \pi_j \sigma_j^2 \sum_{k=1}^{p} \pi_k \sigma_k^2 \right\}$$

$$= \left[ \sum_{j=1}^{p} \pi_j \sigma_j^2 \right]^{-2} \left\{ \sum_{j=1}^{p} \pi_j \kappa_j \sigma_j^4 + 3 \sum_{j=1}^{p} \pi_j \sigma_j^2 \left[ \sigma_j^2 - \sum_{k=1}^{p} \pi_k \sigma_k^2 \right] \right\}$$

Exercise 13:
Show that, if all component densities \( f_j(x) \) in the mixture density defined by Eq. (10.109) have the same mean, \( \bar{x}_j = \bar{x}^0 \), the central moments \( \mu_n \) of the mixture density depend only on the central moments \( \mu_n^j \) of the component densities.

Solution:

1. If all component densities have mean \( \bar{x}^0 \), it follows from Eq. (10.109) that:
\[
E\{x\} = \sum_{j=1}^{p} \pi_j \bar{x}^0 = \bar{x}^0 \sum_{j=1}^{p} \pi_j = \bar{x}^0
\]
Thus, we can consider the zero-mean deviation variable \( z = x - \bar{x}^0 \), noting that its density is given by:
\[
p(z) = \sum_{j=1}^{p} \pi_j \phi_j(z)
\]
where \( \phi_j(z) \) is the density for the component deviation variable \( z_j = x_j - \bar{x}^0 \).

2. The advantage of this representation is that the central moments of \( x \) are the moments of \( z \), i.e.:
\[
E\{(x - \bar{x}^0)^n\} = E\{z^n\} = \sum_{j=1}^{p} \pi_j E\{z_j^n\}
\]
by Eq. (10.108). Note that \( E\{z_j^n\} \) is simply the \( n^{th} \) central moment of the original component variable \( x_j \).

Exercise 14:

Show that the kurtosis \( \kappa(x) \) for the \( p \)-component zero-mean, Gaussian mixture density is given by the result presented in Eq. (10.121). Also, show that \( \kappa(x) \geq 0 \) with \( \kappa(x) = 0 \) if and only if \( \sigma_1 = \sigma_2 = \cdots = \sigma_p \).

Solution:

1. From Eq. (10.112) for the general symmetric, zero-mean case with \( \kappa_j = 0 \) for all \( j \), we have Eq. (10.121) from the results of Exercise 11:
\[
\kappa(x) = \left[ \sum_{j=1}^{p} \pi_j \sigma_j^2 \right]^{-2} \left\{ \sum_{j=1}^{p} \pi_j \sigma_j^2 \left[ \sigma_j^2 - \sum_{k=1}^{p} \pi_k \sigma_k^2 \right] \right\}^{-1}
\]
\[
= \left[ \sum_{j=1}^{p} \pi_j \sigma_j^2 \sum_{k=1}^{p} \pi_k \sigma_k^2 \right]^{-1} \left\{ \sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (\sigma_j^2 - \sigma_k^2)^2 \right\}
\]
\[
= \frac{3}{2} \left( \frac{\sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k (\sigma_j^2 - \sigma_k^2)^2}{\sum_{j=1}^{p} \sum_{k=1}^{p} \pi_j \pi_k \sigma_j^2 \sigma_k^2} \right)
\]
55
2. Since \( \pi_j \geq 0 \) and \( \sigma^2_j \geq 0 \) for all \( j \), it follows that \( \kappa(x) \geq 0 \) for cases. Further, note that to achieve the lower bound \( \kappa(x) = 0 \), it is necessary that all terms \( \sigma^2_j - \sigma^2_k \) vanish, implying that \( \sigma_j = \sigma_k \) for all \( j \) and \( k \).

Exercise 15:

Derive Eq. (10.122) for the first three moments of the slippage model. Next, suppose \( \mu = \lambda \sigma \) and derive a simplified expression for the skewness that does not depend on \( \sigma \). Finally, derive an approximate expression for the skewness in the limit as \( \lambda \to \infty \).

Solution:

1. The slippage model has the distribution:

\[ p(x) = (1 - \epsilon) \phi(x; 0, \sigma^2) + \epsilon \phi(x; \mu, \sigma^2) \]

where \( \phi(x; \mu, \sigma^2) \) represents the Gaussian density with mean \( \mu \) and variance \( \sigma^2 \). The mean of the slippage model is therefore:

\[ E\{x\} = (1 - \epsilon) \cdot 0 + \epsilon \cdot \mu = \epsilon \mu \]

2. The variance is most easily obtained as:

\[ \text{var}\{x\} = E\{x^2\} - [E\{x\}]^2 \]

In this case, the second moment is:

\[ E\{x^2\} = (1 - \epsilon) \sigma^2 + \epsilon \sigma^2 + \epsilon^2 \mu^2 \]

\[ \Rightarrow \text{var}\{x\} = \sigma^2 + \epsilon \mu^2 - \epsilon^2 \mu^2 \]

\[ = \sigma^2 + \epsilon (1 - \epsilon) \mu^2 \]

3. Note: there is an error in the skewness result given in Eq. (10.122). The solution presented here derives the correct result.

4. The skewness is given by:

\[ \gamma(x) = \frac{E\{(x - \bar{x})^3\}}{[E\{(x - \bar{x})^2\}]^{3/2}} \]

To compute the numerator, note that:

\[ E\{(x - \bar{x})^3\} = E\{x^3 - 3x\bar{x} + 3x\bar{x}^2 - \bar{x}^3\} \]

\[ = E\{x^3\} - 3\bar{x}E\{x^2\} + 3\bar{x}^2 E\{x\} - \bar{x}^3 \]

\[ = E\{x^3\} - 3\bar{x}^2 \bar{x}^2 + 3\bar{x}^3 - 3\bar{x}^3 - \bar{x}^3 \]

\[ = E\{x^3\} - 3\bar{x}^2 \bar{x}^2 - \bar{x}^3 \]

\[ = E\{x^3\} - 3\epsilon \mu (\sigma^2 + \epsilon (1 - \epsilon) \mu^2) - \epsilon^3 \mu^3 \]
5. From Eq. (10.116) for the general two-component Gaussian mixture model, note that the slippage model corresponds to \( \alpha = \epsilon, \mu_1 = \mu, \mu_2 = 0, \) and \( \sigma_1 = \sigma_2 = \sigma \). Thus:

\[
E\{x^3\} = \mu_2(\mu_2^3 + 3\sigma_2^3) + 3\epsilon(\mu_1\sigma_1^3 - \mu_2\sigma_2^3) + \alpha(\mu_1^3 - \mu_2^3)
= 3\epsilon\mu\sigma^2 + \epsilon\mu^3
\]

6. Combining these results yields:

\[
E\{(x - \bar{x})^3\} = 3\epsilon\mu\sigma^2 + \epsilon\mu^3 - 3\epsilon\mu\sigma^2 - 3\epsilon^2(1 - \epsilon)\mu^3 - \epsilon^3\mu^3
= \epsilon\mu^3[1 - 3\epsilon(1 - \epsilon) - \epsilon^2]
= \epsilon\mu^3[1 - 3\epsilon + 3\epsilon^2 - \epsilon^3]
= \epsilon\mu^3[1 - 3\epsilon + 2\epsilon^2]
= \epsilon(1 - \epsilon)(1 - 2\epsilon)\mu^3
\]

7. Thus, the corrected skewness result is:

\[
\gamma(x) = \frac{\epsilon(1 - \epsilon)(1 - 2\epsilon)\mu^3}{\sigma^4 + \epsilon(1 - \epsilon)\mu^2}^{3/2}
\]

Note that if \( \epsilon = 0 \) or \( \mu = 0 \), the slippage model reduces to the uncontaminated Gaussian distribution \( N(0, \sigma^2) \) and \( \gamma(x) \) vanishes, as it should. Also, note that for \( 0 < \epsilon < 1/2 \), the skewness has the same sign as \( \mu \), indicating that the slippage model is skewed in the direction of the contamination \( \mu \). For \( \epsilon = 1/2 \), the distribution is bimodal and symmetric, implying zero skewness. For \( 1/2 < \epsilon < 1 \), the roles of the “normal” and “contaminating” terms reverses and so does the sign of the skewness. For \( \epsilon = 1 \), the model reduces to \( N(\mu, \sigma^2) \) and the skewness again vanishes.

Exercise 16:

Suppose \( r = x/y \) where the distribution of \( x \) is arbitrary and \( y \) is normally distributed. If \( x \) and \( y \) are independent, show that \( r \) does not have a finite second moment. (Hint: note that \( E\{r^2\} = E\{x^2/y^2\} \).

Solution:

1. From the hint, note that:

\[
E\{r^2\} = E\{x^2y^{-2}\} = E\{x^2\}E\{y^{-2}\} = (\sigma_x^2 + \mu_x^2)E\{y^{-2}\}
\]

2. To evaluate \( E\{y^{-2}\} \), define \( z = 1/y \) and note that \( E\{y^{-2}\} \) is the second moment of the transformed variable \( z \). A detailed discussion of the non-existence of the moments of the reciprocal Gaussian distribution exhibited by \( z \) is given in Section 12.7.1.
Exercise 17:

Derive Eq. (10.128) for the variance of $x$ in terms of its conditional mean and conditional variance:

$$\text{var}\{x\} = E\{\text{var}\{x|\theta\}\} + \text{var}\{E\{x|\theta\}\}.$$  

Solution:

1. First, note that:

$$\text{var}\{x|\theta\} = E[(x - E\{x|\theta\})^2|\theta]$$

$$= E(x^2 - 2xE\{x|\theta\} + [E\{x|\theta\}]^2|\theta)$$

$$= E(x^2|\theta) - 2E\{x|\theta\}E\{x|\theta\} + [E\{x|\theta\}]^2$$

$$= E(x^2|\theta) - [E\{x|\theta\}]^2$$

2. The first term on the right-hand side of the variance expression given in the problem statement is:

$$E_{\theta}\{\text{var}\{x|\theta\}\} = E\{x^2\} - E_{\theta}\{[E\{x|\theta\}]^2\} \quad \text{(by Eq. (10.47)}$$

3. The second term on the right-hand side of the variance expression is:

$$\text{var}_{\theta}\{E\{x|\theta\}\} = E_{\theta}\{(E\{x|\theta\} - E_{\theta}\{E\{x|\theta\}\})^2\}$$

$$= E_{\theta}\{(E\{x|\theta\} - E\{x\})^2\} \quad \text{(again by Eq. (10.47))}$$

$$= E_{\theta}\{[E\{x|\theta\}]^2 - 2E\{x|\theta\}E\{x\} + [E\{x\}]^2\}$$

$$= E_{\theta}\{[E\{x|\theta\}]^2\} - 2E\{x\}E_{\theta}\{E\{x|\theta\}\} + [E\{x\}]^2$$

$$= E_{\theta}\{[E\{x|\theta\}]^2\} - 2E\{x\}^2 + [E\{x\}]^2$$

$$= E_{\theta}\{[E\{x|\theta\}]^2\} - [E\{x\}]^2$$

4. Combining these results yields:

$$E_{\theta}\{\text{var}\{x|\theta\}\} + \text{var}_{\theta}\{E\{x|\theta\}\} = E\{x^2\} - E_{\theta}\{[E\{x|\theta\}]^2\} + E_{\theta}\{[E\{x|\theta\}]^2\} - [E\{x\}]^2$$

$$= E\{x^2\} - [E\{x\}]^2$$

$$= \text{var}\{x\}$$

Chapter 11

Exercise 1:

Show that the restricted solution $\theta_{RW}$ of the weighted least squares problem discussed in Sec. 11.3.2 satisfies the constraint $C\theta_{RW} = h$.

Solution:
1. The expression for $\theta_{RW}$ is given in Eq. (11.35):

$$\theta_{RW} = \tilde{\theta} + (X^TWX)^{-1}C^T[C(X^TWX)^{-1}C^T]^{-1}(h - C\tilde{\theta})$$

where $\tilde{\theta}$ is the unconstrained weighted least square estimator of $\theta$.

2. By direct computation:

$$C\theta_{RW} = C\tilde{\theta} + C(X^TWX)^{-1}C^T[C(X^TWX)^{-1}C^T]^{-1}(h - C\tilde{\theta})$$

$$= C\tilde{\theta} + h - C\tilde{\theta}$$

$$= h$$

Exercise 2—Note: the equation numbers listed in the original problem statement are not correct; the numbers given here have been corrected.

Show that the condition (11.41) for the M-estimator $\hat{\theta}$ discussed in Sec. 11.4.2 is equivalent to the condition (11.42).

Solution:

1. Note that the matrices $X$ and $W$ and the vector $r$ are defined as:

$$X = \begin{bmatrix}
\phi_1(x_1) & \phi_2(x_1) & \cdots & \phi_p(x_1) \\
\phi_1(x_2) & \phi_2(x_2) & \cdots & \phi_p(x_2) \\
\vdots & \vdots & \ddots & \vdots \\
\phi_1(x_N) & \phi_2(x_N) & \cdots & \phi_p(x_N)
\end{bmatrix} \quad \text{(Eq. (11.11))}$$

$$W = \begin{bmatrix}
w_1 & 0 & \cdots & 0 \\
0 & w_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & w_N
\end{bmatrix}$$

$$r = \begin{bmatrix}r_1 \\
r_2 \\
\vdots \\
r_N\end{bmatrix}$$

2. Combining these results yields:

$$Wr = \begin{bmatrix}w_1r_1 \\
w_2r_2 \\
\vdots \\
wNr_N\end{bmatrix}$$

$$\Rightarrow X^TWr = \begin{bmatrix}
\phi_1(x_1) & \phi_1(x_2) & \cdots & \phi_1(x_N) \\
\phi_2(x_1) & \phi_2(x_2) & \cdots & \phi_2(x_N) \\
\vdots & \vdots & \ddots & \vdots \\
\phi_p(x_1) & \phi_p(x_2) & \cdots & \phi_p(x_N)
\end{bmatrix} \begin{bmatrix}w_1r_1 \\
w_2r_2 \\
\vdots \\
wNr_N\end{bmatrix}$$
\[
\begin{bmatrix}
\sum_{k=1}^{N} \phi_1(x_k) w_k r_k \\
\sum_{k=1}^{N} \phi_2(x_k) w_k r_k \\
\vdots \\
\sum_{k=1}^{N} \phi_p(x_k) w_k r_k
\end{bmatrix}
\]

3. It follows from this result that the \(i^{th}\) component of the vector equation \(X^T W R = 0\) corresponds to the corresponding element of Eq. (11.41).

Chapter 12

Exercise 1—NOTE: this problem is poorly formulated as stated.
This solution is for a corrected version.

Using the same format as the 25 pig transforms considered in Sec. 12.1.1, generate the corresponding 25 plots for the following circle:

\[(x - 0.5)^2 + (y - 0.5)^2 = 0.4^2\]

Solution:

1. The original formulation of this problem is inappropriate because both the \(x\) and \(y\) values in the original problem statement can be non-positive, while the square root and log transformations are only appropriate to positive values. The reformulation given here restricts both \(x\) and \(y\) to the interval \([0.1, 0.9]\); over this range, all five of the transformations considered are well-behaved.

2. Unfortunately, \(R\) does not generate useable plots in a \(5 \times 5\) array, so the requested plots must be displayed in a different format. Here, all of the required plots are displayed in the sequence of four \(3 \times 3\) plots shown in Figs. 3 through 6.

Exercise 2:

A model that arises frequently is the thermally activated model:

\[x(T) = x_0 e^{-E/kT},\]

where \(x_0\) and \(E\) are parameters to be determined from measurements of the observed response \(x(T)\) vs. the absolute temperature \(T\) and \(k\) is Boltzmann’s constant. Derive a linearizing transformation for this model.

Solution:
Figure 3: Plots of circles transformed by square root, linear, and square transformations applied to both the x and y axes.

1. Taking the logarithm of $x(T)$ yields:
   \[ \ln x(T) = \ln x_0 - \frac{E}{kT} \]

2. To linearize this equation, define:
   \[ y = \ln x \]
   \[ s = \frac{1}{T} \]
   \[ \Rightarrow y = a + bs \]

   where $a = \ln x_0$ and $b = -E/k$.

Exercise 3:
Consider the triangular density:

\[ p(x) = \begin{cases} 
0 & x < -1, \\
C(x+1) & -1 \leq x < 0, \\
C(1-x) & 0 \leq x < 1, \\
0 & x \geq 1. 
\end{cases} \]

a. Determine the normalization constant \( C \).

b. Derive an expression for the cumulative distribution function \( F(x) \) for this density.

c. Derive an expression for \( F^{-1}(x) \).

d. Use this result to generate triangularly distributed random numbers.
Figure 5: Plots of circles with the x axis transformed by logarithm, linear, and exponential transformations and the y axes transformed by square root, linear and square transformations.

Solution:

1. To address both parts a and b of this problem, first compute:

   \[ F(x) = \int_{-\infty}^{x} p(z)dz = \int_{-1}^{x} p(z)dz \]

   For \( x \) in the range \( -1 \leq x \leq 0 \), this integral is:

   \[ F(x) = \int_{-1}^{x} C(z + 1)dz \]

   \[ = \left[ C \left( \frac{z^2}{2} + z \right) \right]_{-1}^{x} \]

63
Figure 6: Plots of circles with the x axis transformed by square root, linear and square transformations and the y axes transformed by logarithm, linear, and exponential transformations.

\[
\begin{align*}
F(x) &= \int_{-1}^{0} p(z)dz + \int_{0}^{x} p(z)dz \\
&= C + C \int_{0}^{x} (1 - z)dz \\
&= \frac{C}{2} + C \int_{0}^{x} (1 - z)dz
\end{align*}
\]
\[
C_2 + C \left[ \frac{z - \frac{x^2}{2}}{2} \right]_0 \\
C_2 + C \left( x - \frac{x^2}{2} \right) \\
C \left( \frac{1 + 2x - x^2}{2} \right)
\]

Thus, at \( x = 1 \), we have:

\[ F(1) = C \left( \frac{1 + 2 - 1}{2} \right) = C \]

Since \( F(1) = 1 \) for normalization, this implies \( C = 1 \) and the cumulative distribution is:

\[
F(x) = \begin{cases} 
0 & x \leq -1 \\
\frac{(x+1)^2}{2} & -1 \leq x < 0 \\
\frac{1+2x-x^2}{2} & 0 \leq x < 1 \\
1 & x \geq 1
\end{cases}
\]

2. To compute the inverse of \( F(x) \), consider the equation \( F(x) = p \). First, note that for \(-1 \leq x \leq 0\), the corresponding range of \( p \) is \( 0 \leq p \leq 1/2 \) and the relationship is:

\[
\frac{(x+1)^2}{2} = p \quad \Rightarrow \quad (x+1)^2 = 2p \\
\Rightarrow \quad x + 1 = \sqrt{2p} \\
\Rightarrow \quad x = \sqrt{2p} - 1
\]

For \( 0 \leq x \leq 1 \), the corresponding range of \( p \) is \( 1/2 \leq p \leq 1 \) and the relationship is:

\[
\frac{1 + 2x - x^2}{2} = p \quad \Rightarrow \quad 1 + 2x - x^2 = 2p \\
\Rightarrow \quad x^2 - 2x + 2p - 1 = 0 \\
\Rightarrow \quad x = \frac{2 \pm \sqrt{4 - 4(2p - 1)}}{2} = 1 \pm \sqrt{2(p-1)}
\]

Note that since the range of \( x \) here is \( 0 \leq x \leq 1 \), only the negative sign in the solution is consistent, so we have:

\[ x = 1 - \sqrt{2(1-p)} \text{ for } 1/2 \leq p \leq 1 \]

3. Combining these results gives the final expression for \( F^{-1}(p) \) as:

\[
F^{-1}(p) = \begin{cases} 
\sqrt{2p} - 1 & 0 \leq p \leq 1/2 \\
1 - \sqrt{2(1-p)} & 1/2 \leq p \leq 1
\end{cases}
\]
4. To generate triangularly distributed random numbers, first generate a sequence \( \{u_n\} \) that is uniformly distributed on \([0, 1]\) and then apply the transformation \( t_n = F^{-1}(u_n) \) to obtain the desired sequence \( \{t_n\} \).

Exercise 4:

The logarithmic transformation is widely used. Suppose \( x \) is a random variable with density \( p(x) \) that is nonzero only for \( x \geq 0 \).

a. Derive a general expression for the density \( \phi(z) \) of \( z = \ln x \).

b. Apply this result to the beta distribution with shape parameter \( q = 1 \) (see Chapter 4).

c. The resulting distribution is well known: what is it?

d. Apply this transformation result to the Pareto type I distribution, \( P(f)(k, a) \).

e. How does this distribution differ from that considered in parts [b] and [c]?

Solution:

1. For \( T(z) = \ln z \), it follows that:

\[
\begin{align*}
T'(z) &= 1/z \\
T^{-1}(z) &= e^z
\end{align*}
\]

Thus, the general transformed result is:

\[
\phi(z) = \frac{p(T^{-1}(z))}{|T'(T^{-1}(z))|} = \frac{pe^z}{|1/e^z|} = e^z p(e^z)
\]

2. The beta distribution with \( q = 1 \) has the density:

\[
p(x) = \frac{\Gamma(p+1)}{\Gamma(p)\Gamma(1)} x^{p-1} = px^{p-1}
\]

The transformed result is therefore:

\[
\phi(z) = e^z p(e^z) = pe^z [e^z]^{p-1} = pe^{pz}
\]

Note that the range of \( z \) here is \(-\infty < z < 0\). Re-writing this density in terms of \( \beta = 1/p \), this density corresponds to the exponential distribution, reflected about zero.
3. For the Pareto distribution:

\[ p(x) = \frac{ak^a}{x^{a+1}} \text{ for } x \geq k \]

Applying the log transformation to this distribution gives:

\[ \phi(z) = e^z p(e^z) = \frac{ak^a e^z}{e^{z(a+1)}} = ak^a e^{-az} \]

4. This distribution differs from the previous result in its domain:

- parts b & c: \(-\infty < z < 0\)
- parts d & e: \(k \leq z < \infty\)

**Exercise 5:**

Determine the density \(\phi(z)\) that results when the angular transformation discussed in Sec. 12.3.3 is applied to the uniform distribution on \([0, 1]\). Plot the resulting density.

**Solution:**

1. The angular transformation is:

\[ T(x) = \arcsin(2x - 1) \]

To determine the inverse function \(T^{-1}(z)\), solve:

\[ T(x) = z \Rightarrow \arcsin(2x - 1) = z \]
\[ \Rightarrow 2x - 1 = \sin z \]
\[ \Rightarrow x = \frac{1 + \sin z}{2} \]

The derivative is given by:

\[ T'(x) = \frac{d}{dx} \arcsin(2x - 1) \]
\[ = \frac{1}{\sqrt{1 - (2x - 1)^2}} \cdot \frac{d}{dx}(2x - 1) \]
\[ = \frac{1}{\sqrt{1 - 4x^2 + 4x - 1}} \cdot 2 \]
\[ = \frac{2}{\sqrt{4x - 4x^2}} \]
\[ = \frac{1}{\sqrt{x(1-x)}} \]

67
2. The transformed density is:

\[ \phi(z) = \frac{p(T^{-1}(z))}{|T'(T^{-1}(z))|} \]

The denominator in this expression is given by:

\[ T'(T^{-1}(z)) = T' \left( \frac{1 + \sin z}{2} \right) \]

\[ = \left[ \frac{1 + \sin z}{2} \left( 1 - \frac{1 + \sin z}{2} \right) \right]^{-1/2} \]

\[ = \left[ \frac{1 + \sin z}{2} \cdot \frac{1 - \sin z}{2} \right]^{-1/2} \]

\[ = \left[ \frac{1 - \sin^2 z}{4} \right]^{-1/2} \]

\[ = \left[ \frac{\cos^2 z}{4} \right]^{-1/2} \]

\[ = \frac{2}{\cos z} \]

Thus, the transformed density is:

\[ \phi(z) = \frac{p \left( \frac{1 + \sin z}{2} \right)}{\left| \frac{\cos z}{2} \right|} \]

\[ = \frac{|\cos z|}{2} p \left( \frac{1 + \sin z}{2} \right) \]

3. For the uniform distribution, we have:

\[ p(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

\[ \Rightarrow p \left( \frac{1 + \sin z}{2} \right) = \begin{cases} 1 & 0 \leq \frac{1 + \sin z}{2} \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

\[ = \begin{cases} 1 & 0 \leq 1 + \sin z \leq 2 \\ 0 & \text{otherwise} \end{cases} \]

\[ = \begin{cases} 1 & -1 \leq \sin z \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

Since \(-1 \leq \sin z \leq 1\) for all \(z\), it follows that, for all \(z\):

\[ p \left( \frac{1 + \sin z}{2} \right) = 1 \Rightarrow \phi(z) = \frac{|\cos z|}{2} \]

68
4. Note that the domain of the angular transformation is $[0, 1]$, which is mapped into the interval $[-\pi/2, \pi/2]$. Over this interval, $\cos z \geq 0$, so the transformed density simplifies to:

$$\phi(z) = \frac{1}{2} \cos z \text{ for } -\pi/2 \leq z \leq \pi/2$$

A plot of this transformed density function is shown in Fig. 7.

**Exercise 6:**

The *Pareto type II distribution* or *Lomax distribution* is defined by the following density:

$$p(x) = \frac{a}{(x+1)^{a+1}}, \quad x > 0, \ a > 0.$$
a. Consider the transformation \( z = \ln(x) \) and derive an expression for the density of \( z \). Plot this density.

c. Consider the transformation \( r = 1/x \) and derive an expression for the density of \( r \). Plot this density.

Solution:

1. For the transformation \( z = \ln(x) \), it follows from the results of Exercise 4a that:

\[
\phi(z) = e^z p(e^z) = \frac{ae^z}{(e^z + 1)^{a+1}}
\]

Plots of this density function are shown in Fig. 8 for \( a = 0.5, 1.0 \) and \( 2.0 \).

2. For the reciprocal transformation, it follows from Eq. (12.67) that:

\[
\phi(z) = \frac{p(1/z)}{z^2} = \frac{a}{z^2(1/z + 1)^{a+1}} = \frac{az^{a-1}}{(z + 1)^{a+1}}
\]

Plots of this density function are shown in Fig. 9 for \( a = 0.5, 1.0 \) and \( 2.0 \).

Chapter 13

Exercise 1:

Suppose \( \{x_k\} \) and \( \{y_k\} \) are binary sequences, where \( x_k = 1 \) with probability \( p \) and \( y_k = 1 \) with probability \( q \). Define the population odds ratio as the ratio of the probabilities:

\[
O_{xy} = \frac{p_{00}p_{11}}{p_{01}p_{10}},
\]

where \( p_{ij} \) denotes the joint probability that \( x_k = i \) and \( y_k = j \) for \( i, j = 0 \) or \( 1 \). Show that \( O_{xy} = 1 \) when \( \{x_k\} \) and \( \{y_k\} \) are statistically independent.

Solution:

1. Under independence, the joint probability \( p_{ij} \) may be factored as:

\[
p_{ij} = p_i q_j
\]

where \( p_i \) is the probability that \( x \) assumes its \( i^{th} \) value and \( q_j \) is the probability that \( y \) assumes its \( j^{th} \) value.
2. Thus, the population odds ratio under independence is:

\[ O_{xy} = \frac{p_{00}p_{11}}{p_{01}p_{10}} = \frac{(p_0q_0)(p_1q_1)}{(p_0q_1)(p_1q_0)} = \frac{p_0p_1q_0q_1}{p_0p_1q_0q_1} = 1 \]

**Exercise 2:**

Derive the sum of squares partition from Eq. (13.11):

\[ \sum_{i=1}^{C} \sum_{j=1}^{n} (y_{ij} - \bar{y}_.)^2 = \sum_{i=1}^{C} (\bar{y}_i - \bar{y}_.)^2 + \sum_{i=1}^{C} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2. \]
Solution:

1. First, re-write the difference $y_{ij} - \bar{y}_i$ as:

   \[ y_{ij} - \bar{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y}_.) \]

   \[ \Rightarrow (y_{ij} - \bar{y}_i)^2 = (y_{ij} - \bar{y}_i)^2 + 2(y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y}_.) + (\bar{y}_i - \bar{y}_.)^2 \]

2. From this, it follows that the total sum of squares $SS_T$ may be written as:

   \[ SS_T = \sum_{i=1}^{C} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 \]

   \[ = \sum_{i=1}^{C} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 + 2 \sum_{i=1}^{C} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y}_.) + \sum_{i=1}^{C} \sum_{j=1}^{n} (\bar{y}_i - \bar{y}_.)^2 \]
\[
SS + 2 \sum_{i=1}^{C} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)(\bar{y}_i - \bar{y}) + n \sum_{i=1}^{C} (\bar{y}_i - \bar{y})^2
\]

\[
= SS + SS_B + 2 \sum_{i=1}^{C} (\bar{y}_i - \bar{y}) \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)
\]

3. To simplify this result, note that from Eqs. (13.6) and (13.10):

\[
y_{ij} - \bar{y}_i = \mu + \alpha_i + e_{ij} - \mu - \alpha_i - \frac{1}{n} \sum_{k=1}^{n} e_{ik}
\]

\[
= e_{ij} - \frac{1}{n} \sum_{k=1}^{n} e_{ik}
\]

\[
\Rightarrow \sum_{j=1}^{n} (y_{ij} - \bar{y}_i) = \sum_{j=1}^{n} e_{ij} - \sum_{j=1}^{n} \left[ \frac{1}{n} \sum_{k=1}^{n} e_{ik} \right]
\]

\[
= \sum_{j=1}^{n} e_{ij} - n \left[ \frac{1}{n} \sum_{k=1}^{n} e_{ik} \right]
\]

\[
= \sum_{j=1}^{n} e_{ij} - \sum_{k=1}^{n} e_{ik}
\]

\[
= 0
\]

Thus, the above result reduces to \(SS_T = SS + SS_B\) as claimed.

Exercise 3:

Derive the expectations for the between groups and within groups sums of squares, \(SS_B\) and \(SS_W\), given in Eq. (13.12):

\[
E\{SS_B\} = n \sum_{i=1}^{C} \alpha_i^2 + (C - 1)\sigma_e^2,
\]

\[
E\{SS_W\} = C(n - 1)\sigma_e^2.
\]

Solution:

1. It is simpler to start with the second result first since that builds on the solution of Exercise 2:

\[
E\{SS_W\} = E \left\{ \sum_{i=1}^{C} \sum_{j=1}^{n} (\bar{y}_i - \bar{y})^2 \right\}
\]

\[
= \sum_{i=1}^{C} \sum_{j=1}^{n} E\{(y_{ij} - \bar{y}_i)^2\}
\]

73
From Exercise 2, we have:

\[ y_{ij} - \bar{y}_i = e_{ij} - \bar{\epsilon}_i \]

where \( \bar{\epsilon}_i = \frac{1}{n} \sum_{k=1}^{n} e_{ik} \)

\[ \Rightarrow (y_{ij} - \bar{y}_i)^2 = e_{ij}^2 - 2e_{ij}\bar{\epsilon}_i + \bar{\epsilon}_i^2 \]

2. Since \( E\{e_{ij}\} = 0 \) and \( \text{var}\{e_{ij}\} = E\{e_{ij}^2\} = \sigma_e^2 \), it follows that:

\[ E\{(y_{ij} - \bar{y}_i)^2\} = E\{e_{ij}^2\} - 2E\{e_{ij}\bar{\epsilon}_i\} + E\{\bar{\epsilon}_i^2\} \]

\[ = \sigma_e^2 - 2E\left\{\frac{1}{n} \sum_{k=1}^{n} e_{ik}e_{ij}\right\} + E\left\{\frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} e_{ik}e_{ij}\right\} \]

\[ = \sigma_e^2 - \frac{2}{n} \sum_{k=1}^{n} E\{e_{ij}e_{ik}\} + \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} E\{e_{ij}e_{ik}\} \]

\[ = \sigma_e^2 - \frac{2}{n} \sigma_e^2 + \frac{1}{n^2} \sum_{j=1}^{n} \sigma_e^2 \]

\[ = \sigma_e^2 - \frac{2}{n} \sigma_e^2 + \frac{1}{n} \sigma_e^2 \]

\[ = \sigma_e^2 - \frac{1}{n} \sigma_e^2 \]

3. Thus, for the sum of squares within, we have:

\[ E\{SS_W\} = \sum_{i=1}^{C} \sum_{j=1}^{n} \left[ \sigma_e^2 - \frac{1}{n} \sigma_e^2 \right] \]

\[ = \sum_{i=1}^{C} [n\sigma_e^2 - \sigma_e^2] \]

\[ = C(n-1)\sigma_e^2 \]

4. For the between sum of squares, we have:

\[ E\{SS_B\} = E \left\{ n \sum_{i=1}^{C} (\bar{y}_i - \bar{\bar{y}})^2 \right\} \]

\[ = n \sum_{i=1}^{C} E\{(\bar{y}_i - \bar{\bar{y}})^2\} \]

5. From Eqs. (13.9) and (13.10):

\[ \bar{y}_i - \bar{\bar{y}} = \mu + \alpha_i + \bar{\epsilon}_i - \mu - \frac{1}{C} \sum_{\ell=1}^{C} \bar{\epsilon}_\ell \]

74
\[
\begin{align*}
\alpha_i + \delta_i \\
\text{where } \delta_i = \bar{e}_i - \frac{1}{c} \sum_{\ell=1}^{c} \bar{e}_\ell \\
\Rightarrow (\bar{y}_i - \bar{y}_i)^2 = (\alpha_i + \delta_i)^2 \\
= \alpha_i^2 + 2\delta_i\alpha_i + \delta_i^2
\end{align*}
\]

Taking the expectation of this result gives:
\[E\{(\bar{y}_i - \bar{y}_i)^2\} = \alpha_i^2 + 2\alpha_i E\{\delta_i\} + E\{\delta_i^2\}\]

To simplify this result, note that:
\[E\{\bar{e}_\ell\} = E\left\{\frac{1}{n} \sum_{k=1}^{n} e_{tk}\right\} = \frac{1}{n} \sum_{k=1}^{n} E\{e_{tk}\} = 0\]

Thus:
\[E\{(\bar{y}_i - \bar{y}_i)^2\} = \alpha_i^2 + E\{\delta_i^2\}\]

6. For this expectation:
\[E\{\delta_i^2\} = E\left\{\bar{e}_i^2 - \frac{2\bar{e}_i}{c} \sum_{\ell=1}^{c} \bar{e}_\ell + \frac{1}{c^2} \sum_{\ell=1}^{c} \sum_{m=1}^{c} \bar{e}_\ell \bar{e}_m\right\} = E\{\bar{e}_i^2\} - \frac{2}{c} \sum_{\ell=1}^{c} E\{\bar{e}_\ell \bar{e}_i\} + \frac{1}{c^2} \sum_{\ell=1}^{c} \sum_{m=1}^{c} E\{\bar{e}_\ell \bar{e}_m\}\]

7. For arbitrary \(\ell\) and \(m\), note that:
\[E\{\bar{e}_\ell \bar{e}_m\} = E\left\{\frac{1}{n} \sum_{j=1}^{n} e_{\ell j} \cdot \frac{1}{n} \sum_{k=1}^{n} e_{mk}\right\} = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} E\{e_{\ell j} e_{mk}\}\]

Under the assumption that the errors \(\{e_{ij}\}\) are uncorrelated, it follows that \(E\{e_{\ell j} e_{mk}\} = 0\) unless \(\ell = m\) and \(j = k\). Thus:
\[E\{\bar{e}_\ell \bar{e}_m\} = \frac{1}{n^2} \sum_{j=1}^{n} E\{e_{\ell j} e_{mj}\}\]
\[= \frac{1}{n} E\{e_{\ell j} e_{mj}\}\]
\[= \begin{cases} 
\sigma_e^2/n & m = \ell \\
0 & m \neq \ell 
\end{cases}\]
From this result, we have:

\[
E\{\bar{e}_1^2\} = \frac{\sigma_e^2}{n}
\]
\[
\frac{2}{C} \sum_{\ell=1}^{C} E\{\bar{e}_\ell \bar{e}_\ell\} = \frac{2\sigma_e^2}{nC}
\]
\[
\frac{1}{C^2} \sum_{\ell=1}^{C} \sum_{m=1}^{C} E\{\bar{e}_\ell \bar{e}_m\} = \frac{1}{C^2} \sum_{\ell=1}^{C} E\{\bar{e}_\ell^2\} = \frac{\sigma_e^2}{nC}
\]

Combining these results yields:

\[
E\{\delta_i^2\} = \frac{\sigma_e^2}{n} - \frac{2\sigma_e^2}{nC} + \frac{\sigma_e^2}{nC}
\]
\[
= \frac{\sigma_e^2}{n} \left(1 - \frac{1}{C}\right)
\]
\[
= \sigma_e^2 (C - 1)
\]

This leads to the desired final result, which is:

\[
E\{SS_B\} = n \sum_{i=1}^{C} \alpha_i^2 + nC \cdot \frac{\sigma_e^2 (C - 1)}{nC}
\]
\[
= n \sum_{i=1}^{C} \alpha_i^2 + (C - 1)\sigma_e^2
\]

**Exercise 4:**

Show that the \(C\) components of the least squares solution \(\hat{\theta}\) given in Sec. 13.3.2 are identical with the estimators for \(\mu\) and \(\alpha_1, \alpha_2, \ldots, \alpha_{C-1}\) given in Sec. 13.3 for the ANOVA model.

**Solution:**

1. The key to the equivalence to be shown here lies in recognizing that the observation \(y_k\) in the regression formulation corresponds to the ANOVA observation \(y_{ij}\) if \(k\) defines the \(j^{th}\) data record for which \(x_k = v_i\), which implies \(b_{i,k} = 1\). Thus, from Eq. (13.18),

\[
y_k = y_{ij} \Rightarrow y_{ij} = \mu + \alpha_i + e_k
\]
\[
= \mu + \alpha_i + e_{ij}
\]

which corresponds exactly to Eq. (13.6).
2. Thus, the $C$ components of the regression parameter vector $\theta$ are $\mu$ and $\alpha_1$ through $\alpha_{C-1}$. The remaining ANOVA model parameter $\alpha_C$ follows directly from the constraint in Eq. (13.7):

$$\sum_{i=1}^{C} \alpha_i = 0 \Rightarrow \alpha_C = -\sum_{i=1}^{C-1} \alpha_i$$

Exercise 5:

Show that by defining $D_{i,k}$ as in Eq. (13.31), $y_k$ can be expressed as in Eq. (13.20) for the unbalanced one-way ANOVA problem, from which it follows that the regression formulation of this problem has the same form as the balanced one-way ANOVA problem.

Solution:

1. As in the balanced ANOVA problem, the first step in deriving the regression formulation is to write Eq. (13.18):

$$y_k = \mu + \sum_{i=1}^{C} \alpha_i b_{i,k} + e_k$$

with $b_{i,k}$ defined as in Eq. (13.17).

2. By Eq. (13.26), we have:

$$\sum_{i=1}^{C} n_i \alpha_i = 0 \Rightarrow n_C \alpha_C = -\sum_{i=1}^{C-1} n_i \alpha_i$$

$$\Rightarrow \alpha_C = -\sum_{i=1}^{C-1} \frac{n_i}{n_C} \alpha_i$$

Substituting this result into Eq. (13.18) yields:

$$y_k = \mu + \sum_{i=1}^{C-1} \alpha_i b_{i,k} + \alpha_C b_{C,k} + e_k$$

$$= \mu + \sum_{i=1}^{C-1} \alpha_i b_{i,k} - \sum_{i=1}^{C-1} \frac{n_i}{n_C} b_{i,k} + e_k$$

$$= \mu + \sum_{i=1}^{C-1} \alpha_i \left(b_{i,k} - \frac{n_i}{n_C} b_{C,k}\right) + e_k$$

3. Defining $D_{i,k} = b_{i,k} - (n_i/n_C)b_{C,k}$ as in Eq. (13.31) thus yields Eq. (13.20), just as in the balanced ANOVA problem.
Exercise 6:

a. Show that the Gaussian distribution belongs to the exponential family defined by Eq. (13.46); that is, for \( y \sim \mathcal{N}(\mu, \sigma) \) determine the constants \( \theta \) and \( \phi \) and the functions \( a(\phi) \), \( b(\theta) \), and \( c(y, \phi) \) appearing in this expression.

b. For the Gaussian case with the cannonical link \( g(x) = x \), show that the deviance defined in Eq. (13.52) is equal to the residual sum of squares:

\[
\text{RSS} = \sum_{k=1}^{N} (y_k - x_k^T \beta)^2.
\]

Solution:

1. The exponential family of distributions is defined by densities of the following form:

\[
f(y; \theta, \phi) = \exp\left\{ y\theta - b(\theta) \frac{\phi}{a(\phi)} + c(y, \phi) \right\}
\]

2. The Gaussian density is:

\[
p(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2} \left( \frac{y - \mu}{\sigma} \right)^2 \right\}
\]

\[
= \exp\left\{ -\frac{y^2 - 2\mu y + \mu^2}{2\sigma^2} - \ln[\sigma\sqrt{2\pi}] \right\}
\]

\[
= \exp\left\{ \frac{y\mu - \mu^2/2}{\sigma^2} + \left[ -\frac{y^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2} \right] \right\}
\]

These expressions are the same if we make the following definitions:

\[
\theta = \mu \\
b(\theta) = \theta^2/2 \\
\phi = \sigma^2 \\
a(\phi) = \phi \\
c(y, \phi) = -\frac{y^2}{2\phi} - \frac{1}{2} \ln 2\pi\phi
\]

3. The deviance is defined as:

\[
D(y_k; x_k^T \beta) = 2 \sum_{k=1}^{N} w_k (y_k [g(y_k) - x_k^T \beta] - b(g(y_k)) + b(x_k^T \beta))
\]
Here, we consider the uniformly weighted case $w_k = 1$ for all $k$ and the identity link $g(x) = x$, so the deviance expression simplifies to:

$$D(y_k; x_k^T \beta) = 2 \sum_{k=1}^{N} (y_k[y_k - x_k^T \beta] - y_k^2/2 + (x_k^T \beta)^2/2)$$

$$= 2 \sum_{k=1}^{N} \left[ y_k^2 - x_k^T \beta y_k - \frac{1}{2} y_k^2 + \frac{1}{2} (x_k^T \beta)^2 \right]$$

$$= \sum_{k=1}^{N} [y_k^2 - 2x_k^T \beta + (x_k^T \beta)^2]$$

$$= \sum_{k=1}^{N} (y_k - x_k^T \beta)^2$$

**Exercise 7:**

Show that the variance of the two-component mixture of Poisson distributions discussed in Sec. 13.6.1 is given by:

$$\text{var}\{n\} = \mu + \pi(1 - \pi)(\lambda_1 - \lambda_2)^2,$$

where $\mu = \pi\lambda_1 + (1 - \pi)\lambda_2$ is the mean of the mixture distribution, $\pi$ is the mixing parameter, and $\lambda_1$ and $\lambda_2$ are the parameters associated with the component Poisson distributions on which the mixture is based.

**Solution:**

1. The two-component Poisson mixture density is defined by Eq. (13.66) as:

$$\mathcal{P}\{n\} = \pi \mathcal{P}\{n; \lambda_1\} + (1 - \pi) \mathcal{P}\{n; \lambda_2\}$$

The mean of this distribution is:

$$\mu = E\{n\} = \pi E_1\{n\} + (1 - \pi) E_2\{n\}$$

$$= \pi\lambda_1 + (1 - \pi)\lambda_2$$

2. The variance is most easily computed as:

$$\text{var}\{n\} = E\{n^2\} - [E\{n\}]^2$$

where $E\{n^2\} = \pi E_1\{n^2\} + (1 - \pi) E_2\{n^2\}$

$$= \pi(\text{var}_1\{n\} + [E_1\{n\}]^2) + (1 - \pi)(\text{var}_2\{n\} + [E_2\{n\}]^2)$$

$$= \pi[\lambda_1 + \lambda_1^2] + (1 - \pi)[\lambda_2 + \lambda_2^2]$$

$$= \pi\lambda_1 + (1 - \pi)\lambda_2 + \pi\lambda_1^2 + (1 - \pi)\lambda_2^2$$

$$= \mu + \pi\lambda_1^2 + (1 - \pi)\lambda_2^2$$

$$(E\{n\})^2 = \mu^2 = \pi^2\lambda_1^2 + 2\pi(1 - \pi)\lambda_1\lambda_2 + (1 - \pi)^2\lambda_2^2$$
3. Combining these results yields:

\[
\begin{align*}
\text{var}\{n\} &= \mu + \pi \lambda_1^2 + (1 - \pi)\lambda_2^2 - \pi^2 \lambda_1^2 - 2\pi(1 - \pi)\lambda_1\lambda_2 - (1 - \pi)^2 \lambda_2^2 \\
&= \mu + [\pi - \pi^2]\lambda_1^2 - 2\pi(1 - \pi)\lambda_1\lambda_2 + [(1 - \pi) - (1 - \pi)^2]\lambda_2^2 \\
&= \mu + \pi(1 - \pi)\lambda_1^2 - 2\pi(1 - \pi)\lambda_1\lambda_2 + (1 - \pi)[1 - (1 - \pi)]\lambda_2^2 \\
&= \mu + (1 - \pi)\lambda_2^2 - 2\pi(1 - \pi)\lambda_1\lambda_2 + (1 - \pi)^2 \lambda_2^2 \\
&= \mu + \pi(1 - \pi)(\lambda_1 - \lambda_2)^2
\end{align*}
\]

Chapter 14

Exercise 1:

Suppose \( D \) is a sequence \( \{x_k\} \) of \( N \) random variables drawn from any continuous distribution and let \( R \) denote the sample maximum: \( R = x_{(N)} \). Consider a bootstrap sample \( \{x^*_k\} \) and let \( \tilde{R} = x^*_{(N)} \).

Show that:

\[
P\{\tilde{R} = R\} = 1 - (1 - 1/N)^N \to 1 - e^{-1} \approx 0.632 \quad \text{as} \quad N \to \infty.
\]

(Hint: note that \( P\{\tilde{R} \neq R\} = P\{x_{(N)} \text{ is not drawn from } D \text{ in } N \text{ draws}\}.\)

This result illustrates a difficulty that can arise with the bootstrap method since for any continuous distribution the probability that \( \tilde{R} \) has any specified value is zero.

Solution:

1. The probability of drawing the unique maximum element \( R = x_{(N)} \) is:

\[
P\{\tilde{R} = R\} = 1 - P\{\tilde{R} \neq R\} = 1 - P\{x_{(N)} \text{ not drawn in } N \text{ draws}\}
\]

2. The probability of not drawing \( x_{(N)} \) in \( N \) draws is:

\[
P\{x_{(N)} \text{ not drawn in } N \text{ draws}\} = P\{x_{(N)} \text{ not drawn on draw 1}\} \cdot P\{x_{(N)} \text{ not drawn on draw 2}\} \cdots P\{x_{(N)} \text{ not drawn on draw } N\}
\]

3. The probability that \( x_{(N)} \) is not drawn on draw \( j \) is:

\[
P\{x_{(N)} \text{ not drawn on draw } j\} = 1 - P\{x_{(N)} \text{ drawn on draw } j\} = 1 - 1/N
\]
4. Combining these results gives:

\[ P\{x(N) \text{ not drawn in } N \text{ draws}\} = (1 - 1/N) \cdot (1 - 1/N) \cdots (1 - 1/N) \]

\[ = (1 - 1/N)^N \]

\[ \Rightarrow P\{R = R\} = 1 - (1 - 1/N)^N \]

5. It is a standard result (see, for example, Abramowitz and Stegun [1], p. 70, no. 4.2.21) that:

\[ \lim_{m \to \infty} \left(1 + \frac{z}{m}\right)^m = e^z \]

Letting \( z = -1 \) then yields the result:

\[ \lim_{N \to \infty} \left(1 + \frac{1}{N}\right)^N = e^{-1} \]

Exercise 2:

Recall from Chapter 3 that the probability of drawing \( m \) black balls in a sample of size \( p \) from an urn containing \( Q \) white balls and \( N - Q \) black balls is given by the hypergeometric distribution. For a dataset of size \( N \) containing a single outlier, compute the probability that a subsample of size \( m \) contains this outlier. For a data sequence of length \( N = 100 \), how many of \( B = 1000 \) independent 10\% subsamples can be expected to contain the outlier? Of 50\% subsamples? Of 90\% subsamples?

Solution:

1. Given a dataset of size \( N \) containing \( Q \) outliers, the probability of drawing a sample of size \( m \) with \( k \) outliers is given by Eq. (3.17):

\[ P\{k \text{ outliers}\} = \binom{m}{k} \binom{N - m}{Q - k} / \binom{N}{Q} \]

2. For \( Q = k = 1 \), this probability reduces to:

\[ P\{\text{outlier in sample}\} = \binom{m}{1} \binom{N - m}{0} / \binom{N}{1} \]

\[ = \frac{m!}{1!} \cdot \frac{(N - 1)!}{N!} \]

\[ = \frac{m!}{(m - 1)!} \cdot \frac{1}{N!} \]

\[ = \frac{m}{N} \]
3. For $B$ repeated, independent draws, the result corresponds to a sequence of $B$ Bernoulli trials, each with probability $p = m/N$ that the sample contains the outlier. Thus, the expected number of $B$ independent subsamples containing the outlier is $N_{out} = pB = Bm/N$. For the specific cases considered here, we have:

- 10% subsample: $p = 0.10$, implying 100 subsamples with the outlier
- 50% subsample: $p = 0.50$, implying 500 subsamples with the outlier
- 90% subsample: $p = 0.90$, implying 900 subsamples with the outlier

Chapter 15

Exercise 1:

For the OLS linear regression model, the prediction vector $\hat{y}$ is given by $\hat{y} = Hy$ where $H$ is the hat matrix introduced in Sec. 15.3.1. Prove that the mean of the predictions is equal to the data mean, i.e.:

$$\frac{1}{N} \sum_{k=1}^{N} \hat{y}_k = \frac{1}{N} \sum_{k=1}^{N} y_k.$$

Solution:

1. Note that the OLS solution $\{\hat{y}_k\}$ minimizes:

$$J = \sum_{k=1}^{N} (\hat{y}_k - y_k)^2 = \sum_{k=1}^{N} r_k^2$$

2. Now, suppose the mean of the predictions differs from that of the data values:

$$\frac{1}{N} \sum_{k=1}^{N} \hat{y}_k \neq \frac{1}{N} \sum_{k=1}^{N} y_k$$

$$\Rightarrow \bar{r} = \frac{1}{N} \sum_{k=1}^{N} r_k = \frac{1}{N} \sum_{k=1}^{N} \hat{y}_k - \frac{1}{N} \sum_{k=1}^{N} y_k$$

$$\neq 0$$

3. Under this assumption, write $r_k = \bar{r} + r_k^0$ where:

$$\frac{1}{N} \sum_{k=1}^{N} r_k^0 = 0$$

$$\Rightarrow J = \sum_{k=1}^{N} (\bar{r} + r_k^0)^2$$

82
\[
\sum_{k=1}^{N} \left[ \bar{r}^2 + 2 \bar{r} r_k + (r_k^0)^2 \right] = \sum_{k=1}^{N} \bar{r}^2 + 2 \bar{r} \sum_{k=1}^{N} r^0_k + \sum_{k=1}^{N} (r^0_k)^2 \\
= N \bar{r}^2 + \sum_{k=1}^{N} (r^0_k)^2
\]

4. This result implies that
\[J > \sum_{k=1}^{N} (r^0_k)^2\]
from which it follows that \( J \) is minimized by forcing \( \bar{r} = 0 \), which is always possible in OLS models that include an intercept term.

**Exercise 2:**

Prove that if the variance of the predictions \( \{ \hat{y}_k \} \) is no larger than the variance of the data sequence \( \{ y_k \} \), the maximum achievable \( R^2 \) value defined in Eq. (15.1) is 1.

**Solution:**

1. The \( R^2 \) value is defined as:
\[R^2 = \frac{\sum_{k=1}^{N} (\hat{y}_k - \bar{y})^2}{\sum_{k=1}^{N} (y_k - \bar{y})^2}\]

Under the assumption that the average of the predictions \( \{ \hat{y}_k \} \) is equal to the average of the data values \( \{ y_k \} \), it follows that \( R^2 \) is simply the ratio of the prediction error variance to the data variance. Under the working assumptions of the problem, it follows that \( R^2 \leq 1 \).

**Exercise 3:**

Prove that the OLS parameter estimate vector \( \hat{\theta} \) is given by:
\[\hat{\theta} = V D^{-1} U^T y,\]

where \( U, V, \) and \( D \) are the matrices defined by the singular value decomposition of the data matrix \( X \) given in Eq. (15.21).

**Solution:**

1. First, represent \( X \) in terms of its singular value decomposition as:
\[X = UDV^T\]
2. Next, note that the explicit form of the OLS solution is:

\[ \hat{\theta} = (X^T X)^{-1} X^T y \]

3. Note that:

\[ X^T = (UDV^T)^T = VDU^T \]

\[ \Rightarrow \hat{\theta} = (VDU^T UDV^T)^{-1} VDU^T y \]

\[ = (VD^2 V^T)^{-1} VDU^T y \text{ (since } U^T U = I) \]

\[ = VD^{-2} V^T UDU^T y \text{ (since } V^{-1} = V^T) \]

\[ = VD^{-2} U^T y \]

Exercise 4—NOTE: the term \( \sigma_e^2 \) was omitted from the original problem statement; the corrected version and its solution are given here.

Show that the matrix \( C = \sigma_e^2 (X^T X)^{-1} \) can be written as:

\[ C = \sigma_e^2 VD^{-2} V^T, \]

where \( D \) and \( V \) are the matrices defined by the singular value decomposition of the data matrix \( X \) given in Eq. (15.21) and \( \sigma_e^2 \) is the variance of the model error term \( e_k \).

Solution:

1. From the solution to the previous problem:

\[ X = UDV^T \]

\[ \Rightarrow X^T = VDU^T \]

\[ \Rightarrow X^T X = VDU^T UDV^T \]

\[ = VD^2 V^T \]

2. Thus, we can write the required inverse as:

\[ (X^T X)^{-1} = (VD^2 V^T)^{-1} \]

\[ = (V^T)^{-1} D^{-2} V^{-1} \]

\[ = VD^{-2} V^T \]

3. Substituting this result into the definition of \( C \) yields:

\[ C = \sigma_e^2 (X^T X)^{-1} = \sigma_e^2 VD^{-2} V^T \]

Exercise 5:
It was shown that the condition number for the modified matrix $\tilde{C} = (kI + X^T X)^{-1}$ appearing in the ridge regression problem formulation is given by:

$$\tilde{\kappa} = \frac{k + \lambda_+}{k + \lambda_-},$$

where $k$ is the ridge regression tuning parameter. Show that $\tilde{\kappa}$ decreases with increasing $k$.

**Solution:**

1. Note that:

$$\tilde{\kappa} = \frac{k + \lambda_+}{k + \lambda_-} = \frac{k + \lambda_- + (\lambda_+ - \lambda_-)}{k + \lambda_-} = 1 + \frac{\lambda_+ - \lambda_-}{k + \lambda_-}$$

2. Differentiating with respect to $k$ gives:

$$\frac{d\tilde{\kappa}}{dk} = -\frac{(\lambda_+ - \lambda_-)}{(k + \lambda_-)^2}$$

3. Since $\lambda_+ > \lambda_-$, it follows that this derivative is negative for all $k$, implying that $\tilde{\kappa}$ decreases with increasing $k$.

**Exercise 6:**

Prove that an $N \times N$ idempotent matrix with rank $r$ has $r$ eigenvalues equal to 1 and $N - r$ eigenvalues equal to 0.

**Solution:**

1. Suppose $A$ is an $N \times N$ idempotent matrix and $x$ is an eigenvector of $A$ with eigenvalue $\lambda$. Thus:

$$Ax = \lambda x$$

$$\Rightarrow A^2x = A[Ax] = A[\lambda x] = \lambda Ax = \lambda^2 x$$

2. Since idempotence implies that $A^2 = A$ and eigenvectors are nonzero, this result implies that:

$$\lambda^2 x = \lambda x$$

$$\Rightarrow \lambda^2 = \lambda$$

$$\Rightarrow \lambda(1 - \lambda) = 0$$
3. Since the only two solutions of this quadratic equation are $\lambda = 0$ and $\lambda = 1$, it follows that these are the only possible eigenvalues.

**Exercise 7:**

Show that the change in the OLS parameter vector $\hat{\theta}$ on deletion of the $i^{th}$ observation from the dataset is given by:

$$\hat{\theta} - \hat{\theta}_{(i)} = \frac{(X^T X)^{-1} z_i r_i}{1 - H_{ii}},$$

where $z_i$ is the deleted row of the data matrix $X$, $r_i$ is the residual $r_i = y_i - \hat{y}_i$ associated with the $i^{th}$ response in the full data model, and $H_{ii}$ is the $i, i$ diagonal element of the hat matrix $H$.

**Solution:**

1. The original and deletion-based OLS estimates are:

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

$$\hat{\theta}_{(i)} = (X_{(i)}^T X_{(i)})^{-1} X_{(i)}^T y_{(i)}$$

2. From Eq. (15.34), we have:

$$(X_{(i)}^T X_{(i)})^{-1} = (X^T X)^{-1} + \frac{(X^T X)^{-1} z_i z_i^T (X^T X)^{-1}}{1 - H_{ii}}$$

3. Also, note that:

$$X_{(i)}^T y_{(i)} = X^T y - z_i y_i$$

4. Combining these results gives:

$$\hat{\theta}_{(i)} = (X^T X)^{-1} [X^T y - z_i y_i] + \frac{(X^T X)^{-1} z_i z_i^T (X^T X)^{-1} [X^T y - z_i y_i]}{1 - H_{ii}}$$

$$= \hat{\theta} - (X^T X)^{-1} z_i y_i + \frac{(X^T X)^{-1} z_i z_i^T (X^T X)^{-1} X^T y}{1 - H_{ii}}$$

$$- \frac{(X^T X)^{-1} z_i z_i^T (X^T X)^{-1} z_i y_i}{1 - H_{ii}}$$

$$= \hat{\theta} - (X^T X)^{-1} z_i y_i + \frac{(X^T X)^{-1} z_i z_i^T \hat{\theta}}{1 - H_{ii}} + \frac{(X^T X)^{-1} z_i H_{ii} y_i}{1 - H_{ii}}$$

$$= \hat{\theta} - (X^T X)^{-1} z_i y_i \left[1 + \frac{H_{ii}}{1 - H_{ii}}\right] + \frac{(X^T X)^{-1} z_i \hat{y}_i}{1 - H_{ii}}$$

$$= \hat{\theta} - (X^T X)^{-1} z_i (y_i - \hat{y}_i)$$

$$= \hat{\theta} - (X^T X)^{-1} z_i r_i$$

$$= \hat{\theta} - \frac{(X^T X)^{-1} z_i r_i}{1 - H_{ii}}$$

86
5. From this result, the influence of deletion on the OLS parameter estimate is:
\[
\hat{\theta} - \hat{\theta}_{(i)} = \frac{(X^T X)^{-1} z_i r_i}{1 - H_{ii}}
\]

Exercise 8:
Show that the maximum number of models built and evaluated in forward stepwise regression without backward elimination is \(C(C - 1)/2\).

Solution:
1. Stage 1 of forward stepwise regression without backward elimination builds the univariate models for each covariate. Each of these models is tested for significance relative to the intercept-only model, and if any are found significant at the specified level, the best-fitting of these models is selected as the basis for Stage 2; otherwise, the process terminates without selecting any of the covariates. Stage 1 thus requires building and evaluating \(C\) models.

2. In Stage 2, the best of the Stage 1 models is selected and the other \(C - 1\) covariates are each included as a second predictor in the model and these \(C - 1\) models are evaluated to determine whether any of these additional covariates yields a significant improvement in fit at the specified level. If not, the procedure terminates, selecting the Stage 1 model; otherwise, the procedure goes on to Stage 3.

3. In general, Stage \(j\) begins with a \(j - 1\) component model from Stage \(j - 1\) and evaluates the remaining \(C - j + 1\) covariates as possible additions to the model, either accepting the Stage \(j - 1\) model and terminating, or adding the best fitting new covariate and proceeding on to Stage \(j + 1\). The total number of models evaluated by the end of Stage \(j\) is:

\[
N_j = \sum_{i=1}^{j} (C - i + 1)
\]

\[
= j(C + 1) - \frac{j(j + 1)}{2}
\]

4. If a new covariate is selected at every stage, the procedure can continue until Stage \(C\), where \(C - 1\) covariates have been included in the model and the single remaining covariate is evaluated for possible inclusion. Whether this covariate is included in or excluded from the model, the procedure
necessarily terminates here. Thus, the largest number of models that would need to be evaluated is:

\[ N_C = C(C + 1) - \frac{C(C + 1)}{2} = \frac{C(C + 1)}{2} \]

Chapter 16

Exercise 1:

Show that the expected number of incomplete observations (i.e., observations with \( x_k \) missing, \( y_k \) missing, or both) is \( p(1 - p/4)N \) for the MCAR simulation dataset.

Solution:

1. The MCAR dataset is generated from a set of \( N \) bivariate data samples \( \{(x_k, y_k)\} \) by randomly omitting \( pN/2 \) values from both \( x_k \) and \( y_k \). While the numbers of both of these deletions is fixed, the indices \( k \) are randomly selected in both cases and it can happen that the same index \( k \) is selected for both \( x \)-deletion and \( y \)-deletion. Since each data point can only be deleted once, the effect of these “collisions” is to reduce the total number of deleted observations from \( pN \) to \( pN - C \) where \( C \) is the number of doubly-selected points.

2. Since each selection is made independently, the probability of double selection is:

\[ P\{\text{select both } x_k \text{ and } y_k\} = P\{\text{select } x_k\} \cdot P\{\text{select } y_k\} = \left(\frac{p}{2}\right) \cdot \left(\frac{p}{2}\right) = \frac{p^2}{4} \]

3. Thus, the expected number of double selections from \( N \) observations is \( Np^2/4 \), so the expected number of omitted observations is:

\[ N_{\text{omit}} = pN - Np^2/4 = p(1 - p/4)N \]

Exercise 2:

Show that the complete data log likelihood for the univariate example with nonignorable missing data considered in Sec. 16.7.3 is given by Eq. (16.14):

\[ L_c \{x_k \mid \mu, \sigma^2\} = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{m}{2} \ln \tau^2 - \frac{1}{2\sigma^2} S_2 + \frac{\mu}{\sigma^2} S_1 - \frac{\mu^2}{2\sigma^2} [N - m + m\lambda^2/\tau^2], \]

where \( S_1 \) and \( S_2 \) are the sufficient statistics defined in Eq. (16.15).
Solution:

1. The MNAR model has $x_k$ distributed as:

   $$
   x_k \sim \begin{cases} 
   N(\mu, \sigma^2) & k = 1, 2, \ldots, N - m \ (x_k \text{ not missing}) \\
   N(\lambda \mu, \tau^2 \sigma^2) & k = N - m + 1, \ldots, N \ (x_k \text{ missing})
   \end{cases}
   $$

   $$
   \Rightarrow p_k = \begin{cases} 
   \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{x_k - \mu}{\sigma} \right)^2 \right] & k = 1, 2, \ldots, N - m \\
   \frac{1}{\sqrt{2\pi}\tau \sigma} \exp \left[ -\frac{1}{2} \left( \frac{x_k - \lambda \mu}{\tau \sigma} \right)^2 \right] & k = N - m + 1, \ldots, N
   \end{cases}
   $$

2. Thus, the complete data log likelihood is:

   $$
   L_c(\{x_k\} | \mu, \sigma^2) = \sum_{k=1}^{N} \log p_k
   $$

   $$
   = \sum_{k=1}^{N-m} \left[ -\frac{1}{2} \ln 2\pi - \ln \sigma - \frac{1}{2} \left( \frac{x_k - \mu}{\sigma} \right)^2 \right]
   + \sum_{k=N-m+1}^{N} \left[ -\frac{1}{2} \ln 2\pi - \ln \tau \sigma - \frac{1}{2} \left( \frac{x_k - \lambda \mu}{\tau \sigma} \right)^2 \right]
   $$

   $$
   = \frac{(N - m)}{2} \ln 2\pi - (N - m) \ln \sigma - \frac{1}{2\sigma^2} \sum_{k=1}^{N-m} (x_k - \mu)^2
   $$

   $$
   - \frac{m}{2} \ln 2\pi - m \ln \tau \sigma - \frac{1}{2\tau^2 \sigma^2} \sum_{k=N-m+1}^{N} (x_k - \lambda \mu)^2
   $$

   $$
   = \frac{N}{2} \ln 2\pi - N \ln \sigma - m \ln \tau
   $$

   $$
   - \frac{1}{2\sigma^2} \left[ \sum_{k=1}^{N-m} (x_k - \mu)^2 + \frac{1}{\tau^2} \sum_{k=N-m+1}^{N} (x_k - \lambda \mu)^2 \right]
   $$

   $$
   = \frac{N}{2} \ln 2\pi - N \ln \sigma - m \ln \tau
   $$

   $$
   - \frac{1}{2\sigma^2} \left[ \sum_{k=1}^{N-m} (x_k^2 - 2x_k \mu + \mu^2) + \frac{1}{\tau^2} \sum_{k=N-m+1}^{N} (x_k^2 - 2\lambda \mu x_k + \lambda^2 \mu^2) \right]
   $$

   $$
   = \frac{N}{2} \ln 2\pi - N \ln \sigma - m \ln \tau
   $$

   $$
   - \frac{1}{2\sigma^2} \left[ \sum_{k=1}^{N-m} x_k^2 + \sum_{k=N-m+1}^{N} x_k^2/\tau^2 \right]
   $$

   $$
   + \frac{\mu}{\sigma^2} \left[ \sum_{k=1}^{N-m} x_k + \sum_{k=N-m+1}^{N} (\tau^2 x_k) \right]
   $$

   $$
   - \frac{(N - m)\mu^2 + m\lambda^2 \mu^2 / \tau^2}{2\sigma^2}
   $$

89
3. Defining the sums $S_1$ and $S_2$ as in Eq. (16.15), this result simplifies to Eq. (16.16):

$$L_c(\{x_k\} | \mu, \sigma^2) = -\frac{N}{2} \ln 2\pi - N \ln \sigma - m \ln \tau - \frac{S_2}{2\sigma^2} + \frac{\mu S_1}{\sigma^2} - \frac{(N - m)\mu^2 + m\lambda^2\mu^2/\tau^2}{2\sigma^2}$$

**Exercise 3:**

Derive the maximum likelihood estimators $\hat{\mu}$ and $\hat{\sigma}^2$ shown in Eq. (16.16) for the nonignorable missing data example considered in Sec. 16.7.3.

**Solution:**

1. To solve for $\hat{\mu}$, differentiate the complete data likelihood with respect to $\mu$, giving:

$$\frac{\partial L_c(\{x_k\} | \mu, \sigma^2)}{\partial \mu} = \frac{S_1}{\sigma^2} - \frac{\mu[N - m + m\lambda^2/\tau^2]}{\sigma^2}$$

2. Setting this derivative to zero yields:

$$\frac{S_1}{\sigma^2} = \frac{\hat{\mu}[N - m + m\lambda^2/\tau^2]}{\sigma^2}$$

$$\Rightarrow \hat{\mu} = \frac{S_1}{N - m + m\lambda^2/\tau^2}$$

3. For $\hat{\sigma}^2$, differentiate the complete data likelihood with respect to $\sigma^2$, yielding:

$$\frac{\partial L_c(\{x_k\} | \mu, \sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{S_2}{2\sigma^4} - \frac{\mu S_1}{\sigma^4} + \frac{\mu^2[N - m + m\lambda^2/\tau^2]}{2\sigma^4}$$

4. Setting this derivative to zero yields:

$$\frac{N}{2\sigma^2} = \frac{S_2 - 2\mu S_1 + \mu^2[N - m + m\lambda^2/\tau^2]}{2\sigma^4}$$

$$\Rightarrow \hat{\sigma}^2 = \frac{S_2 - 2\mu S_1 + \mu^2[N - m + m\lambda^2/\tau^2]}{N}$$

5. To simplify this result, note that:

$$S_1 = \hat{\mu}[N - m + m\lambda^2/\tau^2]$$

$$\Rightarrow \hat{\sigma}^2 = \frac{S_2 - 2\hat{\mu}^2[N - m + m\lambda^2/\tau^2] + \hat{\mu}^2[N - m + m\lambda^2/\tau^2]}{N} = \frac{S_2 - [N - m + m\lambda^2/\tau^2]\hat{\mu}^2}{N}$$
Exercise 4:

Compute the limiting values $\mu(\infty)$ and $\sigma^2(\infty)$ for the EM algorithm described in Sec. 16.7.3 for the univariate MNAR problem considered there.

Solution:

1. To obtain the limiting estimator for $\mu$, set $\mu(j) = \mu(j+1) = \mu(\infty)$ in the first equation in Eq. (16.18), which gives:

   \[ \mu(\infty) = \frac{\sum_{k=1}^{N-m} x_k + (m\lambda^2/\tau^2)\mu(\infty)}{N - m + m\lambda^2/\tau^2} \]

   \[ \Rightarrow [N - m + m\lambda^2/\tau^2]\mu(\infty) = \sum_{k=1}^{N-m} x_k + (m\lambda^2/\tau^2)\mu(\infty) \]

   \[ \Rightarrow (N - m)\mu(\infty) = \sum_{k=1}^{N-m} x_k \]

   \[ \Rightarrow \mu(\infty) = \frac{1}{N - m} \sum_{k=1}^{N-m} x_k \]

2. Similarly, setting $\sigma^2(j) = \sigma^2(j+1) = \sigma^2(\infty)$ yields:

   \[ \sigma^2(\infty) = \frac{1}{N} \left[ \sum_{k=1}^{N-m} x_k^2 - (N - m)\mu^2(\infty) + m\sigma^2(\infty) \right] \]

   \[ \Rightarrow N\sigma^2(\infty) = \sum_{k=1}^{N-m} x_k^2 - (N - m)\mu^2(\infty) + m\sigma^2(\infty) \]

   \[ \Rightarrow (N - m)\sigma^2(\infty) = \sum_{k=1}^{N-m} x_k^2 - (N - m)\mu^2(\infty) \]

   \[ \Rightarrow \sigma^2(\infty) = \frac{1}{N - m} \sum_{k=1}^{N-m} x_k^2 - \mu^2(\infty) \]

Exercise 5:

Show that the complete data log likelihood for the bivariate Gaussian example considered in Sec. 16.7.4 is given by Eq. (16.21).

Solution:

1. To obtain the desired result, it is necessary to first write the joint density explicitly in terms of its components. To do this, first note that the inverse
covariance matrix is:

\[ \Sigma^{-1} = \begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix}^{-1} \]

\[ \Sigma^{-1} = \frac{1}{\sigma_x^2 \sigma_y^2 - \rho^2 \sigma_x^2 \sigma_y^2} \begin{bmatrix} \sigma_y^2 & -\rho \sigma_x \sigma_y \\ -\rho \sigma_x \sigma_y & \sigma_x^2 \end{bmatrix} \]

\[ \Sigma^{-1} = \frac{1}{\zeta} \begin{bmatrix} \sigma_y^2 & -\rho \sigma_x \sigma_y \\ -\rho \sigma_x \sigma_y & \sigma_x^2 \end{bmatrix} \]

2. From this result, we have:

\[
\Sigma^{-1}(w - \mu) = \frac{1}{\zeta} \begin{bmatrix} \sigma_y^2 & -\rho \sigma_x \sigma_y \\ -\rho \sigma_x \sigma_y & \sigma_x^2 \end{bmatrix} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}
\]

\[ \Rightarrow (w - \mu)^T \Sigma^{-1}(w - \mu) = \frac{1}{\zeta} [\sigma_y^2 (x - \mu_x)^2 - \rho \sigma_x \sigma_y (x - \mu_x)(y - \mu_y)] \]

\[ -\rho \sigma_x \sigma_y (x - \mu_x)(y - \mu_y) + \sigma_x^2 (y - \mu_y)^2] \]

\[ = \frac{\sigma_y^2 (x - \mu_x)^2 - 2 \rho \sigma_x \sigma_y (x - \mu_x)(y - \mu_y) + \sigma_x^2 (y - \mu_y)^2}{\zeta} \]

3. Taking the log of \( p(x_k, y_k) \) then gives:

\[
\ln p(x_k, y_k) = -\ln 2\pi - \frac{1}{2} \ln \zeta - \frac{1}{2\zeta} \sigma_y^2 (x_k - \mu_x)^2 - 2 \rho \sigma_x \sigma_y (x_k - \mu_x)(y_k - \mu_y) + \sigma_x^2 (y_k - \mu_y)^2]
\]

\[ = -\ln 2\pi - \frac{1}{2} \ln \zeta - \frac{1}{2\zeta} [\sigma_y^2 (x_k^2 - 2 \mu_x x_k + \mu_x^2) - 2 \rho \sigma_x \sigma_y (x_k y_k - \mu_x y_k - \mu_y x_k + \mu_x \mu_y) + \sigma_x^2 (y_k^2 - 2 \mu_y y_k + \mu_y^2)]
\]

\[ = -\ln 2\pi - \frac{1}{2} \ln \zeta - \frac{1}{2\zeta} [\sigma_y^2 x_k^2 - 2 \rho \sigma_x \sigma_y x_k y_k + \sigma_x^2 y_k^2 - 2[(\mu_x \sigma_y^2 - \mu_y \rho \sigma_x \sigma_y)x_k + (\mu_y \sigma_x^2 - \mu_x \rho \sigma_x \sigma_y)y_k] + \mu_x^2 \sigma_y^2 + \mu_y^2 \sigma_x^2 - 2 \rho \mu_x \mu_y \sigma_x \sigma_y]
\]

3. The complete data likelihood is obtained by summing over the \( N \) observations \((x_k, y_k)\) to obtain:

\[ L_c(D|\mu, \Sigma) = \sum_{k=1}^{N} \ln p(x_k, y_k) \]
\[
L_c(D|\mu, \Sigma) = -N \ln 2\pi - \frac{N}{2} \ln \zeta - \frac{1}{2\zeta \Sigma} \left\{ \sigma_y^2 \sum_{k=1}^{N} x_k^2 - 2 \rho \sigma_x \sigma_y \sum_{k=1}^{N} x_k y_k + \sigma_x^2 \sum_{k=1}^{N} y_k^2 \right\} + \frac{1}{\zeta} \left[ (\mu_x \sigma_y^2 - \mu_y \rho \sigma_x \sigma_y) \sum_{k=1}^{N} x_k + (\mu_y \sigma_x^2 - \mu_x \rho \sigma_x \sigma_y) \sum_{k=1}^{N} y_k \right] - \frac{N(\mu_x^2 \sigma_y^2 + \mu_y^2 \sigma_x^2 - 2 \rho \mu_x \mu_y)}{\Sigma} \]

4. Note that these sums correspond to those defined in Eq. (16.22), leading to the final result (Eq. (16.21)):

\[
L_c(D|\mu, \Sigma) = -N \ln 2\pi - \frac{N}{2} \ln \zeta - \frac{1}{2\zeta \Sigma} \left\{ \sigma_y^2 T_{11} - 2 \rho \sigma_x \sigma_y T_{12} + \sigma_x^2 T_{22} - 2[(\mu_x \sigma_y^2 - \mu_y \rho \sigma_x \sigma_y) T_{1} + (\mu_y \sigma_x^2 - \mu_x \rho \sigma_x \sigma_y) T_{2}] + N(\mu_x^2 \sigma_y^2 + \mu_y^2 \sigma_x^2 - 2 \rho \mu_x \mu_y) \right\}
\]

Exercise 6:

Show that the complete data maximum likelihood estimators for the bivariate Gaussian parameters are given by Eq. (16.23).

Solution:

1. The complete data likelihood is given by Eq. (16.20) as:

\[
L_c(D|\mu, \Sigma) = -N \ln 2\pi - \frac{N}{2} \ln \zeta - \frac{1}{2\zeta \Sigma} \left\{ \sigma_y^2 T_{11} - 2 \rho \sigma_x \sigma_y T_{12} + \sigma_x^2 T_{22} - 2[(\mu_x \sigma_y^2 - \mu_y \rho \sigma_x \sigma_y) T_{1} + (\mu_y \sigma_x^2 - \mu_x \rho \sigma_x \sigma_y) T_{2}] + N(\mu_x^2 \sigma_y^2 + \mu_y^2 \sigma_x^2 - 2 \rho \mu_x \mu_y) \right\}
\]

To compute \( \hat{\mu}_x \), solve \( \frac{\partial L_c}{\partial \mu_x} = 0 \). Computing the derivative yields:

\[
\frac{\partial L_c}{\partial \mu_x} = \frac{\partial}{\partial \mu_x} \left[ -\frac{1}{2\zeta} \left\{ -2 \sigma_y^2 T_{11} + 2 \mu_x \rho \sigma_x \sigma_y T_{12} + N\mu_x^2 \sigma_y^2 - 2N \rho \mu_x \mu_y \sigma_x \sigma_y \right\} \right]
\]

\[
= -\frac{1}{2\zeta} \left[ -2 \sigma_y^2 T_{1} + 2 \rho \sigma_x \sigma_y T_{2} + 2N \mu_x \sigma_y^2 - 2N \rho \mu_x \sigma_x \sigma_y \right]
\]

\[
= \frac{\sigma_y^2 (T_{1} - N\mu_x) - \rho \sigma_x \sigma_y (T_{2} - N\mu_y)}{\zeta}
\]

Setting this derivative to zero yields the following equation:

\[
\sigma_y^2 (T_{1} - N\mu_x) - \rho \sigma_x \sigma_y (T_{2} - N\mu_y) = 0
\]
2. Differentiating $L_c$ with respect to $\mu_y$ and setting the result to zero yields an analogous equation, but with $\mu_x \leftrightarrow \mu_y$, $\sigma_x \leftrightarrow \sigma_y$, and $T_1 \leftrightarrow T_2$, giving:

$$\sigma^2_T (T_2 - N \mu_y) - \rho \sigma_x \sigma_y (T_1 - N \mu_x) = 0$$

Rearranging these two equations yields the following pair of simultaneous linear equations for $\mu_x$ and $\mu_y$:

$$-N \sigma^2_{\mu x} + N \rho \sigma_x \sigma_y \mu_y = \rho \sigma_x \sigma_y T_2 - \sigma^2_{\mu y} T_1$$
$$N \rho \sigma_x \sigma_y \mu_x - N \sigma^2_{\mu y} = \rho \sigma_x \sigma_y T_1 - \sigma^2_{\mu y} T_2$$

Multiplying the first of these equations by $\partial L / \partial \mu_x$ and the second by $\sigma_y$, we obtain:

$$-N \rho \sigma_x \sigma_y^2 \mu_x + N \rho^2 \sigma^2_x \sigma_y \mu_y = \rho^2 \sigma^2_x \sigma_y T_2 - \rho \sigma_x \sigma_y^2 T_1$$
$$N \rho \sigma_x \sigma_y^2 \mu_x - N \sigma^2_x \sigma_y \mu_y = \rho \sigma_x \sigma_y^2 T_1 - \sigma^2_x \sigma_y T_2$$

Adding these equations yields:

$$N \sigma^2_x \sigma_y^2 (\rho^2 - 1) \mu_y = (\rho^2 - 1) \sigma^2_x \sigma_y T_2$$
$$\Rightarrow \quad N \mu_y = T_2$$
$$\Rightarrow \quad \hat{\mu}_y = \frac{T_2}{N}$$

Substituting this result into the original equation obtained by setting $\partial L_c / \partial \mu_x = 0$ gives:

$$\sigma^2_T (T_1 - N \mu_x) = 0 \quad \Rightarrow \quad \hat{\mu}_x = \frac{T_1}{N}$$

3. To solve for $\sigma_x$, $\hat{\sigma}_x$, and $\hat{\rho}$, substitute $\hat{\mu}_x$ and $\hat{\mu}_y$ into $L_c(D|\mu, \Sigma)$ to obtain:

$$L_c(D|\mu, \Sigma) = -N \ln 2\pi - \frac{N}{2} \ln \zeta - \frac{1}{2\zeta} \{ \sigma^2_T T_{11} - 2 \rho \sigma_x \sigma_y T_{12} + \sigma^2_T T_{22} \}$$
$$-2 \{(\hat{\mu}_x \sigma^2_y - \hat{\mu}_y \rho \sigma_x \sigma_y)(N \hat{\mu}_x) + (\hat{\mu}_y \sigma^2_y - \hat{\mu}_x \rho \sigma_x \sigma_y)(N \hat{\mu}_y) \}$$
$$+ N(\hat{\mu}_x \sigma^2_y + \hat{\mu}_y \sigma^2_x - 2 \rho \hat{\mu}_x \hat{\mu}_y \sigma_x \sigma_y)$$
$$= -N \ln 2\pi - \frac{N}{2} \ln \zeta - \frac{1}{2\zeta} \{ \sigma^2_T T_{11} - 2 \rho \sigma_x \sigma_y T_{12} + \sigma^2_T T_{22} \}$$
$$-2N[\hat{\mu}_x \sigma^2_y + \hat{\mu}_y \sigma^2_x - 2 \rho \hat{\mu}_x \hat{\mu}_y \sigma_x \sigma_y]$$
$$+ N[\hat{\mu}_x \sigma^2_y + \hat{\mu}_y \sigma^2_x - 2 \rho \hat{\mu}_x \hat{\mu}_y \sigma_x \sigma_y]$$
$$= -N \ln 2\pi - \frac{N}{2} \ln \zeta - \frac{1}{2\zeta} \{ \sigma^2_T T_{11} - 2 \rho \sigma_x \sigma_y T_{12} + \sigma^2_T T_{22} \}$$
$$-N[\hat{\mu}_x \sigma^2_y + \hat{\mu}_y \sigma^2_x - 2 \rho \hat{\mu}_x \hat{\mu}_y \sigma_x \sigma_y]$$
$$= -N \ln 2\pi - \frac{N}{2} \ln \zeta - \frac{N}{2\zeta} \left( \sigma^2_T \left( \frac{T_{11}}{N} - \hat{\mu}_x^2 \right) + \sigma^2_x \left( \frac{T_{22}}{N} - \hat{\mu}_y^2 \right) - 2 \rho \sigma_x \sigma_y \left( \frac{T_{12}}{N} - \hat{\mu}_x \hat{\mu}_y \right) \right) \]
To simplify these expressions, make the following definitions:

\[ S_{11} = \frac{T_{11}}{N} - \bar{\mu}_x^2 \]
\[ S_{12} = \frac{T_{12}}{N} - \bar{\mu}_x \bar{\mu}_y \]
\[ S_{22} = \frac{T_{22}}{N} - \bar{\mu}_y^2 \]

This gives the following expression for the likelihood:

\[
L_c(D|\hat{\mu}, \Sigma) = -N \ln 2\pi - N \ln \frac{\sigma_x^2}{2\pi} - N \ln \frac{\sigma_y^2}{\pi} - \frac{1}{2} \ln \frac{\sigma_x^2 S_{11} + 2\rho \sigma_x \sigma_y S_{12} + \sigma_y^2 S_{22}}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)}
\]

Noting that \( \zeta = \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)} \), this likelihood may be expressed directly in terms of the unknown parameters \( \sigma_x \), \( \sigma_y \) and \( \rho \) as:

\[
L_c(D|\hat{\mu}, \Sigma) = -N \ln 2\pi - N \ln \sigma_x - N \ln \sigma_y - \frac{N}{2} \ln (1 - \rho^2)
\]

\[ = -N \ln 2\pi - N \ln \sigma_x - N \ln \sigma_y - \frac{N}{2} \ln (1 - \rho^2)
\]

\[ = -\frac{NS_{11}}{2\sigma_x^2 (1 - \rho^2)} + \frac{N\rho S_{12}}{\sigma_x \sigma_y (1 - \rho^2)} - \frac{NS_{22}}{2\sigma_y^2 (1 - \rho^2)}
\]

4. To obtain equations for \( \hat{\sigma}_x \), \( \hat{\sigma}_y \), and \( \hat{\rho} \), differentiate the likelihood with respect to each of these parameters and set the results to zero. For \( \sigma_x \), we have:

\[
\frac{\partial L_c}{\partial \sigma_x} = \frac{\partial}{\partial \sigma_x} \left[ -N \ln \sigma_x - \frac{NS_{11}}{2\sigma_x^2 (1 - \rho^2)} + \frac{N\rho S_{12}}{\sigma_x \sigma_y (1 - \rho^2)} \right]
\]

\[ = -\frac{N}{\sigma_x} + \frac{NS_{11}}{\sigma_x^3 (1 - \rho^2)} - \frac{N\rho S_{12}}{\sigma_x^2 \sigma_y (1 - \rho^2)}
\]

\[ = -\frac{N}{\sigma_x} \left[ 1 - \frac{S_{11}}{\sigma_x^2 (1 - \rho^2)} + \frac{\rho S_{12}}{\sigma_x \sigma_y (1 - \rho^2)} \right]
\]

Setting this derivative to zero yields:

\[ 1 = \frac{S_{11}}{\sigma_x^2 (1 - \rho^2)} - \frac{\rho S_{12}}{\sigma_x \sigma_y (1 - \rho^2)} \]

\[ \Rightarrow \sigma_x^2 \sigma_y^2 (1 - \rho^2) = \sigma_x^2 S_{11} - \rho \sigma_x \sigma_y S_{12} \]

By simply reversing the roles of \( x \) and \( y \), we obtain the following equation from setting \( \frac{\partial L_c}{\partial \sigma_y} \) to zero:

\[ \sigma_x^2 \sigma_y^2 (1 - \rho^2) = \sigma_x^2 S_{22} - \rho \sigma_x \sigma_y S_{12} \]
5. Taking the derivative of $L_c$ with respect to $\rho$ yields:
\[
\frac{\partial L_c}{\partial \rho} = \frac{\partial}{\partial \rho} \left[ -\frac{N}{2} \ln(1 - \rho^2) - \frac{NS_{11}}{2\sigma_y^2(1 - \rho^2)} + \frac{N\rho S_{12}}{\sigma_x \sigma_y (1 - \rho^2)} - \frac{N S_{22}}{2\sigma_y^2(1 - \rho^2)} \right]
\]
\[
= \frac{N\rho}{1 - \rho^2} - \frac{N\rho S_{11}}{\sigma_y^2(1 - \rho^2)} + \frac{NS_{12}}{\sigma_x \sigma_y (1 - \rho^2)} \frac{\partial}{\partial \rho} \left( \rho \right) - \frac{N S_{22}}{2\sigma_y^2(1 - \rho^2)}
\]
Note that:
\[
\frac{\partial}{\partial \rho} \left( \frac{\rho}{1 - \rho^2} \right) = \frac{(1 - \rho^2) \cdot 1 - \rho \cdot (-2\rho)}{(1 - \rho^2)^2}
\]
\[
= \frac{1 + \rho^2}{1 - \rho^2}
\]
Thus, it follows that:
\[
\frac{\partial L_c}{\partial \rho} = \frac{N\rho}{1 - \rho^2} - \frac{N\rho S_{11}}{\sigma_y^2(1 - \rho^2)} + \frac{NS_{12}(1 + \rho^2)}{\sigma_x \sigma_y (1 - \rho^2)^2} - \frac{N S_{22}}{2\sigma_y^2(1 - \rho^2)}
\]
\[
= \frac{N}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)^2} \left[ \rho(1 - \rho^2)\sigma_x^2 \sigma_y^2 - \rho \sigma_y^2 S_{11} + S_{12} \sigma_x \sigma_y (1 + \rho^2) - \rho \sigma_y^2 S_{22} \right]
\]

6. To simplify this result, substitute the previous equations involving $\sigma_x$ and $\sigma_y$, re-arranged as:
\[
\sigma_y^2 S_{11} = \sigma_x^2 \sigma_y^2 (1 - \rho^2) + \rho \sigma_x \sigma_y S_{12}
\]
\[
\sigma_x^2 S_{22} = \sigma_x^2 \sigma_y^2 (1 - \rho^2) + \rho \sigma_x \sigma_y S_{12}
\]
Substituting these results into the above equation yields:
\[
\frac{\partial L_c}{\partial \rho} = \frac{N}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)^2} \left[ -2\rho^2 \sigma_x \sigma_y S_{12} + S_{12} \sigma_x \sigma_y (1 + \rho^2) - \rho \sigma_y^2 S_{11} - \rho \sigma_y^2 S_{22} \right]
\]
\[
= \frac{N}{\sigma_x \sigma_y (1 - \rho^2)^2} \left[ S_{12} \sigma_x \sigma_y (1 - \rho^2) - \rho \sigma_x \sigma_y (1 - \rho^2) \right]
\]
Thus, setting this derivative to zero implies:
\[
S_{12} = \rho \sigma_x \sigma_y
\]
Substituting this result back into the $\sigma_y$ equation yields:
\[
S_{12} = \rho \sigma_x \sigma_y \Rightarrow \sigma_y^2 S_{11} - \rho^2 \sigma_x^2 \sigma_y^2 \Rightarrow \sigma_y^2 \sigma_y^2 = \sigma_y^2 S_{11}
\]
\[
\Rightarrow \sigma_y = \sigma_y S_{11}
\]
\[
\Rightarrow \sigma_x = S_{11}
\]
\[
96
\]
By exactly the same reasoning, we also have \( \hat{\sigma}_y^2 = S_{22} \). Defining \( \sigma_{xy} \) as \( \rho \sigma_x \sigma_y \), it follows that \( \hat{\sigma}_{xy} = S_{12} \).

**Exercise 7:**

Show that the conditional expectation of the sums \( T_2, T_{11}, T_{22}, \) and \( T_{12} \) given the observed data and the parameters \( \mu^{(j)} \) and \( \Sigma^{(j)} \) for the bivariate Gaussian example discussed in Sec. 16.7.4 are given by Eq. (16.27).

**Solution:**

1. For \( T_2 \), the E-step conditional expectation is computed in exactly the same was as that for \( T_1 \), i.e.:

\[
T_2^{(j)} = E[T_2 | D_{\text{obs}}, \mu^{(j)}, \Sigma^{(j)}] = \sum_{k \in K_y} y_k + \sum_{k \notin K_y} E\{y_k | x_k, \mu^{(j)}, \Sigma^{(j)}\}
\]

\[
= \sum_{k \in K_y} y_k + (N - N_y) \left( \mu_y^{(j)} - \frac{\sigma_{xy} \mu_x^{(j)}}{\sigma_y^{(j)}} \right) + \left( \frac{\sigma_{xy}}{\sigma_y^{(j)}} \right) \sum_{k \notin K_y} x_k
\]

2. For \( T_{11} \) the conditional expectation is:

\[
T_{11}^{(j)} = E[T_{11} | D_{\text{obs}}, \mu^{(j)}, \Sigma^{(j)}] = \sum_{k \in K_x} x_k^2 + \sum_{k \notin K_x} E\{x_k^2 | y_k, \mu^{(j)}, \Sigma^{(j)}\}
\]

\[
= \sum_{k \in K_x} x_k^2 + \sum_{k \notin K_x} \left[ (E\{x_k | y_k, \mu^{(j)}, \Sigma^{(j)}\})^2 + \text{var}\{x_k | y_k, \mu^{(j)}, \Sigma^{(j)}\} \right]
\]

\[
= \sum_{k \in K_x} x_k^2 + \sum_{k \notin K_x} \left[ \left( \mu_x^{(j)} + \frac{\sigma_{xy} (y_k - \mu_y^{(j)})}{\sigma_y^{(j)}} \right)^2 + \sigma_x^{2(j)} (1 - \rho^{2(j)}) \right]
\]

\[
= \sum_{k \in K_x} x_k^2 + \sum_{k \notin K_x} \left( \mu_x^{(j)} + \frac{\sigma_{xy} (y_k - \mu_y^{(j)})}{\sigma_y^{(j)}} \right)^2 + (N - N_y) \sigma_x^{2(j)} (1 - \rho^{2(j)})
\]

3. The result for \( T_{22}^{(j)} \) is derived analogously and may be obtained by simply interchanging \( x \) and \( y \) in the \( T_{11}^{(j)} \) result.

4. For \( T_{12} \), the conditional expectation is:

\[
T_{12}^{(j)} = E[T_{12} | D_{\text{obs}}, \mu^{(j)}, \Sigma^{(j)}] = \sum_{k \in K_{xy}} x_k y_k + \sum_{k \in K_x \setminus K_y} E\{x_k y_k | x_k, \mu^{(j)}, \Sigma^{(j)}\}
\]
+ \sum_{k \in K_y \setminus K_x} E\{x_k y_k | y_k, \mu^{(j)}, \Sigma^{(j)}\}

= \sum_{k \in K_x} x_k y_k + \sum_{k \in K_x \setminus K_y} x_k E\{y_k | x_k, \mu^{(j)}, \Sigma^{(j)}\}

+ \sum_{k \in K_y \setminus K_x} y_k E\{x_k | y_k, \mu^{(j)}, \Sigma^{(j)}\}

= \sum_{k \in K_x} x_k y_k + \sum_{k \in K_x \setminus K_y} x_k \left( \mu^{(j)} + \frac{\sigma_{xy}^{(j)} (x_k - \mu^{(j)}_x)}{\sigma_{xy}^{(j)}} \right)

+ \sum_{k \in K_y \setminus K_x} y_k \left( \mu^{(j)}_y + \frac{\sigma_{xy}^{(j)} (y_k - \mu^{(j)}_y)}{\sigma_{xy}^{(j)}} \right)

Exercise 8:

Show that the limiting variance for the Healy-Westmacot procedure described in Sec. 16.7.6 is given by:

\[ \sigma^2_{(\infty)} = \frac{1}{N - m} \sum_{k=1}^{N-m} (y_k - x^T_k \beta_{(\infty)})^2, \]

where \( \beta_{(\infty)} \) is the limiting vector of regression coefficients obtained by the procedure.

Solution:

1. Given the limiting parameter estimate vector \( \beta_{(\infty)} \), the limiting variance is obtained by substituting \( \sigma^2_{(j)} = \sigma^2_{(j+1)} = \sigma^2_{(\infty)} \) into Eq. (16.33):

\[
\sigma^2_{(\infty)} = \frac{1}{N} \left[ \sum_{k=1}^{N-m} (y_k - x^T_k \beta_{(\infty)})^2 + m \sigma^2_{(\infty)} \right]
\]

\[ \Rightarrow N \sigma^2_{(\infty)} = \sum_{k=1}^{N-m} (y_k - x^T_k \beta_{(\infty)})^2 + m \sigma^2_{(\infty)} \]

\[ \Rightarrow (N - m) \sigma^2_{(\infty)} = \sum_{k=1}^{N-m} (y_k - x^T_k \beta_{(\infty)})^2 \]

\[ \Rightarrow \sigma^2_{(\infty)} = \frac{1}{N - m} \sum_{k=1}^{N-m} (y_k - x^T_k \beta_{(\infty)})^2 \]