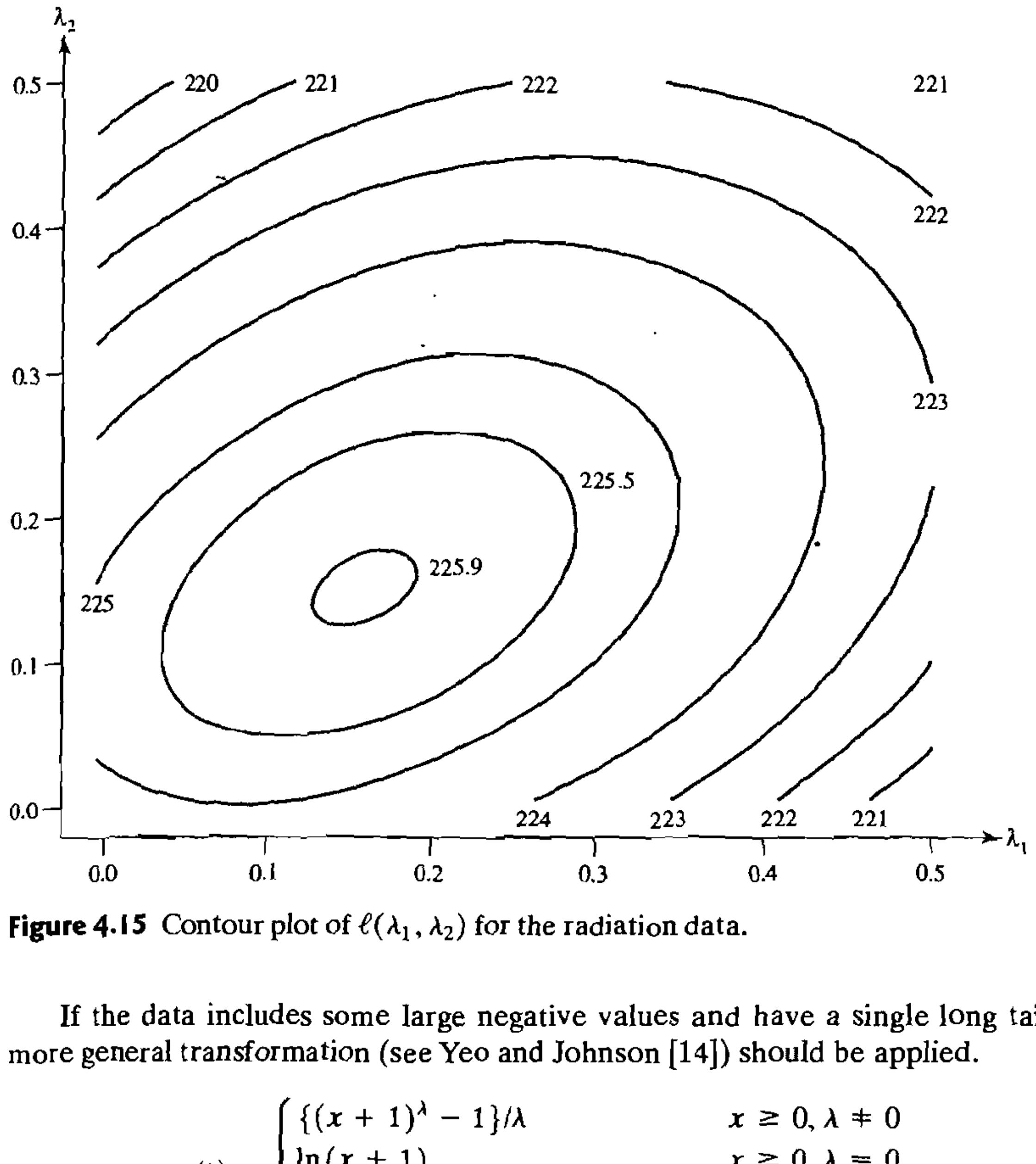

Figure 4.15 Contour plot of $\ell(\lambda_1, \lambda_2)$ for the radiation data.

If the data includes some large negative values and have a single long tail, a more general transformation (see Yeo and Johnson [14]) should be applied.

$$x^{(\lambda)} = \begin{cases} \{(x+1)^\lambda - 1\}/\lambda & x \geq 0, \lambda \neq 0 \\ \ln(x+1) & x \geq 0, \lambda = 0 \\ -\{(-x+1)^{2-\lambda} - 1\}/(2-\lambda) & x < 0, \lambda \neq 2 \\ -\ln(-x+1) & x < 0, \lambda = 2 \end{cases}$$

Exercises

4.1. Consider a bivariate normal distribution with $\mu_1 = 1$, $\mu_2 = 3$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ and $\rho_{12} = -.8$.

(a) Write out the bivariate normal density.

(b) Write out the squared statistical distance expression $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ as a quadratic function of x_1 and x_2.

4.2. Consider a bivariate normal population with $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_{11} = 2$, $\sigma_{22} = 1$, and $\rho_{12} = .5$.

(a) Write out the bivariate normal density.

(b) Write out the squared generalized distance expression $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ as a function of x_1 and x_2.

(c) Determine (and sketch) the constant-density contour that contains 50% of the probability.

4.3. Let $\mathbf{X}$ be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}' = [-3, 1, 4]$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Which of the following random variables are independent? Explain.

(a) X_1 and X_2

(b) X_2 and X_3

(c) (X_1, X_2) and X_3

(d) $\dfrac{X_1 + X_2}{2}$ and X_3

(e) X_2 and $X_2 - \frac{5}{2}X_1 - X_3$

4.4. Let $\mathbf{X}$ be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}' = [2, -3, 1]$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

(a) Find the distribution of $3X_1 - 2X_2 + X_3$.

(b) Relabel the variables if necessary, and find a 2×1 vector $\mathbf{a}$ such that X_2 and $X_2 - \mathbf{a}'\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ are independent.

4.5. Specify each of the following.

(a) The conditional distribution of X_1, given that $X_2 = x_2$ for the joint distribution in Exercise 4.2.

(b) The conditional distribution of X_2, given that $X_1 = x_1$ and $X_3 = x_3$ for the joint distribution in Exercise 4.3.

(c) The conditional distribution of X_3, given that $X_1 = x_1$ and $X_2 = x_2$ for the joint distribution in Exercise 4.4.

4.6. Let $\mathbf{X}$ be distributed as $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}' = [1, -1, 2]$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

Which of the following random variables are independent? Explain.

(a) X_1 and X_2

(b) X_1 and X_3

(c) X_2 and X_3

(d) (X_1, X_3) and X_2

(e) X_1 and $X_1 + 3X_2 - 2X_3$

4.7. Refer to Exercise 4.6 and specify each of the following.

(a) The conditional distribution of X_1, given that $X_3 = x_3$.

(b) The conditional distribution of X_1, given that $X_2 = x_2$ and $X_3 = x_3$.

4.8. (Example of a nonnormal bivariate distribution with normal marginals.) Let X_1 be $N(0, 1)$, and let

$$X_2 = \begin{cases} -X_1 & \text{if } -1 \le X_1 \le 1 \\ X_1 & \text{otherwise} \end{cases}$$

Show each of the following.

(a) X_2 also has an $N(0, 1)$ distribution.

(b) X_1 and X_2 do *not* have a bivariate normal distribution.

Hint:

(a) Since X_1 is $N(0,1)$, $P[-1 < X_1 \le x] = P[-x \le X_1 < 1]$ for any x. When $-1 < x_2 < 1$, $P[X_2 \le x_2] = P[X_2 \le -1] + P[-1 < X_2 \le x_2] = P[X_1 \le -1] + P[-1 < -X_1 \le x_2] = P[X_1 \le -1] + P[-x_2 \le X_1 < 1]$. But $P[-x_2 \le X_1 < 1] = P[-1 < X_1 \le x_2]$ from the symmetry argument in the first line of this hint. Thus, $P[X_2 \le x_2] = P[X_1 \le -1] + P[-1 < X_1 \le x_2] = P[X_1 \le x_2]$, which is a standard normal probability.

(b) Consider the linear combination $X_1 - X_2$, which equals zero with probability $P[|X_1| > 1] = .3174$.

4.9. Refer to Exercise 4.8, but modify the construction by replacing the break point 1 by c so that

$$X_2 = \begin{cases} -X_1 & \text{if } -c \le X_1 \le c \\ X_1 & \text{elsewhere} \end{cases}$$

Show that c can be chosen so that $\text{Cov}(X_1, X_2) = 0$, but that the two random variables are not independent.

Hint:
For $c = 0$, evaluate $\text{Cov}(X_1, X_2) = E[X_1(X_1)]$
For c very large, evaluate $\text{Cov}(X_1, X_2) \doteq E[X_1(-X_1)]$.

4.10. Show each of the following.

(a)

$$\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}' & \mathbf{B} \end{vmatrix} = |\mathbf{A}||\mathbf{B}|$$

(b)

$$\begin{vmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{0}' & \mathbf{B} \end{vmatrix} = |\mathbf{A}||\mathbf{B}| \quad \text{for} \quad |\mathbf{A}| \ne 0$$

Hint:

(a) $\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}' & \mathbf{B} \end{vmatrix} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}' & \mathbf{I} \end{vmatrix}\begin{vmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}' & \mathbf{B} \end{vmatrix}$. Expanding the determinant $\begin{vmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}' & \mathbf{B} \end{vmatrix}$ by the first row (see Definition 2A.24) gives 1 times a determinant of the same form, with the order of $\mathbf{I}$ reduced by one. This procedure is repeated until $1 \times |\mathbf{B}|$ is obtained. Similarly, expanding the determinant $\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}' & \mathbf{I} \end{vmatrix}$ by the last row gives $\begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}' & \mathbf{I} \end{vmatrix} = |\mathbf{A}|$.

(b) $\begin{vmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{0}' & \mathbf{B} \end{vmatrix} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0}' & \mathbf{B} \end{vmatrix} \begin{vmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0}' & \mathbf{I} \end{vmatrix}$. But expanding the determinant $\begin{vmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0}' & \mathbf{I} \end{vmatrix}$ by the last row gives $\begin{vmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0}' & \mathbf{I} \end{vmatrix} = 1$. Now use the result in Part a.

4.11. Show that, if $\mathbf{A}$ is square,

$$\begin{aligned} |\mathbf{A}| &= |\mathbf{A}_{22}||\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}| \quad \text{for } |\mathbf{A}_{22}| \neq 0 \\ &= |\mathbf{A}_{11}||\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}| \quad \text{for } |\mathbf{A}_{11}| \neq 0 \end{aligned}$$

Hint: Partition $\mathbf{A}$ and verify that

$$\begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}' & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0}' & \mathbf{A}_{22} \end{bmatrix}$$

Take determinants on both sides of this equality. Use Exercise 4.10 for the first and third determinants on the left and for the determinant on the right. The second equality for $|\mathbf{A}|$ follows by considering

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0}' & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0}' & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}$$

4.12. Show that, for $\mathbf{A}$ symmetric,

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0} \\ \mathbf{0}' & \mathbf{A}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}' & \mathbf{I} \end{bmatrix}$$

Thus, $(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}$ is the upper left-hand block of $\mathbf{A}^{-1}$.

Hint: Premultiply the expression in the hint to Exercise 4.11 by $\begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}' & \mathbf{I} \end{bmatrix}^{-1}$ and postmultiply by $\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix}^{-1}$. Take inverses of the resulting expression.

4.13. Show the following if $\mathbf{X}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| \neq 0$.

(a) Check that $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{22}||\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}|$. (Note that $|\boldsymbol{\Sigma}|$ can be factored into the product of contributions from the marginal and conditional distributions.)

(b) Check that

$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) &= [\mathbf{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)]' \\ &\quad \times (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}[\mathbf{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)] \\ &\quad + (\mathbf{x}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \end{aligned}$$

(Thus, the joint density exponent can be written as the sum of two terms corresponding to contributions from the conditional and marginal distributions.)

(c) Given the results in Parts a and b, identify the marginal distribution of $\mathbf{X}_2$ and the conditional distribution of $\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2$.

Hint:

(a) Apply Exercise 4.11.

(b) Note from Exercise 4.12 that we can write $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ as

$$\begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1} & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{22}^{-1} \end{bmatrix}$$
$$\times \begin{bmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0}' & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix}.$$

If we group the product so that

$$\begin{bmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0}' & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix}$$

the result follows.

4.14. If $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| \neq 0$, show that the joint density can be written as the product of marginal densities for

$$\underset{(q\times 1)}{\mathbf{X}_1} \quad \text{and} \quad \underset{((p-q)\times 1)}{\mathbf{X}_2} \quad \text{if} \quad \boldsymbol{\Sigma}_{12} = \underset{(q\times(p-q))}{\mathbf{0}}$$

Hint: Show by block multiplication that

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{22}^{-1} \end{bmatrix} \text{is the inverse of } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Then write

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = [(\mathbf{x}_1 - \boldsymbol{\mu}_1)', (\mathbf{x}_2 - \boldsymbol{\mu}_2)'] \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0}' & \boldsymbol{\Sigma}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{bmatrix}$$
$$= (\mathbf{x}_1 - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

Note that $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22}|$ from Exercise 4.10(a). Now factor the joint density.

4.15. Show that $\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}} - \boldsymbol{\mu})'$ and $\sum_{j=1}^{n} (\bar{\mathbf{x}} - \boldsymbol{\mu})(\mathbf{x}_j - \bar{\mathbf{x}})'$ are both $p \times p$ matrices of zeros. Here $\mathbf{x}_j' = [x_{j1}, x_{j2}, \ldots, x_{jp}]$, $j = 1, 2, \ldots, n$, and

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j$$

4.16. Let $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$, and $\mathbf{X}_4$ be independent $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vectors.

(a) Find the marginal distributions for each of the random vectors

$$\mathbf{V}_1 = \tfrac{1}{4}\mathbf{X}_1 - \tfrac{1}{4}\mathbf{X}_2 + \tfrac{1}{4}\mathbf{X}_3 - \tfrac{1}{4}\mathbf{X}_4$$

and

$$\mathbf{V}_2 = \tfrac{1}{4}\mathbf{X}_1 + \tfrac{1}{4}\mathbf{X}_2 - \tfrac{1}{4}\mathbf{X}_3 - \tfrac{1}{4}\mathbf{X}_4$$

(b) Find the joint density of the random vectors $\mathbf{V}_1$ and $\mathbf{V}_2$ defined in (a).

4.17. Let $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$, and $\mathbf{X}_5$ be independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Find the mean vector and covariance matrices for each of the two linear combinations of random vectors

$$\tfrac{1}{5}\mathbf{X}_1 + \tfrac{1}{5}\mathbf{X}_2 + \tfrac{1}{5}\mathbf{X}_3 + \tfrac{1}{5}\mathbf{X}_4 + \tfrac{1}{5}\mathbf{X}_5$$

and

$$\mathbf{X}_1 - \mathbf{X}_2 + \mathbf{X}_3 - \mathbf{X}_4 + \mathbf{X}_5$$

in terms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Also, obtain the covariance between the two linear combinations of random vectors.

4.18. Find the maximum likelihood estimates of the 2×1 mean vector $\boldsymbol{\mu}$ and the 2×2 covariance matrix $\boldsymbol{\Sigma}$ based on the random sample

$$\mathbf{X} = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

from a bivariate normal population.

4.19. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{20}$ be a random sample of size $n = 20$ from an $N_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Specify each of the following completely.

(a) The distribution of $(\mathbf{X}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu})$

(b) The distributions of $\overline{\mathbf{X}}$ and $\sqrt{n}(\overline{\mathbf{X}} - \boldsymbol{\mu})$

(c) The distribution of $(n - 1)\,\mathbf{S}$

4.20. For the random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{20}$ in Exercise 4.19, specify the distribution of $\mathbf{B}(19\mathbf{S})\mathbf{B}'$ in each case.

(a) $\mathbf{B} = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}$

(b) $\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

4.21. Let $\mathbf{X}_1, \ldots, \mathbf{X}_{60}$ be a random sample of size 60 from a four-variate normal distribution having mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Specify each of the following completely.

(a) The distribution of $\overline{\mathbf{X}}$

(b) The distribution of $(\mathbf{X}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu})$

(c) The distribution of $n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu})$

(d) The approximate distribution of $n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu})$

4.22. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{75}$ be a random sample from a population distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. What is the approximate distribution of each of the following?

(a) $\overline{\mathbf{X}}$

(b) $n(\overline{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{X}} - \boldsymbol{\mu})$

4.23. Consider the annual rates of return (including dividends) on the Dow-Jones industrial average for the years 1996–2005. These data, multiplied by 100, are

$-0.6 \quad 3.1 \quad 25.3 \quad -16.8 \quad -7.1 \quad -6.2 \quad 25.2 \quad 22.6 \quad 26.0.$

Use these 10 observations to complete the following.

(a) Construct a Q–Q plot. Do the data seem to be normally distributed? Explain.

(b) Carry out a test of normality based on the correlation coefficient r_Q. [See (4-31).] Let the significance level be $\alpha = .10$.

4.24. Exercise 1.4 contains data on three variables for the world's 10 largest companies as of April 2005. For the sales (x_1) and profits (x_2) data:

(a) Construct Q–Q plots. Do these data appear to be normally distributed? Explain.

(b) Carry out a test of normality based on the correlation coefficient r_Q. [See (4-31).] Set the significance level at $\alpha = .10$. Do the results of these tests corroborate the results in Part a?

4.25. Refer to the data for the world's 10 largest companies in Exercise 1.4. Construct a chi-square plot using all *three* variables. The chi-square quantiles are

0.3518 0.7978 1.2125 1.6416 2.1095 2.6430 3.2831 4.1083 5.3170 7.8147

4.26. Exercise 1.2 gives the age x_1, measured in years, as well as the selling price x_2, measured in thousands of dollars, for $n = 10$ used cars. These data are reproduced as follows:

x_1	1	2	3	3	4	5	6	8	9	11
x_2	18.95	19.00	17.95	15.54	14.00	12.95	8.94	7.49	6.00	3.99

(a) Use the results of Exercise 1.2 to calculate the squared statistical distances $(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$, $j = 1, 2, \ldots, 10$, where $\mathbf{x}_j' = [x_{j1}, x_{j2}]$.

(b) Using the distances in Part a, determine the proportion of the observations falling within the estimated 50% probability contour of a bivariate normal distribution.

(c) Order the distances in Part a and construct a chi-square plot.

(d) Given the results in Parts b and c, are these data approximately bivariate normal? Explain.

4.27. Consider the radiation data (with door closed) in Example 4.10. Construct a Q–Q plot for the natural logarithms of these data. [Note that the natural logarithm transformation corresponds to the value $\lambda = 0$ in (4-34).] Do the natural logarithms appear to be normally distributed? Compare your results with Figure 4.13. Does the choice $\lambda = \frac{1}{4}$ or $\lambda = 0$ make much difference in this case?

The following exercises may require a computer.

4.28. Consider the air-pollution data given in Table 1.5. Construct a Q–Q plot for the solar radiation measurements and carry out a test for normality based on the correlation coefficient r_Q [see (4-31)]. Let $\alpha = .05$ and use the entry corresponding to $n = 40$ in Table 4.2.

4.29. Given the air-pollution data in Table 1.5, examine the pairs $X_5 = NO_2$ and $X_6 = O_3$ for bivariate normality.

(a) Calculate statistical distances $(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$, $j = 1, 2, \ldots, 42$, where $\mathbf{x}_j' = [x_{j5}, x_{j6}]$.

(b) Determine the proportion of observations $\mathbf{x}_j' = [x_{j5}, x_{j6}]$, $j = 1, 2, \ldots, 42$, falling within the approximate 50% probability contour of a bivariate normal distribution.

(c) Construct a chi-square plot of the ordered distances in Part a.

4.30. Consider the used-car data in Exercise 4.26.

(a) Determine the power transformation $\hat{\lambda}_1$ that makes the x_1 values approximately normal. Construct a Q–Q plot for the transformed data.

(b) Determine the power transformations $\hat{\lambda}_2$ that makes the x_2 values approximately normal. Construct a Q–Q plot for the transformed data.

(c) Determine the power transformations $\hat{\boldsymbol{\lambda}}' = [\hat{\lambda}_1, \hat{\lambda}_2]$ that make the $[x_1, x_2]$ values jointly normal using (4-40). Compare the results with those obtained in Parts a and b.

4.31. Examine the marginal normality of the observations on variables $X_1, X_2, \ldots, X_5$ for the multiple-sclerosis data in Table 1.6. Treat the non-multiple-sclerosis and multiple-sclerosis groups separately. Use whatever methodology, including transformations, you feel is appropriate.

4.32. Examine the marginal normality of the observations on variables $X_1, X_2, \ldots, X_6$ for the radiotherapy data in Table 1.7. Use whatever methodology, including transformations, you feel is appropriate.

4.33. Examine the marginal and bivariate normality of the observations on variables X_1, X_2, X_3, and X_4 for the data in Table 4.3.

4.34. Examine the data on bone mineral content in Table 1.8 for marginal and bivariate normality.

4.35. Examine the data on paper-quality measurements in Table 1.2 for marginal and multivariate normality.

4.36. Examine the data on women's national track records in Table 1.9 for marginal and multivariate normality.

4.37. Refer to Exercise 1.18. Convert the women's track records in Table 1.9 to speeds measured in meters per second. Examine the data on speeds for marginal and multivariate normality.

4.38. Examine the data on bulls in Table 1.10 for marginal and multivariate normality. Consider only the variables YrHgt, FtFrBody, PrctFFB, BkFat, SaleHt, and SaleWt.

4.39. The data in Table 4.6 (see the psychological profile data: www.prenhall.com/statistics) consist of 130 observations generated by scores on a psychological test administered to Peruvian teenagers (ages 15, 16, and 17). For each of these teenagers the gender (male = 1, female = 2) and socioeconomic status (low = 1, medium = 2) were also recorded. The scores were accumulated into five subscale scores labeled *independence* (indep), *support* (supp), *benevolence* (benev), *conformity* (conform), and *leadership* (leader).

Table 4.6 Psychological Profile Data

Indep	Supp	Benev	Conform	Leader	Gender	Socio
27	13	14	20	11	2	1
12	13	24	25	6	2	1
14	20	15	16	7	2	1
18	20	17	12	6	2	1
9	22	22	21	6	2	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	11	26	17	10	1	2
14	12	14	11	29	1	2
19	11	23	18	13	2	2
27	19	22	7	9	2	2
10	17	22	22	8	2	2

Source: Data courtesy of C. Soto.

(a) Examine each of the variables independence, support, benevolence, conformity and leadership for marginal normality.

(b) Using all five variables, check for multivariate normality.

(c) Refer to part (a). For those variables that are nonnormal, determine the transformation that makes them more nearly normal.

4.40. Consider the data on national parks in Exercise 1.27.

(a) Comment on any possible outliers in a scatter plot of the original variables.

(b) Determine the power transformation $\hat{\lambda}_1$ the makes the x_1 values approximately normal. Construct a $Q-Q$ plot of the transformed observations.

(c) Determine the power transformation $\hat{\lambda}_2$ the makes the x_2 values approximately normal. Construct a $Q-Q$ plot of the transformed observations.

(d) Determine the power transformation for approximate bivariate normality using (4-40).

4.41. Consider the data on snow removal in Exercise 3.20.

(a) Comment on any possible outliers in a scatter plot of the original variables.

(b) Determine the power transformation $\hat{\lambda}_1$ the makes the x_1 values approximately normal. Construct a $Q-Q$ plot of the transformed observations.

(c) Determine the power transformation $\hat{\lambda}_2$ the makes the x_2 values approximately normal. Construct a $Q-Q$ plot of the transformed observations.

(d) Determine the power transformation for approximate bivariate normality using (4-40).

References

1. Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York: John Wiley, 2003.

2. Andrews, D. F., R. Gnanadesikan, and J. L. Warner. "Transformations of Multivariate Data." *Biometrics*, **27**, no. 4 (1971), 825–840.

3. Box, G. E. P., and D. R. Cox. "An Analysis of Transformations" (with discussion). *Journal of the Royal Statistical Society (B)*, **26**, no. 2 (1964), 211–252.

4. Daniel, C. and F. S. Wood, *Fitting Equations to Data: Computer Analysis of Multifactor Data*. New York: John Wiley, 1980.

5. Filliben, J. J. "The Probability Plot Correlation Coefficient Test for Normality." *Technometrics*, **17**, no. 1 (1975), 111–117.

6. Gnanadesikan, R. *Methods for Statistical Data Analysis of Multivariate Observations* (2nd ed.). New York: Wiley-Interscience, 1977.

7. Hawkins, D. M. *Identification of Outliers*. London, UK: Chapman and Hall, 1980.

8. Hernandez, F., and R. A. Johnson. "The Large-Sample Behavior of Transformations to Normality." *Journal of the American Statistical Association*, **75**, no. 372 (1980), 855–861.

9. Hogg, R. V., Craig. A. T. and J. W. Mckean *Introduction to Mathematical Statistics* (6th ed.). Upper Saddle River, N.J.: Prentice Hall, 2004.

10. Looney, S. W., and T. R. Gulledge, Jr. "Use of the Correlation Coefficient with Normal Probability Plots." *The American Statistician*, **39**, no. 1 (1985), 75–79.

11. Mardia, K. V., Kent, J. T. and J. M. Bibby. *Multivariate Analysis* (Paperback). London: Academic Press, 2003.

12. Shapiro, S. S., and M. B. Wilk. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika*, **52**, no. 4 (1965), 591–611.