7

Multivariate Statistical Models

7.1 Introduction

Often we are not interested merely in a single random variable but rather in the joint behavior of several random variables, for example, returns on several assets and a market index. Multivariate distributions describe such joint behavior. This chapter is an introduction to the use of multivariate distributions for modeling financial markets data. Readers with little prior knowledge of multivariate distributions may benefit from reviewing Sections A.12–A.14 before reading this chapter.

7.2 Covariance and Correlation Matrices

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^{\mathsf{T}}$ be a random vector. We define the expectation vector of $\boldsymbol{Y}$ to be

$$E(\boldsymbol{Y}) = \begin{pmatrix} E(Y_1) \\ \vdots \\ E(Y_d) \end{pmatrix}.$$

The *covariance matrix* of $\boldsymbol{Y}$ is the matrix whose (i,j)th entry is $\mathrm{Cov}(Y_i, Y_j)$ for $i, j = 1, \ldots, N$. Since $\mathrm{Cov}(Y_i, Y_i) = \mathrm{Var}(Y_i)$, the covariance matrix is

$$\mathrm{COV}(\boldsymbol{Y}) = \begin{pmatrix} \mathrm{Var}(Y_1) & \mathrm{Cov}(Y_1, Y_2) & \cdots & \mathrm{Cov}(Y_1, Y_d) \\ \mathrm{Cov}(Y_2, Y_1) & \mathrm{Var}(Y_2) & \cdots & \mathrm{Cov}(Y_2, Y_d) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(Y_d, Y_1) & \mathrm{Cov}(Y_d, Y_2) & \cdots & \mathrm{Var}(Y_d) \end{pmatrix}.$$

Similarly, the *correlation matrix* of $\boldsymbol{Y}$, denoted $\mathrm{CORR}(\boldsymbol{Y})$, has i, jth element $\rho_{Y_i Y_j}$. Because $\mathrm{Corr}(Y_i, Y_i) = 1$ for all i, the diagonal elements of a correlation matrix are all equal to 1. Note the use of "COV" and "CORR" to denote matrices and "Cov" and "Corr" to denote scalars.

D. Ruppert, *Statistics and Data Analysis for Financial Engineering*, Springer Texts in Statistics,
DOI 10.1007/978-1-4419-7787-8_7,

The covariance matrix can be written as

$$\mathrm{COV}(\boldsymbol{Y}) = E\left[\{\boldsymbol{Y} - E(\boldsymbol{Y})\}\{\boldsymbol{Y} - E(\boldsymbol{Y})\}^{\mathsf{T}}\right]. \tag{7.1}$$

There are simple relationships between the covariance and correlation matrices. Let $\boldsymbol{S} = \mathrm{diag}(\sigma_{Y_1}, \ldots, \sigma_{Y_d})$, where σ_{Y_i} is the standard deviation of Y_i. Then

$$\mathrm{CORR}(\boldsymbol{Y}) = \boldsymbol{S}^{-1}\mathrm{COV}(\boldsymbol{Y})\boldsymbol{S}^{-1} \tag{7.2}$$

and, equivalently,

$$\mathrm{COV}(\boldsymbol{Y}) = \boldsymbol{S}\,\mathrm{CORR}(\boldsymbol{Y})\,\boldsymbol{S}. \tag{7.3}$$

The *sample covariance* and *correlation matrices* replace $\mathrm{Cov}(Y_i, Y_j)$ and $\rho_{Y_i Y_j}$ by their estimates given by (A.29) and (A.30).

A *standardized* variable is obtained by subtracting the variable's mean and dividing the difference by the variable's standard deviation. After standardization, a variable has a mean equal to 0 and a standard deviation equal to 1. The covariance matrix of standardized variables equals the correlation matrix of original variables, which is also the correlation matrix of the standardized variables.

Example 7.1. CRSPday covariances and correlations

This example uses the `CRSPday` data set in `R`'s `Ecdat` package. There are four variables, daily returns from January 3, 1969, to December 31, 1998, on three stocks, GE, IBM, and Mobil, and on the CRSP value-weighted index, including dividends. CRSP is the Center for Research in Security Prices at the University of Chicago. The sample covariance matrix for these four series is

```
            ge      ibm    mobil     crsp
ge    1.88e-04 8.01e-05 5.27e-05 7.61e-05
ibm   8.01e-05 3.06e-04 3.59e-05 6.60e-05
mobil 5.27e-05 3.59e-05 1.67e-04 4.31e-05
crsp  7.61e-05 6.60e-05 4.31e-05 6.02e-05
```

It is difficult to get much information just by inspecting the covariance matrix. The covariance between two random variables depends on their variances as well as the strength of the linear relationship between them. Covariance matrices are extremely important as input to, for example, a portfolio analysis, but to understand the relationship between variables, it is much better to examine their sample correlation matrix. The sample correlation matrix in this example is

```
          ge   ibm mobil  crsp
ge    1.000 0.334 0.297 0.715
ibm   0.334 1.000 0.159 0.486
mobil 0.297 0.159 1.000 0.429
crsp  0.715 0.486 0.429 1.000
```

We can see that all sample correlations are positive and the largest correlations are between `crsp` and the individual stocks. GE is the stock most highly correlated with `crsp`. The correlations between individual stocks and a market index such as `crsp` are a key component of finance theory, especially the Capital Asset Pricing Model (CAPM) introduced in Chapter 16.

□

7.3 Linear Functions of Random Variables

Often we are interested in finding the expectation and variance of a linear combination (weighted average) of random variables. For example, consider returns on a set of assets. A *portfolio* is simply a weighted average of the assets with weights that sum to one. The weights specify what fractions of the total investment are allocated to the assets. For example, if a portfolio consists of 200 shares of Stock 1 selling at \$88/share and 150 shares of Stock 2 selling at \$67/share, then the weights are

$$w_1 = \frac{(200)(88)}{(200)(88) + (150)(67)} = 0.637 \quad \text{and} \quad w_2 = 1 - w_1 = 0.363. \tag{7.4}$$

Because the return on a portfolio is a linear combination of the returns on the individual assets in the portfolio, the material in this section is used extensively in the portfolio theory of Chapters 11 and 16.

First, we look at a linear function of a single random variable. If Y is a random variable and a and b are constants, then

$$E(aY + b) = aE(Y) + b.$$

Also,

$$\text{Var}(aY + b) = a^2\text{Var}(Y) \text{ and } \sigma_{aY+b} = |a|\sigma_Y.$$

Next, we consider linear combinations of two random variables. If X and Y are random variables and w_1 and w_2 are constants, then

$$E(w_1X + w_2Y) = w_1E(X) + w_2E(Y),$$

and

$$\text{Var}(w_1X + w_2Y) = w_1^2\text{Var}(X) + 2w_1w_2\text{Cov}(X, Y) + w_2^2\text{Var}(Y). \tag{7.5}$$

Check that (7.5) can be reexpressed as

$$\mathrm{Var}(w_1 X + w_2 Y) = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \tag{7.6}$$

Although formula (7.6) may seem unnecessarily complicated, we will show that this equation generalizes in an elegant way to more than two random variables; see (7.7) below. Notice that the matrix in (7.6) is the covariance matrix of the random vector $\begin{pmatrix} X & Y \end{pmatrix}^{\mathsf{T}}$.

Let $\boldsymbol{w} = (w_1, \ldots, w_d)^{\mathsf{T}}$ be a vector of weights and let $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ be a random vector. Then

$$\boldsymbol{w}^{\mathsf{T}}\boldsymbol{Y} = \sum_{i=1}^{N} w_i Y_i$$

is a weighted average of the components of $\boldsymbol{Y}$. One can easily show that

$$E(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{Y}) = \boldsymbol{w}^{\mathsf{T}}\{E(\boldsymbol{Y})\}$$

and

$$\mathrm{Var}(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{Y}) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i\, w_j\, \mathrm{Cov}(Y_i, Y_j).$$

This last result can be expressed more succinctly using vector/matrix notation:

$$\mathrm{Var}(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{Y}) = \boldsymbol{w}^{\mathsf{T}}\mathrm{COV}(\boldsymbol{Y})\boldsymbol{w}. \tag{7.7}$$

Example 7.2. The variance of a linear combination of correlated random variables

Suppose that $\boldsymbol{Y} = (Y_1\ Y_2\ Y_3)^{\mathsf{T}}$, $\mathrm{Var}(Y_1) = 2$, $\mathrm{Var}(Y_2) = 3$, $\mathrm{Var}(Y_3) = 5$, $\rho_{Y_1,Y_2} = 0.6$, and that Y_1 and Y_2 are independent of Y_3. Find $\mathrm{Var}(Y_1 + Y_2 + 1/2\, Y_3)$.

Answer: The covariance between Y_1 and Y_3 is 0 by independence, and the same is true of Y_2 and Y_3. The covariance between Y_1 and Y_2 is $(0.6)\sqrt{(2)(3)} = 1.47$. Therefore,

$$\mathrm{COV}(\boldsymbol{Y}) = \begin{pmatrix} 2 & 1.47 & 0 \\ 1.47 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix},$$

and by (7.7),

$$\mathrm{Var}(Y_1 + Y_2 + Y_3/2) = \begin{pmatrix} 1 & 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 & 1.47 & 0 \\ 1.47 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \frac{1}{2} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 3.47 \\ 4.47 \\ 2.5 \end{pmatrix}$$
$$= 9.19.$$

□

A important property of a covariance matrix $\mathrm{COV}(\boldsymbol{Y})$ is that it is symmetric and positive semidefinite. A matrix $\boldsymbol{A}$ is said to be positive semidefinite (definite) if $\boldsymbol{x}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{x} \geq 0$ (> 0) for all vectors $\boldsymbol{x} \neq 0$. By (7.7), any covariance matrix must be positive semidefinite, because otherwise there would exist a random variable with a negative variance, a contradiction. A nonsingular covariance matrix is positive definite. A covariance matrix must be symmetric because $\rho_{Y_i Y_j} = \rho_{Y_j Y_i}$ for every i and j.

7.3.1 Two or More Linear Combinations of Random Variables

More generally, suppose that $\boldsymbol{w}_1^{\mathsf{T}}\boldsymbol{Y}$ and $\boldsymbol{w}_2^{\mathsf{T}}\boldsymbol{Y}$ are two weighted averages of the components of $\boldsymbol{Y}$, e.g., returns on two different portfolios. Then

$$\mathrm{Cov}(\boldsymbol{w}_1^{\mathsf{T}}\boldsymbol{Y}, \boldsymbol{w}_2^{\mathsf{T}}\boldsymbol{Y}) = \boldsymbol{w}_1^{\mathsf{T}}\mathrm{COV}(\boldsymbol{Y})\boldsymbol{w}_2 = \boldsymbol{w}_2^{\mathsf{T}}\mathrm{COV}(\boldsymbol{Y})\boldsymbol{w}_1. \tag{7.8}$$

Example 7.3. (Example 7.2 *continued)*

Suppose that the random vector $\boldsymbol{Y} = (Y_1, Y_2, Y_3)^{\mathsf{T}}$ has the mean vector and covariance matrix used in the previous example and contains the returns on three assets. Find the covariance between a portfolio that allocates 1/3 to each of the three assets and a second portfolio that allocates 1/2 to each of the first two assets. That is, find the covariance between $(Y_1 + Y_2 + Y_3)/3$ and $(Y_1 + Y_2)/2$.

Answer: Let

$$\boldsymbol{w}_1 = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}^{\mathsf{T}}$$

and

$$\boldsymbol{w}_2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}^{\mathsf{T}}.$$

Then

$$\mathrm{Cov}\left\{\frac{Y_1 + Y_2}{2}, \frac{Y_1 + Y_2 + Y_3}{3}\right\} = \boldsymbol{w}_1^{\mathsf{T}}\mathrm{COV}(\boldsymbol{Y})\boldsymbol{w}_2$$
$$= \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} 2 & 1.47 & 0 \\ 1.47 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1.157 & 1.490 & 1.667 \end{pmatrix} \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$$

$$= 1.323.$$

□

Let $\boldsymbol{W}$ be a nonrandom $N \times q$ matrix so that $\boldsymbol{W}^{\mathsf{T}}\boldsymbol{Y}$ is a random vector of q linear combinations of $\boldsymbol{Y}$. Then (7.7) can be generalized to

$$\mathrm{COV}(\boldsymbol{W}^{\mathsf{T}}\boldsymbol{Y}) = \boldsymbol{W}^{\mathsf{T}}\mathrm{COV}(\boldsymbol{Y})\boldsymbol{W}. \tag{7.9}$$

Let $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ be two random vectors of dimensions n_1 and n_2, respectively. Then $\boldsymbol{\Sigma}_{Y_1,Y_2} = \mathrm{COV}(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$ is defined as the $n_1 \times n_2$ matrix whose i,jth element is the covariance between the ith component of $\boldsymbol{Y}_1$ and the jth component of $\boldsymbol{Y}_2$, that is, $\boldsymbol{\Sigma}_{Y_1,Y_2}$ is the matrix of covariances between the random vectors $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$.

It is not difficult to show that

$$\mathrm{Cov}(\boldsymbol{w}_1^{\mathsf{T}}\boldsymbol{Y}_1, \boldsymbol{w}_2^{\mathsf{T}}\boldsymbol{Y}_2) = \boldsymbol{w}_1^{\mathsf{T}}\mathrm{COV}(\boldsymbol{Y}_1, \boldsymbol{Y}_2)\boldsymbol{w}_2, \tag{7.10}$$

for constant vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ of lengths n_1 and n_2.

7.3.2 Independence and Variances of Sums

If $Y_1, \ldots, Y_d$ are independent, or at least uncorrelated, then

$$\mathrm{Var}\left(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{Y}\right) = \mathrm{Var}\left(\sum_{i=1}^{n} w_i Y_i\right) = \sum_{i=1}^{n} w_i^2 \mathrm{Var}(Y_i). \tag{7.11}$$

When $\boldsymbol{w}^{\mathsf{T}} = (1/n, \ldots, 1/n)$ so that $\boldsymbol{w}^{\mathsf{T}}\boldsymbol{Y} = \overline{Y}$, then we obtain that

$$\mathrm{Var}(\overline{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Y_i). \tag{7.12}$$

In particular, if $\mathrm{Var}(Y_i) = \sigma^2$ for all i, then we obtain the well-known result that if $Y_1, \ldots, Y_d$ are uncorrelated and have a constant variance σ^2, then

$$\mathrm{Var}(\overline{Y}) = \frac{\sigma^2}{n}. \tag{7.13}$$

Another useful fact that follows from (7.11) is that if Y_1 and Y_2 are uncorrelated, then

$$\mathrm{Var}(Y_1 - Y_2) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2). \tag{7.14}$$

7.4 Scatterplot Matrices

A correlation coefficient is only a summary of the linear relationship between variables. Interesting features, such as nonlinearity or the joint behavior of extreme values, remain hidden when only correlations are examined. A solution to this problem is the so-called scatterplot matrix, which is a matrix of scatterplots, one for each pair of variables. A scatterplot matrix can be created easily with modern statistical software such as `R`. Figure 7.1 shows a scatterplot matrix for the `CRSPday` data set.

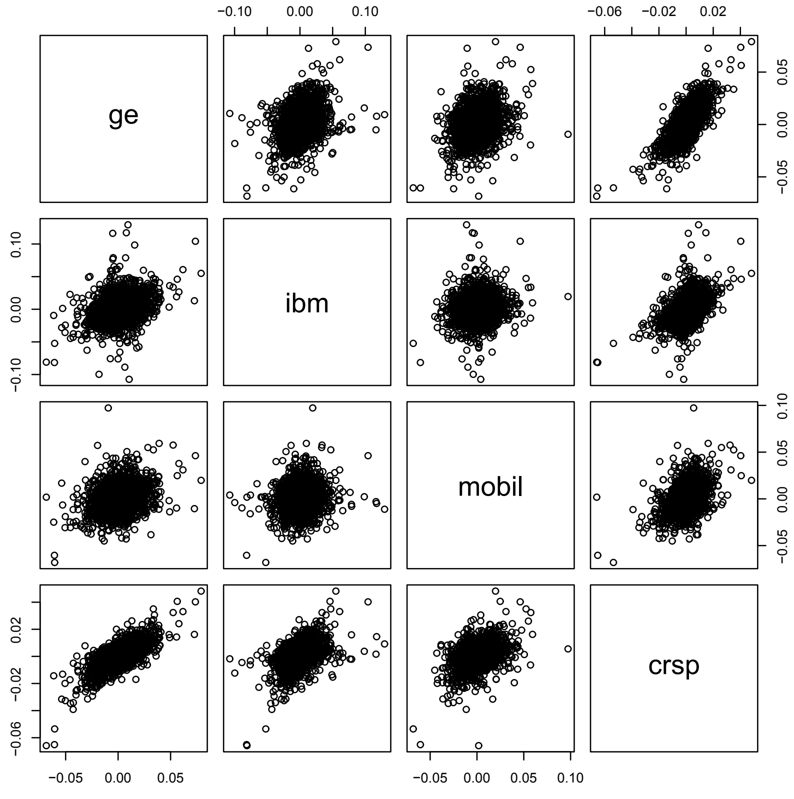

Fig. 7.1. *Scatterplot matrix for the* `CRSPday` *data set.*

One sees little evidence of nonlinear relationships in Figure 7.1. This lack of nonlinearities is typical of returns on equities, but it should not be taken for granted—instead, one should always look at the scatterplot matrix. The strong linear association between GE and `crsp`, which was suggested before by their high correlation coefficient, can be seen also in their scatterplot.

A portfolio is riskier if large negative returns on its assets tend to occur together on the same days. To investigate whether extreme values tend to cluster in this way, one should look at the scatterplots. In the scatterplot for IBM and Mobil, extreme returns for one stock do not tend to occur on the same days as extreme returns on the other stock; this can be seen by noticing that the outliers tend to fall along the x- and y-axes. The extreme-value behavior is different with GE and `crsp`, where extreme values are more likely to occur together; note that the outliers have a tendency to occur together, that is, in the upper-right and lower-left corners, rather than being concentrated along the axes. The IBM and Mobil scatterplot is said to show *tail independence*. In contrast, the GE and `crsp` scatterplot is said to show *tail dependence*. Tail dependence is explored further in Chapter 8.

7.5 The Multivariate Normal Distribution

In Chapter 5 we saw the importance of having parametric families of univariate distributions as statistical models. Parametric families of multivariate distributions are equally useful, and the multivariate normal family is the best known of them.

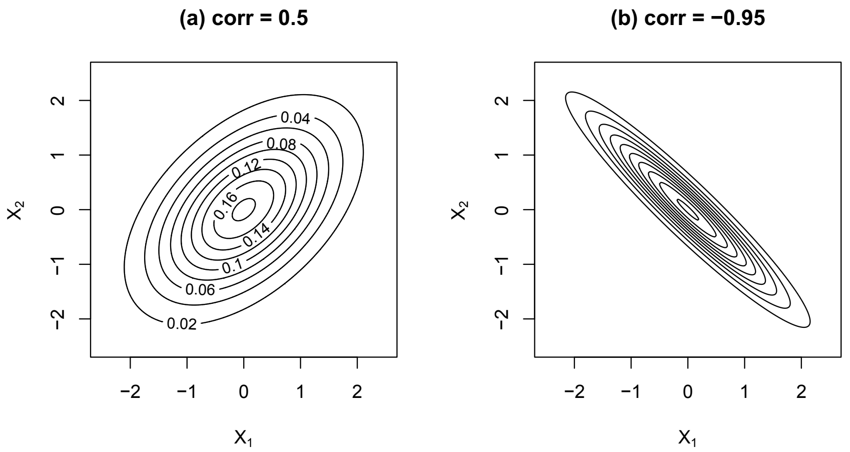

Fig. 7.2. *Contour plots of a bivariate normal densities with $N(0,1)$ marginal distributions and correlations of 0.5 or -0.95.*

The random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\mathsf{T}$ has a d-dimensional *multivariate normal distribution* with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\mathsf{T}$ and covariance matrix $\boldsymbol{\Sigma}$ if its probability density function is

$$\phi_d(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left[\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}}\right] \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right\}, \tag{7.15}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. The quantity in square brackets is a constant that normalizes the density so that it integrates to 1. The density depends on $\boldsymbol{y}$ only through $(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})$, and so the density is constant on each ellipse $\{\boldsymbol{y} : (\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = c\}$. Here $c > 0$ is a fixed constant that determines the size of the ellipse, with larger values of c giving smaller ellipses, each centered at $\boldsymbol{\mu}$. Such densities are called *elliptically contoured*. Figure 7.2 has contour plots of bivariate normal densities. Both Y_1 and Y_2 are $N(0,1)$ and the correlation between Y_1 and Y_2 is 0.5 in panel (a) or -0.95 in panel (b). Notice how the orientations of the contours depend on the sign and magnitude of the correlation. In panel (a) we can see that the height of the density is constant on ellipses and decreases with the distance from the mean, which is (0, 0). The same behavior occurs in panel (b), but, because of the high correlation, the contours are so close together that it was not possible to label them.

If $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^{\mathsf{T}}$ has a multivariate normal distribution, then for *every* set of constants $\boldsymbol{c} = (c_1, \ldots, c_d)^{\mathsf{T}}$, the weighted average (linear combination) $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{Y} = c_1 Y_1 + \cdots + c_d Y_d$ has a normal distribution with mean $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{\mu}$ and variance $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{c}$. In particular, the marginal distribution of Y_i is $N(\mu_i, \sigma_i^2)$, where σ_i^2 is the ith diagonal element of $\boldsymbol{\Sigma}$—to see this, take $c_i = 1$ and $c_j = 0$ for $j \neq i$.

The assumption of multivariate normality facilitates many useful probability calculations. If the returns on a set of assets have a multivariate normal distribution, then the return on any portfolio formed from these assets will be normally distributed. This is because the return on the portfolio is the weighted average of the returns on the assets. Therefore, the normal distribution could be used, for example, to find the probability of a loss of some size of interest, say, 10% or more, on the portfolio. Such calculations have important applications in finding a value-at-risk; see Chapter 19.

Unfortunately, we saw in Chapter 5 that often individual returns are not normally distributed, which implies that a vector of returns will not have a multivariate normal distribution. In Section 7.6 we will look at an important class of heavy-tailed multivariate distributions.

7.6 The Multivariate t-Distribution

We have seen that the univariate t-distribution is a good model for the returns of individual assets. Therefore, it is desirable to have a model for vectors of returns such that the univariate marginals are t-distributed. The multivariate t-distribution has this property. The random vector $\boldsymbol{Y}$ has a multivariate $t_\nu(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ distribution if

$$\boldsymbol{Y} = \boldsymbol{\mu} + \sqrt{\frac{\nu}{W}}\boldsymbol{Z}, \tag{7.16}$$

where W is chi-squared distributed with ν degrees of freedom, $\boldsymbol{Z}$ is $N_d(0, \boldsymbol{\Lambda})$ distributed, and W and $\boldsymbol{Z}$ are independent. Thus, the multivariate t-distribution is a continuous scale mixture of multivariate normal distributions. Extreme values of $\boldsymbol{Z}$ tend to occur when W is near zero. Since $W^{-1/2}$ multiplies all components of $\boldsymbol{Z}$, outliers in one component tend to occur with outliers in other components, that is, there is tail dependence.

For $\nu > 1$, $\boldsymbol{\mu}$ is the mean vector of $\boldsymbol{Y}$. For $0 < \nu \leq 1$, the expectation of $\boldsymbol{Y}$ does not exist, but $\boldsymbol{\mu}$ can still be regarded as the "center" of the distribution of $\boldsymbol{Y}$ because, for any value of ν, the vector $\boldsymbol{\mu}$ contains the medians of the components of $\boldsymbol{Y}$ and the contours of the density of $\boldsymbol{Y}$ are ellipses centered at $\boldsymbol{\mu}$.

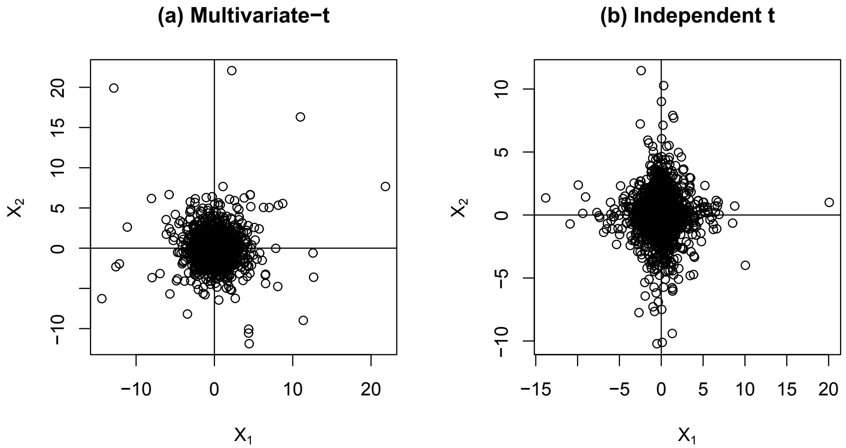

Fig. 7.3. *(a) Plot of a random sample from a bivariate t-distribution with $\nu = 3$, $\boldsymbol{\mu} = (0\ 0)^{\mathsf{T}}$ and identity covariate matrix. (b) Plot of a random sample of pairs of independent $t_3(0,1)$ random variables. Both sample sizes are 2500.*

For $\nu > 2$, the covariance matrix of $\boldsymbol{Y}$ exists and is

$$\boldsymbol{\Sigma} = \frac{\nu}{\nu - 2}\boldsymbol{\Lambda}. \tag{7.17}$$

We will call $\boldsymbol{\Lambda}$ the *scale matrix*. The scale matrix exists for all values of ν. Since the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{Y}$ is just a multiple of the covariance matrix $\boldsymbol{\Lambda}$ of $\boldsymbol{Z}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ have the same correlation matrices, assuming $\nu > 2$ so that the correlation matrix of $\boldsymbol{Y}$ exists. If $\Sigma_{i,j} = 0$, then Y_i and Y_j are uncorrelated, but they are dependent, nonetheless, because of the tail dependence. Tail dependence is illustrated in Figure 7.3, where panel (a) is a plot of 2500 observations from an uncorrelated bivariate t-distribution with marginal distributions that are $t_3(0,1)$. For comparison, panel (b) is a plot of 2500 observations of pairs of independent $t_3(0,1)$ random variables—these pairs do not have a bivariate t-distribution. Notice that in (b), outliers in Y_1 are not associated with outliers in Y_2, since the outliers are concentrated near the x- and y-axes. In contrast, outliers in (a) are distributed uniformly in all directions. The univariate marginal distributions are the same in (a) and (b).

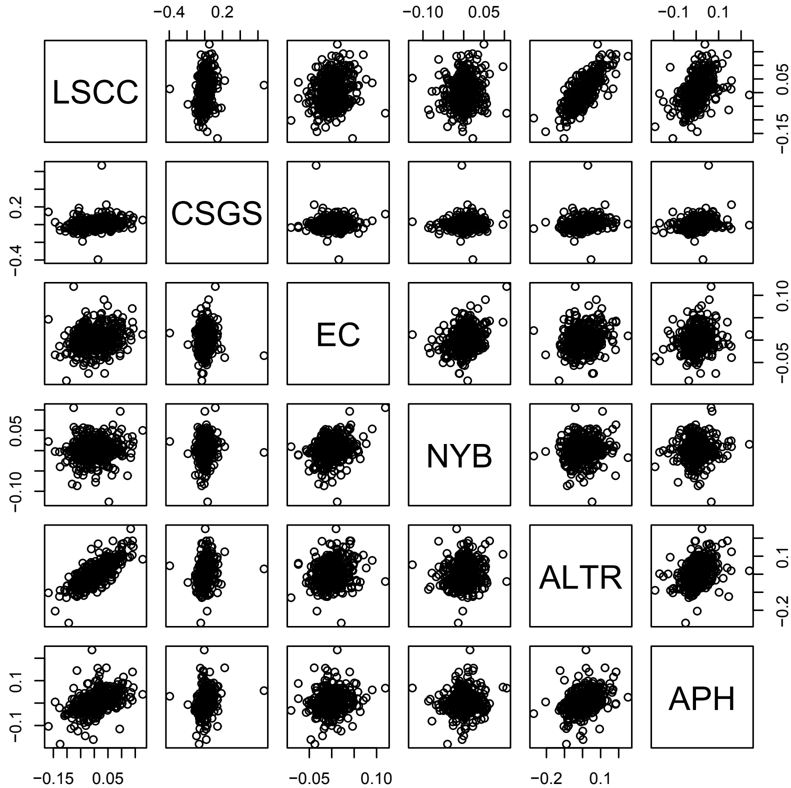

Fig. 7.4. *Scatterplot matrix of* 500 *daily returns on six midcap stocks in* `R`*'s* `midcapD.ts` *data set.*

Tail dependence can be expected in equity returns. For example, on Black Monday, almost all equities had extremely large negative returns. Of course, Black Monday was an extreme, even among extreme events. We would not want to reach any general conclusions based upon Black Monday alone. However, in Figure 7.1, we see little evidence that outliers are concentrated along the axes, with the possible exception of the scatterplot for IBM and Mobil. As another example of dependencies among stock returns, Figure 7.4 contains a scatterplot matrix of returns on six midcap stocks in the `midcapD.ts` data set in `R`'s in `fEcofin` package. Again, tail dependence can be seen. This suggests that tail dependence is common among equity returns and the multivariate t-distribution is a promising model for them.

7.6.1 Using the *t*-Distribution in Portfolio Analysis

If Y has a $t_\nu(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ distribution, which we recall has covariance matrix $\boldsymbol{\Sigma} = \{\nu/(\nu-2)\}\boldsymbol{\Lambda}$, and $\boldsymbol{w}$ is a vector of weights, then $\boldsymbol{w}^\mathsf{T}\boldsymbol{Y}$ has a univariate t-distribution with mean $\boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}$ and variance $\{\nu/(\nu-2)\}\boldsymbol{w}^\mathsf{T}\boldsymbol{\Lambda}\boldsymbol{w} = \boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}$. This fact can be useful when computing risk measures for a portfolio. If the returns on the assets have a multivariate t-distribution, then the return on the portfolio will have a univariate t-distribution. We will make use of this result in Chapter 19.

7.7 Fitting the Multivariate *t*-Distribution by Maximum Likelihood

To estimate the parameters of a multivariate t-distribution, one can use the function `cov.trob` in `R`'s `MASS` package. This function computes the maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ with ν fixed. To estimate ν, one computes the profile log-likelihood for ν and finds the value, $\widehat{\nu}$, of ν that maximizes the profile log-likelihood. Then the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ are the estimates from `cov.trob` with ν fixed at $\widehat{\nu}$.

Example 7.4. Fitting the `CRSPday` *data*

This example uses the data set `CRSPday` analyzed earlier in Example 7.1. Recall that there are four variables, returns on GE, IBM, Mobil, and the CRSP index. The profile log-likelihood is plotted in Figure 7.5. In that figure, one see that the MLE of ν is 5.94, and there is relatively little uncertainty about this parameter's value—the 95% profile likelihood confidence interval is (5.41, 6.55).

AIC for this model is 7.42 plus 64,000. Here AIC values are expressed as deviations from 64,000 to keep these values small. This is helpful when comparing two or more models via AIC. Subtracting the same constant from all AIC values, of course, has no effect on model comparisons.

The maximum likelihood estimates of the mean vector and the correlation matrix are called `$center` and `$cor`, respectively, in the following output:

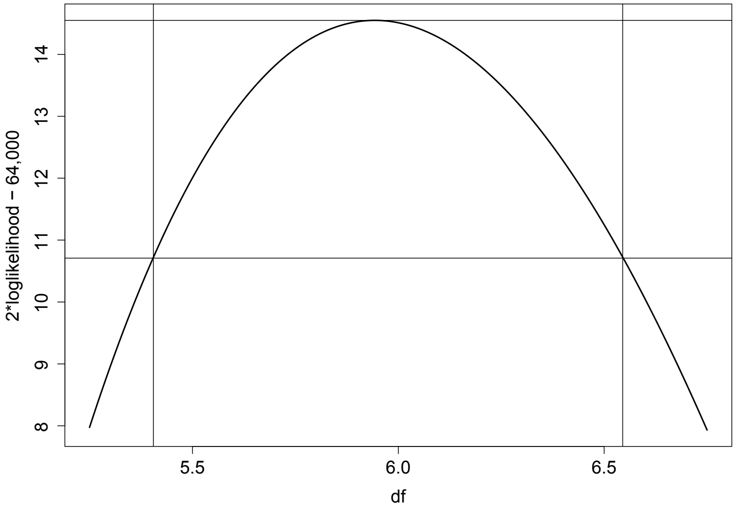

Fig. 7.5. `CRSPday` *data. A profile likelihood confidence interval for* ν*. The solid curve is* $2L_{\max}(\nu)$*, where* $L_{\max}(\nu)$ *is the profile likelihood minus 32,000. 32,000 was subtracted from the profile likelihood to simplify the labeling of the y-axis. The horizontal line intersects the y-axis at* $2L_{\max}(\widehat{\nu}) - \chi^2_{\alpha,1}$*, where* $\widehat{\nu}$ *is the MLE and* $\alpha = 0.05$*. All values of* ν *such that* $2L_{\max}(\nu)$ *is above the horizontal line are in the profile likelihood 95% confidence interval. The two vertical lines intersect the x-axis at 5.41 and 6.55, the endpoints of the confidence interval.*

```
$center
[1] 0.0009424 0.0004481 0.0006883 0.0007693

$cor
       [,1]   [,2]   [,3]   [,4]
[1,] 1.0000 0.3192 0.2845 0.6765
[2,] 0.3192 1.0000 0.1584 0.4698
[3,] 0.2845 0.1584 1.0000 0.4301
[4,] 0.6765 0.4698 0.4301 1.0000
```

These estimates were computed using `cov.trob` with ν fixed at 5.94.

When the data are t-distributed, the maximum likelihood estimates are superior to the sample mean and covariance matrix in several respects—the MLE is more accurate and it is less sensitive to outliers. However, in this example, the maximum likelihood estimates are similar to the sample mean and correlation matrix. For example, the sample correlation matrix is

```
          ge    ibm  mobil   crsp
ge    1.0000 0.3336 0.2972 0.7148
ibm   0.3336 1.0000 0.1587 0.4864
mobil 0.2972 0.1587 1.0000 0.4294
crsp  0.7148 0.4864 0.4294 1.0000
```

□

7.8 Elliptically Contoured Densities

The multivariate normal and t-distributions have *elliptically contoured* densities, a property that will be discussed in this section. A d-variate multivariate density f is elliptically contoured if can be expressed as

$$f(\boldsymbol{y}) = |\boldsymbol{\Lambda}|^{-1/2} g\left\{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right\}, \tag{7.18}$$

where g is a nonnegative-valued function such that $1 = \int_{\Re^d} g\left(\|\boldsymbol{y}\|^2\right) d\boldsymbol{y}$, $\boldsymbol{\mu}$ is a $d \times 1$ vector, and $\boldsymbol{\Lambda}$ is a $d \times d$ symmetric, positive definite matrix. Usually, $g(x)$ is a decreasing function of $x \geq 0$, and we will assume this is true. We will also assume the finiteness of second moments, in which case $\boldsymbol{\mu}$ is the mean vector and the covariance matrix $\boldsymbol{\Sigma}$ is a scalar multiple of $\boldsymbol{\Lambda}$.

For each fixed $c > 0$,

$$\mathcal{E}(c) = \{\boldsymbol{y} : (\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = c\}$$

is an ellipse centered at μ, and if $c_1 > c_2$, then $\mathcal{E}(c_1)$ is inside $\mathcal{E}(c_2)$ because g is decreasing. The contours of f are concentric ellipses as can be seen in Figure 7.6. That figure shows the contours of the bivariate t_4-density with $\boldsymbol{\mu} = (0, 0)^{\mathsf{T}}$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1.1 \\ 1.1 & 1 \end{pmatrix}.$$

The major axis of the ellipses is a solid line and the minor axis is a dashed line.

How can the axes be found? From Section A.20, we know that $\boldsymbol{\Sigma}$ has an *eigenvalue-eigenvector decomposition*

$$\boldsymbol{\Sigma} = \boldsymbol{O}\,\mathrm{diag}(\lambda_i)\,\boldsymbol{O}^{\mathsf{T}},$$

where $\boldsymbol{O}$ is an orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{\Sigma}$ and $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $\boldsymbol{\Sigma}$.

The columns of $\boldsymbol{O}$ determine the axes of the ellipse $\mathcal{E}(c)$. The decomposition can be found in `R` using the function `eigen` and, for the matrix $\boldsymbol{\Sigma}$ in the example, the decomposition is

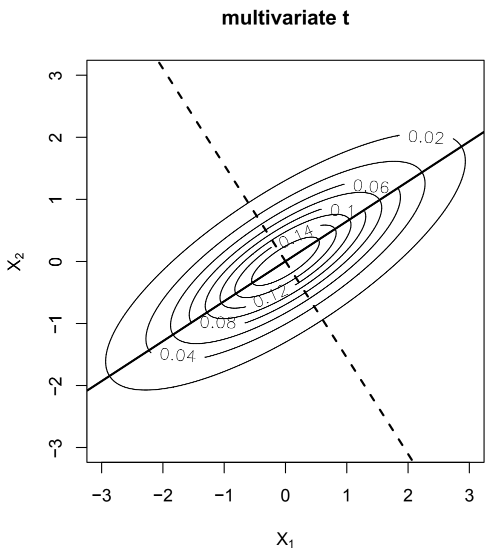

Fig. 7.6. *Contour plot of a multivariate t_4-density with $\mu = (0, 0)^\mathsf{T}$, $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, and $\sigma_{12} = 1.1$.*

```
$values
[1] 2.708 0.292
```

which gives the eigenvalues, and

```
$vectors
       [,1]   [,2]
[1,] -0.841  0.541
[2,] -0.541 -0.841
```

which has the corresponding eigenvectors as columns; e.g., $(-0.841, -0.541)$ is an eigenvector with eigenvalue 2.708. The eigenvectors are only determined up to a sign change, so the first eigenvector could be taken as $(-0.841, -0.541)$, as in the `R` output, or $(0.841, 0.541)$.

If $\boldsymbol{o}_i$ is the ith column of $\boldsymbol{O}$, the ith axis of $\mathcal{E}(c)$ goes through the points $\boldsymbol{\mu}$ and $\boldsymbol{\mu} + \boldsymbol{o}_i$. Therefore, this axis is the line

$$\{\boldsymbol{\mu} + k\,\boldsymbol{o}_i : -\infty < k < \infty\}.$$

Because $\boldsymbol{O}$ is an orthogonal matrix, the axes are mutually perpendicular. The axes can be ordered according to the size of the corresponding eigenvalues. In the bivariate case the axis associated with the largest (smallest) eigenvalue is the major (minor) axis. We are assuming that there are no ties among the eigenvalues.

Since $\boldsymbol{\mu} = 0$, in our example the major axis is $k\,(0.841, 0.541)$, $-\infty < k < \infty$, and the minor axis is $k\,(0.541, -0.841)$, $-\infty < k < \infty$.

When there are ties among the eigenvalues, the eigenvectors are not unique and the analysis is somewhat more complicated and will not be discussed in detail. Instead two examples will be given. In the bivariate case if $\boldsymbol{\Sigma} = \boldsymbol{I}$, the contours are circles and there is no unique choice of the axes—any pair of perpendicular vectors will do. As a trivariate example, if $\boldsymbol{\Sigma} = \mathrm{diag}(1,1,3)$, then the first principle axis is (0,0,1) with eigenvalue 3. The second and third principal axis can be any perpendicular pair of vectors with third coordinates equal to 0. The `eigen` function in `R` returns (0,1,0) and (1,0,0) as the second and third axes.

7.9 The Multivariate Skewed *t*-Distributions

Azzalini and Capitanio (2003) have proposed a skewed extension of the multivariate t-distribution. The univariate special case was discussed in Section 5.7. In the multivariate case, in addition to the shape parameter ν determining tail weight, the skewed t-distribution has a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^{\mathsf{T}}$ of shape parameters determining the amounts of skewness in the components of the distribution. If $\boldsymbol{Y}$ has a skewed t-distribution, then Y_i is left-skewed, symmetric, or right-skewed depending on whether $\alpha_i < 0$, $\alpha_i = 0$, or $\alpha_i > 0$. Figure 7.7 is a contour plot of a bivariate skewed t-distribution with $\boldsymbol{\alpha} = (-1, 0.25)^{\mathsf{T}}$. Notice that, because α_1 is reasonably large and negative, Y_1 has a considerable amount of left skewness, as can be seen in the contours, which are more widely spaced on the left side of the plot compared to the right. Also, Y_2 shows a lesser amount of right skewness, which is to be expected since α_2 is positive with a relatively small absolute value.

Example 7.5. Fitting the skewed t-distribution to `CRSPday`

We now fit the skewed t-model to the `CRSPday` data set using the function `mst.fit` in `R`'s `sn` package. This function maximizes the likelihood over all parameters, so there is no need to use the profile likelihood as with `cov.trob`. The estimates are as follows.

```
$dp$beta
           [,1]      [,2]      [,3]      [,4]
[1,] -0.0001474 -0.001186 3.667e-05 0.0002218

$dp$Omega
          [,1]      [,2]      [,3]      [,4]
[1,] 1.242e-04 4.751e-05 3.328e-05 4.522e-05
```

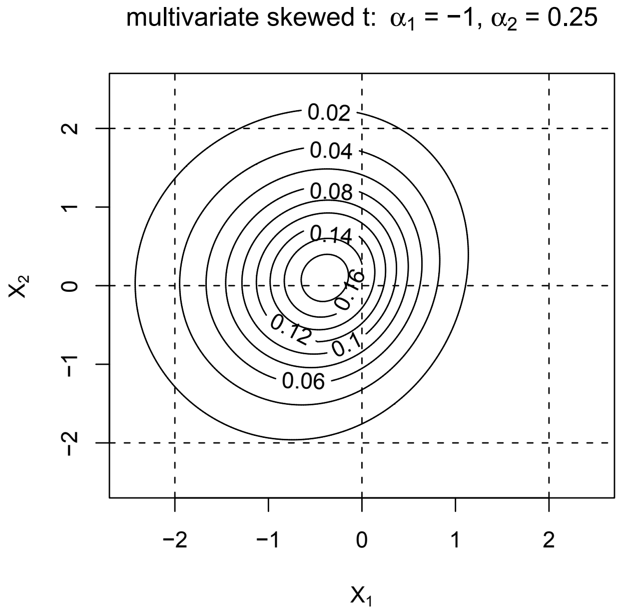

Fig. 7.7. *Contours of a bivariate skewed t-density. The contours are more widely spaced on the left compared to the right because X_1 is left-skewed. Similarly, the contours are more widely spaced on the top compared to the bottom because X_2 is left-skewed, but the skewness of X_2 is relatively small and less easy to see.*

```
[2,] 4.751e-05 1.822e-04 2.255e-05 3.822e-05
[3,] 3.328e-05 2.255e-05 1.145e-04 2.738e-05
[4,] 4.522e-05 3.822e-05 2.738e-05 3.627e-05

$dp$alpha
[1]  0.07929  0.12075  0.03998 -0.01585

$dp$df
[1] 5.8
```

Here `dp$beta` is the estimate of $\boldsymbol{\mu}$, `dp$Omega` is the estimate of $\boldsymbol{\Sigma}$, `dp$alpha` is the estimate of $\boldsymbol{\alpha}$, and `dp$df` is the estimate of ν. Note that the estimates of all components of $\boldsymbol{\alpha}$ are close to zero, which suggests that there is little if any skewness in the data.

AIC for the skewed t-model is 9.06 plus 64,000, somewhat larger than 7.45, the AIC for the symmetric t-model. This result, and the small estimated values of the α_i shape parameters, suggest that the symmetric t-model is adequate for this data set.

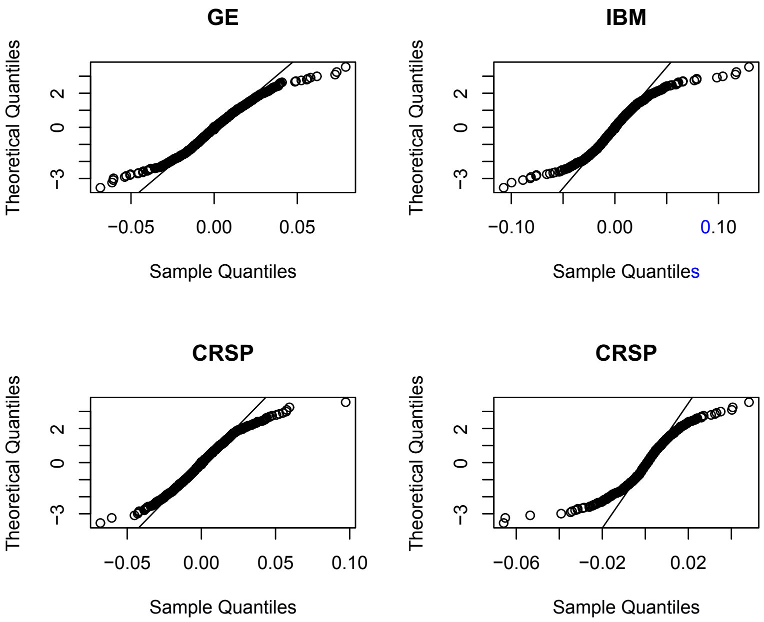

Fig. 7.8. *Normal plots of the four returns series in the* `CRSPday` *data set. The reference lines go through the first and third quartiles.*

In summary, the `CRSPday` data are well fit by a symmetric t-distribution and no need was found for using a skewed t-distribution. Also, normal plots in Figure 7.8 of the four variables show no signs of serious skewness. Although this might be viewed as a negative result, since we have not found an improvement in fit by going to the more flexible skewed t-distribution, the result does give us more confidence that the symmetric t-distribution is suitable for modeling this data set.

□

7.10 The Fisher Information Matrix

In the discussion of Fisher information in Section 5.10, θ was assumed to be one-dimensional. If $\boldsymbol{\theta}$ is an m-dimensional parameter vector, then the Fisher information is an $m \times m$ square matrix, $\mathcal{I}$, and is equal to minus the matrix of expected second-order partial derivatives of $\log\{L(\boldsymbol{\theta})\}$.[1] In other words, the i, jth entry of the Fisher information matrix is

[1] The matrix of second partial derivatives of a function is called its *Hessian matrix*, so the Fisher information matrix is the expectation of minus the Hessian of the log-likelihood.

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = -E\left[\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\log\{L(\boldsymbol{\theta})\}\right]. \tag{7.19}$$

The standard errors are the square roots of the diagonal entries of the inverse of the Fisher information matrix. Thus, the standard error for θ_i is

$$s_{\widehat{\theta}_i} = \sqrt{\{\mathcal{I}(\widehat{\boldsymbol{\theta}})^{-1}\}_{ii}}. \tag{7.20}$$

In the case of a single parameter, (7.20) reduces to (5.19). The central limit theorem for the MLE in Section 5.10 generalizes to the following multivariate version.

Theorem 7.6. *Under suitable assumptions, for large enough sample sizes, the maximum likelihood estimator is approximately normally distributed with mean equal to the true parameter vector and with covariance matrix equal to the inverse of the Fisher information matrix.*

The key point is that there is an explicit method of calculating standard errors for maximum likelihood estimators. The calculation of standard errors of maximum likelihood estimators by computing and then inverting the Fisher information matrix is routinely programmed into statistical software.

Computation of the expectation in $\mathcal{I}(\boldsymbol{\theta})$ can be challenging. Programming the second derivatives can be difficult as well, especially for complex models. In practice, the observed Fisher information matrix, whose i,jth element is

$$\mathcal{I}_{ij}^{\text{obs}}(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\log\{L(\boldsymbol{\theta})\} \tag{7.21}$$

is often used. The observed Fisher information matrix is, of course, the multivariate analog of (5.21). Using observed information obviates the need to compute the expectation. Moreover, the Hessian matrix can be computed numerically by finite differences, for example, using `R`'s `fdHess` function in the `nlme` package.

Inverting the observed Fisher information computed by finite differences is the most commonly used method for obtaining standard errors. The advantage of this approach is that only the computation of the likelihood, or log-likelihood, is necessary, and of course this is necessary simply to compute the MLE.

7.11 Bootstrapping Multivariate Data

When resampling multivariate data, the dependencies within the observation vectors need to be preserved. Let the vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ be an i.i.d. sample of multivariate data. In model-free resampling, the vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are sampled with replacement. There is no resampling of the components within a vector. Resampling within vectors would make their components mutually independent and would not mimic the actual data where the components are dependent. Stated differently, if the data are in a spreadsheet (or matrix) with rows corresponding to observations and columns to variables, then one samples entire rows.

Model-based resampling simulates vectors from the multivariate distribution of the $\boldsymbol{Y}_i$, for example, from a multivariate t-distribution with the mean vector, covariance matrix, and degrees of freedom equal to the MLEs.

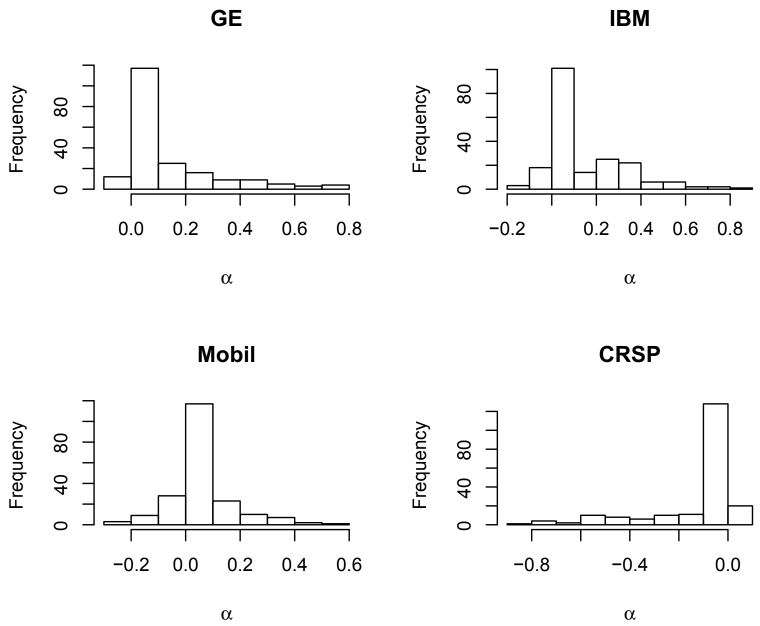

Fig. 7.9. *Histograms of* 200 *bootstrapped values of* $\widehat{\alpha}$ *for each of the returns series in the* `CRSPday` *data set.*

Example 7.7. Bootstrapping the skewed t fit to `CRSPday`

In Example 7.5 the skewed t-model was fit to the `CRSPday` data. This example continues that analysis by bootstrapping the estimator of α for each of the four returns series. Histograms of 200 bootstrap values of $\widehat{\alpha}$ are found in Figure 7.9. Bootstrap percentile 95% confidence intervals include 0 for all four stocks, so there is no strong evidence of skewness in any of the returns series.

Despite the large sample size of 2528, the estimators of α do not appear to be normally distributed. We can see in Figure 7.9 that they are right-skewed for the three stocks and left-skewed for the CRSP returns. The distribution of $\widehat{\alpha}$ also appears heavy-tailed. The excess kurtosis coefficient of the 200 bootstrap values of $\widehat{\alpha}$ is 2.38, 1.33, 3.18, and 2.38 for the four series.

The central limit theorem for the MLE guarantees that $\widehat{\alpha}$ is nearly normally distributed for sufficiently large samples, but it does not tell us how large the sample size must be. We see in this example that in such cases the sample size must be very large indeed since 2528 is not large enough. This is a major reason for preferring to construct confidence intervals using the bootstrap rather than a normal approximation.

A bootstrap sample of the returns was drawn with the following R code. The returns are in the matrix `dat` and `yboot` is a bootstrap sample chosen by taking a random sample of the rows of `dat`, with replacement of course.

```
yboot = dat[sample((1:n),n,replace =T),]
```

□

7.12 Bibliographic Notes

The multivariate central limit theorem for the MLE is stated precisely and proved in textbooks on asymptotic theory such as Lehmann (1999) and van der Vaart (1998). The multivariate skewed t-distribution is in Azzalini and Capitanio (2003).

7.13 References

Azzalini, A., and Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistics Society, Series B*, **65**, 367-389.

Lehmann, E. L. (1999) *Elements of Large-Sample Theory*, Springer-Verlag, New York.

van der Vaart, A. W. (1998) *Asymptotic Statistics*, Cambridge University Press, Cambridge.

7.14 R Lab

7.14.1 Equity Returns

This section uses the data set `berndtInvest` in R's `fEcofin` package. This data set contains monthly returns from January 1, 1987, to December 1, 1987, on

16 equities. There are 18 columns. The first column is the date and the last is the risk-free rate.

In the lab we will only use the first four equities. The following code computes the sample covariance and correlation matrices for these returns.

```
library("fEcofin")
Berndt = as.matrix(berndtInvest[,2:5])
cov(Berndt)
cor(Berndt)
```

If you wish, you can also plot a scatterplot matrix with the following `R` code.

```
pairs(Berndt)
```

Problem 1 *Suppose the four variables being used are denoted by $X_1, \ldots, X_4$. Use the sample covariance matrix to estimate the variance of $0.5X_1 + 0.3X_2 + 0.2X_3$. Include with your work the R code used to estimate this covariance. (Useful* `R` *facts: "*`t(a)`*" is the transpose of a vector or matrix* `a` *and "*`a %*% b`*" is the matrix product of* `a` *and* `b`*.)*

Fit a multivariate-t model to the data using the function `cov.trob` in the `MASS` package. This function computes the MLE of the mean and covariance matrix with a fixed value of ν. To find the MLE of ν, the following code computes the profile log-likelihood for ν.

```
library(MASS)  # needed for cov.trob
library(mnormt)  # needed for dmt
df = seq(2.5,8,.01)
n = length(df)
loglik_max = rep(0,n)
for(i in 1:n)
{
fit = cov.trob(Berndt,nu=df[i])
mu = as.vector(fit$center)
sigma =matrix(fit$cov,nrow=4)
loglik_max[i] = sum(log(dmt(Berndt,mean=fit$center,
     S=fit$cov,df=df[i])))
}
```

Problem 2 *Using the results produced by the code above, find the MLE of ν and a 90% profile likelihood confidence interval for ν. Include your* `R` *code with your work. Also, plot the profile log-likelihood and indicate the MLE and the confidence interval on the plot. Include the plot with your work.*

Section 7.14.3 demonstrates how the MLE for a multivariate t-model can be fit directly with the `optim` function, rather than be profile likelihood.

7.14.2 Simulating Multivariate t-Distributions

The following code generates and plots three bivariate samples. Each sample has univariate marginals that are standard t_3-distributions. However, the dependencies are different.

```
library(MASS)  # need for mvrnorm
par(mfrow=c(1,4))
N =  2500
nu = 3

set.seed(5640)
cov=matrix(c(1,.8,.8,1),nrow=2)
x= mvrnorm(N, mu = c(0,0), Sigma=cov)
w = sqrt(nu/rchisq(N, df=nu))
x = x * cbind(w,w)
plot(x,main="(a)")

set.seed(5640)
cov=matrix(c(1,.8,.8,1),nrow=2)
x= mvrnorm(N, mu = c(0,0), Sigma=cov)
w1 = sqrt(nu/rchisq(N, df=nu))
w2 = sqrt(nu/rchisq(N, df=nu))
x = x * cbind(w1,w2)
plot(x,main="(b)")

set.seed(5640)
cov=matrix(c(1,0,0,1),nrow=2)
x= mvrnorm(N, mu = c(0,0), Sigma=cov)
w1 = sqrt(nu/rchisq(N, df=nu))
w2 = sqrt(nu/rchisq(N, df=nu))
x = x * cbind(w1,w2)
plot(x,main="(c)")

set.seed(5640)
cov=matrix(c(1,0,0,1),nrow=2)
x= mvrnorm(N, mu = c(0,0), Sigma=cov)
w = sqrt(nu/rchisq(N, df=nu))
x = x * cbind(w,w)
plot(x,main="(d)")
```

Note the use of these `R` commands: `set.seed` to set the seed of the random number generator, `mvrnorm` to generate multivariate normally distributed vectors, `rchisq` to generate χ^2-distributed random numbers, `cbind` to bind together vectors as the columns of a matrix, and `matrix` to create a matrix from a vector. In `R`, "`a*b`" is elementwise multiplication of same-size matrices `a` and `b`, and "`a%*%b`" is matrix multiplication of conforming matrices `a` and `b`.

Problem 3 *Which sample has independent variates? Explain your answer.*

Problem 4 *Which sample has variates that are correlated but do not have tail dependence? Explain your answer.*

Problem 5 *Which sample has variates that are uncorrelated but with tail dependence? Explain your answer.*

Problem 6 *Suppose that (X, Y) are the returns on two assets and have a multivariate t-distribution with degrees of freedom, mean vector, and covariance matrix*

$$\nu = 5, \quad \mu = \begin{pmatrix} 0.001 \\ 0.002 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.10 & 0.03 \\ 0.03 & 0.15 \end{pmatrix}.$$

Then $R = (X + Y)/2$ is the return on an equally weighted portfolio of the two assets.

(a) *What is the distribution of R?*
(b) *Write an* `R` *program to generate a random sample of size 10,000 from the distribution of* `R`*. Your program should also compute the 0.01 upper quantile of this sample and the sample average of all returns that exceed this quantile. This quantile and average will be useful later when we study risk analysis.*

7.14.3 Fitting a Bivariate *t*-Distribution

When you run the `R` code that follows this paragraph, you will compute the MLE for a bivariate t-distribution fit to CRSP returns data. A challenge when fitting a multivariate distribution is enforcing the constraint that the scale matrix (or the covariance matrix) must be positive definite. One way to meet this challenge is to let the scale matrix be $A^T A$, where A is an upper triangular matrix. (It is easy to show that $A^T A$ is positive semidefinite if A is any square matrix. Because a scale or covariance matrix is symmetric, only the entries on and above the main diagonal are free parameters. In order for A to have the same number of free parameters as the covariance matrix, we restrict A to be upper triangular.)

```
library(mnormt)
data(CRSPday,package="Ecdat")
Y = CRSPday[,c(5,7)]
loglik = function(par)
{
mu = par[1:2]
```

```
A = matrix(c(par[3],par[4],0,par[5]),nrow=2,byrow=T)
scale_matrix = t(A)%*%A
df = par[6]
f = -sum(log(dmt(Y, mean=mu,S=scale_matrix,df=df)))
f
}
A=chol(cov(Y))
start=as.vector(c(apply(Y,2,mean),A[1,1],A[1,2],A[2,2],4))
fit_mvt = optim(start,loglik,method="L-BFGS-B",lower=c(-.02,-.02,
   -.1,-.1,-.1,2),
   upper=c(.02,.02,.1,.1,.1,15),hessian=T)
```

Problem 7 *Let $\boldsymbol{\theta} = (\mu_1, \mu_2, A_{1,1}, A_{1,2}, A_{2,2}, \nu)$, where μ_j is the mean of the jth variable, $A_{1,1}$, $A_{1,2}$, and $A_{2,2}$ are the nonzero elements of A, and ν is the degrees-of-freedom parameter.*

(a) What does the code `A=chol(cov(Y))` *do?*
(b) Find $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$, the MLE of $\boldsymbol{\theta}$.
(c) Find the Fisher information matrix for $\boldsymbol{\theta}$. (Hint: The Hessian is part of the object `fit_mvt`*. Also, the* `R` *function* `solve` *will invert a matrix.)*
(d) Find the standard errors of the components of $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ using the Fisher information matrix.
(e) Find the MLE of the covariance matrix of the returns.
(f) Find the MLE of ρ, the correlation between the two returns (Y_1 and Y_2).

7.15 Exercises

1. Suppose that $E(X) = 1$, $E(Y) = 1.5$, $\mathrm{Var}(X) = 2$, $\mathrm{Var}(Y) = 2.7$, and $\mathrm{Cov}(X, Y) = 0.8$.
   (a) What are $E(0.2X + 0.8Y)$ and $\mathrm{Var}(0.2X + 0.8Y)$?
   (b) For what value of w is $\mathrm{Var}\{wX + (1-w)Y\}$ minimized? Suppose that X is the return on one asset and Y is the return on a second asset. Why would it be useful to minimize $\mathrm{Var}\{wX + (1-w)Y\}$?
2. Let X_1, X_2, Y_1, and Y_2 be random variables.
   (a) Show that $\mathrm{Cov}(X_1 + X_2, Y_1 + Y_2) = \mathrm{Cov}(X_1, Y_1) + \mathrm{Cov}(X_1, Y_2) + \mathrm{Cov}(X_2, Y_1) + \mathrm{Cov}(X_2, Y_2)$.
   (b) Generalize part (a) to an arbitrary number of X_is and Y_is.
3. Verify formulas (A.24)–(A.27).
4. (a) Show that
$$E\{X - E(X)\} = 0$$
   for any random variable X.
   (b) Use the result in part (a) and equation (A.31) to show that if two random variables are independent then they are uncorrelated.
5. Show that if X is uniformly distributed on $[-a, a]$ for any $a > 0$ and if $Y = X^2$, then X and Y are uncorrelated but they are not independent.

6. Verify the following results that were stated in Section 7.3:

$$E(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{X}) = \boldsymbol{w}^{\mathsf{T}}\{E(\boldsymbol{X})\}$$

and

$$\begin{aligned}\mathrm{Var}(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{X}) &= \sum_{i=1}^{N}\sum_{j=1}^{N} w_i\, w_j\, \mathrm{Cov}(X_i, X_j) \\ &= \mathrm{Var}(\boldsymbol{w}^{\mathsf{T}}\boldsymbol{X})\boldsymbol{w}^{\mathsf{T}}\mathrm{COV}(\boldsymbol{X})\boldsymbol{w}.\end{aligned}$$