Review of Mathematics for Economics:

- **Probability Theory**

- **Linear and Matrix Algebra**

- **Difference Equations**

Marc-André Letendre

McMaster University

Contents

1 Matrix Algebra

1.1 Some Terminology

A **matrix** is a rectangular (*i.e.* two-dimensional) array of elements (*e.g.* numbers, coefficients, variables). When stating the **dimensions of a matrix**, the convention is to indicate the number of rows first, and then the number of columns. For example, if a matrix is said to be 3 by 5 (or 3×5), then it has three rows and 5 columns for a total of 15 elements. Here is an example of a 3×5 matrix of numbers

$$\begin{bmatrix} 9 & 5.1 & 4.3 & 1 & 200.7 \\ 3.8 & 90 & 76.3 & 22.45 & 100 \\ 6 & 85.2 & 5 & 99 & 2 \end{bmatrix}.$$

A **vector** is a one-dimensional array of elements. Simply put, a vector is a matrix with a single column (called a column-vector) or a single row (called a row-vector). We refer to a vector with n elements as an n-vector.

The notational convention I follow in these notes is the following: lowercase boldface type for vectors (*e.g.* $\boldsymbol{y}$) and uppercase boldface type for matrices (*e.g.* $\boldsymbol{X}$).

1.2 Some Operations on Matrices
1.2.1 Addition and Subtraction

Two matrices can be added together if and only if they have the same dimensions. When this dimensional requirement is met, the matrices are said to be **conformable for addition**. For example, only a 2×3 matrix can be added to (subtracted from) a 2×3 matrix. The addition of two matrices is defined as the addition of each pair of corresponding elements. For example, consider two 3×2 matrices $\boldsymbol{A}$ and $\boldsymbol{B}$

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \\ a_{31} + b_{31} & a_{32} + b_{32} \end{bmatrix}. \tag{1}$$

A numerical example is

$$\begin{bmatrix} 1 & 3 \\ 8 & 7 \end{bmatrix} + \begin{bmatrix} 2 & 4 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 7 \\ 9 & 13 \end{bmatrix}. \tag{2}$$

Similarly, for the subtraction of two matrices we have

$$\boldsymbol{A} - \boldsymbol{B} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} \\ a_{21} - b_{21} & a_{22} - b_{22} \\ a_{31} - b_{31} & a_{32} - b_{32} \end{bmatrix}. \tag{3}$$

A numerical example is

$$\begin{bmatrix} 1 & 3 \\ 8 & 7 \end{bmatrix} - \begin{bmatrix} 2 & 4 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 7 & 1 \end{bmatrix}. \tag{4}$$

Note that matrix addition (and subtraction) is commutative

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A} \tag{5}$$

and associative

$$(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) \tag{6}$$

1.2.2 Multiplication

The **multiplication of a matrix by a scalar** is straightforward. We simply multiply every element of the matrix by the given scalar. For example, consider the product of the scalar a with the 3×3 matrix $\boldsymbol{B}$

$$a\boldsymbol{B} = a \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a\, b_{11} & a\, b_{12} & a\, b_{13} \\ a\, b_{21} & a\, b_{22} & a\, b_{23} \\ a\, b_{31} & a\, b_{32} & a\, b_{33} \end{bmatrix}. \tag{7}$$

A numerical example is

$$2 \begin{bmatrix} 6 & 3 & 12 \\ 2 & 7 & 1 \\ 8 & 5 & 9 \end{bmatrix} = \begin{bmatrix} 12 & 6 & 24 \\ 4 & 14 & 2 \\ 16 & 10 & 18 \end{bmatrix}. \tag{8}$$

The multiplication of two matrices is contingent upon the satisfaction of a dimensional requirement. The conformability condition for multiplication is that the column dimension of the first matrix *must* be equal to the row dimension of the second matrix. For example, if $\boldsymbol{A}$ is 4×3 and $\boldsymbol{B}$ is 3×5 we can find the product $\boldsymbol{AB}$ but cannot find $\boldsymbol{BA}$ since the number of columns of $\boldsymbol{B}$ (5) does not equal the number of rows of $\boldsymbol{A}$ (4).

Let matrix $\boldsymbol{A}$ be $m \times \ell$ and matrix $\boldsymbol{B}$ be $\ell \times n$. Then a matrix $\boldsymbol{C}$ defined as $\boldsymbol{C} = \boldsymbol{AB}$ will be $m \times n$. The typical element c_{ij} of matrix $\boldsymbol{C}$ is calculated as follows:

$$c_{ij} = \sum_{h=1}^{\ell} a_{ih} b_{hj}.$$

Let matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ be defined by

$$\boldsymbol{A} = \begin{bmatrix} 6 & 3 & 12 \\ 2 & 7 & 1 \\ 8 & 5 & 9 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 8 & 5 \end{bmatrix}. \tag{9}$$

The matrix $C = AB$ is calculated as follows

$$C = \begin{bmatrix} 6 \times 1 + 3^2 + 12 \times 8 & 6 \times 2 + 3 \times 4 + 12 \times 5 \\ 2 \times 1 + 7 \times 3 + 1 \times 8 & 2 \times 2 + 7 \times 4 + 1 \times 5 \\ 8 \times 1 + 5 \times 3 + 9 \times 8 & 8 \times 2 + 5 \times 4 + 9 \times 5 \end{bmatrix} = \begin{bmatrix} 111 & 84 \\ 31 & 37 \\ 95 & 81 \end{bmatrix} \tag{10}$$

Note that contrary to matrix addition, matrix multiplication is generally *not* commutative. In other words,

$$AB \neq BA. \tag{11}$$

As long as the matrices involved are conformable, matrix multiplication is associative

$$(AB)C = A(BC) \tag{12}$$

and distributive

$$A(B + C) = AB + AC \tag{13}$$

$$(B + C)A = BA + CA. \tag{14}$$

1.2.3   Transposes

When the rows and columns of a matrix A are interchanged we obtain the transpose of A, denoted A' or $A^\top$. Obviously, if matrix A is 6×4, its transpose A' is 4×6. Here are two numerical examples

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}' = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \end{bmatrix}' = \begin{bmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \end{bmatrix}. \tag{15}$$

Transposes have the following properties

$$(A')' = A, \qquad (A + B)' = A' + B', \qquad (AB)' = B'A'. \tag{16}$$

Exercise 1.1 – Matrix Operations

Using the matrices and vector defined below, verify that $(A + B')\, C\, d = \begin{bmatrix} 41 & 117 \end{bmatrix}'$.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & -5 & 4 & 3 \\ -8 & 7 & -6 & 5 \\ 4 & 3 & -2 & 1 \end{bmatrix}, \quad d = \begin{bmatrix} 2 \\ 1 \\ 4 \\ 4 \end{bmatrix}.$$

1.3 Writing a System of Equations Using Matrix Notation

This section shows how we translate a linear system of equations from linear algebra to matrix algebra. As an example, we write a multiple linear regression model using matrix notation.

Consider the system of equations

$$y_1 = \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_k x_{1k} + \ldots + \beta_K x_{1K} + e_1$$

$$y_2 = \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_k x_{2k} + \ldots + \beta_K x_{2K} + e_2$$

$$\vdots$$

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \ldots + \beta_K x_{tK} + e_t \tag{17}$$

$$\vdots$$

$$y_T = \beta_1 x_{T1} + \beta_2 x_{T2} + \ldots + \beta_k x_{Tk} + \ldots + \beta_K x_{TK} + e_T$$

where y_t denotes the tth observation on the dependent variable, x_{tk} denotes the tth observation on the kth explanatory variable, e_t denotes a random error term and β_1, β_2, ..., β_K denote K coefficients. We now want to write system (17) using matrix notation.

The T observations on the dependent variable y are denoted: y_1, y_2, ..., y_t, ..., y_T. These observations are "stacked" in the T-vector $\boldsymbol{y}$ as follows

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_t \\ \vdots \\ y_T \end{bmatrix}. \tag{18}$$

The observations on each of the explanatory variables are also arranged in T-vectors

$$\boldsymbol{x_1} = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{t1} \\ \vdots \\ x_{T1} \end{bmatrix}, \quad \boldsymbol{x_2} = \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \\ \vdots \\ x_{t2} \\ \vdots \\ x_{T2} \end{bmatrix}, \quad ..., \quad \boldsymbol{x_k} = \begin{bmatrix} x_{1k} \\ x_{2k} \\ x_{3k} \\ \vdots \\ x_{tk} \\ \vdots \\ x_{Tk} \end{bmatrix} ..., \quad \boldsymbol{x_K} = \begin{bmatrix} x_{1K} \\ x_{2K} \\ x_{3K} \\ \vdots \\ x_{tK} \\ \vdots \\ x_{TK} \end{bmatrix}. \tag{19}$$

We then define the matrix of regressors $\boldsymbol{X}$ which is formed by horizontally concatenating the vectors above. The matrix of regressors is a $T \times K$ matrix and it can be written as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \boldsymbol{x_k} & ... & \boldsymbol{x_K} \end{bmatrix}$$

or equivalently as

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{1k} & ... & x_{1K} \\ x_{21} & x_{22} & x_{23} & x_{2k} & ... & x_{2K} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{31} & x_{t2} & x_{t3} & x_{tk} & ... & x_{tK} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{T1} & x_{T2} & x_{T3} & x_{Tk} & ... & x_{TK} \end{bmatrix}. \tag{20}$$

Finally, the parameters of the model are grouped in the parameter vector $\boldsymbol{\beta}$ (this is a K-vector) while the error terms are grouped in the vector of error terms $\boldsymbol{e}$ (this is a T-vector)

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_k \\ \vdots \\ \beta_K \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_t \\ \vdots \\ e_T \end{bmatrix}. \tag{21}$$

Using the vectors and matrices defined in (18)-(21), we can simply write the system of T equations (17) as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}. \tag{22}$$

To see this, note that two matrices ($\boldsymbol{A}$ and $\boldsymbol{B}$ say) are equal when the (i, j) element of $\boldsymbol{A}$ equals the (i, j) element of $\boldsymbol{B}$ for all values of the indexes i and j.

Exercise 1.2 – National Income Model

Take a very simple national-income model where only Y and C are endogenous

$$Y = C + I_0 + G_0, \qquad C = a + bY. \tag{23}$$

Show that the matrix equation $\boldsymbol{Ax} = \boldsymbol{d}$, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ -b & 1 \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} Y \\ C \end{bmatrix}, \qquad \boldsymbol{d} = \begin{bmatrix} I_0 + G_0 \\ a \end{bmatrix} \tag{24}$$

is equivalent to the system of equations (23).

1.4 Special Matrices

A square matrix $\boldsymbol{A}$ is said to be **symmetric** if its elements are such that $a_{ij} = a_{ji}$ for all (i,j) combinations. For a symmetric matrix $\boldsymbol{A}$, we have $\boldsymbol{A}' = \boldsymbol{A}$. Here are two symmetric matrices

$$\begin{bmatrix} 3 & 7 \\ 7 & 2 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 8 \end{bmatrix}. \tag{25}$$

A square matrix $\boldsymbol{D}$ is **diagonal** if its elements are such that $d_{ij} = 0$ for all $i \neq j$. Here are some diagonal matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 11 \end{bmatrix}.$$

The **identity matrix** is a special case of a diagonal matrix. The elements on the main diagonal of the identity matrix are all ones and all the off-diagonal elements are zeroes. The identity matrix of order n is denoted $\mathbf{I}_n$. Here are two examples

$$\boldsymbol{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The identity matrix plays a role similar to that of unity in scalar algebra. For example, if the matrix $\boldsymbol{A}$ is $n \times n$, then $\boldsymbol{I}_n \boldsymbol{A} = \boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A}$. Also note that the identity matrix is **idempotent** (it remains unchanged when it is multiplied by itself any number of times).

Just as the identity matrix plays the role of unity, a **null matrix**, or **zero matrix**, denoted by $\boldsymbol{O}$ plays the role of number zero. Note that contrary to the identity matrix, the null matrix does not have to be a square matrix. Here are two null matrices

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{26}$$

Triangular matrices have zeros either above the main diagonal (lower-triangular) or below the main diagonal (upper-triangular).

1.5 Determinants

The determinant of a matrix is particularly useful when we want to determine whether a matrix is invertible or not. In this subsection we discuss determinants. Matrix inversion is discussed later in this chapter.

The **determinant** of a square matrix $\boldsymbol{A}$, denoted $|\boldsymbol{A}|$ is a uniquely defined scalar associated with that matrix. For example, the determinant of a 2×2 matrix $\boldsymbol{A}$ is calculated as follows

$$|\boldsymbol{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}. \tag{27}$$

Here is a numerical example

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ 5 & -7 \end{bmatrix}, \quad |\boldsymbol{A}| = \begin{vmatrix} 2 & 3 \\ 5 & -7 \end{vmatrix} = 2(-7) - 3(5) = -14 - 15 = -29. \tag{28}$$

Whereas calculating second-order determinants as in (28) is easy, the calculation of higher-order determinants can be a burdensome task. To calculate higher-order determinants, we use a method called **Laplace expansion**.[1] The determinant of a square matrix $\boldsymbol{A}$ of dimension n is

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij} |C_{ij}| \quad \text{or} \quad |\boldsymbol{A}| = \sum_{i=1}^{n} a_{ij} |C_{ij}|. \tag{29}$$

where $|C_{ij}|$ is the **cofactor** of element a_{ij}.[2] To understand what is the cofactor of element a_{ij}, we must first introduce the concept of **minor** of element a_{ij}, denoted $|M_{ij}|$. $|M_{ij}|$ is a subdeterminant of $|\mathbf{A}|$ obtained by deleting the ith row and the jth column of $|\boldsymbol{A}|$. A cofactor is simply a minor with a prescribed algebraic sign attached to it

$$|C_{ij}| \equiv (-1)^{i+j} |M_{ij}|. \tag{30}$$

Let's apply formula (29) to calculate a third-order determinant. We want to find the determinant of the 3×3 matrix

$$\boldsymbol{A} = \begin{bmatrix} 5 & 6 & 1 \\ 2 & 3 & 0 \\ 7 & -3 & 0 \end{bmatrix}. \tag{31}$$

---

[1]The explanation of the calculation of higher-order determinants is included in these notes for completeness and is of secondary importance.

[2]In equation (29), the formula on the left corresponds to a Laplace expansion by the ith row while the formula on the right corresponds to an expansion by the jth column. Since the determinant of a matrix is a unique number, it does not matter which row or column is used to perform the expansion.

An expansion by the first row yields

$$|\boldsymbol{A}| = 5(-1)^{1+1}\begin{vmatrix} 3 & 0 \\ -3 & 0 \end{vmatrix} + 6(-1)^{1+2}\begin{vmatrix} 2 & 0 \\ 7 & 0 \end{vmatrix} + 1(-1)^{1+3}\begin{vmatrix} 2 & 3 \\ 7 & -3 \end{vmatrix} = -27. \tag{32}$$

Note that the determinant of a diagonal matrix is simply the product of all the elements on the main diagonal of the matrix. This implies that (i) the identity matrix always has a determinant equal to unity and, (ii) any diagonal matrix with a zero on the main diagonal has a determinant equal to zero. We conclude this subsection by stating six properties of determinants.

1. The interchange of rows and columns does not affect the value of a determinant. In other words, $|\boldsymbol{A}| = |\boldsymbol{A}'|$.

2. The interchange of any two rows (or any two columns) will alter the sign, but not the numerical value, of the determinant.

3. The multiplication of any one row (or one column) by a scalar k will change the value of the determinant k-fold.

4. The addition (subtraction) of a multiple of any row to (from) another row will leave the value of the determinant unaltered. The same holds true if we replace "row" by "column" in the previous statement.

5. If one row (or column) is a multiple of another row (or column), the value of the determinant will be zero. As a special case of this, when two rows (or columns) are identical, the determinant will be zero.

6. The expansion of a determinant by **alien cofactors** (the cofactors of a "wrong" row or column) always yields a value of zero.

1.6   Rank of a Matrix

The rank of a matrix $\boldsymbol{A}$ indicates the maximum number of linearly independent columns (or rows) in $\boldsymbol{A}$. If a matrix $\boldsymbol{A}$ is $m \times n$, then the rank of $\boldsymbol{A}$ can be at most m or n. Put differently, if we let $r(\boldsymbol{A})$ denote the rank of matrix $\boldsymbol{A}$, then

$$r(\boldsymbol{A}) \le min[m, n]. \tag{33}$$

Note that matrix $\boldsymbol{A}$ is said to be of **full rank** if $r(\boldsymbol{A}) = min[m, n]$.

For the product of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we have

$$r(\boldsymbol{AB}) \leq \min\left[r(\boldsymbol{A}), r(\boldsymbol{B})\right]. \tag{34}$$

As an important practical application, recall the matrix of regressors $\boldsymbol{X}$ that was defined in the discussion leading to (22). This matrix is $T \times K$ where we restrict $T > K$. Therefore, this matrix is full rank when $r(\boldsymbol{X}) = K$ meaning that all regressors (columns) are linearly independent. Another important property of the rank of a matrix is that $r(\boldsymbol{X}'\boldsymbol{X}) = r(\boldsymbol{X})$. Noting that $\boldsymbol{X}'\boldsymbol{X}$ is $K \times K$, then we have that $\boldsymbol{X}'\boldsymbol{X}$ is full rank when $\boldsymbol{X}$ is full rank.

To calculate the rank of a matrix we have to find out the number of linearly independent rows and columns. Consider the following matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 \\ 3 & 2 \\ 5 & 7 \end{bmatrix}. \tag{35}$$

Since it has three rows and two columns, we know that it has rank no greater than 2. By inspection, we see that the rows of $\boldsymbol{A}$ are not linearly independent (row 3=row 1+row 2) whereas the two columns of $\boldsymbol{A}$ are. Therefore, $r(\boldsymbol{A}) = 2$ in this example.

Another example is

$$\boldsymbol{B} = \begin{bmatrix} 2 & 3 \\ 4 & 6 \\ 14 & 21 \end{bmatrix}. \tag{36}$$

Again we know that the rank is at most 2. Here we have only one independent row as row 2=2*row 1 and row 3=row 1+3*row 2 and we have only one independent column as column 2=1.5*column 1. So $r(\boldsymbol{B}) = 1$.

A legitimate question at this point is how do we proceed more generally to calculate the rank of a matrix? Suppose you are given a matrix $\boldsymbol{A}$ which is $m \times n$ where $n < m$. The strategy to follow is to find the largest square submatrix of $\boldsymbol{A}$ which has linearly independent rows and columns.[3] If that submatrix is $p \times p$ then $r(\boldsymbol{A}) = p$. For example, if $\boldsymbol{A}$ is 5×4, then start by looking at 4×4 submatrices of $\boldsymbol{A}$ to see if there is one that has linearly independent columns and rows. If there is one then $r(\boldsymbol{A}) = 4$ otherwise look at 3×3 submatrices of $\boldsymbol{A}$.

---

[3]Submatrices are constructed by omitting some of the rows and columns of the matrix.

If you find a 3×3 submatrix that has linearly independent columns and rows then $r(\boldsymbol{A}) = 3$ otherwise look at 2×2 submatrices of $\boldsymbol{A}$ and so on.

If you are used to working with determinants, then you can calculate the rank of a matrix by using the property that the rank of a matrix $\boldsymbol{A}$ is the dimension of the largest square submatrix of $\boldsymbol{A}$ whose determinant is not equal to zero.

1.7   Matrix Inversion

The inverse of matrix $\boldsymbol{A}$, denoted $\boldsymbol{A}^{-1}$, is defined only if $\boldsymbol{A}$ is a square matrix, in which case the inverse is the matrix that satisfies the conditions

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}. \tag{37}$$

Not all square matrices are invertible. In the case where $\boldsymbol{A}$ is a square matrix ($n \times n$ say), $\boldsymbol{A}$ is invertible if and only if it has full rank (that is $r(\boldsymbol{A}) = n$). A square matrix that is not full rank is said to be **singular** and hence, is not invertible.[4]

Here are a few properties of inverses: $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$; $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$; $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix.

The inverse of an invertible matrix $\boldsymbol{A}$ is denoted $\boldsymbol{A}^{-1}$ and can be calculated as

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|} \operatorname{adj}(\boldsymbol{A}) \tag{38}$$

where $\operatorname{adj}(\boldsymbol{A})$ denotes the adjoint of $\boldsymbol{A}$. Note that the adjoint of $\boldsymbol{A}$ is the transpose of the matrix of cofactors of $\boldsymbol{A}$ ($\operatorname{adj}(\boldsymbol{A})=[\operatorname{cof}(\boldsymbol{A})]'$). For the special case where the matrix to invert is 2×2, then the previous formula reduces to

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = (ad - bc)^{-1} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{39}$$

The **inverse of a diagonal matrix** is trivial to calculate. We simply have to invert each of the elements on the main diagonal. For example, for a 3×3 diagonal matrix $\boldsymbol{D}$ we have

$$\begin{bmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & 0 \\ 0 & 0 & d_{33} \end{bmatrix}^{-1} = \begin{bmatrix} 1/d_{11} & 0 & 0 \\ 0 & 1/d_{22} & 0 \\ 0 & 0 & 1/d_{33} \end{bmatrix}. \tag{40}$$

---

[4]In terms of determinants. A matrix with a determinant equal to zero is said to be singular and hence, is not invertible. However, a matrix with a determinant different from zero is nonsingular and is invertible.

Let's work through three numerical examples. First, we apply formula (39) to calculate the inverse of a 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 6 & 5 \end{bmatrix}, \qquad \boldsymbol{A}^{-1} = \frac{1}{5-12} \begin{bmatrix} 5 & -2 \\ -6 & 1 \end{bmatrix} = \begin{bmatrix} -0.7143 & 0.2857 \\ 0.8571 & -0.1429 \end{bmatrix}, \qquad (41)$$

Then we apply formula (40) to calculate the inverse of a 3×3 diagonal matrix

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 8 \end{bmatrix}. \qquad (42)$$

The determinant of this matrix is $24 \neq 0$. Therefore, the inverse of $\boldsymbol{D}$ exists and is given by

$$\boldsymbol{D}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/8 \end{bmatrix}. \qquad (43)$$

Then we apply formula (38) to calculate the inverse of a 3×3 matrix

$$\boldsymbol{B} = \begin{bmatrix} 5 & 6 & 1 \\ 2 & 3 & 0 \\ 7 & -3 & 0 \end{bmatrix}. \qquad (44)$$

The determinant of this matrix is $-27 \neq 0$ as calculated in subsection 1.5 Therefore, the inverse of $\boldsymbol{B}$ exists. The matrix of cofactor of $\boldsymbol{B}$ is

$$\mathrm{cof}(\boldsymbol{B}) = \begin{bmatrix} \begin{vmatrix} 3 & 0 \\ -3 & 0 \end{vmatrix} & -\begin{vmatrix} 2 & 0 \\ 7 & 0 \end{vmatrix} & \begin{vmatrix} 2 & 3 \\ 7 & -3 \end{vmatrix} \\ -\begin{vmatrix} 6 & 1 \\ -3 & 0 \end{vmatrix} & \begin{vmatrix} 5 & 1 \\ 7 & 0 \end{vmatrix} & -\begin{vmatrix} 5 & 6 \\ 7 & -3 \end{vmatrix} \\ \begin{vmatrix} 6 & 1 \\ 3 & 0 \end{vmatrix} & -\begin{vmatrix} 5 & 1 \\ 2 & 0 \end{vmatrix} & \begin{vmatrix} 5 & 6 \\ 2 & 3 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -27 \\ -3 & -7 & 57 \\ -3 & 2 & 3 \end{bmatrix}. \qquad (45)$$

By transposing $\mathrm{cof}(\boldsymbol{B})$ we obtain the adjoint of $\boldsymbol{B}$

$$\mathrm{adj}(\boldsymbol{B}) = \mathrm{cof}(\boldsymbol{B})' = \begin{bmatrix} 0 & -3 & -3 \\ 0 & -7 & 2 \\ -27 & 57 & 3 \end{bmatrix}. \qquad (46)$$

Finally, the inverse of $\boldsymbol{B}$ is

$$\boldsymbol{B}^{-1} = \frac{1}{-27} \begin{bmatrix} 0 & -3 & -3 \\ 0 & -7 & 2 \\ -27 & 57 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0.1111 & 0.1111 \\ 0 & 0.2593 & -0.0741 \\ 1 & -2.1111 & -0.1111 \end{bmatrix}. \qquad (47)$$

1.8   Additional Exercises

Exercise 1.3

Given the matrices and vector below, compute $\boldsymbol{A}'$, $\boldsymbol{b}'$, $\boldsymbol{A}^{-1}$, $\boldsymbol{AI}_2$, $\boldsymbol{I}_2\boldsymbol{A}$, $\boldsymbol{I}_2\boldsymbol{b}$, $\boldsymbol{Ab}$, $\boldsymbol{AA}$, $\boldsymbol{b}'\boldsymbol{b}$, and $\boldsymbol{bb}'$.

$$\boldsymbol{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ -2 & 2 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Exercise 1.4

Find the inverse of each of the following matrices:

$$\boldsymbol{E} = \begin{bmatrix} 4 & -2 & 1 \\ 7 & 3 & 3 \\ 2 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{F} = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 0 & 3 \\ 4 & 0 & 2 \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Exercise 1.5

For the regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, suppose

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 50 & 30 \\ 30 & 84 \end{bmatrix} \qquad \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} 180 \\ 60 \end{bmatrix}$$

Compute $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, the OLS estimates of the parameter vector $\boldsymbol{\beta}$.

Exercise 1.6

Consider the constant-returns-to-scale Cobb-Douglas production function $Y_t = K_t^a L_t^{1-a}$, where Y is output and K and L are capital and labour inputs respectively. A popular econometric model is

$$\ln Y_t = \beta_1 + \beta_2 \ln K_t + \beta_3 \ln L_t + e_t$$

where the constant-returns-to-scale assumption is not imposed (*i.e.* β_2 and β_3 are not re-stricted to sum up to unity). The observations on capital and labour inputs (both logged) are grouped in the matrix of regressors $\boldsymbol{X}$ that also includes a constant regressor. The

observations on output (logged) are grouped in the vector of dependent variables $\boldsymbol{y}$.

$$X = \begin{bmatrix} 1 & 2.3 & 2.5 \\ 1 & 3.3 & 3.4 \\ 1 & 3.2 & 3.1 \\ 1 & 3.1 & 3.1 \\ 1 & 2.7 & 2.8 \\ 1 & 2.9 & 3.0 \\ 1 & 2.3 & 2.4 \\ 1 & 3.2 & 3.3 \\ 1 & 3.0 & 3.1 \\ 1 & 3.0 & 3.1 \end{bmatrix} \qquad y = \begin{bmatrix} 2.4 \\ 3.5 \\ 3.1 \\ 3.3 \\ 2.7 \\ 3.0 \\ 2.3 \\ 3.1 \\ 2.9 \\ 3.0 \end{bmatrix}$$

- Compute $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$, the OLS estimates of the parameter vector $\boldsymbol{\beta}$.

Exercise 1.7

(a) Find the transpose *and* the inverse of

$$A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}.$$

(b) Find the inverse of

$$B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{bmatrix}.$$

(c) Find the determinant *and* the matrix of cofactors of

$$C = \begin{bmatrix} 7 & 2 & -3 \\ 5 & 3 & 6 \\ 0 & 1 & 0 \end{bmatrix}.$$

(d) Calculate the product

$$\begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2 & -1 & -2 \\ 5 & 2 & 1 \end{bmatrix}.$$

Exercise 1.8

Consider an open economy described by the following equations

$$Y = C + I + G + X \tag{1}$$

$$C = a_0 + a_1 Y^d \tag{2}$$

$$Y^d = (1 - \tau)Y \tag{3}$$

$$X = b_0 - b_1 Y \tag{4}$$

where Y, I, G, C, X and Y^d denote income, investment, government spending, consumption, net exports and disposable income respectively. τ, a_i and b_i for $i = 1, 2$ are constant parameters. Note that we assume I and G are exogenous variables.

(a) Eliminate Y^d from the system (1)-(4) and then use matrix notation to write a system of equations in the remaining endogenous variables Y, C and X. Your system must be of the form $\boldsymbol{Az} = \boldsymbol{h}$ were $\boldsymbol{z} = [Y \quad C \quad X]^\top$.

(b) Suppose that $a_0 = 100$, $a_1 = 0.9$, $b_0 = 100$, $b_1 = 0.12$, $\tau = 0.2$, $I = 200$ and $G = 200$. Find the inverse of matrix $\boldsymbol{A}$ and solve for the value of the vector $\boldsymbol{z}$.

Exercise 1.9

Let $\boldsymbol{X}$ and $\boldsymbol{y}$ be defined as

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}$$

1. Calculate $\boldsymbol{X'X}$.

2. Calculate the rank of $\boldsymbol{X}$ and that of $\boldsymbol{X'X}$ and explain how you get to your answer.

3. Calculate the inverse of $\boldsymbol{X'X}$.

4. Write down a 3×4 matrix that is <u>not</u> full rank and explain why it is not full rank.

1.9   Solutions to Exercises

Solution to Exercise 1.3

$$\boldsymbol{A}' = \begin{bmatrix} 3 & -2 \\ 1 & 2 \end{bmatrix}, \quad \boldsymbol{b}' = \begin{bmatrix} 2 & 3 \end{bmatrix}, \quad \boldsymbol{A}^{-1} = \begin{bmatrix} 0.25 & -0.125 \\ 0.25 & 0.375 \end{bmatrix},$$

$$\boldsymbol{AI}_2 = \boldsymbol{A}, \quad \boldsymbol{I}_2\boldsymbol{A} = \boldsymbol{A}, \quad \boldsymbol{I}_2\boldsymbol{b} = \boldsymbol{b},$$

$$\boldsymbol{Ab} = \begin{bmatrix} 9 \\ 2 \end{bmatrix}, \quad \boldsymbol{AA} = \begin{bmatrix} 7 & 5 \\ -10 & 2 \end{bmatrix}, \quad \boldsymbol{b}'\boldsymbol{b} = 13, \quad \boldsymbol{bb}' = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}$$

Solution to Exercise 1.4

$$\boldsymbol{E}^{-1} = \begin{bmatrix} 0.375 & 0.250 & -1.125 \\ -0.125 & 0.250 & -0.625 \\ -0.750 & -0.500 & 3.250 \end{bmatrix}, \quad \boldsymbol{F}^{-1} = \begin{bmatrix} 0.0 & -0.2 & 0.3 \\ -1.0 & 0.6 & 0.1 \\ 0.0 & 0.4 & -0.1 \end{bmatrix},$$

$$\boldsymbol{G}^{-1} = \begin{bmatrix} 1/1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$$

Solution to Exercise 1.5

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 4.0364 & -0.7273 \end{bmatrix}'.$$

Solution to Exercise 1.6

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} -0.1077 & 0.6587 & 0.3783 \end{bmatrix}'.$$

Solution to Exercise 1.7

(a)

$$\boldsymbol{A}' = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}, \quad \boldsymbol{A}^{-1} = \begin{bmatrix} 0.3 & -0.2 \\ -0.1 & 0.4 \end{bmatrix}$$

(b)

$$\boldsymbol{B} = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/7 \end{bmatrix}.$$

(c)

$$|\boldsymbol{C}| = -57, \quad \text{cof}(\boldsymbol{C}) = \begin{bmatrix} -6 & 0 & 5 \\ -3 & 0 & -7 \\ 21 & -57 & 11 \end{bmatrix}$$

(d)

$$\begin{bmatrix} 18 & 0 & -6 \\ 17 & 5 & 1 \end{bmatrix}$$

Solution to Exercise 1.8

(a)

$$\begin{bmatrix} 1 & -1 & -1 \\ -a_1(1-\tau) & 1 & 0 \\ b_1 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y \\ C \\ X \end{bmatrix} = \begin{bmatrix} I+G \\ a_0 \\ b_0 \end{bmatrix}$$

(b)

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 2.5 & 2.5 & 2.5 \\ 1.8 & 2.8 & 1.8 \\ -0.3 & -0.3 & 0.7 \end{bmatrix}, \qquad \boldsymbol{z} = \boldsymbol{A}^{-1}\boldsymbol{h} = \begin{bmatrix} 1,500 \\ 1,180 \\ -80 \end{bmatrix}.$$

Solution to Exercise 1.9

(a)

$$\begin{bmatrix} 3 & 9 \\ 9 & 29 \end{bmatrix}$$

(b) rank($\boldsymbol{X}$)$\leq (2,3)$. Both columns of $\boldsymbol{X}$ are linearly independent. Therefore, rank($\boldsymbol{X}$)=2. rank($\boldsymbol{X}'\boldsymbol{X}$)=rank($\boldsymbol{X}$)=2.

(c) $|\boldsymbol{X}'\boldsymbol{X}| = 6$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{1}{6}\begin{bmatrix} 29 & -9 \\ -9 & 3 \end{bmatrix}$$

(d) Let matrix $\boldsymbol{A}$ be

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 1 & 3 & 5 & 7 \\ 1 & 3 & 5 & 7 \end{bmatrix}$$

Since all rows of $\boldsymbol{A}$ are identical they are evidently not independent. Therefore the rank of this matrix is 1. The rank of the matrix would have to be 3 for it to be full rank.

2 Some Probability Concepts

2.1 Random Variables and PDFs

Almost all economic variables are random variables. A **random variable** is a variable whose value is unknown until it is observed. The notational convention is to denote random variables by uppercase letters (*e.g.* X or Y) and outcomes of these random variables by lowercase letters (*e.g.* x or y).

Random variables are either discrete or continuous. A **discrete random variable** can take only a finite number of values (*e.g.* gender, age, number of years of schooling). A **continuous random variable** can take any real value in an interval on the real number line (*e.g.* stock returns, GDP growth rates).

The value of a random variable is unknown but we can calculate the probability of a given outcome using the **probability density function (PDF)** (also called probability function) of the random variable. In the case of a discrete random variable, the PDF is simply a table listing all possible outcomes and their associated probabilities. For example, consider the random variable X = the number of heads obtained when flipping two coins simultaneously. The PDF of the discrete random variable X, denoted $f(x)$, is then

<div align="center">

Table 2.1

x	$f(x)$
0	1/4
1	2/4
2	1/4

</div>

For the discrete random variable X, the value of the PDF $f(x)$ is the probability that the random variable X takes the value x. Therefore, for a discrete random variable X that can take n possible values: $f(x_i) = Prob(X = x_i)$, $0 \leq f(x_1), f(x_2), \ldots, f(x_n) \leq 1$ and $f(x_1) + f(x_2) + \ldots + f(x_n) = 1$.

For a continuous random variable X, the PDF $f(x)$ can be represented by a function, which can be described graphically by a curve. For example, if X is normally distributed with mean μ_X and variance σ_X^2, its PDF is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[\frac{-(x - \mu_X)^2}{2\sigma_X^2}\right], \qquad -\infty < x < \infty. \tag{48}$$

Figure 2.1 shows the PDF of the continuous random variable X that is normally distributed with $\mu_x = 0$ and $\sigma_X^2 = 1$.

Figure 2.1: PDF of a N(0,1)
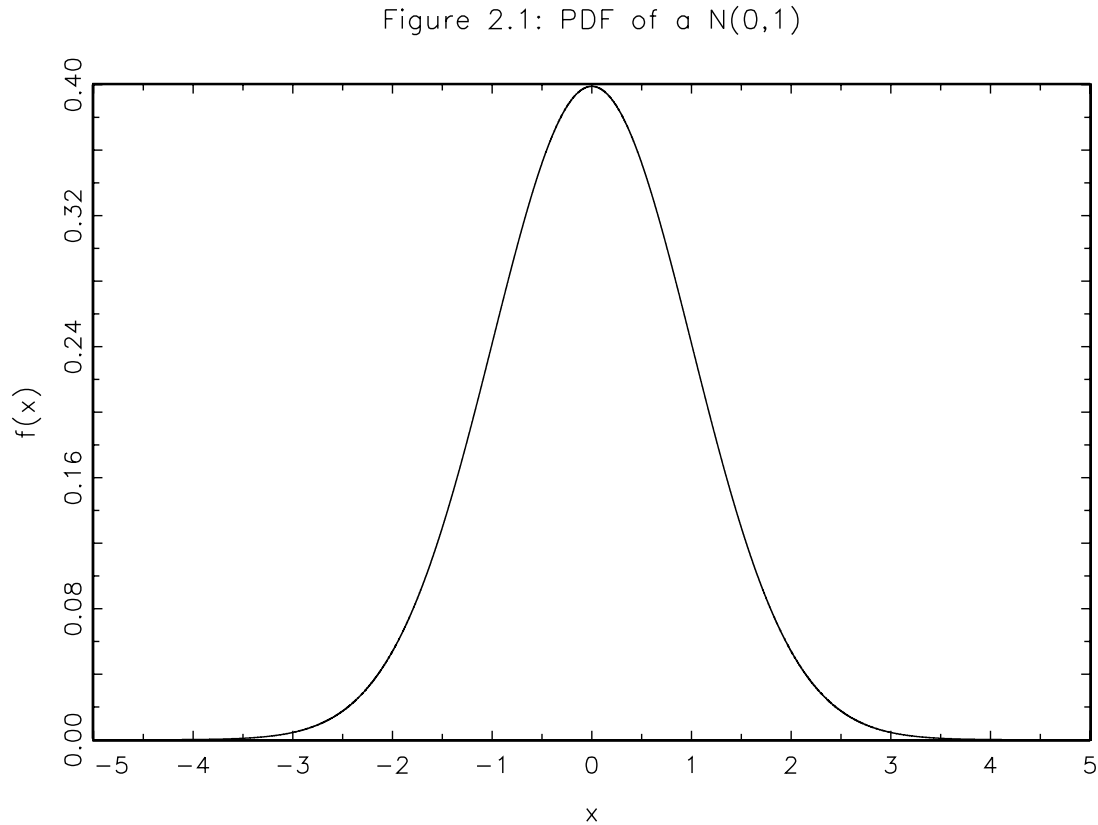
For a continuous random variable, the probability of the random variable taking a value in a certain range corresponds to a specific **area under its PDF**. The total area under a PDF is equal to 1 and the probability that the random variable X takes a value in the interval $[a, b]$, denoted $Prob(a \leq X \leq b)$, is the area under the PDF of X between the values $x = a$ and $x = b$. Using calculus we find that probability as follows

$$Prob(a \leq X \leq b) = \int_a^b f(x) \ dx. \tag{49}$$

Note that since a continuous random variable takes an uncountably infinite number of values, the probability of any one value occurring is zero. In other words, $Prob(X = a) = Prob(a \leq X \leq a) = 0$.

In the above example where X is normally distributed with $\mu_x = 0$ and $\sigma_X^2 = 1$, we have

$$Prob(0.8 \leq X \leq 2) = \int_{0.8}^2 \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-x^2}{2}\right] dx = 0.1891. \tag{50}$$

2.2 Moments of Random Variables

2.2.1 Rules of Summation

The formulae calculating the moments of random variables are more conveniently written using summation notation. Therefore we first review the **rules of summation**. In equations (51)-(58) below, X and Y are random variables and a and b are constants. X can take n possible values and Y can take m possible values where $m \geq n$.

1.
$$x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i. \tag{51}$$

2.
$$\sum_{i=1}^{n} a = n\ a. \tag{52}$$

3.
$$\sum_{i=1}^{n} ax_i = a \sum_{i=1}^{n} x_i. \tag{53}$$

4.
$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i. \tag{54}$$

5.
$$\sum_{i=1}^{n} (a + bx_i) = n\ a + b \sum_{i=1}^{n} x_i. \tag{55}$$

6.
$$\sum_{i=1}^{n} (ax_i + by_i) = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} y_i. \tag{56}$$

7. The **arithmetic mean (average)** of T realizations of the random variable X is
$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_T}{T} = \frac{1}{T} \sum_{i=1}^{T} x_i. \tag{57}$$

8. The order of summation does not matter
$$\sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i, y_i) = \sum_{j=1}^{m} \sum_{i=1}^{n} f(x_i, y_i). \tag{58}$$

We sometimes use abbreviated forms of the summation operator. For example suppose that X can take n possible values. Instead of $\sum_{i=1}^{n} f(x_i)$ we could write $\sum_i f(x_i)$ which reads

"sum over all values of the index i". Or alternatively, $\sum_x f(x)$ which reads "sum over all possible values of X."

Exercise 2.1 – Summation Operator

Express each of the following sums using the summation operator

(a) $y_1 + y_2 + y_3 + y_4 + y_5 + y_6$

(b) $ax_1 + bx_2 + cx_3 + dx_4$

(c) $ax_1 + ax_2 + ax_3 + ax_4$

(d) $y_1 + x_1 + y_2 + x_2 + y_3 + x_3 + x_4 + y_4$

Exercise 2.2 – Summation Operator

Using the sample of data

t	1	2	3	4	5
x_t	2	3	7	0	9
y_t	4	10	1	5	6

Calculate the following

(a) $\sum_{t=1}^{5} y_t$

(b) $\sum_{t=2}^{4} x_t$

(c) $\sum_{t=1}^{5} x_t y_t$

(d) $\sum_{t=1}^{5} x_t \sum_{t=1}^{5} y_t$

(e) $\sum_{t=1}^{5} x_t^2$

(f) $\left(\sum_{t=1}^{5} x_t\right)^2$

2.2.2   Expectation, Variance and the Expectation Operator

The **first moment** of a random variable is its **mean or expectation**. The expectation of a random variable X is the average value of the random variable in an *infinite number of repetitions of the experiment* (or in repeated samples). We denote the expectation of X by $E[X]$.

If X is a discrete random variable with associated PDF $f(x)$, then the expectation of X is

$$E[X] = \sum_x x\, f(x). \tag{59}$$

For example, using the PDF in Table 2.1 we have

$$E[X] = \sum_x x\ f(x) = 0 \times 1/4 + 1 \times 1/2 + 2 \times 1/4 = 1. \tag{60}$$

It is important to distinguish an expectation from a sample average. To illustrate the difference, consider our example where X = the number of heads obtained when flipping two coins simultaneously. The expectation of X is equal to 1 as shown in equation (60). Suppose that we repeat the experiment three times and get the outcomes: Head-Tail, Tail-Head and Head-Head. Therefore the three realizations of X in our sample are 1, 1 and 2. The sample average is $\bar{x} = (1 + 1 + 2)/3 = 4/3$. In general, the sample average will differ from the expectation, as it is the case in this example. However, if we repeat the experiment an infinite number of times, then the sample average will be equal to the expectation in accordance with the definition of an expectation given at the beginning of this subsection.[5]

In the case of a continuous random variable, the expectation is calculated as

$$E[X] = \int x\ f(x)\ dx. \tag{61}$$

To calculate the **expectation of a function** $g(X)$ **of random variable** X, we apply the formula

$$E[g(X)] = \sum_x g(x)f(x) \tag{62}$$

when X is a discrete random variable and

$$E[g(X)] = \int g(x)\ f(x)\ dx \tag{63}$$

when X is a continuous variable.

Note that in general $E[g(X)] \neq g(E[X]])$. We will come back to this below.

To illustrate how to apply formula (62), let us use the PDF of the discrete random variable X displayed in Table 2.1 to calculate $E[3X^2]$.

$$E[3X^2] = (3 \times 0^2)\frac{1}{4} + (3 \times 1^2)\frac{1}{2} + (3 \times 2^2)\frac{1}{4} = 4.5. \tag{64}$$

---

[5]To convince yourself, repeat the experiment a large number of times. After each repetition, recalculate the sample mean. You will see that as the number of repetitions gets large, the sample mean converges to the expectation.

Here are some **properties of the expectation operator**. They are all relatively easy to show.[6] Let a and c denote constants and X denote a random variable:

1. The expectation of a constant

$$E[c] = c. \tag{65}$$

2. Multiplying a random variable by a constant

$$E[cX] = cE[X]. \tag{66}$$

3. Expectation of a sum. Let $g_1(X)$ and $g_2(X)$ be functions of the random variable X, then

$$E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]. \tag{67}$$

4. Combining the previous properties

$$E[a + cX] = a + cE[X]. \tag{68}$$

Often times, economic models imply taking the expectation of non-linear functions. For example, in ECON 723 we often deal with Euler equations. These equations involve taking the expectation of a marginal utility function which is usually non-linear. It turns out that if $g()$ is a concave function we have $E[g(X)] \leq g(E[X])$ whereas we have $E[g(X)] \geq g(E[X])$ if $g()$ is a convex function. This is **Jensen's inequality**. Let's prove Jensen's inequality for the case where $g()$ is convex and differentiable. We first use the convexity of $g(X)$ to write

$$g(x) \geq g(\mu_X) + g'(\mu_X)(x - \mu_X), \quad \forall x \tag{69}$$

where $\mu_X = E[X]$ and $g'(x) = \partial g(x)/\partial x$. Combining equations (62) and (69) we have

$$E[g(X)] = \sum_x g(x)f(x) \geq \sum_x [g(\mu_X) + g'(\mu_X)(x - \mu_X)]f(x). \tag{70}$$

It is straightforward to show that the term to the right of the inequality sign is equal to $g(\mu_X) = g(E[X])$

$$\sum_x [g(\mu_X) + g'(\mu_X)(x - \mu_X)]f(x) = \sum_x g(\mu_X)f(x) + \sum_x g'(\mu_X)(x - \mu_X)f(x) \tag{71}$$

$$= g(\mu_X) + g'(\mu_X)\sum_x xf(x) - g'(\mu_X)\mu_X = g(\mu_X) + g'(\mu_X)\mu_X - g'(\mu_X)\mu_X = g(\mu_X)$$

---

[6]Doing so is an excellent exercise.

Combining (70) and (71) completes the proof that $E[g(X)] \geq g(E[X])$.

Exercise 2.3 − Jensen's Inequality

Suppose that a representative agent has utility function $u(Y) = \ln(Y)$ where Y denotes consumption. Suppose that consumption has the following PDF

Table 2.2

y	$f(y)$
2	0.35
3	0.25
4	0.40

The expected marginal utility of consumption is $E[1/Y]$. Show that $E[1/Y] \neq 1/E[Y]$ in this numerical example.

We complete this section by discussing the second moment of a random variable, its **variance**. The variance of a random variable X is defined as
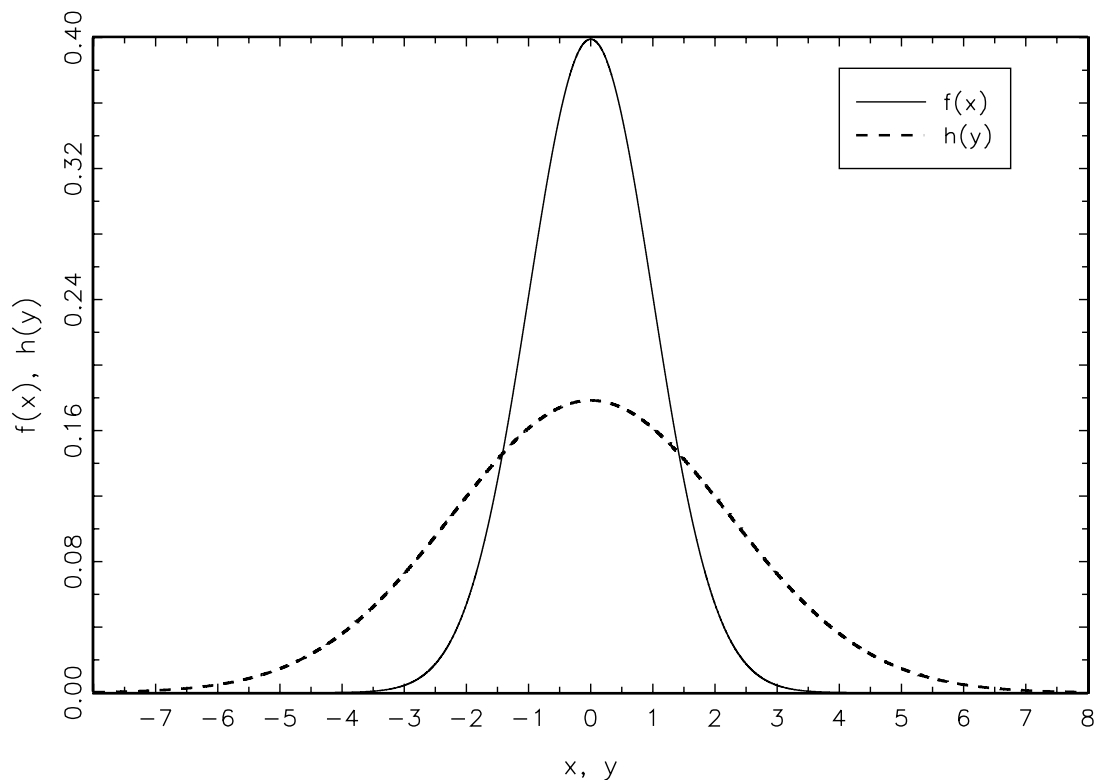
$$var(X) = E(X - E[X])^2. \tag{72}$$

Note that the symbol σ^2 is often used to denote the variance (and the symbol σ to denote the standard deviation) of a random variable.

The variance of a function $g(X)$ of random variable X is calculated as

$$var(g(X)) = E(g(X) - E[g(X)])^2. \tag{73}$$

While the expectation provides information about the center of the distribution of a random variable, the variance provides information about the spread of the distribution. Figure 2.2 compares the PDFs of two random variables: X is normally distributed with expectation 0 and variance 1, and Y is normally distributed with expectation 0 and variance 5. We see that both PDFs are centered around zero but that the PDF of Y (dotted line) is "flatter".

Figure 2.2: PDF of X~N(0,1) and Y~N(0,5)

We can use equation (62) to re-write the variance formula

$$var(X) = E(X - E[X])^2 = \sum_x f(x)(x - E[X])^2 \tag{74}$$

where $f(x)$ is the PDF of the discrete random variable X.

Exercise 2.4 – Variance of Consumption and Marginal Utility

Using the PDF in Table 2.2, calculate the variance of consumption (Y) and marginal utility $(1/Y)$.

Just like the expectation operator, the variance operator has many properties. The ones listed below are all relatively easy to show.[7] Let a and c denote constants and X denote a random variable:

---

[7] Doing so is an excellent exercise.

1. The variance of a constant is zero

$$var(a) = 0. \tag{75}$$

2. Multiplying a r.v. by a constant multiplies its variance by the square of the constant

$$var(cX) = c^2 var(X). \tag{76}$$

3. Combining properties 1 and 2

$$var(a + cX) = c^2 var(X). \tag{77}$$

2.3 Several Random Variables

2.3.1 Joint, Marginal and Conditional PDFs

To make probability statements about more than one random variable we must know their **joint probability density function** (or joint PDF). The joint PDF of two continuous random variables must be represented on a three-dimensional graph. The volume under the surface of the joint PDF is equal to 1.

The joint PDF of two discrete random variables, Y_1 and Y_2 say, is a table reporting possible outcomes of the two variables with their joint probabilities. For example, consider a two-period economy where consumption in period 1 (Y_1) and consumption in period 2 (Y_2) are random variables with joint PDF

Table 2.3: $f(y_1, y_2)$ — Joint PDF of Y_1 and Y_2

$y_1 \backslash y_2$	2	3	4
2	0.20	0.15	0.05
4	0.05	0.15	0.40

The joint PDF above indicates that the probability that the agent consumes 2 units in period 1 and 3 units in period 2 is $f(2, 3) = 0.15$. The probability that the agent consumes 4 units in period 1 and 4 units in period 2 is $f(4, 4) = 0.40$. The concept of joint PDF is not limited to the binary case. For example, if $Y_1, Y_2, \ldots, Y_m$ are random variables, then their joint PDF is denoted $f(y_1, y_2, \ldots, y_m)$.

We calculate the **expectation of a function of several random variables** using a formula similar to that displayed in equation (62). Again, let Y_1 and Y_2 denote two discrete random variables with joint PDF $f(y_1, y_2)$ and let $g(Y_1, Y_2)$ denote a function of Y_1 and Y_2. Then we have

$$E[g(Y_1, Y_2)] = \sum_{y_1} \sum_{y_2} g(y_1, y_2) f(y_1, y_2). \tag{78}$$

If Y_1 and Y_2 are continuous random variables, then

$$E[g(Y_1, Y_2)] = \int_{y_1} \int_{y_2} g(y_1, y_2) f(y_1, y_2) \ dy_1 \ dy_2. \tag{79}$$

Exercise 2.5 – Expected Consumption Growth

Using the joint PDF in Table 2.3, calculate expected gross consumption growth between periods 1 and 2.

Given a joint PDF, we can calculate marginal PDFs and conditional PDFs. A **marginal PDF** indicates the probability of an individual random variable.[8] Let Y_1 and Y_2 be two discrete random variables. To calculate the marginal PDF of Y_1 (denoted $f(y_1)$) using the joint PDF $f(y_1, y_2)$ we "sum out" the *other variable* (Y_2)

$$f(y_1) = \sum_{y_2} f(y_1, y_2). \tag{80}$$

And similarly

$$f(y_2) = \sum_{y_1} f(y_1, y_2). \tag{81}$$

If Y_1 and Y_2 are continuous random variables we have

$$f(y_1) = \int f(y_1, y_2) \ dy_2, \qquad f(y_2) = \int f(y_1, y_2) \ dy_1. \tag{82}$$

Coming back to the two-period consumption example. The marginal PDFs of Y_1 and Y_2 are

---

[8]So in a sense, the PDF displayed in Table 2.1 is a marginal PDF. The notation reflects this fact.

Table 2.4 Marginal PDF of Y_1 (period 1 consumption)

y_1	$f(y_1)$
2	0.40
4	0.60

Table 2.5 Marginal PDF of Y_2 (period 2 consumption)

y_2	$f(y_2)$
2	0.25
3	0.30
4	0.45

Table 2.4 indicates that the consumer will consume 2 units in period 1 with a probability of 0.40. Table 2.5 indicates that the consumer will consume 3 units in period 2 with a probability of 0.30.

A **conditional probability** is the probability of an event conditional on the realization of another event. Consider the case of two discrete random variables, Y_1 and Y_2. Conditional probabilities are calculated from the joint PDF $f(y_1, y_2)$ and the marginal PDF of the conditioning random variable. $Prob[Y_2 = y_2 | Y_1 = y_1]$ denotes the probability that the random variable Y_2 takes the value y_2 *given* that the random variable Y_1 takes the value y_1. We denote the **conditional PDF** by $f(y_2|y_1)$ and calculate it as follows

$$f(y_2|y_1) = Prob[Y_2 = y_2 | Y_1 = y_1] = \frac{f(y_1, y_2)}{f(y_1)} \tag{83}$$

In the two-period consumption example, suppose we want to know the probability that the agent consumes 3 units in period 2 given that he consumes 2 units in period 1. We calculate the probability as

$$Prob(Y_2 = 3 | Y_1 = 2) = \frac{Prob(Y_1 = 2, Y_2 = 3)}{Prob(Y_1 = 2)} = \frac{0.15}{0.40} = 0.375. \tag{84}$$

Note that we could also calculate $Prob(Y_1 = y_1 | Y_2 = y_2)$ if we wanted to.

In the two-period consumption example, the PDF of the random variable Y_2 conditional on $Y_1 = 4$ is

$$\text{Table 2.6 PDF of } Y_2 \text{ conditional on } Y_1 = 4$$

y_2	$f(y_2|4)$
2	0.05/0.60=0.0833
3	0.15/0.60=0.2500
4	0.40/0.60=0.6667

2.3.2 Statistical Independence

We say that two random variables are **statistically independent, or independently distributed**, if knowing the value that one takes does not reveal anything about what value the other may take.

When the random variables Y_1 and Y_2 are statistically independent, their joint PDF is the product of their marginal PDFs, that is

$$f(y_1, y_2) = f(y_1)f(y_2). \tag{85}$$

Similarly, if the random variables $Y_1, Y_2, \ldots, Y_n$ are statistically independent their joint PDF is

$$f(y_1, y_2, \ldots, y_n) = f(y_1)f(y_2)\cdots f(y_n) \tag{86}$$

Combining (83) and (85) we find that if Y_1 and Y_2 are statistically independent random variables then the conditional PDFs reduce to marginal PDFs

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f(y_1)} = \frac{f(y_1)f(y_2)}{f(y_1)} = f(y_2), \qquad f(y_1|y_2) = \frac{f(y_1, y_2)}{f(y_2)} = \frac{f(y_1)f(y_2)}{f(y_2)} = f(y_1). \tag{87}$$

In the two-period consumption example, the random variable Y_1 and Y_2 are not statistically independent. Notice how marginal and conditional PDFs differ by comparing tables 2.5 and 2.6.

2.3.3 Conditional Expectation and Variance

With the conditional PDF of a random variable on hand, we can calculate **conditional moments** (*e.g.* conditional expectation and conditional variance). For example, using the

conditional PDF in Table 2.6 we can calculate the expectation and variance of the random variable Y_2 conditional on $Y_1 = 4$

$$E[Y_2|Y_1 = 4] = 2 \times Prob(Y_2 = 2|Y_1 = 4) + 3 \times Prob(Y_2 = 3|Y_1 = 4) + 4 \times Prob(Y_2 = 4|Y_1 = 4) \tag{88}$$

$$= 2 \times 0.0833 + 3 \times 0.25 + 4 \times 0.6667 = 3.5833,$$

$$var(Y_2|Y_1 = 4) = E[Y_2 - E(Y_2|Y_1 = 4)]^2 = (2 - 3.5833)^2 \times Prob(Y_2 = 2|Y_1 = 4) \tag{89}$$

$$+(3 - 3.5833)^2 \times Prob(Y_2 = 3|Y_1 = 4)$$

$$+(4 - 3.5833)^2 \times Prob(Y_2 = 4|Y_1 = 4) = 0.4097.$$

Exercise 2.6 − Bond Returns

Suppose that in the two-period economy example the gross return on a one-period real bond, denoted $R_{Y_1 = y_1}$, is related to consumption as follows

$$\frac{1}{R_{Y_1 = y_1}} = 0.95 \, E\left[\frac{Y_1}{Y_2}\bigg| Y_1 = y_1\right] \tag{90}$$

where the subscript on R reminds us that the return is conditional on the value of consumption in period 1.

Using the PDF in Table 2.3, calculate the two possible values for the conditional return R and then calculate the unconditional expected bond return R^e

---

2.3.4   Covariance and Correlation

To measure how closely two variables move together, we calculate their covariance defined as

$$cov(Y_1, Y_2) = E[(Y_1 - E[Y_1])(Y_2 - E[Y_2])]. \tag{91}$$

The expression in square brackets in equation (91) is an example of a function of Y_1 and Y_2. Therefore, we can combine equations (78) and (91) to write

$$cov(Y_1, Y_2) = E[(Y_1 - E[Y_1])(Y_2 - E[Y_2])] = \sum_{y_1} \sum_{y_2} (y_1 - E[Y_1])(y_2 - E[Y_2]) \, f(y_1, y_2) \tag{92}$$

when Y_1 and Y_2 are discrete random variables and equations (79) and (91) to write

$$cov(Y_1, Y_2) = E[(Y_1 - E[Y_1])(Y_2 - E[Y_2])] = \int_{y_1} \int_{y_2} (y_1 - E[Y_1])(y_2 - E[Y_2]) \ f(y_1, y_2) \ dy_1 \ dy_2 \tag{93}$$

when Y_1 and Y_2 are continuous random variables.

The sign of the covariance indicates whether the variables are positively or negatively associated. Looking at equations (92) or (93) we see that the covariance between Y_1 and Y_2 will be positive if Y_1 is above its expectation when Y_2 is also above its expectation and *vice versa*. Figures 2.3 and 2.4 show data points drawn from a joint PDF where the random variables Y_1 and Y_2 have positive and negative covariance, respectively.

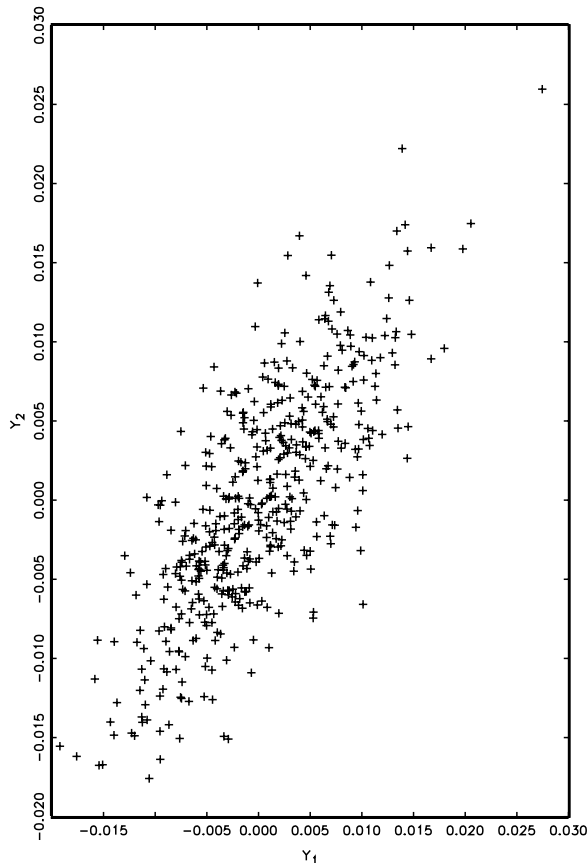Figure 2.3: Realizations of Two Positively Correlated R.V. $cor(Y_1, Y_2)=0.75$

Figure 2.4: Realizations of Two Negatively Correlated R.V. $cor(Y_1, Y_2)=-0.75$
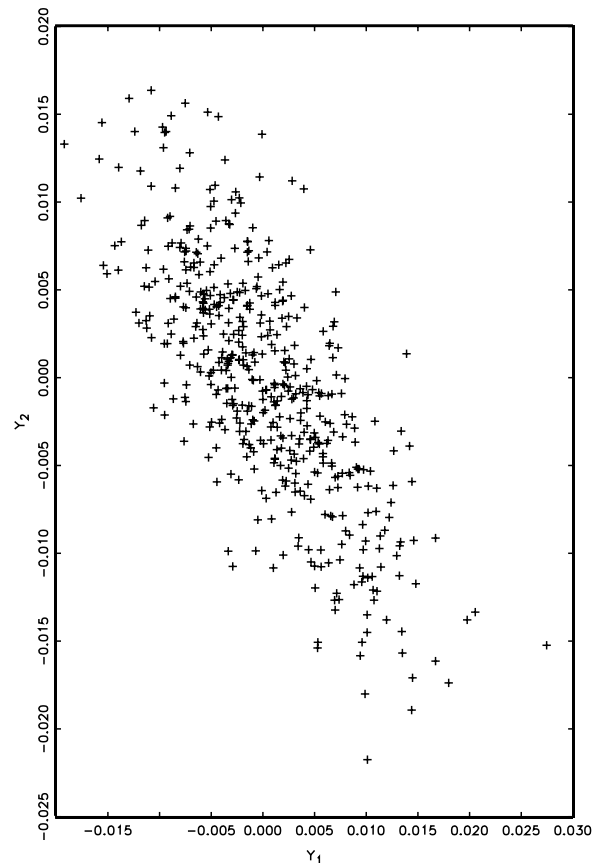
Because the covariance depends on the units in which the variables are measured, it is useful to divide the covariance by the (product of the) standard deviations of the random variables in order to measure the strength of the linear association between the two variables. The result is the familiar formula for the correlation coefficient between two random variables

Y_1 and Y_2

$$cor(Y_1, Y_2) = \frac{cov(Y_1, Y_2)}{\sqrt{var(Y_1)\ var(Y_2)}}. \tag{94}$$

The correlation coefficient lies in the interval $[-1, 1]$. When $cor(Y_1, Y_2)$ equals 1 (-1), we say that the variables Y_1 and Y_2 are perfectly positively (negatively) correlated.

Note that if the random variables Y_1 and Y_2 are statistically independent, then their covariance and correlation are zero. However, the converse is *not* true.

Exercise 2.7 – Covariance and Correlation

Use the joint PDF in Table 2.3 to calculate the covariance and correlation of the random variables Y_1 and Y_2.

2.3.5 Expectation and Variance of a Sum of Random Variables

This short subsection lists a few important rules regarding expectations and variances of a sum of random variables. They can all be derived from equation (78).[9] Let Y_1 and Y_2 be random variables and a and b be constants. The rules are

1. Expectation of a weighted sum of random variables

$$E[aY_1 + bY_2] = aE[Y_1] + bE[Y_2]. \tag{95}$$

2. Variance of a weighted sum of random variables

$$var(aY_1 + bY_2) = a^2 var(Y_1) + b^2 var(Y_2) + 2ab\ cov(Y_1, Y_2). \tag{96}$$

3. Expectation of a product of random variables

$$E[aY_1\ bY_2] = ab\ E[Y_1]\ E[Y_2] + ab\ cov(Y_1, Y_2). \tag{97}$$

In the **special case where Y_1 and Y_2 are independent or uncorrelated** we have

$$var(aY_1 + bY_2) = a^2 var(Y_1) + b^2 var(Y_2), \qquad E[aY_1\ bY_2] = ab\ E[Y_1]\ E[Y_2]. \tag{98}$$

---

[9]Doing so is an excellent exercise.

These rules generalized to the case where there is more than two random variables. For example, let c denote a constant and Y_3 a random variable. Then we have

$$E[aY_1 + bY_2 + cY_3] = aE[Y_1] + bE[Y_2] + cE[Y_3] \tag{99}$$

and

$$var(aY_1 + bY_2 + cY_3) = a^2 var(Y_1) + b^2 var(Y_2) + c^2 var(Y_3) + 2ab \ cov(Y_1, Y_2) \tag{100}$$

$$+2ac \ cov(Y_1, Y_3) + 2bc \ cov(Y_2, Y_3).$$

2.3.6 Random Vectors and Matrices

Closely related to the concepts of expectation, variance and covariance of scalar random variables is that of the expectation of a random vector or matrix and the **variance-covariance matrix** of a vector of random variables.

Recall the vector of error terms in (21). Denote the expectation of element e_i as $E[e_i] = \mu_i$ and let $\boldsymbol{\mu} = [\begin{matrix} \mu_1 & \mu_2 & \dots & \mu_T \end{matrix}]'$. The expectation of a vector (or matrix) is calculated by taking the expectation of each element inside the vector (or matrix). Therefore we have

$$E[\boldsymbol{e}] = \begin{bmatrix} E[e_1] \\ E[e_2] \\ E[e_3] \\ \vdots \\ E[e_T] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_T \end{bmatrix} = \boldsymbol{\mu} \tag{101}$$

By definition, the variance-covariance matrix of $\boldsymbol{e}$ is

$$\text{Var}(\boldsymbol{e}) = E\left[(\boldsymbol{e} - E[\boldsymbol{e}])(\boldsymbol{e} - E[\boldsymbol{e}])'\right]. \tag{102}$$

The variance-covariance matrix has variances of the random variables on the main diagonal and covariance terms on the off diagonal. To see this work out $\text{Var}(\boldsymbol{e})$ when $\boldsymbol{e}$ is a column-vector with two elements.[10]

In the context of the linear regression model where we assume $E[\boldsymbol{e}] = \boldsymbol{0}$ formula (102) simplifies to $\text{Var}(\boldsymbol{e}) = E\left[\boldsymbol{e}\boldsymbol{e}'\right]$.

---

[10]Start by working out the product $(\boldsymbol{e} - E[\boldsymbol{e}])(\boldsymbol{e} - E[\boldsymbol{e}])'$ then calculate the expectation of each element of the resulting matrix.

Let $\boldsymbol{A}$ be a matrix containing deterministic elements. Suppose that the number of columns of $\boldsymbol{A}$ matches the number of rows of $\boldsymbol{e}$. Then the variance-covariance matrix of the product $\boldsymbol{A}\boldsymbol{e}$ is

$$\text{Var}(\boldsymbol{A}\boldsymbol{e}) = E[(\boldsymbol{A}\boldsymbol{e} - E(\boldsymbol{A}\boldsymbol{e}))(\boldsymbol{A}\boldsymbol{e} - E(\boldsymbol{A}\boldsymbol{e}))'] = E[\boldsymbol{A}(\boldsymbol{e} - E(\boldsymbol{e}))(\boldsymbol{e} - E(\boldsymbol{e}))'\boldsymbol{A}'] \quad (103)$$

$$= \boldsymbol{A}E[(\boldsymbol{e} - E(\boldsymbol{e}))(\boldsymbol{e} - E(\boldsymbol{e}))']\boldsymbol{A}' = \boldsymbol{A}\text{Var}(\boldsymbol{e})\boldsymbol{A}'$$

where we make use of the property $E[\boldsymbol{A}\boldsymbol{e}] = \boldsymbol{A}E[\boldsymbol{e}]$ which can easily be derived from (66). Make sure you understand all steps of the derivation in (103). Notice how the variance-covariance formula in (103) differs from that of the scalar case shown in (76).

2.4 Markov Chains

Acknowledgement: this section draws heavily from Ljungqvist and Sargent (2004) chapter 2.

2.4.1 A Few Definitions

A **stochastic process** is a sequence of random variables. We denote a sequence of random variables by $\{X_t\}$ where X denotes a random variable and t denotes a time index (when time is discrete t is an integer). When relevant, the exact starting and ending values of the time index are specified. For example, if the context is such that the time index can only take integer values from 1 and 50 then we write $\{X_t\}_{t=1}^{50}$.

A Markov chain is a stochastic process that satisfies the Markov property. Ljungqvist and Sargent (2004) provide the following definition: A stochastic process $\{X_t\}$ is aid to have the **Markov property** if for all $k \geq 1$ and all t,

$$Prob(x_{t+1}|x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-k}) = Prob(x_{t+1}|x_t). \quad (104)$$

Interpreting X as a state variable[11], then the Markov property means that given the current state (x_t), the realization of the next period state (x_{t+1}) is uncorrelated with past realizations $(x_{t-1}, x_{t-2}$ and so forth). When the order of the Markov chain is not specified, it is to be

---

[11]A classical example given to illustrate the concept of state of the world is the weather: in period t it either rains or shine (two possible states of the world).

understood that the chain is of the first-order (as defined above). In the case of a second-order Markov chain (104) must be replaced by

$$Prob(x_{t+1}|x_t, x_{t-1}, x_{t-2}, \dots, x_{t-k}) = Prob(x_{t+1}|x_t, x_{t-1}). \tag{105}$$

A Markov chain is said to be **time-invariant** (or time-homogenous) if

$$Prob(x_{t+h+1}|x_{t+h}) = Prob(x_{t+1}|x_t), \qquad \forall h > 0. \tag{106}$$

From now on, we work with first-order time-invariant Markov chains. Such a chain is defined by a triplet of objects.

(i) An n-dimensional **state space** consisting of vectors e_i for $i = 1, \dots, n$ where e_i is a unit vector where the ith element is equal to unity and all other elements equal zero. For example, if there are three possible states of the world then $n = 3$ (*e.g.* the GDP growth rate is zero, positive or negative) and the state space is defined by

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{107}$$

(ii) An $n \times n$ **transition matrix** P containing the transition probabilities, that is, the probabilities of moving from one value of the state space to another in one period. The ij^{th} element of P is $P_{ij} = Prob(X_{t+1} = e_j|X_t = e_i)$. Note that P does not have any kind of time dependency here because we focus on time-invariant Markov chains.

Continuing with our $n = 3$ example we have

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

$$= \begin{bmatrix} Prob(X_{t+1} = e_1|X_t = e_1) & Prob(X_{t+1} = e_2|X_t = e_1) & Prob(X_{t+1} = e_3|X_t = e_1) \\ Prob(X_{t+1} = e_1|X_t = e_2) & Prob(X_{t+1} = e_2|X_t = e_2) & Prob(X_{t+1} = e_3|X_t = e_2) \\ Prob(X_{t+1} = e_1|X_t = e_3) & Prob(X_{t+1} = e_2|X_t = e_3) & Prob(X_{t+1} = e_3|X_t = e_3) \end{bmatrix} \tag{108}$$

Matrix P satisfies condition $\sum_{j=i}^{n} P_{ij} = 1$ which makes it a **stochastic matrix**, that is, a matrix that defines the probabilities of moving from each value of the state space to any other in one period.

(iii) A vector of marginal probabilities in period 0 (the initial period) for the n elements of the state space. Let π_0 denote this column vector. Its ith element is $\pi_{0i} = Prob(X_0 = e_i)$. The vector π_0 satisfies $\sum_{i=1}^{n} = \pi_{0i} = 1$

2.4.2 Stationarity

Once the Markov chain is defined, the transition matrix containing the probabilities of moving from each value of the state space to any other in k periods is given by $\boldsymbol{P}^k$ (to illustrate this work out exercise 2.16) and the vector of unconditional probabilities in any period are calculated as

$$
\begin{aligned}
(t=1) \qquad \boldsymbol{\pi}_1' &= Prob(X_1) = \boldsymbol{\pi}_0'\boldsymbol{P} \\
(t=2) \qquad \boldsymbol{\pi}_2' &= Prob(X_2) = \boldsymbol{\pi}_0'\boldsymbol{P}^2 \\
&\vdots \\
(t=k) \qquad \boldsymbol{\pi}_k' &= Prob(X_k) = \boldsymbol{\pi}_0'\boldsymbol{P}^k
\end{aligned}
\tag{109}
$$

where $\boldsymbol{\pi}_t' = Prob(X_t)$ is $1 \times n$ and has i^{th} element $\pi_{ti} = Prob(X_t = \boldsymbol{e}_i)$. Generalizing the equation for $t=1$ above we see that unconditional probability distributions evolve according to

$$
\boldsymbol{\pi}_{t+1}' = \boldsymbol{\pi}_t'\boldsymbol{P}.
\tag{110}
$$

For example, when $n=$ and $t=1$ the above equation becomes

$$
\begin{bmatrix} \pi_{21} & \pi_{22} \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} \pi_{11}P_{11} + \pi_{12}P_{21} & \pi_{11}P_{12} + \pi_{12}P_{22} \end{bmatrix}
\tag{111}
$$

The unconditional distribution is said to be **stationary or invariant** if it satisfies $\boldsymbol{\pi}_{t+1} = \boldsymbol{\pi}_t = \boldsymbol{\pi}$. From equation (110) a stationary distribution must satisfy

$$
\boldsymbol{\pi}' = \boldsymbol{\pi}'\boldsymbol{P} \quad \Rightarrow \quad \boldsymbol{\pi}'(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{0}
\tag{112}
$$

Let's denote the probability distribution the Markov chain will converge to in the infinite future by $\boldsymbol{\pi}_\infty = \lim_{t\to\infty} \boldsymbol{\pi}_t$. Then we have a definition and a theorem regarding asymptotic stationarity in section 2.2.2 of Ljungqvist and Sargent.

Definition: Let $\boldsymbol{\pi}_\infty$ be a unique vector that satisfies $\boldsymbol{\pi}'(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{0}$ [*i.e* $\boldsymbol{\pi}_\infty$ is a stationary distribution]. If for all distributions $\boldsymbol{\pi}_0$ it is true that $\lim_{t\to\infty} \boldsymbol{P}^{t'}\boldsymbol{\pi}_0$ converges to the same $\boldsymbol{\pi}_\infty$, we say that the Markov chain is asymptotically stationary with a unique invariant distribution. Then, we call $\boldsymbol{\pi}_\infty$ a **stationary distribution of $\boldsymbol{P}$**.

Theorem: Let $\boldsymbol{P}$ be a stochastic matrix with $P_{ij} > 0 \; \forall (i,j)$. Then $\boldsymbol{P}$ has a unique stationary distribution, and the process is asymptotically stationary.

The theorem states that as long as the transition probabilities are positive, then $\boldsymbol{P}^{t'}\boldsymbol{\pi}_0$ will eventually converge to a unique $\boldsymbol{\pi}$.

2.4.3 Expectations

We terminate our introduction to Markov chains by revisiting the concept of expectations. Let's first think about $E[X_t] = \sum_{i=1}^{n} x_i \pi_{ti}$. Given the way the $x_i's$ are defined we obviously get $E[X_t] = \pi_t = P^{t'}\pi_0$. Therefore, when the Markov chain is stationary then so is $E[X_t]$.

Suppose we are interested in a variable Y which is related to the state X as follows: $Y_t = X_t'\bar{y}$ where $\bar{y}$ is an $n \times 1$ vector of real numbers. Given the way we defined the state space for X, we know that $Y_t = \bar{y}_i$ when $X_t = e_i$. The unconditional expectation of Y_t is then $E[Y_t] = \pi_t'\bar{y} = \pi_0'P^t\bar{y}$ (make sure you can derive this).

Finally, the conditional expectation of Y_{t+1} given that the state in period t is $x_t = e_i$ is simply the ith row of matrix $P\bar{y}$. That is, $E[Y_{t+1}|X_t = e_i] = (P\bar{y})_i$. For a two-step-ahead forecast we have $E[Y_{t+2}|X_t = e_i] = (P^2\bar{y})_i$ and for a k-step-ahead forecast we have $E[Y_{t+k}|X_t = e_i] = (P^k\bar{y})_i$.

Therefore, once the transition matrix P and initial distribution π_0 are known it is trivial to calculate the conditional and unconditional probability of Y_{t+k}, $k > 0$

$$E[Y_{t+k}] = \pi_0'P^{t+k}\bar{y} \qquad \text{and} \qquad E[Y_{t+k}|X_t] = P^k\bar{y} \tag{113}$$

where the last expression comes from "stacking" the expectations $E[Y_{t+k}|X_t = e_i] = (P^k\bar{y})_i$ for $i = 1, \ldots, n$.

2.5   Additional Exercises

Exercise 2.8

Consider a two-period economy where consumption in period t is given by Y_t where $t = 1,\ 2$. Consumption in both periods is random and its joint PDF is

$y_1 \backslash y_2$	2	3	4
2	0.20	0.08	0.04
3	0.08	0.20	0.08
4	0.04	0.08	0.20

(a) Calculate the expectation and the variance of Y_1.

(b) Let lifetime utlity be $U = \ln(Y_1) + 0.9 \ln(Y_2)$. Calculate the expectation and the variance of U.

(c) Given that $Y_1 = 3$, calculate $E[U|Y_1 = 3]$.

(d) Calculate $cov(Y_1, Y_2)$ and $cor(Y_1, Y_2)$.

(e) Are the random variable Y_1 and Y_2 statistically independent in this example?

Exercise 2.9

Let Y_1 and Y_2 be two discrete random variales. Use Jensen's inequality to find whether

$$E\left[\frac{Y_1}{Y_2}|Y_1 = y_1\right] \times E\left[\frac{Y_2}{Y_1}|Y_1 = y_1\right]$$

is less, greater or equal to 1.

Exercise 2.10

Let Y_1 and Y_2 be two discrete random variales. Prove that

$$var(Y_1 + Y_2) = var(Y_1) + var(Y_2) + 2\ cov(Y_1, Y_2).$$

Exercise 2.11

Consider a two-period economy where consumption in period i is given by Y_i where $i = 1,\ 2$. Consumption in both periods is random and its PDF is

$y_1 \backslash y_2$	2	3	4
1	0.10	0.08	0.02
3	0.50	0.10	0.20

(a) Calculate the expectation and the variance of Y_1.

(b) Verify whether the marginal PDF of Y_2 is equal to the PDF of Y_2 conditional on $Y_1 = 3$.

(c) Let lifetime utlity be $U = Y_1 Y_2$. Calculate the expectations $E[U]$ and $E[U|Y_1 = 3]$.

Exercise 2.12

Let Y_1 and Y_2 be two discrete random variables and let a and b be two constants. The marginal PDF of Y_i is denoted $f(y_i)$ and the PDF of Y_2 conditional on $Y_1 = y_1$ is denoted $f(y_2|y_1)$.

(a) Show that $E[aY_1 + bY_2] = aE[Y_1] + bE[Y_2]$.

(b) Show that $var(aY_2) = a^2 var(Y_2)$.

(c) Show that $f(y_2|y_1) = f(y_2)$ if Y_1 and Y_2 are statistically independent.

(difficult) Show that $E\{E[Y_2|Y_1 = y_1]\} = E[Y_2]$.

Exercise 2.13

Let $a = 2$, $b = 3$ and $c = 4$. The following table lists the *joint probabilities* for the values of two random variables, X and Y.

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 1$	0.15	0.05	0.20
$X = 2$	0.10	0.40	0.10

Calculate the following probabilities:
(a) $Pr(Y = 1)$
(b) $Pr(X = 2)$
(c) $Pr(X = 2|Y = 1)$
(d) $Pr(Y = 1|X = 1)$

Calculate the following expectations:

(e) $E[X]$
(f) $E[Y]$
(g) $E[XY]$
(h) $E[X|Y=1]$
(i) $E[a+bX]$

Calculate the following variances and covariances:
(j) $Var[X]$
(k) $Var[aX+bY]$
(l) $Cov[X,Y]$
(m) $Cov[a+bX,cY]$

Exercise 2.14

Let X_1 and X_2 be two discrete random variables and let a and b be two constants. The joint probability density function (PDF) of X_1 and X_2 is denoted $f(X_1, X_2)$.

(a) Prove that

$$var(aX_1 + bX_2) = a^2 var(X_1) + b^2 var(X_2) + 2ab\ cov(X_1, X_2).$$

(b) Prove that $cov(X_1, X_2) = 0$ if X_1 and X_2 are statistically independent.

Exercise 2.15

Consider a two-period economy where consumption in period i is given by Y_i where $i = 1,\ 2$. Consumption in both periods is random and its PDF is

$y_1 \backslash y_2$	3	4	5
2	0.35	0.15	0.10
4	0.10	0.05	0.25

(a) Calculate the variance of Y_2 conditional on $Y_1 = 4$.

(b) Verify whether the random variables Y_1 and Y_2 are statistically independent.

Exercise 2.16 – Calculating a transition matrix

Suppose there are two states of the world. Write down matrix P for the case $n = 2$ and then calculate P^2. Using a probability tree, verify that the elements of your matrix P^2 are correct.

Exercise 2.17

Let Y_1 and Y_2 be two discrete random variables with joint probability density function (PDF)

$y_1 \backslash y_2$	-3	0.75
-1.5	0.08	0.32
1	0.12	0.48

(a) Prove that Y_1 and Y_2 are statistically independent.

(b) Let Υ be a random vector defined as

$$\Upsilon = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

Calculate the expectation and the variance-covariance matrix of Υ.

2.6   Solutions to Exercises

Solution to Exercise 2.1

(a) $y_1 + y_2 + y_3 + y_4 + y_5 + y_6 = \sum_{t=1}^{6} y_t$

(b) $x_1 y_2 + x_2 y_3 + x_3 y_4 + x_4 y_5 + x_5 y_6 = \sum_{t=1}^{5} x_t y_{t+1}$ or $\sum_{t=2}^{6} x_{t-1} y_t$

(c) $ax_1 + ax_2 + ax_3 + ax_4 = \sum_{t=1}^{4} ax_t$ or $a \sum_{t=1}^{4} x_t$

(d) $y_1 + x_1 + y_2 + x_2 + y_3 + x_3 + x_4 + y_4 = \sum_{t=1}^{4} x_t + \sum_{t=1}^{4} y_t$

(e) $x_1^2 + x_2^2 + x_3^2 + x_4^2 + (y_1 + y_2 + y_3)^2 = \sum_{t=1}^{4} x_t^2 + \left(\sum_{t=1}^{3} y_t\right)^2$

Solution to Exercise 2.2

(a) $\sum_{t=1}^{5} y_t = 26$

(b) $\sum_{t=2}^{4} x_t = 10$

(c) $\sum_{t=1}^{5} x_t y_t = 99$

(d) $\sum_{t=1}^{5} x_t \sum_{t=1}^{5} y_t = 546$

(e) $\sum_{t=1}^{3} x_t^2 = 62$

(f) $\left(\sum_{t=1}^{3} x_t\right)^2 = 144$

(g) $\sum_{t=1}^{4} (x_t/y_t) = 7.8$

(h) $\left(\sum_{t=2}^{4} x_t\right) / \left(\sum_{t=1}^{5} y_t\right) = 0.3846$

Solution to Exercise 2.3

In this question, we have $E[Y] = 3.05$ and therefore $1/E[Y] = 0.3279$. However, applying formula (62) we find

$$E[1/Y] = (1/2) \times 0.35 + (1/3) \times 0.25 + (1/4) \times 0.40 = 0.3583 > 1/E[Y].$$

Solution to Exercise 2.4

Before calculating the variances, we need the expectations. From Exercise 2.3 we have $E[Y] = 3.05$ and $E[1/Y] = 0.3583$. The variance of consumption is

$$var(Y) = (2 - 3.05)^2 \times 0.35 + (3 - 3.05)^2 \times 0.25 + (4 - 3.05)^2 \times 0.40 = 0.7475$$

and the variance of marginal utility is

$$var(1/Y) = (1/2 - 0.3583)^2 \times 0.35 + (1/3 - 0.3583)^2 \times 0.25 + (1/4 - 0.3583)^2 \times 0.40 = 0.0119.$$

Solution to Exercise 2.5

Let the function $g(Y_1, Y_2)$ denote the consumption (gross) growth rate. That is, $g(Y_1, Y_2) = Y_2/Y_1$. Applying formula (78) we find

$$E[g(Y_1, Y_2)] = \frac{2}{2} \times 0.20 + \frac{3}{2} \times 0.15 + \frac{4}{2} \times 0.05 + \frac{2}{4} \times 0.05 + \frac{3}{4} \times 0.15 + \frac{4}{4} \times 0.40 = 1.0625.$$

Solution to Exercise 2.6

If $Y_1 = 2$ then,

$$\frac{1}{R_{Y_1=2}} = 0.95 \ E\left[\frac{Y_1}{Y_2}\Big|Y_1 = 2\right] = 0.95\left[\frac{2}{2} \times 0.5 + \frac{2}{3} \times 0.375 + \frac{2}{4} \times 0.125\right] = 0.8125$$

and therefore $R_{Y_1=2} = 1.2308$.

If $Y_1 = 4$ then,

$$\frac{1}{R_{Y_1=4}} = 0.95 \ E\left[\frac{Y_1}{Y_2}\Big|Y_1 = 4\right] = 0.95\left[\frac{4}{2} \times 0.0833 + \frac{4}{3} \times 0.25 + \frac{4}{4} \times 0.6667\right] = 1.1083$$

and therefore $R_{Y_1=4} = 0.9023$.

The marginal PDF of Y_1 tells us that there is a 40 percent probability that the conditional return takes a value 1.2308 and a 60 percent chance that its takes a value 0.9023. Accordingly, the unconditional expectation of the conditional return is

$$R^e \equiv E[R_{Y_1=y_1}] = 0.40 \times 1.2308 + 0.60 \times 0.9023 = 1.0337.$$

Note that you can also get to the answer

$$R^e = 1.0337$$

by using the property $E[E(X_2|X_1 = x_1)] = E(X_2)$ for any two random variables X_1 and X_2.

Solution to Exercise 2.7

$cov[Y_1, Y_2] = 0.46$ and $cor[Y_1, Y_2] = 0.5779$.

Solution to Exercise 2.8

(a) $E[Y_1] = 3$, $var[Y_1] = 0.64$.

(b) The distribution of the random variable U is

u	$f(u)$
1.316980	0.20
1.681898	0.08
1.722445	0.08
1.940812	0.04
2.010127	0.04
2.087363	0.20
2.346277	0.08
2.375045	0.08
2.633959	0.20

$$E[U] = 2.015751, \qquad var[U] = 0.210210.$$

(c) $E[U|Y_1 = 3] = 2.0638.$

(d) $cov[Y_1, Y_2] = 0.32$ and $cor[Y_1, Y_2] = 0.5.$

(e) No. Two random variables that are statistically independent are always uncorrelated.

Solution to Exercise 2.9

$$E\left[\frac{Y_1}{Y_2}|Y_1 = y_1\right] \times E\left[\frac{Y_2}{Y_1}|Y_1 = y_1\right]$$

$$= E\left[\frac{1}{Y_2}|Y_1 = y_1\right] \times E\left[Y_2|Y_1 = y_1\right]$$

defining $g(x) = 1/x$ we can write

$$= \frac{E[g(Y_2)|Y_1 = y_1]}{g(E[Y_2|Y_1 = y_1])} \geq 1$$

since $g(x) = 1/x$ is a convex function.

Solution − Exercise 2.10

Let $\mu_1 \equiv E[Y_i]$ for $i = 1, 2.$

$$var(Y_1 + Y_2) = \sum_{y_1}\sum_{y_2}[y_1 + y_2 - \mu_1 - \mu_2]^2 f(y_1, y_2)$$

$$= \sum_{y_1}\sum_{y_2}[(y_1 - \mu_1) + (y_2 - \mu_2)]^2 f(y_1, y_2)$$

$$= \sum_{y_1}\sum_{y_2}[(y_1 - \mu_1)^2 + 2(y_1 - \mu_1)(y_2 - \mu_2) + (y_2 - \mu_2)^2]f(y_1, y_2)$$

$$= \sum_{y_1}(y_1 - \mu_1)^2 f(y_1) + 2cov(Y_1, Y_2) + \sum_{y_2}(y_2 - \mu_2)^2 f(y_2)$$

$$= var[Y_1] + 2cov(Y_1, Y_2) + var[Y_2].$$

Solution – Exercise 2.11

(a) $E[Y_1] = 2.6$; $var[Y_1] = 0.64$.

(b) Marginal PDF of Y_2

y_2	$f(y_2)$
2	0.60
3	0.18
4	0.22

Marginal PDF of Y_2 conditional on $Y_1 = 3$

y_2	$f(y_2)$
2	0.625
3	0.115
4	0.25

Marginal PDF of Y_2 is not equal to the conditional PDF of Y_2 conditional on $Y_1 = 3$.

(c) $E[U] = 6.82$; $E[U|Y_1 = 3] = 7.875$.

Solution – Exercise 2.13

(a) $Pr(Y = 1) = 0.25$
(b) $Pr(X = 2) = 0.6$
(c) $Pr(X = 2|Y = 1) = 0.4$
(d) $Pr(Y = 1|X = 1) = 0.375$
(e) $E[X] = 1.6$
(f) $E[Y] = 2.05$
(g) $E[XY] = 3.25$
(h) $E[X|Y = 1] = 1.4$
(i) $E[a + bX] = 6.8$
(j) $Var[X] = 0.24$
(k) $Var[Y] = 0.5475$
(l) $Cov[X, Y] = -0.03$
(m) $Var[aX + bY] = 5.5275$
(n) $Cov[a + bX, cY] = -0.36$

Solution – Exercise 2.15

(a) $E[Y_2|Y_1 = 4] = 4.375$ and $var(Y_2|Y_1 = 4) = 0.734375$.

(b) Statistical independence implies that marginal and conditional PDFs are identical. Showing that $f(Y_2) \neq f(Y_2|Y_1 = 4)$ is sufficient to conclude that Y_1 and Y_2 are not independent.

Solution – Exercise 2.16

$$P^2 = \begin{bmatrix} p_{11}^2 + p_{12}p_{21} & p_{11}p_{12} + p_{12}p_{22} \\ p_{21}p_{11} + p_{22}p_{21} & p_{21}p_{12} + p_{22}^2 \end{bmatrix}$$

Solution – Exercise 2.17

$$E[\Upsilon] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$var(\Upsilon) = \begin{bmatrix} 1.5 & 0 \\ 0 & 2.25 \end{bmatrix}$$

3 Difference Equations

Many economic models give rise to difference equations. It is therefore important to know some of the techniques available to solve them. We restrict our attention to first-order and second-order difference equations. This choice is not as restrictive as it may seem given that a very large number of models generate first or second-order difference equations.

When dealing with difference equations it is implicitly assumed that time is discrete.[12] That is, the "time variable" t takes values like $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$. In general, t can run from $-\infty$ to $+\infty$. You can think of t as indexing months or quarters. For example, $t = 0$ could represent the first quarter of 1947, $t = 1$ the second quarter of 1947 and so on.

The **general form of an nth-order linear difference equation** in the variable y is

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + x_t \tag{114}$$

where $a_0, \ldots, a_n$ are constant coefficients and x_t is called a forcing process and can be any function of time, current and lagged values of other variables, and/or stochastic disturbances.

In solving a difference equation, the objective is to find a time path for the variable y. This time path is represented by the sequence $\{y_t\}_{t=0}^{\infty}$ (often simply denoted $\{y_t\}$). The time path we are looking for should be a function of t (*i.e.* a formula defining the values of y in every time period) which is consistent with the given difference equation as well as with its initial conditions. Also, the time path must not contain any difference expression such as Δy_t. Note that I denote **first differences** as $\Delta y_t \equiv y_t - y_{t-1}$. You should be careful about the Δ notation since some authors prefer defining first differences as $\Delta y_t \equiv y_{t+1} - y_t$.

Most of this chapter discusses **linear** difference equations. In a linear difference equation, no y term (of any period) is raised to a power different from unity or is multiplied by a y term of another period.

3.1 First-Order Difference Equations
3.1.1 Solution by Iteration

If we are provided with a first-order difference equation and an **initial condition** (*i.e.* a known value for y_0), then we can simply iterate forward from that initial period to calculate

---

[12]When time is continuous, we use the expression differential equation.

the entire time path for y. Consider the linear first-order stochastic difference equation

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t \tag{115}$$

where ϵ_t is a random variable. Given an initial condition y_0 and a realization for ϵ_1, we can calculate y_1 by setting $t = 1$ in (115)

$$y_1 = a_0 + a_1 y_0 + \epsilon_1. \tag{116}$$

Similarly, setting $t = 2$ in (115) and using (116) we have

$$y_2 = a_0 + a_1 y_1 + \epsilon_2 = a_0 + a_1[a_0 + a_1 y_0 + \epsilon_1] + \epsilon_2 = a_0(1 + a_1) + a_1^2 y_0 + a_1 \epsilon_1 + \epsilon_2. \tag{117}$$

One more iteration yields

$$y_3 = a_0 + a_1 \left[a_0(1 + a_1) + a_1^2 y_0 + a_1 \epsilon_1 + \epsilon_2 \right] + \epsilon_3 \tag{118}$$

$$= a_0(1 + a_1 + a_1^2) + a_1^3 y_0 + a_1^2 \epsilon_1 + a_1 \epsilon_2 + \epsilon_3.$$

The solutions for y_1, y_2 and y_3 above make clear that further iterations will yield

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \epsilon_{t-i} \tag{119}$$

Consider a first numerical example where an initial condition y_0 is given, the parameters are $a_0 = 0$ and $a_1 = 0.95$ and finally $x_t = 0$ for all t. The equation to use to calculate the time path for y is then

$$y_t = 0.95 y_{t-1}. \tag{120}$$

Again, let's iterate on (120) three times to identify a pattern we can generalize. A first iteration yields $y_1 = 0.95 y_0$. A second iteration yields $y_2 = 0.95 y_1 = 0.95^2 y_0$ and a third iteration yields $y_3 = 0.95 y_2 = 0.95^3 y_0$. Clearly, the solution to the deterministic homogeneous linear first-order difference equation (120) is $y_t = 0.95^t y_0$.

As a second example, consider the (non-homogeneous) difference equation

$$y_t = 1 + 0.95 y_{t-1}. \tag{121}$$

where an initial condition y_0 is provided. A first iteration yields $y_1 = 1 + 0.95 y_0$. A second iteration yields

$$y_2 = 1 + 0.95 y_1 = 1 + 0.95[1 + 0.95 y_0] = 1 + 0.95 + 0.95^2 y_0.$$

A third iteration yields

$$y_3 = 1 + 0.95y_2 = 1 + 0.95[1 + 0.95 + 0.95^2 y_0] = 1 + 0.95 + 0.95^2 + 0.95^3 y_0.$$

After t iterations we have the solution

$$y_t = 0.95^t y_0 + \sum_{i=0}^{t-1} 0.95^i = 0.95^t y_0 + \frac{1 - 0.95^t}{1 - 0.95} \tag{122}$$

where the last equality holds because $-1 < 0.95 < 1$. Figure 3.1 plots the time path described in equation (122), starting from the initial condition $y_0 = 10$. We see that the time path eventually converges to an intertemporal equilibrium where $y_t = 20$.

Figure 3.1: Time path of y
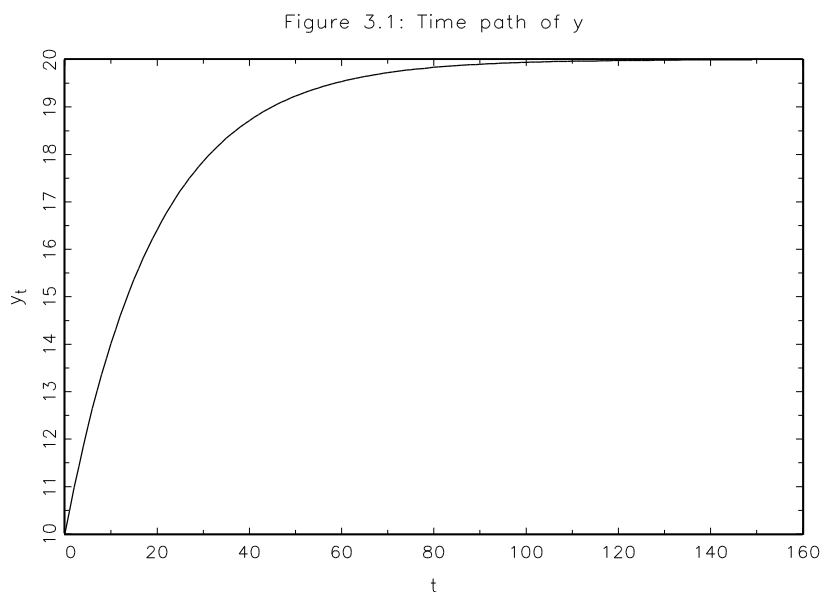
3.1.2 A General Method to Solve n^{th}-Order Difference Equations

Solving higher-order difference equations by the iterative method is difficult because the algebra rapidly becomes very complex. Therefore, a more general method is needed. In this section, we work through the four steps involved in solving linear difference equations. The four steps below explain how the method works to solve an nth-order linear difference equation

Step 1: Form the homogeneous equation to find all n **homogeneous solutions**.

Step 2: Find a **particular solution**.

Step 3: Obtain the **general solution** as the sum of the particular solution and a linear combination of all homogeneous solutions.

Step 4: Eliminate the arbitrary constant(s) by imposing the initial condition(s) on the general solution to get the **definite solution**.[13]

The **particular solution** is *any* solution of the complete nonhomogeneous equation (114). The **homogeneous solution**, denoted y^h, is the general solution of the homogeneous version of (114) given by

$$y_t = \sum_{i=i}^{n} a_i y_{t-i}. \tag{123}$$

The particular solution represents the intertemporal equilibrium level of y while the combination of homogeneous solutions represents the deviations of the time path from that equilibrium.

To see how the method works we apply it to solve the first-order linear difference equation (FOLDE)

$$y_{t+1} = a_0 + a_1 y_t. \tag{124}$$

In step 1, we find the solution to the homogeneous equation

$$y_{t+1} = a_1 y_t. \tag{125}$$

We guess that the solution to (125) is of the form

$$y_t = A\alpha^t \tag{126}$$

where A and α are unknown coefficients. If our guess is correct it must also be the case that $y_{t+1} = A\alpha^{t+1}$. We find the value of α by substituting the guess on both sides of (125) to get

$$A\alpha^{t+1} = a_1 A\alpha^t. \tag{127}$$

Dividing through by $A\alpha^t$ yields $\alpha = a_1$. Using this result in our guess (126) implies that the homogeneous solution should be written as

$$y_t^h = Aa_1^t. \tag{128}$$

---

[13]Note that some authors refer to the "homogeneous solutions" as the "complementary function", to the "particular solution" as the "particular integral" and to the "general solution" as the "arbitrary solution."

In step 2, we have to find a particular solution to equation (124). Remember that the particular solution is *any* solution to the complete nonhomogeneous equation. Therefore, let us use a simple trial solution where y is constant over time. That is $y_t = c$ for all t. Using this trial solution in (124) we can solve for c

$$c = a_0 + a_1 c \quad \Rightarrow \quad c = \frac{a_0}{1 - a_1} \tag{129}$$

when $a_1 \neq 1$. Therefore, the particular solution can be written as

$$y^p = \frac{a_0}{1 - a_1}, \qquad (a_1 \neq 1). \tag{130}$$

Obviously if $a_1 = 1$, then $a_0/(1 - a_1)$ is not defined and some other particular solution must be found. When $a_1 = 1$, we use the trial solution $y_t = ct$. Substituting this trial solution on both sides of (124) we can solve for c

$$c(t + 1) = a_0 + a_1 ct, \quad \Rightarrow \quad c = \frac{a_0}{1 + t - a_1 t}. \tag{131}$$

Given that $a_1 = 1$ we find $c = a_0$ and therefore, the particular solution is

$$y_t^p = \begin{cases} a_0/(1 - a_1), & \text{if } a_1 \neq 1 \\ a_0 t, & \text{if } a_1 = 1 \end{cases}. \tag{132}$$

In **step 3** we combine homogeneous and particular solutions to obtain the general solution. In the present case the general solution is written as

$$y_t = \begin{cases} A a_1^t + \frac{a_0}{1-a_1}, & \text{if } a_1 \neq 1 \\ A + a_0 t, & \text{if } a_1 = 1 \end{cases} \tag{133}$$

In **step 4** we find the value of the unknown coefficient A using the general solution and the initial condition y_0. Setting $t = 0$ in the general solution (133) yields

$$y_0 = A + \frac{a_0}{1 - a_1} \quad \Rightarrow \quad A = y_0 - \frac{a_0}{1 - a_1}, \qquad a_1 \neq 1 \tag{134}$$

$$y_0 = A \quad \Rightarrow \quad A = y_0, \qquad a_1 = 1. \tag{135}$$

Using these results in the general solution (133) yields the **definite solution**

$$y_t = \begin{cases} \left(y_0 - \frac{a_0}{1-a_1}\right) a_1^t + \frac{a_0}{1-a_1}, & \text{if } a_1 \neq 1 \\ y_0 + a_0 t, & \text{if } a_1 = 1 \end{cases} \tag{136}$$

Let us apply the four-step general method to a numerical example. Suppose that the first-order difference equation we want to solve and the initial condition on y are

$$y_{t+1} = 1 + 0.5 y_t, \qquad y_0 = 4. \tag{137}$$

We first proceed to solve the homogeneous version of (137) which is

$$y_{t+1} = 0.5y_t. \tag{138}$$

We guess that a solution to (138) is $y_t = A\alpha^t$. Substituting the guess in (138) yields

$$A\alpha^{t+1} = 0.5A\alpha^t \quad \Rightarrow \quad \alpha = 0.5 \quad \Rightarrow \quad y_t^h = A0.5^t. \tag{139}$$

Second, we find a particular solution to (137). Since $a_1 \neq 1$ in this example, we guess that y is constant over time, $y_t = c$. Substituting this guess in (137) yields

$$c = 1 + 0.5c \quad \Rightarrow \quad c = 2 \quad \Rightarrow \quad y^p = 2. \tag{140}$$

Third, we combine the homogeneous and particular solution to get the general solution

$$y_t = A0.5^t + 2. \tag{141}$$

Fourth, we use the initial condition $y_0 = 4$ to identify A. Setting $t = 0$ in the general solution (141) yields

$$4 = y_0 = A + 2 \quad \Rightarrow \quad A = 2. \tag{142}$$

Therefore, the solution to (137) is

$$y_t = 2 \times 0.5^t + 2. \tag{143}$$

3.1.3   Dynamic Stability of a FOLDE

Recall that the particular solution to a difference equation represents the intertemporal equilibrium level of y while the homogeneous solution represents the deviations of the time path from that equilibrium. Therefore, it is not surprising that the dynamic stability of the equilibrium depends on whether or not the homogeneous solution tends to zero as $t \rightarrow \infty$. The value of α in the homogeneous solution $A\alpha^t$ governs the time path of $A\alpha^t$ as $t \rightarrow \infty$.

We consider three cases. First, if $|\alpha| < 1$ then the expression α^t converges to zero as $t \rightarrow \infty$. In the case where $-1 < \alpha < 0$, the term α^t change signs every period. In such a case, we say that the homogeneous solution displays **oscillations**.

Second, if $|\alpha| > 1$ then the expression α^t increases with t and the homogeneous solution is explosive. In the special case where $\alpha < -1$, the homogeneous solution oscillates explosively.

Third, we have the case where $|\alpha| = 1$. When $\alpha = 1$ the homogeneous solution is always equal to A. This may lead one to conclude that the time path of y will converge since the homogeneous solutions does not explode. It is not the case however. Note that the time path of y will converge to $A + y^p$ and *not* to its long run equilibrium level y^p (unless $A = 0$). Therefore, if we want the time path of y to converge to the equilibrium y^p, we have to rule out the case $\alpha = 1$. Similarly, the case $\alpha = -1$ must also be ruled out.

In summary, the solution $y_t = A\alpha^t + y^p$ is a convergent path if and only if $|\alpha| < 1$.

Exercise 3.1 – A Dynamically Stable Case

Apply the four-step method to solve the following difference equation

$$y_{t+1} = 1 + 0.95y_t, \qquad y_0 = 25.$$

Exercise 3.2 – Unit Root

Apply the four-step method to solve the following difference equation

$$y_{t+1} = 2 + y_t, \qquad y_0 = 5.$$

3.1.4   Linearizing Nonlinear Difference Equations

Not all economic models lead to linear difference equations. There are a couple of ways to deal with nonlinear difference equations. Sometimes a simple operation transforms a nonlinear difference equation in a linear one. For example, suppose the time path of GDP (denoted by y) is described by

$$y_{t+1} = b_0 y_t^{a_1}. \tag{144}$$

If we are willing to study the dynamic properties of the logarithm of GDP instead of GDP, then we can take logs on both sides of (144) to get

$$\ln(y_{t+1}) = \ln(b_0) + a_1 \ln(y_t) \tag{145}$$

Defining $g_t = \ln(y_t)$, $a_0 = \ln(b_0)$ we end up with a linear difference equation in g

$$g_{t+1} = a_0 + a_1 g_t. \tag{146}$$

Obviously, not all nonlinear difference equations have the log-linear form of (144). For example, if we add a constant term so that $y_{t+1} = b_0 y_t^{a_1} + b_1$, then we cannot take logs to "linearize" this nonlinear difference equation.

A strategy used regularly in macroeconomics is to approximate nonlinear equations by linear ones before solving them. Taking a linear approximation yields a linear difference equation that can be solved and studied in the way presented in the previous sections. To present the basic idea we focus here on a first-order approximation.

According to **Taylor's Theorem**, given an arbitrary function $f(y)$, if we know the value of the function at $y = \bar{y}$ and the value of its derivatives at $\bar{y}$ then

$$f(y) = \left[\frac{f(\bar{y})}{0!} + \frac{f'(\bar{y})}{1!}(y - \bar{y}) + \frac{f''(\bar{y})}{2!}(y - \bar{y})^2 + \ldots + \frac{f^{(n)}(\bar{y})}{n!}(y - \bar{y})^n \right] + R_n \tag{147}$$

where n denotes the order of the approximation, $f^{(n)}$ denotes the nth derivative of $f(-)$ and R_n is the remainder term (you can think of it as an approximation error).

When we use a first-order (hence linear) approximation we set $n = 1$ and drop the remainder term to write

$$f(y) \approx f(\bar{y}) + f'(\bar{y})(y - \bar{y}). \tag{148}$$

where we used the fact that $0! = 1$ and $1! = 1$.

Evidently, when we solve difference equations we have at least two variables (taking time into account) appearing in the function to approximate (*e.g.* y_t and y_{t+1}) so we have to adapt the linear approximation to this scenario. Suppose that function $f(-)$ actually depends on y_t and y_{t+1} as it would in the example

$$y_{t+1} = b_1 y_t + b_2 y_t^{b_3}. \tag{149}$$

Then the first-order approximation formula becomes

$$f(y_t, y_{t+1}) \approx f(\bar{y}, \bar{y}) + f_1(\bar{y}, \bar{y})(y_t - \bar{y}) + f_2(\bar{y}, \bar{y})(y_{t+1} - \bar{y}) \tag{150}$$

where the approximation point is $(y_t, y_{t+1}) = (\bar{y}, \bar{y})$ and $f_i(-)$ denotes the partial derivative of $f(-)$ with respect to its ith argument. The natural approximation point is the long run

equilibrium value of y which is usually not that hard to find as long as the difference equation we are solving is stationary. When it is not, then some intermediate steps must be taken to arrive at a stationary nonlinear difference equation. Here we sidestep this issue by assuming that the the equation we want to study (*e.g.* (149)) is stationary.

To see how the first-order linear approximation method works we approximate (149)). First, we define the $f(-)$ function which can be written

$$f(y_t, y_{t+1}) \equiv b_1 y_t + b_2 y_t^{b_3} - y_{t+1}. \tag{151}$$

With this definition, equation (149) is written $f(y_t, y_{t+1}) = 0$.

Second, we take the partial derivatives

$$f_1(y_t, y_{t+1}) = b_1 + b_2 b_3 y_t^{b_3-1}, \qquad f_2(y_t, y_{t+1}) = -1. \tag{152}$$

Third, we apply formula (150)

$$f(y_t, y_{t+1}) \approx \underbrace{[(b_1 - 1)\bar{y} + b_2\bar{y}^{b_3}]}_{f(\bar{y},\bar{y})} + \underbrace{[b_1 + b_2 b_3 \bar{y}^{b_3-1}]}_{f_1(\bar{y},\bar{y})}(y_t - \bar{y}) + \underbrace{[-1]}_{f_2(\bar{y},\bar{y})}(y_{t+1} - \bar{y}) \tag{153}$$

Therefore, our approximation of $f(y_t, y_{t+1}) = 0$ is

$$[(b_1 - 1)\bar{y} + b_2\bar{y}^{b_3}] + [b_1 + b_2 b_3 \bar{y}^{b_3-1}](y_t - \bar{y}) + [-1](y_{t+1} - \bar{y}) = 0. \tag{154}$$

Actually, given that $f(-)$ is defined in such a way that the nonlinear equation being approximated is written $f(y_t, y_{t+1}) = 0$, it is always the case that $f(\bar{y}, \bar{y}) = 0$ which implies that the first term in the equation above is exactly zero since it is exactly equal to $f(\bar{y}, \bar{y})$ as pointed out in (153). Taking this into account our approximation of $f(y_t, y_{t+1}) = 0$ is

$$[b_1 + b_2 b_3 \bar{y}^{b_3-1}](y_t - \bar{y}) + [-1](y_{t+1} - \bar{y}) = 0. \tag{155}$$

Define $a_1 \equiv b_1 + b_2 b_3 \bar{y}^{b_3-1}$ and the variables in deviations from their long-run equilibrium level as $g_t \equiv y_t - \bar{y}$ (and $g_{t+1} \equiv y_{t+1} - \bar{y}$). Then our linearized difference equation is written

$$g_{t+1} = a_1 g_t \tag{156}$$

which has a familiar form.

It is natural to wonder how we actually get the value $\bar{y}$ we are using in the approximation. Remember that we are assuming a stationary nonlinear difference equation. Therefore we

know that in the long-run $y_t = y_{t+1}$. We then solve for $\bar{y}$ by using $y_t = y_{t+1} = \bar{y}$ in (149) which yields

$$\bar{y} = \left[\frac{1 - b_1}{b_2} \right]^{\frac{1}{b_3 - 1}} \tag{157}$$

As an exercise, approximate the nonlinear difference equation

$$y_{t+1} = 0.9 y_t + y_t^{0.5} \tag{158}$$

to find the equivalent of (156) and then solve.

3.1.5 Nonlinear Difference Equations: a Graphical Approach

When we want to study a nonlinear difference equation without approximating it we can simply calculate the time path of y by iterating forward on the nonlinear difference equation starting from an initial condition y_0. However, when we do that we may not be able to figure out the convergence properties of the time path (convergent, divergent, oscillatory). To figure them out, we make use of a graphical tool: the phase diagram.

To explain what a phase diagram is, let us denote the nonlinear relationship between y_{t+1} and y_t as $y_{t+1} = f(y_t)$. For example in equation (144), $f(y_t) = b_0 y_t^{a_1}$. A **phase diagram** plots a **phase line** corresponding to $f()$ together with a 45° line in the (y_{t+1}, y_t)-plan. Figure 3.2 plots four phase diagrams. In panels (a) and (b), the phase lines have positive slopes (everywhere) while they have negative slopes (everywhere) in the other two panels. Starting from an initial condition y_0 on the y_t-axis, we use the phase line to find y_1, y_2 and so on. This is equivalent to iterating forward on the nonlinear difference equation starting from y_0.

In Figure 3.2a, we see that the time path of y will converge to $\bar{y}$ where $\bar{y}$ represents the intertemporal equilibrium level of y and is found at the intersection of the phase line and the 45° line.[14] Note that the time path of y converges to $\bar{y}$ whether $y_0 < \bar{y}$ or $y_0 > \bar{y}$.

In Figure 3.2b, we see that the time path of y diverges from its equilibrium level $\bar{y}$.[15] For example, if $y_0 > \bar{y}$ then we have $y_0 < y_1 < y_2 < y_3 < y_4 < \ldots$ and the time path will tend to infinity (divergent path) as $t \to \infty$. If $y_0 < \bar{y}$, then the time path is also divergent but in the other direction.

---

[14]Recall that, by definition, $y_{t+1} = y_t$ along the 45° line.

[15]Unless the initial condition turns out to be exactly $y_0 = \bar{y}$.

In Figure 3.2c, the time path converges to $\bar{y}$ but is oscillatory. The figure illustrates the phenomenon of **overshooting**. That is, if we start with $y_0 < \bar{y}$, the time path for y will overshoot its equilibrium level as it converges to it. This is seen in the case illustrated in Figure 3.2c since starting from $y_0 < \bar{y}$ implies $y_1 > \bar{y}$.

Finally, Figure 3.2d displays an example of an oscillatory and divergent time path.
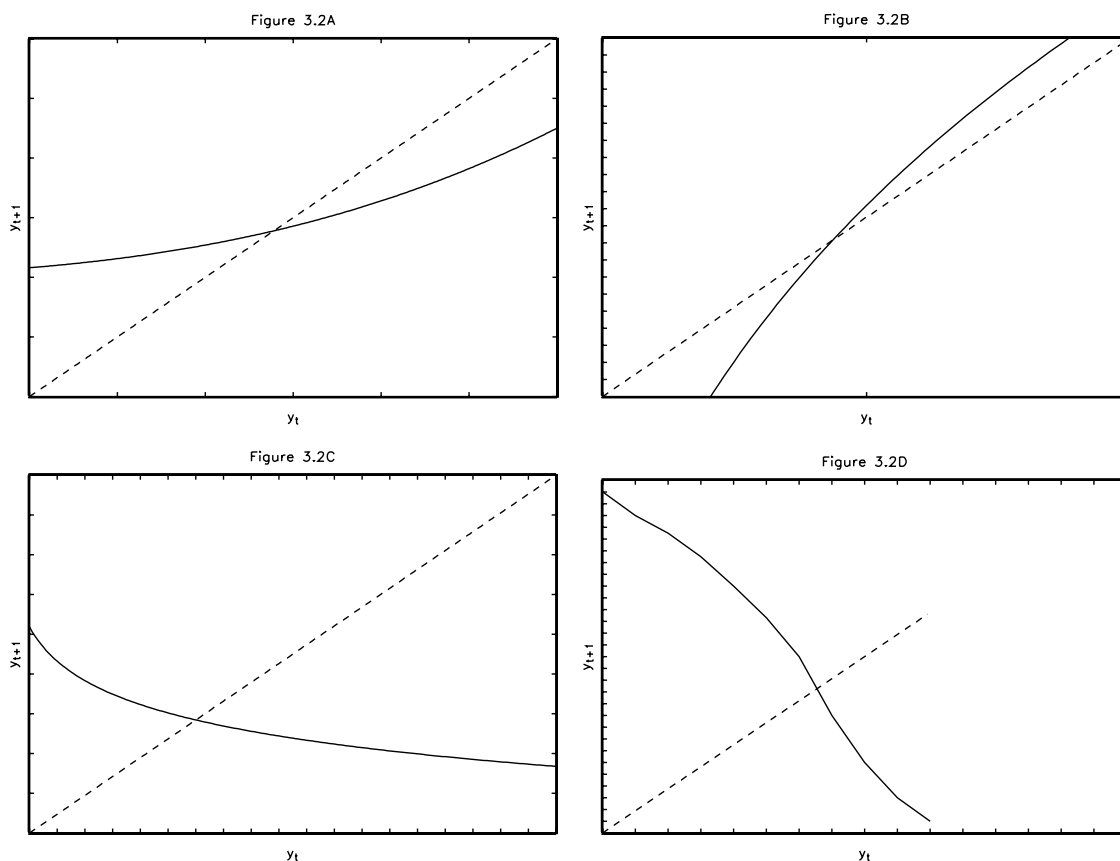
As you have most likely realized by now, the convergence properties of the time path described by a nonlinear difference equation depend on the absolute value of the slope of $f()$ while the algebraic sign of the slope determines whether there will be oscillations or not. When the absolute value of the slope is less than unity (as in Figures 3.2a and 3.2c) then the time path is convergent. When the slope is negative, then the time path is oscillatory.

In general, the phase line may have positively or negatively sloped segments. It may also have segments where the absolute value of its slope is less than unity and others where it is greater than unity. In these cases, the time path will be more complicated but it can still be analyzed using a phase diagram. In these cases, different starting values may imply time

paths that are widely different.

3.1.6 Stochastic Difference Equations and the Method of Undetermined Coefficients

The discussion in subsection 3.1.2 was restricted to deterministic linear difference equations. This section outlines *the method of undetermined coefficients which is useful to find particular solutions when difference equations include stochastic components.* To keep the exposition relatively clear, we will restrict our attention to solving the first-order stochastic difference equation

$$y_{t+1} = a_0 + a_1 y_t + \epsilon_{t+1} \tag{115'}$$

using an initial condition y_0 and where $|a_1| < 1$.

To solve equation (115') we follow the four steps presented in subsection 3.1.2. The first step is to find the homogeneous solution by solving

$$y_{t+1} = a_1 y_t. \tag{159}$$

As usual we guess that the homogeneous solution has the form $y_t = A\alpha^t$. Substituting the guess in the equation above yields $\alpha = a_1$. Therefore the homogeneous solution is $A a_1^t$ where the constant A will be determined in step 4.

In step 2 we have to find a particular solution to equation (115'). This is where the method of undetermined coefficients comes in. Contrary to the deterministic case, it does not make sense to choose $y_t = c$ (y constant over time) as a trial solution since y depends on a random variable (ϵ) which is not constant over time. Instead we use the trial solution

$$y_t = c + \sum_{i=0}^{\infty} \beta_i \epsilon_{t-i} \tag{160}$$

where c and $\beta_0, \beta_1, \ldots$ are undetermined coefficients. If the trial solution is correct, then it must be the case that

$$y_{t+1} = c + \sum_{i=0}^{\infty} \beta_i \epsilon_{t+1-i}. \tag{161}$$

Substituting the trial solution (160)-(161) into (115') yields

$$c + \beta_0 \epsilon_{t+1} + \beta_1 \epsilon_t + \beta_2 \epsilon_{t-1} + \beta_3 \epsilon_{t-2} + \ldots \tag{162}$$

$$= a_0 + a_1 \left[c + \beta_0 \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \ldots \right] + \epsilon_{t+1}$$

Re-arranging (162) to collect like terms, we obtain

$$(c - a_0 - c\,a_1) + (\beta_0 - 1)\epsilon_{t+1} + (\beta_1 - a_1\beta_0)\epsilon_t + (\beta_2 - a_1\beta_1)\epsilon_{t-1} + \ldots = 0 \tag{163}$$

Since equation (163) must hold for *all* possible values of the $\{\epsilon_t\}$ sequence, each of the following conditions must hold

$$\beta_0 - 1 = 0, \quad \beta_1 - a_1\beta_0 = 0, \quad \beta_2 - a_1\beta_1 = 0, \quad \ldots \tag{164}$$

$$c - a_0 - c\,a_1 = 0 \tag{164a}.$$

We can easily solve for the β_i's recursively from (164): $\beta_0 = 1$, $\beta_1 = a_1$, $\beta_2 = a_1^2$, ..., $\beta_i = a_1^i$. Then from (164a) we get $c = a_0/(1 - a_1)$ (recall that we assumed $|a_1| < 1$). Using these results in the trial solution (160) yields the particular solution

$$y_t^p = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i}. \tag{165}$$

In step 3 we simply combine the homogeneous and particular solutions $(y_t = y^h + y^p)$

$$y_t = A a_1^t + \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i}. \tag{166}$$

In the last step, we use the initial condition y_0 to identify A. Setting $t = 0$ in the general solution (166) yields

$$y_0 = A + \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \epsilon_{-i} \tag{167}$$

which implies that

$$A = y_0 - \frac{a_0}{1 - a_1} - \sum_{i=0}^{\infty} a_1^i \epsilon_{-i}. \tag{168}$$

Therefore the definite solution to equation (115) is

$$y_t = \left[y_0 - \frac{a_0}{1 - a_1} - \sum_{i=0}^{\infty} a_1^i \epsilon_{-i} \right] a_1^t + \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i} \tag{169}$$

which after some algebra can be written as

$$y_t = \left[y_0 - \frac{a_0}{1 - a_1} \right] a_1^t + \frac{a_0}{1 - a_1} + \sum_{i=0}^{t-1} a_1^i \epsilon_{t-i} \tag{170}$$

As you can see, the approach followed to solve a stochastic difference equation resembles that used to solve a deterministic difference equation. The main difference lies in the identification of the particular solution (step 2).

3.1.7 Forward and Backward-Looking Solutions

The solution (170) to the stochastic difference equation (115') is what we call a backward-looking solution since it involves *past values of the random variables*. Notice that the solution (119) to (115) obtained by backward repeated substitution is also a backward-looking solution. Recall that the backward repeated substitution method works as follows: start with equation (115), then substitute out y_{t-1} (which introduces a term in y_{t-2}), then substitute out y_{t-2} (which introduces a term in y_{t-3}), and so on.

We can also use a forward repeated substitution method which yields a forward-looking solution. Instead of substituting out lags of y_t, we substitute out leads of y_t. For example, we can rewrite equation (115) as

$$y_{t-1} = -\frac{a_0}{a_1} - \frac{\epsilon_t}{a_1} + \frac{y_t}{a_1}. \tag{171}$$

Updating by one period to get y_t on the left-hand side yields

$$y_t = -\frac{a_0}{a_1} - \frac{\epsilon_{t+1}}{a_1} + \frac{y_{t+1}}{a_1}. \tag{172}$$

To iterate forward on equation (172) we use the fact that $y_{t+1} = (y_{t+2} - a_0 - \epsilon_{t+2})/a_1$ to substitute out y_{t+1} and get

$$y_t = -\frac{a_0}{a_1} - \frac{\epsilon_{t+1}}{a_1} + \frac{1}{a_1}\left[\frac{y_{t+2}}{a_1} - \frac{a_0 + \epsilon_{t+2}}{a_1}\right]. \tag{173}$$

Re-arranging we have

$$y_t = -\frac{a_0}{a_1} - \frac{a_0}{a_1{}^2} - \frac{\epsilon_{t+1}}{a_1} - \frac{\epsilon_{t+2}}{a_1^2} + \frac{y_{t+2}}{a_1^2}. \tag{174}$$

Continuing this iterative process yields

$$y_t = -a_0 \sum_{i=1}^{n} a_1^{-i} - \sum_{i=1}^{n} a_1^{-i}\epsilon_{t+i} + \frac{y_{t+n}}{a_1^n}. \tag{175}$$

Clearly, if $|a_1| > 1$, the first term in (175) converges to $a_0/(1 - a_1)$ while the last term converges to zero as n tends to infinity. Therefore, a particular solution is

$$y_t = \frac{a_0}{1 - a_1} - \sum_{i=1}^{\infty} a_1^{-i}\epsilon_{t+i}. \tag{176}$$

As explained in the discussion around equation (159) the homogeneous solution is $y_t = Aa_1^t$ where A is an unknown constant to be identified in step (4) using the initial condition y_0.

It is worth taking some time to discuss some properties the forward-looking particular solution (176) and of the backward-looking particular solution

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i\epsilon_{t-i}. \tag{177}$$

First, the backward-looking solution depends on past realizations of the random variables while the forward-looking solution depends on future realizations of the random variables. Since these future realizations are not observed, it may seem that the forward-looking solution is of limited interest. As you will see, forward-looking solutions are encountered when solving rational expectations models. In such models, current behaviour of an economic agent depends on its expectation of future events. For example, a consumer's decision regarding consumption/savings in period t most likely depends on its expectations of future income. In such a case, the random variables ϵ_{t+1}, ϵ_{t+2}, ... appearing in the forward-looking particular solution (176) would be replaced by their expectations/forecasts conditional on the information known in period t denoted $E_t\epsilon_{t+1}$, $E_t\epsilon_{t+2}$, ...

Second, let us see how the two solutions depend on a_1. Consider the backward-looking solution first. If $|a_1| < 1$, the expression a_1^i converges toward zero as $i \to \infty$ which implies that the effects of *past* random variables diminishes over time when $|a_1| < 1$. This can be interpreted two ways: (1) the effect of ϵ_{t-2} on y_t is smaller than the effect of ϵ_{t-1} on y_t; (2) the effect of ϵ_{t-1} on y_t is larger than its effect on y_{t+1}. However, if $|a_1| > 1$ then the backward-looking solution explodes since the expression a_1^i converges to infinity as $i \to \infty$ (see the discussion in subsection 3.1.3).

The image is reversed when dealing with the forward-looking solution. If $|a_1| < 1$, then the expression a_1^{-i} tends to infinity as $i \to \infty$ and the forward-looking solution explodes. If $|a_1| > 1$, then the expression a_1^{-i} tends to zero as $i \to \infty$ and the forward-looking solution yields a finite sequence $\{y_t\}$. When $|a_1| > 1$, the effects of *future* random variables diminishes over time. This can be interpreted two ways: (1) the effect of ϵ_{t+2} on y_t is smaller than the effect of ϵ_{t+1} on y_t; (2) the effect of ϵ_{t+2} on y_{t+1} is larger than its effect on y_t.

From a purely mathematical point of view, we cannot say whether the forward-looking or backward-looking solution is most appropriate. However economic theory often suggests that a sequence be bounded (*i.e.* the elements of the sequence cannot tend to infinity).[16] In such a case, we use the forward-looking solution when $|a_1| > 1$ and the backward-looking solution when $|a_1| < 1$. In higher order difference equations where we may have stable (inside the unit circle) and unstable roots (outside the unit circle) we often solve stable roots backwards and unstable roots forwards.

[16]For example, we do not expect real returns on assets, real GDP per capita and several other economic variables to tend to (plus or minus) infinity.

3.2 Vector First-Order Difference Equations

A vector difference equation differs from the difference equations we have seen above in that the equation involves a vector of variables ($\boldsymbol{y}_t$ say) rather that a scalar variable y_t. For example, a vector stochastic first-order difference equation in the vector $\boldsymbol{y}$ could be written

$$\boldsymbol{y}_t = \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \boldsymbol{C} \boldsymbol{w}_t \tag{178}$$

where $\boldsymbol{y}_t$ is an $n \times 1$ vector of (state) variables, $\boldsymbol{A}_1$ is a $n \times n$ matrix, $\boldsymbol{C}$ is a $n \times m$ matrix and $\boldsymbol{w}_t$ is an $m \times 1$ vector of exogenous random variables. As you can see, this vector difference equation is quite similar to the scalar difference equation (115). There are several possible assumptions about the properties of the forcing process $\boldsymbol{w}_t$. We could assume that it is an *i.i.d* process with a specific distribution (*e.g.* multivariate standard normal). A weaker assumption would be $E[\boldsymbol{w}_{t+1}|J_t] = 0$ and $E[\boldsymbol{w}_{t+1}\boldsymbol{w}_{t+1}'|J_t] = I$ where J_t is an information set defined as $J_t = \begin{bmatrix} \boldsymbol{w}_0 & \boldsymbol{w}_1 & \boldsymbol{w}_2 & ... & \boldsymbol{w}_t \end{bmatrix}$.

Suppose an economic model also includes variables that are not part of the state vector $\boldsymbol{y}_t$ appearing in (178) but that are linear combinations of the elements of $\boldsymbol{y}_t$. Let's group these other variables in the vector $\boldsymbol{u}_t$ where

$$\boldsymbol{u}_t = \boldsymbol{G} \boldsymbol{y}_t \tag{179}$$

Equations (178)-(179) form what we call a **linear state-space system**.

Evidently, once a solution/time path $\{\boldsymbol{y}\}_{t=0}^{\infty}$ is identified by solving (178), then we can easily get the time path $\{\boldsymbol{u}\}_{t=0}^{\infty}$ using (179).

Methods to solve vector stochastic linear difference equations exist but are beyond the scope of this math review. Once one knows how to use these methods, then solving an nth-order scalar stochastic linear difference equation is fairly easy since such a difference equation can be written as a vector first-order stochastic linear difference equation. For example, consider the scalar stochastic second-order linear difference equation

$$z_{t+1} = a_0 + a_1 z_t + a_2 z_{t-1} + \epsilon_{t+1}. \tag{180}$$

Defining the following vectors and matrices

$$\boldsymbol{y}_{t+1} = \begin{bmatrix} z_{t+1} \\ z_t \\ 1 \end{bmatrix}, \quad \boldsymbol{A}_1 = \begin{bmatrix} a_1 & a_2 & a_0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{w}_{t+1} = \epsilon_{t+1} \tag{181}$$

we re-write (180) in the format of (178).

3.3   Second-Order Difference Equations

A second-order deterministic linear difference equation is a special case of equation (114) where $n = 2$ and the variable x_t is omitted

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2}. \tag{182}$$

Since the timing convention is arbitrary, we can also write it as

$$y_{t+2} = a_0 + a_1 y_{t+1} + a_2 y_t. \tag{183}$$

The important point to understand is that **a second-order difference equation is defined as one involving a two-period lag in the variable.**[17]

To solve a second-order difference equation we apply the four-step method described in subsection 3.1.2. That is, we need a particular solution (representing the intertemporal equilibrium level of y) and homogeneous solutions (representing deviations from the equilibrium for every period) to derive a general solution. Finding a particular solution to a second-order difference equation is quite similar to finding a particular solution to a first-order difference equation. For this reason, we cover this topic first even though it corresponds to the second step of our four-step solution method.

3.3.1   Particular Solution

As in the first-order case, we posit a trial solution where the variable is constant $y_{t+2} = y_{t+1} = y_t = c$. Using this trial solution in (183) yields

$$c = a_0 + a_1 \ c + a_2 \ c \quad \Rightarrow \quad c = \frac{a_0}{1 - a_1 - a_2} \tag{184}$$

when $a_1 + a_2 \neq 1$. Therefore, the particular solution can be written

$$y^p = \frac{a_0}{1 - a_1 - a_2}, \qquad (a_1 + a_2 \neq 1). \tag{185}$$

Compare this to solution (130).

When $a_1 + a_2 = 1$ the value of c in (184) is undefined. Actually, the process for y_t has a unit-root when $a_1 + a_2 = 1$. As we saw in section 3.1.3, the process for y is not convergent when there is a unit root. Therefore, it is not surprising that the trial solution $y_t = c$ breaks

---

[17]Therefore, equation (182) still is a second-order difference equation when $a_1 = 0$. Obviously, if $a_2 = 0$, then equation (182) reduces to a first-order difference equation.

down in this case. Instead we should work with the trial solution $y_t = ct$ (as long as $a_1 \neq 2$). Using this trial solution on both sides of (183) yields

$$c(t+2) = a_0 + a_1 c(t+1) + a_2 ct \quad \Rightarrow \quad c = \frac{a_0}{2 - a_1} \tag{186}$$

when $a_1 + a_2 = 1$ and $a_1 \neq 2$. Hence, in this case the particular solution is

$$y_t^p = \frac{a_0}{2 - a_1} t, \qquad (a_1 + a_2 = 1, \quad a_1 \neq 2). \tag{187}$$

If it turns out that $a_1 + a_2 = 1$ and $a_1 = 2$ then we should work with the trial solution $y_t = ct^2$ which yields[18]

$$c(t+2)^2 = a_0 + a_1 c(t+1)^2 + a_2 ct^2 \quad \Rightarrow \quad c = \frac{a_0}{2}. \tag{188}$$

In summary, a particular solution to the second-order difference equation (183) is

$$y^p = \begin{cases} \frac{a_0}{1 - a_1 - a_2}, & (a_1 + a_2 \neq 1) \\ \frac{a_0}{2 - a_1} t, & (a_1 + a_2 = 1, \quad a_1 \neq 2) \\ \frac{a_0}{2} t^2, & (a_1 + a_2 = 1, \quad a_1 = 2) \end{cases} \tag{189}$$

3.3.2   Homogeneous Solutions

To find the homogeneous solutions, we must solve the homogeneous version of equation (183) given by

$$y_{t+2} - a_1 y_{t+1} - a_2 y_t = 0. \tag{190}$$

Not surprisingly, we guess that the solution to (190) is of the form $y_t^h = A\alpha^t$. Using this guess in the homogeneous equation (190) yields

$$A\alpha^{t+2} - a_1 A\alpha^{t+1} - a_2 A\alpha^t = 0. \tag{191}$$

Simplifying yields the quadratic equation (called the **characteristic equation of (182) or (183)**)

$$\alpha^2 - a_1 \alpha - a_2 = 0. \tag{192}$$

This characteristic equation yields two **characteristic roots**

$$\alpha_1 = \frac{a_1 - \sqrt{a_1^2 + 4a_2}}{2}, \qquad \alpha_2 = \frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2}. \tag{193}$$

---

[18]The steps leading to the result in (188) make use of the fact that if $a_1 + a_2 = 1$ and $a_1 = 2$ then $a_2 = -1$.

Each of these two characteristic roots yields a valid solution to the homogeneous equation (190). Again, these solutions are not unique. Actually, for any two arbitrary constants A_1 and A_2, the linear combination $A_1\alpha_1^t + A_2\alpha_2^t$ also solves (190). Three possible situations may be encountered in regard to the characteristic roots, depending on the value of $a_1^2 + 4a_2$.

Case 1 (distinct real roots)
When $a_1^2 + 4a_2 > 0$, there will be two distinct real characteristic roots. In that case, α_1 and α_2 are linearly independent and the homogeneous solution is written

$$y_t^h = A_1 \left(\frac{a_1 - \sqrt{a_1^2 + 4a_2}}{2} \right)^t + A_2 \left(\frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2} \right)^t \tag{194}$$

where A_1 and A_2 are arbitrary constants to be determined in step 4.

Case 2 (repeated real roots)
When $a_1^2 + 4a_2 = 0$, the characteristic roots in (193) are repeated

$$\alpha(= \alpha_1 = \alpha_2) = \frac{a_1}{2}. \tag{195}$$

Therefore, a homogeneous solution is $A_1(\frac{a_1}{2})^t$. However, when $a_1^2 + 4a_2 = 0$, there is a second homogeneous solution given by $A_2 \, t \, (\frac{a_1}{2})^t$. Therefore the complete homogeneous solution is

$$y_t^h = A_1 \left(\frac{a_1}{2} \right)^t + A_2 \, t \, \left(\frac{a_1}{2} \right)^t \tag{196}$$

where A_1 and A_2 are arbitrary constants to be determined in step 4.

Case 3 (complex roots)
When $a_1^2 + 4a_2 < 0$, the characteristic roots are conjugate complex. That is, they will be of the form

$$\alpha_1 = \frac{a_1 + i\sqrt{4a_2 - a_1^2}}{2}, \quad \alpha_2 = \frac{a_1 - i\sqrt{4a_2 - a_1^2}}{2} \tag{197}$$

where $i = \sqrt{-1}$. We can use De Moivre's theorem to write the homogeneous solution as

$$y_t^h = \beta_1 r^t cos(\theta t + \beta_2) \tag{198}$$

where β_1 and β_2 are arbitrary constants, $r = \sqrt{-a_2}$, and the value of θ simultaneously satisfies

$$cos(\theta) = \frac{a_1}{2\sqrt{-a_2}}. \tag{199}$$

Note that $\sqrt{-a_2}$ is not an imaginary number when $a_1^2 + 4a_2 < 0$ (because a_2 have to be negative to satisfy the inequality).

3.3.3 The General Solution and Initial Conditions

As usual, the general solution is $y_t = y_t^h + y_t^p$. It has a specific form for each of the three cases discussed above. In the solutions below, A_1, A_2, β_1 and β_2 are arbitrary constants.

Case 1 (distinct real roots)
In case 1, the general solution is

$$y_t = \begin{cases} A_1 \left(\frac{a_1 - \sqrt{a_1^2 + 4a_2}}{2} \right)^t + A_2 \left(\frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2} \right)^t + \frac{a_0}{1 - a_1 - a_2}, & (a_1 + a_2 \neq 1) \\ A_1 \left(\frac{a_1 - \sqrt{a_1^2 + 4a_2}}{2} \right)^t + A_2 \left(\frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2} \right)^t + \frac{a_0}{2 - a_1} t, & (a_1 + a_2 = 1, \quad a_1 \neq 2) \\ A_1 \left(\frac{a_1 - \sqrt{a_1^2 + 4a_2}}{2} \right)^t + A_2 \left(\frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2} \right)^t + \frac{a_0}{2} t^2, & (a_1 + a_2 = 1, \quad a_1 = 2) \end{cases} \quad (200)$$

Case 2 (repeated real roots)
In case 2, the general solution is

$$y_t = \begin{cases} A_1 \left(\frac{a_1}{2} \right)^t + A_2 t \left(\frac{a_1}{2} \right)^t + \frac{a_0}{1 - a_1 - a_2}, & (a_1 + a_2 \neq 1) \\ A_1 \left(\frac{a_1}{2} \right)^t + A_2 t \left(\frac{a_1}{2} \right)^t + \frac{a_0}{2 - a_1} t, & (a_1 + a_2 = 1, \quad a_1 \neq 2) \\ A_1 \left(\frac{a_1}{2} \right)^t + A_2 t \left(\frac{a_1}{2} \right)^t + \frac{a_0}{2} t^2, & (a_1 + a_2 = 1, \quad a_1 = 2) \end{cases} \quad (201)$$

Case 3 (complex roots)
In case 3, the general solution is

$$y_t = \begin{cases} \beta_1 r^t cos(\theta t + \beta_2) + \frac{a_0}{1 - a_1 - a_2}, & (a_1 + a_2 \neq 1) \\ \beta_1 r^t cos(\theta t + \beta_2) + \frac{a_0}{2 - a_1} t, & (a_1 + a_2 = 1, \quad a_1 \neq 2) \\ \beta_1 r^t cos(\theta t + \beta_2) + \frac{a_0}{2} t^2, & (a_1 + a_2 = 1, \quad a_1 = 2) \end{cases} \quad (202)$$

In the solutions above, the arbitrary constants are identified using initial conditions. Since there are two arbitrary constants to solve for, we need two initial conditions. Just like in step 4 explained in subsection 3.1.2, we find values for the arbitrary constants using the general solution and initial conditions.

To illustrate how to identify those arbitrary constants, suppose we are given the initial conditions y_0 and y_1 and that we want to identify A_1 and A_2 in case 1 where $a_1 + a_2 \neq 1$. At $t = 0$ and $t = 1$ the general solution becomes

$$y_0 = A_1 + A_2 + \frac{a_0}{1 - a_1 - a_2}, \qquad y_1 = A_1 \alpha_1 + A_2 \alpha_2 + \frac{a_0}{1 - a_1 - a_2} \quad (203)$$

where α_1 and α_2 are given by (193). The latter two equations are linear in A_1 and A_2 and can easily be solved for these two constants. For example, the first equation in (203) can be

rewritten as

$$A_1 = y_0 - A_2 - \frac{a_0}{1 - a_1 - a_2} \tag{204}$$

and this result can be used in the second equation to eliminate A_1. This yields

$$A_2 = \frac{1}{\alpha_2 - \alpha_1} \left[y_1 - \alpha_1 y_0 + \frac{a_0(\alpha_1 - 1)}{1 - a_1 - a_2} \right]. \tag{205}$$

Finally, plugging (205) back into (204) yields

$$A_1 = y_0 - \frac{a_0}{1 - a_1 - a_2} - \frac{1}{\alpha_2 - \alpha_1} \left[y_1 - \alpha_1 y_0 + \frac{a_0(\alpha_1 - 1)}{1 - a_1 - a_2} \right]. \tag{206}$$

3.3.4 Dynamic Stability

As in the case of first-order difference equations, the convergence of the time path of y depends exclusively on whether y^h tends toward zero as t tends to infinity. The condition for dynamic stability closely resembles the condition stated in section 3.1.3.

For all three cases of characteristic roots, the time path will converge to a (stationary or moving) intertemporal equilibrium (regardless of what the initial conditions might be) **if and only if the absolute value of every root is less than 1**.

Exercise 3.3 – A Dynamically Unstable Case

Apply the four-step method to solve the following difference equation

$$y_{t+2} = 1 + 0.5y_{t+1} + 0.6y_t, \qquad y_0 = 1, y_1 = 2.$$

Exercise 3.4 – A Special Case

Apply the four-step method to solve the following difference equation

$$y_{t+2} = 1 + 0.8y_t, \qquad y_0 = 7.$$

3.4   Additional Exercises

Exercise 3.5

A standard neo-classical growth model where physical capital fully depreciates each period yields the following first-order difference equation in the capital stock (K)

$$K_{t+1} = 0.297 K_t^{0.3}.$$

(a) Draw the phase diagram corresponding to the above nonlinear first-order difference equation.

(b) What are the dynamic properties of the time path of capital (convergent/divergent, smooth/oscillatory)?

(c) Using the trial solution $K_t = c$ for all t, solve for the intertemporal equilibrium level of K (a particular solution).

Taking natural logs, we can actually derive a difference equation that is linear in $\ln(K)$

$$\ln(K_{t+1}) = -1.214 + 0.3 \ln(K_t).$$

(d) Solve the (log) linear difference equation above for the time path of $\ln(K)$, given an initial condition K_0.

(e) What is the intertemporal equilibrium level of $\ln(K)$?

(f) Are your answers in (c) and (f) consistent?

Exercise 3.6

Use the initial conditions $y_0 = 10$ to solve the first-order difference equation

$$y_{t+1} = 4 + 0.2 y_t.$$

Exercise 3.7

Use the initial conditions $y_0 = 20$ to solve the first-order difference equation

$$y_{t+1} = 6 + \frac{1}{3} y_t.$$

Is the time path of y convergent? Is it oscillatory?

Exercise 3.8

Use the initial conditions $y_0 = 2$ to solve the first-order difference equation

$$y_{t+1} = 1 + y_t.$$

Is the time path of y convergent? Is it oscillatory?

Exercise 3.9

Use the initial conditions $y_0 = 2$ to solve the first-order difference equation

$$y_{t+1} = 4 + 1.5y_t.$$

Is the time path of y convergent? Is it oscillatory?

Exercise 3.10

Use the initial conditions $y_0 = 4$ to solve the first-order difference equation

$$y_{t+1} = 9 - 2y_t.$$

Is the time path of y convergent? Is it oscillatory?

Exercise 3.11

Consider the first-order difference equation

$$y_{t+1} = 10 + \frac{1}{2}y_t.$$

Find the initial condition (y_0) that is consistent with the solution

$$y_t = 20 + 11 \left(\frac{1}{2}\right)^t.$$

Exercise 3.12

Consider the first-order difference equation

$$y_{t+1} = a_0 + a_1 y_t, \qquad y_0 = -\frac{1}{3}.$$

Using the solution

$$y_t = \frac{32}{3} - 11 \left(\frac{1}{4}\right)^t$$

find the values of a_0 and a_1.

Exercise 3.13

A standard real business cycle model where physical capital fully depreciates each period yields the following stochastic first-order difference equation in the capital stock (K)

$$K_{t+1} = 0.297 e^{\varepsilon_{t+1}} K_t^{0.3}$$

where $\{\varepsilon_t\}$ is a sequence of *i.i.d.* random variables. Use the method of undetermined coefficients to derive an expression for the time path of $\ln(K)$ given an initial condition $K_0 = 1.3$. Is the path explosive?

Exercise 3.14

Consider the money demand function

$$m_t - p_t = \omega - \gamma(p_{t+1}^e - p_t), \qquad \gamma > 0$$

where m_t is the logarithm of the nominal money supply in period t, p_t is the logarithm of the price level in t and p_{t+1}^e is the logarithm of the price level expected in period $t+1$.

To keep things simple, we assume that $m_t = m + \varepsilon_t$ where m is a constant and ε_t is a mean-zero random variable. In addition, we assume that the agents have perfect foresight, which implies that $p_{t+1}^e = p_{t+1}$. Thus, the equation to solve can be written as

$$p_{t+1} - (1 + 1/\gamma)p_t = -(m - \omega)/\gamma - \varepsilon_t/\gamma.$$

(a) Find the homogeneous solution.

(b) Use the method of undetermined coefficients and the forward-looking trial solution

$$p_t = c + \sum_{i=0}^{\infty} \beta_i \varepsilon_{t+i}$$

to find a particular solution. Note that c, β_0, β_1, ... are undetermined coefficients.

(c) Is the particular solution explosive? Would the particular solution be explosive if we had used a backward-looking trial solution?

(d) Is the time path for p convergent?

Exercise 3.15

Use the initial conditions $y_0 = 13$ and $y_1 = 11.3$ to solve the second-order difference equation

$$y_{t+2} = 3 + 0.9y_{t+1} - 0.2y_t.$$

Is the time path of y convergent?

Exercise 3.16

Use the initial conditions $y_0 = 5/4$ and $y_1 = 3$ to solve the second-order difference equation

$$y_{t+2} = 4 - 6y_{t+1} - 9y_t.$$

Is the time path of y convergent?

Exercise 3.17

Use the initial conditions $y_0 = 4$ and $y_1 = 3$ to solve the second-order difference equation

$$y_{t+2} = -2 + 0.2y_{t+1} - 0.01y_t.$$

Is the time path of y convergent?

Exercise 3.18

Use the initial conditions $y_0 = 2$ and $y_1 = 2.5$ to solve the second-order difference equation

$$y_{t+2} = 0.4y_{t+1} + 0.6y_t.$$

Is the time path of y explosive?

Exercise 3.19

Consider the second-order difference equation

$$y_{t+2} = 3 + 0.7y_{t+1} + 0.3y_t.$$

(a) Find the homogeneous, particular and general solutions.

(b) What should the initial conditions y_0 and y_1 be to insure that the time path of y converges to the particular solution despite the existence of a unit root?

Exercise 3.20

Suppose that an overlapping generations model yields the following first-order difference equation in the capital-labour ratio (k)

$$k_{t+1} = \frac{(1-\mu)\beta}{1+\beta}k_t^\mu, \quad 0 < \mu < 1, \quad 0 < \beta < 1, \quad k_0 > 0.$$

(a) Sketch the phase diagram corresponding to the above nonlinear first-order difference equation. Clearly indicate on the graph the intertemporal equilibrium level of k.

(b) What are the dynamic properties of the time path of the capital-labour ratio (convergent/divergent, smooth/oscillatory)?

(c) Take natural logs to obtain a difference equation that is linear in $\ln(k)$.

(d) Solve the log-linear difference equation you derived in (c) for the time path of $\ln(k)$, given an initial condition $k_0 > 0$.

(e) What is the intertemporal equilibrium level of $\ln(k)$?

Exercise 3.21
Consider the following second-order difference equation

$$k_{t+2} = 0.8k_{t+1} + 0.2k_t \tag{1}$$

where the initial conditions $k_0 = 2$ and $k_1 = 1$.

(a) Solve the difference equation (1) to find an expression for the time path of k.

(b) Is the time path of k convergent? Explosive? Oscillatory? Explain.

Exercise 3.22
Consider the linear difference equation $1.5y_{t+1} = 2 + 0.5(y_{t+1} - y_t)$

(a) Solve this difference equation using the initial condition $y_0 = 2.333$.

(b) Is the time path for y implied by your solution convergent? Explain.

(c) Is the time path for y implied by your solution oscillatory? Explain.

3.5   Solutions to Exercises

Solution to Exercise 3.1

The time path for y is described by the equation $y_t = 20 + 5 \times 0.95^t$.

Solution to Exercise 3.2

The time path for y is described by the equation $y_t = 5 + 2t$. If you graph the time path of y, you will notice that y_t does not converge to its equilibrium path $y_t^p = 2t$. This occurs because the condition $|a_1| < 1$ is not satisfied in this exercise .
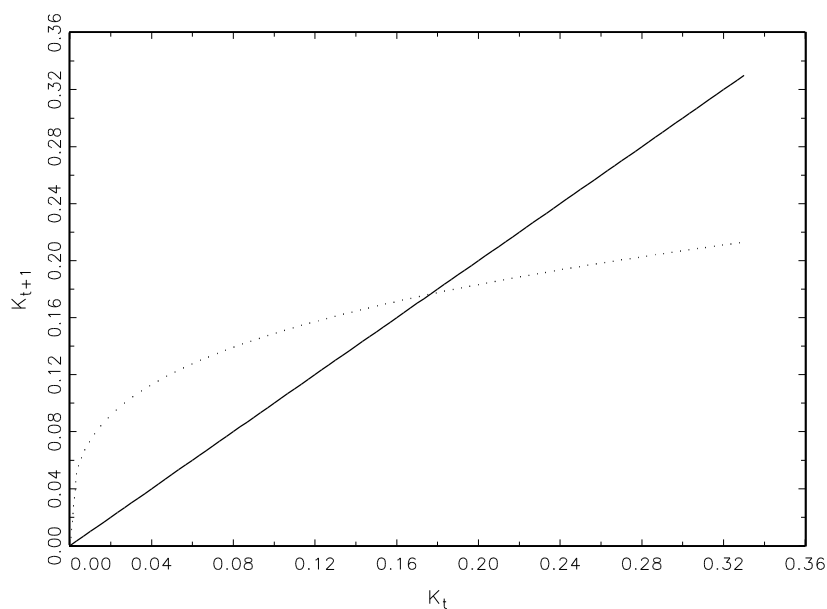
Solution to Exercise 3.3

The time path for y is described by the equation $y_t = -10 - 0.5942 \times 0.5639^t + 11.5942 \times 1.0639^t$.

Solution to Exercise 3.4

You should find that the time path for y is described by the equation $y_t = 5 + 2(\sqrt{0.8})^t$.

Solution to Exercise 3.5

(a)

(b) Smooth and convergent.

(c) The equilibrium level of K is $0.297^{\frac{1}{0.7}} = 0.17652$.

(d) $\ln(K_t) = -1.7343 + [\ln(K_0) + 1.7343] \times 0.3^t$.

(e) The equilibrium level of $\ln(K)$ is given by its particular solution -1.7343.

(f) Yes they are since $\ln(0.17652) = -1.7343$.

Solution to Exercise 3.6

$$y_t = 5 + 5 \times 0.2^t$$

Solution to Exercise 3.7

$$y_t = 9 + 11 \times (1/3)^t$$

Since $a_1 = 1/3 > 0$ the time path of y is not oscillatory. Since $|a_1| < 1$ the time path converges to its equilibrium level (9 in this example).

Solution to Exercise 3.8

$$y_t = 2 + t.$$

Since $a_1 = 1 > 0$ the time path of y is not oscillatory. Since $|a_1|$ is not less than 1 the time path does not converge to its equilibrium (in this example the equilibrium path is $y_t = t$).

Solution to Exercise 3.9

$$y_t = -8 + 10 \times 1.5^t.$$

Since $a_1 = 1.5 > 0$ the time path of y is not oscillatory. Since $|a_1| > 1$ the time path is explosive.

Solution to Exercise 3.10

$$y_t = 3 + (-2)^t.$$

Since $a_1 = -2 < 0$ the time path of y is oscillatory. Since $|a_1| > 1$ the time path is explosive.

Solution to Exercise 3.11

The initial condition must be $y_0 = 31$.

Solution to Exercise 3.12

By inspecting the solution, we immediately have that $a_1 = 1/4$. Using the fact that the particular solution is $32/3$ and that $a_1 = 1/4$ we find $a_0 = 8$.

Solution to Exercise 3.13

$$\ln(K_t) = \left[2 - \sum_{i=0}^{\infty} 0.3^i \varepsilon_{-i}\right] 0.3^t - 1.7343 + \sum_{i=0}^{\infty} 0.3^i \varepsilon_{t-i}.$$

The path is not explosive since $|0.3| < 1$.

Solution to Exercise 3.14

(a) $y_t^h = A(1 + 1/\gamma)^t$.

(b)

$$c = m - \omega, \quad \beta_0 = \frac{1}{1+\gamma}, \quad \beta_1 = \frac{\gamma}{(1+\gamma)^2}, \quad \beta_2 = \frac{\gamma^2}{(1+\gamma)^3}, \ldots, \quad \beta_i = \frac{\gamma^i}{(1+\gamma)^{i+1}}$$

which implies the particular solution

$$y_t^p = m - \omega + \sum_{i=0}^{\infty} \frac{\gamma^i}{(1+\gamma)^{i+1}} \varepsilon_{t+i} = m - \omega + \frac{1}{\gamma} \sum_{i=0}^{\infty} \left(\frac{\gamma}{1+\gamma}\right)^{i+1} \varepsilon_{t+i}.$$

(c) The forward-looking particular solution is not explosive since $\gamma > 0$ implies that $0 < \gamma/(1+\gamma) < 1$. The backward-looking solution would have been explosive.

(d) The time path for p predicted by the general solution

$$y_t = A(1 + 1/\gamma)^t + m - \omega + \frac{1}{\gamma} \sum_{i=0}^{\infty} \left(\frac{\gamma}{1+\gamma}\right)^{i+1} \varepsilon_{t+i}$$

will be explosive in general since $1 + 1/\gamma > 1$ implies that $(1 + 1/\gamma)^t$ tend to infinity as $t \to \infty$. However, if the initial condition is such that $A = 0$, then the time path for p is convergent.

Solution to Exercise 3.15

$$y_t = 10 + 2 \times 0.4^t + 0.5^t.$$

Both roots are less than one in absolute value, therefore the time path of y is convergent.

Solution to Exercise 3.16

$$y_t = \frac{1}{4} + (-3)^t - 1.9167t(-3)^t$$

The time path of y is not convergent since $|-3| > 1$.

Solution to Exercise 3.17

$$y_t = -2.4691 + 6.4691 \times 0.1^t + 48.2222t \times 0.1^t$$

The time path of y is convergent since $|0.1| < 1$.

Solution to Exercise 3.18

$$y = 2.3125 - 0.3125(-0.6)^t$$

The path is not explosive. As $t \to \infty$, y_t will approach 2.3125.

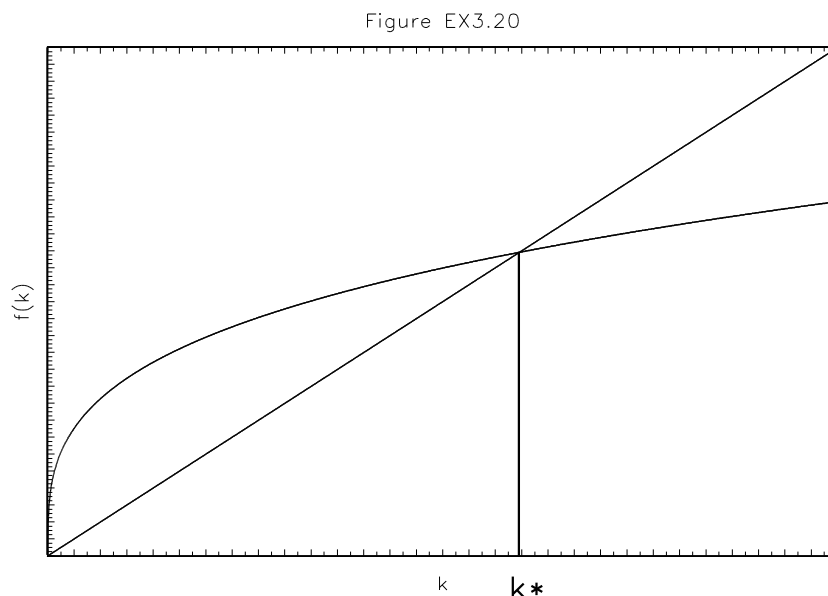**Solution to Exercise 3.19**

(a)

$$y_t^h = A_1(-0.3)^t + A_2 \qquad y_t^p = 2.308t.$$

$$y_t = 2.308t + A_1(-0.3)^t + A_2.$$

(b) The initial conditions must be $y_0 = 0$ and $y_1 = 2.308$.

Solution to Exercise 3.20

(a) Let $f(k) \equiv (1 - \mu)\beta(1 + \beta)^{-1} k^\mu$. Since $f'(k) > 0$ for $k > 0$ and $f''(k) < 0$ for $k > 0$, the function $f(k)$ is concave. Moreover, the Inada conditions indicate that the slope of the function is infinitely steep when $k \to 0$ and is zero when $k \to \infty$.

Figure EX3.20

(b) Using phase diagrams, we can easily show that the time path of k is smooth and convergent for any $k_0 > 0$.

(c) To simplify notation, define $\tilde{k}_t \equiv \ln k_t$ and $a_0 = \ln([1-\mu]\beta[1+\beta]^{-1})$. Taking logs yields $\tilde{k}_{t+1} = a_0 + \mu\tilde{k}_t$.

(d) Definite solution:

$$\tilde{k}_t = \frac{a_0}{1-\mu} + \left[\tilde{k}_0 - \frac{a_0}{1-\mu}\right]\mu^t$$

(e) The intertemporal equilibrium level of $\ln(k)$ is given by the particular solution $a_0/(1-\mu)$.

Solution to Exercise 3.21

(a) $k_t = 1.1667 + 0.8333(-0.2)^t$.

(b) We need both roots less than unity in absolute value for the time path of k to converge to its intertemporal equilibrium (zero in the current case). This condition is not satisfied since one of the roots equals unity exactly.

The time path will not be explosive since $\lim(-0.2)^t = 0$ as $t \to \infty$.

The path is occilatory since there is a negative root.

Solution to Exercise 3.22

(a) $y_t = 1.333 + (-0.5)^t$

(b) Yes, as indicated by the root test: $|\alpha| = |-0.5| = 0.5 < 1$.

(c) Yes, because the root -0.5 is negative.

4 References

Chiang, A.C., *Fundamental Methods of Mathematical Economics*, Third Edition, 1984, McGraw-Hill.

Davidson, R. and J.G. MacKinnon, *Estimation and Inference in Econometrics*, 1993, Oxford University Press.

Enders, W., *Applied Econometric Time Series*, 1995, John Wiley and Sons.

Hill, R.C., Griffiths, W.E. and G.G. Judge, *Undergraduate Econometrics*, Second Edition, 2001, John Wiley and Sons.

Smith, G.W., *MAcroeconomics Lecture Notes*, 1999, mimeo Queen's University.