<u>**Sample Chapter 3**</u>

Value-at-Risk

Theory and Practice

Glyn A. Holton

This is Chapter 3 of Glyn Holton's book **Value-at-risk: Theory and Practice**. For more information, visit www.value-at-risk.net.

Chapter 3

Probability

3.1. MOTIVATION

VaR measures are inherently probabilistic. A central question that VaR addresses is this: If a portfolio comprises holdings in various instruments, how is its market risk determined by theirs? In the parlance of probability, the question becomes: If a random variable is defined as a function of other random variables, how is its distribution determined by theirs?

As it relates to market risk, the question is addressed in Chapter 10, which discusses transformation procedures. That chapter draws on various techniques from this chapter, including:

- techniques for characterizing the distribution of a linear polynomial of a random vector;
- techniques for characterizing the distribution of a quadratic polynomial of a random vector;
- the central limit theorem;
- the inversion theorem.

Concepts from the present chapter underlie statistics and time series analysis, which are the topics of Chapter 4. They, in turn, are used in Chapter 7 to design inference procedures.

Finally, the present chapter describes principal component analysis, which is used in Chapter 9 with a category of portfolio remappings called, transparently, "principal component remappings."

3.2. PREREQUISITES

We assume familiarity with basic notation and concepts from probability. If E is an event, we denote its probability $Pr(E)$. You should be familiar with **random variables** and **random vectors**. A random vector $\boldsymbol{X}$ can be thought of as an n-dimensional vector of random variables X_i all defined on the same sample space. When we present general definitions or results for random vectors, these also apply to random variables.

It is important to distinguish between a random vector $\boldsymbol{X}$ and a realization of that random vector, which we may denote $\boldsymbol{x}$. The **realization** is an element of the range of the random vector.

You should be familiar with **discrete** and **continuous** distributions for random vectors. You should be comfortable working with **probability functions** (PFs), **probability density functions** (PDFs), and **cumulative distribution functions** (CDFs). You should be familiar with **joint distributions**, **conditional distributions**, and **marginal distributions**.

We may think of random vectors as being "equivalent" in several senses. We distinguish between two of these. Random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are **equal**, denoted $\boldsymbol{X} = \boldsymbol{Y}$, if they both take on the same value with probability 1. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ simply have the same probability distribution, we denote this relationship $\boldsymbol{X} \sim \boldsymbol{Y}$. We also use the symbol $\sim$ to indicate what a random variable represents, say: $X \sim$ tomorrow's 3-month USD Libor rate.

You should know what it means for two or more components of a random vector $\boldsymbol{X}$ to be **independent**. In particular, if n components X_i are independent, their joint CDF and marginal CDFs satisfy:

$$\Phi(x_1, x_2, \ldots, x_n) = \Phi_1(x_1)\Phi_2(x_2)\cdots\Phi_n(x_n). \tag{3.1}$$

for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$. Similarly, their joint PDF and marginal PDFs satisfy:

$$\phi(x_1, x_2, \ldots, x_n) = \phi_1(x_1)\phi_2(x_2)\cdots\phi_n(x_n). \tag{3.2}$$

for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$.[1]

3.3. PARAMETERS

Parameters describe random vectors much as we might use height or age to describe a person. Formally, a **parameter** is a function that is applied to a random vector's probability distribution. It may take on real, vector, or matrix values. A standard deviation, mean vector, or covariance matrix are all examples

[1]For technical reasons, we should qualify [3.2] and say that it may fail to hold on a set of values for $\boldsymbol{X}$ of probability 0.

of parameters. In this section, we describe parameters for random variables. In Section 3.4, we extend the discussion to parameters for random vectors.

Expectation

Let X be a random variable. We denote the **expected value**, **expectation**, or **mean** of X as either μ or $E(X)$. If X is discrete, we define its expectation as

$$E(X) = \sum_x x\phi(x), \qquad [3.3]$$

where ϕ is the PF of X. If X is continuous, we replace the summation with an integral and define

$$E(X) = \int_{-\infty}^{\infty} x\phi(x)dx, \qquad [3.4]$$

where ϕ is the PDF of X.

Expectation is used to define a number of other parameters, but first we must discuss expectations of functions of random variables.

Expectation of a Function of a Random Variable

Suppose X is a random variable and f is a function from $\mathbb{R}$ to $\mathbb{R}$. Then $f(X)$ is a new random variable[2] whose probability distribution we can, at least in theory, infer from that of X. We do not need the probability distribution of $f(X)$ in order to determine the expectation $E[f(X)]$. This can be obtained directly from the probability distribution of X using the formula

$$E[f(X)] = \sum_x f(x)\phi(x) \qquad [3.5]$$

or

$$E[f(X)] = \int_{-\infty}^{\infty} f(x)\phi(x)dx, \qquad [3.6]$$

depending upon whether X is discrete or continuous. In a sense, [3.5] and [3.6] are generalizations of [3.3] and [3.4].

[2]Technically, f must be measurable for $f(X)$ to be a random variable.

Variance and Standard Deviation

Variance is a parameter that measures how dispersed a random variable's probability distribution is. In Exhibit 3.1, two PDFs have a mean of 0. The one on the left is more dispersed than the one on the right. It has a higher variance.

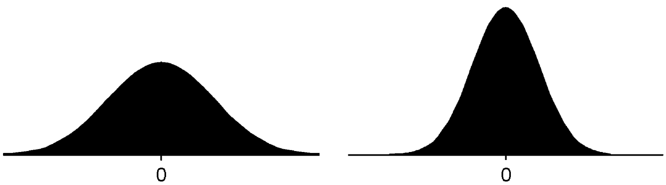

Exhibit 3.1 These graphs illustrate the notion of variance. Both PDFs have an expectation of 0, but the one on the left is more dispersed than the one on the right. It has a higher variance.

The **variance**, denoted σ^2 or $var(X)$, of a random variable X is defined as an expectation of a function of X:

$$var(X) = E[(X - \mu)^2]. \tag{3.7}$$

Standard deviation, denoted σ or $std(X)$, is the positive square-root of variance.

Skewness

Skew or **skewness** is a measure of asymmetry in a random variable's probability distribution. Both PDFs in Exhibit 3.2 have the same mean and standard deviation. The one on the left is positively skewed. The one on the right is negatively skewed.

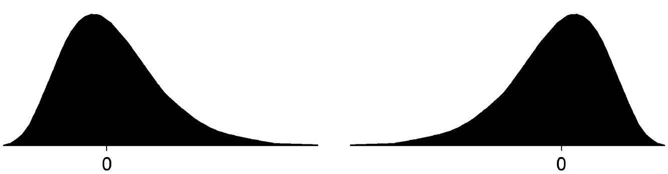

Exhibit 3.2 These graphs illustrate the notion of skewness. Both PDFs have the same mean and variance. The one on the left is positively skewed. The one on the right is negatively skewed.

The skewness of a random variable X is denoted η_1 or $skew(X)$. It is defined as

$$skew(X) = \frac{E[(X - \mu)^3]}{\sigma^3}. \tag{3.8}$$

KURTOSIS

Kurtosis is another parameter that describes the shape of a random variable's probability distribution. Consider the two PDFs in Exhibit 3.3. Both have a mean and skewness of 0. Which would you say has the greater standard deviation? It is impossible to say. The distribution on the right is more peaked at the center, which might lead us to believe that it has a lower standard deviation. It has fatter tails, which might lead us to believe that it has a greater standard deviation. If the effect of the peakedness exactly offsets that of the fat tails, the two distributions may have the same standard deviation. The different shapes of the two distributions illustrates kurtosis. The distribution on the right has a greater kurtosis than the distribution on the left.

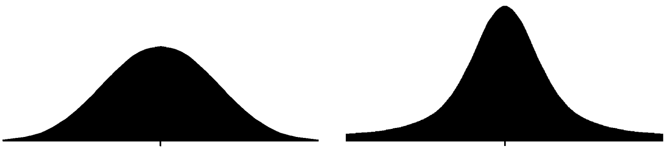

Exhibit 3.3 These graphs illustrate the notion of kurtosis. The PDF on the right has higher kurtosis than the PDF on the left. It is more peaked at the center, and it has fatter tails.

The **kurtosis** of a random variable X is denoted η_2 or $kurt(X)$.[3] It is defined as

$$kurt(X) = \frac{E[(X - \mu)^4]}{\sigma^4}. \qquad [3.9]$$

If a distribution's kurtosis is greater than 3, it is said to be **leptokurtic**. If its kurtosis is less than 3, it is said to be **platykurtic**. Leptokurtosis is associated with distributions that are simultaneously "peaked" and have "fat tails." Platykurtosis is associated with distributions that are simultaneously less peaked and have thinner tails. In Exhibit 3.3, the distribution on the left is platykurtic. The one on the right is leptokurtic.

QUANTILES

Consider a random variable X with CDF Φ. A q-quantile of X is any value x such that $Pr(X \leq x) = q$. A q-quantile need not exist. If it does exist, it need not be unique.[4] In most VaR applications, all q-quantiles exist and are unique for

[3] The use of subscripts in the notation η_1 and η_2 for skewness and kurtosis is unfortunate because it can lead to confusion if subscripts are also employed to distinguish between different random variables. We use the notation because it is well established.

[4] We could force uniqueness by defining the q-quantile as the supremum of all values satisfying the definition provided in the text.

$q \in (0, 1)$. In such cases, a q-quantile is a parameter and equals the inverse CDF evaluated at q. For this reason, we denote a q-quantile as $\Phi^{-1}(q)$.

Moments

For any positive integer k, the k^{th} **moment** of a random variable X is defined as

$$\mu'_k = E(X^k). \qquad [3.10]$$

Its k^{th} **central moment** is defined as

$$\mu_k = E[(X - \mu)^k], \qquad [3.11]$$

where $\mu = E(X)$. Based upon our earlier definitions, the expectation and variance of a random variable are its first moment and second central moment. Its skewness and kurtosis are scaled third and fourth central moments.

For any $n > 0$, a random variable's first n moments convey the same information as its first n central moments—each can be derived from the other. See Exercise 3.15.

We say a random variable X is **bounded** if there exists a number a such that $Pr(|X| > a) = 0$. If a random variable is bounded, all its moments exist. If it is unbounded, specific moments may or may not exist. However, if the k^{th} moment of X exists, then all moments of order less than k must also exist.

Exercises

3.1 PDFs for two continuous random variables are illustrated in Exhibit 3.4. Assume probability density is 0 for both distributions outside the graphed regions. Where possible, indicate which random variable has the greater:
   a. expectation,
   b. standard deviation,
   c. skewness,
   d. kurtosis, and
   e. .25-quantile.

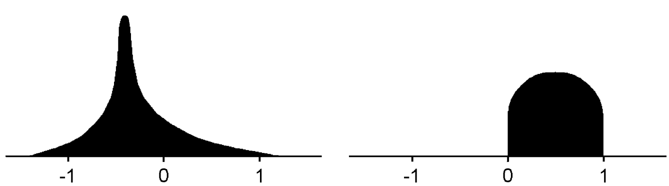

Exhibit 3.4 PDFs for two continuous random variables.

3.2 Consider a discrete random variable Y, which represents the number of "heads" that will be obtained in three flips of a fair coin. It has PF

$$\phi(y) = \begin{cases} 0.125 & y = 0 \text{ or } 3 \\ 0.375 & y = 1 \text{ or } 2. \\ 0 & \text{otherwise} \end{cases} \quad [3.12]$$

a. Calculate the mean of Y.
b. Calculate the variance of Y.
c. Calculate the standard deviation of Y.
d. Calculate the skewness of Y.
e. Calculate the kurtosis of Y.
f. Calculate a .10 quantile of Y.
g. Calculate a .875 quantile of Y.

3.3 Consider a continuous random variable Z with PDF

$$\phi(z) = \begin{cases} 0.5 & 1 < z < 3 \\ 0 & \text{otherwise} \end{cases}. \quad [3.13]$$

a. Calculate the mean of Z.
b. Calculate the variance of Z.
c. Calculate the standard deviation of Z.
d. Calculate the skewness of Z.
e. Calculate the kurtosis of Z.
f. Calculate a .10-quantile of Z.
g. Calculate a .875-quantile of Z.

3.4 Consider the random variable $W = Z^2$, where Z is defined as in the previous exercise.
a. Calculate the mean of W.
b. Calculate the variance of W.
c. Calculate the standard deviation of W.

3.5 True or false: If a continuous random variable X has a symmetric distribution, $\phi(x - \mu) = \phi(-x - \mu)$, it must have 0 skewness.

3.6 In general, for any random variable X and any constant b, $E(bX) = bE(X)$. Prove this result for the case X is discrete.

3.7 In general, for any random variable X and any constant a, $E(X + a) = E(X) + a$. Prove this result for the case X is continuous.

3.8 In general, for any random variable X and any constant b, $std(bX) = |b|std(X)$, where $|b|$ indicates the absolute value of b. Prove this result for the case X is discrete. Use your result from Exercise 3.6.

3.9 In general, for any random variable X and any constant a, $std(X + a) = std(X)$. Prove this result for the case X is continuous. Use your result from Exercise 3.7.

3.4. PARAMETERS OF RANDOM VECTORS

The **expectation** of an n-dimensional random vector $\boldsymbol{X}$ is a vector which we denote either $\boldsymbol{\mu}$ or $E(\boldsymbol{X})$. Its components are the expectations of the marginal distributions of the X_i:

$$E(\boldsymbol{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}. \qquad [3.14]$$

Expectation of a Function of a Random Vector

Let $\boldsymbol{X}$ be an n-dimensional random vector and f a function from $\mathbb{R}^n$ to $\mathbb{R}$ that defines a random variable $f(\boldsymbol{X})$. We may generalize [3.5] and [3.6] to calculate the mean of $f(\boldsymbol{X})$. If $\boldsymbol{X}$ is discrete with PF ϕ, we have

$$E[f(\boldsymbol{X})] = \sum_{\boldsymbol{x}} f(\boldsymbol{x})\phi(\boldsymbol{x}). \qquad [3.15]$$

If $\boldsymbol{X}$ is continuous with PDF ϕ, this becomes

$$E[f(\boldsymbol{X})] = \int_{\mathbb{R}^n} f(\boldsymbol{x})\phi(\boldsymbol{x})d\boldsymbol{x} \qquad [3.16]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\phi(x_1, x_2, \ldots, x_n)dx_1dx_2\cdots dx_n. \qquad [3.17]$$

Joint Moments

Joint moments generalize moments. Let X_i and X_j be components of a random vector. We define their (k, l) **joint moment** as

$$E\left[X_i^k X_j^l\right]. \qquad [3.18]$$

We define their (k, l) **joint central moment** as

$$E[(X_i - \mu_i)^k (X_j - \mu_j)^l]. \qquad [3.19]$$

We define the $\boldsymbol{n}^{\text{th}}$ **moments** of a random vector $\boldsymbol{X}$ as all its joint moments $E[X_i^k X_j^l]$ for which $k + l = n$. We define its $\boldsymbol{n}^{\text{th}}$**central moments** as all its joint central moments $E[(X_i - \mu_i)^k (X_j - \mu_j)^l]$ for which $k + l = n$.

COVARIANCE

We are primarily interested in the (1, 1) joint central moment, which we call **covariance**. For the i^{th} and j^{th} components of $\boldsymbol{X}$, we denote covariance $\Sigma_{i,j}$ or $cov(X_i, X_j)$. By definition, covariance is symmetric, with $\Sigma_{i,j} = \Sigma_{j,i}$. Also, the covariance of any component X_i with itself is that component's variance:

$$\Sigma_{i,i} = E[(X_i - \mu_i)(X_i - \mu_i)] = E[(X_i - \mu_i)^2] = \sigma_i^2. \qquad [3.20]$$

We summarize all the covariances of a random vector $\boldsymbol{X}$ with a **covariance matrix**:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \Sigma_{1,3} & \cdots & \Sigma_{1,n} \\ \Sigma_{2,1} & \Sigma_{2,2} & \Sigma_{2,3} & & \\ \Sigma_{3,1} & \Sigma_{3,2} & \Sigma_{3,3} & & \vdots \\ \vdots & & & \ddots & \\ \Sigma_{n,1} & & \cdots & & \Sigma_{n,n} \end{pmatrix}. \qquad [3.21]$$

By the symmetry property of covariances, the covariance matrix is symmetric. Intuitively, covariance is a metric of the tendency of two components of a random vector to vary together, or co-vary. The magnitude of a covariance depends upon the standard deviations of the two components. To obtain a more direct metric of how two components co-vary, we scale covariance to obtain **correlation**.

CORRELATION

The **correlation**, $\rho_{i,j}$ or $cor(X_i, X_j)$, of the i^{th} and j^{th} components of a random vector $\boldsymbol{X}$ is defined as

$$\rho_{i,j} = \frac{cov(X_i, X_j)}{\sigma_i \sigma_j}. \qquad [3.22]$$

By construction, a correlation is always a number between -1 and 1. Correlation inherits the symmetry property of covariance: $\rho_{i,j} = \rho_{j,i}$. From [3.20] and [3.22], $\rho_{i,i} = 1$, which indicates that a random variable co-varies perfectly with itself. If X_i and X_j are independent, their correlation is 0. The converse is not true. As with covariances, we can summarize all the correlations of a random vector $\boldsymbol{X}$ with a symmetric **correlation matrix**:

$$\boldsymbol{\rho} = \begin{pmatrix} \rho_{1,1} & \rho_{1,2} & \rho_{1,3} & \cdots & \rho_{1,n} \\ \rho_{2,1} & \rho_{2,2} & \rho_{2,3} & & \\ \rho_{3,1} & \rho_{3,2} & \rho_{3,3} & & \vdots \\ \vdots & & & \ddots & \\ \rho_{n,1} & & \cdots & & \rho_{n,n} \end{pmatrix}. \qquad [3.23]$$

Exercises

3.10 Use [3.17] to prove that, if the components X_1 and X_2 of a two-dimensional random vector $\boldsymbol{X}$ are independent, then

$$E(X_1X_2) = E(X_1)E(X_2). \tag{3.24}$$

3.11 Consider the two-dimensional discrete random vector $\boldsymbol{Q}$ with PF

$$\phi(\boldsymbol{q}) = \begin{cases} 0.3 & \boldsymbol{q} = (1, 0) \\ 0.1 & \boldsymbol{q} = (1, 3) \\ 0.2 & \boldsymbol{q} = (2, 1) \\ 0.1 & \boldsymbol{q} = (0, 3) \\ 0.3 & \boldsymbol{q} = (3, 2) \end{cases}. \tag{3.25}$$

Calculate $\rho_{1,2}$.

3.12 Give an example of a two-dimensional random vector whose components have 0 covariance but are not independent.

3.5. LINEAR POLYNOMIALS OF RANDOM VECTORS

We now consider formula [1.10], which we used in some of the examples of Chapter 1. Let $\boldsymbol{X}$ be a random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Define random variable Y as a linear polynomial

$$Y = \boldsymbol{bX} + a \tag{3.26}$$

of $\boldsymbol{X}$, where $\boldsymbol{b}$ is an n-dimensional row vector and $a \in \mathbb{R}$. The mean and variance of Y are given by

$$E(Y) = \boldsymbol{b\mu} + a, \tag{3.27}$$

$$var(Y) = \boldsymbol{b\Sigma b}'. \tag{3.28}$$

Formulas [3.27] and [3.28] are general. They require no additional assumptions about $\boldsymbol{X}$ whatsoever.

Vector Linear Polynomials

Formulas [3.27] and [3.28] generalize for vector-valued polynomials. Let $\boldsymbol{Y}$ be an m-dimensional random vector defined as a linear polynomial

$$\boldsymbol{Y} = \boldsymbol{bX} + \boldsymbol{a} \tag{3.29}$$

of an n-dimensional random vector $\boldsymbol{X}$. Here, $\boldsymbol{b}$ is an $m \times n$ matrix and $\boldsymbol{a}$ is an m-dimensional vector. If $\boldsymbol{X}$ has mean vector $\boldsymbol{\mu}_X$ and covariance matrix $\boldsymbol{\Sigma}_X$, then $\boldsymbol{Y}$

has mean vector and covariance matrix

$$\boldsymbol{\mu}_Y = \boldsymbol{b}\boldsymbol{\mu}_X + \boldsymbol{a}, \qquad [3.30]$$

$$\boldsymbol{\Sigma}_Y = \boldsymbol{b}\boldsymbol{\Sigma}_X\boldsymbol{b}'. \qquad [3.31]$$

Exercises

3.13 Suppose $\boldsymbol{X}$ is a three-dimensional random vector with the parameters shown in Exhibit 3.5. Let $Y = 10 + X_1 + 3X_2 - 2X_3$. Calculate the mean and standard deviation of Y using [3.27] and [3.28].

Component	Mean	Standard Deviation	Correlations		
			X_1	X_2	X_3
X_1	–4	1.1	1.0		
X_2	0	0.7	0.3	1.0	
X_3	5	0.4	0.1	–0.2	1.0

Exhibit 3.5 Assumptions for Exercise 3.13.

3.14 Suppose a random variable Z is equal to the sum of two other random variables A and B which are related by the functional relationship $B = A^2 - 2A - 4$. Both A and B have a standard deviation of 3. Their correlation is 0.25. What is the standard deviation of Z?

3.15 Use [3.27] to prove that, in general:

$$var(X) = E(X^2) - E(X)^2. \qquad [3.32]$$

3.16 Consider a three-dimensional random vector $\boldsymbol{X}$. Its first two components, X_1 and X_2, are uncorrelated. They have standard deviations of 5 and 4, respectively. If $X_3 = 2X_1 - 3X_2$, what is the correlation between X_1 and X_3?

3.6. PROPERTIES OF COVARIANCE MATRICES

Covariance matrices are always positive semidefinite. To see why, let $\boldsymbol{X}$ be any random vector with covariance matrix $\boldsymbol{\Sigma}$, and let $\boldsymbol{b}$ be any constant row vector. Define the random variable

$$Y = \boldsymbol{b}\boldsymbol{X}. \qquad [3.33]$$

By [3.28], the variance of Y is

$$var(Y) = \boldsymbol{b}\boldsymbol{\Sigma}\boldsymbol{b}'. \qquad [3.34]$$

The variance of any random variable Y must be nonnegative, so expression [3.34] is nonnegative. Recall from Section 2.7 that a symmetric matrix $\boldsymbol{\Sigma}$ is positive semidefinite if $\boldsymbol{b}\boldsymbol{\Sigma}\boldsymbol{b}' \geq 0$ for all row vectors $\boldsymbol{b}$. A covariance matrix is necessarily symmetric, so we conclude that all covariance matrices $\boldsymbol{\Sigma}$ are positive semidefinite.

We shall call a random vector **nonsingular** or **singular** according to whether its covariance matrix is positive definite or singular positive semidefinite.

Singular Random Vectors

Suppose random vector $\boldsymbol{X}$ is singular with covariance matrix $\boldsymbol{\Sigma}$. There exists a row vector $\boldsymbol{b} \neq \boldsymbol{0}$ such that $\boldsymbol{b}\boldsymbol{\Sigma}\boldsymbol{b}' = 0$. Consider the random variable $\boldsymbol{b}\boldsymbol{X}$. By [3.28],

$$var(\boldsymbol{b}\boldsymbol{X}) = \boldsymbol{b}\boldsymbol{\Sigma}\boldsymbol{b}' = 0. \qquad [3.35]$$

Since random variable $\boldsymbol{b}\boldsymbol{X}$ has 0 variance, it must equal some constant a. This argument is reversible, so we conclude that a random vector $\boldsymbol{X}$ is singular if and only if there exists a row vector $\boldsymbol{b} \neq \boldsymbol{0}$ and a constant a such that

$$\boldsymbol{b}\boldsymbol{X} = a. \qquad [3.36]$$

Dispensing with matrix notation, this becomes

$$b_n X_n + \cdots + b_2 X_2 + b_1 X_1 = a. \qquad [3.37]$$

Since $\boldsymbol{b} \neq \boldsymbol{0}$, at least one component b_i is nonzero. Without loss of generality, assume $b_1 \neq 0$. Rearranging [3.37], we obtain

$$X_1 = \left(-\frac{b_n}{b_1}\right)X_n + \cdots + \left(-\frac{b_2}{b_1}\right)X_2 + a, \qquad [3.38]$$

which expresses component X_1 as a linear polynomial of the other components X_i. We conclude that a random vector $\boldsymbol{X}$ is singular if and only if one of its components is a linear polynomial of the other components. In this sense, a singular covariance matrix indicates that at least one component of a random vector is extraneous.

If one component of $\boldsymbol{X}$ is a linear polynomial of the rest, then all realizations of $\boldsymbol{X}$ must fall in a plane within $\mathbb{R}^n$, where $m < n$. The random vector $\boldsymbol{X}$ can be thought of as an m-dimensional random vector sitting in a plane within $\mathbb{R}^n$. This is illustrated with realizations of a singular two-dimensional random vector $\boldsymbol{X}$ in Exhibit 3.6.

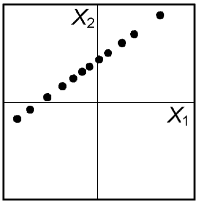

Exhibit 3.6 Realizations of a singular two-dimensional random vector $\boldsymbol{X}$. Component X_2 is a linear polynomial of component X_1.

If a random vector $\boldsymbol{X}$ is singular, but the plane it sits in is not aligned with the coordinate system of $\mathbb{R}^n$, we may not immediately realize that it is singular from its covariance matrix $\boldsymbol{\Sigma}$. A simple test for singularity is to calculate the determinant $|\boldsymbol{\Sigma}|$ of the covariance matrix. If this equals 0, $\boldsymbol{X}$ is singular.

Once we know that $\boldsymbol{X}$ is singular, we can apply a change of variables to eliminate extraneous components X_i and transform $\boldsymbol{X}$ into an equivalent m-dimensional random vector $\boldsymbol{Y}$, $m < n$. The change of variables will do this by transforming (rotating, shifting, etc.) the plane that realizations of $\boldsymbol{X}$ sit in so that it aligns with the coordinate system of $\mathbb{R}^n$. Such a change of variables is obtained with a linear polynomial of the form:

$$\boldsymbol{X} = \boldsymbol{k}\boldsymbol{Y} + \boldsymbol{d}. \tag{3.39}$$

Consider a three-dimensional random vector $\boldsymbol{X}$ with mean vector and covariance matrix

$$\boldsymbol{\mu}_{\boldsymbol{X}} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{X}} = \begin{pmatrix} 25 & 8 & 10 \\ 8 & 4 & 2 \\ 10 & 2 & 5 \end{pmatrix}. \tag{3.40}$$

We note that $\boldsymbol{\Sigma}$ has determinant $|\boldsymbol{\Sigma}| = 0$, so it is singular. We propose to transform $\boldsymbol{X}$ into an equivalent two-dimensional random vector $\boldsymbol{Y}$ using a linear polynomial of the form [3.39]. For convenience, let's find a transformation such that $\boldsymbol{Y}$ will have mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}$:

$$\boldsymbol{\mu}_{\boldsymbol{Y}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{Y}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.41}$$

We first solve for $\boldsymbol{k}$. By [3.31]:

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{k}\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{k}' = \boldsymbol{k}\boldsymbol{I}\boldsymbol{k}' = \boldsymbol{k}\boldsymbol{k}', \tag{3.42}$$

so we seek a factorization $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{k}\boldsymbol{k}'$. Applying the Cholesky factorization and discarding an extraneous column of 0's, as described in Section 2.7, we obtain:

$$\boldsymbol{k} = \begin{pmatrix} 5.0 & 0.0 \\ 1.6 & 1.2 \\ 2.0 & -1.0 \end{pmatrix}. \tag{3.43}$$

Solving next for $\boldsymbol{d}$, by [3.30]:

$$\boldsymbol{\mu}_X = \boldsymbol{k}\boldsymbol{\mu}_Y + \boldsymbol{d}, \tag{3.44}$$

$$\Rightarrow \boldsymbol{d} = \boldsymbol{\mu}_X - \boldsymbol{k}\boldsymbol{\mu}_Y \tag{3.45}$$

$$= \boldsymbol{\mu}_X - \boldsymbol{k}\boldsymbol{0} \tag{3.46}$$

$$= \boldsymbol{\mu}_X. \tag{3.47}$$

Accordingly, our transformation is

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 5.0 & 0.0 \\ 1.6 & 1.2 \\ 2.0 & -1.0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}. \tag{3.48}$$

Exhibit 3.7 illustrates how this change of variables transforms the plane in which $\boldsymbol{X}$ sits so that it aligns with the coordinate system of $\mathbb{R}^2$.

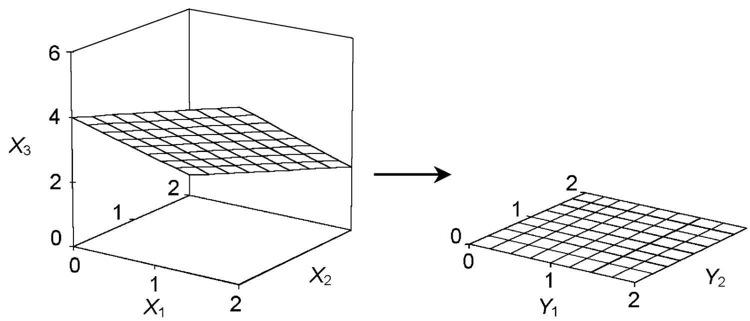

Exhibit 3.7 Our change of variables transforms the plane that realizations of $\boldsymbol{X}$ sit in so that it aligns with the coordinate system of $\mathbb{R}^2$. The third extraneous component of $\boldsymbol{X}$ "drops out."

Multicollinear Random Vectors

Suppose we are analyzing the risk in a natural gas trading portfolio. Random variables represent tomorrow's values for each price the portfolio is exposed to. The portfolio holds New York Mercantile Exchange (NYMEX) Henry Hub futures out to 24 months, so there are 24 futures prices. It also has forward positions out to 18 months for 30 delivery points, for another 540 prices. In total, our model depends upon a vector of 564 random variables!

Based upon a time series analysis of historical price data, we construct a 564 × 564 covariance matrix for our random vector of prices. Gazing at the 318,096 variances and covariances of our matrix, we wonder: Do we really need all these numbers?

Intuitively, we know that the random variables are interdependent. Prices for 6-month and 7-month Transco Zone 2 delivery are highly correlated. So are

3-month prices for adjacent Transco Zones 1 and 2. Because of such interdependencies, it is conceivable that our random vector is singular, but this is probably not the case. Singularity arises infrequently in applications. A more common situation is "almost" singularity, which is known as **multicollinearity**.

We illustrate with two four-dimensional random vectors. Random vector $\boldsymbol{X}$ is singular. Its first three components X_1, X_2, and X_3 are uncorrelated, each with mean 0 and standard deviation 1. The fourth component X_4 equals $X_1 + X_2 + X_3$. The covariance matrix for $\boldsymbol{X}$ is

$$\boldsymbol{\Sigma}_X = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 3 \end{pmatrix}. \qquad [3.49]$$

Random vector $\boldsymbol{Z}$ is multicollinear. Like $\boldsymbol{X}$, its first three components Z_1, Z_2, and Z_3 are uncorrelated, each with mean 0 and standard deviation 1. The fourth component Z_4 equals $Z_1 + Z_2 + Z_3 + E$, where E is a "noise" random variable that is uncorrelated with Z_1, Z_2, and Z_3 and has mean 0 and standard deviation .001. Except for the addition of "noise" E, our random vector $\boldsymbol{Z}$ is identical to our random vector $\boldsymbol{X}$. Its covariance matrix is

$$\boldsymbol{\Sigma}_Z = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 3.000001 \end{pmatrix}. \qquad [3.50]$$

The covariance matrix of $\boldsymbol{X}$ is singular. It has determinant 0. The covariance matrix of $\boldsymbol{Z}$ is not singular, but with a determinant of .000001, it is "almost" singular. The random variable Z_4 is almost a linear polynomial of Z_1, Z_2, and Z_3, but not quite. We added just enough random "noise" to make it linearly independent. We say a random vector is **multicollinear** if it is "almost" singular in this sense.

Realizations of a multicollinear random vector tend to cluster near a plane within $\mathbb{R}^n$. They don't all lie in that plane, but they "almost" do. This is illustrated with realizations of a two-dimensional multicollinear random vector $\boldsymbol{Z}$ in Exhibit 3.8.

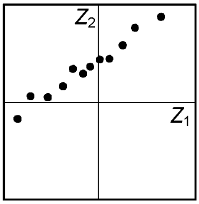

Exhibit 3.8 Realizations of a multicollinear two-dimensional random vector $\boldsymbol{Z}$. Component Z_2 is "almost" a linear polynomial of component Z_1.

We may think of a random vector $\boldsymbol{Z}$ as being "almost" singular if its covariance matrix has a determinant $|\boldsymbol{\Sigma}_{\boldsymbol{Z}}|$ close to 0. In practical applications, the magnitude of this determinant will depend upon the units in which components of $\boldsymbol{Z}$ are measured. A more reasonable test for multicollinearity is to consider the determinant $|\boldsymbol{\rho}_{\boldsymbol{Z}}|$ of the correlation matrix of $\boldsymbol{Z}$. This determinant will always be in the interval [0, 1]. If it is close to 0, this is an indication of multicollinearity. Obviously, if it equals 0, $\boldsymbol{Z}$ is singular.

As we have seen, the dimensionality of a singular random vector $\boldsymbol{X}$ can be reduced with a simple change of variables. No information is lost, as we only eliminate extraneous random variables. Multicollinearity is more problematic. Reducing the dimensionality of a multicollinear random vector $\boldsymbol{Z}$ requires an approximation that somehow identifies and discards minor randomness that is preventing the covariance matrix from being singular.

This is the situation we face with our natural gas portfolio. We feel confident that the natural gas market can reasonably be modeled with less than 564 random variables, but we can't arbitrarily discard random variables! If our covariance matrix isn't singular, how can we replace our 564 random variables with a smaller set that conveys essentially the same information? Principal component analysis will provide a solution.

Exercises

3.17 Below are described four three-dimensional random vectors: $\boldsymbol{W}$, $\boldsymbol{V}$, $\boldsymbol{X}$, and $\boldsymbol{Y}$. Assuming their second moments exist, which of the random vectors has a singular covariance matrix?

a. Components V_1 and V_2 are independent. Component $V_3 = 2V_1 - 5V_2 + 1$.

b. Components W_1 and W_2 are independent. Component $W_3 = W_1 - log(W_2)$.

c. Components X_1, X_2, and X_3 represent next year's total returns for three different companies' common stocks.

d. Components Y_1 and Y_2 represent tomorrow's prices for the nearby 3-month Treasury bill and 3-month Eurodollar futures. Component Y_3 represents tomorrow's price difference between those two futures.

3.18 True or false:

a. A covariance matrix is singular if and only if it is positive definite.

b. A covariance matrix is nonsingular if and only if it is positive semidefinite.

c. Every random vector has a positive semidefinite covariance matrix.

3.19 Which of the following covariance matrices $\boldsymbol{\Sigma}$ are singular? Which are multicollinear?

a.

$$\begin{pmatrix} 1.21 & 2.31 & 1.32 \\ 2.31 & 4.41 & 2.52 \\ 1.32 & 2.52 & 2.65 \end{pmatrix} \qquad [3.51]$$

b.

$$\begin{pmatrix} 1.69 & 1.82 & 0.91 \\ 1.82 & 1.97 & 0.98 \\ 0.91 & 0.98 & 0.58 \end{pmatrix} \qquad [3.52]$$

c.

$$\begin{pmatrix} 8.19 & 1.20 & 1.68 \\ 1.20 & 2.68 & 2.24 \\ 1.68 & 2.24 & 6.40 \end{pmatrix} \qquad [3.53]$$

3.20 Consider a singular random vector $\boldsymbol{X}$ with mean vector and covariance matrix

$$\boldsymbol{\mu}_{\boldsymbol{X}} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{X}} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 13 & 1 \\ -1 & 1 & 2 \end{pmatrix}. \qquad [3.54]$$

Transform $\boldsymbol{X}$ into an equivalent two-dimensional random vector $\boldsymbol{Y}$ with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}$:

$$\boldsymbol{\mu}_{\boldsymbol{Y}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{Y}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \qquad [3.55]$$

3.7. PRINCIPAL COMPONENT ANALYSIS

With principal component analysis, we transform a random vector $\boldsymbol{Z}$ with correlated components Z_i into a random vector $\boldsymbol{D}$ with uncorrelated components D_i. This is called an **orthogonalization** of $\boldsymbol{Z}$.

Principal component analysis can be performed on any random vector $\boldsymbol{Z}$ whose second moments exist, but it is most useful with multicollinear random vectors. Principal component analysis takes the plane in which realizations of a multicollinear random vector "almost" sit and aligns it with the coordinate system of $\mathbb{R}^n$.

The components of $\boldsymbol{D}$ that are perpendicular to the transformed plane have small, almost trivial standard deviations. Discarding these components provides a lower-dimensional approximate representation for $\boldsymbol{Z}$. This is illustrated with realizations of a multicollinear two-dimensional random vector $\boldsymbol{Z}$ in Exhibit 3.9:

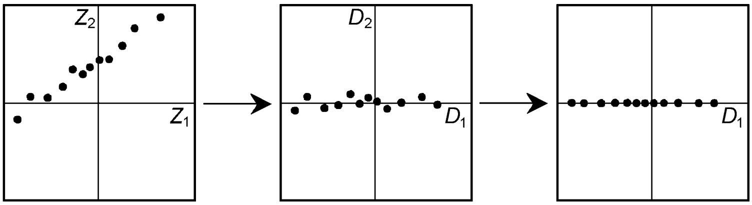

Exhibit 3.9 Principal component analysis can be used to reduce the dimensionality of a multicollinear random vector. Realizations for a multicollinear two-dimensional random vector $\boldsymbol{Z}$ are illustrated in the left graph. Principal component analysis transforms $\boldsymbol{Z}$ into an equivalent multicollinear random vector $\boldsymbol{D}$ that is aligned with the coordinate system of $\mathbb{R}^2$. Realizations of $\boldsymbol{D}$ are shown in the middle graph. Discarding the second component D_2 of $\boldsymbol{D}$ transforms $\boldsymbol{D}$ into a one-dimensional approximate representation of the two-dimensional $\boldsymbol{Z}$. Realizations of this representation are shown in the right graph.

EXAMPLE: EUROPEAN CURRENCIES. Suppose today is June 30, 2000. We consider a random vector $\boldsymbol{Z}$ whose components represent the simple price returns that specific European currencies will realize versus the US dollar (USD) over the upcoming trading day:

$$\boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7 \end{pmatrix} \sim \begin{pmatrix} \text{Swiss franc (CHF) price return} \\ \text{Danish krone (DKK) price return} \\ \text{Euro (EUR) price return} \\ \text{British pound (GBP) price return} \\ \text{Greek drachma (GRD) price return} \\ \text{Norwegian krone (NOK) price return} \\ \text{Swedish krona (SEK) price return} \end{pmatrix}. \qquad [3.56]$$

Exhibit 3.10 graphs 18 months of daily exchange-rate data drawn from the period immediately following the launch of the new EUR currency. In our data, the EUR weakens following its launch, and the remaining European currencies—those that did not join the EUR on January 1, 1999—weaken in sympathy. All the currencies track the EUR, but the GBP does so the least. It is less correlated with the EUR and loses value more slowly.

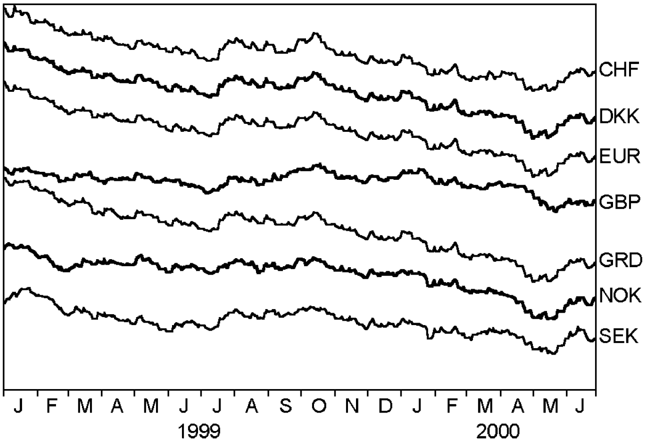

Exhibit 3.10 Historical exchange rates versus the USD for the period January 1, 1999 through June 30, 2000. Exchange rates are presented as USD/unit of currency, so a rising curve indicates a strengthening currency. Exchange rates are individually scaled so they all fit on the graph.

We assume $\boldsymbol{\mu}_Z = \mathbf{0}$.[5] Based upon a time series analysis of the historical price data, we construct a covariance matrix for $\boldsymbol{Z}$

$$\boldsymbol{\Sigma}_Z = \begin{pmatrix} .00004256 & .00003912 & .00003900 & .00001737 & .00003776 & .00002726 & .00002735 \\ .00003912 & .00003994 & .00003929 & .00001602 & .00003865 & .00002829 & .00002848 \\ .00003900 & .00003929 & .00003935 & .00001599 & .00003829 & .00002793 & .00002810 \\ .00001737 & .00001602 & .00001599 & .00002105 & .00001524 & .00001256 & .00001175 \\ .00003776 & .00003865 & .00003829 & .00001524 & .00003961 & .00002782 & .00002805 \\ .00002726 & .00002829 & .00002793 & .00001256 & .00002782 & .00002959 & .00002603 \\ .00002735 & .00002848 & .00002810 & .00001175 & .00002805 & .00002603 & .00003220 \end{pmatrix}. \quad [3.57]$$

The corresponding correlation matrix is:

$$\boldsymbol{\rho}_Z = \begin{pmatrix} 1 & .9488 & .9530 & .5804 & .9197 & .7682 & .7388 \\ .9488 & 1 & .9911 & .5523 & .9717 & .8229 & .7940 \\ .9530 & .9911 & 1 & .5554 & .9698 & .8184 & .7894 \\ .5804 & .5523 & .5554 & 1 & .5276 & .5031 & .4512 \\ .9197 & .9717 & .9698 & .5276 & 1 & .8124 & .7855 \\ .7682 & .8229 & .8184 & .5031 & .8124 & 1 & .8434 \\ .7388 & .7940 & .7894 & .4512 & .7855 & .8434 & 1 \end{pmatrix}. \quad [3.58]$$

The correlations are all positive. Several exceed 0.90. The one between DKK and EUR exceeds 0.99. The smallest is a respectable 0.45 between GBP and SEK.

[5] An alternative would be to derive a mean vector based upon interest rate parity.

With such pronounced interdependencies between its components, we expect $\boldsymbol{Z}$ to be multicollinear, and it is. The correlation matrix has determinant $|\boldsymbol{\rho}| = .0000045$.

To define principal components of $\boldsymbol{Z}$, we calculate orthonormal[6] eigenvectors $\boldsymbol{v}_i$ of the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{Z}$. We arrange these as the columns of a matrix:

$$\boldsymbol{v} = \begin{pmatrix} .4331 & -.2211 & -.2704 & .1110 & .7702 & -.2897 & -.0306 \\ .4323 & -.0401 & -.2237 & -.0083 & -.1764 & .4874 & -.7020 \\ .4288 & -.0566 & -.2312 & -.0005 & -.1269 & .4873 & .7112 \\ .1929 & -.7592 & .6074 & .0542 & -.1208 & .0005 & -.0013 \\ .4245 & .0026 & -.2402 & -.0362 & -.5658 & -.6635 & .0198 \\ .3324 & .3361 & .4134 & -.7622 & .1549 & -.0261 & .0057 \\ .3373 & .5070 & .4753 & .6343 & .0270 & -.0118 & .0048 \end{pmatrix}. \quad [3.59]$$

The eigenvectors are graphed in Exhibit 3.11. Corresponding eigenvalues λ_i are also indicated.

The eigenvectors may be thought of as "modes of fluctuation" of random vector $\boldsymbol{Z}$. We observed in our historical data a tendency for the European currencies to move together. This is reflected in the first eigenvector. It describes a broad move in all the currencies, with the GBP participating about half as much as the other currencies. The second eigenvector has the GBP moving in opposition to the NOK and SEK, with the CHF moving modestly with the GBP. The third eigenvector describes the GBP, NOK, and SEK moving together in opposition to the other currencies. The remaining eigenvectors describe other "modes of fluctuation."

If the eigenvectors $\boldsymbol{v}_i$ are modes of fluctuations of $\boldsymbol{Z}$, then $\boldsymbol{Z}$ is a random combination of those modes of fluctuation:

$$\boldsymbol{Z} = D_1\boldsymbol{v}_1 + D_2\boldsymbol{v}_2 + \cdots + D_7\boldsymbol{v}_7 = \boldsymbol{v}\boldsymbol{D}. \quad [3.60]$$

The D_i are the principal components of $\boldsymbol{Z}$. They are random variables that define each mode of fluctuation's random contribution to $\boldsymbol{Z}$. The D_i are uncorrelated with variances equal to the eigenvalues of their corresponding eigenvectors. The vector $\boldsymbol{D}$ of principal components has mean $\boldsymbol{\mu}_{\boldsymbol{D}} = \boldsymbol{0}$ and covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{D}} = \begin{pmatrix} .00020719 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .00001469 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .00001305 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .00000471 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .00000315 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .00000117 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .00000035 \end{pmatrix}. \quad [3.61]$$

[6]A set of vectors is **orthonormal** if they are orthogonal and normalized (of length 1).

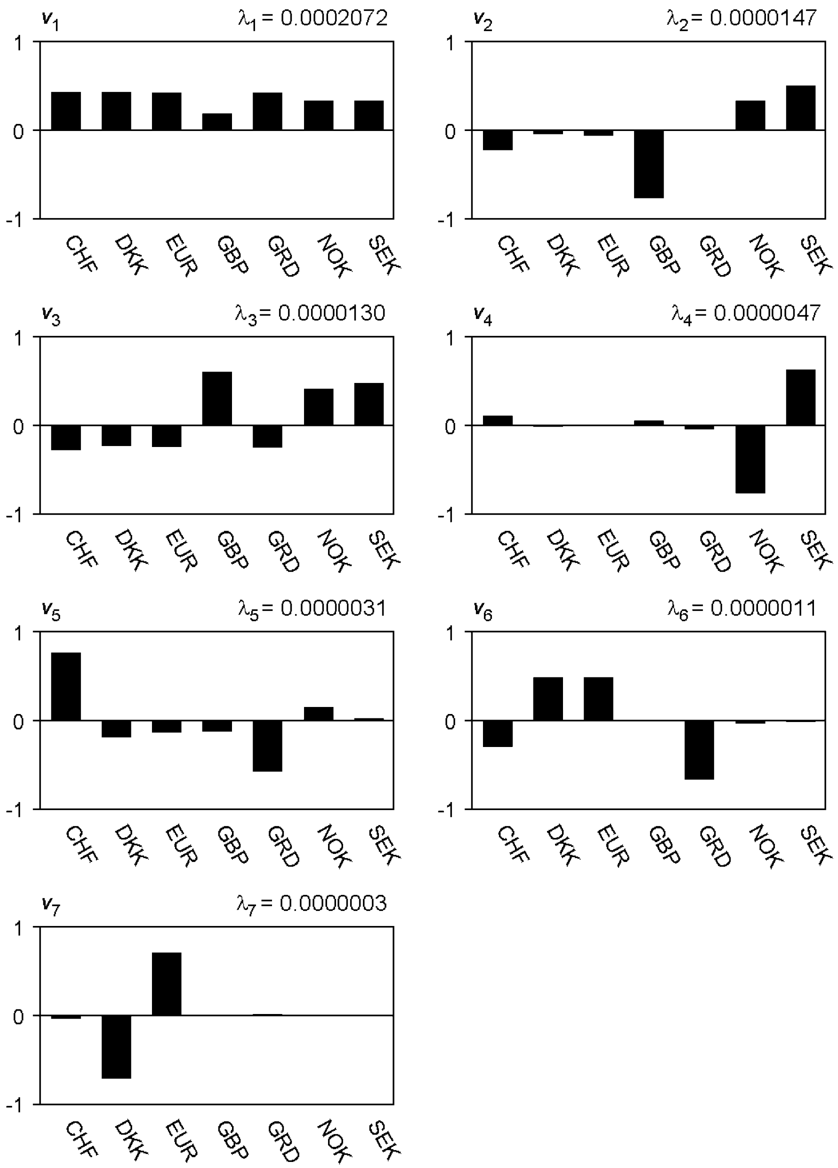

Exhibit 3.11 Eigenvectors $\boldsymbol{v}_i$ of covariance matrix [3.57]. Corresponding eigenvalues λ_i are also indicated.

We have ordered our principal components according to their variances. From our covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{D}}$, we see that the first three principal components are more significant than the rest. The last two principal components, D_6 and D_7, have variances that are less than 1% of the variance of D_1. Their contribution to random vector $\boldsymbol{Z}$ is trivial.

We can approximate $\boldsymbol{Z}$ by discarding from [3.60] insignificant principal components. The more we discard, the simpler—and cruder!—will be our approximation. If we want to be aggressive in our approximation, we can discard the contributions of the last four principal components, and approximate $\boldsymbol{Z}$ with just the first three. A more accurate approximation can be obtained by discarding only the last two. For this example, we pursue the more aggressive course. We define

$$\widetilde{\boldsymbol{Z}} = D_1\boldsymbol{v}_1 + D_2\boldsymbol{v}_2 + D_3\boldsymbol{v}_3. \tag{3.62}$$

and approximate $\boldsymbol{Z}$ with $\widetilde{\boldsymbol{Z}}$. Like $\boldsymbol{Z}$, $\widetilde{\boldsymbol{Z}}$ has mean vector $\boldsymbol{0}$. Its covariance matrix is obtained from [3.61] and [3.62] using [3.31]:

$$\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{Z}}} = \begin{pmatrix} .00004054 & .00003971 & .00003948 & .00001763 & .00003893 & .00002728 & .00002694 \\ .00003971 & .00003940 & .00003912 & .00001595 & .00003872 & .00002837 & .00002853 \\ .00003948 & .00003912 & .00003884 & .00001594 & .00003844 & .00002800 & .00002811 \\ .00001763 & .00001595 & .00001594 & .00002099 & .00001503 & .00001281 & .00001159 \\ .00003893 & .00003872 & .00003844 & .00001503 & .00003809 & .00002795 & .00002820 \\ .00002728 & .00002837 & .00002800 & .00001281 & .00002795 & .00002678 & .00002830 \\ .00002694 & .00002853 & .00002811 & .00001159 & .00002820 & .00002830 & .00003030 \end{pmatrix}. \tag{3.63}$$

Comparing this covariance matrix with [3.57], you can judge for yourself the quality of our approximation.

Principal Components

Our example informally introduced principal components. Now let's formalize them. Consider an n-dimensional random vector $\boldsymbol{Z}$ with mean $\boldsymbol{\mu}_{\boldsymbol{Z}}$ and nonsingular covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{Z}}$. We construct principal components in such a manner that the first accounts for as much of the variability of $\boldsymbol{Z}$ as possible. The second accounts for as much of the remaining variability of $\boldsymbol{Z}$ as possible, and so on.

Specifically, the first principal component D_1 is defined as

$$D_1 = \boldsymbol{v}_1'(\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}}), \tag{3.64}$$

where $\boldsymbol{v}_1$ has unit length and is selected to maximize the variance of D_1. This is achieved by setting $\boldsymbol{v}_1$ equal to the normalized first eigenvector of $\boldsymbol{\Sigma}_{\boldsymbol{Z}}$—the eigenvector with the largest eigenvalue. In this case, the variance of D_1 equals that eigenvalue, λ_1.

The second principal component D_2 is defined as

$$D_2 = \boldsymbol{v}_2'(\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}}), \tag{3.65}$$

where $\boldsymbol{v}_2$ is selected from the set of all n-dimensional unit vectors that are orthogonal to $\boldsymbol{v}_1$ in such a manner as to maximize the variance of D_2. This is achieved by setting $\boldsymbol{v}_2$ equal to the normalized second eigenvector of $\boldsymbol{\Sigma_Z}$—the eigenvector with the second largest eigenvalue. The variance of D_2 equals that eigenvalue, λ_2.

Proceeding in this manner, we define the remaining principal components. There will be m principal components D_i, each one corresponding to a normalized eigenvector $\boldsymbol{v}_i$ of $\boldsymbol{\Sigma_Z}$. We can represent

$$\boldsymbol{Z} = D_1\boldsymbol{v}_1 + D_2\boldsymbol{v}_2 + \cdots + D_m\boldsymbol{v}_m + \boldsymbol{\mu_Z} = \boldsymbol{vD} + \boldsymbol{\mu_Z}. \qquad [3.66]$$

The vector of principal components $\boldsymbol{D}$ has mean $\boldsymbol{\mu_D} = \boldsymbol{0}$ and covariance matrix

$$\boldsymbol{\Sigma_D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix}. \qquad [3.67]$$

If $\boldsymbol{\Sigma_Z}$ is nonsingular, the number m of principal components equals the dimensionality n of $\boldsymbol{Z}$. If $\boldsymbol{\Sigma_Z}$ is singular, some of its eigenvalues will equal 0, and the number m of principal components will be less than the dimensionality n of $\boldsymbol{\Sigma_Z}$. In this case, [3.66] will have reduced the dimensionality of the singular $\boldsymbol{Z}$ in the same manner as that described in Section 3.6.

Choice of Weights

Principal component analysis is best performed on random variables whose standard deviations are reflective of their relative significance for an application. This is because principal component analysis depends upon both the correlations between random variables and the standard deviations of those random variables. If we were to change the standard deviations of a set of random variables but leave their correlations the same, this would change their principal components. In a sense, principal component analysis uses standard deviation as a metric of significance. If one random variable has a standard deviation that far exceeds the rest, that random variable will dominate the first eigenvector.

Unfortunately, there may be no correspondence between a random variable's standard deviation and its significance. Standard deviations depend upon the units in which a random variable is measured. Suppose a random variable reflects the time it takes for some event to occur, and if the random variable is measured in days, it has a standard deviation of 13.5. If the standard deviation is measured in hours, it is 324. Measured in minutes, it becomes 19,440. Certainly, the 19,440 standard deviation is no more significant than the 13.5 standard deviation, but principal component analysis will treat it as more significant!

If we use principal components only to orthogonalize a random vector, this will not be a problem. No information is lost. It will be a problem if principal components are discarded to form an approximation. In this case, information is lost. Before we discard principal components that appear "insignificant," we should make sure that they truly are insignificant.

There are various solutions to this problem. We might insist that all random variables be measured in the same units, but this is not always feasible. If one random variable represents temperature and another represents volume, these are fundamentally different quantities. Also, identical units do not necessarily correspond to identical significance. Suppose we are analyzing blood samples for lead, and we have a random variable for each component of the blood. All components are measured in parts per million (ppm). Measured in ppm, the standard deviation of lead will be trivial compared to standard deviations for other constituents of the blood. Yet, the lead component is the most important random variable!

Alternatively, we might apply principal component analysis to normalized random variables:

$$Z_i^* = \frac{Z_i}{\sigma_i}. \qquad [3.68]$$

With this approach, we effectively apply principal component analysis to the random variables' correlation matrix. This represents a different weighting from that obtained by measuring all random variables in identical units, but not necessarily a better one.

Any solution may be reasonable in certain contexts and unreasonable in others. Each one weights the random variables in some manner. There is no objective way to assign weights, just as there is no objective way to assign "significance." Weights and "significance" can and should vary from one application to another. When we use principal components to reduce the dimensionality of a random vector, there is subjectivity in the process.

Exercises

3.21 Consider a random vector $\boldsymbol{Z}$ with mean and covariance matrix

$$\boldsymbol{\mu}_Z = \begin{pmatrix} 2.1 \\ 0.4 \\ 1.6 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_Z = \begin{pmatrix} 13.82 & 10.73 & 12.21 \\ 10.73 & 16.82 & 1.74 \\ 12.21 & 1.74 & 18.18 \end{pmatrix}. \qquad [3.69]$$

a. Calculate the determinant of the corresponding correlation matrix.
b. Is $\boldsymbol{Z}$ singular, multicollinear, or neither of these?
c. 🖩[7] Calculate the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}_Z$.

[7] The symbol 🖩 indicates that this is a computationally involved problem that may require computing tools more sophisticated than a spreadsheet.

d. Represent $\boldsymbol{Z}$ in terms of its principal components as in [3.66].
e. What is the covariance matrix $\boldsymbol{\Sigma_D}$ of the vector of principal components $\boldsymbol{D}$?
f. Construct an approximation $\widetilde{\boldsymbol{Z}}$ for $\boldsymbol{Z}$ based on the first two principal components of $\boldsymbol{Z}$.
g. Construct the covariance matrix of $\widetilde{\boldsymbol{Z}}$. Compare your result with the covariance matrix of $\boldsymbol{Z}$.

3.8. UNIFORM AND RELATED DISTRIBUTIONS

Until now, we have avoided mentioning any standard families of distributions such as the uniform, normal, or chi-squared families of distributions. This has been intentional. All the results we have discussed so far are general and assume no particular distributions. We emphasized this fact by not incorporating any standard distributions into the discussion. Let's reiterate. Formulas [3.30] and [3.31] for the mean and variance of a linear polynomial of a random vector are entirely general. So are principal component analysis, the fact that covariance matrices are positive semidefinite, and every other result we have presented so far. None assumes any particular distribution.

To construct probabilistic models, it is useful to consider standard families of distributions. We assume passing familiarity with a variety of distributions. See Evans et al. (1993) for summary information on important distributions. For our purposes, uniform distributions play a particularly important role, as do normal and related distributions. We review important properties of these in this and the next section. The distributions we consider in this section, along with a shorthand notation for each, are the:

1. uniform distribution: $U(a, b)$,
2. multivariate uniform distribution: $U_n(\Omega)$.

Uniform Distribution

A **uniform distribution** has constant probability density on an interval (a, b) and zero probability density elsewhere. The distribution is specified by two parameters: the end points a and b. We denote the distribution $U(a, b)$. Its PDF is

$$\phi(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b, \\ 0 & \text{otherwise,} \end{cases} \qquad [3.70]$$

which is illustrated in Exhibit 3.12

Exhibit 3.12 PDF of a uniform distribution $U(a, b)$.

A $U(a, b)$ random variable has CDF and inverse CDF:

$$\Phi(x) = \frac{x-a}{b-a} \qquad a < x < b, \tag{3.71}$$

$$\Phi^{-1}(q) = a + (b-a)q \quad 0 < q < 1. \tag{3.72}$$

The expectation, standard deviation, skewness, and kurtosis of a $U(a, b)$ random variable are:

$$\mu = \frac{a+b}{2}, \tag{3.73}$$

$$\sigma = \frac{b-a}{2\sqrt{3}}, \tag{3.74}$$

$$\eta_1 = 0, \tag{3.75}$$

$$\eta_2 = \frac{9}{5}. \tag{3.76}$$

Multivariate Uniform Distribution

Let $\Omega \subset \mathbb{R}^n$ be a bounded region with volume (area) $v(\Omega)$. The **multivariate uniform** distribution on Ω is denoted $U_n(\Omega)$ and has PDF

$$\phi(\boldsymbol{x}) = \begin{cases} \dfrac{1}{v(\Omega)} & \text{if } \boldsymbol{x} \in \Omega \\ 0 & \text{if } \boldsymbol{x} \notin \Omega \end{cases}. \tag{3.77}$$

In applications, the distribution $U_n((0, 1)^n)$ often arises. If $\boldsymbol{X} \sim U_n((0, 1)^n)$, its components X_i are independent random variables, each with marginal distribution $U(0, 1)$.

Exercises

3.22 Answer the following questions:

a. If $U \sim U(0, 1)$, how is $V = 1 - U$ distributed?

b. If $U \sim U(0, 1)$, how is $W = bU + a$ distributed for arbitrary constants a, b?

3.23 Suppose $\boldsymbol{U} \sim U_n((0, 1)^n)$. What is $Pr(U_i > 0.5 \text{ for all } i)$?

3.24 Suppose $\boldsymbol{V} \sim U_n(\Omega)$ where the region Ω is indicated in Exhibit 3.13.

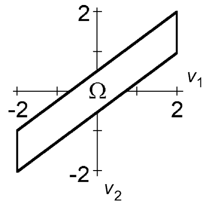

Exhibit 3.13 Region Ω for Exercise 3.24.

a. Is component V_1 marginally uniformly distributed?
b. Is component V_2 marginally uniformly distributed?
c. Are components V_1 and V_2 independent?

3.9. NORMAL AND RELATED DISTRIBUTIONS

We now consider normal and related distributions. Particular families we consider, along with shorthand notation for each, are:

1. normal: $N(\mu, \sigma^2)$,
2. lognormal: $\Lambda(\mu, \sigma^2)$,
3. chi-squared: $\chi^2(\nu, \delta^2)$,
4. joint-normal: $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Normal Distributions

A **normal distribution** is specified by two parameters: a mean μ and variance σ^2. We denote it $N(\mu, \sigma^2)$. Its PDF is

$$\phi(x) = \frac{exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{\sigma\sqrt{2\pi}}. \qquad [3.78]$$

This is graphed in Exhibit 3.14:

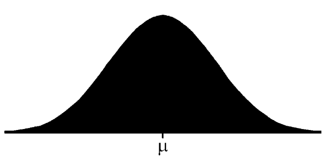

Exhibit 3.14 PDF of a normal distribution.

Irrespective of its mean or standard deviation, every normal distribution has skewness and kurtosis

$$\eta_1 = 0, \quad [3.79]$$

$$\eta_2 = 3. \quad [3.80]$$

With a kurtosis of 3, normal distributions fall precisely between platykurtosis and leptokurtosis. Distributions that have lower kurtosis than a normal distribution are platykurtic. Those that have higher kurtosis are leptokurtic.

A linear polynomial of a normal random variable is also normal. If $X \sim N(\mu, \sigma^2)$, then

$$bX + a \sim N(b\mu + a, (b\sigma)^2). \quad [3.81]$$

for any constants $a, b \in \mathbb{R}$. This means that any $N(\mu, \sigma^2)$ random variable X can be expressed as a linear polynomial of some $N(0, 1)$ random variable Z:

$$X = \sigma Z + \mu. \quad [3.82]$$

We call $N(0, 1)$ the **standard normal distribution**.

It has been proven that there is no closed-form expression for the CDF Φ of a normal distribution. The function exists. It simply cannot be expressed in terms of other standard functions. In practice, it and its inverse Φ^{-1} are approximated to many decimal places using computer algorithms. See Patel (1996).

Based upon [3.82], it follows that any quantile of an $N(\mu, \sigma^2)$ distribution occurs a distance from its mean μ that is a fixed multiple of σ. For example, the .90-quantile $\Phi_Z^{-1}(.90)$ of a standard normal variable Z is obtained from a standard normal table as 1.282. For any $N(\mu, \sigma^2)$ random variable X, the .90-quantile $\Phi_X^{-1}(.90)$ occurs 1.282 standard deviations σ above its mean μ because, by [3.82],

$$.90 = Pr(Z \leq 1.282) \quad [3.83]$$

$$= Pr((X - \mu)/\sigma \leq 1.282) \quad [3.84]$$

$$= Pr(X \leq 1.282\sigma + \mu). \quad [3.85]$$

The result is independent of the values of μ and σ. Accordingly, the .90-quantile of any $N(\mu, \sigma^2)$ random variable is 1.282 standard deviations σ greater than its mean μ. Results for other quantiles are shown in Exhibit 3.15:

q	$\Phi_X^{-1}(q)$
.50	μ
.90	$\mu + 1.282\sigma$
.95	$\mu + 1.645\sigma$
.975	$\mu + 1.960\sigma$
.99	$\mu + 2.326\sigma$

Exhibit 3.15 Selected quantiles of an $N(\mu, \sigma^2)$ distribution.

Because a normal distribution is symmetrical about its mean, the .10, .05, .025, and .01 quantiles can be obtained by replacing plus signs with minus signs in Exhibit 3.15.

Lognormal Distributions

A random variable X is **lognormally distributed** if the natural logarithm of X is normally distributed. A lognormal distribution may be specified by its mean μ and variance σ^2. Alternatively, it may be specified by the mean m and variance s^2 of the normally distributed $log(X)$. We denote a lognormal distribution $\Lambda(\mu, \sigma^2)$, but its PDF is most easily expressed in terms of m and s:

$$\phi(x) = \begin{cases} \dfrac{exp\left(-\dfrac{1}{2}\left(\dfrac{log(x) - m}{s}\right)^2\right)}{xs\sqrt{2\pi}} & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad . \qquad [3.86]$$

A lognormal distribution is illustrated in Exhibit 3.16.

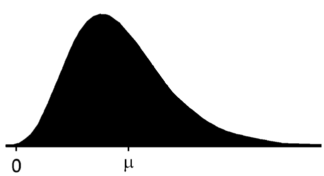

Exhibit 3.16 The PDF of a lognormal distribution.

The expectation, standard deviation, skewness, and kurtosis of a lognormal distribution are, in terms of m and s,

$$\mu = exp([2m + s^2]/2), \qquad [3.87]$$

$$\sigma = \sqrt{exp[2m + 2s^2] - exp[2m + s^2]}, \qquad [3.88]$$

$$\eta_1 = [exp(s^2) + 2]\sqrt{exp(s^2) - 1}, \qquad [3.89]$$

$$\eta_2 = exp(4s^2) + 2exp(3s^2) + 3exp(2s^2) - 3. \qquad [3.90]$$

If we know μ and σ instead of m and s, we can convert between these with

$$m = log\left(\frac{\mu^2}{\sqrt{\sigma^2 + \mu^2}}\right), \qquad [3.91]$$

$$s = \sqrt{log[(\sigma/\mu)^2 + 1]}. \qquad [3.92]$$

The reverse conversion is provided by [3.87] and [3.88].

As with the normal distribution, the CDF of a lognormal distribution exists but cannot be expressed in terms of standard functions. It can be valued using a standard normal table. Let $X \sim \Lambda(\mu, \sigma^2)$ with corresponding parameters m and s. Then $X = exp(sZ + m)$ for some $Z \sim N(0, 1)$. Denote the CDFs of X and Z as Φ_X and Φ_Z. By [3.82] and the definition of the lognormal distribution:

$$\Phi_X(x) = Pr(X \leq x) \quad [3.93]$$

$$= Pr(log(X) \leq log(x)) \quad [3.94]$$

$$= Pr(sZ + m \leq log(x)) \quad [3.95]$$

$$= Pr\left(Z \leq \frac{log(x) - m}{s}\right) \quad [3.96]$$

$$= \Phi_Z\left(\frac{log(x) - m}{s}\right), \quad [3.97]$$

which can be looked up in a standard normal table. Note that step [3.94] depends critically on the monotonicity of the *log* function. As indicated in Exhibit 3.17, $a \leq b$ if and only if $log(a) \leq log(b)$.

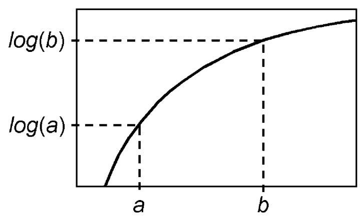

Exhibit 3.17 Because the *log* function is monotone, $a \leq b$ if and only if $log(a) \leq log(b)$.

Chi-Squared Distributions

Suppose Z is a standard normal random variable. How is Z^2 distributed? The answer is a chi-squared distribution. More generally, let $Z_1, Z_2, \ldots Z_\nu$ be ν independent standard normal random variables, and let $\delta_1, \delta_2, \ldots \delta_\nu$ be constants. Then the random variable

$$X = (Z_1 + \delta_1)^2 + (Z_2 + \delta_2)^2 + \cdots + (Z_\nu + \delta_\nu)^2 \quad [3.98]$$

has a chi-squared distribution with ν **degrees of freedom** and **noncentrality parameter**[8]

$$\delta^2 = \sum_{i=1}^{\nu} \delta_i^2. \quad [3.99]$$

[8] Treatment of the noncentrality parameter is not standardized in the literature. Some authors define the parameter as in [3.99] but denote it simply δ. Others define the parameter differently, for example, taking a square root in [3.99] or dividing the sum by 2.

We denote a chi-squared distribution $\chi^2(\nu, \delta^2)$. If $\delta^2 = 0$, the distribution is said to be **centrally chi-squared**. Otherwise, it is said to be **noncentrally chi-squared**. The PDF for a central chi-squared distribution is

$$\phi(x) = \begin{cases} \dfrac{x^{(\nu-2)/2} exp(-x/2)}{2^{\nu/2}\Gamma(\nu/2)} & x > 0, \\ 0 & \text{otherwise.} \end{cases} \quad [3.100]$$

For noncentral chi-squared distributions, this generalizes to

$$\phi(x) = \begin{cases} \dfrac{exp[-(x+\delta^2)/2]}{2^{\nu/2}} \displaystyle\sum_{j=0}^{\infty} \dfrac{x^{j-1+\nu/2}\delta^{2j}}{\Gamma(j+\nu/2)2^{2j}j!} & x > 0, \\ 0 & \text{otherwise,} \end{cases} \quad [3.101]$$

where $\Gamma(\,)$ denotes the gamma function[9]

$$\Gamma(y) = \int_0^\infty e^{-z} z^{y-1} dz. \quad [3.102]$$

The PDF of a chi-squared distribution is illustrated in Exhibit 3.18.

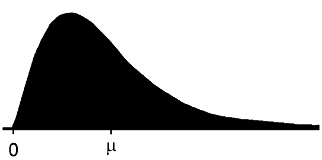

Exhibit 3.18 PDF of a chi-squared distribution.

The expectation, standard deviation, skewness, and kurtosis of a chi-squared distribution are

$$\mu = \nu + \delta^2, \quad [3.103]$$

$$\sigma = \sqrt{2(\nu + 2\delta^2)}, \quad [3.104]$$

$$\eta_1 = \frac{2^{3/2}(\nu + 3\delta^2)}{(\nu + 2\delta^2)^{3/2}}, \quad [3.105]$$

$$\eta_2 = 3 + \frac{12(\nu + 4\delta^2)}{(\nu + 2\delta^2)^2}. \quad [3.106]$$

[9] The gamma function is defined for any $y > 0$. It is related to the factorial function by $\Gamma(y) = (y-1)!$ for $y \in \mathbb{N}$.

Joint-Normal Distributions

Let $\boldsymbol{X}$ be an n-dimensional random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Suppose the marginal distribution of each component X_i is normal. Let Y be a random variable defined as a linear polynomial

$$Y = \boldsymbol{bX} + a \qquad [3.107]$$

of $\boldsymbol{X}$. Based upon [3.27] and [3.28], we can calculate the mean μ_Y and standard deviation σ_Y of Y. Knowing only that the marginal distributions of the X_i are normal, there is little more that we can say about the distribution of Y. However, there is an additional condition that we can impose upon $\boldsymbol{X}$ that will cause Y to be normally distributed. That condition is joint-normality.

The definition of joint-normality is almost trivial. A random vector $\boldsymbol{X}$ is said to be **joint-normal** if every nontrivial linear polynomial Y of $\boldsymbol{X}$ is normal. We denote the n-dimensional joint-normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ as $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\boldsymbol{\Sigma}$ is positive definite, it has PDF

$$\phi(\boldsymbol{x}) = \frac{exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^n\,|\boldsymbol{\Sigma}|}}, \qquad [3.108]$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. Exhibit 3.19 illustrates a joint-normal distribution in two random variables X_1 and X_2. If we define $Y = X_1 + X_2$, then Y is normal.

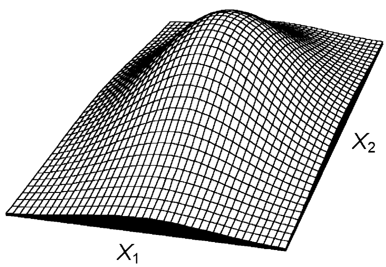

Exhibit 3.19 The PDF of a joint-normal distribution.

Now let's illustrate how a random vector may fail to be joint-normal despite each of its components being marginally normal.[10] Let $\boldsymbol{X}$ be a two-dimensional random vector with components X_1 and X_2. Let X_1 and Z be independent $N(0, 1)$ random variables, and define $X_2 = sign(X_1)|Z|$.[11] In this case, both X_1 and X_2

[10] See Stoyanov (1987) for more counterexamples relating to the joint-normal distribution.

[11] The *sign* function is defined as:

$$sign(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}.$$

are $N(0, 1)$, but the vector $\boldsymbol{X}$, whose PDF is illustrated in Exhibit 3.20, is not joint-normal. In this case, the random variable $Y = X_1 + X_2$ is not normal. Instead, it has the PDF illustrated in Exhibit 3.21.

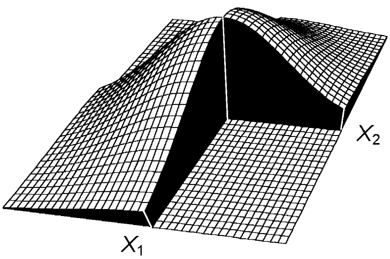

Exhibit 3.20 This PDF illustrates how a random vector $\boldsymbol{X}$ can have two components that are both marginally normal but not be joint-normal.

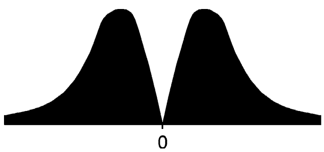

Exhibit 3.21 The PDF of $Y = X_1 + X_2$ is illustrated where X_1 and X_2 are components of random vector $\boldsymbol{X}$, whose PDF is illustrated in Exhibit 3.20. This example illustrates that a linear polynomial of normal random variables need not be normal.

A random vector is joint-normal with uncorrelated components if and only if the components are independent normal random variables.

A property of joint-normal distributions is the fact that marginal distributions and conditional distributions are either normal (if they are univariate) or joint-normal (if they are multivariate). Specifically, let $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Select k components. Without loss of generality, suppose these are the first k components X_1, $X_2, \ldots X_k$. Let $\boldsymbol{X}_1$ be a k-dimensional vector comprising these components, and let $\boldsymbol{X}_2$ be an $(n-k)$-dimensional vector of the remaining components. These partition $\boldsymbol{X}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ into sub-vectors and sub-matrices as follows

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{pmatrix}. \qquad [3.109]$$

The marginal distribution of $\boldsymbol{X}_1$ is $N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{1,1})$ and that of $\boldsymbol{X}_2$ is $N_{n-k}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{2,2})$. If $\boldsymbol{\Sigma}_{2,2}$ is positive definite, the conditional distribution of $\boldsymbol{X}_1$, given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$, is

$$N_k\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{1,2}\boldsymbol{\Sigma}_{2,2}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{1,1} - \boldsymbol{\Sigma}_{1,2}\boldsymbol{\Sigma}_{2,2}^{-1}\boldsymbol{\Sigma}_{2,1}\right). \qquad [3.110]$$

If $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{b}$ is a constant $m \times n$ matrix, and $\boldsymbol{a}$ is an m-dimensional constant vector, then

$$\boldsymbol{bX} + \boldsymbol{a} \sim N_m(\boldsymbol{b\mu} + \boldsymbol{a}, \boldsymbol{b\Sigma b}'). \qquad [3.111]$$

This generalizes property [3.81] of one-dimensional normal distributions.

Exercises

3.25 Answer the following questions. If the answer is some nonstandard distribution or cannot be determined from the information provided, say so.

a. If $N \sim N(1, 4)$, how is $M = 3N + 5$ distributed?
b. If $L \sim \Lambda(1, 3)$, how is $G = log(L)$ distributed?
c. If $N \sim N(2, 6)$, how is $E = e^N$ distributed?
d. If $N_1 \sim N(0, 9)$ and $N_2 \sim N(2, 1)$ have correlation 0.3, how is $M = N_1 + 3N_2$ distributed?
e. If $N_1 \sim N(1, 1)$ and $N_2 \sim N(0, 4)$ are independent, how is $M = 2N_1 + N_2$ distributed?
f. If $N \sim N(0, 1)$, how is $H = N^2$ distributed?
g. If $N_1 \sim N(0, 1)$ and $N_2 \sim N(0, 1)$ are independent, how is $H = N_1^2 + (N_2 + 5)^2$ distributed?
h. If $X \sim \chi^2(1, 0)$, how is $C = \sqrt{X}$ distributed?

3.26 Suppose $Z \sim N(0, 1)$. Use a standard normal table to determine $Pr(Z \leq 1.15)$.

3.27 Suppose $Z \sim N(0, 1)$. Use a standard normal table to determine $Pr(Z \leq -0.51)$. (Hint: Use the symmetry of the normal distribution to find a solution.)

3.28 Suppose $X \sim N(5, 7)$. Use a standard normal table to determine $Pr(X \leq 8)$.

3.29 Suppose $X \sim N(2, .09)$. Use a standard normal table to determine the .90 quantile $\Phi^{-1}(.90)$ of X.

3.30 Suppose $X \sim N(100, 36)$. Use a standard normal table to determine the .15 quantile $\Phi^{-1}(.15)$ of X.

3.31 Suppose $X \sim N(\mu, \sigma^2)$. Use a standard normal table to determine the .10, .70, and .80 quantiles of X.

3.32 Suppose $X \sim \Lambda(1.1, .0625)$. Use a standard normal table to determine $Pr(X \leq .9)$.

3.33 Suppose $X \sim \Lambda(1.05, 0.01)$. Use a standard normal table to determine the .75 quantile $\Phi^{-1}(.75)$ of X.

3.10. MIXTURES OF DISTRIBUTIONS

Consider an experiment. You will flip a fair coin. If it comes up heads, you will draw a number from an $N(0, 4)$ distribution.[12] If it comes up tails, you will draw the number from an $N(0, 9)$ distribution. The number X that results from your experiment has a **mixed normal distribution** with PDF

$$\phi(x) = \frac{1}{2}\frac{exp(-x^2/8)}{2\sqrt{2\pi}} + \frac{1}{2}\frac{exp(-x^2/18)}{3\sqrt{2\pi}}. \qquad [3.112]$$

This is the weighted average of the PDFs of the two normal distributions. More generally, consider m random variables X_k, each with PDF ϕ_k. Define m weights $\xi_k > 0$ that sum to 1. Then the random variable X that has PDF

$$\phi(x) = \sum_{k=1}^{m} \xi_k \phi_k(x) \qquad [3.113]$$

is said to have a **mixed distribution**.

Parameters of Mixed Distributions

Consider a random variable X with a mixed distribution as described above. The X_k have means μ_k and standard deviations σ_k. Then X has mean μ and standard deviation σ given by

$$\mu = \sum_{k=1}^{m} \xi_k \mu_k, \qquad [3.114]$$

$$\sigma = \sqrt{\left(\sum_{k=1}^{m} \xi_k E\left(X_k^2\right)\right) - \mu^2} = \sqrt{\left(\sum_{k=1}^{m} \xi_k \left(\sigma_k^2 + \mu_k^2\right)\right) - \mu^2}. \qquad [3.115]$$

Calculating a q-quantile of X requires that we solve a nonlinear system of equations, which can be done with Newton's method. The q-quantile is that value x such that

$$\Phi_X(x) = \sum_{k=1}^{m} \xi_k \Phi_{X_k}(x) = q, \qquad [3.116]$$

so we seek probabilities $q_1, q_2, \ldots, q_m$ such that

$$\sum_{k=1}^{m} \xi_k q_k = q, \qquad [3.117]$$

[12] We discuss random variate generation in Chapter 5.

while

$$\Phi_{X_1}^{-1}(q_1) = \Phi_{X_2}^{-1}(q_2) = \cdots = \Phi_{X_m}^{-1}(q_m). \qquad [3.118]$$

The desired q-quantile x of X then equals any of—all of—these:

$$x = \Phi_X^{-1}(q) = \Phi_{X_k}^{-1}(q_k). \qquad [3.119]$$

These conditions are motivated for the case $m = 2$ in Exhibit 3.22.

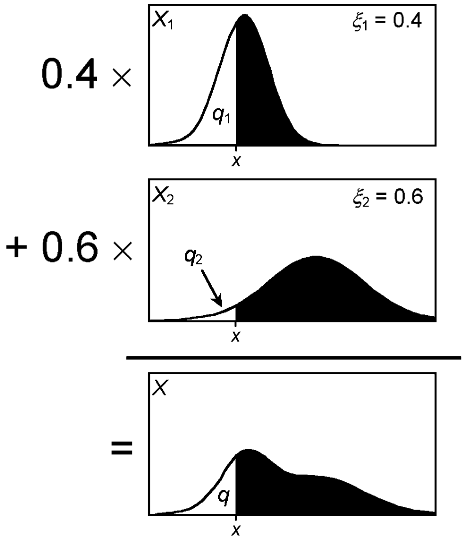

Exhibit 3.22 Random variable X is a mixture of normal random variables X_1 and X_2. Respective weights are $\xi_1 = 0.4$ and $\xi_2 = 0.6$. To find a q-quantile of X, we must find probabilities q_1 and q_2 such that $\xi_1 q_1 + \xi_2 q_2 = q$ while $\Phi_{X_1}^{-1}(q_1) = \Phi_{X_2}^{-1}(q_2)$.

Mixed-Normal Distributions

Since a normal distribution is defined by a mean and standard deviation, a mixed-normal distribution $N^m(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\xi})$ is defined with a vector $\boldsymbol{\mu}$ of means, a vector $\boldsymbol{\sigma}^2$ of variances, and a vector $\boldsymbol{\xi}$ of weights:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}, \quad \boldsymbol{\sigma}^2 = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_m^2 \end{pmatrix}, \quad \boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{pmatrix}, \qquad [3.120]$$

where the weights $\xi_k > 0$ sum to 1.

Mixed-normal distributions are useful for modeling multimodal or leptokurtic distributions. Exhibit 3.23 illustrates PDFs for two mixed-normal distributions.

The first is weighted 0.6 in an $N(-1, 1)$ distribution and 0.4 in an $N(2, 1)$ distribution to achieve a bimodal distribution. The second is evenly weighted in $N(0, 1)$ and $N(0, 9)$ distributions to achieve a leptokurtic distribution.

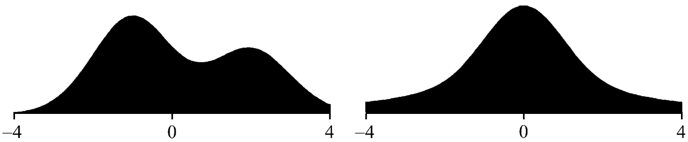

Exhibit 3.23 PDFs for two mixed normal distributions are illustrated. The first is weighted 0.6 in an $N(-1, 1)$ distribution and 0.4 in an $N(2, 1)$ distribution to achieve a bimodal distribution. The second is evenly weighted in $N(0, 1)$ and $N(0, 9)$ distributions to achieve a leptokurtic distribution.

Mixed Joint-Normal Distributions

While our discussion of mixed distributions has focused on random variables, similar concepts generalize for random vectors.

Market professionals often observe that market correlations seem exaggerated during large market swings. This phenomenon can be modeled with a mixture of joint-normal distributions—one with low variances and modest correlations and the other with high variances and more extreme correlations.

Consider vectors of n-dimensional mean vectors, $n \times n$ covariance matrices, and scalar weights:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_m \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_2 \\ \vdots \\ \boldsymbol{\Sigma}_m \end{pmatrix}, \quad \boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{pmatrix}, \qquad [3.121]$$

where the weights $\xi_k > 0$ sum to 1. These define a **mixed joint-normal distribution** $N_n^m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\xi})$ with PDF

$$\phi(\boldsymbol{x}) = \sum_{k=1}^{m} \xi_k \phi_k(\boldsymbol{x}), \qquad [3.122]$$

where $\phi_k(\boldsymbol{x}) \sim N_n(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$.

Exercises

3.34 Derive formulas [3.114] and [3.115].

3.35 Consider random variable $X \sim N^2(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\xi})$, where:

$$\boldsymbol{\mu} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\sigma}^2 = \begin{pmatrix} 4 \\ 9 \end{pmatrix}, \quad \boldsymbol{\xi} = \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix}. \qquad [3.123]$$

Calculate the mean, standard deviation, and .25-quantile of X.

3.11. MOMENT-GENERATING FUNCTIONS

The **moment-generating function** (MGF) of a random variable X is defined as:

$$M_X(w) = E(e^{wX}) \quad [3.124]$$

for $w \in \mathbb{R}$. We call it the moment-generating function because it provides a means of calculating the moments of X. If the MGF is finite on an open interval about $w = 0$, then all the moments of X exist and the k^{th} moment of X equals the k^{th} derivative with respect to w of the MGF evaluated at $w = 0$. Heuristically, we motivate this result by applying the Taylor series expansion for the exponential function in our definition [3.124]:

$$M_X(w) = E\left(1 + wx + \frac{(wx)^2}{2!} + \frac{(wx)^3}{3!} + \frac{(wx)^4}{4!} + \cdots\right). \quad [3.125]$$

If $M_X(w)$ is finite on some interval about the point $w = 0$, it can be shown that the expectation of the sum equals the sum of the expectations

$$M_X(w) = 1 + E(wX) + E\left(\frac{(wX)^2}{2!}\right) + E\left(\frac{(wX)^3}{3!}\right) + E\left(\frac{(wX)^4}{4!}\right) + \cdots \quad [3.126]$$

$$= 1 + wE(X) + \frac{w^2E(X^2)}{2!} + \frac{w^3E(X^3)}{3!} + \frac{w^4E(X^4)}{4!} + \cdots \quad [3.127]$$

You may confirm that the k^{th} derivative of [3.127] with respect to w evaluating at $w = 0$ yields the k^{th} moment $E(X^k)$.

Let X be a random variable and $a, b \in \mathbb{R}$. Define a new random variable $Y = bX + a$. By definition [3.124], the MGF for the new random variable is related to the MGF of X by

$$M_Y(w) = e^{aw}M_X(bw). \quad [3.128]$$

More generally, suppose $\boldsymbol{X}$ is an n-dimensional random vector with independent components X_i; $\boldsymbol{b}$ is an n-dimensional row vector $(b_1 \;\; b_2 \;\; \ldots \;\; b_n)$, and $a \in \mathbb{R}$. Define the random variable $Y = \boldsymbol{bX} + a$. The MGF of Y is

$$M_Y(w) = e^{aw}M_{X_1}(b_1w)M_{X_2}(b_2w)\cdots M_{X_n}(b_nw). \quad [3.129]$$

A uniform, $U(a, b)$, random variable has MGF

$$M(w) = \frac{e^{bw} - e^{aw}}{w(b-a)}. \quad [3.130]$$

Those of $N(\mu, \sigma^2)$ or $\chi^2(\nu, \delta^2)$ random variables are, respectively,

$$M(w) = exp\left(\mu w + \frac{\sigma^2 w^2}{2}\right), \qquad [3.131]$$

$$M(w) = \frac{exp[\delta^2 w/(1-2w)]}{(1-2w)^{\nu/2}}, \quad w < 1/2. \qquad [3.132]$$

The MGF for a lognormal random variable is derived by Leipnik (1991). It is complicated, so we do not present it here.

Exercises

3.36 🖩 Consider a two-dimensional random vector $\boldsymbol{Z}$, whose components are independent and both have $U(0, 1)$ marginal PDFs. Let $Y = Z_1 + Z_2$. Use moment-generating functions to calculate $E(Y^2)$.

3.12. QUADRATIC POLYNOMIALS OF JOINT-NORMAL RANDOM VECTORS

Formulas [3.27] and [3.28] provide general expressions for the mean and variance of a linear polynomial of a random vector. What about the mean and variance of a quadratic polynomial of a random vector? Unfortunately, no general formulas exist. However, if we restrict our attention to quadratic polynomials of a joint-normal random vector, there are expressions for the mean, variance, and all moments. To obtain these, we generalize an earlier result regarding chi-squared distributions.

Simplified Representation

In Section 3.9, we saw that the specific form [3.98] of a quadratic polynomial of a joint standard normal random vector has a chi-squared distribution. Generalizing this, we shall demonstrate that *any* quadratic polynomial of *any* joint-normal random vector can be expressed as a linear polynomial of independent chi-squared and normal random variables. Specifically, let $\boldsymbol{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ positive-definite.[13] Define Y as a quadratic polynomial of $\boldsymbol{X}$:

$$Y = \boldsymbol{X}'\boldsymbol{c}\boldsymbol{X} + \boldsymbol{b}\boldsymbol{X} + a. \qquad [3.133]$$

[13]We lose no generality by assuming $\boldsymbol{\Sigma}$ is positive-definite. If $\boldsymbol{\Sigma}$ were singular, we could perform dimensional reduction as described in Section 3.6 to obtain a positive-definite joint-normal random vector.

Let $\boldsymbol{z}$ be the Cholesky matrix of $\boldsymbol{\Sigma}$, and define $\boldsymbol{u}$ as a square matrix whose rows comprise orthonormal eigenvectors of $\boldsymbol{z}'\boldsymbol{c}\boldsymbol{z}$. By construction, $\boldsymbol{u}$ is orthogonal: $\boldsymbol{u}^{-1} = \boldsymbol{u}'$. Define the change of variables

$$\dot{\boldsymbol{X}} = \boldsymbol{u}\boldsymbol{z}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}). \quad [3.134]$$

Then, by [3.30] and [3.31], $\dot{\boldsymbol{X}}$ is joint-normal with mean vector

$$E(\dot{\boldsymbol{X}}) = E[\boldsymbol{u}\boldsymbol{z}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})] \quad [3.135]$$

$$= \boldsymbol{u}\boldsymbol{z}^{-1}(E[\boldsymbol{X}] - \boldsymbol{\mu}) \quad [3.136]$$

$$= \boldsymbol{u}\boldsymbol{z}^{-1}(\boldsymbol{0}) \quad [3.137]$$

$$= \boldsymbol{0} \quad [3.138]$$

and covariance matrix $\dot{\boldsymbol{\Sigma}}$

$$\dot{\boldsymbol{\Sigma}} = (\boldsymbol{u}\boldsymbol{z}^{-1})\boldsymbol{\Sigma}(\boldsymbol{u}\boldsymbol{z}^{-1})' \quad [3.139]$$

$$= \boldsymbol{u}\boldsymbol{z}^{-1}\boldsymbol{\Sigma}(\boldsymbol{z}')^{-1}\boldsymbol{u}' \quad [3.140]$$

$$= \boldsymbol{u}\boldsymbol{I}\boldsymbol{u}' \quad [3.141]$$

$$= \boldsymbol{I}. \quad [3.142]$$

Accordingly, $\dot{\boldsymbol{X}} \sim N_m(\boldsymbol{0}, \boldsymbol{I})$. Applying our change of variables [3.134]:

$$Y = \boldsymbol{X}'\boldsymbol{c}\boldsymbol{X} + \boldsymbol{b}\boldsymbol{X} + a \quad [3.143]$$

$$= (\boldsymbol{z}\boldsymbol{u}^{-1}\dot{\boldsymbol{X}} + \boldsymbol{\mu})'\boldsymbol{c}(\boldsymbol{z}\boldsymbol{u}^{-1}\dot{\boldsymbol{X}} + \boldsymbol{\mu}) + \boldsymbol{b}(\boldsymbol{z}\boldsymbol{u}^{-1}\dot{\boldsymbol{X}} + \boldsymbol{\mu}) + a \quad [3.144]$$

$$= \dot{\boldsymbol{X}}'(\boldsymbol{u}\boldsymbol{z}'\boldsymbol{c}\boldsymbol{z}\boldsymbol{u}')\dot{\boldsymbol{X}} + [(2\boldsymbol{\mu}'\boldsymbol{c} + \boldsymbol{b})\boldsymbol{z}\boldsymbol{u}']\,\dot{\boldsymbol{X}} + (\boldsymbol{\mu}'\boldsymbol{c}\boldsymbol{\mu} + \boldsymbol{b}\boldsymbol{\mu} + a), \quad [3.145]$$

so Y has form

$$Y = \dot{\boldsymbol{X}}'\dot{\boldsymbol{c}}\dot{\boldsymbol{X}} \; + \; \dot{\boldsymbol{b}}\dot{\boldsymbol{X}} + \; \dot{a}, \quad [3.146]$$

where

$$\dot{\boldsymbol{c}} = \boldsymbol{u}\boldsymbol{z}'\boldsymbol{c}\boldsymbol{z}\boldsymbol{u}', \quad [3.147]$$

$$\dot{\boldsymbol{b}} = (2\boldsymbol{\mu}'\boldsymbol{c} + \boldsymbol{b})\boldsymbol{z}\boldsymbol{u}', \quad [3.148]$$

$$\dot{a} = \boldsymbol{\mu}'\boldsymbol{c}\boldsymbol{\mu} + \boldsymbol{b}\boldsymbol{\mu} + a. \quad [3.149]$$

Recall that we defined $\boldsymbol{u}$ as a matrix whose rows comprise orthonormal eigenvectors of $\boldsymbol{z}'\boldsymbol{c}\boldsymbol{z}$. This means, by the spectral theorem of linear algebra, that the matrix $\dot{\boldsymbol{c}}$ is diagonal with diagonal elements equal to the eigenvalues of $\boldsymbol{z}'\boldsymbol{c}\boldsymbol{z}$. Consequently, Y depends upon no cross terms of the form $\dot{c}_{i,j}\dot{X}_i\dot{X}_j$. We can write [3.145] as

$$Y = \sum_{i=1}^{m} \left(\dot{c}_{i,i}\dot{X}_i^2 + \dot{b}_i\dot{X}_i\right) + \dot{a}, \quad [3.150]$$

and conclude that Y depends upon each of the variables $\dot{X}_i$ in one of four ways:

- No dependence: $\dot{c}_{i,i} = 0$ and $\dot{b}_i = 0$.
- Linear dependence: $\dot{c}_{i,i} = 0$ and $\dot{b}_i \neq 0$, so Y depends upon a term $\dot{b}_i \dot{X}_i$.
- Central quadratic dependence: $\dot{c}_{i,i} \neq 0$ and $\dot{b}_i = 0$, so Y depends upon a term $\dot{c}_{i,i} \dot{X}_i^2$.
- Noncentral quadratic dependence: $\dot{c}_{i,i} \neq 0$ and $\dot{b}_i \neq 0$, so Y depends upon a term $\dot{c}_{i,i} \dot{X}_i^2 + \dot{b}_i \dot{X}_i$.

In the last case, "completing the squares" results in a dependence of the form

$$\dot{c}_{i,i} \left(\dot{X}_i + \frac{\dot{b}_i}{2\dot{c}_{i,i}} \right)^2 . \qquad [3.151]$$

Y is a linear polynomial of independent random variables, each of which is standard normal, central chi-squared with one degree of freedom, or noncentral chi-squared with one degree of freedom and noncentrality parameter $(\dot{b}_i / 2\dot{c}_{i,i})^2$.

Since a linear polynomial of independent normal random variables is itself normal, all normal terms can be combined into one. A general expression for Y is

$$Y = \left(\sum_{k=1}^{n} \gamma_k Q_k \right) + \beta Q_0 + \alpha \qquad [3.152]$$

where the Q_k are chi-squared with one degree of freedom, noncentrality parameters are obtainable from [3.151], and Q_0 is standard normal. The constants γ_k, β, and α can be calculated directly from the terms $\dot{\boldsymbol{c}}$, $\dot{\boldsymbol{b}}$, and $\dot{a}$.

EXAMPLE. Consider random vector $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 2 & 5 \end{pmatrix}. \qquad [3.153]$$

Let

$$Y = \boldsymbol{X}'\boldsymbol{c}\boldsymbol{X} + \boldsymbol{b}\boldsymbol{X} + a, \qquad [3.154]$$

where

$$\boldsymbol{c} = \begin{pmatrix} 3 & 6 & -3 \\ 6 & 16 & -6 \\ -3 & -6 & 3 \end{pmatrix}, \qquad [3.155]$$

$$\boldsymbol{b} = (18 \quad 32 \quad -12), \qquad [3.156]$$

$$a = 12. \qquad [3.157]$$

We wish to express Y as a linear polynomial of independent chi-squared and normal random variables. To do so, we construct the Cholesky matrix $\boldsymbol{z}$ of $\boldsymbol{\Sigma}$,

$$\boldsymbol{z} = \begin{pmatrix} 1.4142 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.7071 & 2.0000 & 0.7071 \end{pmatrix}, \tag{3.158}$$

and a matrix $\boldsymbol{u}$ with rows equal to orthonormal eigenvectors of $\boldsymbol{z}'\boldsymbol{c}\boldsymbol{z}$:

$$\boldsymbol{u} = \begin{pmatrix} 0.0000 & 1.0000 & 0.0000 \\ 0.7071 & 0.0000 & -0.7071 \\ 0.7071 & 0.0000 & 0.7071 \end{pmatrix}. \tag{3.159}$$

We define the change of variables $\dot{\boldsymbol{X}} \sim N_3(\boldsymbol{0}, \boldsymbol{I})$ for $\boldsymbol{X}$

$$\dot{\boldsymbol{X}} = \boldsymbol{u}\boldsymbol{z}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}), \tag{3.160}$$

and obtain

$$Y = \dot{\boldsymbol{X}}'\dot{\boldsymbol{c}}\dot{\boldsymbol{X}} + \dot{\boldsymbol{b}}\dot{\boldsymbol{X}} + \dot{a}, \tag{3.161}$$

where

$$\dot{\boldsymbol{c}} = \boldsymbol{u}\boldsymbol{z}'\boldsymbol{c}\boldsymbol{z}\boldsymbol{u}' = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{3.162}$$

$$\dot{\boldsymbol{b}} = (2\boldsymbol{\mu}'\boldsymbol{c} + \boldsymbol{b})\boldsymbol{z}\boldsymbol{u}' = (0 \quad 12 \quad 6), \tag{3.163}$$

$$\dot{a} = \boldsymbol{\mu}'\boldsymbol{c}\boldsymbol{\mu} + \boldsymbol{b}\boldsymbol{\mu} + a = 5. \tag{3.164}$$

Multiplying [3.161] out:

$$Y = 4\dot{X}_1^2 + 3\dot{X}_2^2 + 12\dot{X}_2 + 6\dot{X}_3 + 5. \tag{3.165}$$

We complete the squares for terms involving $\dot{X}_2$ to obtain

$$Y = 4\dot{X}_1^2 + 3(\dot{X}_2 + 2)^2 + 6\dot{X}_3 - 7. \tag{3.166}$$

We have expressed Y as a linear polynomial of three independent random variables:

- $\dot{X}_1^2 \sim \chi^2(1, 0)$,
- $(\dot{X}_2 + 2)^2 \sim \chi^2(1, 4)$,
- $\dot{X}_3 \sim N(0, 1)$.

Moments

We have seen that a random variable Y that is a quadratic polynomial of a random vector $\boldsymbol{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be expressed as a linear polynomial of independent chi-squared and normal random variables. Based upon this representation, we may apply [3.129] to obtain the MGF of Y. From this, we can calculate the moments of Y. The details of the derivation are covered by Mathai and Provost (1992). Results, based upon notation introduced earlier in this section, are as follows.

Define, for positive integers k,

$$g^{[k]} = \begin{cases} \boldsymbol{\mu}'\boldsymbol{c}\boldsymbol{\mu} + \boldsymbol{b}\boldsymbol{\mu} + a + \sum_{j=1}^{m} \dot{c}_{j,j} & k = 0, \\ \dfrac{(k+1)!}{2} \sum_{j=1}^{m} \dot{b}_j^2 (2\dot{c}_{j,j})^{k-1} + \dfrac{k!}{2} \sum_{j=1}^{m} (2\dot{c}_{j,j})^{k+1} & k > 0, \end{cases} \qquad [3.167]$$

where any undefined term 0^0 is set equal to 0. The r^{th} moment of Y is given by[14]

$$\begin{aligned} E(Y^r) = \sum_{r_1=0}^{r-1} \Bigg[\binom{r-1}{r_1} g^{[r-1-r_1]} \sum_{r_2=0}^{r_1-1} \Bigg[\binom{r_1-1}{r_2} g^{[r_1-1-r_2]} \\ \times \sum_{r_3=0}^{r_2-1} \Bigg[\binom{r_2-1}{r_3} g^{[r_2-1-r_3]} \cdots \Bigg]\Bigg]\Bigg], \end{aligned} \qquad [3.168]$$

where any empty product is interpreted as equaling 1. Based upon [3.168],

$$E(Y) = g^{[0]}, \qquad [3.169]$$

$$E(Y^2) = g^{[1]} + \binom{1}{1} g^{[0]} E(Y), \qquad [3.170]$$

$$E(Y^3) = g^{[2]} + \binom{2}{1} g^{[1]} E(Y) + \binom{2}{2} g^{[0]} E(Y^2), \qquad [3.171]$$

$$E(Y^4) = g^{[3]} + \binom{3}{1} g^{[2]} E(Y) + \binom{3}{2} g^{[1]} E(Y^2) + \binom{3}{3} g^{[0]} E(Y^3), \qquad [3.172]$$

$$\begin{aligned} E(Y^5) = g^{[4]} + \binom{4}{1} g^{[3]} E(Y) + \binom{4}{2} g^{[2]} E(Y^2) \\ + \binom{4}{3} g^{[1]} E(Y^3) + \binom{4}{4} g^{[0]} E(Y^4), \end{aligned} \qquad [3.173]$$

and so forth according to a similar pattern.

EXAMPLE. Consider the random variable Y defined by [3.154] in our last example. Let's calculate its first five moments. Based upon results from that example, we calculate values for $g^{[k]}$ as indicated in Exhibit 3.24. Moments of Y are calculated from these by [3.169] through [3.173]. Results are indicated in Exhibit 3.25.

[14] We employ the notation $\binom{a}{b} = \frac{a!}{b!(a-b)!}$.

k	$g^{[k]}$
0	12
1	230
2	3,320
3	78,384
4	2,352,768

Exhibit 3.24 Using formula [3.167], $g^{[k]}$ values are calculated for the random variable Y defined by [3.154]. Inputs for the calculations are obtained from [3.153], [3.155], [3.156], [3.157], [3.162], and [3.163].

r	$E(Y^r)$
1	12
2	374
3	13,328
4	615,900
5	33,217,840

Exhibit 3.25 Moments of portfolio value are indicated for the random variable Y defined by [3.154]. Computations are performed according to [3.169] through [3.173]. Inputs are the values of Exhibit 3.24.

Other Parameters

We can calculate any central moment of Y. This is simply a matter of multiplying out the formula for the desired central moment and substituting in values for moments. Consider the third central moment:

$$\mu_3 = E[(Y-\mu)^3] \quad [3.174]$$
$$= E[Y^3 - 3Y^2\mu + 3Y\mu^2 - \mu^3] \quad [3.175]$$
$$= E(Y^3) - 3E(Y^2)\mu + 3E(Y)\mu^2 - \mu^3 \quad [3.176]$$
$$= E(Y^3) - 3E(Y^2)\mu + 2\mu^3. \quad [3.177]$$

where $\mu = E(Y)$. The variance of Y is, by our result from Exercise 3.15,

$$\sigma^2 = E(Y^2) - E(Y)^2. \quad [3.178]$$

The skewness η_1 and kurtosis η_2 are obtained as

$$\eta_1 = \frac{E[(Y-\mu)^3]}{\sigma^3} = \frac{E(Y^3) - 3\mu E(Y^2) + 3\mu^2 E(Y) - \mu^3}{\sigma^3}, \quad [3.179]$$

$$\eta_2 = \frac{E[(Y-\mu)^4]}{\sigma^4} = \frac{E(Y^4) - 4\mu E(Y^3) + 6\mu^2 E(Y^2) - 4\mu^3 E(Y) + \mu^4}{\sigma^4}. \quad [3.180]$$

Quantiles of Y can be approximated using the Cornish-Fisher (1937) expansion, which we discuss in the next section. They can be calculated exactly using the inversion theorem that we discuss in Section 3.15.

Exercises

3.37 🖩 Consider random vector $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 5 & 0 & 2 \\ 0 & 1 & 2 \\ 2 & 2 & 5 \end{pmatrix}. \qquad [3.181]$$

Let

$$Y = \boldsymbol{X}'\boldsymbol{c}\boldsymbol{X} + \boldsymbol{b}\boldsymbol{X} + a, \qquad [3.182]$$

where

$$\boldsymbol{c} = \begin{pmatrix} 2 & 8 & -4 \\ 8 & 31 & -16 \\ -4 & -16 & 8 \end{pmatrix}, \qquad [3.183]$$

$$\boldsymbol{b} = (20 \quad 90 \quad -43), \qquad [3.184]$$

$$a = 29. \qquad [3.185]$$

Express Y as a linear polynomial of independent chi-squared and normal random variables.

3.38 Calculate the mean and standard deviation of the random variable Y of the previous exercise.

3.13. THE CORNISH-FISHER EXPANSION

The **cumulants** of a random variable X are conceptually similar to its moments. They are defined, somewhat abstrusely, as those values κ_r such that the identity

$$exp\left(\sum_{r=1}^{\infty} \frac{\kappa_r t^r}{r!}\right) = \sum_{r=0}^{\infty} \frac{E(X^r)t^r}{r!} \qquad [3.186]$$

holds for all t. Cumulants of a random variable X can—see Stuart and Ord (1994)—be expressed in terms of its mean $\mu = E(X)$ and central moments $\mu_r = E[(X-\mu)^r]$. Expressions for the first five cumulants are

$$\kappa_1 = \mu, \qquad [3.187]$$

$$\kappa_2 = \mu_2, \qquad [3.188]$$

$$\kappa_3 = \mu_3, \qquad [3.189]$$

$$\kappa_4 = \mu_4 - 3\mu_2^2, \qquad [3.190]$$

$$\kappa_5 = \mu_5 - 10\mu_3\mu_2. \qquad [3.191]$$

Suppose X has mean 0 and standard deviation 1. Cornish and Fisher (1937) provide an expansion for approximating the q-quantile, $\Phi_X^{-1}(q)$, of X based upon its cumulants. Using the first five cumulants, the expansion is

$$\begin{aligned}\Phi_X^{-1}(q) \approx\ & \Phi_Z^{-1}(q) + \frac{\Phi_Z^{-1}(q)^2 - 1}{6}\kappa_3 + \frac{\Phi_Z^{-1}(q)^3 - 3\Phi_Z^{-1}(q)}{24}\kappa_4 \\ & - \frac{2\Phi_Z^{-1}(q)^3 - 5\Phi_Z^{-1}(q)}{36}\kappa_3^2 + \frac{\Phi_Z^{-1}(q)^4 - 6\Phi_Z^{-1}(q)^2 + 3}{120}\kappa_5 \\ & - \frac{\Phi_Z^{-1}(q)^4 - 5\Phi_Z^{-1}(q)^2 + 2}{24}\kappa_3\kappa_4 + \frac{12\Phi_Z^{-1}(q)^4 - 53\Phi_Z^{-1}(q)^2 + 17}{324}\kappa_3^3,\end{aligned} \quad [3.192]$$

where $\Phi_Z^{-1}(q)$ is the q-quantile of $Z \sim N(0, 1)$. Although [3.192] applies only if X has mean 0 and standard deviation 1, we can still use it to approximate quantiles if X has some other mean μ and standard deviation σ. Simply define the **normalization** of X as

$$X^* = \frac{X - \mu}{\sigma}, \quad [3.193]$$

which has mean 0 and standard deviation 1. Central moments of X^* can be calculated from central moments of X with

$$\mu_r^* = \frac{\mu_r}{\sigma^r}, \quad [3.194]$$

where $\sigma = \sqrt{\mu_2}$ is the standard deviation of X. Apply the Cornish-Fisher expansion to obtain the q-quantile x^* of X^*. The corresponding q-quantile x of X is then

$$x = x^*\sigma + \mu. \quad [3.195]$$

EXAMPLE. Let's use the Cornish-Fisher expansion to approximate the .10-quantile of the random variable Y defined by [3.154] in our earlier example. From the first five moments of Y provided in Exhibit 3.25, we calculate the central moments of Y and the central moments and cumulants of the normalization Y^* of Y. Results are indicated in Exhibit 3.26.

r	μ_r	μ_r^*	κ_r^*
1	0	0.00000	0.00000
2	230	1.00000	1.00000
3	3,320	0.95180	0.95180
4	237,084	4.48174	1.48174
5	9,988,768	12.45066	2.93265

Exhibit 3.26 Central moments of Y, central moments of Y^*, and cumulants of Y^* are calculated from formulas [3.11], [3.194], and [3.187] through [3.191].

Applying the Cornish-Fisher expansion [3.192] yields the .10-quantile of Y^* as –1.123. Applying [3.195], we obtain the .10-quantile of Y as -5.029.

Exercises

3.39 Using a spreadsheet and inputs from Exhibit 3.26, reproduce the results from the example of this section.

3.14. CENTRAL LIMIT THEOREM

The normal distribution is useful for modeling various random quantities, such as people's heights, asset returns, and test scores. This is no coincidence. If a process is additive—reflecting the combined influence of multiple random occurrences—the result is likely to be approximately normal. This follows from the **central limit theorem**.

Let $\boldsymbol{X}$ be an n-dimensional random vector with independent and identically distributed (IID) components X_i. It doesn't matter what their common distribution is as long as its mean μ and standard deviation σ exist. Let $\overline{X}_n$ be the random variable equal to the average of the X_i. By [3.27] and [3.28], $\overline{X}_n$ has mean μ and standard deviation $\sigma/\sqrt{n}$. Accordingly, the normalized average

$$\overline{X}_n^* = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \qquad [3.196]$$

has mean 0 and standard deviation 1. The central limit theorem tells us $\overline{X}_n^*$ is approximately $N(0, 1)$. Specifically, it states that, for any constant x,

$$\lim_{n\to\infty} Pr\left(\overline{X}_n^* \leq x\right) = \Phi(x), \qquad [3.197]$$

where $\Phi(x)$ is the CDF of the standard normal distribution.

Exhibit 3.27 illustrates the PDF for the profit and loss (P&L) that will be realized by purchasing and holding for 1 month EUR 30,000 of a particular at-the-money 3-month call option on Euribor futures. The limited downside risk of the options strategy is evident in the skewed P&L distribution. It has skewness 0.96 and kurtosis 3.90.

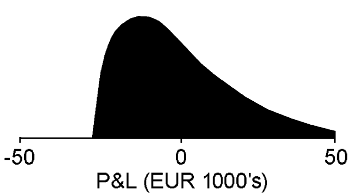

Exhibit 3.27 PDF for the P&L that will be realized by purchasing EUR 30,000 of at-the-money 3-month call options on Euribor futures and holding the position for 1 month. The PDF is based upon market conditions on May 1, 2000.

Suppose random changes in Euribor are independent from one month to the next. We repeat our options strategy every month for 18 months. At the start of every month, we purchase at-the-money 3-month options and liquidate them at the end of the month. Repeating this process for 18 consecutive months yields a total P&L for the 18 months whose PDF is graphed in Exhibit 3.28.[15]

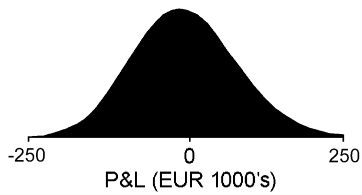

Exhibit 3.28 PDF for the P&L that will be realized by rolling 3-month call options monthly for 18 months.

The P&L distribution for the 18-month strategy is not skewed like that of the 1-month strategy. It does not afford the same protection against downside risk. With skewness of just 0.23, kurtosis of 3.05, and a familiar "bell" shape, it is almost normal. Rolling options for 18 months offers a P&L distribution that is little different from that which could be obtained by just holding the underlying futures.

Our example illustrates the central limit theorem. With the 1-month strategy, we randomly draw a P&L from the probability distribution of Exhibit 3.27. With our 18-month strategy, we independently draw from that distribution 18 times. The 18-month P&L is the sum of these.

There are many versions of the central limit theorem.[16] Several of these place additional restrictions on the X_i but do not require that they be identically distributed. The additional restrictions vary, but are generally designed to prevent one or a handful of random variables from dominating the average, which might happen if one random variable has a standard deviation far greater than the rest.

In Exhibit 3.29, probability distributions are illustrated for five independent random variables X_i. All five distributions have mean 0 and standard deviation 1 and are dramatically non-normal. They were selected arbitrarily, but their normalized average $\overline{X}_5^*$ is approximately normal.

[15]This and the analysis of Exhibit 3.27 were performed with the Monte Carlo method, which we describe in Chapter 5.

[16]See Spanos (1999) for a detailed discussion including historical notes.

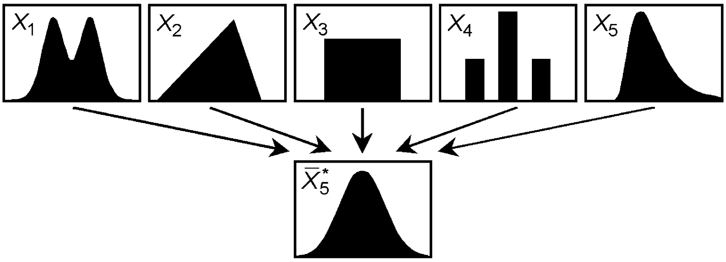

Exhibit 3.29 The central limit theorem is illustrated in the case of five arbitrarily selected independent random variables. Random variables X_1, X_2, X_3, and X_5 are continuous, so their PDFs are shown; X_4 is discrete, so its PF is shown. The normalized average $\overline{X}_5^*$ is approximately $N(0, 1)$. All graphs indicate the interval $[-3, 3]$ on the x-axis.

Exhibit 3.30 provides summary information on the distributions of Exhibit 3.29.

Random Variable	Mean	Standard Deviation	Skewness	Kurtosis	Description
X_1	0.00	1.00	0.00	1.89	continuous
X_2	0.00	1.00	-0.41	2.41	continuous
X_3	0.00	1.00	0.00	1.80	continuous
X_4	0.00	1.00	0.00	2.00	discrete
X_5	0.00	1.00	1.62	7.89	continuous
$\overline{X}_5^*$	0.00	1.00	0.11	3.03	continuous

Exhibit 3.30 Summary information on the distributions of Exhibit 3.29. All distributions were constructed with mean 0 and standard deviation 1. This was done primarily to standardize the graphs in Exhibit 3.29, but having uniform standard deviations does aid convergence to a normal distribution.

Other versions of the central limit theorem modestly weaken the independence assumption for the X_i. The central limit theorem generalizes to multiple dimensions.

Exercises

3.40 Let $\overline{X}_n^*$ be the normalized average of n independent $U(-1, 1)$ random variables. Based upon your intuition:

a. How large do you think n must be for the PDF of $\overline{X}_n^*$ to have the same general shape as the PDF of a standard normal distribution?

b. How large do you think n must be for the kurtosis of $\overline{X}_n^*$ to match the 3.0 kurtosis of a normal distribution to one decimal place?

3.41 Suppose Y equals a sum of 20 independent random variables. One is $U(-10, 10)$. The rest are $U(-1, 1)$. Is Y approximately normal?

3.42 To gain insight into the last exercise, use your intuition to sketch the PDFs of:
   a. a $U(-10, 10)$ random variable;
   b. a sum of a $U(-10, 10)$ and 1 independent $U(-1, 1)$ random variables;
   c. a sum of a $U(-10, 10)$ and 2 independent $U(-1, 1)$ random variables;
   d. a sum of a $U(-10, 10)$ and 3 independent $U(-1, 1)$ random variables.

3.15. THE INVERSION THEOREM

Much of this chapter has been devoted to studying linear polynomials of random vectors. Results have included:

- formulas [3.30] and [3.31] for calculating the means and covariances of linear polynomials of random vectors;
- the use of moment-generating functions to calculate moments of linear polynomials of independent random variables;
- the definition that a linear polynomial of a joint-normal random vector is normal;
- the fact that a quadratic polynomial of a joint-normal random vector can be expressed as a linear polynomial of independent chi-squared and normal random variables;
- the central limit theorem describing certain linear polynomials of random variables as being approximately normal.

In this section, we present an inversion theorem, which is primarily of theoretical interest. We shall use it for the practical purpose of evaluating the CDF of a linear polynomial of independent random variables.

Imaginary Random Variables

To define characteristic functions, we must extend the notion of random variables into the complex plane. Let U_1 and U_2 be real random variables, and let $i = \sqrt{-1}$. Then

$$U = U_1 + iU_2 \qquad [3.198]$$

is a **complex random variable**. We define its expectation as

$$E(U) = E(U_1) + iE(U_2). \qquad [3.199]$$

Characteristic Functions

Characteristic functions are similar to MGFs. We define the **characteristic function** of a random variable X as

$$\Psi(w) = E(e^{iwX}), \quad [3.200]$$

where w is real and $i = \sqrt{-1}$. If X is continuous,

$$\Psi(w) = E(e^{iwX}) = \int_{-\infty}^{\infty} \phi(x) e^{iwx} dx. \quad [3.201]$$

If $\boldsymbol{X}$ is a random vector with independent components X_i, and Y is a linear polynomial of $\boldsymbol{X}$

$$Y = \boldsymbol{bX} + a \quad [3.202]$$

with $\boldsymbol{b}$ and a a real row vector and scalar, then, analogous to [3.129] for MGFs,

$$\Psi_Y(w) = e^{aiw}\Psi_{X_1}(b_1 w)\Psi_{X_2}(b_2 w) \cdots \Psi_{X_n}(b_n w). \quad [3.203]$$

A uniform, $U(a, b)$, random variable has characteristic function

$$\Psi(w) = \frac{e^{ibw} - e^{iaw}}{iw(b-a)}. \quad [3.204]$$

Characteristic functions for $N(\mu, \sigma^2)$ and $\chi^2(\nu, \delta^2)$ random variables are, respectively,

$$\Psi(w) = exp\left(i\mu w - \frac{\sigma^2 w^2}{2}\right), \quad [3.205]$$

$$\Psi(w) = \frac{exp[\delta^2 iw/(1-2iw)]}{(1-2iw)^{\nu/2}}. \quad [3.206]$$

The characteristic function for a lognormal random variable is derived by Leipnik (1991). It is complicated, so we do not present it here.

Inversion Theorem

The CDF of a random variable is uniquely determined by its characteristic function. If two random variables have the same characteristic function, they have the same CDF. An **inversion theorem** provides the CDF of a random variable X in terms of its characteristic function:

$$\Phi(x) = \frac{1}{2} + \frac{1}{2\pi}\int_{0}^{\infty} \frac{\Psi(-w)e^{ixw} - \Psi(w)e^{-ixw}}{iw} dw. \quad [3.207]$$

Exercises

3.43 Determine the characteristic function for the following random variables:

a. $X \sim N(1, 4)$;

b. $Y = 3Q + R + 5$, where $Q \sim U(0, 1)$ and $R \sim \chi^2(2, 1)$ are independent;

c. $Z = X_1^2 + X_2^2$, where $X_1 \sim X_2 \sim N(0, 1)$ are independent.

3.44 Use the characteristic function [3.205] of the normal distribution and [3.203] to prove that, if $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then

$$X_1 + X_2 \sim N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right). \quad [3.208]$$

3.16. QUANTILES OF QUADRATIC POLYNOMIALS OF JOINT-NORMAL RANDOM VECTORS

Consider a random variable Y that is a quadratic polynomial of a joint-normal random vector $\boldsymbol{X}$. We can approximate its quantiles using the Cornish-Fisher expansion. Alternatively, if exact quantiles are required, we may employ the inversion theorem in a manner described by Imhof (1961) and Davies (1973). This provides the CDF of Y. From this, we can calculate quantiles.

The CDF of a Quadratic Polynomial of a Joint-Normal Random Vector

As described in Section 3.12, express Y as a linear polynomial of independent random variables,

$$Y = \left(\sum_{k=1}^{m} \gamma_k Q_k\right) + \beta Q_0 + \alpha, \quad [3.209]$$

where $Q_0 \sim N(0, 1)$ and $Q_k \sim \chi^2(1, \delta_k^2)$ for $k > 0$. Based upon this representation, the characteristic function of Y is calculated from [3.203], [3.205], and [3.206] as

$$\Psi(w) = \frac{exp\left(iw\alpha - \dfrac{w^2\beta^2}{2} + iw\displaystyle\sum_{k=1}^{m} \frac{\gamma_k \delta_k^2}{1 - 2iw\gamma_k}\right)}{\displaystyle\prod_{k=1}^{m} \sqrt{1 - 2iw\gamma_k}}. \quad [3.210]$$

Inversion theorem [3.207] provides an expression for the CDF of Y in terms of this characteristic function

$$\Phi(y) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{\Psi(-w)e^{iyw} - \Psi(w)e^{-iyw}}{iw}dw, \tag{3.211}$$

but it involves an integral that is not amenable to standard techniques of numerical integration, such as the trapezoidal rule or Simpson's rule. Employing the algebra of complex numbers, the theorem can be reformulated as

$$\Phi(y) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{Im(e^{-iwy}\Psi(w))}{w}dw. \tag{3.212}$$

Substituting characteristic function [3.210] into [3.212] and simplifying yields

$$\Phi(y) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{e^A sin(B+C)}{D}dw, \tag{3.213}$$

where

$$A = -\frac{w^2}{2}\left(\beta^2 + 4\sum_{k=1}^{m}\frac{\gamma_k^2\delta_k^2}{1+4\gamma_k^2w^2}\right), \tag{3.214}$$

$$B = w\left(\alpha - y + \sum_{k=1}^{m}\frac{\gamma_k\delta_k^2}{1+4\gamma_k^2w^2}\right), \tag{3.215}$$

$$C = \frac{1}{2}\sum_{k=1}^{m} tan^{-1}(2\gamma_k w), \tag{3.216}$$

$$D = w\left(\prod_{k=1}^{m}\left(1+4\gamma_k^2w^2\right)\right)^{1/4}, \tag{3.217}$$

and tan^{-1} denotes the inverse tangent function with output in radians. The integral in [3.213] appears cumbersome, but it entails no imaginary numbers and its integrand is easily evaluated by a computer. To solve the integral numerically, two problems must be addressed:

- The integrand has form $0/0$ as w approaches 0.
- The interval of integration is unbounded.

To solve the first problem, we apply l'Hôpital's rule to obtain[17]

$$\lim_{w\to 0}\frac{e^A sin(B+C)}{D} = \alpha - y + \sum_{k=1}^{m}\gamma_k\left(\delta_k^2 + 1\right). \tag{3.218}$$

[17] I am indebted to Arcady Novosyolov for this simplification.

We solve the second problem with the approximation

$$\int_{0}^{\infty} \frac{e^{A} sin(B+C)}{D} dw \approx \int_{0}^{u} \frac{e^{A} sin(B+C)}{D} dw, \qquad [3.219]$$

where $u < \infty$ is chosen sufficiently large. Valuing this integral is one instance where the trapezoidal rule provides consistently superior results to Simpson's rule. By selecting an appropriate value for u, we can make the error in approximation [3.219] arbitrarily small. The solution is essentially exact.

Quantiles of a Quadratic Polynomial of a Joint-Normal Random Vector

Since we can evaluate the CDF $\Phi(y)$ of Y, we can now calculate any q-quantile of Y. Consider a specific value q. We seek that value y such that $\Phi(y) = q$. We might find this by evaluating Φ at a range of values for y and finding which one yields a probability closest to q. A faster and more systematic approach is to use the secant method of Section 2.12.

The secant method requires two seed values $y^{[1]}$ and $y^{[2]}$. Subsequent values $y^{[3]}$, $y^{[4]}$, $y^{[5]}$, ... are obtained with the recursive equation

$$y^{[i]} = y^{[i-1]} - \frac{\left[\Phi\left(y^{[i-1]}\right) - q\right]\left(y^{[i-1]} - y^{[i-2]}\right)}{\Phi\left(y^{[i-1]}\right) - \Phi\left(y^{[i-2]}\right)}. \qquad [3.220]$$

The resulting sequence of values should converge to the q-quantile y of Y.

Example. Consider again the random variable Y that is a quadratic polynomial of a joint-normal random vector $\boldsymbol{X}$ as defined by [3.154]. We have considered this random variable in several examples. Let's find its .10-quantile.

Based upon representation [3.166] and formula [3.213], we can use the trapezoidal rule to evaluate $\Phi(y)$ at any point y. For this purpose, we use approximation [3.219] with $u = 1$. We partition $[0, u]$ into 500 subintervals to apply the trapezoidal rule. Consider seed values $y^{[1]} = 0$ and $y^{[2]} = 1$. Applying the trapezoidal rule to each, we obtain

- $\Phi(0) = 0.21752$,
- $\Phi(1) = 0.24546$.

Applying the secant method, we obtain the results indicated in Exhibit 3.31.

The .10-quantile of Y is -5.004, which is exact to the number of decimal places shown. Compare this with the -5.029 approximation we obtained for the same quantile using the Cornish-Fisher expansion in Section 3.13.

i	$y^{[i]}$	$\Phi(y^{[i]})$
1	0.000	0.21752
2	1.000	0.24546
3	−4.207	0.11565
4	−4.835	0.10322
5	−4.997	0.10013
6	−5.004	0.10000

Exhibit 3.31 Results of applying the secant method to evaluate the .10-quantile of Y. Each iteration requires the evaluation of Φ using the trapezoidal rule.

Exercises

3.45 🖩 Independently reproduce the results of the example of this section.

3.17. FURTHER READING

DeGroot (1986) covers basic probability theory. Feller (1968, 1971) is a more advanced treatment. Spanos (1999) is a formal text with a detailed treatment of the central limit theorem. See Johnson (1998) for an elementary discussion of principal component analysis. Multicollinearity is discussed in econometrics texts, typically in relation to regression analysis. Judge *et al.* (1988) offers an alternative treatment relating to principal component analysis. See Mathai and Provost (1992) for quadratic polynomials of joint-normal random vectors. Stuart and Ord (1994) offers an authoritative if somewhat inaccessible treatment of various topics, including the Cornish-Fisher expansion, characteristic functions, and the inversion theorem.